UACE2025

CONFERENCE PROCEEDINGS

ISSN: 2408-0195

Edited by
Michael I. Taroudakis

UNDERWATER ACOUSTICS conference & exhibition series

HALKIDIKI, GREECE | 15-20.06.2025



8th Underwater Acoustics Conference and Exhibition

UACE2025

15-20 June, 2025 | Halkidiki, Greece

Conference Proceedings

Edited by

Michael I. Taroudakis

University of Crete and Foundation for Research and Technology-Hellas

Heraklion, Crete, Greece

ISSN: 2408-0195

Table of Contents

Preface xiii
Organising Committee xiv
Scientific Committee xv
Session Organisers xvi

Structured Session 01. Acoustics in Ocean Observation Systems 1

Machine/Deep Learning categorisation of sub-kilohertz Arctic soundscapes 3
Jonathan Cleverly, Philippe Blondel, Hanne Sagen, Espen Storheim, Matthew Dzieciuch

Change Detection in Synthetic Aperture Sonar Imagery Using Segment Anything Model 11
William Hedlund, Per Abrahamsson, Louise Fuchs

Structured Session 02. Acoustics in polar environments 19

Variability of acoustic signals near a Polar front at the Western Barents Sea 21
Altan Turgut, Jeffrey Schindall

Climate-driven variability of underwater noise propagation in Hornsund fjord, Svalbard 29
Pavani Vithana Madugeta Vidanamesthrige, Natalia Gorska, Oskar Glowacki

Structured Session 03. Advances in acoustic measurement systems: Technologies and applications 35

Insights and Feedback from a Long-Term Deployment of Automated Acoustic Monitoring Systems for Marine Mammal Detection 37
Samuel Antoine

Structured Session 04. Comprehensive Nuclear-Test-Ban Treaty Monitoring, and its Civil and Scientific Applications 43

Deep-ocean soundscapes: The impact of atmospheric carbon dioxide 45
Gianluca Audone, Philippe Blondel, Matthew Nunes, Chris Budd Obe

Monitoring underwater eruptions using networks of hydrophones in the SOFAR channel 53
Sara Bazin, Pierre-Yves Raumer, Aude Lavayssière, Hugo Fauvel, Romain Safran, Vaibhav Vijay Ingale, Jonathan Tanrin, Julien Vergoz, Jean-Yves Royer, Jun Korenaga

Structured Session 05. Distributed Fiber-Optic Sensing technology for underwater acoustical monitoring 59

Analysis of passive DAS data for observation of microseisms and extracting Scholte wave dispersion in shallow water 61
Hefeng Dong, Michael Taroudakis

Structured Session 06. Enhancing underwater acoustic sensing through machine learning 67

Neural Network-Informed Acoustics Sensors for Optimal Coherent Processing: Benchmarking in a Forward-Modelled Fluctuating Ocean Environment 69
Antoine Blachet, Xavier Cristol, Thomas Mahiout, Dominique Fattaccioli, Gaultier Real

The urgent call on data management: are we capable to store valuable (meta)data for naval application? 77
Sonia Papili

Power and accuracy trade-offs for machine learning methods applied to detection of underwater sound sources 83
William Butler, Harrison Smith, Marios Impraimakis, Andrew Barnes, Alan Hunter

An Investigation of Machine Learning Capabilities for Cavitation Detection 91
Dale Smith, Oscar Carter

Structured Session 07. Inverse Problems in Acoustical Oceanography 97

Underwater Passive Source Localization based on a Data Derived Replica Library 99
Baptiste Menetrier, Gnouregma Bazile Kinda, Valérie Labat, Samuel Pinson, Abdel Boudraa

Sequential Measurement Point Selection For Geoacoustic Inversion Using Unscented Kalman Filtering 105
Siobhán Correnty, Ylva Ljungberg Rydin, Magnus Lundberg Nordenvad, Isaac Skog, Martin Östberg

Range-Dependent Source Localization Methods in Shallow-Water environments 111
Jing Guo, Juan Zeng, Li Ma

Structured Session 08. Bioacoustics and Soundscape 133

Automated Classification of Vocalisations from Wild and Captive Seal Populations 135
William Butler, Harrison Smith, Wil Lloyd, Marios Impraimakis, Andrew Barnes, Alan Hunter

Advancing Marine Monitoring Through Edge-Optimised Deep Learning: Hydrotwin's Approach to Underwater Acoustic Surveillance 143
Astrid van Toor, Erica Cruz, Guilherme Beleza Vaz

North Atlantic Right Whale Up-call Localization with a Four-Element Acoustic Array on a Slocum Glider 153
Michael MacGillivray, Mae Seto, S. Bruce Martin, Laird Bolt

Correlation of NARW Vocalizations with Behaviours based on Analysis of Tagged Whale Data towards more Informed Localization 159
Marina Antipina, Mae Seto, Jay Kirkham, Sarah Fortune

Structured Session 09. Modeling techniques for underwater acoustic scattering and propagation (including 3D effects) 165

A Machine Learning Approach for Model Selection in Underwater Acoustic Propagation 167
Finn Henman, Stewart Haslinger, Joshua Wakefield

RAINDROP: A Framework for Real-Time Holistic Underwater Acoustic Maps 175
Tiago Gomes, Guilherme Vaz

Measurements of sea-floor backscattering strength of very low frequency acoustic waves in western Mediterranean Sea; a comparison to GABIM [1] 185
Iannis Bennaceur, Nathan Ivkovic, Xavier Cristol, Adrien Rondeau, Benoit Theckes

Time-frequency structure of normal modes in range-dependent waveguides: numerical modeling and field experiment 191
Danila Sidorov, Devi Bodjona, Dmitrii Kovzel, Vladimir Bezotvetnykh, Andrey Lunkov

Sound scattering by nested cylindrically symmetric and linearly invariant anomalies 197
Sven Ivansson

Structured Session 10. New frontiers in fisheries acoustics: applications, analysis, decision and management tools 205

In situ measurements of Lateral target strength (TS) of Atlantic bluefin tuna (Thunnus thynnus) from purse seiners 207
Vicente Puig-Pons, Sergio Morell-Monzó, Hector Peña, Isabel Pérez-Arjona, Victor Espinosa

Advances in biomass estimation of Nile tilapia aquaculture with acoustic methods 213
Luis Lorenzo Carrillo La Rosa, Vicente Puig-Pons, Sergio Morell-Monzó, Isabel Pérez-Arjona, Víctor Espinosa

Characterisation of aggregated Baltic herring – technique for assessment of possible vessel avoidance reaction 219
Aleksander Żytko, Natalia Gorska, Dezhang Chu, Beata Schmidt

Using open FEM/BEM software for modeling fish target strength 225
Marek Moszyński, Isabel Pérez Arjona, Víctor Espinosa Roselló, Anderson Ladino Velasquez

Improving Fish School Detection Using Phase Information in Continuous-Wave Underwater Acoustic Signals: A Study Based on Cylindrical Sonar Data 231
Yohann Gourret, Karl Thomas Hjelmervik, Tommi Brander, Hector Pena, Henrik Berg, Bjarte Berntsen

Structured Session 12. Observing the Oceans Acoustically 237

A Submodular Approach to Acoustic Sensor Placement using the MARLIN Digital Twin 239
Edward Clark, Eric Schoof, Alan Hunter, Chris Manzie

Structured Session 13. Pile driving noise -Radiated noise from offshore wind construction 247

Evaluating the Applicability of Scaling Laws for Mitigated Pile Driving Noise to Highly Frequency Dependent Mitigation Systems 249
Jonas von Pein, Matthias Ram

Underwater noise in offshore impact piling including pile-soil contact relaxation 259
Khairina Anindya Canny, Yaxi Peng, Athanasios Tsetas, Apostolos Tsouvalas

Acoustic emissions from offshore monopile installation using jetting technology 267
Matej Simurda, Areti Ioanna Tsigkinopoulou

The impact of pile driving noise on North Atlantic Right Whale detection using coherently beamformed hydrophone arrays 275
Vincent E. Premus, Phil Abbot, Ted Abbot, Eric Illich, John Browning, Vitaly Kmelnitsky, Jake Friese, Aaron Logan

Structured Session 15. Seabed Acoustics 283

Scattering and reflection from soft gassy sediments: Geoacoustic modeling approach 285
Anatoliy Ivakin

Structured Session 16. Sonar performance modeling and verification: Active and passive sonar 293

The future evolution of mid-frequency active sonar performance in a changing ocean 295
Xavier Cristol, Bruno Chalindar, Nicolas Fisichella, Herald Rabeson

Comparisons Between Code- and Time-Division-Multiplexing for MIMO Sonar Signal Processing 301
Michaela Murzyniec, Yannik Steiniger, Sven Schröder

Decoupling Environmental Complexity: Direct Analysis in Sonar Performance Space 309
Jonas Halse Rygh, Karl Thomas Hjelmervik

Mapping the Impact of Climate Change on Indicators of Acoustic Propagation in the Atlantic Ocean 317
Victor Oppeneer, Nolwenn Risser, Aniello Russo, Giacomo Giorli, Robbert van Vossen

J-divergence detection currency after thresholding a Rician signal in Gaussian noise 325
Douglas Abraham

Structured Session 17. Target Echo Strength – Measurements and Modelling 333

Target echo strength of layered media with an active surface 335
Kevin Hostombe, Delf Sachau

TES Modelling of TESSMEX Object M32 and M32T 343
Elizabeth Chilvers

Porting a ray tracing method for determining the target echo strength (TES) of underwater objects to GPUs - a field report 349
Ralf Burgschweiger, Ingo Schäfer, Delf Sachau, Jan Ehrlich

Structured Session 18. Towards Automatic Target Recognition. Detection, Classification and Modelling 357

Development of an airborne circular sonar for research and teaching 359
Blair Bonnett, Kaushikk V. N., Sudhanshu Apte, Thomas Fickenscher

Current state of the openSTB sonar simulator 367
Blair Bonnett, Holger Schmaljohann, Sudhanshu Apte, Thomas Fickenscher

A First Comparison of GANs and Diffusion Models for Generating Sidescan Sonar Images 375
Yannik Steiniger, Benjamin Lehmann

Transformers and Ensembles for Object Detection in Sidescan Sonar Images 385
Yannik Steiniger

Structured Session 19. Uncertainty quantification and machine learning in signal processing 393

Source Localization in Shallow Water Using a Neural Network Based on Range-Dependent Sound Speed Profile Model 395
Jing Guo, Juan Zeng, Li Ma

Structured Session 20. Underwater acoustic calibration, measurement, and standards 413

Analysis of signal processing methods applied in wide band free-field calibrations of hydrophones 415
Guanghui Jia, Yi Chen, Liuqing Yang

Calibration and its correction method for hydrophones in the air chamber 423
Liuqing Yang, Yi Chen, Boyue Wu

The Effect of Temperature on Acoustic Transducer Sensitivity and Transmitting Response 429
Caitlin Kean, William Slater

Application of 2E-2U Method for Free-Field Underwater Acoustic Calibration of Hydrophones and Projectors in a Reverberant Laboratory Test Tank 439
Ata Can Corakci, Alper Biber, Talha Saydam, Osman Said Biskin, Serkan Aksoy

Commercial Broad Range Calibration by Multi-method Synthesis 447
Nicolas Wood

Structured Session 21. Underwater Communications and Networking 457

A Novel GRU-CNN Model for Self-Interference Cancellation in IBFD Underwater Communication Systems 459
Manar Alhamnday, Bilal Jebur, Charalampos Tsimenidis, Shahid Mumtaz

Robust channel comparison metrics for underwater physical layer authentication 471
Davide Eccher, Paolo Casari

A Lightweight AI-powered Framework for Multimodal Underwater Networks 479
Filippo Donegà, Filippo Bragato, Filippo Campagnaro, Michele Zorzi

Error Correction Code for FSCM-based Acoustic Underwater Communication 487
Fabian Steinmetz, Julia Wüpperling, Bernd-Christian Renner

Target detection in an underwater acoustic network: Lake experiment 497
Lu Shen, Yuriy Zakharov, Benjamin Henson, Nils Morozs, Benjamin Sherlock, Paul Mitchell

Structured Session 22. Underwater noise in the EU Marine Strategy Framework Directive: implementation, monitoring, assessment and measures 505

Long-term measurements of ambient sounds in Cambridge Bay (Canada), 2015–2024 – Implications for extending the MSFD to Arctic waters 507
Philippe Blondel, Rhys Belcher, Dylan Cooper

A BEM model for acoustic radiation due to vibrating underwater vehicle in the sea acoustic waveguide 515
John Prospathopoulos, Kostas Belibassakis

Structured Session 23. Underwater noise modelling and measurements 523

Near-Field Testing and Evaluation Method for Stream Noise of Linear Array Based on Laser Doppler Velocimetry 525
Zeng Cao, Xiaojian Wu, Yi Chen

A deep-sea ambient noise prediction method with few training samples based on VMD and LSTM 533
Guoli Song, Xinyi Guo, Qianchu Zhang, Bo Yuan, Qunyan Ren, Li Ma

Structured Session 25. Unmanned vehicles for underwater acoustic surveillance and monitoring 539

Autonomous Underwater Inspection of Inland Waterways with a Small Uncrewed Surface Vessel for Police and Emergency Response 541
Alfie Anthony Treloar, Zuhayr Rymansaib, Alan Hunter

Reinforcement Learning for Marine Path Planning 547
Edward Clark, Alfie Anthony Treloar, Alan Hunter, Marcus Donnelly

Acoustic Surveillance Using an Unmanned Blended Wing Glider 555
Scott Mahar, Gerald D'Spain

Concept demonstration of an ASW barrier using low-cost unmanned systems 565
Jeroen van de Sande, Peter Beerens, Koen Blom, Mathieu Colin, Martijn van Riet, Otto Koetsier

Regular Session 27. Radiated noise from Ships and Surface Platforms 573

Soundscape modelling in a marine area designated for floating offshore wind turbine (FOWT) development in northern Portugal 575
Marcos Ricardo Souza, Tiago Gomes, Guilherme Vaz

The impact of electric propulsion systems on underwater radiated noise 585
Andrea Grech La Rosa, Tom Smith, Peng Wu

Cavitation on small craft propellers and its contribution to underwater radiated noise 593
Tom Smith, Andrea Grech La Rosa, Solomon McMorran, Giles Piggott, Joao Gaivota

Regular Session 28. Signal and Image Processing 601

Hierarchical Multi-Target Tracking in Underwater Environments 603
Joshua Wakefield, Stewart Haslinger, Jason Ralph

Dynamic region method for improving adaptive quantization during sonar data transmission 611
Hyunjae Song, Minho Kim, Youngsub Lim, Youngshin Kim

Regular Session 29. Underwater Communication Navigation and Positioning 617

Uplink multi-user underwater acoustic orthogonal frequency division multiplexing communication based on vertical vector sensor array 619
Wei Men, Bowen Dong, Lu Zhang, Yanan Tian

Real-Time Underwater Localization System Based on Position-Keeping USVs for Multiple AUVs 625
Youngsun Moon, Minho Kim, Gwangho Moon, Seokjeong Yong, Myungjun Kim

Regular Session 30. Underwater Sensors Technology 631

Determination of Matched Filter Broadband Beampattern for Arbitrary Antenna Geometries 633
Christoph Zimmer, Johannes Groen, Dieter Kraus, Nils Theuerkauf, Anton Kummert

Quantum Spin States in Water: Ortho-Para Transition Theory and Its Impact on Environmental and Industrial Technologies 639A
S.M. Pershin, I. Bjørnø, M.Ya. Grishin

Poster session 639

Distance estimation of near-surface moving targets by deep sonar 641
Xingyuan Pei, XiaoChuan Ma, Chao Feng, Yu Liu, Xiaomei Wang

Clutter Reduction in Active Sonar Using a Clustering Method 647
MyoungJun Cheong, Juho Kim, Hyeon-Deok Cho

Authors' Index 653

Preface

The Underwater Acoustics Conference and Exhibition 2025-UACE2025 was held with great success in Halkidiki, Macedonia, Greece 15-20 June 2025. The conference topics covered all areas of underwater acoustics, and the attendance was exceptionally high as the conference participants were more than 260 from 27 countries.

On behalf of the Conference Chairmen and the Organizing Committee of UACE 2025, I have the pleasure to present to the scientific community the volume of the Conference Proceedings which is available freely accessible through the Conference website. Note that UACE Proceedings have been registered under SCOPUS since 2023.

The papers that have been submitted for inclusion to the Conference Proceedings reflect the high quality of the presentations and I hope that they will constitute important references for on-going and future research work in underwater acoustics

On the occasion of the release of the UACE 2025 proceedings, and on behalf of the Conference chairmen and the Organizing Committee, I wish to express once more my sincere appreciation to all scientists who have contributed to the success of the Conference. These include in particular, the members of the Scientific Committee of UACE2025, the Structured Session organizers and of course all speakers and participants. I am most grateful to all of them, and I hope that they will continue their support to the UACE series of Conferences.

Finally, I would like to thank the Office of Naval Research Global for their continuous support and financial contributions throughout the history of the UACE Conferences.

A special thank you is due to IACM-FORTH, for the support to the Conference, the UACE technical team members Yiota Rigopoulou and Yorgis Androulakis for their valuable contribution to all the organizational details of the conference, Dr. Panagiotis Papadakis for his assistance in the organization of the awards' contests, as well as CCBS Greece which assists us in the organization of all UACE Conferences.

Michael Taroudakis
Professor Emeritus
Conference co-chairman

Organising Committee

- David Bradley, ARL, Penn State University, USA (Co-chairman)
- Ross Chapman, University of Victoria, Canada (Co-chairman)
- Michael Taroudakis, University of Crete & FORTH, Greece (Co-chairman)
- Philippe Blondel, University of Bath, UK
- Mike Buckingham, MPL, Scripps, USA
- Nick Chotiros, University of Texas at Austin, USA
- George Haralabus, Formerly with CTBTO; incoming Programme Manager at CMRE (Centre for Maritime Research and Experimentation).
- Anthony Lyons, CCOM/JHC, University of New Hampshire, USA
- Qihu Li, Institute of Acoustics, Acad. Sinica, China

Technical Manager

- Yorgis Androulakis, IACM - FORTH

Secretariat

- Yiota Rigopoulou, IACM – FORTH

Partner Agency in Organization

- Cretan Conference and Business Services Greece (CCBS Greece)

Supported by

- Office of Naval Research (ONR)
- IACM-FORTH

Scientific Committee

- David Bradley, ARL, Penn State University, USA (Co-chairman)
- Ross Chapman, University of Victoria, Canada (Co-chairman)
- Michael Taroudakis, University of Crete & FORTH, Greece (Co-chairman)
- Michael Ainslie, Jasco Applied Sciences, Germany
- Philippe Blondel, University of Bath, UK
- Julien Bonnel, Woods Hole Oceanographic Institute, USA
- Mike Buckingham, MPL, Scripps, USA
- Andrea Caiti, ISME/DII, Italy
- Chi-Fang Chen, National Taiwan University, Taiwan
- Nick Chotiros, University of Texas at Austin, USA
- Paul Cristini, University Aix Marseille, France
- Stan Dosso, University of Victoria, Canada
- Victor Espinosa, Polytechnic University of Valencia, Spain
- Alexander Gavrilov, Curtin University, Australia
- Peter Gerstoft, MPL, Scripps, USA
- Roy Edgar Hansen, NDRE, Norway
- George Haralabus, Formerly with CTBTO; incoming Programme Manager at CMRE (Centre for Maritime Research and Experimentation).
- Sergio Jesus, Univ. Algarve, Faro, Portugal
- Tony Lyons, University of New Hampshire, USA
- Qihu Li, Institute of Acoustics, Acad. Sinica, China
- Alain Maguer, CMRE, Italy
- Zoi-Heleni Michalopoulou, NJIT, USA
- Jennifer Miksis-Olds, University of New Hampshire, USA
- Konstantinos Pelekanakis, TNO, Netherlands
- Pavel Petrov, V.I.Il`ichev Pacific Oceanological Institute, Russia
- Purnima Ratilal, Northeastern Univ., Boston, USA
- Jean-Pierre Sessarego, CNRS-LMA, Marseille, France
- Martin Siderius, Portland State University, USA
- Emmanuel Skarsoulis, FORTH, Greece
- Alessandra Tesei, CMRE, Italy

Session Organizers

- Megan Ballard, University of Texas at Austin
- David Barclay, Dalhousie University
- Philippe Blondel, University of Bath, UK
- Julien Bonnel, Woods Hole Oceanographic Institution
- Cheng Chi, Institute of Acoustics, Chinese Academy of Science
- Nick Chotiros, University of Texas at Austin
- Mathieu Colin, TNO, The Netherlands
- Paul Cristini, CNRS HDR
- Hefeng Dong, Norwegian University of Science and Technology
- Stan Dosso, University of Victoria
- Victor Espinosa, Universitat Politècnica de València
- Victor Evora, Naval Undersea Warfare Center Newport
- Nathalie Favretto-Cristini, CNRS HDR
- Alexander Gavrilov, Curtin University, Australia
- Kay Gemba, Naval Postgraduate School, USA
- Giacomo Giorli, Centre for Maritime Research and Experimentation, NATO-STO
- Oskar Glowacki, Polish Academy of Sciences
- Johannes Groen, Atlas Elektronik, GmbH
- Roy Edgar Hansen, Norwegian Defence Research Establishment and University of Oslo
- Georgios Haralabus, Formerly with CTBTO; incoming Programme Manager at CMRE (Centre for Maritime Research and Experimentation).
- Kevin Heaney, OASIS Inc, USA
- Bruce Howe, School of Ocean and Earth Science and Technology (SOEST), University of Hawaii at Mānoa
- Boris Katsnelson, University of Haifa
- Kevin Lee, Univ. Texas at Austin, USA
- Alain Maguer, Centre for Maritime Research and Experimentation, NATO-STO
- Zoi-Heleni Michalopoulou, New Jersey Institute of Technology

- Jennifer Miksis-Olds, UNH Earth, Oceans, & Space
- Paul Mitchell, University of York, UK
- David Nunn, Dstl Porton Down, UK
- Tiago Oliveira, Centre for Maritime Research and Experimentation, NATO-STO
- Yan Pailhas, Centre for Maritime Research and Experimentation, NATO-STO
- Jonas von Pein, DHI
- Konstantinos Pelekanakis, Underwater Operations & Seabed Security, TNO
- Pavel S. Petrov, Instituto de Matemática Pura e Aplicada
- Vincent E. Premus, Ocean Acoustical Services and Instrumentation Systems (OASIS)
- Mark Prior, CTBTO
- Aristides Prospathopoulos, Hellenic Centre for Marine Research
- Purnima Ratilal, Northeastern Univ., Boston, USA
- Stephen Robinson, Acoustics Group, National Physical Laboratory Teddington, UK
- Hanne Sagen, Nansen Environmental and Remote Sensing Center
- Matthew Schinault, Northeastern University
- Evgenii Sidenko, Centre for Marine Science and Technology, Curtin University
- Martin Siderius, Portland State University
- Peter Sigray, Royal Institute of Technology (KTH)
- Emmanuel Skarsoulis, IACM-FORTH, Greece
- Will Slater, Naval Undersea Warfare Center
- Espen Storheim, Nansen Environmental and Remote Sensing Center (NERSC)
- Jaroslaw Tegowski, Institute of Oceanography, University of Gdansk, Poland
- Alessandra Tesei, NATO - STO CMRE, Italy
- Charalampos Tsimenidis, Nottingham Trent University
- Lora van Uffelen, University of Rhode Island
- Peng Wang, Institute of Acoustics, Chinese Academy of Sciences
- Narada Warakagoda, Forsvarets Forskningsinstitutt | FFI
- Angeliki Xenaki, IACM-FORTH, Greece

- Mario Zampolli, CTBTO
- Pengfei Zhang, Institute of Acoustics, Chinese Academy of Sciences

Structured Session
01. Acoustics in Ocean Observation Systems

Organised by Jaroslaw Tegowski, Philippe Blondel and Hanne Sagen

Machine/Deep Learning categorisation of sub-kilohertz Arctic soundscapes

Jonathan Cleverly[1], Philippe Blondel[1], Hanne Sagen[2], Espen Storheim[2], and Matthew Dzieciuch[3]

[1]University of Bath, Bath, UK
[2]Nansen Environmental and Remote Sensing Center, Bergen, Norway
[3]Scripps Institution of Oceanography, San Diego, USA

Contact: Jonathan Cleverly, Department of Physics, University of Bath, Claverton Down, Bath, United Kingdom, BA2 7AY, jgc60@bath.ac.uk

Abstract: *Arctic soundscapes are being modified by climate change, which is greatly amplified in the region. Cryophony (sounds from sea ice processes) will become more variable as ice floes become increasingly fragmented. These changes to the sea ice will also result in shifts in temporospatial patterns of marine mammal vocalisations and anthropogenic sounds. These markers of the state of the Arctic Ocean are monitored using passive acoustic technologies, however there are still no standard practices for exploring soundscapes in this region. Here we investigate Machine/Deep Learning (ML/DL) approaches for categorising deep-water Arctic soundscapes. Recordings from hydrophone moorings deployed along the Nansen Basin during the "Coordinated Arctic Acoustic Thermometry Experiment" (CAATEX, 2019-2020) have been considered for this study. We utilise AVES (Animal Vocalisation Encoder based on Self-Supervision) to identify sounds within recordings with broad descriptors. Training datasets for ML/DL algorithms usually consider a broader frequency range (beyond 20 kHz), but it is not always feasible to use these higher sampling rates, thus the robustness of algorithms requires testing with lower sample rate data (976 Hz here), where the frequency content of sounds is not always fully recorded. To study multiple sound sources at once (e.g. whale songs, anthropogenic sounds), we consider longer context windows ('snippets') of 120 seconds, currently seldom considered in acoustic ML problems. These techniques will be crucial for avoiding current bottlenecks in data processing, in particular in the Arctic, enabling more in-depth studies for marine mammal conservation and industrial regulation.*

Keywords: *Arctic, Nansen Basin, Soundscape Analysis, Machine Learning, Remote Sensing*

1. INTRODUCTION

The study of soundscapes is a crucial part of marine ecosystem management [1]. However, methodologies for assessing biodiversity and human impacts in temperate waters are not necessarily applicable to polar waters. Cryophony leads to significant variability in soundscapes [2], raising issues in using traditional methods, such as comparing third-octave bands.

The Nansen Basin of the Arctic Ocean remains understudied with regard to soundscapes. While under-ice hydrophone moorings have been successfully deployed in the region [3], there are still no universally effective strategies to automatically identify and separate features in recordings. ML/DL techniques have grown in popularity for detecting specific sounds within datasets, but these require large amounts of labelled data, which can be time-consuming to produce. Arctic marine mammal vocalisations and tones from sea ice can vary significantly in frequency and duration. Additionally, the acoustic 'surface duct' of ice-covered waters results in sounds warping and attenuating greatly during propagation. Hence, it is perhaps more suitable to first study soundscapes with a holistic approach, rather than searching for specific time-frequency signatures.

Here, we have considered a pretrained ML model, AVES [4], to process recordings taken during CAATEX [5] and categorise individual snippets of recordings by broad features, such as anthropophony, biophony and cryophony, then by a second set of category-specific descriptors. Locations of hydrophone moorings used throughout CAATEX are shown in Figure 1, as well as sea ice area fraction in April 2020 across the region. By developing a robust soundscape classifier, larger networks of hydrophones deployed on moorings (e.g. [6]) can be processed efficiently and automatically, freeing up valuable time for more focussed, source-specific research.

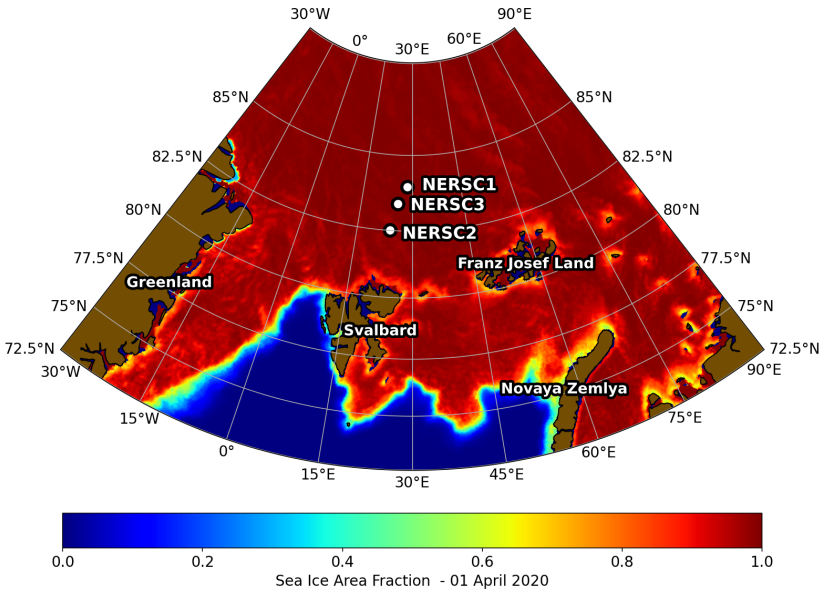


Figure 1: Location of moorings used during CAATEX. Sea ice area fraction in April 2020 taken from the Copernicus programme [7,8]. Map created using Cartopy [9].

2. DATA AND METHODOLOGY

2.1. SOUND PROCESSING AND ANNOTATION PROCESS

CAATEX recordings were taken between September 2019 and September 2020, every 12 hours for 45 minutes and 12 seconds, with a sample rate of 976 Hz. Only recordings from the shallowest hydrophone on each mooring were considered here, with depths ranging between 55-85 m. Recordings were truncated into 120-s snippets for input into AVES, often encompassing several acoustic sources. Partial annotation of the NERSC2 dataset was completed manually, selecting 3,000 snippets at random and labelling them by the following categories: 'Anthro' (anthropophony, i.e. sounds of human origins), 'Bio' (biophony, i.e. sounds associated with marine life, like whale vocalisations), 'Cryo' (cryophony, i.e. sounds associated with ice processes), 'Geo' (geophony, i.e. earthquakes) and 'Noise' (everything else). A second tier of annotations was completed to further categorise soundscape features, with biophonic activity split into five further categories, whereas anthropophonic and cryophonic activities were split into two categories each, transients and tones, for the 3,000 annotations (Table 1).

For manual inspection, spectrograms were computed with a segment length of 0.5 s with 50% overlap and Hann window convolution. Throughout the CAATEX deployment, a thermometry signal was transmitted every 3 days, dominating the soundscape for the first 18 minutes of recording time, so snippets covering this time were omitted from the study.

Category	Sub-category	Description	No. labels
Anthro	0	Consistent tones, e.g. shipping	209
	1	Impulsive sounds, e.g. seismic airguns (in concurrence with tones)	54
Bio	0	Bearded seal mating calls (distinctive trilled downsweeps)	308
	1	Repetitive downsweeps, terminating above 300 Hz	495
	2	Low frequency tones, below 100 Hz	53
	3	Mixed, few non-overlapping calls	637
	4	Mixed, many overlapping calls	213
Cryo	0	Impulsive ice cracking	283
	1	Tones from ice shearing or hums	59
Geo	0	Seismic activity	21
Noise	0	Absence of any identifiable or discernible signals	668

Table 1: Summary of the sound categories used in this study, with their respective numbers.

2.2. IMPLEMENTATION OF AVES

AVES is a transformer-based ML algorithm built primarily for encoding animal vocalisations [4], but here we utilise it for categorising sounds of man-made and geophysical origin as well. Based on the HuBERT model for human speech [10], AVES is pretrained on the large unlabelled

audio datasets: AudioSet, FSD50K and VGGSound, covering a wide-range of sounds, both in air and water, and processing their waveform representations.

A recent *torchaudio*-based version of AVES has been released as a standalone Python package, *esp-aves*, and it has been used for this study [11]. Example configurations for AVES are available online and we selected *aves-base-all.torchaudio* ('AVES-all') as it had been trained on all segments segments in the pretraining datasets.

The *AVES-all* model consists in a 12-layer Convolutional Neural Network encoder with a 768-unit transformer for feature extraction. For both the 'Category' and 'Sub-category' runs of this study, the *AVES-all* model was trained using 1,000 of the annotated recording snippets, selected randomly, and then tested against the remaining 2,000. In both instances, the model was trained up to 30 epochs, with a default learning rate of 10^{-3} and batch size of 50 (i.e. 50 recordings read into the model at a time). Afterwards, the model was applied to the entire NERSC2 dataset (training snippets omitted) to observe temporal patterns of each sound source.

3. RESULTS AND DISCUSSION

3.1. TRAINING AND TESTING *AVES* WITH BROAD CATEGORY LABELS

The performance of *AVES-all* after 30 train/test epochs with broad category labels is presented in Figure 2. This model was able to recognise biophony with the highest success for all metrics (accuracy, precision, recall, Intersection over Union (Jaccard Index) and F1 score). The high number of samples for this category should however be compared with the numbers for other categories, although it is justified by the wide variety of vocalisations observed throughout the dataset. Snippets containing geophonic signals were labelled with moderate success despite the low sample size, although a significant number of snippets in the category were labelled as containing biophony, resulting from vocalisations concurrent with seismic events.

Most snippets featuring cryophony were correctly labelled, but the numbers of false positives and negatives were high. False positives may have occurred due to clipping within a snippet, appearing as a single intense transient above all other levels. Sometimes cryophony and biophony were coincident, resulting in false positives/negatives between the two categories.

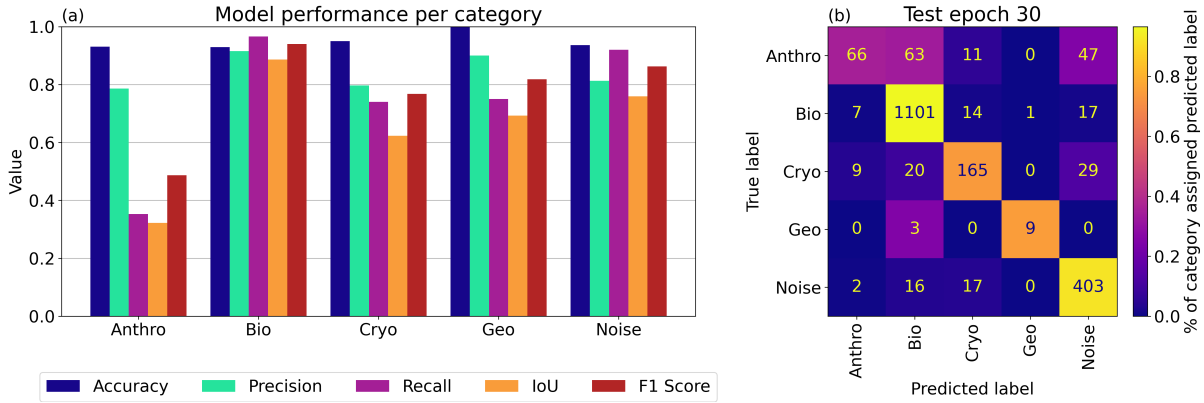


Figure 2: Model performance after training/testing for 30 epochs with broad category labels. (a) Performance metrics. IoU ('Intersection over Union') is also known as the Jaccard Index. (b) Confusion matrix of true vs. predicted labels.

The model struggled at correctly labelling snippets featuring anthropophonic signals, but the number of false positives was low. False negatives are as a result of anthropophony being largely faint against ambient noise except for the start and end of deployment (close proximity of the icebreaker deploying and retrieving the moorings), amongst other occurrences. Mislabelling of the noise category may have occurred due to the presence of single acoustic signatures of unknown origin, be they transients or short tones.

3.2. TRAINING AND TESTING *AVES* WITH SUB-CATEGORY LABELS

The performance of *AVES-all* after 30 train/test epochs with sub-category labels is presented in Figure 3. As expected for a model built for encoding animal vocalisations, *AVES-all* was better at correctly labelling sub-categories concerning temporally short events rather than continuous processes. The model was able to label most anthropophonic transients ('Anthro1'), but performance was poor for anthropophonic tones ('Anthro0'), though it should be noted that only a few isolated instances of anthropophony were included in the annotated dataset.

For biophonic activity, *AVES-all* was most effective at labelling files containing bearded seal mating calls ('Bio0'), perhaps due to these call types being the most consistent in time-frequency signatures. The majority of snippets featuring trains of downsweep calls ('Bio1') were labelled correctly, with lower precision and a significant number of mislabellings into other biophonic sub-categories. Low frequency calls ('Bio2') were labelled with good precision, but with low recall, in part due to calls being warped during propagation, causing some confusion with anthropophonic transients and tones. For mixed biophonic sub-categories, the model was more successful in labelling snippets with non-overlapping calls ('Bio3'), but with a notable amount of false positives and negatives. Snippets containing overlapping calls ('Bio4') were poorly labelled, with more false positives than true positives. Notably, 'Bio1', 'Bio3' and 'Bio4' shared many false negatives and positives, so if these sub-categories were merged together as a broader '$>$ 100 Hz calls' sub-category, the model may have been able to perform better.

Performance decreased for cryophonic events. While a decent precision was achieved, instances of cryophony were missed for both sub-categories, resulting in a lower recall. For cryophonic transients ('Cryo0'), the majority of false negatives were labelled as 'Noise' because transients were faint, and other false negatives arose along other biophonic sub-categories,

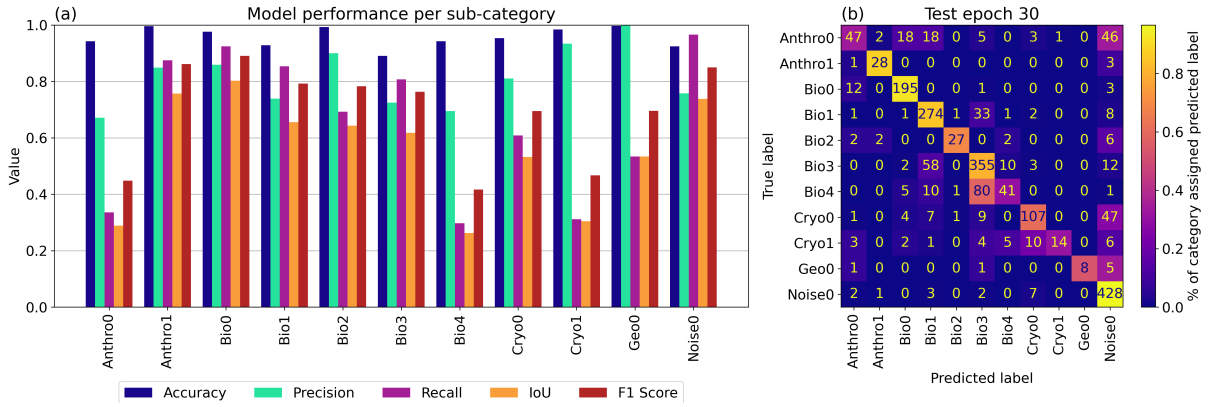


Figure 3: Model performance after training/testing for 30 epochs with sub-category labels. (a) Performance metrics. (b) Confusion matrix of true vs. predicted labels.

due to co-occurrence. For cryophonic tones ('Cryo1'), many false negatives were labelled as 'Cryo0', likely because ice tones were rarely separate from transient events. Geophony snippets showed lower recalls, despite remaining a single category, with most false negatives labelled as noise. Snippets labelled as 'Noise' were recalled well, although there were a significant number of false positives across all other sub-categories, likely due to the sounds being faint.

3.3. PROCESSING THE ENTIRE DATASET WITH *AVES*

Temporal patterns of soundscape categorisation across the NERSC2 dataset are given in Figure 4 for both broad category and sub-category labels. The 'summed weighted power spectral density (PSD) index' ($swPSD$), an adaptation of the 'bioacoustic index' used in ecoacoustic studies, is computed using *scikit-maad* [12]. It is plotted here for the frequency bands of 10 - 100 Hz and 100 - 488 Hz (the highest frequency achievable here), highlighting the variance of acoustic energy above the ambient noise floor [13].

Between 100 Hz and 488 Hz, $swPSD$ shows a near consistent level in acoustic activity, regardless of what sources are present, with slight drops from December 2019 to mid-January 2020 and from mid-May to mid-June 2020. This coincides with time frames along the dataset labelled as predominantly featuring biophony. In winter months, 'Bio1', 'Bio2' and 'Bio4'

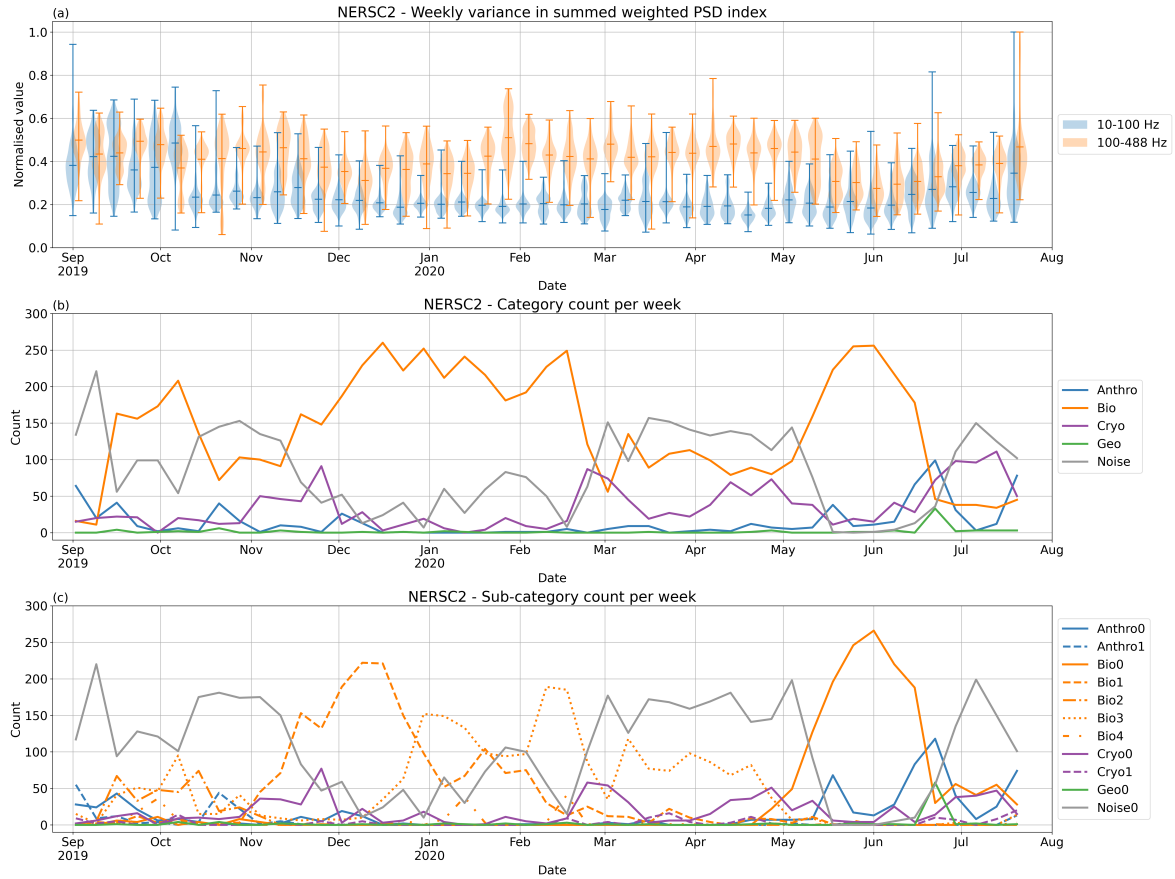


Figure 4: (a) Weekly variance in summed weighted PSD spectrum, across the frequency bands 10-100 Hz and 100-488 Hz. Weekly count of (b) broad category and (c) sub-category labels across entire NERSC2 dataset.

were the prevailing biophonic categories and they are attributed to bowhead whales. This is comparable to the vocal activity of bowhead whales observed in the Fram Strait during winter months [2, 14]. Bearded seal downsweep calls were observed only from mid-May to July 2020, which is comparable, but perhaps more restricted than other regions of the Arctic [15]. Sub-100 Hz calls were only observed between mid-September to November 2019, potentially from rorqual whale species visiting the region.

Anthropophony was largely restricted to the summer and autumn months, as shown not only by the increase in label count across these months, but from an increase in $swPSD$ when computed across 10-100 Hz. Some anthropophonic tones were noted up to December 2019, with sparse anthropophony next identified in March 2020. This is consistent with AIS records.

Along the dataset, increases in cryophony were often in line with increases in ambient noise. As presented in the confusion matrices in Figures 2 and 3, many cryophonic events may have been mislabelled as just noise. Around April 2020, ice cracking sounds may have been so frequent that they just accumulated into a raised ambient noise level. The number of snippets labelled as geophonic remained low across most of the dataset. The increase in geophonic snippets in June 2020 was partly from the presence of seismic events, but also from mislabelling of loud, modulating anthropophonic tones.

Comparing the two approaches of labelling the dataset, temporal patterns of sub-category counts are largely in agreement with their combined categories. If trained on more samples of isolated anthropophony, cryophony and geophony, the performance of *AVES* for identifying long-duration events, something it was not initially designed for, would improve.

4. CONCLUSIONS

AVES is a ML model originally built to identify animal vocalisations using audio waveforms. We show here that it can be used to broadly and automatically annotate large datasets from Arctic, ice-covered waters. Running the model across the entire deployment period identifies the temporal patterns of prevalent anthropophony, biophony and cryophony, along with isolated seismic events. The model allows for a quick first pass of annotation of acoustic datasets and it will next be applied to future datasets where temporal soundscape contributions are unknown. To further improve the applications of this approach, a set of models, e.g. *AVES* or other ML models, can be set to run in synchronicity to identify individual categories, assigning multiple and increasingly complex labels to recordings.

5. ACKNOWLEDGEMENTS

The CAATEX data used in this paper are used with permission from Principal Investigators of the CAATEX projects: Dr. Hanne Sagen, NERSC, (Project No. 280531) funded by the Research Council of Norway, and Dr. Matthew Dzieciuch, Scripps Institution of Oceanography, University of California (Project No. N00014-18-1-2698) funded by the Office of Naval Research. The HiAOOS project is funded by the European Union Horizon Europe Programme (Grant Agreement No. 101094621, UK participants supported by UKRI Grant No. 10071903). JC would like to acknowledge Gagan Narula (Earth Species Project) for his prompt correspondence and great advice in getting set up with *AVES*. This work has been completed as part of JC's EPSRC-funded PhD studentship (EP/W524712/1) supervised by PB.

REFERENCES

[1] Duarte, C. M., *et al.* : "The soundscape of the Anthropocene ocean", *Science* , **371**, eaba4658 **(2021)**

[2] Cleverly, J., Blondel, Ph., Sagen, H., Dzieciuch, M. and Storheim, E. : "Automatic identification of biophonics and sea ice processes in large datasets from the High Arctic Ocean", *International Conference on Underwater Acoustics, ICUA 2024, Bath, UK*, **46 (2024)**

[3] Worcester, P., Dzieciuch, M. and Sagen, H. : "Ocean Acoustics in the Rapidly Changing Arctic", *Acoustics Today*, **16**(1), 55-64 **(2020)**

[4] Hagiwara, M. : "AVES: Animal Vocalization Encoder based on Self-Supervision", *arXiv pre-print*, arXiv:2210.14493v1 **(2022)**

[5] NERSC : "CAATEX - Coordinated Arctic Acoustic Thermometry Experiment" https://caatex.nersc.no/ (last accessed: May 2025)

[6] NERSC : "HiAOOS - High Arctic Ocean Observation System" https://hiaoos.eu/ (last accessed: May 2025)

[7] Copernicus Marine Service : "Arctic Ocean Sea Ice Analysis and Forecast", https://doi.org/10.48670/moi-00004 (last accessed: May 2025)

[8] Williams, T., Korosov, A., Rampal, P. and Ólason, E. : "Presentation and evaluation of the Arctic sea ice forecasting system neXtSIM-F", *The Cryosphere*, **15**, 3207-3227 **(2021)**

[9] Met Office : "Cartopy: a cartographic Python library with a Matplotlib interface", https://scitools.org.uk/cartopy/docs/latest/ (last accessed: May 2025)

[10] Hsu, W.-N., *et al.*: "HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units", *arXiv pre-print*, arXiv:2106.07447 **(2021)**

[11] Narula, G. : "esp-aves 1.0.0", Earth Species Project **(2025)**, https://pypi.org/project/esp-aves/ (last accessed: May 2025)

[12] Ulloa, J. S., Haupert, S., Latorre, J. F., Aubin, T. and Sueur, J. : "scikit-maad: An open-source and modular toolbox for quantitative soundscape analysis in Python", *Methods in Ecology and Evolution*, **12**(12), 2273-2500 **(2021)**

[13] Bradfer-Lawrence, T., *et al.* : "The Acoustic Index User's Guide: A practical manual for defining, generating and understanding current and future acoustic indices", *Methods in Ecology and Evolution*, **00**, 1 - 11 **(2024)**

[14] Ahonen, H., Stafford, K. M., de Steur, L., Lydersen, C. and Wiig, Ø. : "The underwater soundscape in western Fram Strait: Breeding ground of Spitsbergen's endangered bowhead whales", *Marine Pollution Bulletin*, **123**, 97-112 **(2017)**

[15] Escobar-Amado, C. D., Badiey, M. and Pecknold, S. : "Automatic detection and classification of bearded seal vocalizations in the northeastern Chukchi Sea using convolutional neural networks", *The Journal of the Acoustical Society of America*, **151**(1), 299-309 **(2022)**

Change Detection in Synthetic Aperture Sonar Imagery Using the Segment Anything Model

William Hedlund[1], Per Abrahamsson[1], and Louise Rixon Fuchs[3]

[1]Saab AB, SE-581 88 Linköping, Sweden

William Hedlund, Bröderna Ugglas gata, SE-581 88 Linköping, Sweden
william.hedlund3@saabgroup.com

Abstract: *Recent advancements in artificial intelligence have led to the development of foundation models, which are large, general-purpose models capable of performing a wide range of tasks across domains. This paper investigates the use of the Segment Anything Model (SAM), a foundation model for image segmentation, for change detection in synthetic aperture sonar (SAS) imagery. The method is entirely training-free and applies SAM to segment bi-temporal SAS images, which are then compared to identify changes. Three prompting strategies are evaluated: two use grids of points with varying spacings, and one uses a change map generated by a traditional method. The approach is tested on both real and synthetically generated datasets, with the synthetic data created through image composition to simulate changes. Results show that the change-map and fine-grid methods achieve similar precision and recall, while the coarse grid results in more missed detections but higher precision. On real data, the SAM-based approach outperforms a traditional log-ratio method by improving both precision and recall, reducing false alarms caused by misregistration and environmental noise. These results demonstrate SAM's adaptability to new domains and its potential as a viable, training-free alternative for SAS change detection, supporting broader applications of deep learning in marine research and environmental monitoring.*

Keywords: *change detection, synthetic aperture sonar, segmentation, Segment Anything Model*

1. INTRODUCTION

Change detection (CD) refers to the task of identifying differences between two observations of the same scene taken at different times. In the underwater domain, this is typically performed by comparing synthetic aperture sonar (SAS) images collected by autonomous underwater vehicles (AUVs) on repeat-pass surveys. As AUV and SAS technologies continue to improve in coverage speed and resolution, the volume of collected data increases significantly, exceeding the capacity of manual review and demanding reliable automated processing techniques [1]. A key application of SAS-based surveying is mine reconnaissance, where regular repeat-pass imaging in high-traffic areas enables detection of newly introduced or moved objects. Automated change detection in such contexts supports maritime safety by aiding early threat identification.

Early methods for CD in SAS imagery, matched new detections to known historical objects. Modern approaches enhance this by co-registering repeated SAS images, analysing differences in backscatter intensity to produce coherence maps [1]. These methods are sensitive to co-registration errors, environmental changes, and threshold tuning, limiting their robustness in complex, cluttered underwater environments. Furthermore, they operate solely on pixel-level intensity comparisons, without leveraging higher-level contextual information that human operators often use to distinguish meaningful changes from irrelevant noise.

In other remote sensing domains such as satellite and aerial imagery, deep learning has surpassed traditional methods for change detection [2]. However, deep learning models typically require large annotated datasets, which are costly or impractical to obtain in SAS applications due to operational constraints and the sensitive nature of much sonar data. Deep learning has been applied to SAS imagery primarily for target classification, seabed characterization, and image enhancement [3, 4, 5]. Despite these advances, the application of deep learning for change detection in SAS imagery remains limited. This highlights a need for methods that can leverage deep learning's capabilities while addressing the challenges of scarce labelled data.

In recent years, models trained on massive datasets have given rise to foundation models, which are large, pre-trained models with capabilities spanning a wide range of tasks. These models exhibit strong generalization and often require little or no fine-tuning to adapt to new domains. One such model is the Segment Anything Model (SAM) [6], which performs image segmentation by using different types of input prompts such as points, boxes, or masks to guide and condition the model's behaviour to focus on specific regions or features in the image. SAM has shown adaptability across domains, including medical imaging, microscopy, and satellite imagery. This raises the question of whether these capabilities can also benefit the domain of sonar imagery and change detection, where the visual modality differs significantly from natural images.

2. METHOD

The work of this paper focuses on leveraging a foundation model to perform training-free change detection in SAS imagery. To evaluate the results, the proposed method will be compared to a traditional algorithmic log-ratio approach described in Section 3. Prerequisites include data collection and preprocessing of the images, these components are briefly described in this section.

2.1. DATA COLLECTION AND PREPARATION

The data used in this study was collected using Saab's AUV62-MR system, an autonomous underwater platform equipped with dual SAS sensors. The surveys were conducted over two separate missions with a one-day interval. Between passes, several objects were manually introduced to simulate real changes. These included a rock, a metal pipe, and a bundle of rope, placed in diverse environments such as textured sand and gravel. The SAS system produced high-resolution imagery at approximately 9 cm^2 per pixel, as seen in Figure 1.

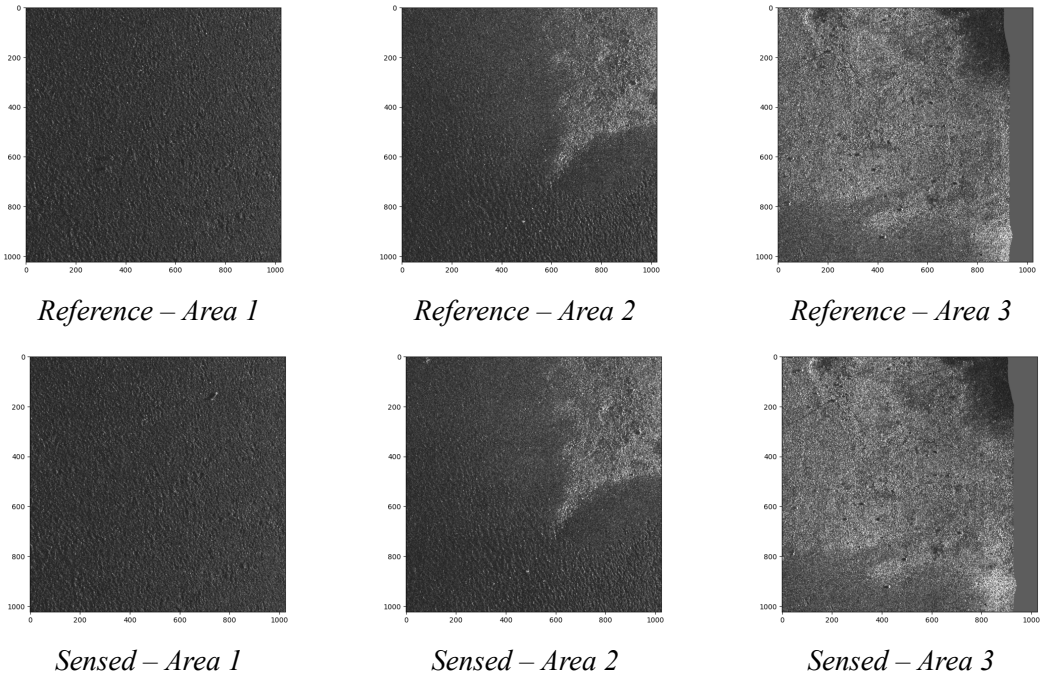


Figure 1: Areas where real change is introduced. The images are approximately 30×30 m. In each case, the changed object is visible in the image: top right for Area 1, top left for Area 2, and top left for Area 3.

Because real change data is costly and time-consuming to collect, a synthetic dataset was also constructed. This involved segmenting real objects from prior missions and inserting them into unchanged image backgrounds using cut-and-paste augmentation [7]. This approach enables controlled evaluation of the change detection method across a wider variety of object types and scenes.

Co-registration is essential to correct for navigational drift and ensure spatial alignment between bi-temporal images. We adopted a multistage registration approach based on the methods described in [8, 9]. First, SIFT keypoints are extracted and matched, with RANSAC applied to reject outliers and achieve a coarse global alignment. This is followed by localized refinement using normalized cross-correlation in the Fourier domain. The final transformation is computed on a per-tile basis, enabling sub-pixel precision and allowing compensation for local distortions caused by terrain variation.

3. TRADITIONAL LOG-RATIO-BASED CHANGE DETECTION

Given two co-registered SAS intensity images $I_1, I_2 \in \mathbb{R}^{H\times W}$, the *log-ratio operator* is defined to highlight regions of change while reducing the influence of multiplicative noise:

$$D(x,y) = \left|\log\left(\frac{I_1(x,y)}{I_2(x,y)}\right)\right| \tag{1}$$

The resulting *difference image* D is transformed using pixel-level saliency [10, 11] as:

$$\text{Sal}(a_m) = \sum_{n=0}^{255} f_n \left|a_m - a_n\right| \tag{2}$$

where f_n is the frequency of pixel intensity a_n in the image histogram, and $|a_m - a_n|$ is the absolute intensity difference. This saliency transformation emphasizes perceptually significant regions, reducing speckle noise. The saliency-enhanced image is then thresholded using *Otsu's method* [12], which determines the optimal threshold T_{Otsu} by maximizing inter-class variance. The final binary *change map* CM is defined such that $CM(x,y) = 1$ if $D(x,y) > T_{\text{Otsu}}$, and $CM(x,y) = 0$ otherwise. To reduce noise and speckle artifacts, a median filter is applied to the resulting binary map CM.

4. SAM-BASED OBJECT-LEVEL CHANGE DETECTION

In contrast to pixel-based differencing, the proposed method performs change detection at the object level using SAM. Each image I_1 and I_2 is segmented independently. We evaluate three prompting strategies: (1) a **sparse grid** of point prompts with 32-pixel spacing, (2) a **dense grid** of point prompts with 16-pixel spacing, and (3) prompts derived from the traditional method, using the **change map as input** to focus on more salient regions of the image.

SAM returns a set of binary object masks for each image, denoted as $\mathcal{S}_1 = \{M_1^i\}$ and $\mathcal{S}_2 = \{M_2^j\}$. Change detection is performed by comparing these masks using the Intersection over Union (IoU), defined as $\text{IoU}(M_1^i, M_2^j) = \frac{|M_1^i \cap M_2^j|}{|M_1^i \cup M_2^j|}$. An object M_1^i is marked as changed if its maximum IoU with any object in the other image is below 0.5. Symmetrically, unmatched masks from $\mathcal{S}_2$ are also considered changes. To reduce false positives, segmented objects with fewer than 100 pixels are discarded. Morphological operations are then applied to refine the boundaries of the masks.

5. PERFORMANCE METRICS FOR CHANGE DETECTION

Evaluation of change detection is performed per pixel, as each pixel is classified as either changed or unchanged. Due to the inherent class imbalance where changed pixels are much less frequent, metrics such as: precision, recall, and F1 score are used. Precision is defined as $\text{Precision} = \frac{TP}{TP+FP}$, recall as $\text{Recall} = \frac{TP}{TP+FN}$, and the F1 score as $\text{F1 Score} = \frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}$. While overall accuracy is frequently reported, it can be misleading for imbalanced datasets. The F1 score provides a more balanced evaluation by combining both precision and recall.

6. RESULTS

Evaluation is performed on two datasets: synthetically generated data for quantitative comparison of prompt-methods on the SAM-based method, and qualitative comparison on the real data comparing the traditional method and selected prompt-method.

6.1. QUANTITATIVE EVALUATION – SYNTHETIC DATA

As shown in Table 1, the method using a change map as input achieves the highest F1 score (0.73), indicating a good balance between precision (0.68) and recall (0.78). The dense grid method (16-pixel spacing) follows closely with an F1 score of 0.71, with notably higher recall (0.82) but lower precision (0.62), suggesting more detected changes but also more false positives. The sparse grid method (32-pixel spacing) achieves the highest precision (0.71), indicating fewer false positives, but has the lowest recall (0.49), resulting in many missed changes and the lowest F1 score (0.58). This shows that the prompt method has a significant impact on the results.

Point Prompt Method	Precision	Recall	F1 Score
Change map	0.68	0.78	0.73
Sparse grid (32 px)	0.71	0.49	0.58
Dense grid (16 px)	0.62	0.82	0.71

Table 1: Performance metrics on synthetic data for different SAM prompt strategies.

6.2. QUALITATIVE EVALUATION – REAL DATA

For this qualitative comparison, we focus on the method that uses the change map from the traditional approach, as it demonstrated the best trade-off between precision and recall in the quantitative evaluation. The traditional method was tuned on synthetic data using a median blur kernel size of 3×3 and a minimum region size of 5×5. A larger blur kernel removed thin objects entirely, while omitting it resulted in excessive false alarms. SAM was configured with a confidence threshold of 0.8.

The resulting change maps are shown in Figure 2. In Area 1, the SAM-based method captures both the changed object and its shadow completely, which is desirable for accurate interpretation. The predicted change map includes three false alarms, which were found to be caused by shadows resulting from an offset in the collection path rather than true scene changes. In contrast, the traditional method produces a fragmented and incomplete mask of the object, with numerous false alarms distributed across the image. In Area 2, the SAM-predicted mask contains no false alarms and clearly highlights the relevant change, in contrast to the traditional method, where the change is difficult to perceive due to false alarms.

While grid-based prompt strategies were also evaluated, they are not included in this qualitative comparison. These methods generally showed higher recall but introduced more false alarms, consistent with the quantitative results. These observations reinforce that prompt selection is a critical factor in SAM-based change detection. Point prompts targeting salient regions

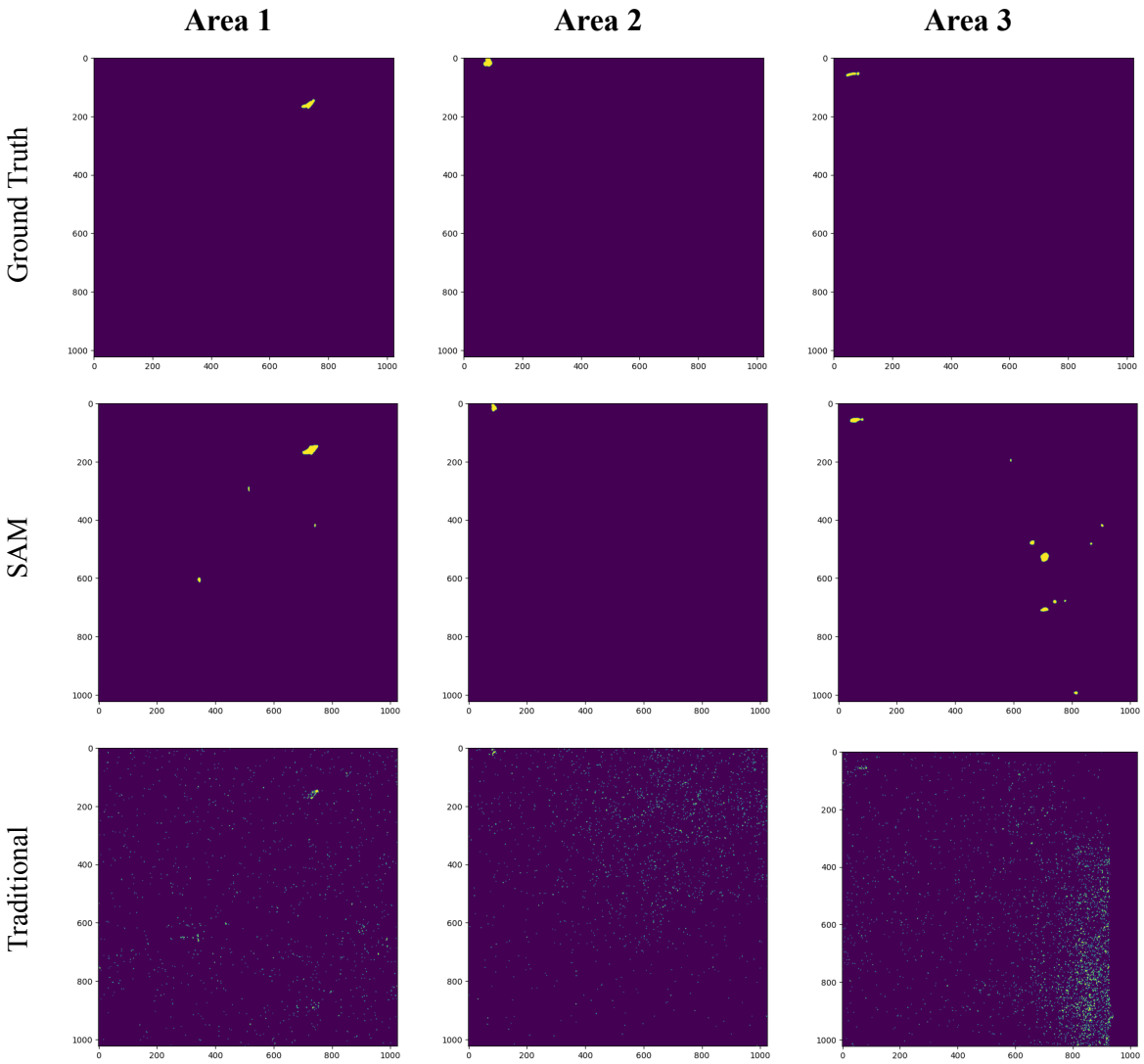


Figure 2: Comparison of change detection methods across three test areas. Columns correspond to Area 1, Area 2, and Area 3. Rows show ground truth (top), SAM-based predictions (middle), and traditional method results (bottom).

offer a strong balance between precision and recall, improving change mask quality while reducing false detections.

7. CONCLUSION

This work demonstrates the potential of adapting foundation models, specifically SAM, for change detection in SAS imagery. Despite the domain mismatch between natural and sonar images, the results show that SAM can be effectively applied using selected point prompts, particularly when guided by salient regions from a change map. This indicates a promising approach for low-data scenarios where training deep models is impractical. However, the real-world dataset used in this study is limited in size, making it difficult to draw definitive conclusions.

Future work could focus on developing methods that incorporate semantic understanding and are robust to variations in survey trajectories and viewing angles. Such advancements would further support the practical deployment of change detection systems in complex underwater environments.

REFERENCES

[1] Crawford, A., Hansen, R., Johnson, S., King, J., Midtgaard, Ø., Sæbø, T., Salas, J., Shouldice, E., Smith, S., Sternlicht, D., and Woods, D.: "Improvements to Automated Change Detection Tools for SAS Images," in *Proceedings of OCEANS 2022*, IEEE, 2022.

[2] Shafique, A., Cao, G., Khan, Z., Asad, M., and Aslam, M.: "Deep Learning-Based Change Detection in Remote Sensing Images: A Review," *Remote Sensing*, vol. 14, no. 4, pp. 871, 2022.

[3] Williams, D.P.: "Acoustic-Color-Based Convolutional Neural Networks for UXO Classification with Low-Frequency Sonar," in *Proceedings of the Underwater Acoustics Conference*, 2019.

[4] Sun, Y.-C., Gerg, I.D., and Monga, V.: "Iterative, Deep Synthetic Aperture Sonar Image Segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[5] Fuchs, L.R., Larsson, C., and Gällström, A.: "Deep Learning-Based Technique for Enhanced Sonar Imaging," in *Proceedings of the 5th Underwater Acoustics Conference & Exhibition (UACE)*, Hersonissos, Crete, Greece, pp. 1021–1028, 2019.

[6] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., and Girshick, R.: "Segment Anything," arXiv:2304.02643, 2023. `https://arxiv.org/abs/2304.02643`

[7] Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E.D., Le, Q.V., and Zoph, B.: "Simple Copy-Paste Is a Strong Data Augmentation Method for Instance Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2918–2928, 2021.

[8] G-Michael, T., Marchand, B., Tucker, J.D., Marston, T.M., Sternlicht, D.D., and Azimi-Sadjadi, M.R.: “Image-Based Automated Change Detection for Synthetic Aperture Sonar by Multistage Coregistration and Canonical Correlation Analysis,” *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 592–612, 2016.

[9] Wang, V.T. and Hayes, M.P.: “Synthetic Aperture Sonar Track Registration Using SIFT Image Correspondences,” *IEEE Journal of Oceanic Engineering*, vol. 42, no. 4, pp. 901–913, 2017.

[10] Zhai, Y. and Shah, M.: “Visual Attention Detection in Video Sequences Using Spatiotemporal Cues,” in *Proceedings of the 14th ACM International Conference on Multimedia*, pp. 815–824, 2006.

[11] Zhang, Y., Wang, S., Wang, C., Li, J., and Zhang, H.: “SAR Image Change Detection Using Saliency Extraction and Shearlet Transform,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 12, pp. 4701–4710, 2018.

[12] Otsu, N.: “A Threshold Selection Method from Gray-Level Histograms,” *Automatica*, vol. 11, pp. 23–27, 1975.

Structured Session
02. Acoustics in polar environments

Organised by Espen Storheim, Lora van Uffelen and Oskar Glowacki

Variability of acoustic signals near a Polar front at the Western Barents Sea

Altan Turgut[1], and Jeffrey Schindall[2]

[1]SoundOceanAI, 1110 S. Washington St., Falls Church, VA 22046, USA
[2]Naval Research Laboratory, Acoustics Div., Washington, DC 20375

Altan Turgut, 1110 S. Washington St., Falls Church, VA 22046, USA,
alturgut@yahoo.com

Abstract: *Analysis of mid-frequency (0.7 to 4 kHz) acoustic data from the October 2022 Barents Sea experiment exhibited the effects of the Polar front, tidal currents, and tidal heights on acoustic travel-time fluctuations. Broadband acoustic model predictions revealed that the primary contribution to the travel-time fluctuations were due to the dynamics of a Polar front in the experimental area. Similar travel-time fluctuations were also calculated using HYCOM-predicted sound speed profiles capturing the dynamics of the Polar front. This implied that use of HYCOM would facilitate a prediction capability of acoustic travel-time fluctuations near a Polar front. In addition, up to 15 dB higher acoustic Transmission Loss (TL) was observed during a major storm event. Measured acoustic Scintillation Index (SI) indicated an unexpected decrease of SI values during this storm event. This SI decrease can be explained by scattering and attenuation of surface-interacting high-angle arrivals and propagation of only more stable low-angle arrivals at a 10 km range. This phenomenon is further investigated by simulating additional scattering mechanisms due to bottom roughness and diffuse background internal waves. Further investigation is underway to include an additional attenuation mechanism due to near-surface bubble layer as well as a 3D out-of-plane scattering.*

Keywords: *Polar front, Travel-time fluctuations, Scintillation Index*

INTRODUCTION

Amplitude and travel-time variability of mid-frequency acoustic signals near a Polar front is of interest given the scarcity of previous mid-frequency acoustic experiments and theoretical work at high latitudes. Previous experimental and theoretical work by Lynch et al.[1] investigated the effect of diffuse internal waves on low frequency (224 Hz) amplitude and travel-time fluctuation near a Polar Front (PF) in the Barents Sea. Experimental difficulties limited their acoustic data to a cross-front path in which a combination of frontal and internal wave/internal tide signals is present. At lower latitudes, similar studies have been reported to investigate the acoustic scattering by internal waves, surface gravity waves, and fish[2,3]. This paper presents results from a unique experiment providing acoustic and oceanographic data near a PF in the Western Barents Sea covering both cross-front and along-front paths. Therefore, it focuses on the direct effects of a Polar front on the mid-frequency acoustic propagation. In addition, it uses the HYbrid Coordinate Ocean Model (HYCOM) in the acoustic propagation modelling to investigate the travel-time and amplitude fluctuation due to the Polar Front dynamics. Additional effects such as sea-surface roughness, bottom roughness, and diffuse background internal waves are also included in the acoustic propagation modeling.

EXPERIMENT

The experiment took place near a PF south of Svalbard (Norway) during October 7-17, 2022. Nine oceanographic moorings and three acoustic moorings were deployed within a 25 km by 25 km area. Oceanographic measurements included current, hydrographic, and microstructure observations. Acoustic measurements included transmission of Linear Frequency Modulated (LFM) and Continuous Wave (CW) signals, received at a ~10 km range in cross-front and along-front paths. The source mooring uses an ITC 2010 source, covering a frequency band of 0.7-4.2 kHz. Mid-frequency acoustic data were collected for one-week duration on two 32-channel Vertical Line Arrays (VLAs) with a 77.5 m aperture. Figure 1a shows the locations of the acoustic source and two VLAs in the experimental area. Figure 1b shows the sound-speed at 2 m depth calculated from HYCOM temperature and salinity predictions for October 10, 2022.

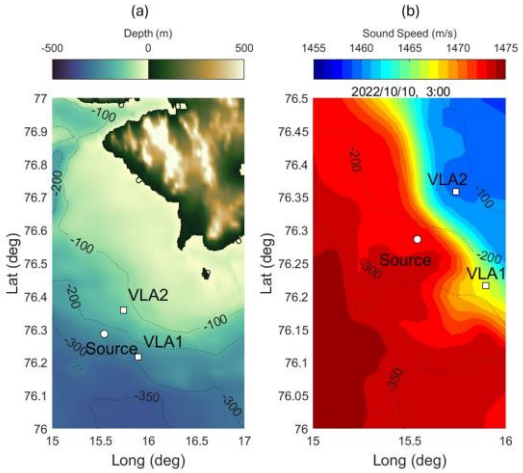


Fig.1: Experimental area. a) acoustic mooring locations and b) sound-speed at 2 m depth calculated from HYCOM temperature and Salinity data for Oct 10, 2022.

The colder Arctic water intrusion (low-sound-speed region near Svalbard) in the cross-front path is well predicted by the HYCOM model that is consistent with the Microstructure Vertical Profiler (MVP) measurements. In Figs. 2a and 2b, sound speeds from the MVP measurements and HYCOM predictions are compared along the MVP tracks that was collected during October 11, 2022. In Figs. 2c and 2d, HYCOM predicted sound-speed fields are shown along Source/VLA2 and Source/VLA1 paths, respectively. Figures 2e and 2f show the HYCOM predicted sound-speed fields from Oct 10 to Oct 14, 2022 at 2 m depth along Source/VLA2 and Source/VLA1 paths, respectively. Note in Fig. 2e the intrusion of cold Atlantic water during 10/11/2022 in the Source/VLA2 path with a modulation at M2 tidal period. Time series of temperature and salinity data from the oceanographic moorings also displayed similar variability related to tidal motion with the Arctic and Atlantic waters moving in and out of the Source/VLA2 path.

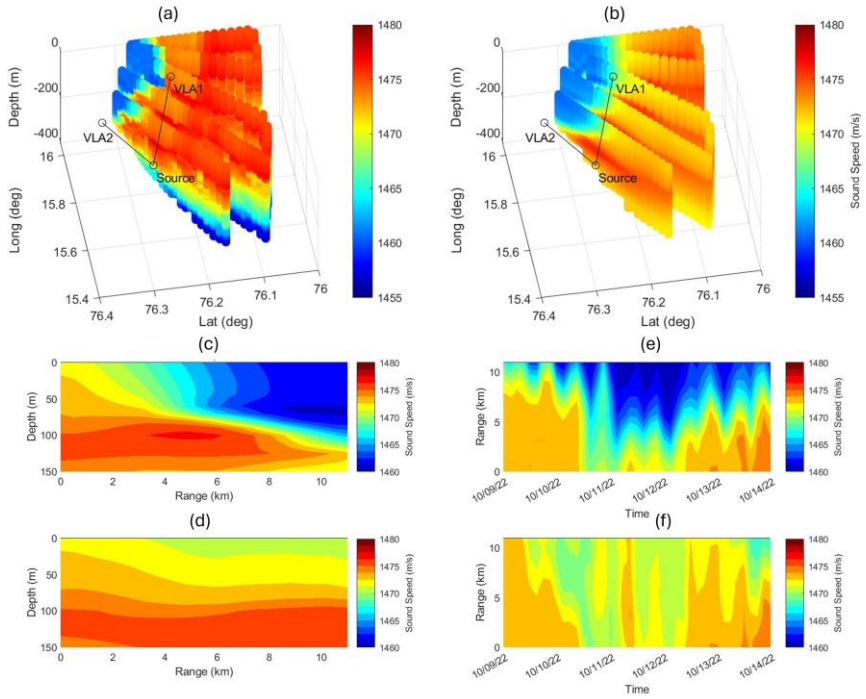


Fig.2: a) Sound-speed field along the survey track calculated from the MVP measurements, b) sound speed along the survey track predicted from HYCOM, c) sound speed along the source/VLA2 path predicted from HYCOM, d) sound speed along the source/VLA1 path predicted from HYCOM, e) sound speed at 2 m depth along the source/VLA1 path calculated from HYCOM, and f) sound speed at 2 m depth along the source/VLA2 path calculated from HYCOM.

The ten-day survey was impacted by two major storms with winds around 20 m/s that generated surface waves with significant wave heights above 4 m. Bottom pressure was dominated by semidiurnal fluctuations with a measured tidal-height less than 1.2 m. In this work, starting from Oct 9, 2022 at 18:20UTC, four days of acoustic measurements including the first major storm event are presented. Figure 3 shows 700-1200 Hz, 2-sec long LFM pulses transmitted at every 30 second for four days and received at a VLA1 hydrophone (90 m deep) and at a VLA2 hydrophone (110 m deep). Figs. 3a and 3d show the Matched-Filter (MF) outputs, Figs. 3b and 3e show the peak (MF) outputs, and Figs. 3c and 3f show the relative travel-times for the VLA1 and VLA2, respectively. Note in

Figs. 3a and 3d the stripping of high-angle late arrivals by the first storm event around 50th hour (Oct 12, 2022). In Fig. 1e, peak MF output shows a 15 dB reduction during the storm event at VLA2 hydrophone that is close to the bottom. This was not observed clearly at the VLA1 hydrophone since it was near the mid-water column (see Fig. 3b). Similarly, up to ~50 ms travel-time fluctuations were observed at the VLA2 hydrophone with a M2 tidal period. Travel-time fluctuations and dominance of M2 tidal period were less obvious at the VLA1 hydrophone (see Fig. 3c).

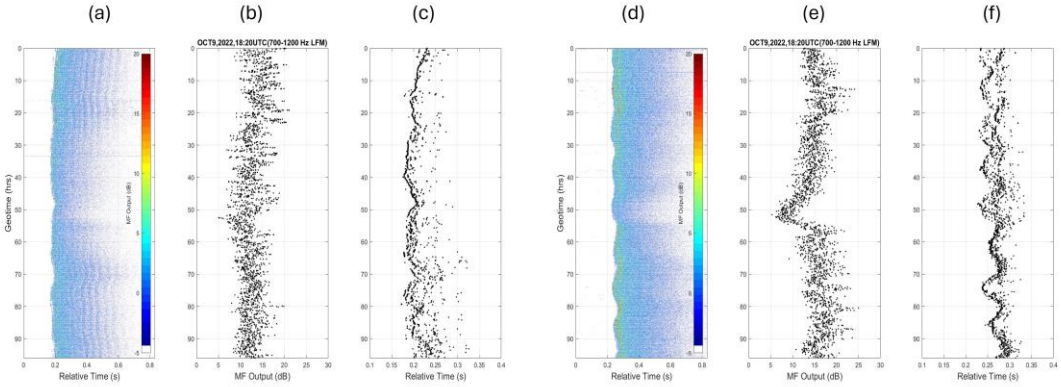


Fig.3: a,d) Matched-filter outputs of 700-1200 Hz LFM pulses, b,d) peak matched-filter outputs, and c,f) relative travel-time at peak MF outputs at VLA1 and VLA2, respectively

Figure 4 shows the measured Scintillation Index (SI) of 2-second long 1151 Hz and 2101 Hz CW signals calculated for every 15-minute duration (30 pings with 30-second repetition interval).

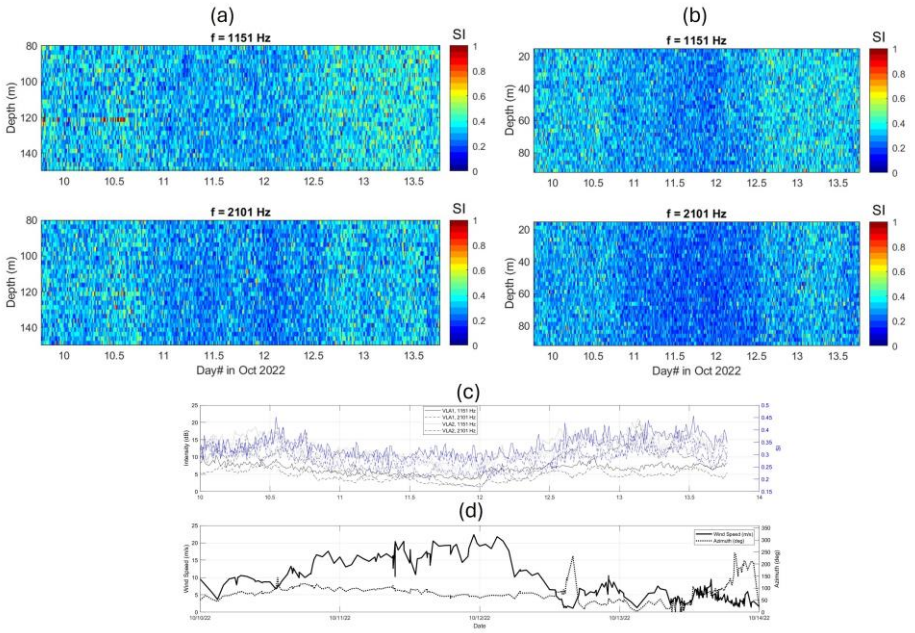


Fig. 4: a,b) Measured SI of 1151 Hz and 2101 Hz CW signals received at VLA1 and VLA2, respectively, b) depth-averaged acoustic intensity and SI, d) measured wind speed and direction on-board RV HU Sverdrup II.

The scintillation index values were calculated using the formula

$$SI = \left\langle \left(I - \langle I \rangle \right)^2 \right\rangle / \langle I \rangle^2 , \quad (1)$$

where I is the signal intensity and $\langle I \rangle$ is the average intensity. Note that both measured acoustic intensity and SI are reduced during the major storm event when the wind speed increased to 20 m/s. Decreasing SI at high seas seems to be against the intuition. This will be discussed in the next section with a major numerical simulation effort.

NUMERICAL SIMULATIONS

To investigate the travel-time and amplitude fluctuations due to the Polar Front dynamics, the HYCOM model outputs are used in the acoustic propagation modelling. The HYCOM ocean model[4] is two-way coupled to the Community Ice CodE (CICE). It has 1/25° (~2.5 km in the study area) horizontal resolution and 41 vertical hybrid coordinate layers. It uses the Navy Coupled Ocean Data Assimilation (NCODA) system for assimilation of satellite surface height anomalies, temperature, and sea ice concentration as well as available in-situ vertical temperature and salinity profiles from XBTs, CTDs, Argo floats, moored buoys and gliders. Observations are assimilated on a 24-hour update cycle. The system is forced with atmospheric momentum and heat fluxes from the 0.17625° 3-hourly Navy Global Environmental Model (NAVGEM). Detailed analysis of oceanographic measurements, modelling, and data processing is available in Jarosz et al.[5] and Schulman at al.[6]

A rough-surface[7] and a 3D[8] Paracolic Equation (PE) model were used to calculate the travel-time and amplitude fluctuations in the area. In Figure 5, a broadband rough-surface PE model was used to investigate the effects of high seas and PF dynamics on broadband (700-850 Hz) acoustic signals. To simulate the sea-surface roughness, JONSWAP spectra were used with different wind speeds and a 100 km fetch. The pulse signal $p(t)$ is obtained by a Fourier synthesis of CW runs within 700 Hz and 850 Hz. Then, transmission loss of the pulse signal defined as $TL(t) = -20\log_{10}\left(\left|p(t)\right|^2\right)$.

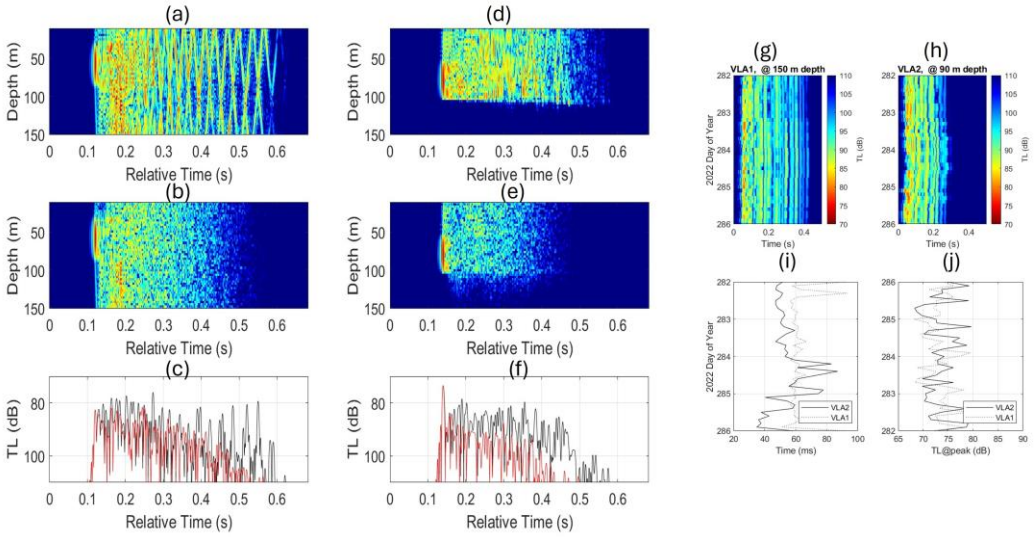


Fig. 5: a,b,c) comparison of 700-850 Hz signals received at VLA1 for flat surface and rough surface (wind speed: 20 m/s), d,e,f) comparison of 700-850 Hz signals received at VLA2 for flat surface and rough surface (wind speed: 20 m/s), g) modelled travel-time fluctuations at VLA1 using HYCOM predictions, h) modelled travel-time fluctuations at VLA2 using HYCOM predictions, i) relative arrival times at VLA1 and VLA2, and j) modelled TL fluctuations at VLA1 and VLA2.

In 5a, 5b, and 5c, comparison of broadband signals for the source/VLA1 path are shown for both a flat surface and a rough surface with a wind speed of 20 m/s. In 5d, 5e, and 5f, comparison of broadband signals for the source/VLA2 path are shown for a flat surface and rough surface with a wind speed of 20 m/s. Figures 5g and 5h show the modelled travel-time fluctuations due to the PF dynamics predicted by HYCOM data at every 3 hours (see Figs. 2e and 2f). Note in Fig. 5i that cross-front path has more travel-time fluctuations after the cold Arctic water intrusion starts on Year Day 284 (Oct 12, 2022). The model predicts ~40 ms travel-time fluctuations from HYCOM-predicted PF dynamics that is comparable with those of the measurements. The model also predicts a 3 ms travel-time variation for an 80 cm/s tidal current and a travel-time value of 0.3 ms for a 1.2 m tidal height. Also note in Fig. 5j that, there is ~ 10 dB TL fluctuations on both cross-front and along-front paths.

In Fig.6, probability density functions (pdfs) of measured SIs are compared with modelled SIs for 1151 Hz and 2101 Hz signals. As a reference, median SIs (50 percentile)

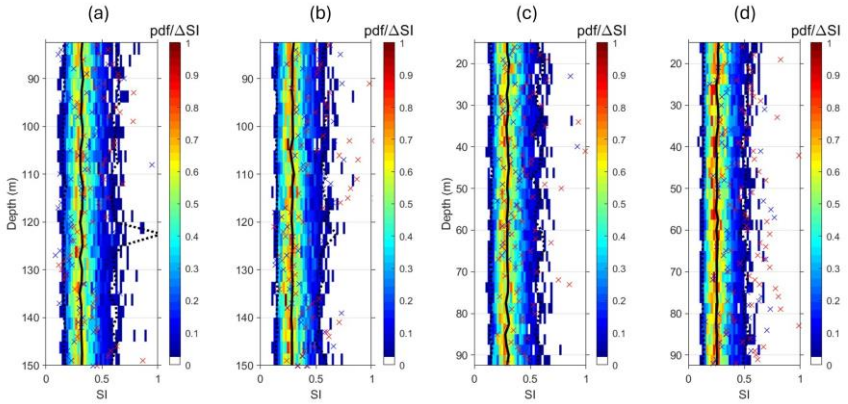


Fig. 6: a,b) Normalized pdfs of measured SI at VLA1 for 1151 Hz and 2101 Hz, respectively, c,d) normalized pdfs of measured SI at VLA2 for 1151 Hz and 2101 Hz, respectively. Model-predicted SIs are overlayed for a 5 m/s wind (blue x marks) and a 20 m/s wind (red x marks).

are plotted as black lines. In addition, 2.5 and 97.5 percentiles are plotted as dotted black lines. Median SI values of measured SIs were ~0.35 that did not show a depth dependence and indicated an unsaturated regime at VLA1 (12 km range) and VLA2 (10 km range). The modelled SI values were generated by rough-surface PE using 30 rough see-surface realizations for a 5 m/s wind (blue x marks) and 20 m/s wind (red x marks). The modelled SI values are slightly larger than the measured SIs. In addition, results show slight increase of SI values for the 20 m/s wind. Additional scattering mechanisms including rough-bottom and diffuse internal waves were also added in the rough-surface PE model (not shown in this paper) that did not change the results of SI values being slightly higher at higher winds. Additional attenuation mechanisms for the high angle arrivals (e.g., near-surface bubble layer, 3D scattering) might be needed to verify the measurements.

Increase of TL during the storm event was also studied using a 3D rough-surface PE model to investigate the interaction of surface gravity waves as well as a range/azimuth dependent bottom near a PF. Figure 7 shows the TL at 10 km range for a 700 Hz CW signal. Figures 7a and 7b show the TL predictions for a flat sea-surface using the HYCOM data for Oct 10 and Oct 13, respectively. Figure 7c compares the depth-averaged TL for these two cases. Considering the minor TL differences and relatively minor azimuthal dependence, dynamics of PF did not produce significant horizontal refraction. Figures 7d and 7e show the TL predictions for a rough sea surface generated by a 20 m/s wind coming from NW (45 deg. azimuth) direction during Oct 10 and Oct 13, respectively. Figure 7f compares the depth-averaged TL for these two cases. Difference between predicted TL for Oct 10 and 13 seems to be minor. However, there is a significant azimuthal dependency of TL due to interaction of directional surface gravity waves and range/azimuth dependent bathymetry. Further analysis of measured acoustic data will be conducted to investigate the azimuthal dependency due to surface gravity waves and highly variable bathymetry.

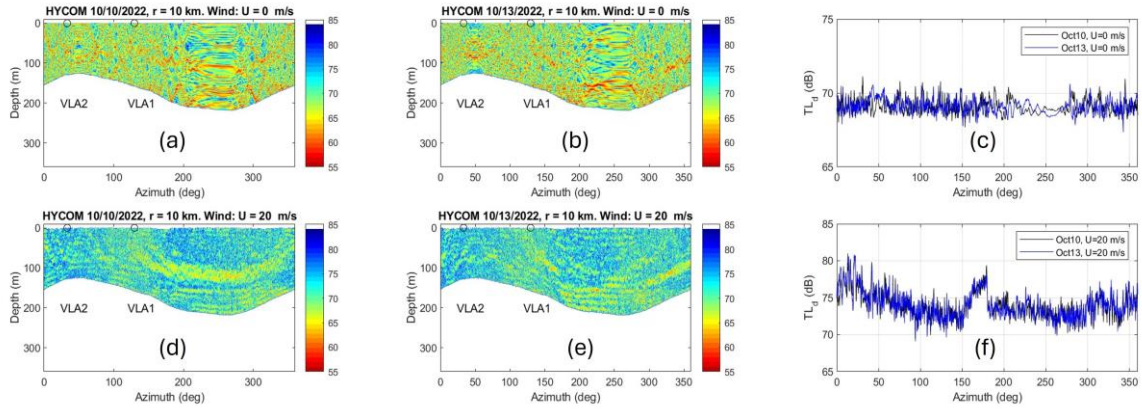


Fig. 7: a,b) TL predictions for a flat surface using the HYCOM data for Oct 10 and Oct 13, c) depth-averaged TL for a flat surface using the HYCOM data for Oct 10 and Oct 13, d,e) TL predictions for a rough surface using the HYCOM data for Oct 10 and Oct 13, f) depth-averaged TL for a rough surface using the HYCOM data for Oct 10 and Oct 13,

SUMMARY

This paper presents extensive field measurement and numerical simulation results to better understand the acoustic travel-time and amplitude fluctuations in the vicinity of a Polar front. Several experimental observations have been made during a four-day acoustic and oceanographic measurements. First, up to 50 ms of travel-time variations were observed especially after the Polar front moved into the Source/VLA2 propagation path. Second, a slight decrease of the Scintillation Index values was observed during a major storm event. Numerical simulations showed that higher sea states produced slightly increased SI values that contradicts with the observations. Inclusion of additional scattering due to bottom roughness and diffuse internal waves did not change the fact that higher sea states produced slightly increased SI values. Inclusion of a near surface bubble layer as an additional attenuation mechanism, and 3D modelling to include additional (out-of-plane) scattering mechanism was left for the future work. Extensive mid-frequency

acoustic propagation modelling and use of 3D HYCOM data provided a realistic model/data comparison as well as a sonar prediction capability. Broadband acoustic model predictions revealed that the primary contribution to the travel-time fluctuations were due to the dynamics of a Polar front in the experimental area. In addition, acoustic modelling with HYCOM data produced amplitude fluctuations comparable to those of measurements. This implied that use of HYCOM would facilitate a prediction capability of acoustic travel-time and amplitude fluctuations near a Polar front.

ACKNOWLEDGEMENTS

This work was supported by the US Office of Naval Research. We thank the PI Ewa Jarosz for the collection and analysis of oceanographic data. The help from the captain and crew of the RV HU Sverdrup II is greatly appreciated.

REFERENCES

[1] **J. F. Lynch, G. Jin, R. Pawlowicz, D. Ray, A. J. Plueddemann, C-S Chiu, J. H. Miller, R. H. Bourke, A. R. Parsons, and R. Muenchuthors,** Acoustic travel-time perturbations due to shallow-water internal waves and internal tides in the Barents Sea Polar Front: Theory and experiment, *J. Acoust. Soc. Am. 99 (2), pp. 803-821, 1996.*

[2] **D. E. Weston,** 'Mechanisms of ocean acoustic attenuation: Scattering by internal solitons, by sea surface waves, and by fish, *J. Acoust. Soc. Am. 92, pp. 3435–3437, 1992.*

[3] **D. Rubenstein and M. H. Brill, '***Acoustic variability due to internal waves and surface waves in shallow water, in Ocean Variability and Acoustic Propagation,* Editors: J. Potter and A. Warn-Varnas, Kluwer Academic, *pp. 215–228, Dordrecht, 1991.*

[4] **R. Bleck,** An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Model 4: pp 55–88, 2002.

[5] **E. Jarosz, A. Turgut, H. Wikesekera, I. Schulman, and C. Luecke,** Observations on dynamics of the Polar Front south of Svalbard, Norway, *AGU Ocean Sciences Meeting,* HE43A-08, Feb 2024, New Orleans, LA.

[6] **I. Schulman, E. Jarosz, S. Cayula1, and E. J. Metzger,** Dynamics of the Polar Front in the southwestern area of Svalbard, Norway, Ocean Dynamics,74, pp 637–653. 2024.

[7] **M. D. Collins, A. Coury A., and W. L. Siegmann,** Beach acoustics. J. Acoust. Soc. Am., 87 97(5), pp. 2767–2770, 1995.

[8] **M. D. Collins and S. A. Chin-Bing,** A three-dimensional parabolic equation model that includes the effects of rough boundaries, J. Acoust. Soc. Am., 87, pp. 1104–1109. 1990.

CLIMATE-DRIVEN VARIABILITY OF UNDERWATER NOISE PROPAGATION IN HORNSUND FJORD, SVALBARD

Pavani Vithana Madugeta Vidanamesthrige[1], Natalia Gorska[1], and Oskar Glowacki[2]

[1]Institute of Oceanology, Polish Academy of Sciences, Powstańców Warszawy 55, 81-712 Sopot, Poland
[2]Institute of Geophysics, Polish Academy of Sciences, Księcia Janusza 64, 01-452 Warszawa, Poland

Pavani Vithana Madugeta Vidanamesthrige, Institute of Oceanology, Polish Academy of Sciences, Powstańców Warszawy 55, 81-712 Sopot, Poland. pavani@iopan.pl

Abstract: *Arctic fjords are particularly sensitive to climate warming due to the accelerated melt of glaciers and sea ice. The resulting meltwater delivery impacts the spatial and temporal variability of sound speed conditions. Changes in sound speed profiles modify underwater noise propagation – a key factor determining noise pollution. Moreover, the sea ice loss causes an increase in shipping traffic, which is a main source of anthropogenic noise. Here, we investigate the long-term variability of noise propagation conditions in Hornsund fjord, Svalbard. The archival CTD data collected from 2001 to 2019 along the fjord axis have been used for sound speed calculation. We used the Bellhop model to predict noise propagation in different years. The results show that the long-term variability of noise propagation in Hornsund is driven primarily by the intensity of glacier melting and shelf-fjord water exchange. Different underwater acoustic channels are identified. We also analyze and discuss the variable transmission loss of the shipping noise and its potential impact on marine mammals. As ocean warming continues, we suggest that the climate-driven changes in the propagation of anthropogenic noise will impact the noise pollution in Hornsund.*

Keywords: *underwater noise propagation, Arctic fjords, climate warming*

1. INTRODUCTION

Arctic fjords are particularly sensitive to climate warming due to the accelerating retreat of glaciers and sea ice melting (e.g., [1]). Processes of ice loss can influence the marine noise pollution in two different ways. First, the sea ice loss causes an increase in shipping traffic, which is a main source of anthropogenic noise in the fjords (e.g., [2]). Second, meltwater delivery modifies the thermohaline structure of the water column, which changes sound speed profiles (e.g., [3]). For example, in the Arctic, cold and less saline water near the surface creates an upward refracting sound speed profile (near-surface sound channel; [4]). Such changes in sound speed profiles influence underwater noise propagation, which is a key factor determining the noise pollution (e.g., [5]). Underwater noise pollution affects marine mammals, causing behavioural disturbance, hearing damage, and masking effects (e.g., [5]). For this reason, it is important to understand the impact of climate warming on underwater noise propagation in Arctic fjords.

Here, we discuss the impact of climate-driven warming on the underwater noise propagation in Hornsund fjord, Svalbard. Two contrasting years are presented to demonstrate this relationship: 2001 and 2014.

2. STUDY AREA

Hornsund is a glacierized fjord located in the southernmost part of West Spitsbergen (red box in Fig. 1). It is approximately 10 km wide and extends about 35 km [6]. Fifteen marine-terminating glaciers flow into Hornsund. Glacier meltwater runoff, resulting from the surface melting of ice, is the major freshwater input to the fjord [7]. Moreover, the fjord's mouth topography facilitates the water exchange with the West Spitsbergen Shelf, specifically water from two different currents: cold and less saline water from the Spitsbergen Polar Current and warm and more saline water from the West Spitsbergen Current (Fig. 1a) [8]. The warming of fjord waters, along with atmospheric warming, intensifies meltwater runoff and submarine glacier melting in the fjord (e.g., [1]).

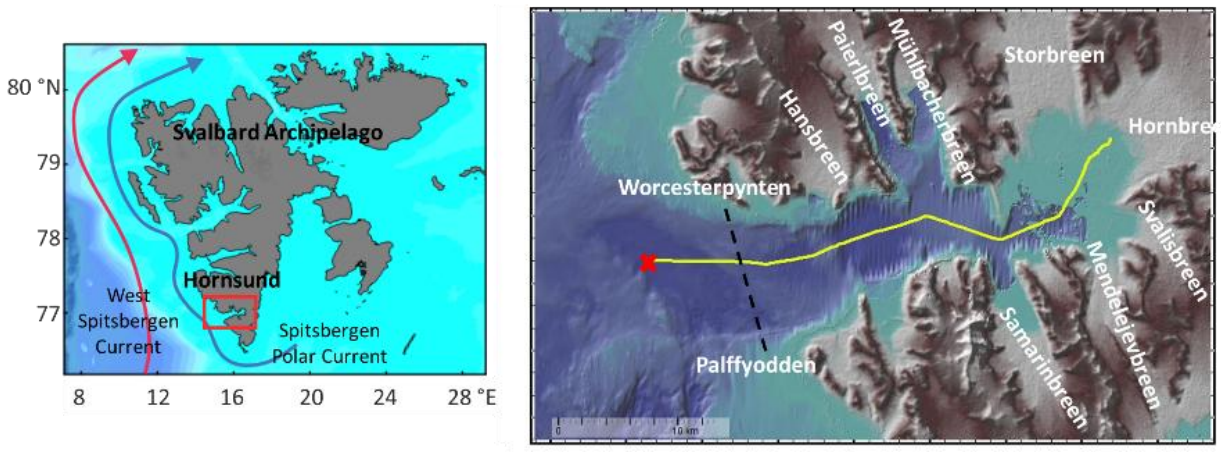


Fig.1: The study area. (a) Svalbard archipelago, the location of Hornsund fjord (red box) and main currents (adopted after Nilsen et al. [9]). (b) Detailed map of Hornsund fjord with positions of the CTD transect (yellow line). The red cross indicates the position of the passenger ship in modelling of noise transmission loss.

In Hornsund, summer is the main shipping season. The ship traffic is dominated by passenger ships, the number of which continuous to grow [10]. The increasing ship traffic certainly results in higher noise levels in the fjord that potentially impact the marine life. Hornsund is home to several cetaceans species [11]. In this study, we limit our analysis to the arctic endemic beluga whale (*Delphinapterus leucas*) and the summer resident minke whale (*Balaenoptera acutorostrata*), both of which are sensitive to shipping noise [12].

3. METHODS

3.1 Sound speed calculation

The CTD data, collected along the fjord axis in late July of 2001 (138 stations) and 2014 (82 stations), were used for sound speed calculation (Fig. 1b). For this purpose, we used Chen and Millero formula adopted by UNESCO [13].

3.2 Transmission loss and received noise level calculation

The beam tracing Bellhop model was used to calculate transmission losses in the study area [14, 15]. Incoherent transmission loss calculations were performed using the Gaussian ray bundle approximation based on the sound speed calculation in 2001 and 2014. The vertical and horizontal resolution of the model were $\lambda/4$ and λ, respectively, where λ is the sound wavelength. The number of beams was set to 28400, with the emitting angles ranging from -80 to 80 degrees with respect to the sea surface. The Thorpe equation was used for the frequency-dependent attenuation in seawater. The homogeneous silt sediment was considered for Hornsund [16]. The seafloor's sound speed and density were set to 1570 m/s and 1.6 g/cm^3, respectively [17]. The frequency-dependent seafloor attenuation was set to 0.1 dB/m/kHz [18]. Bathymetry data for the model were provided by the Norwegian Hydrographic Service. We assumed a flat sea surface. The passenger ship was considered as the noise source located 2 m below the sea surface in the fjord’s mouth (Fig. 1b, red cross). We considered one harmonic component of frequency spectrum of the ship noise at 1 kHz.

The sonar equation was applied to calculate the received noise level (RL) at receiver depth of 5 m below the sea surface along the fjord (see Eq. 1), using 152 dB re 1µPa as the ship's source level (SL) at 1 kHz [19], and transmission loss (TL) calculated from the beam tracing Bellhop model.

$$RL = SL - TL \quad (1)$$

Studying the possible impact of the shipping noise on marine mammals required the consideration of hearing thresholds: 95 dB re 1µPa and 80 dB re 1µPa for belugas and minke whales, respectively [20, 21].The visualizations of the results were done using the MATLAB software (Version 2020b).

4. RESULTS AND DISCUSSION

4.1 Variability of near-surface sound channels in Hornsund

Fig. 2 illustrates two contrasting scenarios of sound speed profiles in Hornsund. In 2001, the barely visible near-surface sound channel was observed only in the easternmost part of the

fjord (pink rectangle in Fig. 2a). In contrast, in 2014, more prominent near-surface sound channel developed along the entire fjord (pink rectangle in Fig. 2b).

The fact that Hornsund is a glacierized fjord is crucial, particularly because cold and fresh glacier meltwater forms a near-surface sound channel in Hornsund during the warmer summer periods, which was demonstrated in Vidanamesthrige et al. [22]. Accordingly, the lack of a prominent near-surface sound channel in 2001 could be due to reduced glacier meltwater input, likely caused by the cold water flowing from the Spitsbergen Polar Current into the fjord and the relatively low air temperature in that year [8, 23]. Meanwhile, it was shown in Vidanamesthrige et al. [22] that in 2014 long and prominent near-surface sound channel was formed as a result of intense glacier meltwater that was almost certainly caused by the inflow of warm water from the West Spitsbergen Current (WSC) into the fjord in combination with atmospheric warming [8, 22].

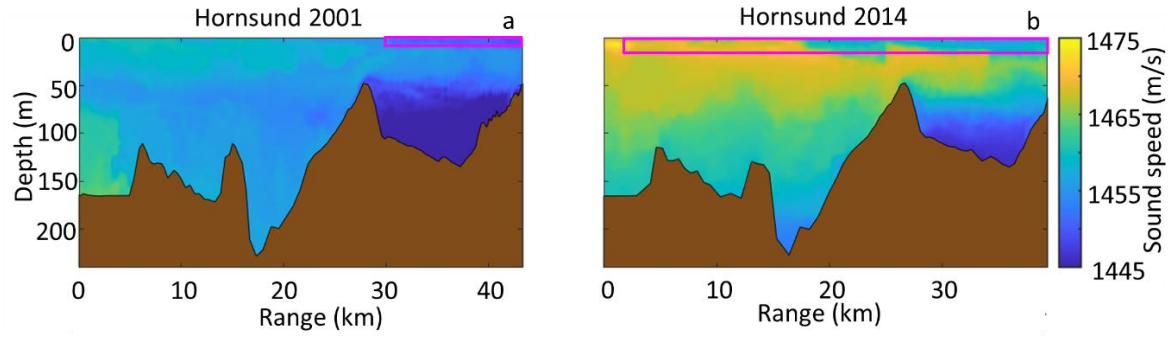


Fig. 2: Sound speed profiles (a and b) along Hornsund in 2001 (left panel) and 2014 (right panel). Pink rectangles indicate the near-surface sound channels.

4.2 Variability of the shipping noise level in Hornsund and its potential consequences

Fig. 3 illustrates the received noise level of the passenger ship along the Hornsund fjord. In 2014, the ship noise level was higher near the surface compared to 2001 (Fig. 3). This higher level resulted from the lower transmission loss due to the more prominent long near-surface sound channel in 2014 (see Fig. 2).

Consequently, in 2014, the ship noise exceeded the hearing threshold of the beluga whale (black dashed line) up to 5 km from the ship, while in 2001 this range was almost two times smaller (~3 km). Moreover, for minke whales, ship noise exceeded their hearing threshold (green dashed line) up to ~23 km from the ship in 2014, which was a larger noise detection range than in 2001 (~17 km). These findings suggest that the prominent near-surface sound channel in 2014 contributed to expanding the ship noise detection range for both species. Consequently, ship noise impact on both species could be higher that year.

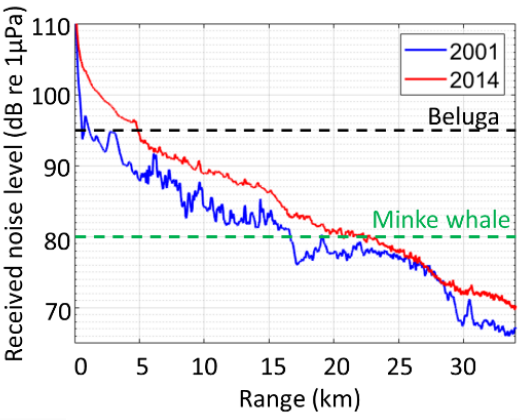


Fig. 3: Range-dependent, received levels of the underwater noise produced at 1 kHz by a passenger ship in Hornsund in 2001 (blue) and 2014 (red). The source and receiver depths were set to 2 and 5 m, respectively. The black and green dashed lines indicate the hearing threshold levels at 1 kHz for the beluga and minke whales, respectively.

5. CONCLUSIONS

The climate-driven variability of near-surface sound channels in Hornsund was evidenced by the lack of a channel in 2001 and its pronounced presence in warmer 2014 year. A more prominent long near-surface sound channel in 2014 resulted in reduced sound transmission loss, leading to increased ship noise level in the fjord compared to 2001. Therefore, the ship noise detection ranges for Beluga and Minke whales were increased near the surface in 2014. The results demonstrate that climate-driven warming of Hornsund can make beluga and minke whales more sensitive to the shipping noise.

ACKNOWLEDGEMENTS

We thank the scientific team of the Observational Oceanography Laboratory of the Physical Oceanography Department of the IOPAS for pre-processed CTD data. This work was supported by the IOPAS (statutory activity) and the Tricity Doctoral School, Poland. OG has been supported by the National Science Centre, Poland (grant 2021/43/D/ST10/00616) and the Ministry of Science and Higher Education of Poland (subsidy for the Institute of Geophysics, Polish Academy of Sciences). The cost of participation in the conference for MVPV was financed by the Polish National Agency for Academic Exchange in STER mTSDPAN project (BPI/STE/2023/1/00008).

REFERENCES

[1] **Błaszczyk, M., et al.**, The Response of Tidewater Glacier Termini Positions in Hornsund (Svalbard) to Climate Forcing, 1992–2020, *Journal of Geophysical Research: Earth Surface*, 128(5), pp. e2022JF006911, 2023.

[2] **PAME**, The Increase in Arctic Shipping 2013-2024, Arctic Shipping Status Report #1, 2025.

[3] **Glowacki, O., Moskalik, M., and Deane, G.B.**, The Impact of Glacier Meltwater on the Underwater Noise Field in a Glacial Bay, *Journal of Geophysical Research: Oceans*, 121(12), pp. 8455-8470, 2016.

[4] **Kutschale, H.**, Arctic Hydroacoustics, *Arctic*, 22(3), pp. 246-264, 1969.
[5] **PAME,** Underwater Noise in the Arctic: A State of Knowledge Report, *Protection of the Arctic Marine Environment (PAME) International Secretariat*, 2019.
[6] **Moskalik, M., et al.**, Bathymetry and Geographical Regionalization of Brepollen (Hornsund, Spitsbergen) Based on Bathymetric Profiles Interpolations, *Polish Polar Research*, 34(1), pp. 1-22, 2013.
[7] **Błaszczyk, M., et al.**, Freshwater Input to the Arctic Fjord Hornsund (Svalbard), *Polar Research*, 38, 2019.
[8] **Promińska, A., Falck, E., and Walczowski, W.**, Interannual Variability in Hydrography and Water Mass Distribution in Hornsund, an Arctic Fjord in Svalbard, *Polar Research*, 37(1), pp. 1495546, 2018.
[9] **Nilsen, F., et al.**, A Simple Shelf Circulation Model: Intrusion of Atlantic Water on the West Spitsbergen Shelf, *Journal of Physical Oceanography*, 46(4), pp. 1209-1230, 2016.
[10] **Kystdatahuset**, *Available from: https://kystdatahuset.no/*, 2025.
[11] **Bengtsson, O., Lydersen, C., and Kovacs, K.M.**, Cetacean Spatial Trends from 2005 to 2019 in Svalbard, Norway, *Polar Research*, 41, 2022.
[12] **Erbe, C., et al.**, The Effects of Ship Noise on Marine Mammals—a Review, *Frontiers in Marine Science*, 6, 2019.
[13] **Chen, C.-T. and Millero, F.J.**, Sound Speed in Seawater at High Pressures, *J. Acoust. Soc. Am*, 62(3), pp. 1129-1135, 1977.
[14] **Porter, M.B.**, The Bellhop Manual and User's Guide: Preliminary Draft, *Heat, Light, and Sound Research, Inc., La Jolla, CA, USA, Tech. Rep*, 260, 2011.
[15] **Porter, M.B. and Bucker, H.P.**, Gaussian Beam Tracing for Computing Ocean Acoustic Fields, *J. Acoust. Soc. Am.*, 82(4), pp. 1349-1359, 1987.
[16] **Drewnik, A., et al.**, From the Worm's Point of View. I: Environmental Settings of Benthic Ecosystems in Arctic Fjord (Hornsund, Spitsbergen), *Polar Biology*, 39(8), pp. 1411-1424, 2016.
[17] **Hamilton, E.L. and Bachman, R.T.**, Sound Velocity and Related Properties of Marine Sediments, *J. Acoust. Soc. Am.*, 72(6), pp. 1891-1904, 1982.
[18] **Hamilton, E.L.**, Compressional-Wave Attenuation in Marine Sediments, *Geophyics*, 37(4), pp. 620-646, 1972.
[19] **Veirs, S., Veirs, V., and Wood, J.D.**, Ship Noise Extends to Frequencies Used for Echolocation by Endangered Killer Whales, *PeerJ*, 4, pp. e1657, 2016.
[20] **Erbe, C., et al.**, Communication Masking in Marine Mammals: A Review and Research Strategy, *Mar Pollut Bull*, 103(1-2), pp. 15-38, 2016.
[21] **Southall, B.L., et al.**, Marine Mammal Noise Exposure Criteria: Updated Scientific Recommendations for Residual Hearing Effects, *Aquatic Mammals*, 45(2), pp. 125-232, 2019.
[22] **Vidanamesthrige, P.V.M., Gorska, N., and Głowacki, O.**, Long-Term Variability of Sound Speed Conditions in Hornsund Fjord, Svalbard, *Oceanologia*, under positive review.
[23] **Wawrzyniak, T. and Osuch, M.**, A 40-Year High Arctic Climatological Dataset of the Polish Polar Station Hornsund (Sw Spitsbergen, Svalbard), *Earth System Science Data*, 12(2), pp. 805-815, 2020.

Structured Session
03. Advances in acoustic measurement systems: Technologies and applications

Organised by Alessandra Tesei, Matthew Schinault and Purnima Ratilal

Insights and Feedback from Long-Term Deployment of an Automated Acoustic Monitoring System for Marine Mammal Detection

Samuel Antoine[1]

[1]RTSYS, ZA de Kerloïc, Rue Jean-Noël Jégo, 56850 Caudan, France

Samuel Antoine, RTSYS, ZA de Kerloïc, Rue Jean-Noël Jégo, 56850 Caudan, France.
Email : santoine@rtsys.fr

Abstract: *Rtsys, a French industrial SME specializing in underwater acoustic instrumentation, has leveraged the recent advancements in deep neural network algorithms and embedded processors to pioneer real-time, onboard advanced sound signal analysis. This paper presents our inaugural large-scale implementation of such a system, conducted in collaboration with a client in 2024. The project aimed to ensure continuous acoustic monitoring approximately 40 km offshore, focusing on the automatic detection of marine mammals around an offshore wind farm construction site. The data gathered is crucial for adjusting construction schedules and vessel routes based on animal presence.*

We developed a fleet of seven offshore buoys equipped with acoustic measurement and automatic data processing chains, capable of transmitting results via satellite communication while being energy autonomous. The processed data was stored in a database, accessible to the operator via a custom web interface that also allowed for remote configuration of the buoys. Over the twelve-month project duration, we accumulated over 1000 buoy-days of operation.

The primary challenges encountered false positive detections primarily triggered by environmental and anthropogenic noise sources, and sporadic satellite communication interruptions in the operational area. Despite these challenges, the mission was largely successful: system availability was high, enabling the detection of a significant number of marine mammals. The buoys demonstrated resilience against storms and maintained energy self-sufficiency.

This experience underscores the potential of advanced acoustic monitoring systems in marine environments and highlights the importance of robust design and reliable communication infrastructure. The insights gained from this project will inform future deployments and technological improvements in the field of marine acoustic monitoring.

Keywords: *Marine mammals detection, Passive acoustic monitoring, Neural network, Artificial intelligence, Acoustic surveillance*

1. INTRODUCTION

Automated sound analysis methods have traditionally relied on signal processing techniques such as energy detection and matched filtering. However, these conventional approaches frequently demonstrate limited efficacy in detecting and classifying complex acoustic signals in high-noise marine environments.

Recent advancements in machine learning, particularly convolutional neural networks (CNNs), have shown superior performance in pattern recognition tasks. When first adapted for underwater acoustics, these techniques were primarily employed for offline analysis of large datasets. However, such methods typically require substantial computational resources. Over the past decade, advancements in both algorithmic efficiency and hardware capabilities have significantly reduced operational constraints, including:

- Power consumption requirements
- Processing latency
- Implementation costs

These improvements have enabled the development of embedded, real-time underwater acoustic monitoring systems. Automated processing allows for continuous analysis of large-scale acoustic datasets while minimizing human intervention requirements. By implementing onboard, real-time analysis, these systems can transmit relevant acoustic information through low-bandwidth communication channels (e.g., satellite or radio links).

This represents a significant advancement in marine acoustic monitoring capabilities, enabling continuous surveillance across extensive offshore areas that were previously inaccessible to traditional monitoring methods. The technology has potential applications in both environmental monitoring and maritime security operations.

2. OFFSHORE CONSTRUCTION MONITORING APPLICATION

During offshore construction projects, measures must be implemented to prevent harm to protected species. For the construction of the *Coastal Virginia Offshore Wind (CVOW)* and *Vineyard Wind* projects, our client, RPS (USA), commissioned us to design an acoustic monitoring solution to detect marine mammals.

In this project, acoustic monitoring complements visual surveillance. The goal is to detect these animals to avoid initiating pile driving in their presence, thereby preventing harm from underwater noise and reducing collision risks with vessels.

3. OPERATIONAL CONSTRAINTS

The monitoring system was deployed in offshore construction zones located approximately 40 km from shore, presenting several technical challenges:

- Continuous acoustic data acquisition capabilities
- Automated detection and classification algorithms
- Requirements for long-range data transmission
- Near real-time alert transmission protocols
- Operator interface for detection visualization

4. SYSTEM ARCHITECTURE

The monitoring system was structured into four primary functional components:

1. Acoustic Signal Acquisition Module
2. Automated Detection and Classification System
3. Long-Range Communication Interface
4. Operator Visualization Platform

4.1. Acoustic signal acquisition

The system was designed with the following technical specifications:

- Minimum bandwidth requirement of 64 kHz
- Low self-noise characteristics to maximize detection range
- Optimized power consumption for extended deployment durations

To meet these requirements, the system incorporated a high-frequency hydrophone and a specialized acquisition board designed for marine acoustic applications. This hardware configuration supported GPS-synchronized recording capabilities, enabling temporal alignment across multiple distributed monitoring units.

4.2. Automated signal analysis

The detection system utilized deep learning-based classifiers adapted from computer vision applications. The classifiers were developed and trained by an environmental monitoring and research company specializing in marine bioacoustics, specifically optimized to analyse spectrogram representations of underwater acoustic signals.

The embedded system generates multiple spectrogram representations in real-time using varying Fourier transform parameters. Time-frequency regions of interest are then extracted, pre-processed, and analysed by the neural network architecture.

Given the computational demands of the classification algorithms, the primary acquisition board's processing capabilities were insufficient for real-time operation. Therefore, spectrogram computation and neural network inference were offloaded to a dedicated processing unit, while the primary board handled raw data acquisition and streaming.

4.3. Long-Range Communication

Due to the offshore deployment location, the system incorporated Iridium satellite communication capabilities. This communication method was selected based on its balance between power efficiency and sufficient bandwidth for transmitting compressed acoustic data and detection alerts.

4.4. Operator Interface

The system was designed to provide detection alerts to operators in near-real time of the initial detection, accompanied by supporting evidence for validation purposes. To achieve this requirement, a centralized data processing architecture was implemented, featuring:

- Automated data reception and processing pipeline
- Structured database for detection event storage
- Web-based interface for visualization of acoustic parameters and spectrogram representations

4.5. Offshore buoy

The acquisition, analysis, and communication modules were integrated into buoys designed for open-sea conditions with a weight of about 1.2 tons. To ensure energy autonomy, each buoy was equipped with 400W of solar panels and two lead-acid batteries. For ease of maintenance, all electronics were housed in a single watertight enclosure, allowing for quick removal and replacement. The hard drive was placed in a separate compartment on the buoy's access door for easy swapping at sea.

5. DEPLOYMENT OPERATIONS

The buoys were shipped from France in containers and assembled in Norfolk, Virginia. RPS handled mooring line design and deployment. In early May 2024, the first buoys were deployed offshore of Norfolk, with construction activities beginning in mid-May and lasting until September.

In October 2024, the buoys were retrieved and redeployed for the Vineyard Wind project off Nantucket Island (USA), with operations continuing until December. After a brief pause, the buoy fleet was redeployed in May 2025 offshore of Norfolk, Virginia, for the next phase of the CVOW project, expected to last until September.
Over one year, the cumulative deployment duration reached 1,000 buoy-days.



Figure 1: Deployment of some of the buoys for CVOW, picture credit to RPS TetraTech PSO Group.

6. RESULTS AND CONCLUSION

6.1. Marine Mammal Detections

The buoys successfully detected numerous marine mammals, including dolphins and whales. However, the system experienced false positive detections primarily triggered by environmental and anthropogenic noise sources. Particularly challenging were continuous vessel engine sounds, which exhibited complex frequency and amplitude modulations during engine speed variations. RTSYS plans to improve performance by enhancing mechanical isolation between the buoy and hydrophone and by training neural networks with additional noise data, including vessels engine noise.

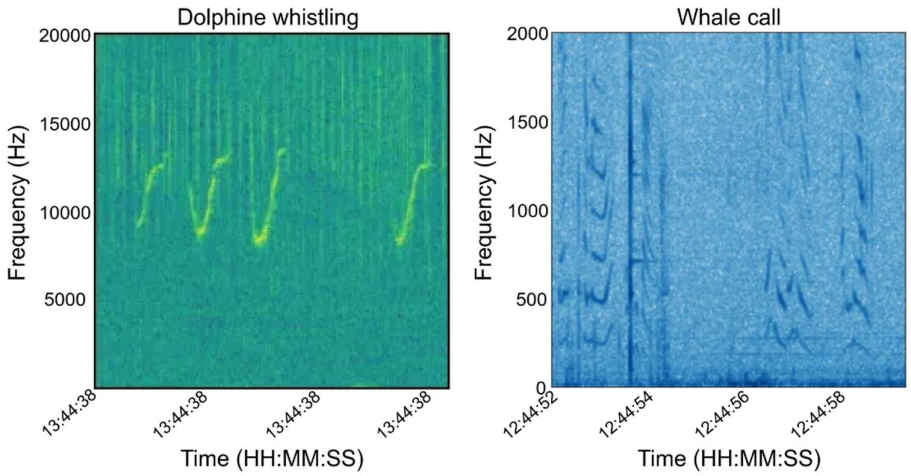


Figure 2: Sample of spectrograms sent by the buoys to the operators during deployment as supporting evidence.

6.2. Reliability

The buoy fleet maintained high availability, allowing construction to proceed as scheduled. Satellite communication proved less stable than expected, with occasional bandwidth limitations even under favorable weather conditions. Message transmission had to be adapted based on available bandwidth, by reducing packet size.

The RubhyAI buoy demonstrated suitability for long-term open-sea deployment due to its robustness, stability, and ample onboard energy. During the Vineyard Wind project, conducted by 41° Nord in a storm-prone area, the buoy fleet maintained high availability.

7. ACKNOWLEDGEMENTS

We extend our gratitude to RPS for their trust in this project and to the crew of the « Go Freedom » for their buoy deployment efforts.

Structured Session
04. Comprehensive Nuclear-Test-Ban Treaty Monitoring, and its Civil and Scientific Applications

Organised by Georgios Haralabus, Mario Zampolli, Tiago Oliveira and Mark Prior

Deep-ocean soundscapes: The impact of atmospheric carbon dioxide

Gianluca Audone[1,2], Philippe Blondel[3], Matthew Nunes[4], and Chris Budd OBE[4]

[1]Dipartimento di Scienze Matematiche "Giuseppe Luigi Lagrange", Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129, Turin, Italy
[2]Gruppo Nazionale per il Calcolo Scientifico INdAM,Piazzale Aldo Moro 5, 00185,Rome, Italy
[3]Department of Physics, University of Bath, Bath, BA2 7AY, UK
[4]Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK

Gianluca Audone, gianluca.audone@polito.it

Abstract: *The Comprehensive Nuclear Test Ban Treaty Organization manages a global International Monitoring System featuring 11 hydroacoustic stations installed in the deep-sea sound channel. Depending on when each station is commissioned, continuous measurements have accumulated up to 20 years of sound pressure data at frequencies reaching 100 Hz. These extended datasets provide a valuable foundation for exploring the long-term impacts of climate variability. This study uses data from CTBT stations Cape Leeuwin in the Indian Ocean and Wake Island in the Pacific Ocean, to investigate the impact of climate change on deep-ocean soundscapes. We analyse long-term acoustic data alongside atmospheric carbon dioxide (CO_2) concentration trends to understand their relationship. Continuous measurements from these stations provide a unique opportunity to assess climatic effects over extended periods. Multi-scale analysis of Sound Pressure Level (SPL) and Hjorth's Mobility is used to correlate acoustic parameters with CO_2 concentrations. This approach allows us to detect seasonal changes and long-term climatic variations. Our findings reveal strong negative correlations between CO_2 trends and both SPL and Hjorth's Mobility at Cape Leeuwin and Wake Island. The Intergovernmental Panel on Climate Change (IPCC) has highlighted the difficulty of gathering comprehensive long-term data to assess the impact of climate change in the deep sea (>1000 m). This research addresses this challenge by linking sound pressure levels at depth with atmospheric CO_2 concentrations. The observed relationships suggest that increasing CO_2 concentrations may be associated with decreasing sound levels and alterations in the frequency content of the underwater soundscape. Sound is an Essential Ocean Variable, and this study emphasises the importance of understanding the Earth's climate system through deep-ocean acoustics.*

Keywords: *Deep-ocean acoustics, Sound Pressure Level, Atmospheric CO_2 coupling, CTBTO IMS hydrophones, Low-frequency absorption*

1. INTRODUCTION

Sound is the primary carrier of information in the sea. Marine mammals, fish, and invertebrates depend on it for communication, navigation, and foraging, while human activities, such as shipping, the extraction of seabed resources and sonar, continuously shape the ambient soundscape [1, 2]. Detecting climate-driven changes within this noisy environment is therefore essential to both safeguarding the ecosystem and regulating anthropogenic actions.

The Intergovernmental Panel on Climate Change (IPCC) reported that warming has been strongest so far in the upper 0–700 m of the ocean; quantifying trends below $1{,}000$ m remains notoriously difficult because sustained observations are sparse [3]. However, the deep-sea sound channel (SOFAR) offers an acoustic window into this blind spot. Because the speed of the sound is coupled to temperature, salinity, and pressure, even subtle hydrographic changes can imprint detectable changes on the propagation of low frequency [1]. Moreover, increasing atmospheric CO_2 alters seawater absorption, reducing attenuation below 100 Hz and potentially lowering ambient Sound Pressure Level (SPL) while reshaping the spectra [4]. Ainslie showed that seasonal Sea Surface Temperature (SST) variability can modulate acoustic levels at mid-latitudes [6]; however, direct long-term links between deep-ocean acoustics and atmospheric climate drivers remain largely unexplored because suitable records are scarce.

The global hydroacoustic network of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) with its eleven IMS stations, deployed in the SOFAR channel and operating quasi-continuously since the early 2000s, now holds up to two decades of broadband data (up to 100 Hz) [7]. In this study we exploit two stations that span contrasting oceanographic regimes: **Cape Leeuwin (HA01)**, a temperate, mid-latitude site sensitive to the Leeuwin Current and Southern Ocean variability, and **Wake Island (HA11)**, a tropical North Pacific site influenced by pronounced seasonal cycles and the El Niño–Southern Oscillation (ENSO).

Our objectives are to:

1. quantify correlations between multiscale acoustic metrics and globally averaged atmospheric CO_2 concentrations from 2004–2023;

2. assess whether geographically distant basins share coherent acoustic responses that may constitute a global fingerprint of climate change.

2. DATA PROCESSING

Acoustic waveforms from CTBTO hydroacoustic stations are sampled at 250 Hz with 24-bit resolution, providing usable frequency content up to 100 Hz with a theoretical dynamic range of 144 dB. Each daily file is accompanied by metadata specifying per-channel sensitivity and gain coefficients; raw analogue-to-digital counts are converted to instantaneous sound pressure by applying these coefficients and the manufacturer-supplied hydrophone inverse response [8]. The resulting pressure series is high-pass filtered at 1 Hz to remove hydrostatic offsets.

Calibrated data are segmented into contiguous 60-s records ($N = 15{,}000$ samples), tapered with a Hann window to reduce spectral leakage, and transformed to the frequency domain using the Fast Fourier Transform (FFT). One-Third-octave band power spectral densities are then computed according to ANSI S1.11-2014 [5]; bands with centre frequencies from 1.12 to 100

Hz are retained. The broadband Sound Pressure Level (SPL) is obtained by

$$\mathrm{SPL} = 10\log_{10}\Big(\sum_k P_k\Big) \quad \text{dB re } 1\,\mu\text{Pa}^2 \tag{1}$$

where P_k denotes the mean-squared sound pressure in the kth one-third octave band.

Hjorth's Mobility denoted by $\mathcal{M}$ is computed as a spectral-moment proxy for mean frequency. Although originally formulated in the time domain, Mobility can be expressed through spectral moments via Parseval's theorem. Let $\hat{p}(f_k)$ be the FFT of a 60-s window and σ_n^2 the variance of the nth temporal derivative; then

$$\mathcal{M} = \sqrt{\sigma_2^2/\sigma_0^2}, \tag{2}$$

where

$$\sigma_0^2, = \sum_k |\hat{p}(f_k)|^2, \qquad \sigma_2^2, = \sum_k \omega_k^2\,|\hat{p}(f_k)|^2$$

with $\omega_k = 2\pi f_k$. Equations (2) make clear that both SPL and Mobility are derived consistently from the calibrated frequency spectrum, ensuring comparability with other studies [11, 12]. The 1-min metrics are then averaged to weekly means, the cadence used throughout the correlation analyses. Weekly aggregation suppresses short impulsive events while retaining intra-annual variability.

Atmospheric CO_2 concentrations are taken from the weekly Global Greenhouse Gas Reference Network product of NOAA's Mauna Loa Observatory[1].

3. METHODOLOGY

The analysis follows these three successive stages —spectral exploration, multi-scale decomposition, and lag-resolved correlation— applied independently to the four one-third-octave bands that visual inspection showed to be most informative: band 13 (19.9 Hz), band 14 (25.1 Hz), band 19 (63.1 Hz) and band 20 (79.4 Hz). Centre frequencies follow the ANSI S1.11 definitions [5].

Spectral exploration. The spectrogram analysis revealed (i) a persistent baleen-whale chorus near 20 Hz, (ii) a broad attenuation feature between 60 Hz and 80 Hz, and (iii) an absence of long-lived instrumental tones. These observations motivated restricting out analyses to the four bands listed above. Figure 1 shows representative spectrograms for each station.

Time-series decomposition. Each acoustic series is decomposed using Seasonal–Trend decomposition with LOESS (STL) [9]. Let $y_{b,t}$ denote either SPL or $\mathcal{M}$ in band b during week t; STL represents

$$y_{b,t} \;=\; T_{b,t} + S_{b,t} + R_{b,t},$$

where $T_{b,t}$ is the long-term trend, $S_{b,t}$ is the seasonal component, and $R_{b,t}$ is the remainder. The same procedure is applied to the weekly atmospheric CO_2 record from Mauna Loa, yielding directly comparable trend and seasonal sub-series. Residual autocorrelation and quantile-quantile (Q-Q) plots confirmed that seasonality had been adequately removed.

[1] `https://gml.noaa.gov/ccgg/trends/mlo.html`

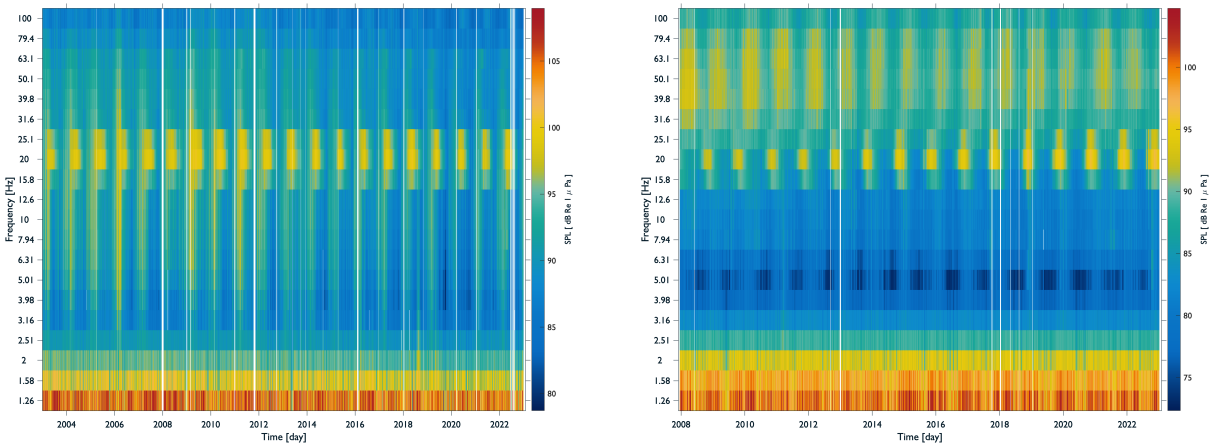


Figure 1: Daily Welch spectrograms for Cape Leeuwin (left) and Wake Island (right). Persistent energy near 20 Hz corresponds to baleen-whale calling, while a basin-scale attenuation feature is visible around 60–80 Hz.

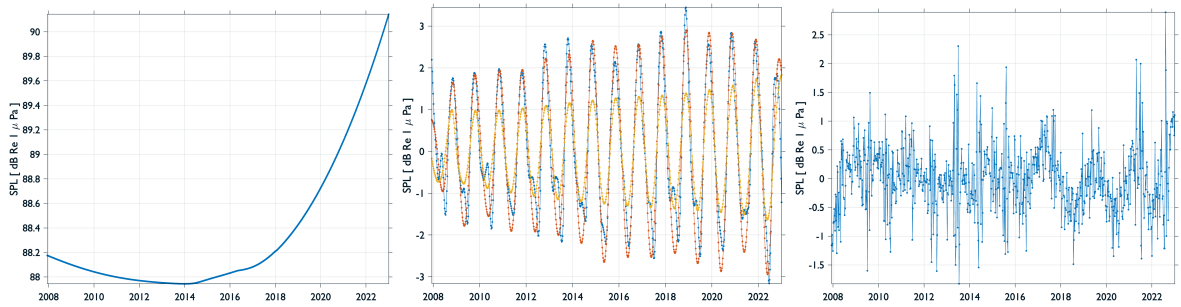


Figure 2: STL decomposition of weekly SPL at Wake Island, band 13. Left to right: long-term trend $T_{13,t}$, seasonal component $S_{13,t}$ (52-week period), and remainder $R_{13,t}$.

Correlation analysis. Long-term associations are quantified with the Pearson product moment coefficient $r(\ell)$ between the trend component of each acoustic metric and the trend in atmospheric CO_2. Phase alignment within the annual cycle is examined by finding the lag that maximises the absolute Pearson correlation between the seasonal components of the acoustic series and the CO_2 record.

This integrated workflow—targeted spectral filtering, STL separation and lag-resolved correlation —provides a compact yet comprehensive basis for detecting and attributing CO_2-driven fingerprints in the deep-ocean acoustic environment while rigorously controlling for seasonality and autocorrelation.

4. ANALYSIS

The weekly time series of Sound Pressure Level (SPL) and Hjorth's Mobility, $\mathcal{M}$, are examined in the four diagnostic one-third-octave bands—13 (19.9 Hz), 14 (25.1 Hz), 19 (79.4 Hz) and 20 (100 Hz), where the numbers in parentheses are the centre frequencies of each band. The analysis follows the procedure set out in Sec. 3. To investigate the statistical significance of the trends we fitted a linear and a quadratic model.

At Cape Leeuwin the decline in SPL is coherent across every frequency considered. Specifically, in band 13, the SPL decreases by 0.15 dB re 1 dB re $1\,\mu\text{Pa}^2$ per year. In band 14, the

decrease is 0.14 dB re 1 dB re $1\,\mu\text{Pa}^2$ per year. Band 19 experiences a decrease of 0.22 dB re 1 dB re $1\,\mu\text{Pa}^2$ per year, and band 20 shows a decrease of 0.16 dB re 1 dB re $1\,\mu\text{Pa}^2$, see Figure 3. These findings indicate a consistent downward trend in SPL across all examined bands. The p-value is equal to zero for all bands. Quadratic models showed no improvements over the linear ones.

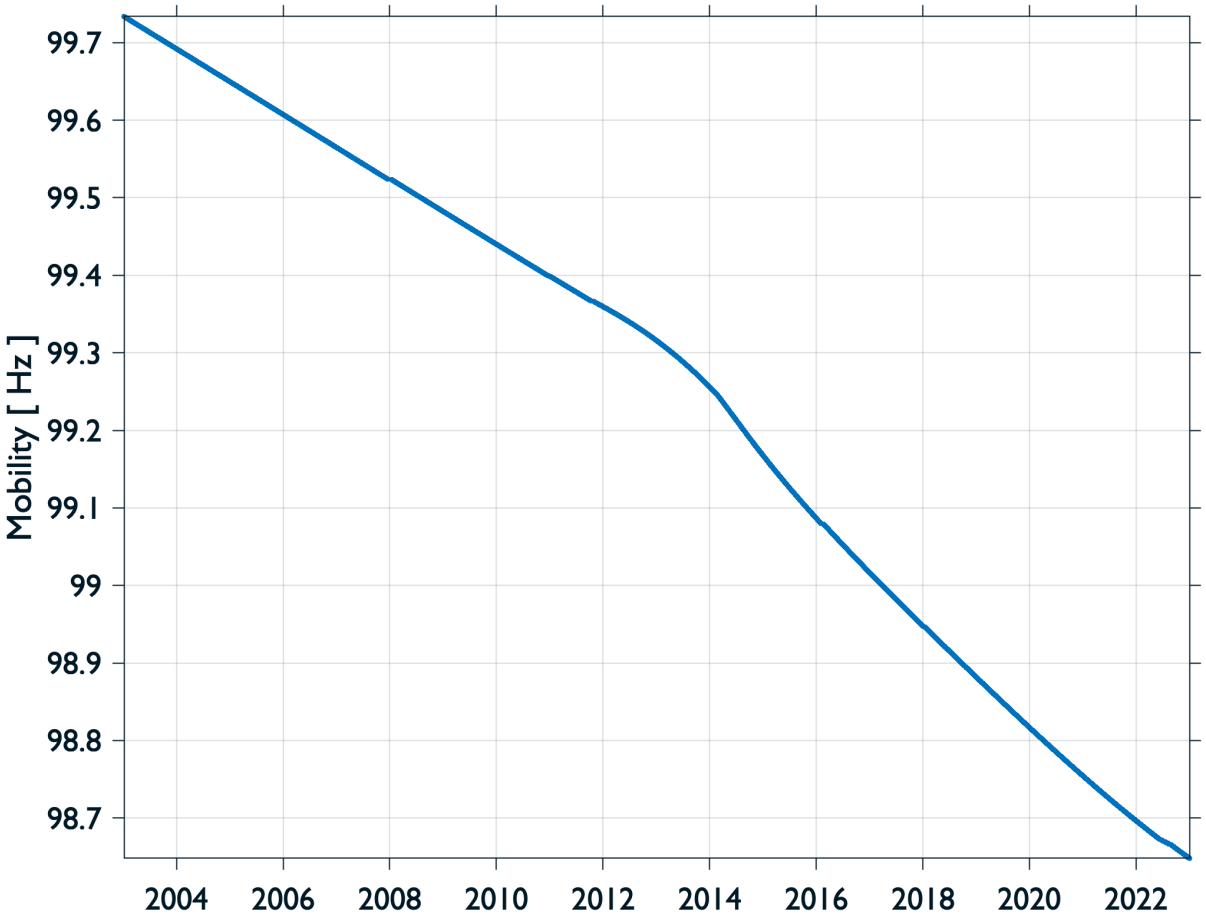


Figure 3: Long-term trend in Hjorth's Mobility band 20 (centre 100 Hz) at Cape Leeuwin (H01W1). This figure displays the trend in Hjorth's Mobility for band 20 at Cape Leeuwin. We point out the strong linear decrease of 1 Hz. This trend provide insights into the long-term evolution variability of underwater sound at this location.

Hjorth's mobility shows a decrease by 0.0017 Hz per year in band 13, while bands 14 shows a decrease of 0.0198 Hz per year. We see an increase of 0.0152 Hz per year in band 19, and a decrease by 0.0572 Hz per year in band 20. These results highlight distinct trends in mobility across different bands, with some bands showing an increase and others a decrease over time. In this case quadratic models gave slightly better fit than the linear models, particularly for bands 13 and 19. This suggests a non-linear relationship in mobility to be explored.

Wake Island Sound Pressure Level trends for bands 13 and 14 are best described by a quadratic model while bands 19 and 20 need more complex model for the trend. A linear model is fitted to have an estimate of the trend. Band 13, the SPL decreases by 0.09 dB re $1\,\mu\text{Pa}^2$ per year. Band 19 experiences a decrease of 0.06 dB re $1\,\mu\text{Pa}^2$ per year, and band 20 shows a decrease of 0.06 dB re $1\,\mu\text{Pa}^2$ per year. These findings indicate a consistent downward trend in SPL across these examined bands. For band 14, the linear model shows a statistically significant increase in SPL over time with a p-value of 0.00016, with an increase of 0.01 dB re $1\,\mu\text{Pa}^2$ per year.

Hjorth's mobility bands 13 is best described by a linear model and we see an increase of 0.0102 Hz per year. All the other bands do not show significant changes to model the trend.

Lag-resolved correlation To quantify the link with atmospheric carbon loading, Pearson coefficients between the acoustic trends and the Mauna Loa CO_2 trend are evaluated. At Cape Leeuwin, SPL shows consistent strong negative correlations across all bands ranging from -0.84 to -0.94. Band 20 is the only to present a lag effect with 2 and a half months. Mobility's trends only shows negative strong correlations (-0.9) for bands 14 and 20 and strong positive correlation (0.96) for band 19. There is no lag effect.

Wake Island shows no lag for both SPL and Mobility. SPL band 13 has correlation of 0.7 while band 14 show -0.7 correlation. Band 20 has -0.9 correlation. Mobility's band 13 has correlation of 0.9, band 14 correlation of -0.8, band 19 of 0.5 and lastly band 20 with 0.3.

Seasonal variability SPL at Cape Leeuwin shows for band 13 and 14 a positive cross correlation of 0.88 between two to three weeks. Bands 19 and 20 show respectively negative and positive correlation with an absolute value of 0.5 at five and two month respectively. Mobility has strong negative correlation -0.8 for band 13 with 3 months lag. Band 20 show negative correlation -0.6 with a month lag, band 19 a weaker positive correlation of 0.47 at 5 month lag. Mobility's band 14 show no correlation.

Wake Island SPL show negative correlation -0.91 with a month lag for band 13, -0.83 with two and a half month lag for band 14 and bands 19 and 20 with -0.77 with four and a half months lag. Mobility has correlation of 0.86 with no lag for band 13, we find the same correlation for band 14 with a lag of one month and a half. Weak correlation -0.3 is found for bands 19 and 20.

5. CONCLUSIONS AND FUTURE WORK

Weekly records from the Cape Leeuwin and Wake Island IMS hydrophones provide a first view of how the deep-ocean acoustic environment may be evolving in a period of rapidly rising atmospheric CO_2. At both sites the 20-year trend in low-frequency (< 100 Hz) Sound Pressure Level is downward, with fitted slopes of roughly -0.2 dB yr^{-1} at Cape Leeuwin and -0.1 dB yr^{-1} at Wake Island, although one biologically dominated band at the latter shows a small increase. Hjorth's Mobility, a proxy for mean spectral frequency, changes in a broadly consistent manner, but its band-to-band diversity cautions that source composition and local conditions also play a role. Correlations with the Mauna Loa CO_2 record are negative and peak at lags close to zero weeks, behaviour that is qualitatively consistent with higher carbon concentrations but that cannot, on its own, exclude alternative explanations such as basin-scale variations in shipping, baleen-whale population dynamics, or instrumentation biases.

No systematic change is detected above about 50 Hz, in line with theory that predicts chemically mediated absorption to be strongest below 100 Hz. Even so, the data originate from only two stations, and their interpretation is complicated by uncertainties in long-term sensor calibration, by incomplete knowledge of regional source statistics and by the possibility of subtle environmental influences (e.g. mesoscale circulation) that fall outside the present analysis framework.

Several extensions could therefore help to clarify the emerging picture. Repeating the band-resolved workflow at the remaining IMS hydrophones—and at other observatories with decade-scale coverage—would test whether the Cape–Wake behaviour is regionally confined or part

of a larger-scale pattern. Combining acoustic records with ship-tracking data, earthquake catalogues and modern source-separation techniques should allow a more explicit attribution of the long-term level change to anthropogenic, geophysical and biological components. Parallel work linking low-frequency trends to independent observations of marine fauna—whether passive-acoustic call detections, fisheries statistics—could shed light on ecological consequences. Finally, embedding calibrated SPL and Mobility diagnostics in coupled ocean – acoustic – biogeochemical models will be essential if deep-ocean sound is to become a quantitative indicator of carbon-cycle change.

Viewed in that context, hydroacoustic archives represent a largely untapped but promising complement to physical and chemical oceanographic datasets.

6. ACKNOWLEDGEMENTS

GA acknowledges the support of the Fondazione Compagnia di San Paolo within the framework of the Artificial Intelligence Call for Proposals, AIxtreme project (ID Rol: 71708). GA was supported by a PhD studentship from the Engineering and Physical Sciences Research Council (EPSRC, UK) to the University of Bath's Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), project #EP/L015684/1m, and by the National Physical Laboratory. GA acknowledges the support of V. Livina (NPL) in downloading and organizing the data, and S.-H. Cheong (NPL) in the initial data processing. The views expressed in the paper are those of the authors and do not necessarily represent those of the CTBTO.

REFERENCES

[1] Woolfe, K. F., Lani, S., Sabra, K. G., and Kuperman, W. A.: "Monitoring deep-ocean temperatures using acoustic ambient noise: passive acoustic tomography in the ocean", *Geophys. Res. Lett.* **42**, 2878–2884 **(2015)**.

[2] Ainslie, M. A. and McColm, J. G.: "A simplified formula for viscous and chemical absorption in sea water", *J. Acoust. Soc. Am.* **103**, 1671–1672 **(1998)**.

[3] Barros, V. R. *et al.*, *Climate Change 2014: Impacts, Adaptation, and Vulnerability*, Contribution of WG II to the Fifth Assessment Report of the IPCC (Cambridge University Press, 2014).

[4] Reeder, D. B. and Chiu, C. S.: "Ocean acidification and its impact on ocean noise: phenomenology and analysis", *J. Acoust. Soc. Am.* **128**, EL137–EL143 **(2010)**.

[5] ANSI/ASA S1.11-2014: "Specification for octave-band and fractional-octave-band analog and digital filters" (Acoustical Society of America, New York, 2014).

[6] Ainslie, M. A., Andrew, R. K., Howe, B. M., and Mercer, J. A.: "Temperature-driven seasonal and longer-term changes in basin-averaged deep-ocean ambient sound at 63–125 Hz", *J. Acoust. Soc. Am.* **149**, 2531–2545 **(2021)**.

[7] Prior, M., Brown, D., Haralabus, G., and Stanley, J.: "Long-term monitoring of ambient noise at CTBTO hydrophone stations", in *Proc. 11th European Conf. on Underwater Acoustics* (Edinburgh, 2012).

[8] Robinson, S. P., Lepper, P. A., and Hazelwood, R. A.: *Good Practice Guide for Underwater Noise Measurement*. National Measurement Office, Marine Scotland, and The Crown Estate. NPL Good Practice Guide No. 133, ISSN 1368–6550, 2014.

[9] Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I.: "STL: a seasonal-trend decomposition procedure based on LOESS", *J. Off. Stat.* **6**, 3–73 **(1990)**.

[10] Brockwell, P. J. and Davis, R. A., *Time Series: Theory and Methods*, 2nd ed. (Springer, 1991).

[11] Hjorth, B.: "EEG analysis based on time-domain properties", *Electroencephalogr. Clin. Neurophysiol.* **29**, 306–310 **(1970)**.

[12] Hjorth, B.: "The physical significance of time-domain descriptors in EEG analysis", *Electroencephalogr. Clin. Neurophysiol.* **34**, 321–325 **(1973)**.

Monitoring underwater eruptions using networks of hydrophones in the SOFAR channel

Sara Bazin[1], Pierre-Yves Raumer[1], Aude Lavayssière[1,2], Hugo Fauvel[1,3], Romain Safran[1], Vaibhav V. Ingale[1,4], Jonathan Tanrin[1], Julien Vergoz[3], Jean-Yves Royer[1], Jun Korenaga[5]

[1]Geo-Ocean, University of Brest, CNRS, Ifremer, F-29280, Plouzané, France
[2]now at Géoazur, University Côte d'Azur, IRD, CNRS, F-06560, Valbonne, France
[3]CEA, DAM, DIF, F-91297, Arpajon, France
[4]now at IGPP, Scripps Inst. of Oceanography, Univ of California, San Diego, CA, USA
[5]Dept. of Earth and Planetary Sciences, Yale University, New Haven, CT, USA

Contact author: Sara Bazin (sara.bazin@univ-brest.fr)

Abstract: *Most volcanic eruptions occurring in the ocean remain undetected. Hydroacoustic monitoring, a widely used method among seismologists due to its sensitivity to T-phases associated with the low-level seismic activity in remote oceanic regions, can address this knowledge gap. Indeed, near volcanic areas, such as spreading ridges or undersea volcanoes, acoustic data recorded by moored hydrophones display numerous impulsive signals of short durations (5-15 s) and high energy (up to 100 Hz), that we attribute to interactions between hot lava and cold seawater. Most likely, thermal cracking is the process that generates these signals, which can be used as proxies to detect and locate new lava flows. These impulsive events were found in three different networks located in the Indian Ocean: (i) the large-scale OHA-SISBIO network, composed of 6-9 autonomous hydrophones encompassing 1000-km-long sections of the three Indian mid-oceanic ridges; (ii) the small-scale MAHY network, with 4 autonomous hydrophones 50 km away from a new-born volcano off Mayotte Island, Mozambique Channel; and (iii) the hydroacoustic stations of the International Monitoring System, consisting of cabled-to-shore triplets of hydrophones. In all these areas and data, we identified clusters of impulsive events, within clusters of seismic events, related to volcanic activity. To further optimize eruption monitoring, we developed automatic AI-based detection, classification and localisation tools. Implementing such automatic detection of volcanic events in real-time hydrophones, as the IMS stations, would be a major advance for organizing in-situ studies of active submarine volcanoes. When real-time stations are not available, we deployed innovative autonomous hydrophones equipped with small messengers to enable data retrieval from a small vessel of opportunity, should a major cluster of events occur in a monitored area.*

Keywords: *Hydroacoustics, Volcanic activity, Impulsive events, Thermal cracking.*

1. INTRODUCTION

Underwater earthquakes generate T-waves by conversion of seismic energy into low-frequency hydroacoustic energy at the seafloor. Recording T-waves with a network of hydrophones has been widely used by seismologists since the seminal work of Fox et al. [1], particularly for monitoring the low-level seismic activity along mid-ocean ridges [e.g. 2]. This is because they can propagate in the SOFAR channel over thousands of km and their sources can be localized by trilateration from arrival-times detected on at least 3 hydrophones. T-wave sources do not necessarily match epicenters but seafloor acoustic radiators of unknown size. However, in the context of spreading ridges where diking and faulting events are shallow, they provide a very useful and precise indication on where active processes take place (Fig. 1). T-waves commonly last several minutes (Fig. 2a), but hydroacoustic data also display short and energetic impulsive signals (5-15s), indicating that their energy (up to 100 Hz) is released directly into the water (H-waves, Fig. 2b) and has not travelled into the solid crust. We attribute these impulsive events to interactions between lava and seawater that can be used to track submarine eruption.

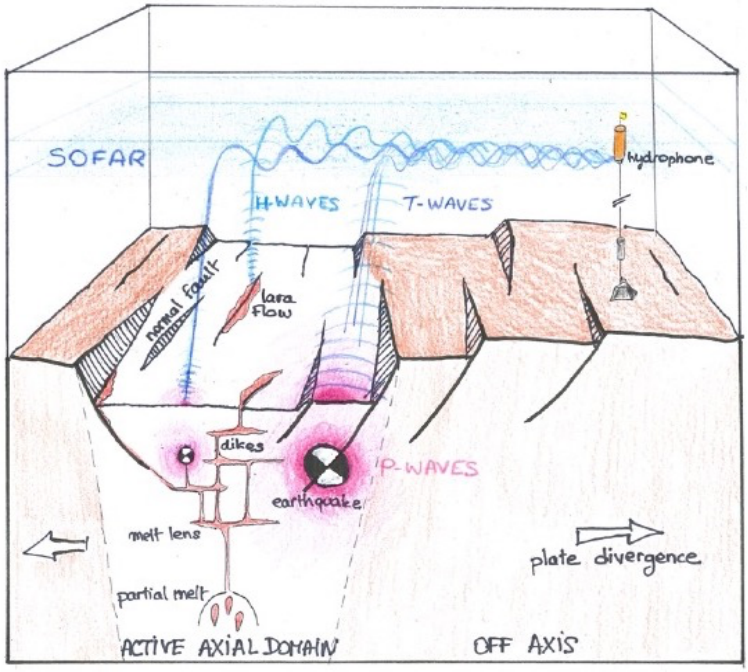


Fig.1: Mid-ocean ridges acoustic waves that can be captured by moored hydrophones.

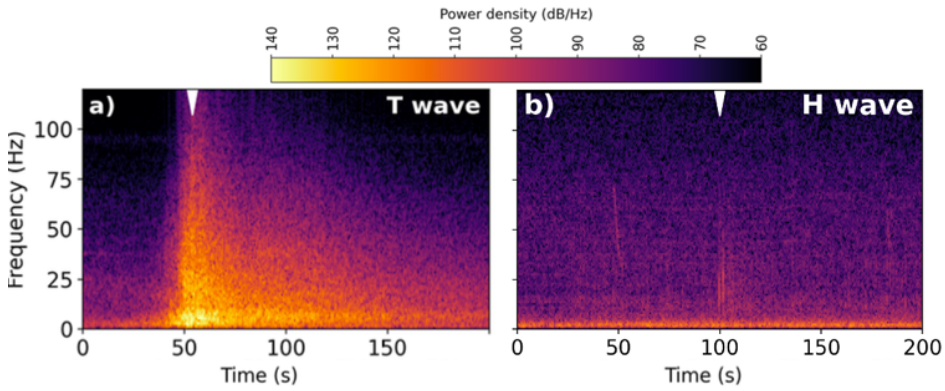


Fig.2: Spectrograms of a T- and a H-wave detected at mid-ocean ridges, after [3].

2. INSIGHTS FROM LONG-TERM MONITORING OF MID-OCEANS RIDGES

Mid-ocean ridges create and shape the ocean floor through a complex interplay of volcanic and tectonic processes, which remain difficult to unravel when the low-level seismicity is lacking. Investigating seismic clusters by T-wave monitoring helps fill this gap. For instance, along the 3 Indian mid-oceanic ridges, we identified 94 seismic clusters in the 2010 to 2020 land-based catalogue from the International Seismological Center (ISC; Fig. 3). Among them, we selected the 15 biggest clusters, spanning over the 3 contrasting ridges (ultra-slow, slow and intermediate spreading rates). The analysis of these large seismic clusters with continuous hydroacoustic data, recorded from 2012 to 2020 by the large-scale OHA-SISBIO network, yielded a catalogue of 38,910 hydroacoustic events [2], 50 times larger than the ISC catalogue, and with an average location accuracy in the order of 400 m. The spatio-temporal distribution of events within each swarm reflects its origin, either magmatic, tectonic, or a complex interplay of both. Five out of the 15 selected swarms included clusters of impulsive events, interpreted as submarine eruptions. However, they lack in situ evidence to confirm them.

T-wave sources can also be characterized by their Source Level (SL), which measures their acoustic level in dB re 1 µPa @1m. It is estimated from the sound levels recorded at the receivers (RL) corrected for the transmission loss (TL) along the acoustic paths. Comparisons of SLs with seismic magnitudes yield empirical relationships between these parameters (for example SL = 6.25 mb + 186 along the Southwest Indian Ridge [2]).

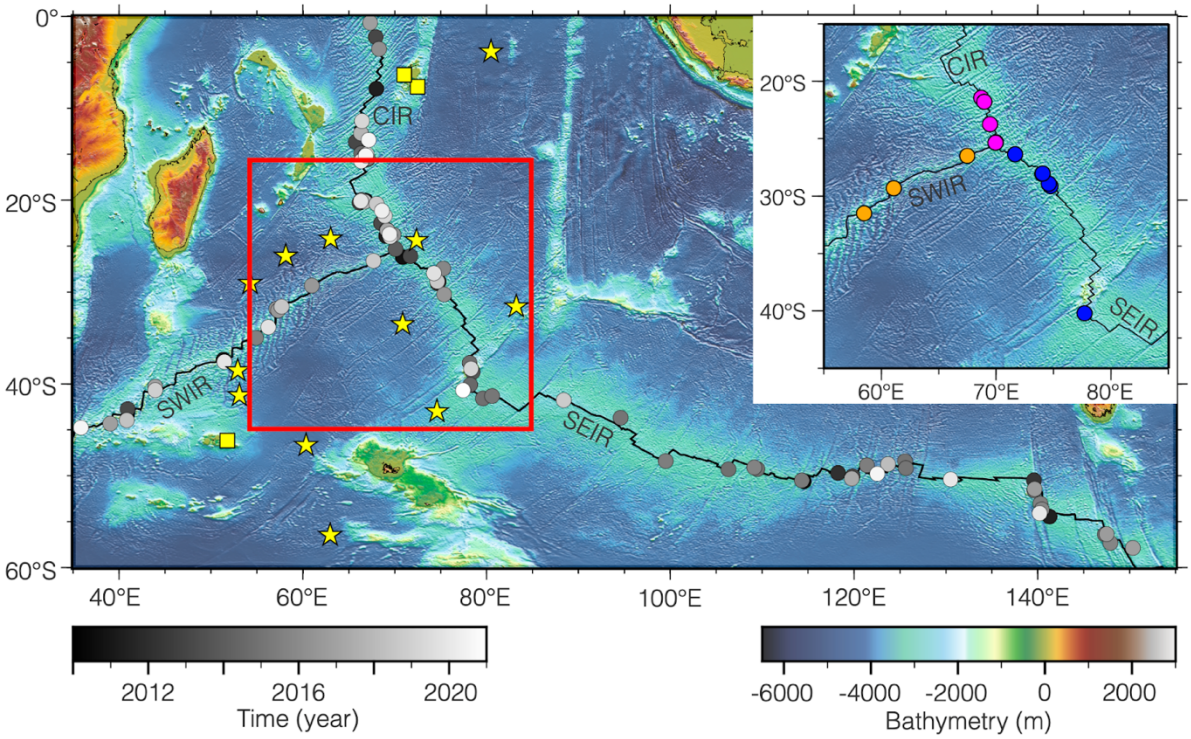


Fig.3: 2010-2020 ISC clusters along the 3 Indian ridges shaded with respect to time. Among them, 15 clusters (inset) along the ultra-slow Southwest (orange), slow Central (pink), and intermediate Southeast (blue) Indian ridges were analyzed using the OHA-SISBIO (yellow stars) and IMS (yellow squares) hydrophone networks [2].

3. INSIGHTS FROM SMALL-SCALE ACOUSTIC MONITORING OF AN UNDERWATER VOLCANO

In 2019, a new volcano was discovered at 3500 m depth, 50 km east of Mayotte Island in the Mozambique Channel. The intense seismicity associated with this event is monitored by the REVOSIMA (MAyotte VOlcano and Seismic Monitoring) network. Since 2020, it includes the MAHY (MAyotte HYdroaoustic) array with 4 hydrophones moored in the SOFAR channel for monitoring sounds generated by the volcanic activity. It indeed detected several clusters of impulsive signals that we interpret as lava flow emplacements, confirmed by a towed-camera transect over glossy lava fields (Fig. 4). The events occurred in swarms of 4 to 7 days, and stopped after December 4, 2020, which would end the eruption [4]. Based on the lava thickness derived from successive ship-borne multibeam bathymetric surveys and the duration of the lava flow events, the average effusion rate can be estimated to 14 m^3/s, compatible with that derived from land-GPS data and deflation rate over time. Within the northern cluster, events migrated towards the Northeast at a velocity of ~3 m/h, probably representative of the end of the eruptive phase. In addition, a 40 s-long and low-amplitude signal recorded by all 4 hydrophones on October 25, 2020 is interpreted as a submarine landslide, moving down-slope at a speed of ~26 m/s from the top of a lava pile over a distance of nearly 1 km.

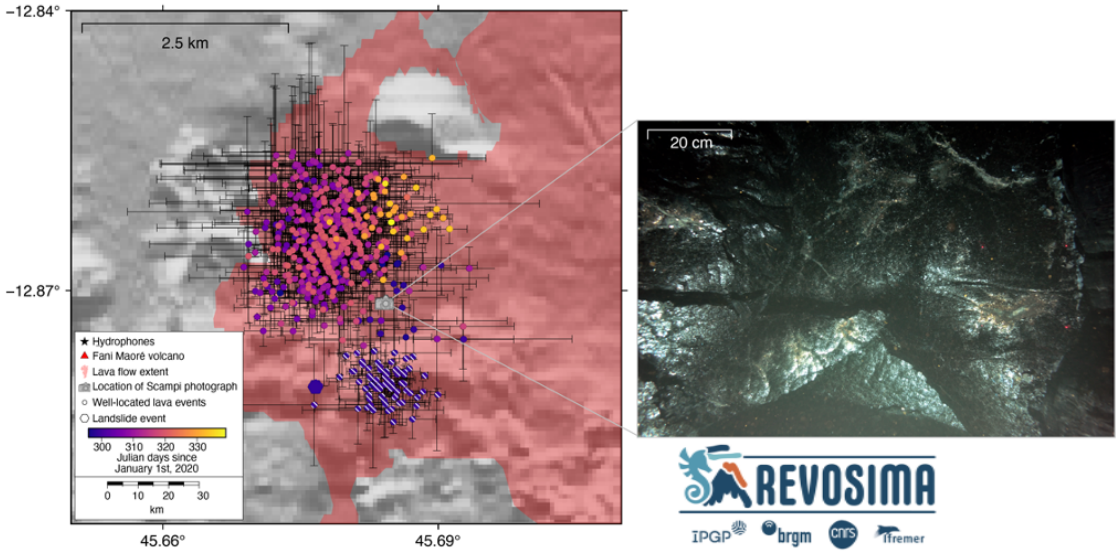


Fig.4: Spatio-temporal evolution of the impulsive events near the Fani Maoré volcano. An underwater photo reveals fresh lava flows associated with ~5-cm wide cracks, from [4].

4. NEW TOOLS FOR PASSIVE ACOUSTIC MONITORING

To optimize the hydroacoustic data analysis, we implemented automated algorithms based on machine learning to detect and identify T- and H- phases in hydrophone spectrograms. This Time Spectrogram Segmentation Network (TiSSNet) model is designed to process spectrograms of arbitrary duration with 128 bins of frequency as inputs and provides a probability of an event occurring at each time step as output [3]. Along with the TiSSNet modules and notebooks, available on GitHub, a common benchmarking framework for the community is also provided to evaluate and compare performances of

other hydroacoustic detection tools. This annotated data set will be further improved and updated. Future developments include automatic trilateration using various hydrophone network geometries.

In addition, we are implementing an automatic categorization tool by type of source for the 6 real-time IMS stations (Fig. 6, [7]). Major volcanic events could thus be detected in near real-time, facilitating rapid intervention studies on active undersea volcanoes and enabling new scientific discoveries in the field of volcanology.

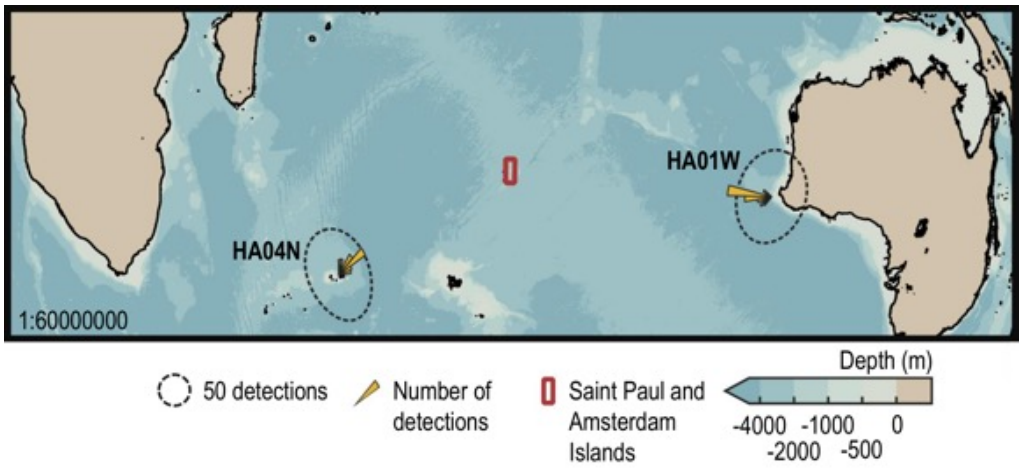


Fig.6: Real-time detection of major in April 2024 by IMS triplets at Cape Leewin and Crozet Isl. We since confirmed a submarine eruption along the Southeast Indian ridge.

Data recorded by the French pool of moored hydrophones will soon be distributed, using standardized FDSN Web Services, through the open database of the European Plate Observing System. Specific metadata regarding the data format and units, and internal clock drift are now implemented in the Obsinfo toolbox (in GitHub) in the hope that it will foster hydroacoustic data sharing among a broader scientific community [5].

Finally, to facilitate access to data recorded by autonomous hydrophones, should a major event of interest occur during a monitoring period, we developed a new hydrophone equipped with releasable glass spheres to allow data retrieval from any type of boat (Fig. 5, [6]). Three of them are currently deployed off Mayotte Island.

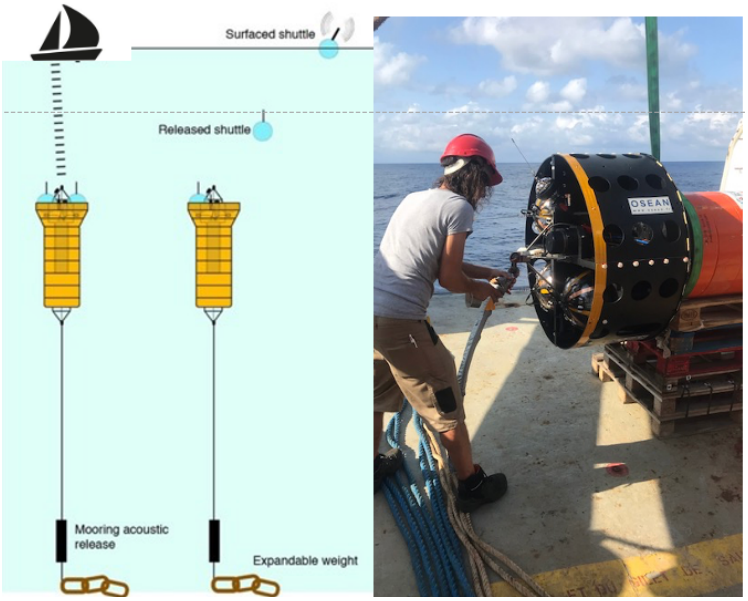


Fig.5: Autonomous HYDROBS mooring equipped with data messengers.

5. QUANTIFICATION OF THERMAL CRACKING IN LAVA

The lava events have an average SL of ~190 dB, similar to the energy released by a 300 in^3 seismic air-gun at 140 bar (~50 kJ, [4]). Several mechanisms have been proposed to explain such impulsive events observed during underwater eruptions [8]: collapse of lava tubes, steam bursts, or lava quenching. In Fig. 4, the photographed lava surface is glossy, without any fragmented pieces that an explosion would produce. There are no bubble effects in the spectrograms and the seafloor is below gas critical-depth which excludes the formation and implosion of bubbles. We therefore favor thermal cracking while fresh lava quenches. The elastic strain energy density released by a thermal crack can be estimated with Young's modulus (E~100 GPa), the maximum tensile strength (σ_T ~10 MPa), and Poisson ratio (ν) in basalt: $U=\sigma_T^2/E(1-\nu)$ ~1 kJ/m^3. A lava volume of ~50 m^3 undergoing thermal cracking could then generate an impulsive event releasing 50 kJ of acoustic energy. This rough estimate is consistent with in situ observations.

6. ACKNOWLEDGEMENTS

Our data comes from the OHASISBIO (doi 10.18142/229) and MAHY (doi 10.15778/RESIF.1T2018) networks. We thank the Preparatory Commission for the Comprehensive Nuclear Test Ban Treaty Organization for providing access to the IMS data; the Commission is not responsible for the views expressed by the authors.

REFERENCES

[1] **Fox, C. G., H. Matsumoto, and T. K. Lau**, Monitoring Pacific Ocean seismicity from an autonomous hydrophone array, *J. Geophys. Res.,* 106, 4183–4206, 2001.

[2] **Ingale, V. V., S. Bazin, J.-A. Olive, A. Briais, and J.-Y. Royer**, Hydroacoustic study of a seismic swarm in 2016–2017 near the Melville transform fault on the Southwest Indian Ridge, *Bull. Seis. Soc. Am.*, 1-19, 2023.

[3] **Raumer, P.Y., S. Bazin, D. Cazau, V.V. Ingale, A. Lavayssière and J.-Y. Royer**, An open source hydroacoustic benchmarking framework for geophonic signal detection. *Seismica*, doi 10.26443/seismica.v3i2.1344, 2024.

[4] **Lavayssière, A., S. Bazin, and J.-Y. Royer**, Hydroacoustic Monitoring of Mayotte Submarine Volcano during Its Eruptive Phase. *Geosciences*, doi 10.3390/geosciences14060170, 2024.

[5] **Crawford, W., M. Schmidt-Aursch, and the DSN OBS standard action group**, Station metadata curation for amphibian experiments. *ORFEUS & Geo-INQUIRE Workshop*, Helsinki, 2024.

[6] **Royer, J.-Y., Y. Hello, and M. Yegikyan, 2017.** HYDROBS: a long-term autonomous mooring for passive acoustic monitoring. *AGU Fall Meeting, Am. Geophys. Union*. doi: https://agu.confex.com/agu/fm17/meetingapp.cgi/Paper/249119

[7] **Fauvel, H., S. Oger, S. Bazin, D. Cazau, and J. Vergoz**, Automatic identification of sources detected by the hydroacoustic stations of the International Monitoring. *Science and Technology conference*, CTBTO, Vienna, Austria, 2025.

[8] **Tepp, G., and R. P. Dziak**, The seismo-acoustics of submarine volcanic eruptions, *J. Geophys. Res. Solid Earth,* 126, e2020JB020912, 2021.

Structured Session
05. Distributed Fiber-Optic Sensing technology for underwater acoustical monitoring

Organised by Alexander Gavrilov, Evgenii Sidenko and Hefeng Dong

Analysis of passive DAS data for observation of microseisms and extracting Scholte wave dispersion in shallow water

Hefeng Dong[1], and Michael Taroudakis[2]

[1]Norwegian University of Science and Technology, NTNU, NO7491Trondheim, Norway
[2]University of Crete, and FORTH Vassilika Vouton 70013 Heraklion, Crete, Greece

Contact author: Hefeng Dong, Department of Electronics, NTNU, NO-7491, Trondheim, Norway. hefeng.dong@ntnu.no

Abstract: *In recent years, the emerging technology of Distributed Acoustic Sensing (DAS), measuring the strain in a fibre-optic (FO) cable caused by acoustic waves, has been used for monitoring and imaging of trains, pipelines, earthquakes, whale and oil fields, etc. DAS systems have the advantage over traditional receiving systems since it can provide densely sampled array along the FO cable and wide-band signals, especially low frequency response. In this study, A patch of two-hour continuous strain data from the Tampnet DAS dataset released from "The Global DAS Month of February 2023 is analysed. The ambient noise is dominated by periodic shoaling ocean surface waves and contains the interaction of ocean oppositely propagated ocean wind-waves. The primary and secondary microseisms are separated by frequency-wavenumber decomposition due to the good low-frequency response of the DAS system. Greens' functions from the passive strain data are retrieved by passive seismic interferometry. Dispersive Scholte waves are extracted from the Green's function using time-frequency analysis. The shear-wave velocity profile in the upper sediment layer is estimated by inverting the dispersion curve of the Scholte wave in a nonlinear geoacoustic inversion algorithm.*

Keywords: *Distributed acoustic sensing, microseisms, frequency-wavenumber analysis, Scholte wave dispersion, shear wave velocity*

1. INTRODUCTION

In recent years, the emerging technology of Distributed Acoustic Sensing (DAS), measuring the strain in a fibre-optic (FO) cable down to nano strain levels caused by acoustic waves, has been used for monitoring and imaging of trains, pipelines, earthquakes, whale and oil fields, etc. DAS systems have the advantage over traditional receiving systems since it can provide densely sampled array along the FO cable and wide-band signals, especially low frequency response.

DAS data are generated by an interrogator that transmits laser pulses into a fibre and interrogates the Rayleigh backscattering caused by density fluctuations in the fibre. These density fluctuations are detected as phase changes in the backscattered light. The interrogator calculates the time-differentiated phase change of the backscattered response from consecutive sweeps and converts it to the longitudinal strain of the corresponding fibre position. The number of channels over which the phase differentiation is performed is called the gauge length (GL) [1].

DAS technology has been applied to many disciplines both on land and in underwater, including sub-surface monitoring, earthquake seismology, geophysics exploration, passive acoustic monitoring of ships and monitoring and tracking whales [2-6]. The existing FO telecom cable deployed on the seafloor along the coastal area can be used for monitoring underwater infrastructure and underwater noise generated by ships and marine life. A few km long DAS array can detect ocean surface gravity waves in the coastal environment, sense the dynamic pressure produced by the surface gravity wave down to the seafloor, and record seafloor-bounded Scholte waves [3].

In this paper, a patch of two-hour continuous strain data from the Tampnet DAS dataset released from “The Global DAS Month of February 2023” [7] are analysed. The good low-frequency response of DAS system in shallow water is demonstrated by frequency-wavenumber analysis. The primary and secondary microseisms are generated by ocean surface gravity waves and the interaction of oppositely propagated wind-waves reaching the seafloor to convert into seafloor-bound Scholte waves are observed. The Scholte wave is obtained from the Green’s function retrieved by passive seismic interferometry and the shear wave velocity profile is estimated by using the dispersion property of the Scholte wave.

2. DAS DATASET

DAS dataset used in this study is from “The Global DAS Month of February 2023” [7]. It was a Global DAS Month campaign and a total of 32 individual DAS systems acted jointly as a global seismic monitoring network and 32 sub-datasets were generated for free download. The Tampnet sub-dataset is selected in this study. The dataset was recorded by OptoDAS interrogator from Alcatel Submarine Networks (ASN) from a telecom cable with landing in Lowestoft, UK and extended NE to the Southern North Sea. The FO cable layout and the bathymetry of the area are illustrated in Fig. 1. The recorded strain data were from the first 80 km of the cable deployed at a water depth from 20 m to 50 m. The related parameters of the dataset are listed in Table 1. The dataset is not entirely continuous in time but contains patches of continuous data related to seismic events. A patch of around 2-hour continuous strain data recorded on 10th of February is selected and analysed.

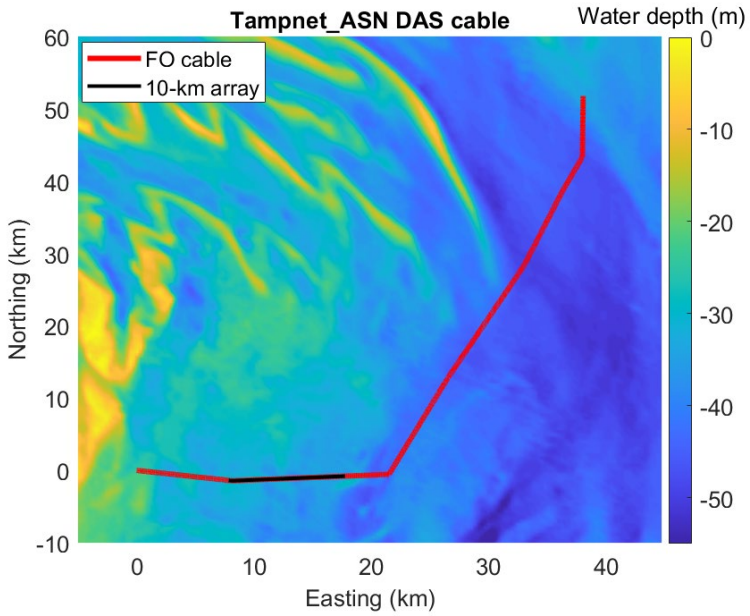


Fig.1: Tampnet FO cable layout and the bathymetry.

Parameter	Value
FO cable length [m]	80,000
Gaugh length [m]	20
Channel spacing [m]	20
Number of channels	40000
Sampling frequency [Hz]	125

Table 1: DAS dataset parameters.

3. DAS DATA PROCESSING AND RESULTS

1) Detection of Ocean surface waves

Using a long DAS array perpendicular to the coastline, the good low-frequency response of the DAS system can be used to detect the ocean surface gravity waves. Fig. 2 (a) plots 5-minute raw stain data from a 10 km DAS array (black part in Fig. 1) which is perpendicular to the shoreline. It shows periodic oscillations mainly propagating towards the shore. PM and SM are separated by a frequency-wavenumber decomposition as shown in Fig. 2 (b). Figure 3 is an enlarged version of Fig. 2 (b) at two different frequency bands. PM in Fig. 3 (a) shows the asymmetric dispersive wave trains in the frequency band between 0.02 and 0.5 Hz, propagating landward with stronger amplitude and oceanward with weaker amplitude that corresponds to a reflection off the shore. The red curve in Fig. 2 (b) and Fig. 3 (a) is the theoretical dispersion curve of the linear gravity wave in shallow water with a depth of h = 30 m given in (1).

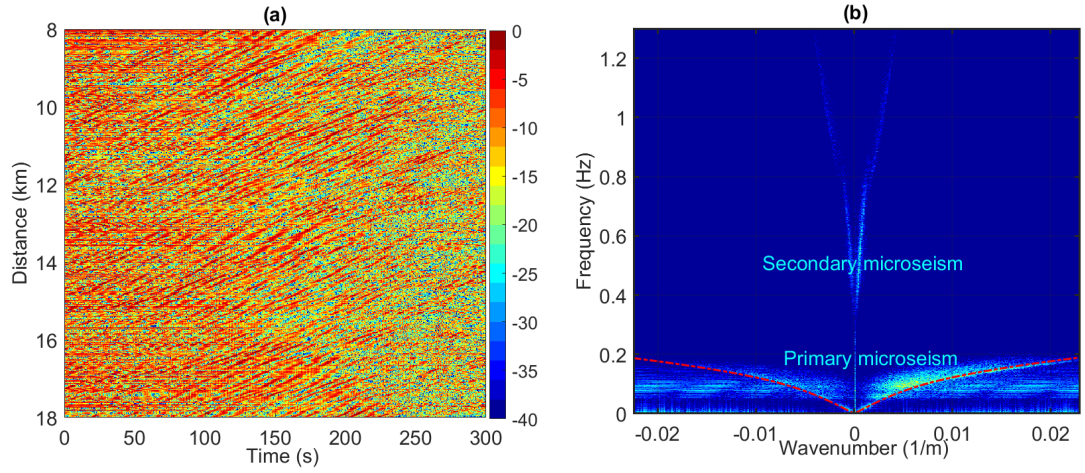


Fig.2: Microseism analysis of strain data from a 10-km array shown in Fig. 1. (a) Raw strain data. (b) Frequency-wavenumber decomposition of the strain data. Red curve is the theoretical dispersion of ocean surface gravity wave in shallow water (h=30 m).

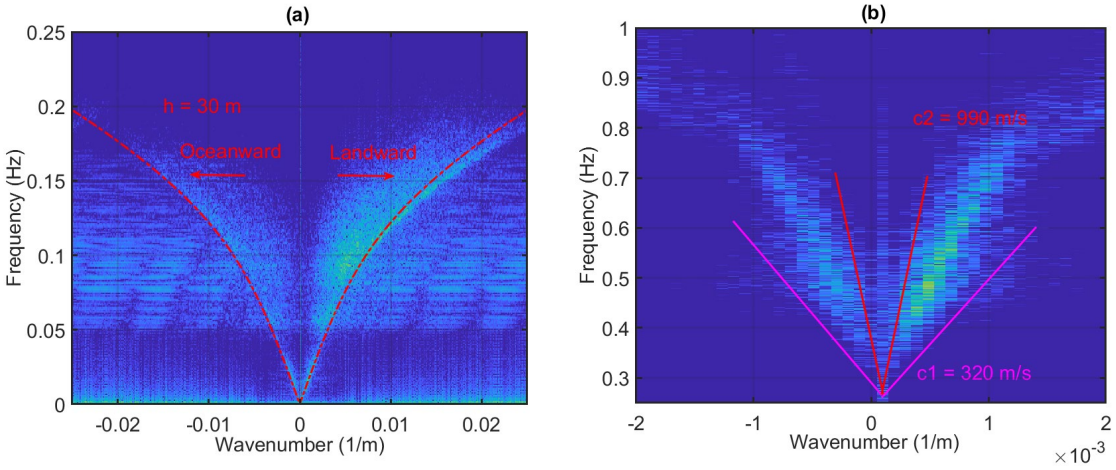


Fig. 3: Enlarged version of Fig. 2 (b). (a) PM decomposition and the dispersion curve of the gravity wave. (b) SM decomposition showing seafloor bounded Scholte wave.

$$\omega = \sqrt{g \cdot k \cdot \tanh(k \cdot h)}, \tag{1}$$

where ω is angular frequency, g is gravity acceleration, k is wavenumber, and h is water depth. The red curve fits the wave trains well. The interaction of wind waves in the opposite directions results in SM in the frequency band of 0.25 - 1.0 Hz, that reaches the seafloor and converts to Scholte wave with symmetric components traveling with the speed between the elastic speed of shallow sediments (200 - 500 m/s) and the acoustic speed of water (1500 m/s) [3].

2) Estimation of sediment shear wave velocity

The received passive strain dada are used to estimate the shear wave velocity profile in the sediment by performing the passive seismic interferometry. A bandpass filter of 0.6 -

5.0 Hz is applied to the passive strain data from a 1.2 km array, that is 60 DAS channels, to remove the high frequency components. The bandpass filtered strain data plotted in Fig. 4 (a) show the evenly distributed ocean noise. The passive seismic interferometry is performed on the bandpass filtered data to retrieve Green's functions [8]. The averaged Green's functions are shown in Fig. 4 (b). The Green's functions contain dispersive Scholte waves. The dispersion curve of the Scholte waves is extracted from the retrieved Green's functions by applying time-frequency analysis, τ-p transform and plotted in Fig. 5 (a), which is named as measured data.

The estimated dispersion curve is used as input data to a non-linear geoacoustic inversion algorithm, ASSA, to estimate the seabed shear wave velocity profile. An earth model with one horizontally homogenous sediment layer overlaying half-space is selected, where the shear wave velocities in the sediment and the half-space, and the sediment layer thickness are estimated parameters, while other geoacoustic parameters are kept constant during the inversion. The predicted dispersion curve is plotted in Fig. 5 (a), and it fits the measured dispersion curve very well. The estimated shear wave velocity profile is plotted in Fig. 5 (b) which indicate a soft sediment layer. The scatter plot of the objective function values for the estimated parameters are displayed in Fig. 5 (c) which gives a qualitative sensitivity of the estimated parameters and a rough measure of the uncertainties of the estimated values. All the three parameters are well estimated as their plots appear like "Tornado". The inversion results indicate a soft sediment layer with the shear wave velocity of 360 m/s and the thickness of 65 m, and the shear wave velocity in the layer beneath is about 410 m/s.

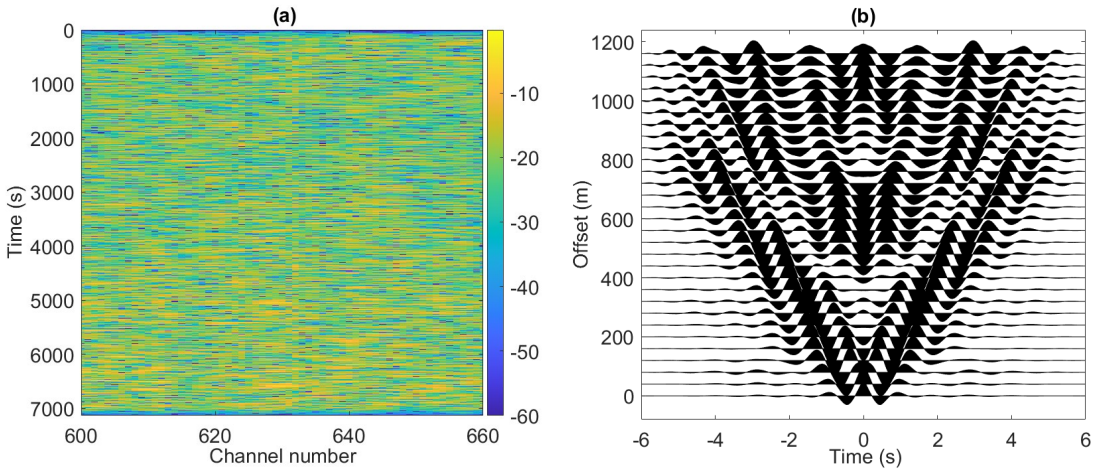


Fig. 4: (a) Bandpass filtered 60 channels of 2-hour strain data. (b) Averaged Green's functions of 60 channels retrieved by passive seismic interferometry.

4. CONCLUSION

A two-hour passive DAS strain data from the Tampnet DAS dataset are analysed to detect the ocean surface gravity waves and estimate shear wave velocity of the shallow sediments. Frequency-wavenumber decomposition separates the PM and SM thanks to the good low-frequency performance of the DAS system. This low frequency analysis can be used to detect the surface wave generated by vessels to characterize the vessel's speed and dimension. Green's functions are retrieved by performing passive seismic interferometry,

which contains dispersive Scholte waves. Shear wave velocity profile of the shallow sediments is estimated using the extracted dispersion curve of the Scholte wave by ASSA inversion. The inversion results indicate a soft sediment layer and the penetration depth around 100 m.

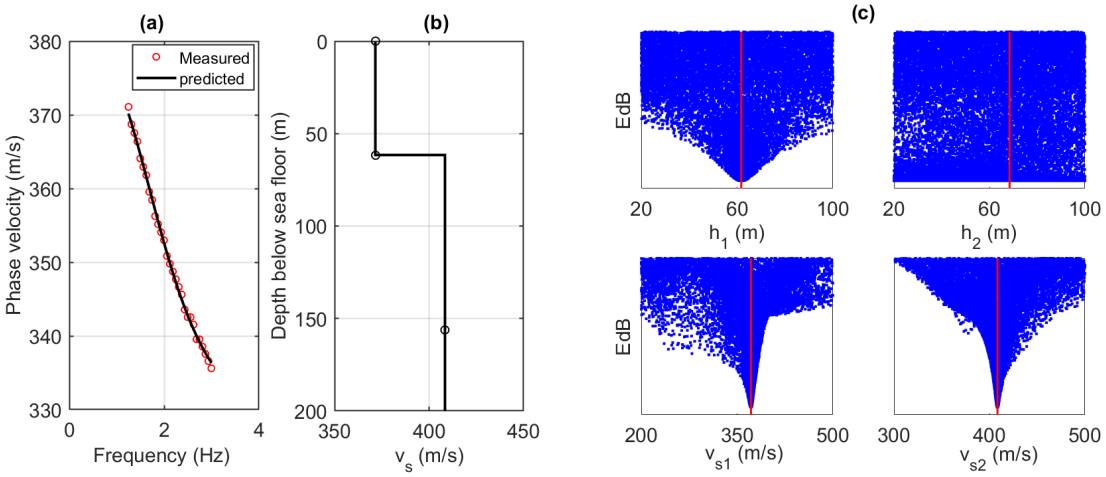


Fig. 5: ASSA inversion results. (a) Estimated and predicted dispersion curve. (b) Estimated shear wave velocity profile. (c) Scatter plot of the objective function values of the estimated parameters.

5. ACKNOWLEDGEMENTS

This work is partially supported by the SFI Centre for Geophysical Forecasting under grant 309960.

REFERENCES

[1] **Hartog, A. H.,** *An introduction to distributed optical fibre sensors*, 1st ed.: CRC Press, 2017.

[2] **Dou, S. *et al.*,** Distributed acoustic sensing for seismic monitoring of the near surface: A traffic-noise interferometry case study, Scientific Reports, 7, pp. 1–12, 2017.

[3] **Lindsey, N. J. *et al.*,** Illuminating seafloor faults and ocean dynamics with dark fiber distributed acoustic sensing. *Science* 366, pp. 1103–1107, 2019.

[4] **Landrø, M. *et al.*,** Sensing whales, storms, ships and earthquakes using an arctic fibre optic cable. *Sci. Rep.* 12, 19226, 2022.

[5] **Bouffaut, L. *et al.*,** Eavesdropping at the speed of light: distributed acoustic sensing of baleen whales in the Arctic, 2022.

[6] **Taweesintananon, K. *et al.*,** Distributed acoustic sensing of ocean-bottom seismo-acoustics and distant storms: A case study from Svalbard, Norway. *Geophysics*, 88(3), pp. B135–B150. 2023.

[7] **Wuestefeld, A. *et al.*,** The Global DAS Month of February 2023. Seismological Research Letters 95 (3), pp. 1569–1577, 2024.

[8] **Wu, G. *et al.*,** Sehar-wave tomography using ocean ambient noise with interference, Remote Sens, 12, 2696, 2020.

Structured Session
06. Enhancing underwater acoustic sensing through machine learning

Organised by Cheng Chi

Neural Network-Informed Acoustic Sensors for Optimal Coherent Processing: Benchmarking in a Forward-Modelled Fluctuating Ocean Environment

Antoine Blachet[1], Xavier Cristol[1], Thomas Mahiout[1], Dominique Fattaccioli[2], and Gaultier Real[3]

[1]Thales Defence Mission Systems, Sophia Antipolis, France
[2]DGA Naval Systems, Toulon, France
[3]CMRE, La Spezia SP, Italy (Formerly DGA Naval Systems)

Keywords: Internal waves, Sound propagation modelling, Artificial Neural Networks, Array processing

Abstract

Acoustic signal propagation in fluctuating underwater environments, particularly those influenced by internal waves (IW), presents significant challenges for maintaining coherence over large aperture arrays. Vertical hydrophone arrays are typically used to evaluate the vertical Mutual Coherence Function (MCF) of acoustic pressure.

In this work, we investigate strategies to mitigate acoustic vertical loss of coherence in the context of signal detection and direction of arrival (DOA) estimation. We explore the use of artificial neural networks (ANNs) to adaptively inform array signal processing based on spatial oceanographic measurements. This is formulated as a time series prediction task for estimating array coherence length from multivariate environmental data, enabling real-time optimization of processing strategies.

Building on previous work using canonical correlation analysis (CCA) on the ALMA 2017 dataset, our approach extends that study by leveraging realistic synthetic data to gain better control and interpretability of the acoustic effects induced by large-scale, sparsely observed oceanographic processes. This strategy enables validation of propagation hypotheses, expansion of dataset volume, and reduction of uncertainties from noise, interference, and limited observations.

Using a parabolic equation model in a 2D fluctuating environment based on the ALMA 2017 setup [1, 2], we simulate acoustic propagation through IW-induced 2D sound speed profiles (SSPs). We validate the acoustic and oceanographic models by comparing beamformed outputs and mutual coherence functions with real data.

These synthetic datasets are then used to train and benchmark various ANN architectures. Our approach validate the previous CCA-based method and extends it by introducing a learning-based framework that accounts for complex non-linear relationships within the acoustic data. Results highlight the value of acoustic modeling in supporting data-driven methods, offering controlled training data and guiding optimal ANN design. Future work includes scaling to larger synthetic datasets, adapting the method to different sonar geometries, and further ANN optimization for IW-dominated environments.

1. INTRODUCTION

The propagation of acoustic signals in underwater environments influenced by internal waves (IW) [3, 4, 5] poses significant challenges for large-aperture arrays, where spatial coherence is essential. IW-induced fluctuations in the sound speed profile (SSP) lead to phase decorrelation and interferences, which degrade detection and direction-of-arrival (DOA) performance. Vertical linear arrays (VLAs) are commonly used to characterize these effects through measurements of the vertical Mutual Coherence Function (MCF).

To address acoustic decoherence, we investigate data-driven strategies that adapt array processing based on environmental inputs. Specifically, we use artificial neural networks (ANNs) to predict vertical coherence length from SSP profiles, formulating the problem as a multivariate time series prediction that enables real-time tuning of beamforming algorithms to the optimum subarray length.

This work extends previous studies using Canonical Correlation Analysis (CCA) on ALMA 2017 data [6, 7], by introducing a synthetic modeling framework for benchmarking ANN architectures. Our focus is on predicting the vertical coherence length, a key parameter for optimizing array processing strategies — its use for adaptive beamforming has already been demonstrated in prior work [6, 7]. Internal waves are difficult to observe in full detail due to the challenges of spatially sampling the ocean environment. To gain better control over this sparse information and improve data diversity, we generate synthetic datasets using a parabolic equation (PE) model driven by IW-induced SSPs. These simulations replicate ALMA 2017 conditions and are validated against real beamformed outputs and coherence functions.

The synthetic approach supports exploration of potential ANN types and architectures, and enables large-scale dataset generation for training and validation. We benchmark three architectures: a standard Feed-Forward ANN (FF), a Long Short-Term Memory (LSTM) network to capture temporal dependencies, and a CCA model as a comprehensive baseline.

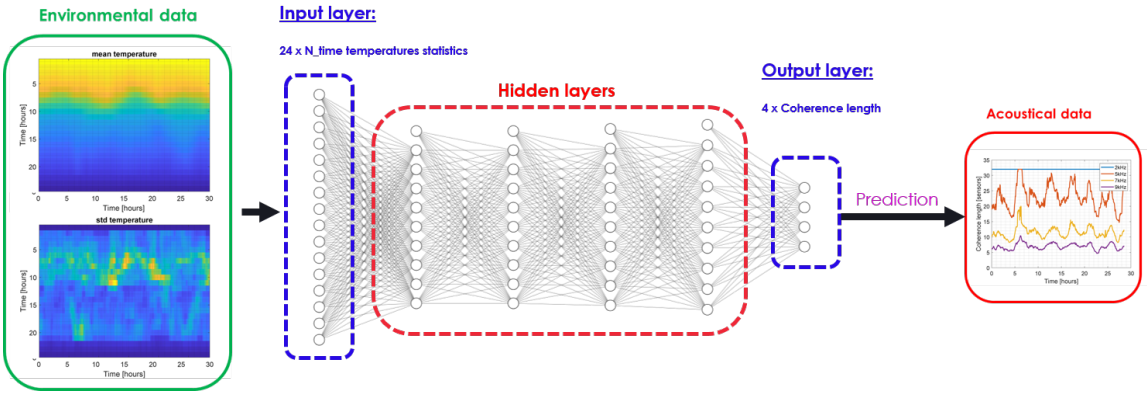


Figure 1: Illustration of an ANN (here a FF) for predicting array vertical coherence length. Input: multivariate temperature measurements (mean and standard deviation in a moving window). Output: coherence length at specific frequencies.

2. FORWARD MODELING & DATASET GENERATION

2.1. ALMA 2017 EXPERIMENTAL SETUP

The ALMA 2017 experiment involved a 4-column VLA with 32 sensors per column. The array Nyquist frequencies are 24 kHz temporally and 4.6 kHz vertically. The system was deployed near the north of Corsica, near a deep underwater canyon, 16.5 km from an artificial source, which emitted a comb of pure-tone signals. Concurrent thermistor measurements provided 24 samples over the first 150 m of the thermocline every three seconds. These data revealed strong temporal fluctuations due to IW, as documented in [6, 7]. Figure 2 illustrates the experimental setup used in this study.

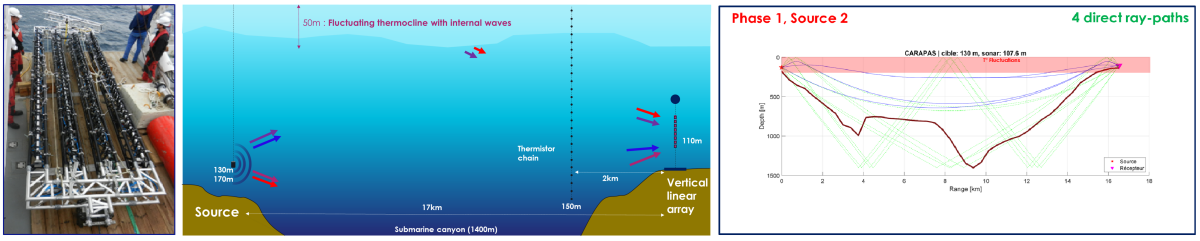


Figure 2: Left: Rx VLA, center: ALMA2017 experimental setup. Right: eigenrays computed with the mean SSP.

Ray tracing was first employed to model sound propagation under the assumption of no SSP variation. We computed eigenrays with an SSP built from the mean temperature data. As shown in Figure 2, the dominant propagation paths are equivalent to a deep-water scenario. The eigenrays include two dominant pairs with closely spaced angles of arrival, causing slowly time-varying multipath fading. Ray tracing without 2D fluctuations but using real SSP measurements extrapolated under the thermocline showed that the scale of this phenomenon is about several hours.

2.2. MODELING THE ACOUSTIC WAVEFIELD WITH 2D SOUND SPEED FLUCTUATIONS

To benchmark AI techniques, it was necessary to increase the volume of coordinated temperature-acoustic data beyond the available experimental signals. To this end, we first generated 3D numerical simulations of internal waves in the ALMA 2017 configuration, and then simulated the propagation of acoustic waves through these channel disturbances. This method was used to augment the ALMA2017 dataset with longer sequences and improved diversity of IW, while gaining control over large-scale oceanographic processes for training and testing AI techniques.

We first used the Thales proprietary model ONDIN: the experimental mean temperature profile over the thermistor chain was extended for deeper immersions under the assumption of a constant temperature. This profile was then used to calculate a Brunt-Väisälä frequency profile. Sequences of vertical modes of displacement were then calculated for a range of frequencies. The stochastic amplitudes of these modes were distributed along the Garrett-Munk spectrum [3, 5] in frequency and were drawn pseudo-randomly. We assumed a statistically uniform, isotropic distribution of IW arrivals horizontally. The result of this process is a complete set of

distorted isothermal surfaces over time.

In a second step, we simulated temperatures at the thermistor chain sensors (Figure 3) and the associated acoustic pressures at the hydrophones of the vertical array using the Thales model PATPATROIL (Figure 4). Propagation from the source to the array was computed with a split-step Fourier solution to the Standard Parabolic Equation [8], modified to account for horizontal SSP fluctuations. To support our approach, we deliberately simulated stronger internal waves than those observed in the real data to benchmark performance under intensified conditions.

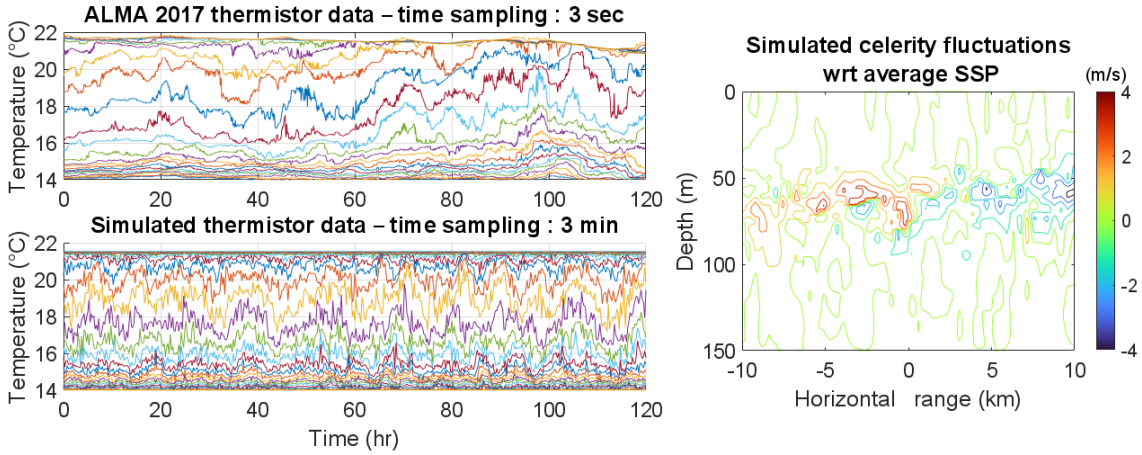


Figure 3: Top left: ALMA2017 thermistor data, bottom left: simulated thermistor data, right: example of simulated 2D SSP anomaly used to perturb propagation (note: we simulated intensified internal waves for benchmarking purposes).

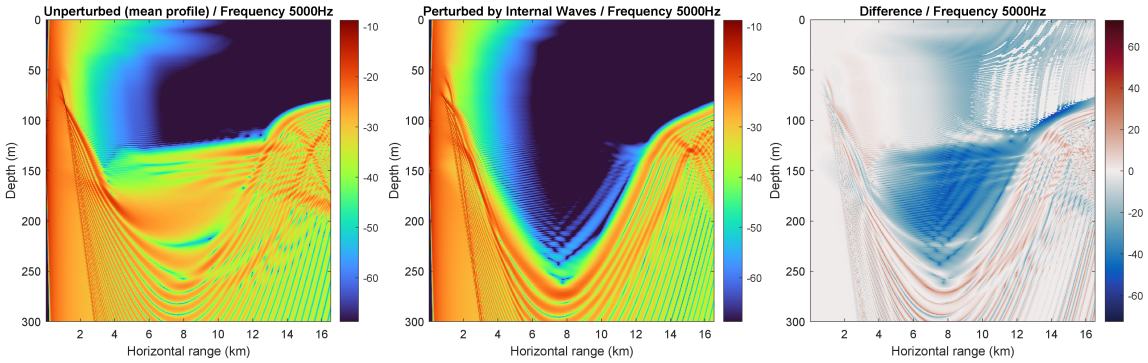


Figure 4: Example of transmission loss at 5 kHz (left: reference, middle: perturbed with 2D SSP fluctuations, right: difference). Computed with PE in the 2D vertical plane source-receiver (No cylindrical spreading).

3. COMPARISON: BEAMFORMING & ESTIMATED MUTUAL COHERENCE

Vertical beamforming was applied for analyzing the characteristics of received acoustic signals. Using a Kaiser ($\beta = 4$) window, we processed both simulated and real acoustic data identically, estimating the signal power in DOA every 3 minutes. Figure 5 presents waterfall plots of the beamformed data at different frequencies of the tone comb. The simulated data and real observations exhibit consistent arrival angles, ranging from $+10°$ to $-10°$. In both cases,

we observe DOA fluctuations induced by internal waves on a time scale of about 5–10 minutes. Minor differences are observed: at frequencies between 5 and 9 kHz, the simulated dataset shows a higher variance in DOA, which we attribute to stronger simulated internal waves. At 2 kHz, the real data appears more perturbed, and a re more difficult to interpret due to limited antenna directivity at this frequency. We compared the beamformed noise only data in Figure 6, similar spatial distribution of energy is observed with a noise notch, filled with some remaining energy at horizontal arrival in the real data, possibly caused by low level interfering sources (biological, trafic, source vessel).

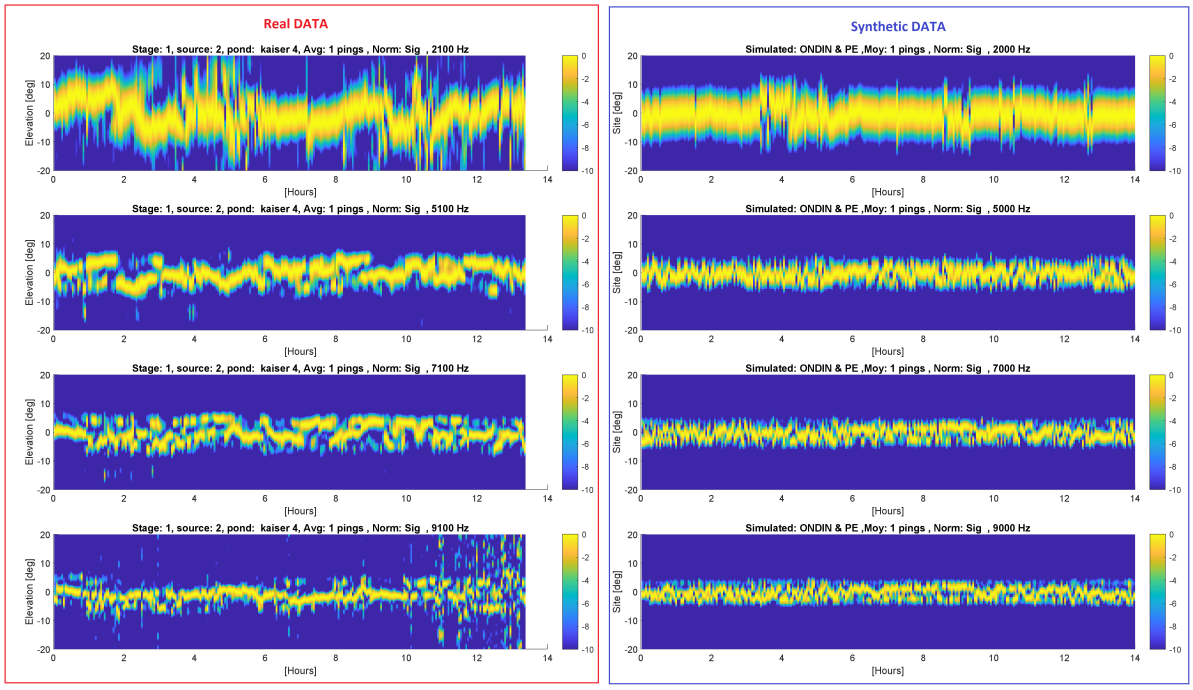


Figure 5: Beamformed VLA data (left: ALMA2017 real data, right: synthetic data), normalized relative to the maximum signal level.

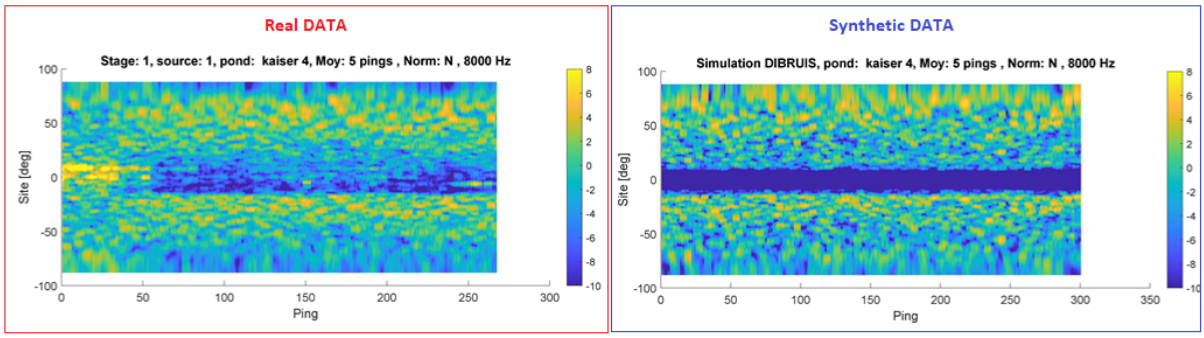


Figure 6: Example of vertically beamformed noise-only data at 8 kHz (left: ALMA2017 data, right: simulated additive ambient noise). On noise data a 5 pings moving average was applied to estimate noise power level.

We evaluate the Mutual Coherence Function (MCF) and the mean coherence length of signals received across the array. Figure 7 compares the MCF derived from real and simulated data. Coherence length is computed from averaging the MCF function in time (moving average

window of 30 min) and between sensors after aligning the matrix diagonal, then estimating de the 3dB width. We observe good agreement above 5 kHz, where the source signal is dominant. One hypothesis is a higher amount of correlated noise induced by interfering sources at low frequencies. Generally, the simulated data exhibit less variation between sensors, while the real data contains clusters of spatially close sensors with highly correlated pressure signals.

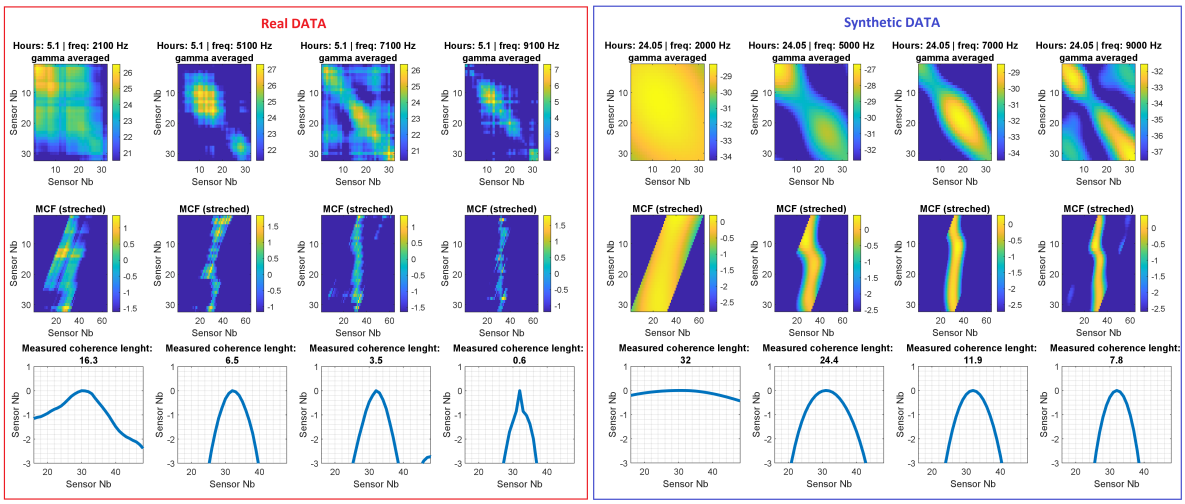


Figure 7: Left: ALMA2017 data, Right: simulated data. Top row: averaged array covariance matrices, middle row: MCF functions, bottom row: estimated coherence length.

4. VERTICAL COHERENCE LENGTH PREDICTION THROUGH MACHINE LEARNING

We evaluate and compare a range of neural network architectures to identify the most suitable model for predicting vertical coherence length from SSP time series. The tested architectures include feedforward (FF) networks, recurrent networks based on Long Short-Term Memory (LSTM), and Canonical Correlation Analysis (CCA) after applying Principal Component Analysis (PCA). For FF networks, we use the ReLU activation function to capture complex non-linear relationships within the data. All networks are trained using the Mean Squared Error (MSE) loss function, with frequency-normalized error to account for spectral variability. Optimization is performed using the ADAM algorithm. For FF and LSTM models, a grid search is conducted over key hyperparameters, including learning rate, dropout rate, number and size of hidden layers, and sequence length (for LSTM). FF Inputs are flatten multivariate individual time samples, and LSTM inputs are multivariate time series of 54 minutes. Approximately 30 hours of simulated data — covering several cycles of internal wave variability — are used for training. Each approach presents specific advantages: FF networks are computationally efficient and expressive; LSTMs capture temporal dependencies; and CCA serves as a comprehensive baseline.

The results, presented in Figure 8, indicate that all networks achieve a success rate (defined as a relative error below 10%) exceeding 50% in most cases. Furthermore, a stronger correlation is observed between FF/LSTM predictions and ground truth values, demonstrating their ability to accurately model underwater fluctuation behavior. Among all architectures, LSTM provided the best overall performance, with the lowest mean square error. These outcomes validate the effectiveness of the proposed approach in fully controlled environments dominated by internal

waves.

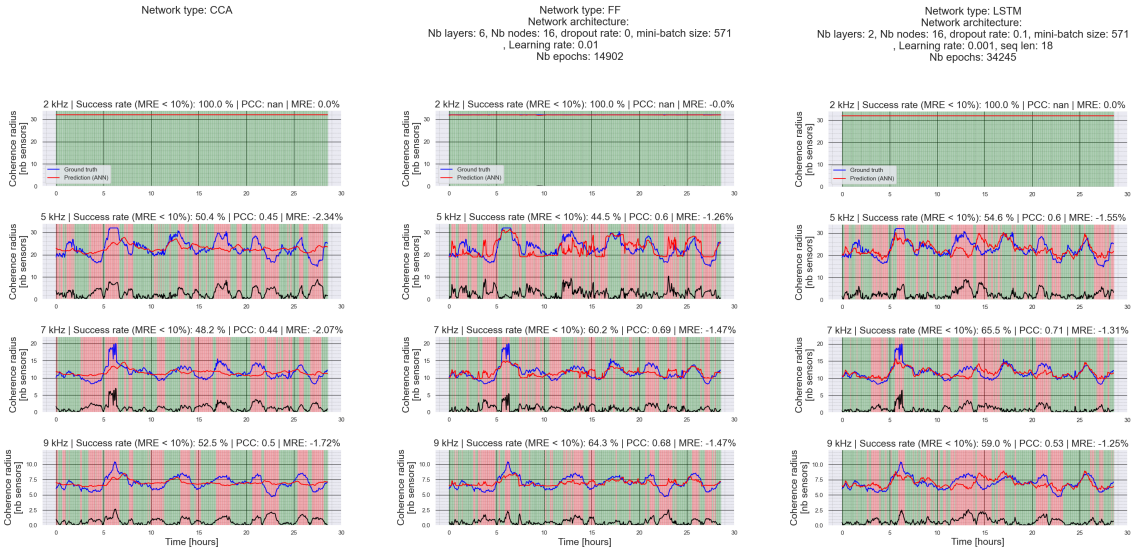


Figure 8: Comparison of the three neural networks versus ground truth at different frequencies (left: PCA & CCA, middle: Feed-Forward, right: LSTM). Blue curve: ground truth, Red curve: prediction, Black curve: error, green background: MRE < 10%, red background: MRE > 10% , PCC: mean moving correlation between prediction and ground truth, MSE: Mean square error, MRE: Mean relative error.

5. CONCLUSION AND PERSPECTIVES

This study illustrates the value of combining underwater acoustic modeling with observational data. The agreement between real and simulated datasets validates the modeling approach and supports key assumptions about underwater signal propagation. The synthetic data successfully captured essential features of the measured signals, providing a useful resource for designing and studying neural networks, but of also for training and evaluation.

Benchmarking different ANN architectures on this dataset demonstrated potential improvements over the baseline method (CCA). Although performance gains were relatively modest, both feedforward and LSTM architectures achieved slightly better correlation and lower MSE than CCA, suggesting enhanced ability to capture complex, non-linear relationships and temporal dependencies. In addition, the simulations confirm and validate the results obtained in previous studies using CCA on real data, reinforcing the relevance of this approach in internal wave dominant environments.

This work is part of ongoing research. Future efforts will focus on expanding the dataset to longer time series, incorporating non-stationary internal waves, validating new architectures on real data, and exploring generalization to varied environments and alternative sonar configurations.

6. ACKNOWLEDGEMENTS

The authors would like to thank Thales DMS for financing this research through self R&D funding and permission to publish. We thank DGA for the ALMA2017 experimental data. Dr. Alexandre L'Her (ENSTA) for discussions and Dr. Iannis Bennaceur for his help with reading the ALMA2017 raw data. Artificial Neural Network was performed with the open source library PyTorch developed by Meta AI [9].

REFERENCES

[1] D. Fattaccioli and G. Real, "A brief five-year progress report on the dga alma project: scientific program and projects, at-sea campaigns and data collected, system evolutions and improvements," in *Proceeding of the Underwater Acoustics Conference and Exhibition (UACE)*, 2019.

[2] G. Real and D. Fattaccioli, "Acoustic laboratory for marine applications: Overview of the alma system and data analysis," *OCEANS 2018 MTS/IEEE Charleston*, pp. 1–7, 2018.

[3] C. Garrett and W. Munk, "Space-time scales of internal waves," *Geophysical Fluid Dynamics*, vol. 3, no. 3, pp. 225–264, 1972.

[4] R. Dashen, S. M. Flatté, W. H. Munk, K. M. Watson, and F. Zachariasen, *Sound transmission through a fluctuating ocean*. Cambridge University Press, 2010.

[5] J. A. Colosi, *Sound Propagation through the Stochastic Ocean*. Cambridge University Press, 2016.

[6] A. L'Her *et al.*, "Towards improved sonar performance using environment-informed sparse sub-array processing," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023.

[7] A. L'Her, "Propagation acoustique en milieu fluctuant: des phénomènes océanographiques au traitement du signal," Ph.D. dissertation, ENSTA Bretagne, 2024.

[8] F. B. Jensen, W. A. Kuperman, M. B. Porter, H. Schmidt, and A. Tolstoy, *Computational ocean acoustics*. Springer, 2011, vol. 2011.

[9] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035.

The Urgent call on data management: are we capable to store valuable (meta)data for naval application?

Sonia Papili[1]

[1]Royal Military Academy, Robotic & autonomous Systems. Graaf Jansdijk 1, 8380 Brugge. Belgium. Email: Sonia.papili@mil.be

Sonia Papili. [1]Royal Military Academy, Robotic & autonomous Systems. Graaf Jansdijk 1, 8380 Brugge. Belgium. Email: Sonia.papili@mil.be

Abstract: *Technology is increasing rapidly and new concepts as robotics, artificial intelligence (AI) or augmented reality are entering the daily life. Nevertheless, the maritime world is not immune to the disrupting advancing in technology and information production especially regarding shipping industry, military, and monitoring methodologies. The basic will is automatization, efficiencies, and safety at affordable costs. Therefore, in parallel, other aspects are evolving, mostly related to interoperability and data management. An enormous amount of information is produced and the necessity to store in an appropriate way is demanding. In this contest, an interdisciplinary environmental data model for mine countermeasure application is proposed where heterogeneous environmental information is integrated as a base structure for a future data management system. The attention was posed not only on measurable data related to various disciplines such as physics, chemistry, geology, biology, and acoustics, but also on historical data and subjective interpretation such as descriptions and lesson learned.*
In the model a distinction was made between "humans" and "machines". The humans are seen as entities that must collect and store data, meanwhile the machines are seen as entities that query information. In this way the model shows what to take into consideration when to collect and store data, how to integrate data following rules and style conventions, how to build a conceptual table with heterogeneous information, what

are the tools to allow the interrogation of measured, historical, and subjective-descriptive data.
The possibility to automatically access heterogeneous data volume opens a broad spectrum of possibilities, from a characterization of a marine environment to the representation of a marine environment by virtual reality, to a sensor performance prediction, to risk analysis.

Keywords: *data model, historical information, descriptive data, descriptive information.*

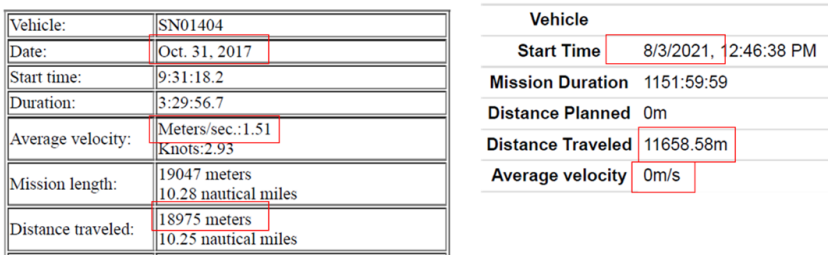
1. INTRODUCTION

Although large amount of information is available within different disciplines in marine science, collecting the right information for naval purposes is a demanding and tedious process. As highlighted by Yan et al. in 2014 the difficulties encountered are mainly due to the heterogeneity of marine data being multi-scale, multi-temporal and multi-semantic.
Several tentative are on-going to tackle this matter following different strategies.
The United Nations are working on the Global Ocean Observing System (GOOS) with the idea to achieve integrated observations and data sharing.
Yan et al. in 2014 used machine learning model to disentangle the difficulties refining the correspondence relationship between data.
A new International Hydrographic Organization (IHO) is attempting a series of standards developments based on a new universal hydrographic data model, S-100 (Ponce, 2018),
In this framework and considering the new challenges in maritime monitoring, an interdisciplinary data model is hereby presented.
The model clarifies how to collect and store heterogeneous data from various disciplines such as physics, chemistry, geology, biology, and acoustics.
It also identifies tools and methodologies necessary to have proper interrogations of measured, historical and subjective-descriptive data related to marine environments. Moreover, it highlights how personal and subjective perspective can affect descriptive information and metadata collection.
Hereafter, the different topics are illustrated separately. A section on heterogeneous data is illustrated, subsequently a section on historical data and then a section on subjective and descriptive data are discussed. Finally, an overview on the full model is given.

2. HETEROGENEOUS DATA

To investigate the feasibility and the modalities for constructing an interdisciplinary data model for environmental characterization, information on different domains was extracted from databases and scientific literature and subsequently collected into a comprehensive table. This table is subdivided using a pyramidal structure containing different level of information. The first level describes scientific domain such as physics, chemistry, geology, acoustics, computing and others; the second level specifies the scientific fields such as sedimentology, bio-communities, hydrodynamics, sonar properties, vehicle properties; and the last level describes the parameters and the metadata important from an environmental perspective.
The extraction of information from scientific literature and free accessible databases presented major difficulties. Lack of standardization results in discharging of information, and duality in wording creates substantial confusion. Common example of lack of

standardization is often encountered among software's or among updated versions of the same software giving different formats for the same entity (Fig.1). Duality in wording is often encountered among different disciplines where different meanings are attributed to the same word. It is often a semantic problem that can create confusion and subsequently errors. As example it is reported the use of the word "resolution" in relation to sonar images. Some authors refer to the word resolution as "pixel resolution" whilst others refer to it as the "theoretical resolution" of the sonar system.
Therefore, not only a descriptive analysis on what data to collect, but also a detailed analysis on how to collect and store the data is considered in the model.



Vehicle:	SN01404
Date:	Oct. 31, 2017
Start time:	9:31:18.2
Duration:	3:29:56.7
Average velocity:	Meters/sec.:1.51 Knots:2.93
Mission length:	19047 meters 10.28 nautical miles
Distance traveled:	18975 meters 10.25 nautical miles

Vehicle	
Start Time	8/3/2021, 12:46:38 PM
Mission Duration	1151:59:59
Distance Planned	0m
Distance Traveled	11658.58m
Average velocity	0m/s

Fig.1.Example of lack of standardization. Different formats are here illustrated for the same software in a time range of four years. Differences in entities like "date", "distance travelled" and "average velocity" are evident.

It would be good policy to use always the same name convention and format convention for data collection and analysis.

The application of Rules and Style conventions of the International System of Units (SI) if an example of it. Other examples could be the Climate and Forecast (CF) Conventions, CF standard Names, ACDD Conventions. Furthermore, important is to record which applied method is used and to note which analysis and ancillary information were done.

3. HISTORICAL DATA

Historical data brings a high potential amount of information.

The possibility to have historical data coupled with recent data will allow to have a broader dataset for deep learning and automatic processing. The term 'historical' does not necessarily refer to information made in the past, but eventually to data observed with technologies or software retained obsolete and not more in use. In that case, some recommendations are listed to facilitate a straightforward integration of data.

Conformity between the data is a primary requirement for a good integration of data. Within the context of this model, conformity is related to three main domains: methodology to collect and rework the gathered information, geographical positioning systems and collection of metadata.

Conformity in methodology

To reclassify historical laboratory results to modern one, a good method is to perform modern laboratory analysis to a set of historical samples. The new results would give a scale of comparison between new and ancient way of classification.

An interesting example was made by Houziaux et al. in 2008 during the review and the digitalization of the samples collection of prof. G. Gilson. Sixty-two sandy samples from the collection were re-analyzed in laboratory for grain size analysis. The resulting

measurements were confronted with those made by prof. Gilson and a comparison scale was made.

Conformity in wordings

To verify a conformity in ancient and modern scientific language, a good method is to perform a new description on already described historical data and compare the two descriptions.

Conformity in geographical positioning

When the position accuracy is not satisfying the modern techniques, it is necessary to build a re-sampling strategy. A reference area is re-sampled and subsequently used to verify the historical data with current data.

An interesting approach to estimate the value of ancient maps is proposed by Leriche et al. in 2004. She calculated a Reliability Index (RI). This index weighs three parameters: the scale of the map, the data acquisition method and the geographical positioning system method. Using this method, she could estimate the reliability of ancient maps.

Next to conformity, historical data need to be labelled by a *quality rank coefficient* expressing the reliability of the information and to assign a coefficient of uncertainty giving a weight to the information self.

A good example on how to attribute a quality rank is given by Kint et al. in 2020 when under the umbrella of the TILES project (TILES consortium 2018a) they have introduced data uncertainty for horizontal positioning, sampling methods and age of information. Quality rank ranging from 1 (very uncertain) to 5 (very certain) called quality flags were associated to historical measurements. Those quality ranks were converted into corresponding uncertainty percentages. Those percentages were associated to the historical data to weigh the precision of the information.

4. SUBJECTIVE DESCRIPTION AND LESSON LEARNED

With the idea to develop a method to feed subjective information into the model, an experiment was performed to weigh how individual background, and subjective perspective may affect a scientific description. Voluntaries holding different background as engineers, ecologists, military officers and schoolteachers, gently offered their contribution to this semantic experiment. They were required to make an environmental description watching two video fragments showing the seafloor of the Belgian Continental Shelf. The simulation was built in two phases. A free description of the videos was requested during the first phase whilst a questionary was provided during the second phase. The resulting descriptions were analysed following different criteria. Differences and analogies were highlighted. Attention was posed on the use of the vocabulary and on the syntax. Among all the result of the analysis, interesting is to highlight that a conversational style using first personal pronoun is used when the subject is not confident with his/her observation (I see, I guess, I think, I have no idea). Different semantic is used to describe the same entity: “broken shells, shell detritus, shell fragments”. Differences in styles: structured or lose. This analysis together with the previous ones highlights the important of using dedicated style conventions also for descriptive information.

5. MODEL

This model (Fig. 2) highlights several aspects related to data management. It clarifies how to collect and store heterogeneous data. It also identifies tools and methodologies

necessary to have proper interrogations of measured, historical and subjective-descriptive data related to marine environments and gives insights on how human and machine skills can contribute to achieve the best from heterogeneous collection of information.

The model is separated in two sections that eventually can cross each other: the first section focusses on “humans”, the second on “machines”.

The humans are seen as entities that must collect and store data, meanwhile the machines are seen as entities that query information.

When referring to collecting and storing of information, attention is posed to integrate data following rules and style conventions. Mostly, conventions are related to data format, measuring units and avoiding duality in wording. When historical information is stored, the introduction of policies for the conservation of historical data would allow a smooth integration into future datasets. Annotation on used technology, used software and used formats are essential for this integration.

Dedicated template would facilitate the process to collect descriptive information. Annotation of time, encounter problems and used techniques would be valuable ancillary information.

Considering what is necessary to have to be able to interrogable information, the attention in the model is posed on the availability of structured data, conformity and quality rank coefficient for historical data and dedicated vocabularies for descriptive data.

Particularly, scientific vocabularies, classification of random wordings, use of the adjectives and adverbs and synonyms needs to be integrated in computer capabilities.

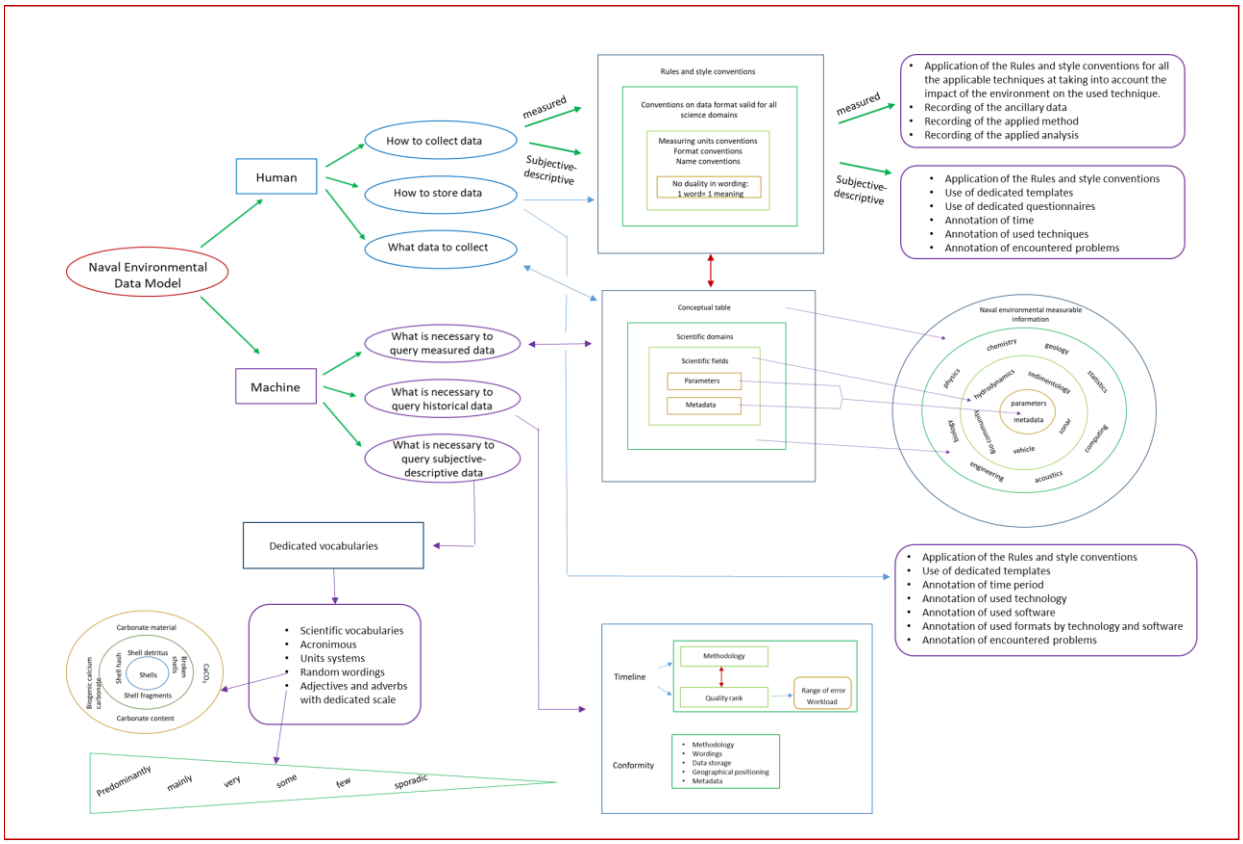


Fig.2. Conceptual model. A framework wherein measured, subjective and historical data are harmonized together

6. CONCLUSION

The model results on a list of insights and recommendation for future application in deep learning. A unique approach is presented that has the potentiality to be a key concept into the marine data management. The insights presented can be used at different levels and for different purposes in the maritime world.
The possibility to automatically access heterogeneous data volume opens a broad spectrum of possibilities, from a characterization of a marine environment to the representation of a marine environment through virtual reality, to a sensor performance tools to risk analysis. If dedicated vocabularies would be developed, automatic understanding of subjective information could be evaluated during the experimentation of new methodologies at sea or new technologies.

7. ACKNOWLEDGEMENTS

The author thanks the volunteers taking part to the semantic investigation, the department of synthesis-evaluation lesson learned for the contribution on extracting information from technical lesson-learned reports.
This paper contributes to the projects DAP19-07 and DAP 22-16 funded by the Belgian Ministry of Defense, Royal Higher Institute for Defense.

REFERENCES

- **[1] Yan,W., Jiajin, L., Yun, Z**. A multianalyzer machine learning model for marine heterogeneous data schema mapping. *Scientific World Journa*l. doi: 10.1155/2014/248467. (2014)
- **[2] UNESCO**. United Nations Decade of Ocean Science for Sustainable Development (2021-2030). *https://en.unesco.org/ocean-decade*
- **[3] Ponce, R.** The Maritime World Enters the Fourth Industrial Revolution. *Sea-Technology. (*2018*).*
- **[4] IHO.** Standards for Hydrographic Surveys. *International Hydrographic Bureau*. Monaco. Special Publication n° 44 (2008).
- **[5] Houziaux, J.-S., F. Kerckhof, K. Degrendele, M. Roche, A. Norro**,. The Hinder Banks: yet an important area for the Belgian marine biodiversity? *Belgian Science Policy* D/2008/1191/7. (2008)
- **[6] Leriche A., Boudouresque C-F., Bernard G., Bonhomme P., Denis J**. A one-century suite of seagrass bed maps: can we trust ancient maps? *Estuarine, Coastal and Shelf Science. Elsevier*. 59. 353-362 (2004).
- **[7] Kint L., Hademenos V., De Mol R., Stafleu J., van Heteren S., Van Lancker V**. Uncertaintly assessment applied to marine subsurface datasets. *Quarterly Journal of Engineering Geology and Hydrogeology. 54*. (2020). *https://doi.org/10.1144/qjegh2020-028*

Power and Accuracy Trade-Offs for Machine Learning Methods Applied to Detection of Underwater Sound Sources

William Butler[1], Harrison Smith[2], Marios Impraimakis[1], Andrew Barnes[1], and Alan Hunter[1]

[1]University of Bath, Bath, BA2 7AY, United Kingdom
[2]Celtic Sea Power, Hayle, TR27 4DD, United Kingdom

Contact Author: William Butler, wmb34@bath.ac.uk

Keywords: *Machine Learning, CNN, Low-Power.*

1. INTRODUCTION

There are many autonomous marine drones currently deployed for a range of commercial, defense and marine conservation applications. Two categories commonly discussed are Uncrewed Surface Vehicles (USVs) and Uncrewed Underwater Vehicles (UUVs). These uncrewed vehicles often operate autonomously at sea for extended periods of time without refueling, yet due to their relatively small size they have limited power on board. For example, the Kongsberg SEAGLIDER may operate for up to 10 months, with a total power supply of only 17 MJ [1] (17 MJ $\approx$ 4722 Wh). In order to improve smart autonomous decision-making, as well as optimising the storage of data recorded by the onboard sensors, Artificial Intelligence (AI) can be used for rapid, real-time automated detection and classification of underwater noise-emitting objects. However, AI models often rely on large, deep neural networks that require extensive computational resources to run.

Much work has been done on edge computation, often utilising cloud resources [2]. For applications where cloud computing is not viable due to either a lack of internet connectivity at sea or security concerns associated with transmitting data, there has been less research. In addition, many of the models described above require pre-processing. In the case of audio classification, often spectrograms are used with Computer Vision models. These require the computation of Fourier transforms, which contribute to the overall energy cost of the system, not just the model itself.

However, while 2D CNN-based architectures such as ResNet have been widely successful [3], including for classifying spectrograms, CNNs with 1D-kernels can be used on time-series data such as the raw audio waveform. These models bypass the need for frequency decomposition via Fourier transform. In this work, two lightweight CNN models with similar architectures

Class	Count	Description
S1	741	Multiple Harmonic Tonals
S2	520	Non-tonal Bursts
S3	363	Harmonic Pulses
SP4	134	Dyadic Pulse Train
SP5	170	"Zipper" Pattern Pulse Train

Table 1: Seal call class densities and descriptions

are presented to compare the power and accuracy trade-offs between a 2D-kernel CNN model classifying spectrograms and a 1D-kernel CNN model classifying raw waveforms, as well as their respective pre-processing computational cost.

2. METHODOLOGY

The data used in this paper was provided by Celtic Sea Power (CSP). Multiple 2-hour recordings were taken with an AudioMoth [4] and HydroMoth [5] hydrophone with a sample rate of 96kHz. The audio was manually tagged by CSP employees and verified by a marine biologist. Seal calls were tagged with a start and end time and then split into 5 classes labeled S1, S2, S3, SP4, and SP5 according to a modified version of Nowak's seal vocalisation classes ([6], [7]). Table 1. shows the class densities in the dataset.

For each call tagged and labeled according to the five classes mentioned in Section 2., a 10-second slice of the 2 hour audio recordings is taken, randomly centering the call. Gaps between calls with a duration of 2 to 10 seconds are used to compute a background noise floor for each location by taking the median across each frequency. More information on the slicing and background noise floor generation, as well as the call types themselves can be found in the paper titled "Automated detection and classification of vocalisations from wild and captive seal populations" in this conference proceedings [7].

2.1. 1D CNN PRE-PROCESSING

Minimal pre-processing on the raw wav files is performed. 20 wav files are randomly sampled from each class, from the training set, and a mean and standard deviation of the amplitudes across those sampled calls is computed ($\mu \approx 0.00$, $\sigma \approx 0.07$). These values are then used to perform Z-score normalisation on the raw waveform. No other pre-processing is performed; this normalised time series is used as the input to the 1D CNN model. This input is a waveform sampled at 96kHz over a 10-second window, resulting in a single-channel input vector of shape (1,960000).

2.2. 2D CNN PRE-PROCESSING

To generate 2D inputs, a Short-Time Fourier Transform (STFT) is computed for each 10-second audio segment using a 2048-point window, a hop size of 1024 samples, a Hann window, and a sampling rate of 96kHz. The absolute value of the spectrogram, computed using the

Python Librosa package [8], is taken to obtain the magnitude of the frequencies. This returns a 1 channel, real-valued spectrogram of shape (1025, 936), after slicing the first and last time frame to account for any potential artifacts. A frequency-dependent background noise floor is estimated for each location by taking the median amplitude (in linear scale) from non-call segments as described above. This noise profile is then divided from the linear spectrogram for each input sample. After noise subtraction, the spectrogram is converted to the decibel (dB) scale with a reference of 1.0. From these dB spectrograms, a class-wise mean and standard deviation is computed across 20 randomly selected samples from each class ($\mu \approx -0.29, \sigma \approx 6.78$). These statistics are then used to perform Z-score normalisation on all spectrograms. The resulting normalised dB-scale spectrograms are used as input to the 2D CNN model.

2.3. MODEL ARCHITECTURES AND TRAINING

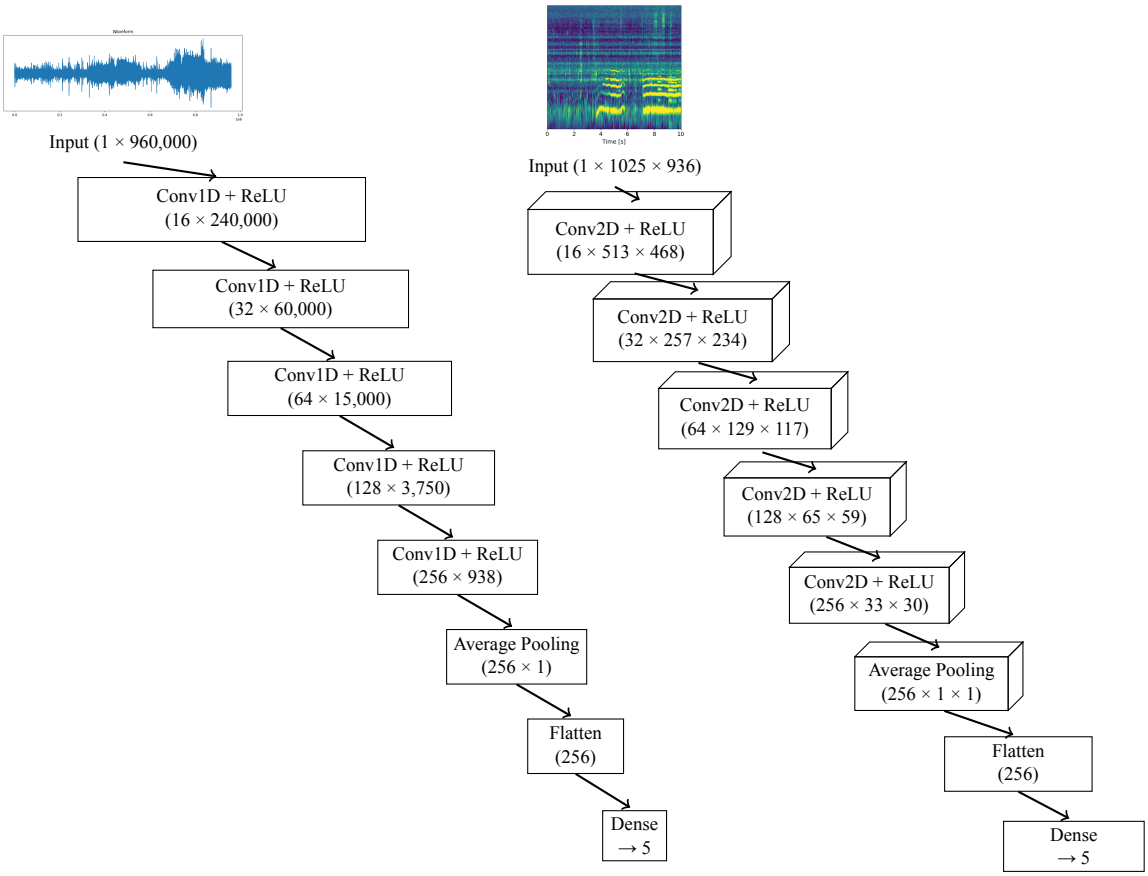


Figure 1: Comparison of 1D CNN for raw waveform input (left) and 2D CNN for spectrogram input (right). Dimensions indicate (channels × temporal length) for 1D and (channels × height × width) for 2D layers.

The raw waveform and spectrogram models share the same architecture, differing only in convolution type. The waveform model uses 1D convolutions with kernel size 9, stride 4 and 4 padding; the spectrogram model uses 2D convolutions with 3×3 kernels, stride 2 and 1 padding. This keeps both the parameter count and the downsampling rate consistent across models. Both models have five convolutional layers with ReLU activations at each layer and channels increasing from 16 to 256. After the convolutional stack, both models include an adaptive aver-

age pooling (Average Pooling) layer followed by a fully connected dense layer with 5 output nodes corresponding to the 5 classification categories. As PyTorch's CrossEntropyLoss integrates softmax internally, no explicit output activation was used. Fig 1. shows the architectures side by side with their channels and latent space dimensionality at each step.

Both models were trained under identical conditions to ensure a fair comparison. Cross-entropy loss was used with class weighting to address data imbalance. The dataset was split into training, validation, and test sets with ratios of 60:20:20. The Adam optimiser [9] with a learning rate of 0.0001 was used, and training was conducted over 100 epochs, with 10 iterations of training for each model. The best model was selected on the basis of the validation loss across epochs for each iteration of training.

2.4. METHODS FOR ESTIMATING/CALCULATING POWER CONSUMPTION

Floating Point Operations, or FLOPs was used to benchmark power consumption as it is agnostic to hardware specifications. To compute the FLOPs of the models themselves, two Python packages were used: ptflops [10] and fvcore from Facebook Research [11]. The results of these were compared and averaged to compute the estimated FLOPs of the model. The normalisation step consists of simply subtracting the mean and dividing by the standard deviation for each sample, giving a complexity of O(n) FLOPs.

The STFT computation is more challenging to benchmark as there are numerous implementations as well as hardware specific optimisations. Here, the number of FLOPs is estimated from the asymptotic computational cost of an FFT [12] over P samples within Q windows, i.e.

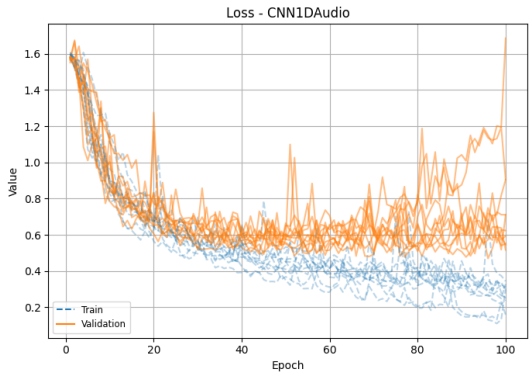
$$C_{\text{STFT}} \approx Q \times (2.5\, P \log_2 P) \tag{1}$$

is the estimated cost in FLOPs of the STFT.

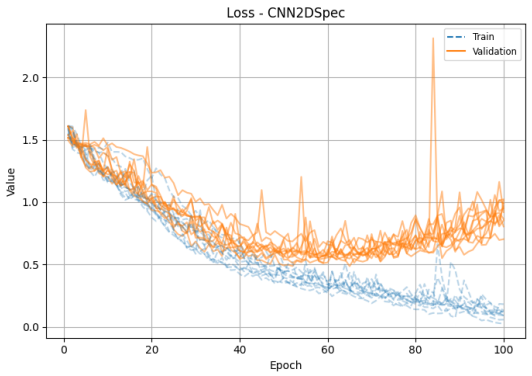
Finally, the background noise subtraction and conversion to dB scale need to be estimated. As background subtraction is just a division per sample; it is computed in O(n) FLOPs. The conversion to dB is not generally performed in FLOPs, depending on implementation, and may use methods such as lookup tables. However, it is assumed that this is generally O(n) complexity and so for this paper, only the STFT cost is used as all other costs are in the order of O(n) and thus are minimal compared to the model complexity and the Fourier transforms.

3. RESULTS

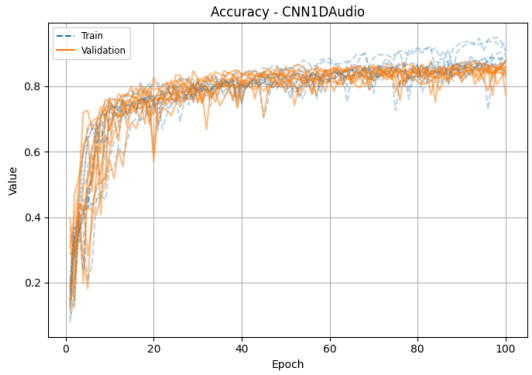
The total FLOPs for the STFT for $Q = 938$ windows and $P = 2048$ samples per window gives an estimated compute cost of just over 52.8 MFLOPs. This is an upper estimate, and hardware-specific implementations will very likely require fewer operations. Ptflops calculated 1163.43 MFLOPS and 1187.53 MFLOPs while fvcore calculated 1140.63 MFLOPS and 1164.85 for the 1D CNN and 2D CNN respectively. That gives an average of $\sim$1152 MFLOPs for the 1D model and $\sim$1176 for the 2D model.



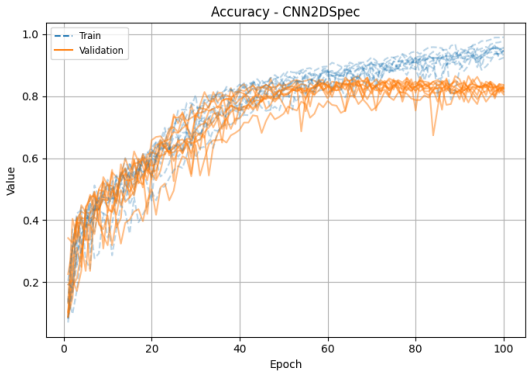
(a) 1D CNN training and validation loss plots over 10 iterations



(b) 2D CNN training and validation loss plots over 10 iterations



(c) 1D CNN training and validation accuracy plots over 10 iterations



(d) 2D CNN training and validation accuracy plots over 10 iterations

Figure 2: Comparison of 1D vs 2D CNN validation metrics.

Class	Precision	Recall	F1-score	Support
S1	**0.92**	0.77	0.84	149
S2	0.83	0.84	0.83	104
S3	0.65	0.78	0.71	72
SP4	0.83	0.87	0.84	27
SP5	0.82	**0.93**	**0.86**	34
Accuracy		*0.812*		386
Macro avg	0.812	0.837	0.816	386
Weighted avg	0.831	0.812	0.813	386

Table 2: Average test data classification report for 1D CNN trained on waveform input

Class	Precision	Recall	F1-score	Support
S1	**0.87**	0.83	**0.85**	149
S2	0.86	0.84	**0.85**	104
S3	0.68	0.70	0.69	72
SP4	0.78	**0.86**	0.82	27
SP5	0.80	0.84	0.82	34
Accuracy		*0.813*		386
Macro avg	0.799	0.816	0.804	386
Weighted avg	0.822	0.813	0.812	386

Table 3: Average test data classification report for 2D CNN trained on spectrograms

Model	Test Accuracy	Model MFLOPs	Pre-Processing MFLOPs	Total MFLOPs
1D-Kernel CNN	0.812	1152	Negligible	1152
2D-Kernel CNN	0.813	1176	<52.8	<1228

Table 4: Model accuracies and compute costs at inference, including pre-processing

4. DISCUSSION AND CONCLUSION

Despite the imbalanced dataset, the use of class weighting in the optimiser proved effective. With the exception of the S3 class, which both models are noticeably worse at classifying, performance was relatively consistent across all categories. The poorer performance of the S3 class is expected as this category may be more acoustically heterogeneous and comprised of multiple functionally distinct call types [7].

The 2D-kernel CNN trained on spectrogram inputs does not significantly outperform the 1D-kernel CNN trained on raw waveform inputs in terms of overall accuracy (0.813, 0.812 respectively) despite the increased pre-processing. As demonstrated in Fig. 2, which shows the performance of each model through training, both models exhibit initially relatively stable training dynamics without significant signs of overfitting. However, at around epoch 50-60 both models begin to spike in variance and begin to overfit.

Interestingly, despite similar architectures with nearly equal parameter counts and inference FLOPs, the 1D model converges faster, while the 2D model's validation performance improves more slowly, eventually achieving slightly better accuracy. However, given the class imbalance and challenges associated with S3, accuracy alone is insufficient to evaluate the models. Tables 2 and 3, which present class-level metrics, show that the 2D-kernel CNN slightly outperforms the 1D model on the two largest classes, S1 and S2, while the 1D model performs better on the smaller SP4 and SP5 classes. This trend may be due to the acoustic characteristics of the call types themselves, particularly if the SP4 and SP5 classes benefit from the phase information ommitted from the spectrograms in this paper. However, the slower training curve as well as the increased overfitting observed from in 2D CNN during later epochs may suggest that the architecture itself is less robust to imbalanced classification problems.

While the STFT adds a non-trivial computational cost of approximately 4–5% of total FLOPs at inference, this overhead is relatively minor. In scenarios where power consumption must be reduced such as in long-duration field deployments of UUVs or USVs, the loss of accuracy when choosing a 1D CNN is also minimal and may be justified. Furthermore, if primary target classes are sparsely represented, a waveform-based model may conserve energy while improving performance on critical but sparse acoustic signals since the 1D CNN appears to handle the minority classes (SP4 and SP5) more effectively. However, it remains unclear whether these performance differences stem from fundamental properties of the representation of the input signal and convolutional dimensionality (i.e. 1D vs. 2D inputs/kernels), if they are signal specific to SP4 and SP5, or if they are simply artifacts of variability due to the poorly performing and acoustically heterogeneous S3 class.

Further work is needed to determine whether the observed trends in minority class performance are due to signal-specific characteristics, class imbalance, incidental variability or other unknown factors. A clearer understanding of these factors could guide the development of more efficient, task-specific models that balance accuracy and resource constraints.

5. ACKNOWLEDGEMENTS

The authors would like to thank Seiche Water Technology Group and the ART-AI CDT for funding and supporting William Butler's PhD.

REFERENCES

[1] SEAGLIDER - kongsberg maritime - PDF catalogs | documentation | boating brochures. [Online]. Available: https://pdf.nauticexpo.com/pdf/kongsberg-maritime/seaglider/31233-105713.html

[2] D. Rosendo, A. Costan, P. Valduriez, and G. Antoniu, "Distributed intelligence on the edge-to-cloud continuum: A systematic literature review," *Journal of Parallel and Distributed Computing*, vol. 166, pp. 71–94, 2022.

[3] T. S. Prajwal and I. A K, "A comparative study of resnet-pretrained models for computernbsp;vision," in *Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing*, ser. IC3-2023. New York, NY, USA: Association for Computing Machinery, 2023, p. 419–425. [Online]. Available: https://doi.org/10.1145/3607947.3608042

[4] A. P. Hill, P. Prince, J. L. Snaddon, C. P. Doncaster, and A. Rogers, "Audiomoth: A low-cost acoustic device for monitoring biodiversity and the environment," *HardwareX*, vol. 6, p. e00073, 2019.

[5] T. A. Lamont, L. Chapuis, B. Williams, S. Dines, T. Gridley, G. Frainer, J. Fearey, P. B. Maulana, M. E. Prasetya, J. Jompa *et al.*, "Hydromoth: Testing a prototype low-cost acoustic recorder for aquatic environments," *Remote Sensing in Ecology and Conservation*, vol. 8, no. 3, pp. 362–378, 2022.

[6] L. J. Nowak, "Observations on mechanisms and phenomena underlying underwater and surface vocalisations of grey seals," *Bioacoustics*, vol. 30, no. 6, pp. 696–715, 2021.

[7] W. Butler, H. Smith, W. Lloyd, M. Impraimakis, A. Barnes, and A. Hunter, "Automated classification of vocalisations from wild and captive seal populations," June 2025, to appear in Proceedings of the Underwater Acoustics Conference and Exhibition (UACE 2025).

[8] B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th Python in Science Conference*, 2015, pp. 18–25.

[9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization. arxiv [preprint](2014)," *arXiv preprint arXiv:1412.6980*, vol. 5, 2014.

[10] V. Sovrasov. (2018–2024) ptflops: A flops counting tool for neural networks in pytorch framework. Accessed: May 2025. [Online]. Available: https://github.com/sovrasov/flops-counter.pytorch

[11] Facebook AI Research, "fvcore: A library for efficient model analysis," https://github.com/facebookresearch/fvcore, 2025, accessed: May 2025. [Online]. Available: https://github.com/facebookresearch/fvcore

[12] M. Frigo and S. G. Johnson, "The design and implementation of fftw3," *Proceedings of the IEEE*, vol. 93, no. 2, pp. 216–231, 2005.

An Investigation of Machine Learning Capabilities for Cavitation Detection

D. Smith[1] and O. Carter[1]

[1]QinetiQ, Rosyth, Scotland

D. Smith: 2 Aquarius Court, Rosyth, KY11 2DW, Scotland, dasmith6@qinetiq.com
O. Carter: 2 Aquarius Court, Rosyth, KY11 2DW, Scotland, ocarter@qinetiq.com

Abstract: *Due to environmental concerns, there is a growing need to reduce the radiated noise of marine traffic. The noise from marine vessels comprises multiple sources, with propeller cavitation typically being the dominant source when present. Cavitation can be avoided by operating the vessel at lower speeds. However, the speed at which cavitation occurs may not be explicit and may change over the life of the vessel. Therefore, in order to ensure noise levels are minimised, there is a need to detect when cavitation occurs and take the necessary corrective action. In this contribution, the capability of machine learning in the detection of propeller cavitation is evaluated. Using time series signals as the input feature, the performance of two models is evaluated. Special attention is given to the signal to noise ratio to more fully understand the capabilities of the models and to determine if any signal conditioning is required to increase performance. A number of signal statistics are also evaluated to identify additional features that may accurately characterise the presence of cavitation. The present work enables a rapid detection of cavitation and may allow for vessels to avoid this noisy operating state and hence reduce the impact on the marine environment.*

Keywords: *Propeller Cavitation, Classification, Detection.*

1. INTRODUCTION

With the increasing awareness of the impact of noise from shipping on marine life, there is a growing need to reduce the noise from ships. The noise radiated from ships is comprised of multiple complex sources with propeller cavitation typically being the dominant source. Whilst efforts can be made to reduce cavitation at the design stage, its control is more difficult for in-service ships. It is typically suggested to operate at lower speeds to avoid cavitation. However, the speed at which cavitation occurs is not always evident due to its dependence on local conditions at the propeller blade. Further, for variable pitch propellers, cavitation can occur, even at low rotational speeds. Therefore, in order to mitigate the effects of ship noise, there exists a need to detect the presence of cavitation during ship operations.

Cavitation can be detected following analysis of measured vibration or acoustic pressure in a location near the propeller. Whilst cavitation manifests as an increase in measured level, advanced signal processing methods are usually applied to distinguish cavitation from other sources of increased noise. Traditionally, DEMON processing has been used for identifying cavitation, requiring operator knowledge in order select appropriate frequency bandwidths [1]. More recently, Cyclostationarity has been proposed [2, 3], whilst removing the operator input, tuning is still required to select the appropriate signal overlap and peak-picking parameters.

To overcome the need for an operator to analyse measured signals, we investigate the use of Machine Learning (ML) for detecting the presence of cavitation. In particular, we investigate the utility of time-series classification methods in order to overcome the need for advanced signal processing of the measured data. Previous work [4] highlighted the challenges of time-series classification for cavitation detection using readily available ML models. One of the challenges identified was the effect of Signal to Noise Ratio (SNR) in the data. In this contribution, we build on the previous work with an investigation on ML approaches more suited for time series classification, evaluating their performance across a range of SNRs.

2. METHODOLOGY

2.1. DATA

For the present work, synthetic data has been used in the evaluation of the ML methods. Synthetic data has been used as it allows close control of specific properties of the data. In particular, synthetic data has been generated covering a range of SNRs. Six data sets were generated. For five of these data sets, the data within each was constructed with a fixed SNR. Namely, data sets were generated for SNRs of -30, -20, -10, 0, and +10 dB. The final data set was constructed from a range of SNRs between -30 and +10 dB.

Hull vibration or hydrophone measurements made in the proximity of a cavitating propeller are characterised by an increased level across a broad frequency range, modulated by the propeller blade rate and its harmonics. To avoid a model simply training on an increased level and not the underlying features of a cavitating signal, there was not an overall increase in levels between non-cavitating and cavitating signals.

The synthetic data is generated by the summation of continuous and broadband components. Where cavitation is present, the broadband component is modulated. The continuous component is modelled as Gaussian white noise and controlled the SNR in the overall signal. The

broadband component is filtered Gaussian white noise and the modulated component is produced by a summation of harmonics at multiples of the blade rate. To increase the variability, the modulating component was scaled according to a cavitation index, randomly generated in the range zero to one. For signals where no cavitation is required, the modulating component is excluded.

Across the analysis, 500 samples were nominally used, split 80/20 for training/testing. Further, an equal split of cavitating/non-cavitating signals was used throughout.

2.2. MACHINE LEARNING METHODS

To avoid the need for data preprocessing, the present work investigated the capabilities of time series classification methods. To this end, the Time Series Forest (TSF) [5] was first evaluated. The TSF is a relatively simple approach to extend the Random Forest algorithm to support time series classification. The TSF is based on identifying a number of random intervals of the time series signal and using the properties of the signal in these intervals as the input features for the random forest model [6]. In particular, the mean, standard deviation and slope of each interval is computed and used as input features. Fig. 1 shows a schematic of the model feature generation. The number of input features is therefore $3N$, with N the number of segments.

The kurtosis and skewness have been investigated to understand if they can better characterise the presence of cavitation within a signal. These measures have previously been used within time series classification [8, 9]. The TSF classifier is therefore modified, with additional segment features of kurtosis and skewness. Thus, increasing the number of features to $5N$.

In the subsequent analysis, the TSF model hyperparameters were tuned for each evaluation. In particular, the minimum number of windows and minimum window length were tuned. The default hyperparameters for the underlying random forest model were not modified as they had a minimal impact on model performance.

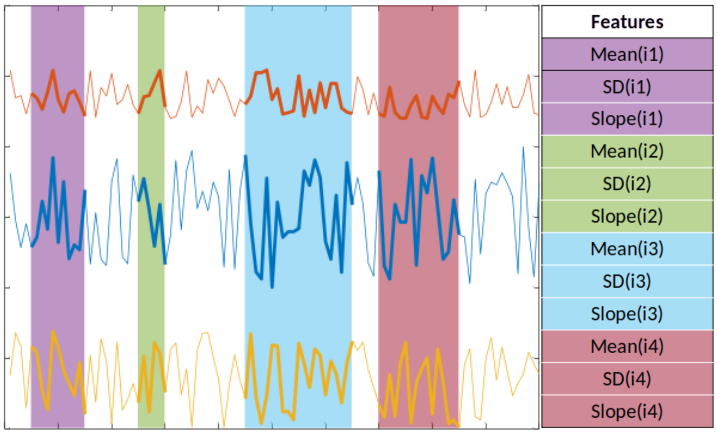


Figure 1: Schematic of the Time Series Forest

A Feed Forward Convolutional Neural Network was adopted to compare with the TSF approach utilising more training parameters and a larger input dimension of the full time series signal. The Fully Convolutional Network (FCN) architecture from [10] was adopted as a simple test bed to analyse full time waveform binary classification. The structure of this network allows for a progressive feature learning order; basic, complex and refined. This ordering aims to prevent early compression or loss of features. Global Average pooling is used to compress

features into predictions and helps to avoid over-fitting.

3. RESULTS & DISCUSSION

Fig. 2 compares the F1-score for the unmodified TSF and the TSF with kurtosis and skewness as additional segment features. The performance of both models is seen to increase as the SNR increases. In the case of the lowest SNR data, the unmodified TSF outperforms the modified model. At increasing SNR, the modified TSF shows improved performance, indicating the utility of the kurtosis and skewness as input features. This aligns to the findings of a separate feature importance analysis carried out using a Random Forest Classifier.

For the data set comprised of data across a range of SNRs, a relatively poor performance is observed, particularly for the unmodified TSF. Inclusion of the kurtosis and skewness shows a significant improvement in model performance. The results for the case are expanded in the confusion matrices shown in Fig. 3. The unmodified TSF generally shows the model has not been able to learn well, essentially guessing the classification. On the other hand, modified TSF shows an improved performance in predicting both classes. Given the increased variability of this data, increasing the data size is suggested to improve the overall performance for this case.

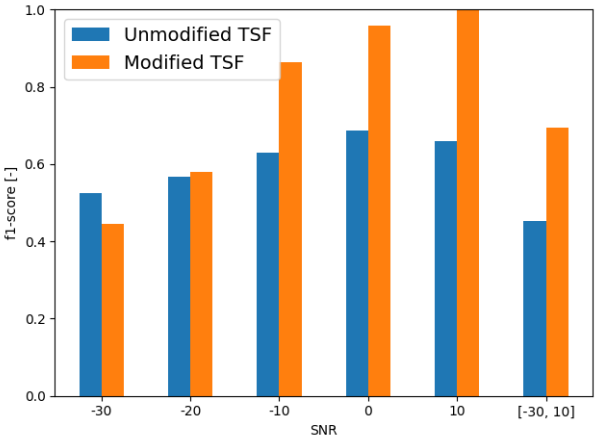
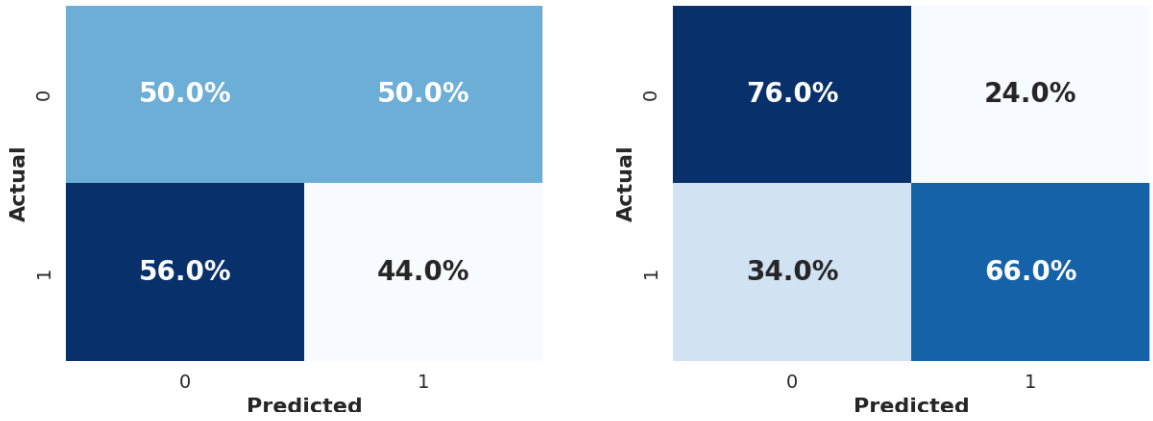


Figure 2: TSF performance across SNR range.

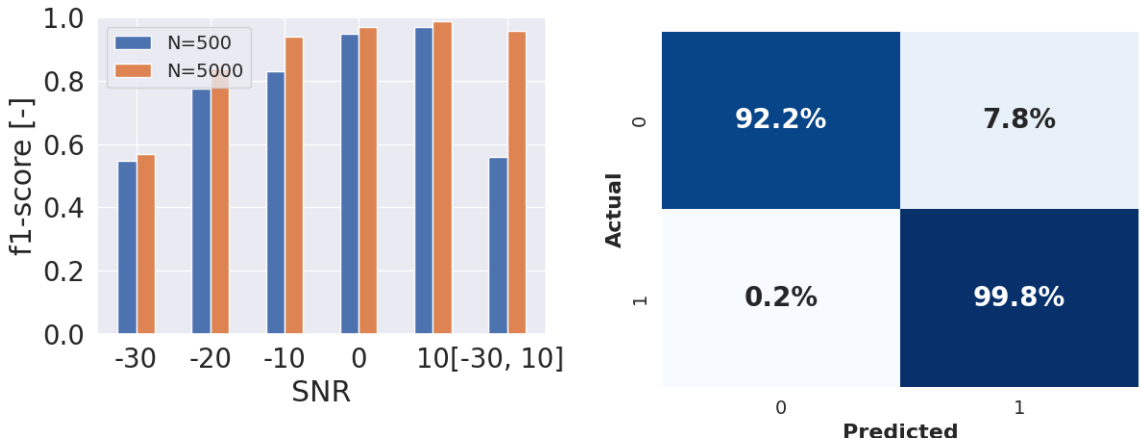
Fig. 4 shows the F1-score for the FCN and the corresponding confusion matrix for the data across a SNR range. The FCN performance shows a similar trend to that observed for the TSF, with increasing performance as the SNR increases. However, the FCN shows a greater performance at lower SNR.

Using the nominal sample size of 500, the performance of the FCN for the data comprising a range of SNR was relatively poor. It was found that by increasing the sample size, it was possible to drastically improve this, with the corresponding confusion matrix showing the ability to accurately classify both cases. For the data with fixed SNR, the nominal sample size was found suitable. This difference in performance, as identified for the TSF is suggested to be due to the increased variability within this data. The precise sample count required to increase the performance for this case has not yet been determined, with the larger value used to demonstrate the capabilities. The precise data requirements will be the focus of future work.



(a) Unmodified TSF *(b) Modified TSF*

Figure 3: Confusion matrices for the unmodified and modified TSF.



(a) F1 score for different sample sizes. *(b) Confusion matrix, [-30,10] dB range with N = 5000.*

Figure 4: FCN performance across SNR range for varying sample sizes.

4. CONCLUSIONS

The capability of two Machine Learning methods for detecting cavitation has been demonstrated in the present work. Using synthetic data to control the noise level, the capability of a Time Series Forest and a Fully Convolutional Network has been explored.

The TSF showed increasing performance as the SNR increased, indicating that some level of signal conditioning may be required for very low values of SNR. The utility of kurtosis and skewness in aiding in the detection of classification was explored. Modifying the TSF, the kurtosis and skewness were found not to be beneficial in circumstances of very low SNR. However, as the SNR increased, they demonstrated significant improvements in the TSF classification capability.

The FCN generally showed improved performance when compared to the TSF. However, when evaluating the data whose signals were constructed from a range of SNRs, the performance was comparable to the TSF. FCN performance increased dramatically when the number

of samples was increased which may not always be a feasible option in practice. However, the precise data requirements should be studied further.

Further work will apply the developed models to a measured data set to evaluate their capabilities to real data. Further work will also focus on comparing the two explored models and identifying the most suitable approach for transitioning to a deployable service as a cavitation detection system.

REFERENCES

[1] D.B, A.M, V. P. S and S. K. Ambat, " Performance Evaluation of the DEMON Processor for Sonar", *2022 IEEE Region 10 Symposium*, 2022.

[2] J. Antoni "Cyclostationary by examples", *Mechanical Systems and Signal Processing*, Vol. 23, No. 4, 2009.

[3] D. Hanson, J. Antoni, G. Brown and R. Emslie, " Cyclostationarity for Passive Underwater Detection of Propeller Craft: A Development of DEMON Processing", *Proceedings of Acoustics 2008*, 2008.

[4] D. Smith, B. McKinklay and R. Potter: "Using Machine Learning for the Detection of Propeller Cavitation" in *Proceeding of the Institute of Acoustics*, Vol. 46, Pt. 1, 2024.

[5] H. Deng, G. Runger, E. Tuv, M. Vladimir, "A time series forest for classification and feature extraction", *Information Sciences*, vol. 239, 2013.

[6] L. Breiman "Random Forests", *Machine Learning*, Vol. 45, No. 1, 2001.

[7] J. Faouzi. "Time Series Classification: A review of Algorithms and Implementations", *Ketan Kotecha. Machine Learning (Emerging Trends and Applications)*, Proud Pen, In Press, 978-1-8381524-1-3.

[8] R. Tiwari, D. J. Bordoloi and A. Dewangan, "Blockage and cavitation detection in centrifugal pumps from dynamic pressure signal using deep learning algorithm", *Measurement*, vol. 173, no. 108676, 2021.

[9] P. Phukpattaranont, N. Thiamchoo, P. Neranon, "Real-time identification of noise type contaminated in surface electromyogram signals using efficient statistical features", *Medical Engineering Physics*, vol. 131, no. 104232, 2024.

[10] Z. Wang, W. Yan and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline". *2017 International joint conference on neural networks (IJCNN)* (pp. 1578-1585), IEEE, 2017.

Structured Session
07. Inverse Problems in Acoustical Oceanography

Organised by Julien Bonnel and Stan Dosso

Underwater Passive Source Localization based on a Data Derived Replica Library

Baptiste Menetrier[1], Gnouregma Bazile Kinda[2], Valérie Labat[1], Samuel Pinson[1], and Abdel Boudraa[1]

[1]Arts et Métiers Institute of Technology, École navale, IRENav, BCRM Brest, CC 600, Brest, 29240, France
[2]DGA Naval Systems, Brest, 29200

Baptiste Menetrier, Arts et Métiers Institute of Technology, École navale, IRENav, BCRM Brest, CC 600, Brest, 29240, France. Email: baptiste.menetrier@ecole-navale.fr

Abstract: *Passive localization of underwater acoustic source using a sparse sensor network is a challenging task due to the complex nature of the ocean environment. Matched Field Processing (MFP) techniques have drawn considerable attention in the literature as a potential solution to this challenge. Generally speaking, this technique can be viewed as a beamformer which matches the measured data with dictionary replicas to estimate the source location. However, its successful implementation in real-life operating conditions remains limited due to sensitivity to environmental mismatch. In the case of low-frequency signals, such as radiated ship noise, uncertainties in waveguide properties such as bathymetric features or seabed parameters can lead to significant localization errors. To address these challenges, MFP methods based on data-derived replicas have been recently explored. To overcome the limitations of cross-correlation-based data-derived replicas, this paper investigates the potential of Relative Transfer Function (RTF) vectors as localization features in an MFP approach. The distance between the RTF vector associated with an unknown radiator and the library vectors is evaluated using the so called Hermitian angle. The key idea of the method is to exploit the environmental diversity to construct replicas that smoothly map to source locations. The potential of this approach is evaluated through simulations. A synthetic dataset is generated using the KRAKEN normal mode propagation code in a deep water environment and the ultra-low-frequency domain. Propagated time series are synthesized using a ship noise model as the source signal, and performance is assessed across different noise scenarios, including additive white Gaussian noise and interference from propagated signals. Results demonstrate the feasibility of the proposed method.*

Keywords: *Passive source localization, deep water, ultra-low-frequency, relative transfer function*

1. INTRODUCTION

Source localization has been a longstanding topic of interest in underwater acoustics. While the classical matched field processing (MFP) technique has received significant attention in the literature, its application under realistic operating conditions remains limited [1, 2]. To mitigate the well-known environmental mismatch problem associated with MFP methods, model-based replicas can be replaced by data-derived replicas [3]. Traditional approaches construct the measured replica library using cross-correlation vectors computed between sensor pairs [4]. The position of an unknown source is then estimated by comparing its cross-correlation vector to those in the library through a second cross-correlation measure.

The relative transfer function (RTF), developed for audio and speech applications, is defined as the ratio between the transfer functions of a sensor and a reference sensor [5] and has been shown to effectively fingerprint acoustic environments, making it well-suited for source localization tasks [6]. Instead of the cross-correlation vector, in this work we investigate the potential of RTF vector in a data-derived replica-based MFP framework. This vector is estimated from the signal covariance matrix [7, 8], and source localization is achieved by comparing RTF vectors, which are complex quantities, using the mean Hermitian angle [8].

After a brief presentation of the method, its performance is assessed using simulated data in a realistic range-dependent deep-water environment where ship noise signals are propagated in the 5–50 Hz frequency band using normal mode theory. We then present a comprehensive analysis of the results. Robustness is evaluated under various white Gaussian noise conditions, and performance is quantified using the mainlobe-to-sidelobe ratio (MSR) and root mean square error (RMSE). The robustness against interferences is also tested in the presence of blue whale Z-call signals propagated in the same waveguide.

2. METHODS

We assume a network composed of N receivers. The noisy pressure signal $p_n(l,k)$ recorded by the n-th sensor can be expressed in the short-time Fourier transform (STFT) domain by:

$$p_n(l,k) = h_n(k)s(l,k) + v_n(l,k), \tag{1}$$

where the source signal and noise at the k-th frequency bin ($k \in \{1,\dots,K\}$) and the l-th frame ($l \in \{1,\dots,L\}$) are denoted by $s(l,k)$ and $v_n(l,k)$ respectively. The acoustic transfer function (ATF) relating the source and the n-th receiver is denoted by h_n and is assumed to be frequency-dependent and time-invariant. Equation (1) is extended to vector notation to represent signals received by the N sensors. The noisy pressure vector is noted $\mathbf{p}(l,k) = [p_1(l,k), p_2(l,k), \dots, p_N(l,k)]^T$. Vectors $\mathbf{h}(k)$ and $\mathbf{v}(l,k)$ are defined analogously, and the vector of propagated signals is denoted by $\mathbf{z}(l,k) = \mathbf{h}(k)s(l,k)$.

The RTF is defined as the ratio between an ATF and a given reference ATF. Without loss of generality, we assume the first receiver to be the reference receiver and the RTF vector is defined by:

$$\mathbf{d}(k) = \left[\frac{h_1(k)}{h_1(k)}\ \frac{h_2(k)}{h_1(k)}\ \cdots\ \frac{h_N(k)}{h_1(k)}\right]^T = \frac{1}{h_1(k)}\mathbf{h}(k). \tag{2}$$

Assuming that signal and noise are mutually uncorrelated, the spatial covariance matrix of the noisy signals $\mathbf{R_p} \in \mathbb{C}^{N\times N}$ is given by:

$$\mathbf{R_p}(k) = \mathbb{E}\{\mathbf{p}(l,k)\mathbf{p}^H(l,k)\} = \mathbf{R_z}(k) + \mathbf{R_v}(k) \tag{3}$$

where $\mathbb{E}\{\bullet\}$ is the expectation operator and $\{\bullet\}^H$ denotes the Hermitian transpose operator. $\mathbf{R_z}$ and $\mathbf{R_v}$ are respectively the propagated signal and noise covariance matrices. Signal and noise are assumed to be realizations of ergodic processes and covariance matrices $\mathbf{R_p}$ and $\mathbf{R_v}$ are estimated by the corresponding sample covariance matrices $\hat{\mathbf{R}}_\mathbf{p}$ and $\hat{\mathbf{R}}_\mathbf{v}$ respectively evaluated on signal plus noise and noise only periods. Assuming noise to be stationary and using relation (3), the propagated signal covariance matrix is given by:

$$\hat{\mathbf{R}}_\mathbf{z}(k) = \hat{\mathbf{R}}_\mathbf{p}(k) - \hat{\mathbf{R}}_\mathbf{v}(k). \tag{4}$$

In the single source scenario, the RTF vector is proportional to the unique non-zero eigenvalue of $\mathbf{R_z}$ and can be estimated as the best rank 1 approximation of $\mathbf{R_z}$ given by [7]:

$$\hat{\mathbf{d}}(k) = \frac{\mathbf{v_{max}}(k)}{\mathbf{e_1^T v_{max}}(k)}, \tag{5}$$

where $e_1 = \begin{bmatrix}1 & \mathbf{0}_{1\times(N-1)}\end{bmatrix}^T$ with $\{\bullet\}^T$ denoting the transpose operator and $\mathbf{v_{max}}$ is the principal eigenvector of $\hat{\mathbf{R}}_\mathbf{z}$.

RTF vectors associated with a set of known source positions (x, y, z) are stored and form the library of measured replicas used in the MFP algorithm. These library replicas, $\mathbf{d}^{(l)}(x, y, z)$, are compared to the RTF vector $\mathbf{d}^{(e)}$ of an unknown passing by acoustic radiator, referred to as the event RTF vector. The distance between two vectors is evaluated using the mean Hermitian angle $\bar{\theta}(x, y, z)$ [8] over the K common frequency bins shared by the library sources and the event source:

$$\bar{\theta}(x,y,z) = \frac{1}{K}\sum_{k=1}^{K} \arccos\left(\frac{|\mathbf{d}^{(l)}(x,y,z,k)^H \mathbf{d}^{(e)}(k)|}{\|\mathbf{d}^{(l)}(x,y,z,k)\|\|\mathbf{d}^{(e)}(k)\|}\right). \tag{6}$$

The estimated source position $\hat{\mathbf{x}}_\mathbf{s}$ $(\hat{x}_s, \hat{y}_s, \hat{z}_s)$ is defined as the position minimizing the distance $\bar{\theta}$.

3. SIMULATION SETUP

To assess the potential of the proposed method, a synthetic dataset is generated for a realistic range-dependent deep-water environment. The bathymetric and sound speed profiles used are shown in Fig. 1. The seabed is modeled as a semi-infinite layer of coarse sediment with density $\rho = 2.37$ g cm^{-3}, compression wave celerity $c_p = 2122$ m s^{-1} and attenuation $\alpha = 0.88$ dB λ^{-1}.

The source signals are designed to replicate the typical width and shape of ship noise tonal lines in the time-frequency domain. This can be achieved by incorporating the stochastic nature of ship noise periodicity. Accordingly, the source signal is modeled as a quasi-periodic signal with dominant frequency f_0. It is represented by the convolution of a quasi-periodic series of Dirac delta functions $\delta(t)$ and a pulse $e(t)$ [9]:

$$s(t) = e(t) * \sum_{k=-\infty}^{+\infty} \delta\left(t - kT_0 - \sum_{m=-\infty}^{k} \psi_m\right). \tag{7}$$

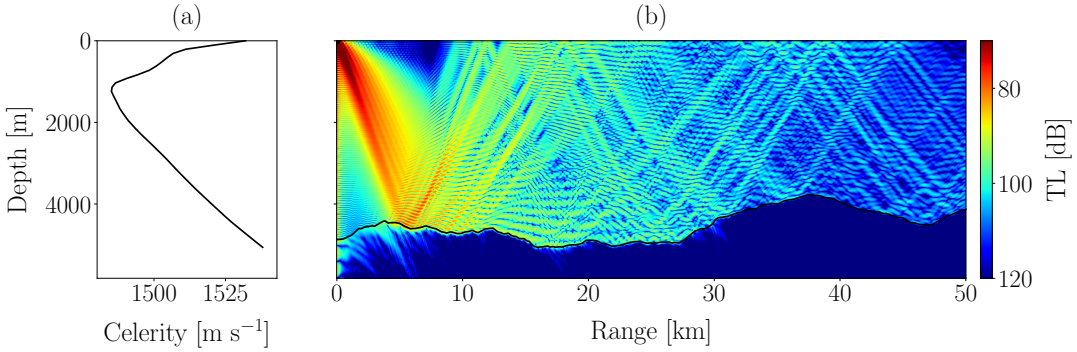


Figure 1: Deep water environment used for broadband modeling. (a) Sound celerity profile, (b) transmission loss (TL) in dB at 20 Hz derived with KRAKEN (coupled modes).

where $T_0 = \frac{1}{f_0}$ and ψ_m denotes the zero-mean discrete Gaussian random fluctuation of the m-th period. The pulse $e(t)$ is defined as the inverse Fourier transform of the function $E(f) = [1 - f^2/f_c^2 + i1/Qf/fc]^{-1}$ where $f_c = 30\,\text{Hz}$, $Q = 0.9$ and i is the imaginary unit. As described in [9], Equation (7) can be expressed as a sum of signal harmonics $s_n(t)$:

$$s(t) = \sum_{n=-\infty}^{+\infty} \hat{e}(n\omega_0) \times \exp\left[\mathrm{i}n\omega_0 \left(t - \int_{-\infty}^{t} \psi(u)\mathrm{d}u\right)\right] = \sum_{n=-\infty}^{+\infty} s_n(t) \tag{8}$$

The random variable $n\omega_0\psi(t)$ represents the instantaneous frequency (IF) fluctuation and can be characterized by its autocorrelation function $R_{\delta_f}(\tau)$. This autocorrelation function is here assumed to be Gaussian and is defined by $R_{\delta_f}(\tau) = \sigma^2 \exp\left[-\tau^2/2\tau_c^2\right]$ where $n\omega_0\sigma$ denotes the standard deviation of the IF fluctuation $n\omega_0\psi(t)$, while τ_c represents its correlation time. The IF fluctuation of the n-th harmonic $n\omega_0\psi(t)$ is build according to its second-order statistics and the corresponding time-domain signal $s(t)$ is derived using Equation (8). Ship noise signals are simulated in the ultra-low-frequency (ULF) band ranging from 5 to 50 Hz. As illustrated in Fig. 2.a and Fig. 2.b, the simulated signal replicates the harmonic thickness observed from real data.

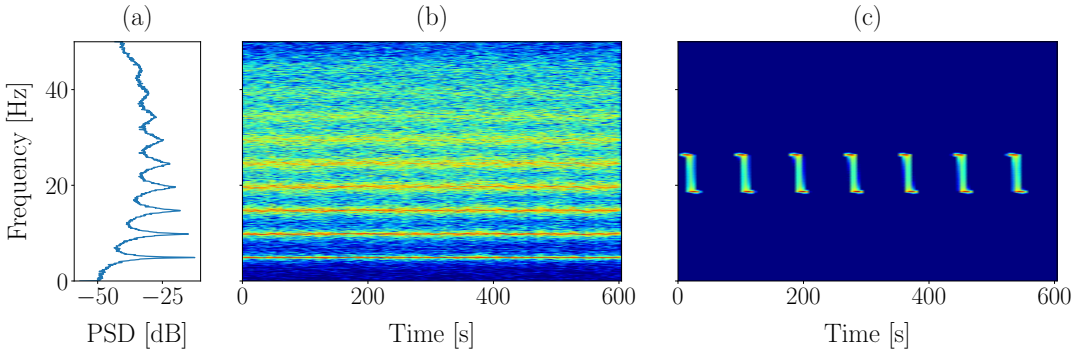


Figure 2: Time-frequency characteristics of ship noise signal ($f_0 = 4.9$ Hz, $\sigma = 0.09\ f_0$, $\tau_c = 0.07\ T_0$) and Z-call signal derived with Hann windows, 2048 samples and 75 % overlap. (a) Ship noise power spectral density (PSD). (b) Ship noise spectrogram. (c) Z-call spectrogram.

Broadband Green's functions are computed using a normal mode propagation model [10]

and synthetic propagated time series are obtained by Fourier synthesis. The range-dependent bathymetric profile is sampled at 200 m intervals, and the coupled-mode formulation is employed with the first 100 modes for each frequency.

4. RESULTS

The network is composed of six receivers randomly distributed within a 5 kilometers circle centered on the origin of the $(\mathcal{O}, \mathbf{e_x}, \mathbf{e_y})$ frame and located 1 m above the seafloor. The library dataset is generated with the ship noise signal described earlier with $f_0 = 4.9$ Hz, $\sigma = 0.09\ f_0$ and $\tau_c = 0.07\ T_0$. The source to localize is positioned at $\mathbf{x_s}$ $(x = 30$ km$, y = 23$ km$, z = 5$ m$)$ and emits a ship noise signal with parameters $f_0 = 4.65$ Hz, $\sigma = 0.12\ f_0$ and $\tau_c = 0.30\ T_0$. The search area is a 1 km $\times$ 1 km grid, centered on $\mathbf{x_s}$ and divided into 20 m $\times$ 20 m cells.

The robustness of the proposed method is first evaluated in the presence of additive white Gaussian noise (AWGN). Localization performance is quantified using the RMSE of the estimated position, while the ambiguity of the localization solution is assessed through the MSR [11]. A Monte Carlo analysis is conducted over 100 iterations for various signal-to-noise ratios (SNRs). In each iteration, an independent realization of Gaussian white noise v_n is added to the propagated signals, and the noise covariance matrix $\hat{\mathbf{R}}_\mathbf{v}$ is estimated from a separate noise-only realization. An illustration of the ambiguity surface generated by the proposed method is shown in Fig. 3.b for an SNR of 0 dB. The ambiguity surface exhibits a well-defined mainlobe centered at the true source position, along with two relatively low sidelobes. The dynamic range of the surface is satisfactory, with an MSR of 1.9 dB. As illustrated in Fig. 3.a, both the RMSE and the MSR improve consistently with increasing SNRs. Notably, for SNRs exceeding 0 dB, the source can be localized with an RMSE below 20 m, corresponding to ambiguity surfaces with MSRs higher than 2 dB.

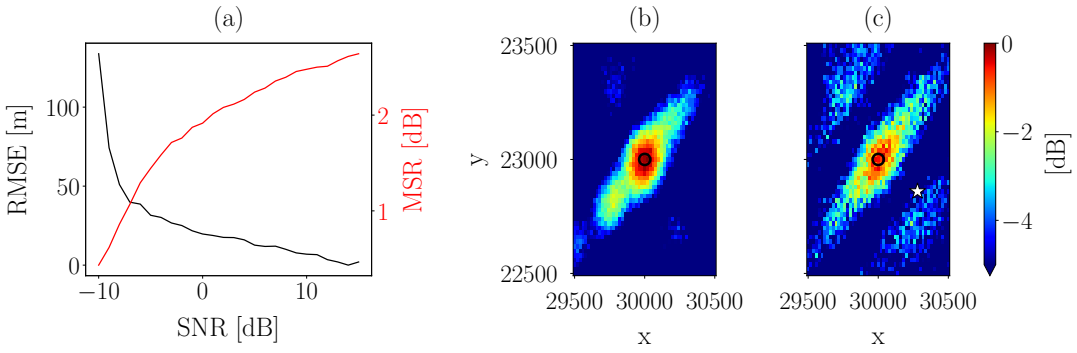


Figure 3: (a) RMSE [m] and MSR [dB] across increasing SNRs. (b) Ambiguity surface with AWGN (SNR = 0 dB), (c) ambiguity surface with AWGN (SNR = 0 dB) and interference z-call signal (SIR = -5 dB). The black circle on (b) and (c) denotes the true target position, white star on (c) is the position of the interfering blue whale.

The robustness is further assessed in the presence of interfering blue whale Z-call signals, propagated from the position $\mathbf{x}_{\text{bw}}$ $(x = 30.28$ km$,\ y = 22.86$ km$,\ z = 5$ m$)$. Z-calls can be modeled as frequency-modulated signals $s_{\text{z-call}}(n) = a(n)\text{e}^{\text{i}\phi(n)}$ where the phase $\phi(n)$ is defined by the parametric model proposed by [12] with $f_c = 22.5$ Hz, $U = 3.2$ Hz, $L = -4.5$ Hz, $\alpha = 1.8$ Hz and $M = 10$ s (Fig. 2.c). The AWGN SNR is fixed at 0 dB, and the received

Z-call signal is scaled to achieve a signal-to-interference ratio (SIR) of -5 dB. In the presence of the Z-call signal, the accuracy of RTF estimation degrades within the frequency band of the interference, leading to a noisier ambiguity surface and a lower MSR (Fig. 3.c). Nevertheless, a distinct mainlobe is still visible, and the method achieves reliable localization performance with an RMSE of 44 m and an MSR of 1.4 dB (evaluated over 100 iterations).

5. CONCLUSIONS

This paper presented a passive source localization method based on a data-driven library of RTF vectors. Its performance was evaluated using simulated data generated in a realistic, range-dependent deep-water environment, in the ULF band. Source signals were carefully modeled to replicate the main time-frequency characteristics of ship noise, improving relevance of the simulation to real-world scenarios. The results demonstrate the feasibility and effectiveness of the RTF-based MFP framework, showing accurate source position estimation even under low SNR conditions. Furthermore, the method was evaluated in the presence of interfering blue whale Z-call signals. Despite the additional interference, accurate localization performance is maintained, demonstrating the robustness of the proposed approach.

REFERENCES

[1] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky. An overview of matched field methods in ocean acoustics. *IEEE J. Oceanic Eng.*, 18:401–424, 1993.

[2] A. B Baggeroer. Why did applications of MFP fail, or did we not understand how to apply MFP? In *Proc. 1st Int. Conf. and Exhib. on Underwater Acoustics. Greece*, pages 41–49, 2013.

[3] L. T. Fialkowski et al. Matched-field processing using measured replica fields. *J. Acoust. Soc. Am*, 107:739–746, 2000.

[4] C. M. A. Verlinden, J. Sarkar, W. S. Hodgkiss, W. A. Kuperman, and K. G. Sabra. Passive acoustic source localization using sources of opportunity. *J. Acoust. Soc. Am*, 138:EL54–EL59, 2015.

[5] S. Gannot, D Burshtein, and E. Weinstein. Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Trans. Signal Process.*, 49:1614–1626, 2001.

[6] A. Bross and S. Gannot. Training-based multiple source tracking using manifold-learning and recursive expectation-maximization. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 31:1124–1140, 2023.

[7] R. Serizel, M. Moonen, B. Van Dijk, and J. Wouters. Low-rank approximation based multichannel wiener filter algorithms for noise reduction with application in cochlear implants. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 22:785–799, 2014.

[8] R. Varzandeh, M. Taseska, and E. A. P Habets. An iterative multichannel subspace-based covariance subtraction method for relative transfer function estimation. In *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)*, pages 11–15, 2017.

[9] S. Pinson. Modeling of a ship sound recorded by an ocean bottom seismometer. In *Forum Acusticum 2025*, 2025. Accepted paper.

[10] M. B. Porter. Porter - 2001 - The KRAKEN Normal Mode Program, 2001.

[11] T. Zhang, D. Zhou, L. Cheng, and W. Xu. Correlation-based passive localization: Linear system modeling and sparsity-aware optimization. *J. Acoust. Soc. Am*, 154:295–306, 2023.

[12] F. X. Socheleau et al. Automated detection of antarctic blue whale calls. *J. Acoust. Soc. Am*, 138:3105–3117, 2015.

Sequential Measurement Point Selection For Geoacoustic Inversion Using Unscented Kalman Filtering

Siobhán Correnty[1], Ylva Ljungberg Rydin[1], Magnus Lundberg Nordenvad[1], Isaac Skog[2], and Martin Östberg[1]

[1]Swedish Defence Research Agency (FOI), Sweden
[2]KTH Royal Institute of Technology, Sweden

Siobhán Correnty, Swedish Defence Research Agency (FOI), Gullfossgatan 6, 164 40 Kista, Sweden, siobhan.correnty@foi.se

Abstract: *The problem of geoacoustic inversion using acoustic transmission loss measurements is considered. A recursive Bayesian method based on the unscented Kalman filter is proposed to sequentially estimate the unknown sound propagation model parameters and select the measurement points that minimize the estimation uncertainty. The proposed method is evaluated in a simulation study emulating the estimation of geoacoustic seabed parameters in shallow waters. The results show that the proposed sequential measurement point selection method improves estimation performance compared to a fixed preplanned measurement point selection strategy. Moreover, the distribution of the selected measurement points over time shows that the proposed method tends to choose measurement points associated with a lower sound speed.*

Keywords: *Sequential decision making, inverse modeling, acoustic propagation, unscented Kalman filtering*

1. INTRODUCTION

Sound propagation in shallow waters is highly affected by the seabed and the underlying bedrock's constitution. Therefore, detailed information on sediment layering and physical properties is needed to predict transmission loss accurately. Many methods exist to acquire such information, from costly methods, such as seismic surveying and sediment sample analysis, to more crude and cost-efficient methods, such as inverse modeling using transmission loss measurements. This paper focuses on the latter, i.e., inverse modeling, and how the collection of measurements can be done efficiently. Specifically, the paper focuses on sequentially selecting the measurement points to minimize the uncertainty in the estimated model parameters.

To select the measurement points sequentially, the applied inverse modeling techniques must quantify the uncertainty in the estimated model parameters. Preferably, recursively updating parameter estimates should also be possible as more measurements are gathered. Moreover, it should be possible to include prior knowledge about the model parameters.

Traditional optimization methods for inverse modeling, such as Simulated Annealing [1] and genetic algorithms [2], generally only provide a point estimate of the model parameters and do not support recursive updating of the parameter estimates as mode measurement points are collected; prior knowledge about the model parameters can be included by adding regularizations to the loss function used in the optimization. Therefore, Bayesian methods have recently been developed for inferring the parameters in underwater sound propagation loss models [3]. These methods enable the uncertainty in the estimated parameters to be quantified and prior knowledge about the model parameters to be included in a structured way.

Sequential measurement point selection to optimize the information gathered about a system or physical process is a well-established research field. However, to the authors' knowledge, no work has been done on sequential measurement point selection in the context of underwater propagation loss modeling. Nor have the benefits of sequentially selecting the measurement points to minimize the uncertainty about the model parameters been analyzed. To that end, we propose a recursive Bayesian framework based upon the unscented transform to: *(i)* recursively estimate the parameters in the sound propagation model from transmission loss measurements; and *(ii)* select the measurement points that minimize the covariance of the parameter estimates. The proposed method is evaluated in a simulation in which the parameters of the fluid sediment layer are determined. The results indicate that the proposed method can significantly reduce the uncertainty in the estimated parameters compared to using predefined measurement points.

2. PROBLEM FORMULATION AND PROPOSED SOLUTION

Consider the problem of estimating the parameters θ in a sound propagation model from a sequence of transmission loss measurements $y_{1:k}$. The measurements are collected with an array of hydrophones at a fixed location and a movable transmitter that takes on the positions $s_{1:k}$. Here y_i denotes the i:th measurement collected with the transmitter at position s_i.

From a Bayesian perspective, the posterior distribution $\pi(\theta|y_{1:k}, s_{1:k})$ of the parameters θ can be recursively calculated from the measurements via the Bayesian recursions

$$\pi(\theta|y_{1:k}, s_{1:k}) = \frac{\ell(y_k|\theta, s_k)\pi(\theta|y_{1:k-1}, s_{1:k-1})}{\int \ell(y_k|\theta, s_k)\pi(\theta|y_{1:k-1}, s_{1:k-1})d\theta}, \qquad \pi(\theta|y_{1:-1}, s_{1:-1}) \equiv \pi(\theta). \quad (1)$$

Here $\ell(y_k|\theta, s_k)$ denotes the measurement likelihood functions, which depend on the sound

propagation model (forward model), the position of the transmitter s_k, and the statistical distribution of the measurement noise. Further, $\pi(\theta)$ is the prior distribution of the model parameters.

Although seemingly simple, the Bayesian recursion is generally impossible to solve analytically for all but linear models with Gaussian distributions. A range of numerical and approximate methods of different computational complexity that can be used to solve the Bayesian recursions exists, see the list of methods in [4]. Here, the unscented Kalman filter (UKF) is used (see [5] for details), in which the likelihood is modeled as

$$\ell(y_k|\theta, s_k) = \mathcal{N}(y_k; f(\theta; s_k), R), \tag{2}$$

where $\mathcal{N}(\mu, \Sigma)$ denotes a multivariate normal distribution with mean μ and covariance Σ. Further, $f(\theta; s_k)$ denotes the sound propagation (forward) model. Here, it has been explicitly highlighted that the forward model, and so also the measurement likelihood, depend on the location s_k of the transmitter. In the UKF, the posterior is, in every recursion, approximated as

$$\pi(\theta|y_{1:k}, s_{1:k}) \approx \mathcal{N}(y_k; \hat{\theta}_k, P_k). \tag{3}$$

where the mean and covariance are recursively updated as

$$\hat{\theta}_k = \hat{\theta}_{k-1} + P_{\theta y} P_{yy}^{-1} (y_k - \hat{y}_k) \tag{4a}$$

$$P_k = P_{k-1} - P_{\theta y} P_{yy}^{-1} P_{\theta y}^{\top}. \tag{4b}$$

Here the predicted measurement $\hat{y}_k$ and covariances $P_{\theta y}$ and P_{yy} are calculated as in Alg. 1.

Note that the covariance update in (4b) and the unscented transform in Alg. 1 do not depend on the measurements $y_{1:k}$ directly, but instead indirectly via $\hat{\theta}$ in the unscented transform. Therefore, given the current parameter estimate (mean) $\hat{\theta}_k$ and the parameter covariance P_k, these expressions can be used to calculate an approximation of P_{n+k}, the parameter covariance n steps ahead, given a sequence of of measurement positions $s_{k+1:k+n}$. The next measurement point is selected as

$$s_{k+1} = \underset{\bar{s}_{k+1} \in \Omega_{k+1}}{\arg\min}\ L(\bar{s}_{k+1}), \qquad L(\bar{s}_{k+1}) = \min_{\bar{s}_{k+1:k+n} \in \Omega_{k+1:k+n}} \det\left(\bar{P}_{n+k}(\bar{s}_{k+1}, \bar{s}_{k+2:k+n})\right), \tag{5}$$

where $\bar{P}_{n+k}(\bar{s}_{k+1:k+n})$ is the approximation of the parameter covariance calculated assuming the future measurement point sequence $\bar{s}_{k+1:k+n}$. Further, Ω_k is the set of feasible measurement points at time step k. To get a stable behavior in the selection of measurement points and allow the system to plan multiple steps ahead, $n \gg 1$ is typically used. A block diagram illustrating the sequential estimation and measurement process is shown in Fig. 1.

Algorithm 1 Unscented Transform

Require: $\hat{\theta}_{k-1}$, P_{k-1}, $f(\cdot; p_k)$, R

1: Generate sigma points Θ_i, weights $w_i^{(m)}, w_i^{(c)}$ from $\mathcal{N}(\hat{\theta}_{k-1}, P_{k-1})$
2: Compute $Y_i = f(\Theta_i; p_k)$
3: $\hat{y}_k = \sum_i w_i^{(m)} Y_i$
4: $P_{yy} = \sum_i w_i^{(c)} (Y_i - \hat{y}_k)(Y_i - \hat{y}_k)^{\top} + R$
5: $P_{\theta y} = \sum_i w_i^{(c)} (\Theta_i - \hat{\theta})(Y_i - \hat{y}_k)^{\top}$
6: **return** $\hat{y}_k$, $P_{\theta y}$, P_{yy}

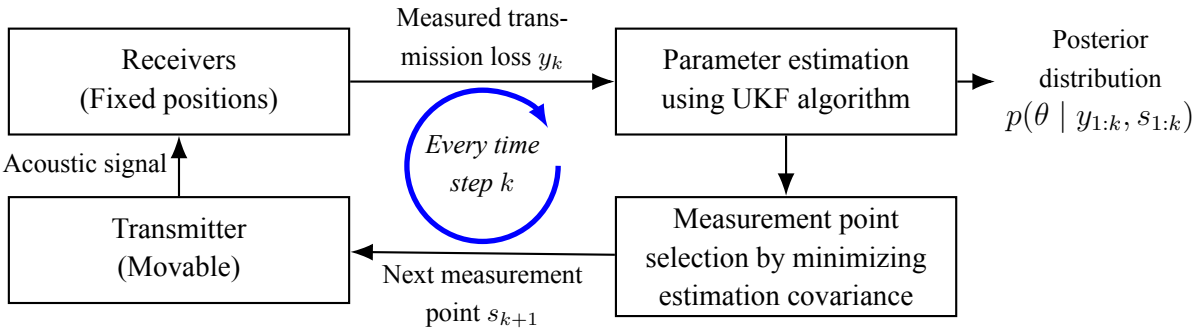


Figure 1: Proposed parameter estimation and measurement point selection framework.

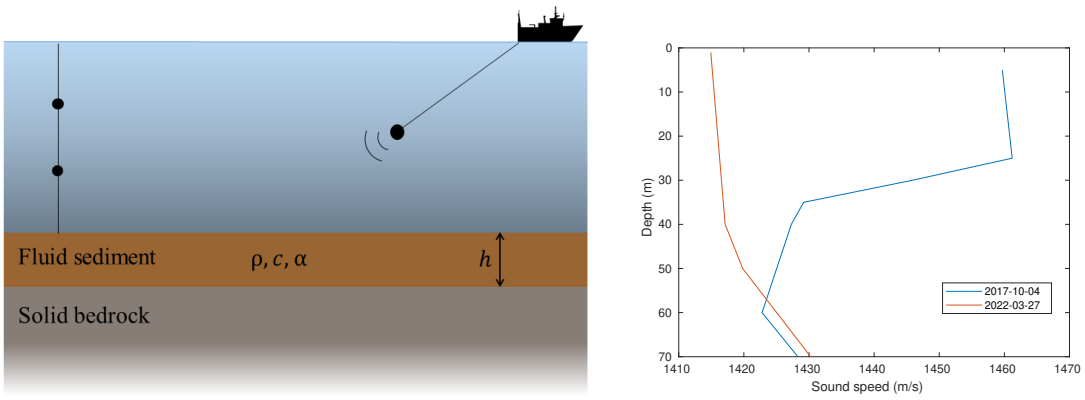
3. SIMULATION EXPERIMENT

A simulation study was conducted to evaluate the performance of the proposed parameter estimation and measurement point selection method. The simulation setup is illustrated in Fig. 2. A transmitter, towed by a boat, emits a signal at each time step k. A vertical array with two hydrophone elements receives the signal at a fixed location. The transmitter moves at a constant horizontal speed away from the receiver, and measurements are collected every 100 meters.

The sound propagation environment is modeled horizontally stratified, with a fluid sediment layer overlying a solid bedrock half-space. The water depth is set to 70 meters. The fluid sediment layer is characterized by the unknown parameters, the density ρ, the sound speed c_p, and the attenuation α_p. The two-way travel time $2h/c_\mathrm{p}$ in the sediment is assumed to be known. Further, the acoustic properties of the bedrock are assumed to have the known values $\rho = 2700$ kg/m^3, $c_\mathrm{p} = 5500$ m/s, $c_\mathrm{s} = 1800$ m/s, $\alpha_\mathrm{p} = 0.15$ dB/λ, and $\alpha_\mathrm{s} = 1$ dB/λ.

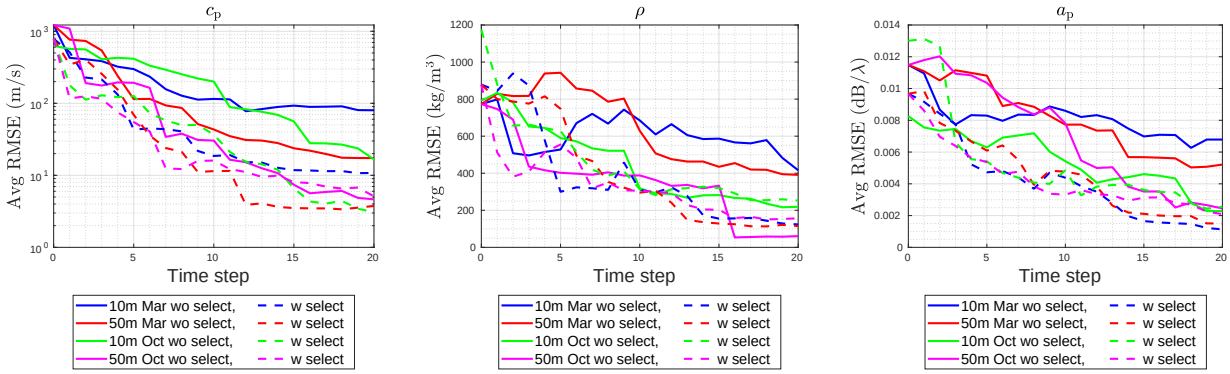
The simulations use two different sound speed profiles (SSP) recorded in the Baltic Sea, see Fig. 2. The profiles, labeled as 2017-10-04 and 2022-03-27, represent a typical downward refracting summer/fall condition and upward refracting winter/spring conditions, respectively. Transmission loss is simulated using the wave number integration program RPRESS [6], and a monochromatic transmitted signal with a frequency of 150 Hz is assumed. That is, RPRESS is used as the forward model $f(\theta; s_k)$.

The transmitter depth is initialized at either 10 or 50 meters, at a horizontal distance of 100 meters from the receiver. Two measurement point selection methods are considered. In the first,



(a) Experimental setup *(b) Sound speed profiles*

Figure 2: Experimental setup and SSP used in the simulation experiments.



(a) Sound speed *(b) Density* *(c) Attenuation*

Figure 3: RMSE of the estimated parameters as a function of time, wo indicates without measurement selection and w indicates with selection.

a preplanned fixed-depth measurement point selection scheme is considered. In the second, the depth of the measurement points is selected at each time step k using the proposed method. The transmitter is constrained to the interval 5 to 65 meters, and its depth can change by ± 5 meters between consecutive measurement points, i.e., $\Omega_k = \{s_{k-1} - 5, s_{k-1}, s_{k-1} + 5\} \cap (5, 65)$. Further, the planning horizon $n = 5$.

Given these conditions, eight scenarios are simulated: all combinations of the two sound speed profiles, two initial depths, and parameter estimation *with* and *without* measurement point selection. In each scenario, the true parameter values are set to $c_p^* = 1400$ m/s, $\rho^* = 1400$ kg/m^3, and $\alpha_p^* = 0.4$ dB/λ. Further, the prior was selected as

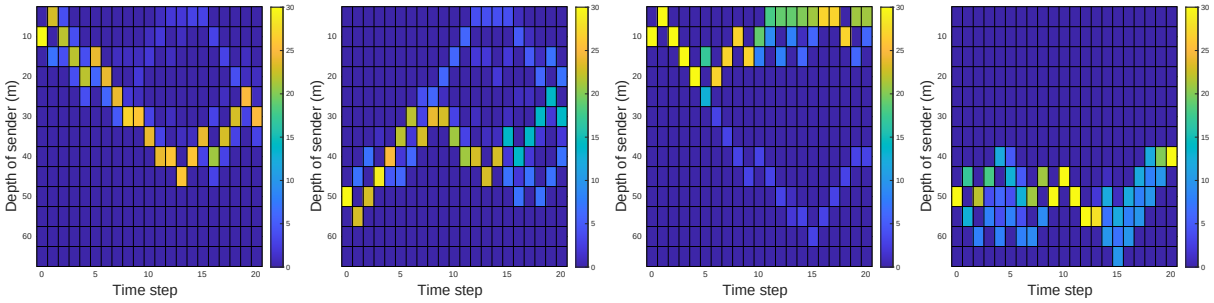
$$\pi(\theta) \sim \mathcal{N}\left(\begin{bmatrix}1400 & 1400 & 0.4\end{bmatrix}^\top, \text{diag}(30^2, 30^2, 0.1^2)\right).$$

Each scenario was evaluated using 30 Monte Carlo simulation runs. The performance was assessed using the root mean square error (RMSE) of the parameter estimates. Figure 3 shows the RMSE as a function of time for all scenarios. From the figure, it can be seen that the unknown parameters can be estimated well both with and without measurement point selection. However, the proposed sequential measurement point selection method renders significant improvements compared to a preplanned measurement point scheme. With the upward refracting March SSP (2022-03-27), improvements can be noted for both starting depths. With the October sound speed profile (2017-10-04), the estimate of c_p is better when the transmitter depth is allowed to vary and the starting depth of the transmitter is set to 10 meters. However, the other two parameters are estimated equally well with or without measurement point selection. When the starting depth of the transmitter is set to 50 meters, the parameter a_p is more accurately estimated using the sequential measurement point selection method.

To illustrate how the method selects transmitter depths over time, Fig. 4 shows the distribution of depths chosen as a function of time step. A comparison with the corresponding SSP in Fig. 2 indicates that depths associated with lower sound speeds are frequently selected.

4. DISCUSSION AND CONCLUSIONS

The results from the simulation study indicate that the proposed recursive Bayesian estimation framework, combined with sequential measurement point selection, improves the accuracy



(a) SSP 2022-03-27, Starting depth 10m. *(b) SSP 2022-03-27, Starting depth 50m.* *(c) SSP 2017-10-04, Starting depth 10m.* *(d) SSP 2017-10-04 Starting depth 50m.*

Figure 4: Distribution of the selected transmitter depth as a function of time.

of geoacoustic parameter estimation. The unknown sediment parameters could be accurately estimated in all simulated scenarios. However, the benefit of adapting the transmitter depth becomes clear when comparing the sequentially selected and preplanned (fixed-depth) measurement point selection schemes.

The benefit of the measurement point selection was most pronounced in the upward refracting SSP case, where the proposed method consistently reduced the estimation error across all parameters and starting depths. In the downward refracting SSP case, the effect depended more on the initial conditions. In addition, the distribution of the selected depths over time shows that the method tends to choose measurement points associated with a lower sound speed.

Overall, the results indicate that the proposed sequential selection of measurement points can significantly reduce estimation uncertainty in underwater acoustic inversion problems. Future research should investigate more complex inversion problems and consider uncertainties in the positions of the transmitter and receivers. Moreover, the case with a non-horizontally stratified sound propagation environment should be considered.

REFERENCES

[1] S. Kirkpatrick, Jr. C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[2] S. Katoch, S. S. Chauhan, and V. Kumar. A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, 80:8091–8126, 2021.

[3] G. Zheng, H. Zhu, X. Wang, S. Khan, N. Li, and Y. Xue. Bayesian inversion for geoacoustic parameters in shallow sea. *Sensors (Basel)*, 2020.

[4] D. Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience, 2006.

[5] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

[6] S. Ivansson. RPRESS—a family of computer programs for seismo-acoustic wave-fields in range-independent fluid–solid media: User's guide. Technical Report FOA-D-95-00169-2.2-SE, Department of Guidance & Control, Materials and Underwater Sensors, National Defence Research Establishment, 1995.

Range-Dependent Source Localization Methods in Shallow-Water environments

Jing Guo[1, 2], Juan Zeng[1], Li Ma[1]

[1] *Key Laboratory of Underwater Acoustic Environment, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China*

[2] *University of Chinese Academy of Sciences, Beijing 100049, China*

***Abstract* —** In shallow-water environments, the model mismatch due to the spatiotemporal variations in environmental parameters, particularly the sound speed profile (SSP) is a major challenge for matched-field processing (MFP) in passive source localization. In this paper, the sound speed variation in shallow-water environments affected by internal waves is analyzed, and two range-dependent equivalent SSP models, piece-wise linear equivalent model and periodic equivalent SSP model are proposed. These two models are incorporated into the MFP localization algorithm, and both simulation and experimental results, derived from sea trials conducted in the South China Sea, validate their effectiveness. Compared to traditional MFP methods, the algorithm based on the proposed range-dependent equivalent SSP models achieves smaller errors in source range estimation, demonstrating that a range-dependent equivalent SSP model is crucial for improving the source range estimation performance of MFP localization algorithms.

***Key Words* —** source localization, matched-field processing, range dependence, sound speed profile, shallow water

1 Introduction

Passive underwater source localization has been extensively researched for decades, leading to continuous advancements in related techniques. Among the various techniques, matched field processing (MFP) [1] remains a widely used and representative approach. While MFP performs well in stable ocean environments, its effectiveness often degrades in complex and variable ocean conditions. Unlike the large-scale and relatively range-independent deep-sea environment, shallow water is characterized by complex ocean dynamics, resulting in highly variable spatiotemporal sound speed profiles (SSPs). In shallow-water waveguides, range-dependent SSPs and bottom undulations can induce mode coupling among acoustic normal modes [2], which poses challenges for traditional passive localization methods. To address performance degradation caused by environmental model mismatch, numerous adaptive processors for MFP-based localization with improved tolerance to parameter mismatch have been proposed. For example, the Neighborhood Location Constraints method (MV-NLC) [3] enhances robustness to environmental mismatch by introducing constraints to preserve the mainlobe. The Optimal Uncertain Field Processor (OUFP) [4] estimates the source position by maximizing the posterior probability density function of the source location. The Environmental Perturbation Constraints method (MV-EPC) [5] constrains the processor using the statistical characteristics of replica signal covariance matrices computed over a range of perturbed environmental parameters. The Feature Extraction Method (FEM) [6], combined with the Multiple Signal Classification (MUSIC) algorithm, improves robustness by extracting features from the incoherent subspace of the acoustic field.

Despite the advancements, the aforementioned improved MFP algorithms face significant limitations in practical applications. Algorithm-level modifications to the processor do not directly resolve the underlying issue of model mismatch. To address the localization performance degradation caused by environmental parameter uncertainties, Collins and other researchers [7][8] proposed the focalization approach, which jointly estimates both the environmental parameters and the source location.

Among the various ocean environmental parameters, minor errors in seabed geoacoustic properties (such as sound speed, density, and attenuation) typically have limited impact on localization performance. In contrast, mismatch in the water-column SSP can lead to significant degradation in localization accuracy [9]. Therefore, in addition to source range and depth, focalization methods also incorporate the water-column SSP and other unknown or uncertain environmental parameters into the optimization process as variables. A commonly used approach represents the vertical structure of the SSP, as a weighted sum of a background profile and several modes derived from empirical orthogonal functions (EOFs) [10]. This formulation allows for efficient characterization of time-varying SSPs and has thus been adopted in focalization frameworks. Cristiano et al. [11] employed EOFs to construct a range-independent SSP model for joint estimation of the SSP parameters (i.e., EOF coefficients) and source location. For sources at short ranges ($\leqslant$2 km), the EOFs extracted from SSP measurements and the estimated EOF coefficients reflect the temporal variability of the SSP. This approach achieved improved localization performance compared to traditional methods that assume a known SSP. However, as range increases (e.g., 10 km) and source frequency rises, the range dependence of environmental SSP becomes increasingly significant. Consequently, MFP localization methods that consider only temporal variations of the SSP but ignore its range dependence suffer severe performance degradation [12][13].

Accurate passive source localization in shallow-water environments requires effective mitigation of environmental mismatch, particularly due to range-dependent variability in the environmental SSP. The main factor contributing to SSP fluctuations in shallow-water environments is internal wave activity. Internal waves with tidal-cycle characteristics introduce significant spatiotemporal variations in the SSP, which can severely degrade MFP performance. In this study, two range-dependent equivalent SSP models are developed for shallow-water regions of the South China Sea, where internal tides are prevalent. Localization on simulated and experimental data demonstrates the feasibility and effectiveness of range-dependent MFP localization algorithms. The results show that environment-adaptive equivalent SSP models can significantly

enhance localization performance in the presence of internal tidal variability.

2 Sound speed profile models

2.1 Environmental SSP in internal wave environment

In a linear internal wave environment, the motion of seawater particles is typically assumed to be an adiabatic process unaffected by external heat sources. In the two-dimensional (range-depth) domain, the displacement of water particles $\eta(r,z,t)$, can be represented as a superposition of discrete vertical modes [14]

$$\eta(r,z,t) \simeq \sum_{n=1}^{N}\sum_{m=1}^{M} a_{mn}\psi_m e^{i(k_{mn}r-\omega_n t)}\,, \tag{1}$$

where ω_n denotes the frequency of the internal wave, and ψ_m represents the vertical function of the mth mode, characterizing the depth-dependent distribution. Since the vertical modes vary weakly with frequency, they are typically assumed to be independent of ω_n. The horizontal wavenumber of the mth mode at frequency ω_n is denoted by k_{mn}, and a_{mn} is the corresponding modal amplitude, which is determined by the internal wave energy spectrum. The resulting temperature variation due to particle displacement can be approximated as [15]:

$$\frac{\partial T}{\partial t} = -\eta\frac{\partial T}{\partial z}. \tag{2}$$

Integrating Eq. (2) allows the water column temperature profile in a linear internal wave environment to be expressed as a superposition of multiple basis functions [15][16]:

$$T(r,z,t) = T_0(z) + \sum_{m=1}^{M}\sum_{n=1}^{N}\frac{a_{mn}}{i\omega_n}e^{i(k_{mn}-\omega_n t)}\,\varphi_T^m\,, \tag{3}$$

$$\varphi_T^m = \frac{dT_0}{dz}\psi_m\,, \tag{4}$$

where T_0 denotes the background temperature profile, and φ_T^m represents the mth basis function of temperature profile.

In shallow-water environments, the sound speed in the water column exhibits an approximately linear correlation with seawater temperature. Therefore, the SSP can similarly be represented as a combination of a background profile and a superposition of multiple basis functions, and is approximated as:

$$C(r,z,t) \simeq C_0(z) + \sum_{m=1}^{M}\varphi_C^m\sum_{n=1}^{N}A_{(m,\omega_n)}e^{i(k_{mn}r-\omega_n t)}\,, \tag{5}$$

$$\varphi_C^m = \frac{\partial C_0}{\partial z}\psi_m \,, \tag{6}$$

where, $C_0(z)$ denotes the background SSP, φ_C^m is the mth SSP basis function, and $A_{(m,\omega_n)}$ is the corresponding modal amplitude. The horizontal wavenumber k characterizes the spatial variation, while the internal wave frequency ω describes its temporal variability. For different internal wave modes, the nonlinear dispersion relationship between k and ω varies accordingly [14]. In most studies investigating internal wave, the range is often fixed at $r = 0$ to simplify the model—thus focusing solely on the temporal characteristics associated with internal wave frequency. In contrast, if considering the temporal variation of sound speed as a phase modulation superimposed along the spatial domain, then under ideal conditions, the range-dependent SSP in a shallow-water linear internal wave environment can be expressed as:

$$C(r,z) \simeq C_0(z) + \sum_{m=1}^{M} \varphi_C^m \sum_{n=1}^{N} A_{(m,\omega_n)} e^{i(k_{mn} r - \Theta_t^n)} \,, \tag{7}$$

$$\Theta_t^n = -\omega_n t \,. \tag{8}$$

When $t = 0$, the equivalent temporal phase Θ_t^n is zero. In solving inverse problems for focalization methods, an excessive number of model parameters can lead to an exponential increase in computational complexity during multi-parameter optimization. Moreover, model parameter coupling can further degrade the convergence rate of the algorithm. Therefore, constructing an equivalent SSP model with a limited number of parameters that still captures the essential range-dependent characteristics of the environment SSP is critical for the efficiency and effectiveness of range-dependent MFP localization algorithms.

2.2 Range-independent equivalent SSP model

In conventional source localization methods, range-independent equivalent SSPs are commonly constructed using EOFs as basis functions to reduce environmental model mismatch [10]. The range-independent equivalent SSP model can be expressed as:

$$C(z) \simeq C_0(z) + \sum_{m=1}^{M} \varphi_C^m A_m \,, \tag{9}$$

where φ_C^m and A_m denote the mth basis function and corresponding coefficient,

respectively. The background SSP C_0 can be estimated by averaging a set of historical SSP measurements. From the perspective of acoustic field effects, internal wave environments often exhibit strong range-dependent fluctuations in the SSP, which can induce severe coupling between acoustic normal modes [2]. Since range-independent equivalent SSP model cannot capture such features, its effectiveness in mitigating environmental mismatch is substantially limited.

2.3 Range-dependent equivalent SSP model

Conventional MFP algorithms based on range-independent equivalent SSP model primarily focus on capturing the vertical structure and temporal variability of the SSP. By incorporating range-dependent characteristics into these models, it becomes possible to approximate the range dependence of the SSP. As discussed in the analysis of sound speed variability in shallow-water environments, internal waves induce SSP variations with an implicit spatial periodicity. Within a limited range interval, these variations can be approximated as slowly varying linear trends [17]. Therefore, the range-dependent SSP in a linear internal wave environment can be approximated using a piecewise linear model consisting of L segments, and is expressed as:

$$\hat{C} \simeq \left[\widehat{\boldsymbol{C}}_1, \cdots, \widehat{\boldsymbol{C}}_l, \cdots, \widehat{\boldsymbol{C}}_L\right], \tag{10}$$

$$\widehat{\boldsymbol{C}}_l \simeq C_0 + \begin{bmatrix} e_{l,1} & \cdots & e_{l,m} & \cdots & e_{l,M} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_m \\ \vdots \\ \xi_M \end{bmatrix}. \tag{11}$$

Here, ξ_m denotes the mth order basis function, and $e_{l,m}$ represents its corresponding coefficient in the lth segment. The basis functions ξ_m can be chosen as EOFs extracted from historical SSP data, which contain environmental prior information. Assuming a linear equivalence condition—i.e., the sound speed variation within each segment can be approximated as a slow linear trend—the range-dependent equivalent SSP model described by Eq. (11) can be referred to as a *linear equivalent SSP model*. The linear equivalent SSP model has advantage over conventional range-independent model in describing the complex SSP in shallow water environments.

However, the choice of segmentation interval remains empirical and depends on environmental prior knowledge, due to the discretization introduced by the piecewise segmentation.

Some researchers have employed harmonic analysis methods to study isotherm fluctuations induced by internal tides in shallow water environments. Since isotherms and isospeed lines are closely related, harmonic analysis of isotherms is effectively equivalent to that of sound speed structures. Essentially, the harmonic analysis approach decomposes the SSP into a product of an equivalent mode $\varphi_c^{'}$, which integrates the characteristics of all basis functions, and a series of harmonic components of different orders:

$$C(t,z) \simeq C_0(z) + \varphi_C^{'} \sum_{n=1}^{N} A_n^{'} sin(\omega_n^{'} t + \Theta_n^{'}) . \quad (12)$$

Here, $A_n^{'}$, $\omega_n^{'}$, and $\Theta_n^{'}$ represent the amplitude, frequency, and phase of the nth harmonic component, respectively. Variations in these harmonic components correspond to fluctuations of isospeed contours. The harmonic frequencies ($\omega_n^{'}$) obtained through harmonic analysis are typically associated with the frequencies of internal tidal constituents, and they are related—but not identical—to the internal wave frequencies ω_n. The range dependence of the SSP arises from the combined variations of multiple internal tidal constituents, resulting in a compound spatiotemporal periodicity of the SSP. In certain shallow-water environments dominated by specific tidal constituents—for example, the diurnal tide in the South China Sea [18]—the range dependence of the environmental SSP is primarily governed by that dominant tidal constituent or by the combined influence of several major tidal constituents.

Assuming that the temporal and spatial variations of the sound speed can be approximated as linear changes [19][20], an equivalent tidal constituent can be constructed using only the first-order equivalent vertical function and a set of corresponding coefficients based on the basis function decomposition and harmonic analysis methods. By adding an equivalent compensation, the range-dependent equivalent SSP model with periodic variation characteristics can be expressed as:

$$C(r,z) \simeq C_0(z) + \varphi_C^{'}(A_0^{'} + A_1^{'} sin(\frac{2\pi}{T_r^{'}} r + \Theta^{'})) , \quad (13)$$

$$T_r^{'} = \frac{2\pi}{\omega^{'}}, \tag{14}$$

where, $\varphi_C^{'}$ denotes the equivalent basis function, and $A_0^{'}$, $A^{'}$, $\omega^{'}$, and $\Theta^{'}$ represent the compensated bias, amplitude, frequency, and phase of the equivalent tidal constituent, respectively. $T_r^{'}$ is the wavelength of the equivalent tidal constituent, corresponding to the horizontal periodicity of the SSP. The range-dependent equivalent SSP model based on the equivalent tidal constituent, as expressed in Equation (13), is hereafter referred to as the *periodic equivalent SSP model*. Compared to the linear equivalent SSP model, the periodic equivalent SSP model provides a more continuous and effective representation of the spatial periodicity of the environmental SSP, particularly in regions dominated by internal tidal waves. However, the modeling approach based on the equivalent tidal constituent also limits the applicability.

3 Range-dependent MFP algorithm

In this section, the impact of different equivalent SSP models on the performance of MFP-based source localization will be investigated through numerical simulations. Simulations are performed under shallow-water conditions representative of the South China Sea, following the experimental setup. The widely used *Bartlett* processor is employed to localize simulated acoustic sources,

$$F(\alpha, \beta, r, z) = \frac{1}{N_f} \sum_{j=1}^{N_f} \left| V_{r,z}{}^{H} \boldsymbol{X} \right|^2 . \tag{15}$$

Signal composed of multiple frequency points is emitted from a source located at a depth r and a range z. N_f denotes the number of frequency points to be calculated, $\boldsymbol{X}$ represents the observed acoustic field vector measured by the receiving array, and $V_{r,z}$ is the computed replica field. α represents parameters related to SSP and β represents other environmental parameters. The normalized two-dimensional ambiguity surface over distance and depth is defined as follows:

$$Surface(r, z) = 10\, log_{10}\left(\frac{F(\alpha_{opt}, \beta_{opt}, r, z)}{max\{F(\alpha_{opt}, \beta_{opt}, r, z)\}}\right) . \tag{16}$$

Acoustic fields are computed using the range-dependent acoustic model (RAM), and genetic algorithm (GA) is employed as the optimization method, with the acoustic field

correlation $F(\alpha, \beta, r, z)$ serving as the objective function. The GA parameters are set as follows: generation gap of 0.9, crossover probability of 0.9, and mutation probability of 0.05. Multi-parameter optimization is performed via multiple parallel searches, with a population size of 40 individuals evolving over 50 generations, resulting in a total of 2000 parameter vectors evaluated.

In the summer of 2020, the Institute of Acoustics, Chinese Academy of Sciences, conducted a fixed-point acoustic fluctuation experiment in the South China Sea. Figure 1(a) presents the measured 30-hour SSP measurements ensemble from the experiment. The internal wave activity during the experiment is dominated by the diurnal tide (24 h). A segment of the SSP measurements is transformed into a range-dependent SSP using an internal wave propagation speed of 1 m/s [18], representing the simulation environment of the South China Sea, as illustrated in Fig. 1(c). The first three EOFs extracted from the SSP measurements are shown in Fig. 1(b). The first EOF accounts for approximately 78% of the variance, thus serving as the basis function in both the linear equivalent SSP model and the periodic equivalent SSP model sufficiently characterizes the vertical structure of SSP. The range-independent equivalent SSP model uses all the first three EOFs.

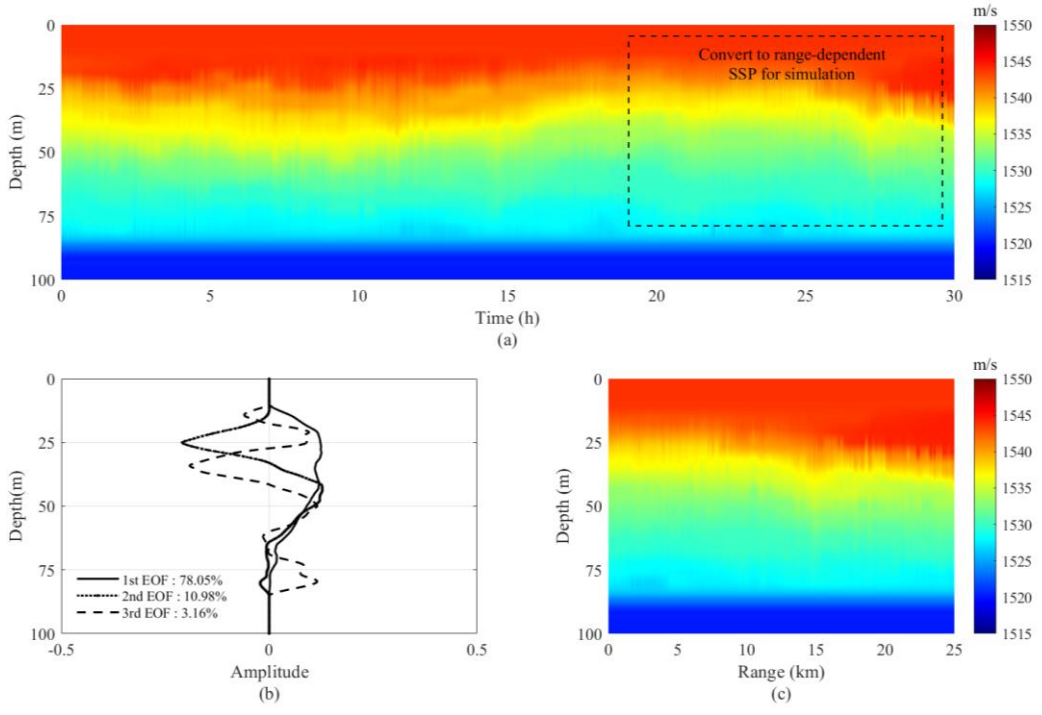


Fig. 1. Sound speed profile information from the South China Sea.

(a) 30-hour SSP measurements; (b) First three EOFs; (c) 25-km simulated SSP.

It is noteworthy that the dominant tidal constituent in the shallow waters of the South China Sea is the diurnal tide, which has a relatively long wavelength (approximately 77–86 km [18]). Since the 20-km SSP along the acoustic propagation path accounts for only about one-third of the diurnal tide wavelength, higher-frequency tidal constituents tend to dominate the sound speed variability over such shorter ranges. Therefore, the search range for the equivalent tidal wavelength (T') in the periodic equivalent SSP model should be adjusted accordingly. For example, in the South China Sea experimental environment, the semidiurnal tide—aside from the dominant diurnal tide—has the second-largest energy contribution, with a wavelength of approximately 40 km [18]. Accordingly, the wavelength T' of the equivalent tidal constituent is constrained to a search range of 20–60 km. In the case of the linear equivalent SSP model, a segment interval of $\Delta r = 4$ km is selected, resulting in six linear-equivalent profiles and a total of seven model parameters. Table 1 lists other relevant parameters, including seabed properties, vertical line array (VLA) configuration, and source information. The maximum acoustic propagation range is set to 24 km, with a range step size of 10 m. Table 2 summarizes the parameter search ranges and estimation results for each algorithm under the South China Sea simulation scenario.

Tab. 1. Simulation environment parameter settings.

Seabed			Vertical line array			Source		
sound speed	density	attenuation	element count	element spacing	depth coverage	range	depth	signal frequency
m/s	g/cm3	dB/λ	\	m	m	km	m	Hz
1650	1.75	0.2	16	1.5	32.5~55	20	55	320 340 360 380 400

Tab. 2. Parameter search ranges and localization results in the South China Sea simulation.

Source localization algorithms	Parameter search range of equivalent SSP model							Source location	
	A_1	A_2	A_3	A_0'	A_1'	T_r' (km)	Θ'	depth (m)	range (km)
	[-15,15]	[-10,10]	[-5,5]	[-10,10]	[5,15]	[20,60]	$(-\pi, \pi]$	[20,80]	[16,24]
Range-independent MFP	-10.11	-3.54	-4.93	\	\	\	\	41	16.4
Periodic MFP	\	\	\	-0.38	13.93	25.46	-0.12 π	48	19.4
\	e_1	e_2	e_3	e_4	e_5	e_6	e_7	depth (m)	range (km)
	[-25,25]	[-25,25]	[-25,25]	[-25,25]	[-25,25]	[-25,25]	[-25,25]	[20,80]	[16,24]
Linear MFP ($\Delta r = 4$ km)	8.43	12.72	-24.16	-18.97	-14.35	12.21	3.31	56	19.6
Linear MFP ($\Delta r = 6$ km)	8.08	-19.87	-19.21	14.17	24.89	\	\	55	19.6

Figures 2 and 3 show the localization ambiguity surfaces and reconstructed equivalent SSPs obtained by different algorithms. The actual source location and the estimated position are marked with symbols “O” and “X” on the ambiguity surfaces, respectively. The range-independent MFP algorithm [Fig. 2 (d)], which does not account for the range-dependence of SSP, fails to locate the source, resulting in large localization error. The two range-dependent equivalent SSP models introduce horizontal variability, which helps mitigate the mismatch between the assumed and environmental SSPs, thereby enhancing localization performance. Compared to the range-independent MFP algorithm, which yields a source range estimate of 16.4 km, the linear MFP algorithm ($\Delta r = 4$km) and the periodic MFP algorithm provide more accurate estimates of 19.6 km and 19.4 km, respectively, closer to the actual source range of 20 km. By incorporating range-dependent features in the SSP, the localization error is significantly reduced from 3600 m to 400 m. The equivalent SSPs reconstructed by the range-dependent algorithms capture main features of the environmental SSP variations and better account for mode coupling effects, resulting in improved localization accuracy over the range-independent MFP algorithm.

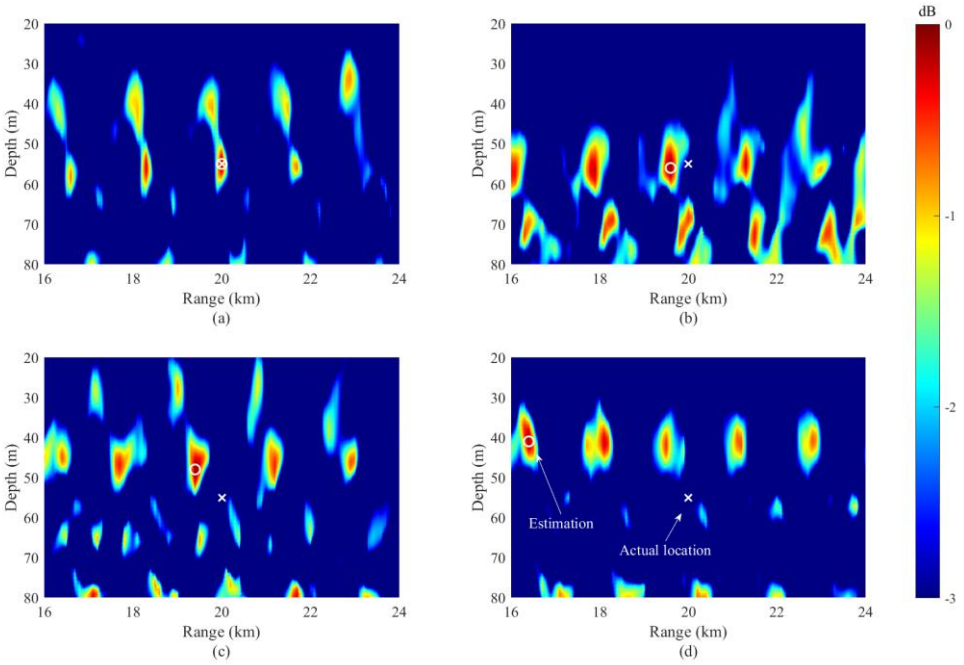


Fig. 2. Localization ambiguity surfaces for different algorithms in simulation.

(a) MFP without mismatch; (b) Linear MFP;

(c) Periodic MFP; (d) Range-independent MFP.

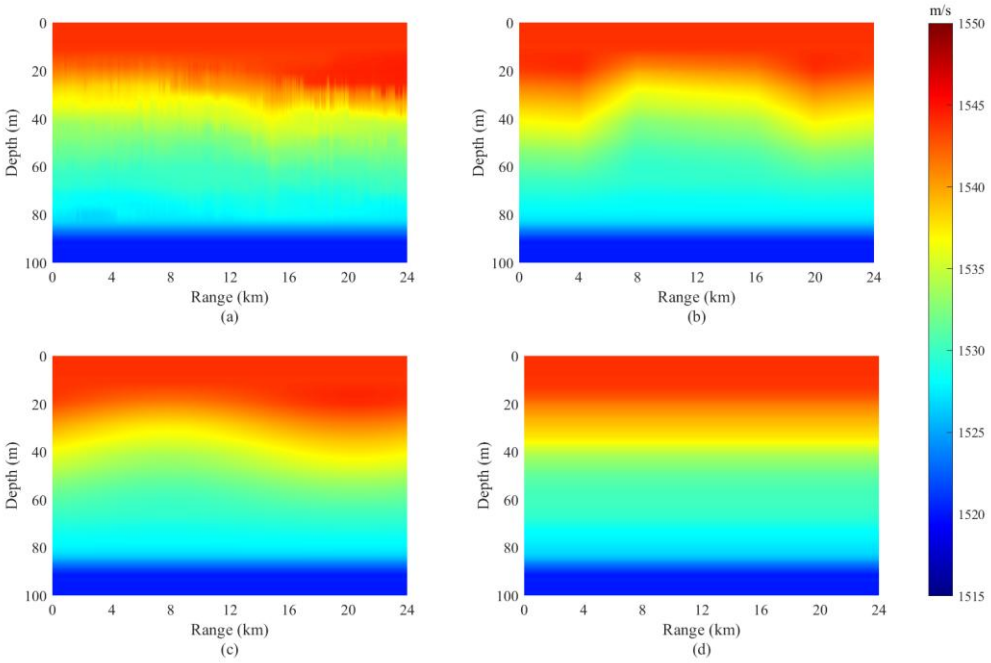


Fig. 3. Reconstructed equivalent sound speed profiles in simulation.

(a) SSP of the simulation environment; (b) Linear equivalent SSP;

(c) Periodic equivalent SSP; (d) Range-independent SSP.

The continuously varying periodic SSP model incorporates certain physical features (e.g., tidal periodicity), and its closer alignment with environmental characteristics enables improved localization performance [Fig. 3(c)]. In contrast, the linear equivalent SSP model produces discrete SSP segments [Fig. 3(b)], and strongly depends on the chosen segment interval. The reconstructed equivalent SSPs from the two range-dependent MFP localization algorithms reflect the general trend of environmental variability in simulation, but discrepancies remain when compared to the SSP of the simulation environment, especially for the linear equivalent SSP with a segment interval of $\Delta r = 4$ km. To further investigate the effect of segmentation choice, linear MFP algorithm is performed with a larger segment interval of $\Delta r = 6$ km (i.e., four segments), which better matches the trend of SSP variation in the simulation environment. Figure 4 presents the localization ambiguity surface and the reconstructed equivalent SSP for the linear MFP algorithm with $\Delta r = 6$ km. The corresponding parameter search ranges and localization results are listed in Table 2.

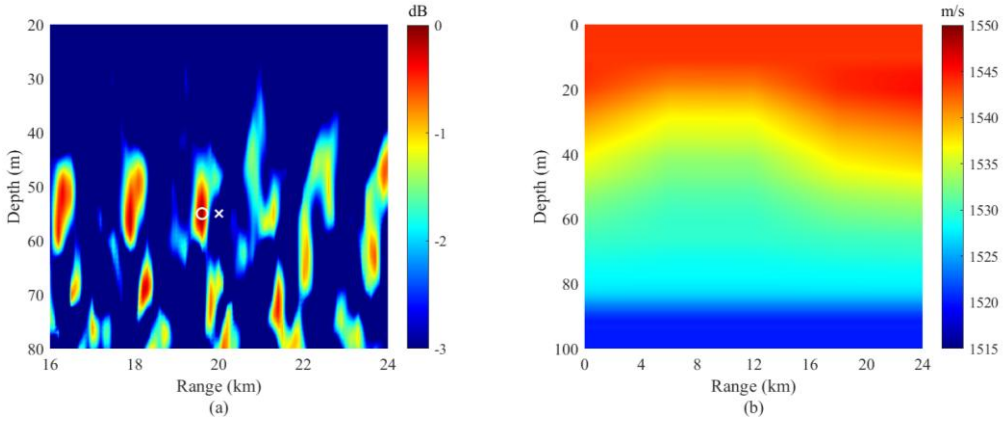


Fig. 4. Results of the linear MFP algorithm with segment interval $\Delta r = 6$ km.
(a) Localization ambiguity surface; (b) Reconstructed equivalent SSP.

While the four-segment linear MFP successfully localized the source, its ambiguity surface [Fig. 4(a)] exhibits more dispersed sidelobe patterns compared to the six-segment result with $\Delta r = 4$ km [Fig. 2(b)]. This suggests that denser segmentation can provide stronger fitting capability. However, due to the discrete nature of the segmentation approach, the equivalent SSP reconstructed with the more suitable interval ($\Delta r = 6$ km) offers a better match to the overall trend of the environmental

SSP [Fig. 4(b)]. Furthermore, figure 5 shows a comparison of the first-order EOF coefficient curves reconstructed by different algorithms, indicating that both the linear MFP ($\Delta r = 6$ km) and the periodic MFP provide estimates more consistent with the environmental EOF curve in simulation.

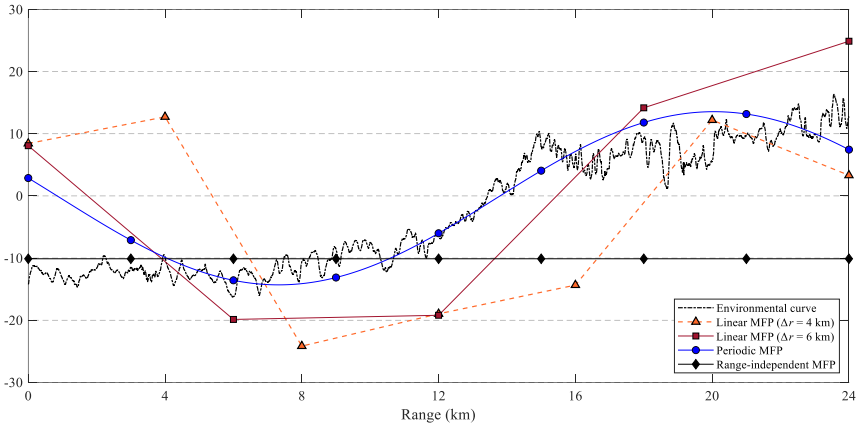


Fig. 5. First-order EOF coefficient curves reconstructed by different algorithms.

Although the range-dependent equivalent SSP models can characterize range dependence of the environmental SSP, the primary goal of the range-dependent MFP localization algorithm is not to invert the actual environmental SSP. Instead, the algorithm aims to optimize the model parameters to maximize the field matching function $F(\alpha, \beta, r, z)$. Under the given model assumptions, the algorithm searches within the model parameter space to find the set that maximizes the objective function. Whether the reconstructed equivalent SSP accurately represents the environmental SSP depends heavily on the accuracy of environmental modeling and the availability of sufficient prior information to solve the nonlinear multi-parameter optimization problem, and both of which are difficult to achieve in real-world environments. In summary, the localization results in the South China Sea simulation environments demonstrate that, in shallow-water environment with internal waves, the range-dependent MFP algorithms based on the linear and periodic equivalent SSP models achieve better localization performance than the commonly used range-independent MFP algorithm.

4 Experiment Validation

As a supplement to the previously described experiment in the South China Sea, the deployment of the experimental equipment is illustrated in Fig. 6. The acoustic source was a low-frequency transducer suspended 55 m below the vessel, transmitting linear frequency modulated (LFM) signals with a frequency range of 320–400 Hz, a duration of 10 seconds, and a repetition interval of 45 seconds. A vertical line array consisting of 16 hydrophones spaced 1.5 m apart was deployed approximately 20 km from the source to serve as the receiver. During the experiment, the hydrophone depths exhibited periodic fluctuations due to large-scale internal wave activity. The uppermost hydrophone was located at a depth of 32.5 m, and the bottom hydrophone at 55 m, with both remaining relatively stable at these depths during the stable period. Due to experimental constraints, the array was deployed in shallow water and spanned only about one-quarter of the total water depth. Source localization is conducted using three algorithms while the array remained in a stable configuration: a range-independent MFP algorithm and two range-dependent MFP algorithms based on the linear and periodic equivalent SSP models, respectively.

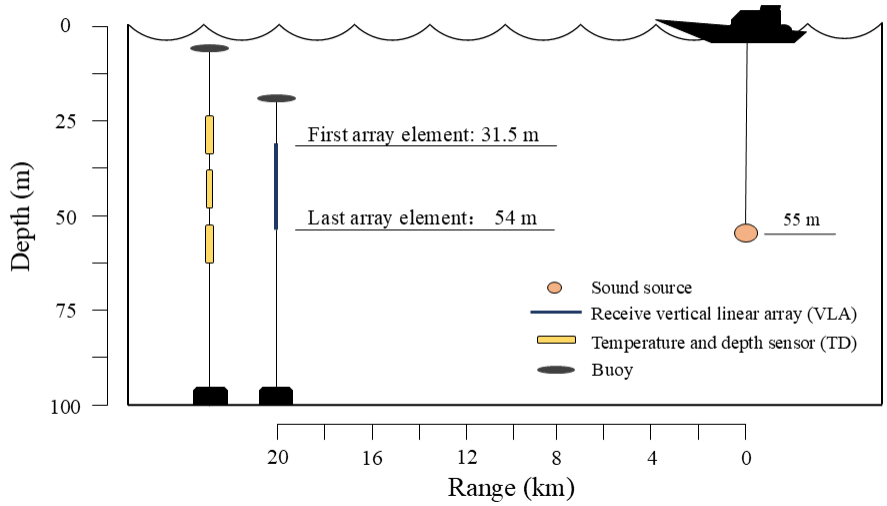


Fig. 6. The equipment deployment of the experiment in the South China Sea.

In the numerical simulation, seabed parameters are assumed to be known. However, in real environments, these parameters—such as seabed sound speed, density,

and attenuation—are typically unknown and should be jointly estimated along with the SSP, source depth, and source range. To solve this nonlinear multi-parameter optimization problem, genetic algorithm (GA) with multiple parallel search groups was employed. The GA parameters are set with a generation gap of 0.9, a crossover probability of 0.9, and a mutation probability of 0.05. A total of 10,000 parameter vectors are evaluated during the optimization process. The segmentation intervals for the linear equivalent SSP models are set to $\Delta r = 4$ km and $\Delta r = 6$ km, consistent with the numerical simulation and satisfying the conditions for linear approximation. The equivalent tidal wavelength $T_r^{'}$ for the periodic equivalent SSP model is selected based on the tidal constituents observed in the environment, consistent with the search range used in the numerical simulations. All the parameter search ranges and the localization results for a representative experimental signal are summarized in Table 3.

Tab. 3. Parameter search ranges and localization results in the South China Sea experiment.

			Range-independent MFP	Periodic MFP			Linear MFP ($\Delta r = 4$ km)	Linear MFP ($\Delta r = 6$ km)
Equivalent SSP Model parameters	A_1	[-15,15]	13.94	\	e_1	[-25,25]	8.69	6.93
	A_2	[-10,10]	-3.92	\	e_2	[-25,25]	17.49	14.24
	A_3	[-5,5]	3.09	\	e_3	[-25,25]	-11.97	-7.09
	A_0'	[-10,10]	\	7.74	e_4	[-25,25]	-15.19	-16.97
	A_1'	[5,15]	\	6.71	e_5	[-25,25]	19.88	10.02
	T_r' (km)	[20,60]	\	34.07	e_6	[-25,25]	14.58	\
	Θ''	$(-\pi, \pi]$	\	0.86 π	e_7	[-25,25]	1.27	\
Seabed parameters	sound speed (m/s)	[1550,1650]	1586	1559			1554	1597
	density (g/cm3)	[1,2]	1.74	1.48			1.39	1.58
	attenuation (dB/λ)	[0,1]	0.29	0.81			0.91	0.75
Source location	depth (m)	[20,80]	59	57			51	59
	range (km)	[16,24]	18.8	19.4			20.2	20.7

The equivalent SSPs reconstructed by the range-dependent and range-independent MFP algorithms, along with the estimated localization ambiguity surfaces, are shown in Fig. 7. The algorithmic estimates and the actual source position are marked by “O” and “X” on the ambiguity surfaces, respectively. The range-dependent MFP algorithms based on the linear and periodic equivalent SSP models account for the range dependence of the environmental SSP. Compared to the 1200 m range estimation error of the range-independent MFP, the range estimation errors are reduced to 200 m, 700 m, and 600 m for the linear MFP with $\Delta r = 4$ km, linear MFP with $\Delta r = 6$ km, and periodic MFP, respectively. The introduction of range-dependent equivalent SSP models significantly improved the range estimation performance of the MFP algorithms. Both the linear and periodic MFP algorithms effectively approximate the mode coupling effects induced by internal waves. Figure 8 presents the equivalent SSPs reconstructed by the four localization algorithms.

Notably, the equivalent SSPs reconstructed by the linear MFP with $\Delta r = 6$ km and the periodic MFP exhibit similar trends along the range [Fig. 8(b) (c)], and their range estimation errors are comparable. However, during the optimization process—where the objective function is the field matching function $F(\alpha, \beta, r, z)$—the linear MFP with $\Delta r = 4$ km benefits from a greater number of model parameters, resulting in a smaller range estimation error of 200 m. Overall, the processing of experimental data confirms the necessity of considering the range dependence of environmental SSP. Establishing an appropriate range-dependent equivalent SSP model can enhance the localization performance of MFP algorithms in shallow-water internal wave environments.

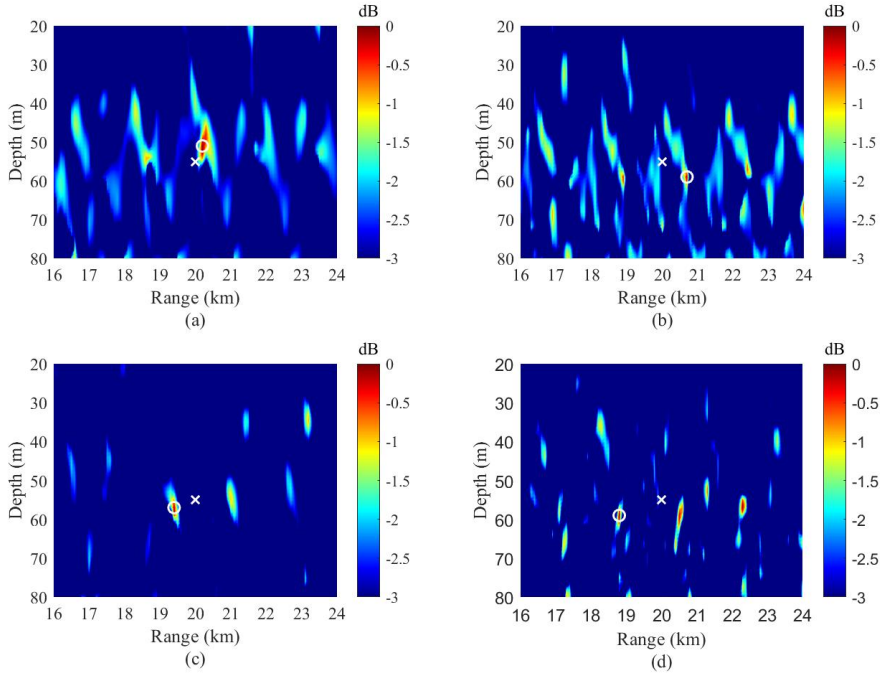


Fig. 7. Localization ambiguity surfaces for different algorithms in experiment. (a) Linear MFP ($\Delta r = 4$ km); (b) Linear MFP ($\Delta r = 6$ km); (c) Periodic MFP; (d) Range-independent MFP.

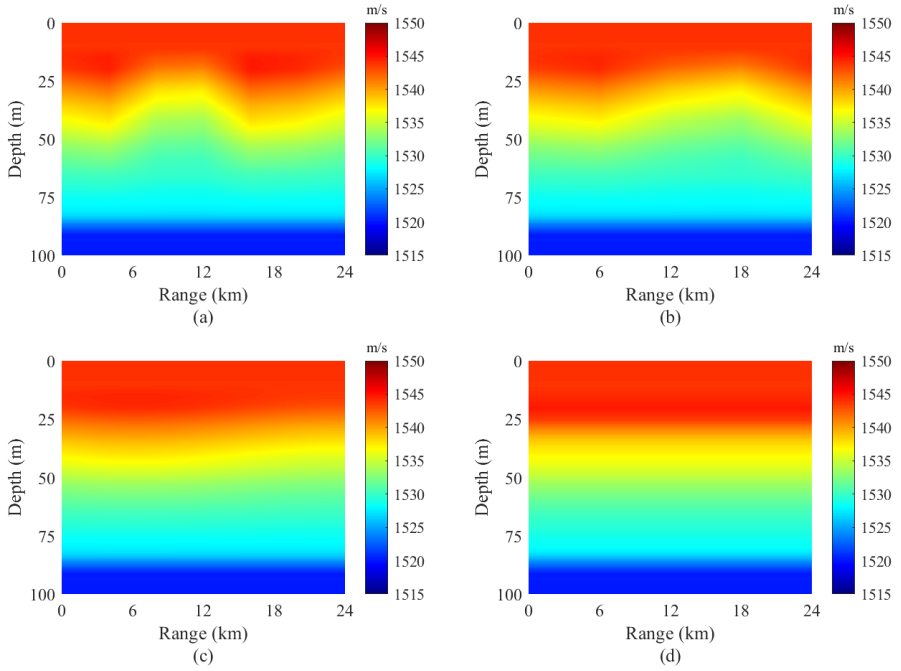


Fig. 8. Reconstructed equivalent sound speed profiles in experiment. (a) Linear equivalent SSP ($\Delta r = 4$ km); (b) Linear equivalent SSP ($\Delta r = 6$ km); (c) Periodic equivalent SSP; (d) Range-independent SSP.

5 Conclusion

The temporal variations of the SSPs are represented using basis functions such as EOF, where vertical fluctuations are captured through the estimation of corresponding coefficients. To account for the range dependence of the environmental SSP associated with internal waves, two range-dependent equivalent SSP models are investigated: (1) a piecewise linear equivalent SSP model that approximates variations by using multiple segments over distance, and (2) a continuous periodic equivalent SSP model that represents range dependence based on the dominant tidal constituents of the environment. The segmentation interval in the linear equivalent SSP model and the wavelength search range in the periodic equivalent SSP model are determined by environmental prior knowledge. Both range-dependent MFP algorithms—based on the linear and periodic equivalent SSP models—outperform the conventional range-independent MFP method, which neglects spatial variability in the environmental SSP. Numerical simulations and experimental results from the South China Sea confirm the feasibility and effectiveness of incorporating range-dependent equivalent SSP models to improve MFP-based source localization in shallow-water internal wave environments.

References

[1] H. P. Bucker, “Use of calculated sound fields and matched-field detection to locate sound sources in shallow water,” *J. Acoust. Soc. Am.*, vol. 59, no. 2, pp. 368–373, 1976, doi: 10.1121/1.380872.

[2] Y.F. Li, R. M. Guo, and H. F. Zhao, ”Sparse reconstruction of acoustic interference fringes in shallow water and internal wave environment,” *Acta Physica Sinica*, vol. 72, no. 7, pp. 074301, 2023, doi: 10.7498/aps.72.20221932

[3] Schmidr H, Baggeroer A B, Kuperman W A, et al. Environmentally tolerant beamforming for high-resolution matched field processing: deterministic

mismatch[J]. Journal of the Acoustical Society of America, 1990, 88(4): 1851–1862.

[4] Richardson A M, Nolte L W. A posteriori probability source localization in an uncertain sound speed, deep ocean environment[J]. Journal of the Acoustical Society of America, 1991, 89(5): 2280–2284.

[5] Shorey J A, Nolte L W, Krolik J L. Computationally efficient Monte Carlo estimation algorithms for matched field processing in uncertain ocean environments[J]. Journal of Computational Acoustics, 1994, 2(3): 285–314.

[6] Seong W, Byun S H. Robust matched field-processing algorithm based on feature extraction[J]. IEEE Journal of Oceanic Engineering, 2002, 27(3): 642–652.

[7] Michael D. Collins, W. A. Kuperman. Focalization: Environmental focusing and source localization[J]. *J. Acoust. Soc. Am.*, 1991; 90: 1410-1422.

[8] Ralph N. Baer, Michael D. Collins. Source localization in the presence of internal waves[J]. *J. Acoust. Soc. Am.*, 2005; 118: 3117-3121.

[9] Rachel M. Hamson, Richard M. Heitmeyer. Environmental and system effects on source localization in shallow water by the matched-field processing of a vertical array[J]. *J. Acoust. Soc. Am.*, 1989; 86: 1950-1959

[10] Lester R. LeBlanc, Forster H. Middleton. An underwater acoustic sound velocity data model[J]. *J. Acoust. Soc. Am.*, 1980; 67: 2055-2062.

[11] Cristiano Soares, Martin Siderius, and Sergio M. Jesus. Source localization in a time-varying ocean waveguide[J]. *J. Acoust. Soc. Am.*, 2002; 112: 1879-1889.

[12] Peter Gerstoft, Donald F. Gingras. Parameter estimation using multifrequency range-dependent acoustic data in shallow water[J]. *J. Acoust. Soc. Am.*, 1996; 99: 2839-2850.

[13] A.G. Sazontov, I.P. Smirnov. Source Localization in an Acoustic Waveguide with Inaccurately Known Parameters Using Matched Processing in the Mode Space[J]. *Acoust. Phys*., 2019; 65: 450-459.

[14] T. C. Yang and K. Yoo, "Internal wave spectrum in shallow water: measurement and comparison with the Garrett-Munk model," *IEEE Journal of Oceanic Engineering*, vol. 24, no. 3, pp. 333-345, 1999, doi: 10.1109/48.775295.

[15] W. Song, T. Hu, S. M. Guo, L. Ma, and L. L. Lu, "A methodology to achieve the basis for the expansion of the sound speed profile," *Acta Acustica*, vol. 39, no. 1, pp. 11–18, 2014, doi: 10.15949/j.cnki.0371-0025.2014.01.001

[16] K. Qu, S. C. Piao, and F. Q. Zhu, "A noval method of constructing shallow water sound speed profile based on dynamic characteristic of internal tides," *Acta Phys. Sin.*, vol. 68, no. 12, pp. 124302, 2019, doi: 10.7498/aps.68.20181867.

[17] Harry A and DeFerrari, "Effects of horizontally varying internal wavefields on multipath interference for propagation through the deep sound channel," *J. Acoust. Soc. Am.*, vol. 56, no. 1, pp. 40–46, 1974.

[18] Zhen Wang, Tao Hu, W. B. Wang, S. M. Guo, and Li Ma, "Characteristics of coastal internal waves and acoustic-energy fluctuations in the South China Sea in summer," *ACTA ACUSTICA*, vol. 49, no. 1, pp. 67-77, 2024, doi: 10.12395/0371-0025.2023147.

[19] G. H. Ji, and Z. L. Li, "The coherence-time of matched-field processing with the presence of the linear internal waves in the South China Sea," *Acta Acustica*, vol. 36, no.1, pp. 1-7, 2011, doi: 10.15949/j.cnki.0371-0025.2011.01.001.

[20] P. Qian, W. M. Gan, H. Q. Niu, G. H. Ji, Z. L. Li, and G. J. Li, "A feature-compressed multi-task learning U-Net for shallow-water source localization in the

presence of internal waves," *Applied Acoustics*, vol. 211, no.4, pp. 109530, 2023, doi: 10.1016/j.apacoust.2023.109530.

Structured Session
08. Bioacoustics and Soundscape

Organised by Jennifer Miksis-Olds and Giacomo Giorli

Automated Classification of Vocalisations from Wild and Captive Seal Populations

William Butler[1], Harrison Smith[2], Wil Lloyd[1], Marios Impraimakis[1], Andrew Barnes[1], and Alan Hunter[1]

[1]University of Bath, Bath, BA2 7AY, United Kingdom
[2]Celtic Sea Power, Hayle, TR27 4DD, United Kingdom

Contact Author: William Butler, wmb34@bath.ac.uk
Keywords: *seal vocalisation, bioacoustics, deep learning.*

1. INTRODUCTION

Grey seals (*Halichoerus grypus*) and Harbour seals (*Phoca vitulina*) are key marine predators distributed across the North Atlantic, with significant populations in the U.K. where they are protected under several conservation frameworks ([1], [2], [3]). As the U.K. rapidly expands offshore infrastructure, particularly floating offshore wind farms, understanding how these developments interact with marine species such as grey seals is increasingly critical ([4], [5]). Passive acoustic monitoring (PAM) provides a non-invasive method to assess seal presence, behaviour, and potential disturbance responses, especially in offshore regions beyond visual monitoring range. Grey seal vocalisations are known to vary seasonally and contextually ([6], [7]), but classification has been limited by observational constraints and variation in wild populations [8]. While recent work has begun to map acoustic repertoires using spectrogram features [9], the behavioural significance of individual call types remains poorly understood.

Some notable work has been done on the automated detection and classification of marine mammal vocalizations. PAMGuard's Whistle and Moan Detector [10] is a widely used tool for the detection of odontocete vocalisations through noise reduction and thresholding techniques. More recently, several works have successfully utilised neural network approaches to classify marine mammals ([11], [12]), [13]), though focus has generally been on whales and other types of odontocetes rather than seals.

Here, we introduce a novel contribution to grey and harbour seal bioacoustics: we present and empirically validate two previously undocumented call types, termed SP4 and SP5 through expert-tagged recordings, and establish their consistency across multiple vocalising individuals. We then use these labels to train a spectrogram-based deep learning model, contributing a scalable pipeline for future automated classification. This work supports acoustic ecology efforts in the southwest UK, where dialectal variation may exist [14], and informs conservation planning amid expanding offshore energy development.

Class	Count	Description
S1	741	Multiple Harmonic Tonals
S2	520	Non-tonal Bursts
S3	363	Harmonic Pulses
SP4	134	Dyadic Pulse Train
SP5	170	“Zipper” Pattern Pulse Train

Table 1: Seal call class densities and descriptions

2. METHODOLOGY

The data used in this paper was provided by Celtic Sea Power (CSP). The data is comprised of 2-hour, uninterrupted recordings of captive and wild seals, taken at the Cornish Seal Sanctuary and in-situ at a location near to Godrevy island, U.K. using the HydroMoth [15] and AudioMoth [16] hydrophones with a sample rate of 96kHz. Near Godrevy, the audio was taken from a “Camera Frame” and a “Lobster Pot” situated around 30 meters apart and contains vocalisations of Grey Seals. Audio from the Cornish Seal Sanctuary was taken from the Common Seal Pool and Grey Seals Pool.

2.1. DATA LABELLING AND CLASS DESCRIPTIONS

Seal vocalizations were manually tagged by CSP employees with a start time and end time, checked by the marine biologist, and classified into one of five types of seal calls. The call types were labelled S1, S2, S3, SP4, and SP5, expanding upon Nowak’s classification of S1, S2, and S3 [8]. The class density of the calls is described in Table 1. with a total of 1928 seal calls tagged and classified. Calls that were not classified into one of these five categories were removed and are not included in the subsequent figures in this work, nor used in training any of the models described later.

The S1 and S2 classes were both categorised as described in Nowak’s paper. S1 calls are described as tonal, with multiple harmonic elements. S2 calls are repeated short bursts of vocalisation, lacking any distinct tonal elements. It is theorised that these may be mating calls [8].

A subset of tagged signals was provisionally classified as S3-type vocalisations, following the definitions established by Nowak who describes S3 calls as short, pulsed sounds with distinct harmonic structure and no observable air exhalation during production. In the Nowak dataset, S3 vocalisations predominantly occur in repetitive sequences (dyads or pulse trains) with inter-pulse intervals typically shorter than 1 second, often around 0.5 seconds. In contrast, the majority of sounds classified at Godrevy as S3 were singular pulses, occasionally occurring in pairs or triplets, or as isolated sounds with longer inter-call intervals (e.g., 2 seconds). Although a few examples displayed characteristics consistent with Nowak’s descriptions such as sequences of 4–5 evenly spaced pulses with clear structure, the overall pattern in our dataset is less rhythmically structured and lacks consistent evidence of paired airflow dynamics.

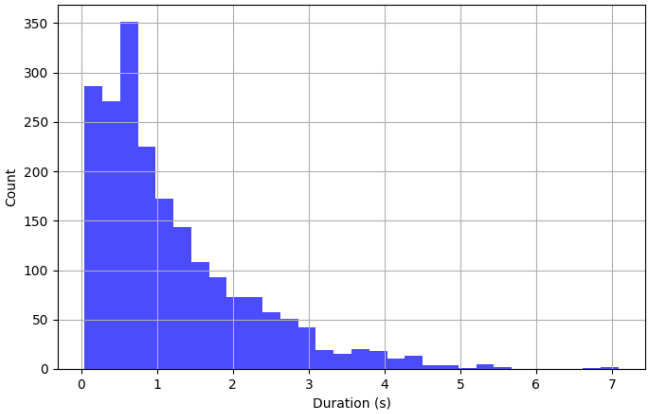


Figure 1: Histogram of call durations.

2.1.1. SP4 - RHYTHMIC DYADIC PATTERN

During the labelling process, a distinct acoustic signal—likely produced by a grey seal—was classified as "SP4", where "P" denotes *Provisional* categorisation. These calls typically last 1.5–2 seconds (ranging from 1–4+ seconds) and comprise 10–30 discrete, rhythmic elements. Each element features a steady tone with spectral peaks around 1 kHz, followed by a rapid downsweep to 250 Hz. The evenly spaced elements create a drumming or clapping-like rhythm, reminiscent of a seal clapping/striking its body as observed out of water. Closer analysis reveals a dyadic structure: elements occur in pairs with slightly shorter gaps within pairs than between them, producing a syncopated rhythm that occasionally shifts to even spacing mid-call.

2.1.2. SP5 - ZIPPER PATTERN PULSE TRAIN

Another distinct signal, labelled "SP5" was identified at Godrevy. Lasting ∼0.5 seconds (occasionally up to 1s), SP5 comprises 5–10 evenly spaced spectral peaks (0.5–1 kHz), and acoustically resembles a zipper or dragging a finger along metal coils, suggesting a tonal, percussive quality. SP5-type vocalisations are hypothesised to originate from a phocids (seals) as, based on the high intensity and propagation, they are unlikely caused by invertebrates or small fish which typically produce weaker, more broadband sounds. Known UK vocal fish (e.g. *Chelidonichthys* spp., *Sciaenidae*) do not match SP5's spectral or rhythmic features. Additionally, SP5 sometimes co-occurs or transitions into low-frequency "grunt" or "sniff" sounds typical of seals. These characteristics support a seal origin, though concurrent visual confirmation is recommended for species attribution.

2.2. DATA PRE-PROCESSING

Fig. 1 presents a histogram of call durations across all sites, showing that most last less than 3–4 seconds, with few exceeding 5 seconds and none over 10. Therefore, a 10-second time window was chosen for spectrogram generation.

For tagged seal calls, a random offset between 0 and (10 – call duration) seconds was applied before the labelled start time, and a 10-second audio clip was extracted. To estimate the noise floor at each site, 2–10 second gaps between tagged calls were used. Linear spectrograms from these segments were stacked horizontally and the median was computed for each frequency.

Spectrograms were generated using Python's Librosa library [17] with a Short-Time Fourier Transform (STFT) using an FFT window size of 2048, a Hann window, and a hop size of 1024, at a 96 kHz sample rate. Absolute FFT magnitudes were used, and the first and last frames were trimmed to avoid artifacts. This process yielded square spectrograms of shape (1025, 936), about 4× larger than ResNet's input (224×224), which will be introduced and used later in this paper. It is assumed there is acceptable information loss upon 4× downsampling. Finally, linear spectrograms were divided by the site-specific, frequency dependent noise floors and converted to dB scale (ref=1.0)

Several transformations were defined as part of the dataloaders for the fine tuning, including scaling, normalisation and resizing. ResNet expects a 3 channel input with values normalised to a mean of 0.485, 0.456 and 0.406 and standard deviation of 0.229, 0.224 and 0.225 for each RGB channel, respectively. The matrices generated from the STFT are 1 channel, so we take the mean of the ImageNet means (0.449) and standard deviations (0.226), as the matrices are duplicated across 3 channels to match the input dimensions expected by ResNet.

The standard deviation of the dB values, after noise reduction, is estimated by randomly sampling 20 seal and background noise spectrograms from each recording session at each location. The average standard deviation across all locations was found to be 6.6dB (2 s.f.). We assume a mean of 0 after the background noise subtraction in the dataset here and calculate appropriate minimum and maximum dB values of -13.11dB and 16.09dB for the following Min-Max scaling, to ensure the spectrograms approximate the mean and standard deviation of the data ResNet was initially trained on:

$$y = \text{clip}\left(\frac{x - \text{dB}_{\text{min}}}{\text{dB}_{\text{max}} - \text{dB}_{\text{min}}},\ 0,\ 1\right) \tag{1}$$

Finally, the data are normalised using the mean of 0.449 and standard deviation of 0.226 across all 3 channels and reshaped using linear interpolation to 224x224. Fig. 2 shows the original background subtracted spectrograms, the transformed inputs to the model and histograms of the pixel intensities after the transformations.

2.3. RESNET MODEL AND FINE-TUNING

A pretrained ResNet18 model from PyTorch [18], originally trained on ImageNet [19], was selected due to its broad success in computer vision tasks across domains [20]. The model was fine-tuned for a 5-class classification task corresponding to seal call classes S1, S2, S3, SP4, and SP5. The final fully connected layers of ResNet were replaced with the following sequence:

- Dropout (rate = 0.5)
- Fully Connected Layer (512 neurons)
- ReLU activation
- Dropout (rate = 0.5)
- Fully Connected Layer (5 neurons for class scores)

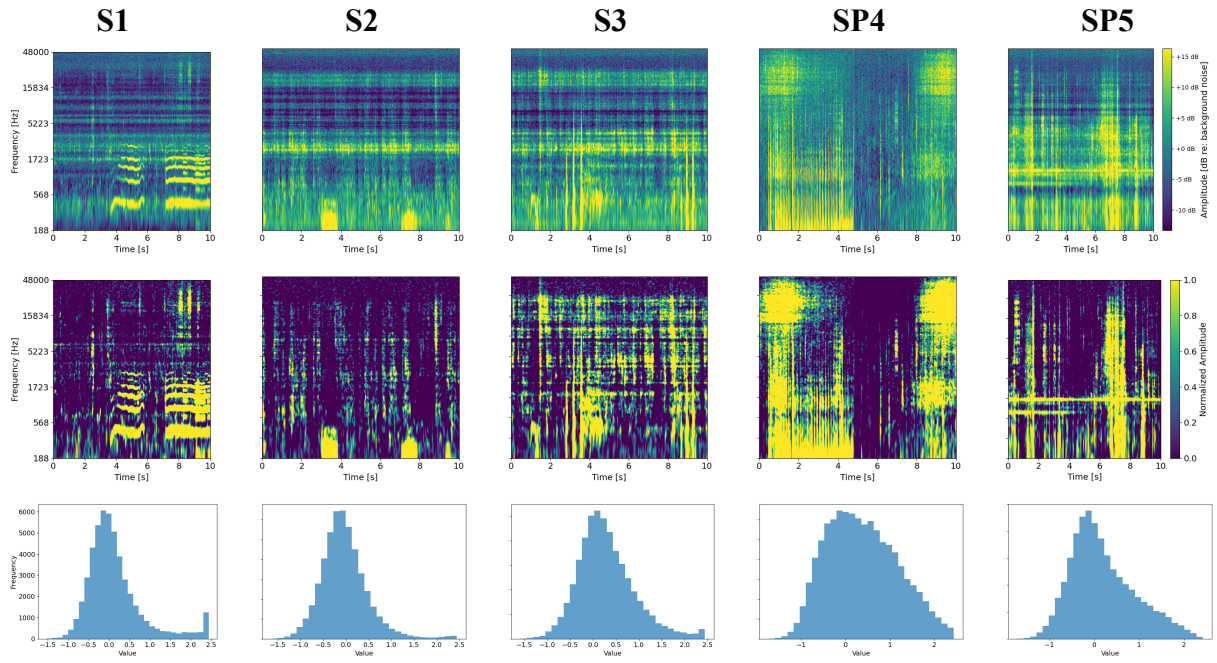


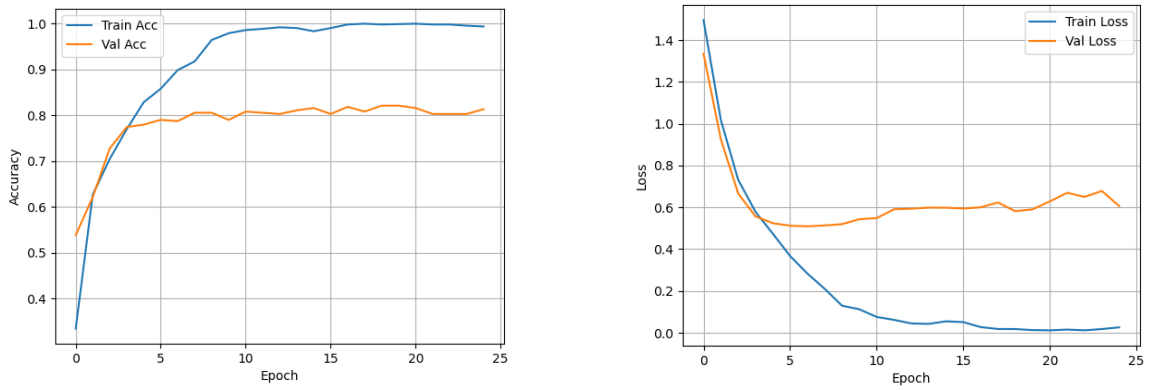
Figure 2: Examples of all five call types. Top row: Spectrograms after background subtraction. Middle row: Spectrograms after transformations. Bottom row: Histogram of intensities used as input to ResNet. Spectrograms are resampled to logarithmic frequency bins for visualization.

No output activation function was used, as PyTorch's CrossEntropyLoss implicitly applies softmax

The ResNet model was fine-tuned over 25 epochs. The data were split into training, evaluation and testing datasets with a ratio of 60:20:20, stratified according to the class densities. The CrossEntropyLoss function, weighted according to the class densities, was the criterion. Adam optimiser [21] was used with a learning rate of 1e-4. The model with the lowest validation loss, computed at each epoch, was selected. The initial layers were frozen to mitigate against overfitting, leaving only the last convolutional (Pytorch's "Layer 4") and fully connected layers trainable.

3. RESULTS

Fig. 3 present the model performance in terms of accuracy and loss over each epoch, and shows the model converging quickly before overfitting, despite freezing initial layers and high dropout rates (0.5) in the final layers. The final accuracy was 0.77. Fig. 4 presents the confusion matrix for the final model and reveals class S3 likely shares many similar features to class S1 and S2, evidenced by 42% of true S3 calls being misclassified as either S2 or S1. In contrast, only 4% of S3 calls are misclassified as SP4 and none classified as SP5. SP4 and SP5 are much more distinct with only 2 instances of SP4 misclassified (one each as S1 and S3) while all instances of SP5 were correctly identified. Fig 4. also highlights that S1 and S2 recall and precision are negatively impacted by the S3 class. It is clear from both Fig. 4 and Table 2 which show the model performance per-class that this model finds the S3 class significantly harder to model. The proposed SP4 and SP5 classes had the highest F-1 scores across all classes.



(a) Training and validation accuracy *(b) Training and validation loss*

Figure 3: Training curves for ResNet model.

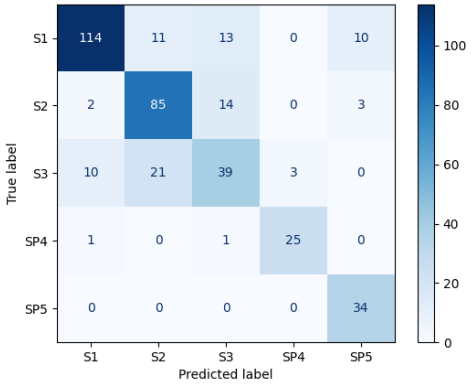


Figure 4: Test data confusion matrix

Class	Precision	Recall	F1-score	Support
S1	**0.90**	0.77	0.83	148
S2	0.73	0.82	0.77	104
S3	0.58	0.53	0.56	73
SP4	0.89	0.93	**0.91**	27
SP5	0.72	**1.00**	0.84	34
Accuracy		*0.77*		386
Macro avg	0.76	0.81	0.78	386
Weighted avg	0.78	0.77	0.77	386

Table 2: Classification report for fine-tuned ResNet

4. DISCUSSION AND CONCLUSION

As noted by Tadeo et al. [9], the S3 vocalisation category exhibits a high degree of variability, often cross-correlating with different vocalisation classes identified in previous studies. This variability complicates consistent classification and may reflect either ambiguity in our labelling or intrinsic heterogeneity within the S3 call type itself. It is therefore plausible that S3 represents a weaker clustering of vocal behaviour given it has the lowest F1-score of just 0.56 (next lowest is 0.77). Further work is needed to reassess the categorisation of these signals, particularly as we remain confident that the singular noise events originate from seals, despite their divergence from more stereotyped S3 structures. Given this variability, we acknowledge a lower level of confidence in the classification of S3 vocalisations during the labelling process. While some calls clearly match the acoustic profile of S3, others may represent contextually or functionally distinct call types. Additionally, observation conditions may have contributed to the ambiguity.

Despite these limitations, it is evident that the ResNet model performed significantly worse at classifying S3 vocalisations having both the lowest recall (0.58, next lowest 0.72) and precision (0.53, next lowest 0.77), compared to S1 and S2 as well as the proposed SP4 and SP5 classes. This supports the hypothesis that the S3 class represents a heterogeneous or poorly defined category, potentially encompassing multiple vocalisation types. In contrast, the higher classification performance observed for SP4 and SP5 suggests that these newly proposed categories

correspond to acoustically and structurally distinct vocalisation types that form well-separated clusters in the feature space. These findings underscore the need for further investigation into the acoustic properties and behavioural context of SP5 vocalisations. Additionally, future work should consider integrating supplementary classifications from existing literature to refine or subdivide the ambiguous S3 category and enhance the overall taxonomy of seal vocal behaviour.

5. ACKNOWLEDGEMENTS

The authors would like to thank Seiche Water Technology Group and the ART-AI CDT for funding and supporting William Butler's PhD.

REFERENCES

[1] D. Bowen, "Halichoerus grypus. the iucn red list of threatened species 2016: e. t9660a45226042," 2016.

[2] A. J. Hall and D. J. Russell, "Gray seal: Halichoerus grypus," in *Encyclopedia of marine mammals*. Elsevier, 2018, pp. 420–422.

[3] Joint Nature Conservation Committee, "Species interest features – special areas of conservation," n.d., accessed: 2025-05-29. [Online]. Available: https://sac.jncc.gov.uk/species/

[4] G. D. Hastie, P. Lepper, J. C. McKnight, R. Milne, D. J. Russell, and D. Thompson, "Acoustic risk balancing by marine mammals: anthropogenic noise can influence the foraging decisions by seals," *Journal of Applied Ecology*, vol. 58, no. 9, pp. 1854–1863, 2021.

[5] F. Chen, G. I. Shapiro, K. A. Bennett, S. N. Ingram, D. Thompson, C. Vincent, D. J. Russell, and C. B. Embling, "Shipping noise in a dynamic sea: a case study of grey seals in the celtic sea," *Marine Pollution Bulletin*, vol. 114, no. 1, pp. 372–383, 2017.

[6] S. Asselin, M. O. Hammill, and C. Barrette, "Underwater vocalizations of ice breeding grey seals," *Canadian Journal of Zoology*, vol. 71, no. 11, pp. 2211–2219, 1993.

[7] Y. P. Pozo Galván, M. Pérez Tadeo, M. Pommier, and J. O'Brien, "Static acoustic monitoring of harbour (phoca vitulina) and grey seals (halichoerus grypus) in the malin sea: A revolutionary approach in pinniped conservation," *Journal of Marine Science and Engineering*, vol. 12, no. 1, p. 118, 2024.

[8] L. J. Nowak, "Observations on mechanisms and phenomena underlying underwater and surface vocalisations of grey seals," *Bioacoustics*, vol. 30, no. 6, pp. 696–715, 2021.

[9] M. Pérez Tadeo, M. Gammell, and J. O'Brien, "First steps towards the automated detection of underwater vocalisations of grey seals (halichoerus grypus) in the blasket islands, southwest ireland," *Journal of Marine Science and Engineering*, vol. 11, no. 2, p. 351, 2023.

[10] D. Gillespie, M. Caillat, J. Gordon, and P. White, "Automatic detection and classification of odontocete whistles," *The Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 2427–2437, 2013.

[11] Y. Shiu, K. Palmer, M. A. Roch, E. Fleishman, X. Liu, E.-M. Nosal, T. Helble, D. Cholewiak, D. Gillespie, and H. Klinck, "Deep neural networks for automated detection of marine mammal species," *Scientific reports*, vol. 10, no. 1, p. 607, 2020.

[12] M. Thomas, B. Martin, K. Kowarski, B. Gaudet, and S. Matwin, "Marine mammal species classification using convolutional neural networks and a novel acoustic representation," in *Joint European conference on machine learning and knowledge discovery in databases*. Springer, 2019, pp. 290–305.

[13] D. Duan, L.-g. Lü, Y. Jiang, Z. Liu, C. Yang, J. Guo, and X. Wang, "Real-time identification of marine mammal calls based on convolutional neural networks," *Applied Acoustics*, vol. 192, p. 108755, 2022.

[14] S. M. Van Parijs, P. J. Corkeron, J. Harvey, S. A. Hayes, D. K. Mellinger, P. A. Rouget, P. M. Thompson, M. Wahlberg, and K. M. Kovacs, "Patterns in the vocalizations of male harbor seals," *The Journal of the Acoustical Society of America*, vol. 113, no. 6, pp. 3403–3410, 2003.

[15] T. A. Lamont, L. Chapuis, B. Williams, S. Dines, T. Gridley, G. Frainer, J. Fearey, P. B. Maulana, M. E. Prasetya, J. Jompa *et al.*, "Hydromoth: Testing a prototype low-cost acoustic recorder for aquatic environments," *Remote Sensing in Ecology and Conservation*, vol. 8, no. 3, pp. 362–378, 2022.

[16] A. P. Hill, P. Prince, J. L. Snaddon, C. P. Doncaster, and A. Rogers, "Audiomoth: A low-cost acoustic device for monitoring biodiversity and the environment," *HardwareX*, vol. 6, p. e00073, 2019.

[17] B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th Python in Science Conference*, 2015, pp. 18–25.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[20] T. S. Prajwal and I. A K, "A comparative study of resnet-pretrained models for computernbsp;vision," in *Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing*, ser. IC3-2023. New York, NY, USA: Association for Computing Machinery, 2023, p. 419–425. [Online]. Available: https://doi.org/10.1145/3607947.3608042

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization. arxiv [preprint](2014)," *arXiv preprint arXiv:1412.6980*, vol. 5, 2014.

Advancing Marine Monitoring Through Edge-Optimised Deep Learning: Hydrotwin's Approach to Underwater Acoustic Surveillance

Astrid van Toor[1] and Guilherme Beleza Vaz[2]

[1,2]blueOasis, Ericeira, Portugal

Astrid van Toor; avtoor@blueoasis.pt; blueOasis, Rua do Norte 16, 2655-321 Ericeira, Portugal

Abstract: *Effective marine bioacoustics monitoring is essential for biodiversity conservation and ecosystem health assessment. This paper introduces Hydrotwin, a scalable multi-model acoustic monitoring platform that integrates underwater sound measurement with edge-optimised deep learning for real-time marine species detection and anthropogenic noise analysis in computationally constrained deployments. The system employs two specialised convolutional neural networks: Model A for vessel classification across low (≤112Hz) to medium (112-2500Hz), and high-frequency (2500-5000Hz) bands plus dolphin whistle detection (5000-20000Hz), and Model B for blue and fin whale call identification in the 5-125Hz spectrum. Model development leveraged extensive multi-year data collection efforts comprising 206 hours of expert-labelled recordings across Portuguese waters and the Azores for vessel/dolphin detection, and 1880 hours of baleen whale data. Through spectrogram analysis, multi-band frequency decomposition, and optimised wavelet feature extraction, our ultra-lightweight models (8K-120K parameters) are optimised for high-precision performance: 76-78% precision for vessel detection, 94% precision for dolphin whistles, 70% precision for blue whale ABZ calls, and 77% precision for fin whale pulse calls. We emphasise methodological rigour through temporally unbiased validation splits that prevent data leakage and provide realistic field performance assessment, contrasting with traditional random splits that artificially inflate metrics to unrealistic 90%+ ranges. Real-time acoustic processing occurs in sliding windows, with detection events logged alongside environmental data from integrated sensors measuring wave height, wind speed, water temperature, and dissolved oxygen for subsequent correlation analysis. Field deployments demonstrate Hydrotwin's capacity for autonomous three-month monitoring cycles while supporting conservation policies through precise identification of anthropogenic disruptions and marine mammal presence patterns in diverse ocean environments.*

Keywords: *deep learning, digital bioacoustics, passive acoustic monitoring, internet of things, audio signal processing*

1. BACKGROUND

Marine environments face unprecedented threats [1, 2], demanding scalable bioacoustics monitoring solutions to assess biodiversity and support conservation. Underwater passive acoustic monitoring (PAM) offers a non-invasive approach for continuous surveillance [3, 4, 5], but transitioning deep learning (DL) based PAM from research prototypes to operational systems faces critical hurdles. Current DL approaches often report high accuracies based on biased evaluation protocols that don't always account for temporal correlations in acoustic data [4, 5]. While the literature provides streamlined benchmarks frameworks for blue and fin whales [3, 6], there is a lack of a wider standardised benchmark in the detection of marine acoustic events [4, 5]. Furthermore, deploying sophisticated DL models on computationally-constrained edge devices remains a significant challenge, often requiring performance trade-offs through quantization or pruning [7]. While specific DL architectures have shown promise for tasks like marine mammal or vessel detection [4, 5, 7, 8, 9], the gap towards reliable, edge-optimised, and robustly validated field systems persists.

Like any supervised AI system, labelled data plays a crucial role. The algorithms have been trained using extensive datasets of acoustic data collected in Portugal, labelled by trained people for the detection of dolphin whistles, whale calls, and three classes of vessels based on their dominant frequency and observation of Lloyds' mirror effect.

To address these challenges, we introduce Hydrotwin (HT), a scalable acoustic monitoring platform integrating underwater sound measurement with edge-optimised Convolutional Neural Network (CNN) models for real-time marine species detection and anthropogenic noise analysis. Key features include class-specific frequency band optimisations, a minimal DL layer architecture for computational efficiency, and real-world noise incorporation for enhanced robustness. Field deployments across diverse marine environments (such as Italy, Madeira, Azores, and coastal mainland Portugal) demonstrate Hydrotwin's capacity to detect marine mammals and vessel activity, supporting conservation policies by identifying anthropogenic disruptions.

Labelled data plays a crucial role in the conception of such algorithms. The algorithms presented in this article have been trained using extensive datasets of internal and external acoustic data of continuous measurements labelled by experts for the detection of dolphin whistles, various vessel classes, and whale calls [10].

Section 2 details Hydrotwin's system architecture, our acoustic feature extraction methodology, and our model development approach; Section 3 describes the evaluation framework and results; Section 4 presents a discussion of applications and Section 5 concludes a discussion with future directions.

2. METHODOLOGY

2.1. SYSTEM ARCHITECTURE

Hydrotwin implements a modular IoT (Internet of Things) architecture that integrates acoustic sensors with environmental monitoring components for comprehensive marine surveillance. The system consists of three primary layers: (1) data acquisition, (2) data processing, and (3) visualisation and analysis.

The data acquisition layer collects underwater acoustic signals through hydrophones de-

ployed at strategic monitoring locations. These acoustic sensors are complemented by environmental sensors that record parameters including wave height, wind speed, water temperature, and dissolved oxygen levels. This multi-modal sensing approach provides essential context for acoustic interpretation: wave height and wind speed directly influence mid-frequency ambient noise levels that can mask vessel signatures, water temperature affects sound velocity and thus detection ranges, while dissolved oxygen levels indicate habitat suitability for marine fauna.

The data processing layer performs real-time acoustic analysis directly on deployment units, minimising data transmission requirements and enabling autonomous operation. Hydrotwin runs its processing layer on internally developed hardware; HT-S with a Spotter [11] oceanographic buoy for remote edge deployments and cellular data streams, and HT-C with a cable to the mainland designed for high-performance operations where data flow is essential (including access to live acoustic recordings). Within this hardware exists HT-Intelligence; a software layer that analyses the audio data and outputs relevant information about the soundscape and the classification of acoustic sources. To support this, each device contains an ARM Cortex-A73 CPU with just 512MB RAM optimised for neural network inference. Our Artificial Intelligence (AI) models are optimised to operate efficiently on these computationally-constrained platforms and the cycles can be adapted according to the availability of energy resources.

Audio processing occurs in sliding, overlapping windows with detection events and environmental parameters immediately represented in our web-app dashboard for analysis and environmental data correlation. The complete deployment platform was designed for autonomous operation with minimal maintenance, capable of three months of continuous monitoring between service intervals.

Hydrotwin currently employs two specialised multi-label detection deep learning pipelines optimised for edge deployment: Model A for vessels and dolphins, and Model B for blue and fin whales. Both datasets were constructed from continuous underwater measurements, featuring a reflection of the real-world suitable for representative modelling and analysis through temporally unbiased validation splits; by separating the evaluation sets based on recording sessions rather than random sampling, we prevent data leakage that artificially inflates performance metrics through temporal correlation between adjacent windows. When compared to commonly reported metrics in literature that use random validation splits traditional in ML domains, our unbiased approach along with constrained edge-optimisations may produce lower metrics which we consider a more realistic assessment of field performance. Further context is presented in the results section.

Model A (vessels and dolphins) is built with internally collected data that encapsulates target classes of vessels across low, medium, and high frequency bands, and dolphin whistles. This data consists of 206 hours of data recorded over three years from multiple locations in mainland Portugal and the Azores, comprehensively labelled by in-house experts and collaborators. The data is split stratified based on class presence per 10 minute recordings (Table 1), to ensure a temporally separated even distribution of the training data. These recordings are further split into segments of 4 seconds with 50% overlap to prepare the CNN input.

Class	Train	Val	Test
Background	22,110	3,791	2,930
Fast Vessel	10,635	1,400	2,038
Low-Med Vessel	146,885	19,341	19,225
High-Freq Vessel	142,538	14,346	14,750
Dolphin Whistles	18,149	2,386	2,071
Total	**297,149**	**35,138**	**35,166**

Table 1: Model A Class Distribution

Class	Train	Val
Background	362,944	312,730
Blue Whale ABZ	64,846	28,931
Fin Whale Pulse	26,677	6,301
Fin/Blue Downsweep	31,615	5,581
Total	**473,620**	**351,374**

Table 2: Model B Class Distribution

Model B is targeting the identification of blue and fin whale calls operating on the lower-frequency components of the acoustic spectrum. The data for Model B originates from extensive multi-year and location labelling efforts by [10] for a total of 1880 hours of recordings. Model B follows the data distribution protocol set out by the BioDCASE benchmark challenge for baleen whales [6]. Longer, 11.8-second recording windows were selected to accommodate the longer temporal characteristics of whale vocalisations than the shorter, more transient acoustic signatures of vessel noise and dolphin whistles. The same 50% overlap is used. The final distribution is presented in Table 2.

2.2. FEATURE EXTRACTION AND MODEL DEVELOPMENT

This section provides a high-level overview of the distinct processing pipelines for Model A (vessels and dolphins) and Model B (blue and fin whales).

For Model A, we implemented a multi-band frequency analysis approach that divides the acoustic spectrum into four key regions as per our labelling boundaries: low-frequency vessel bands (≤112Hz), medium-frequency vessel bands (112-2500Hz), high-frequency vessel bands (2500-5000Hz), and dolphin whistle bands (5000-20000Hz). This division enables specialised processing optimised for each acoustic signature type. The feature extraction pipeline combines a log-power mel-scale spectrogram representation (48000Hz sampling rate, 60 mel bands, 2048-point FFT, 1024-point hop length) with band-specific energy ratio calculations normalised against total energy, and whistle prominence detection through adaptive peak finding in high-frequency bands wrapped into a single feature stack for input into our CNN model. This feature processing is applied to an 18-second snippet in Fig. 1.

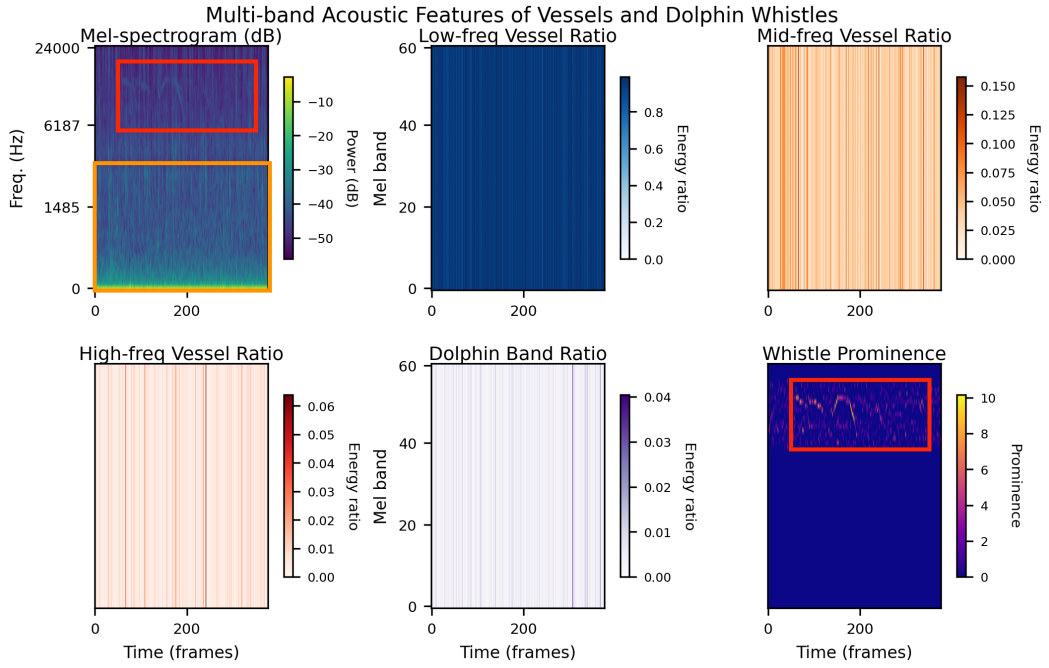


Figure 1: 18-second feature stack frame with CNN processing for vessels and dolphin whistles. Red box presents dolphin whistles, the orange box represents mid to high frequency vessel signals. In production these are 4-second windows.

For Model B, we developed a specialised low-frequency processing pipeline for blue and fin whale calls operating in the 5-125Hz frequency range. The feature extraction pipeline combines a low-frequency log-power spectrogram representation (250Hz sampling rate, 1024-point FFT providing 0.244Hz frequency resolution, 2-second window with 0.2-second hop) with bandpass filtering (5-125Hz) to isolate whale vocalisations from higher-frequency noise. Percentile-based normalisation (5th-95th percentile range) and background subtraction with a Gaussian filter to smoothly subtract the noise floor are applied to ensure robust amplitude scaling across recordings with varying signal-to-noise ratios and environmental conditions. Wavelet features have previously proven to be highly effective for identifying low frequency whale calls [12]. Alongside the spectrogram and its first derivate, we therefore include a computationally optimised wavelet layer that processes only the approximation coefficients and first-level detail coefficients per frequency bin. The feature stack for Model B is presented in Fig. 2.

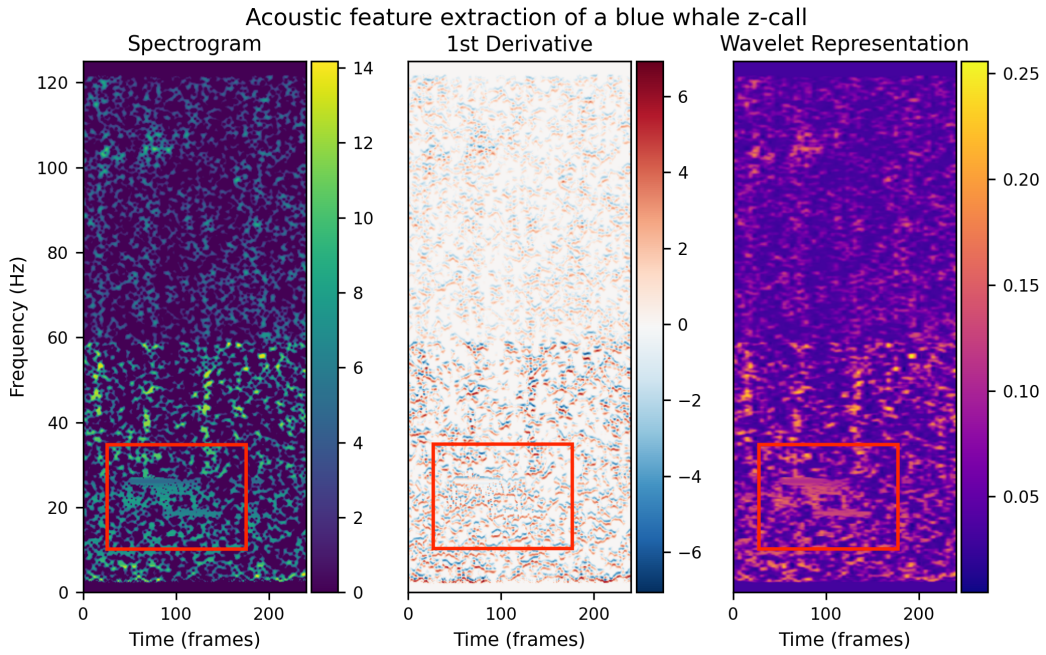


Figure 2: CNN Feature stack for blue and fin whale calls on a 48-second frame. Example shows a blue whale ABZ call [10]. In production these are 12-second windows.

For both models we normalise the data by calculating the mean and standard deviation on the training set and apply these metrics on the evaluation sets and production data, to avoid data leakage. Both models were trained using a focal loss function that addresses the severe class imbalance by emphasising challenging examples. As we are dealing with a multiclass problem, to avoid over prediction on the background class in frames where background observations are present, the background class is activated only when no other classes were detected. The exception to this is when there is a high confidence for background (>75%) with no other classes exceeding this confidence.

Our CNN architectures prioritise computational efficiency through strategic design choices including separable convolutions for more compact computations than traditional convolution layers (Model B only), global average pooling layers before the classification heads for final parameter reductions, batch normalisation for training stability throughout the layers, progressive regularisation (Model A only), spatial dimension reduction, and TensorFlow Lite quantization

to optimise inference speed and RAM usage. These optimisations result in models with 8K-120K parameters - a 97-99% reduction compared to standard and edge-optimised CNNs [13]. This extreme efficiency is what enables real-time inference on energy-constrained marine IoT devices such as Hydrotwin.

3. VALIDATION AND RESULTS

Performance evaluation followed a precision-focused approach, recognising that in ocean sustainability, conservation, and security as major applications and drivers for Hydrotwin, false positives often incur higher costs than false negatives. We implemented class-specific detection thresholds optimised for precision while maintaining acceptable recall rates for key target species. An example of our approach for threshold analysis is presented in Fig. 3.

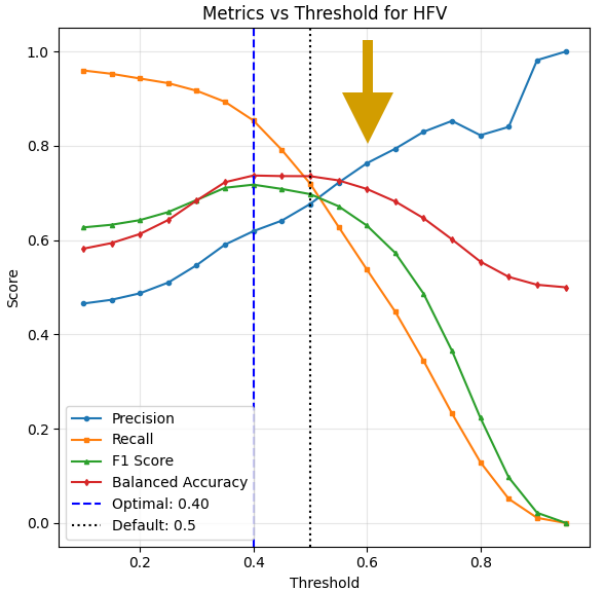


Figure 3: Threshold analysis approach demonstrated in the High-Freq. Vessel (HFV) class on the evaluation set. "Optimal" is maximised F1, the arrow highlights our selected threshold of 0.6 for optimised precision.

Results from our vessel and dolphin detection model (Fig. 4) demonstrate strong precision for vessel detection (76-78%) and exceptional precision for dolphin whistle detection (94%). The notably lower precision for background classification (18%) reflects our conservative approach to minimise false negatives of target classes. Recall rates varied considerably across classes, with background detection achieving 75% recall the remaining classes ranged from 21-57%. This performance profile aligns with our design priorities to minimise false alarms for species of conservation interest. The whale detection model (Fig. 5) shows high values in background discrimination (92% precision, 98% recall) as expected from this highly imbalanced dataset. These values need to be considered in the context of the rare classes; a good precision for blue whale Z-calls (70%) and fin whale pulse calls (77%) with suboptimal recall detections of 28% and 44% respectively – reflecting the over-prediction of the background class. The downsweep call detection performed poorly (10% precision), consistent with challenges reported in benchmark studies [3, 6] due to acoustic similarity to background noise patterns. For

context we provide the results of a traditional random stratified validation split on our vessel and dolphin data in Fig. 6, which achieved highly unrealistic metrics around the 90% range before any specialised tuning due to temporal data leakage. With this we aim to highlight the importance of unbiased data splits.

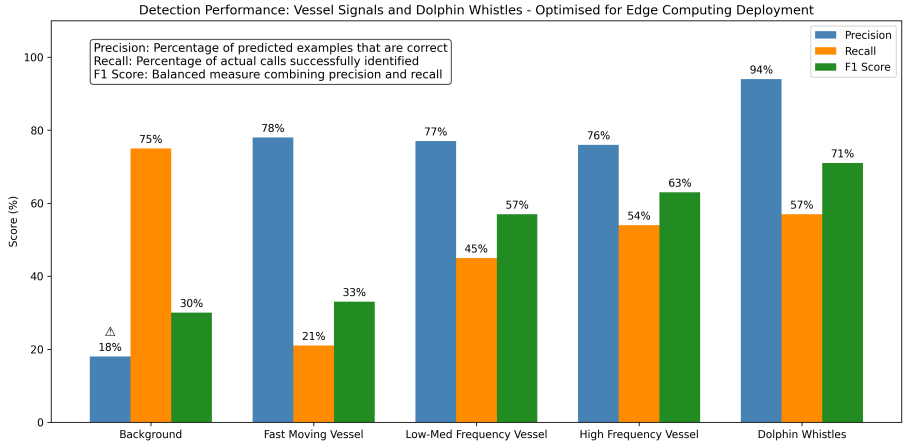


Figure 4: Detection performance of vessel and dolphin targets optimised for precision

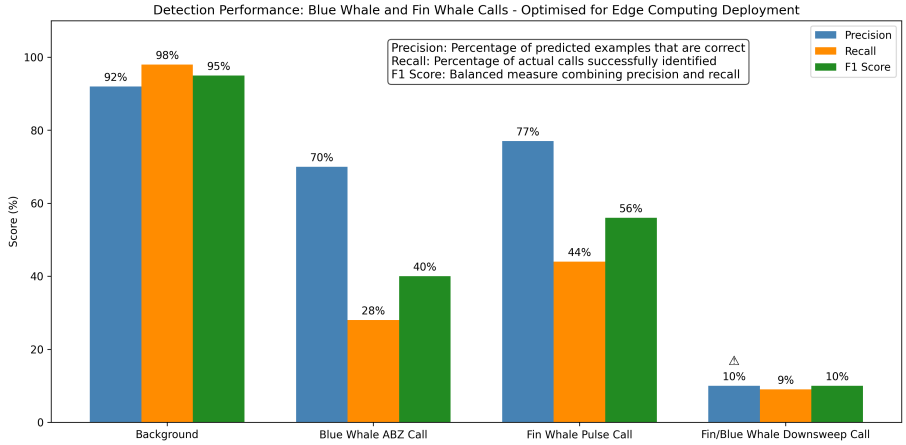


Figure 5: Detection performance for baleen whale targets optimised for precision

4. APPLICATIONS

Hydrotwin enables ocean sustainability and security applications through its real-time acoustic monitoring capabilities in the areas below. As more data, deployments, and classes become available, this list is constantly expanding.

- Species Monitoring: Hydrotwin continuously tracks marine mammal presence, providing temporal distribution data for cetaceans and acoustically active fish. Detection events are automatically logged with environmental context, supporting habitat analysis.

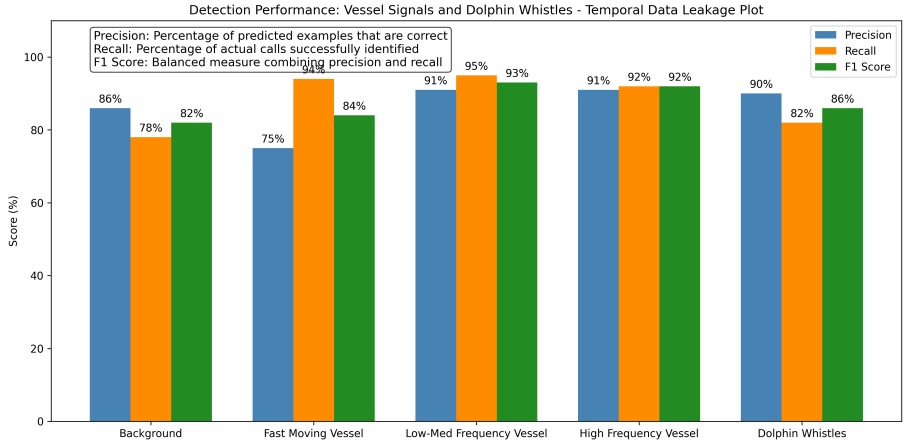


Figure 6: Evaluation results of a highly biased test set due to temporal data leakage

- Anthropogenic Noise Assessment: By classifying vessel acoustic signatures across frequency bands, the system quantifies noise contributions to marine soundscapes. This enables assessment of potential impacts on marine life and supports evidence-based policy decisions for traffic management in sensitive habitats.
- Illegal Activity Detection: By deploying Hydrotwin-intelligence in illegal shipping and fishing zones or other illicit activity routes, it can detect and trigger warnings for unexpected activity.

4.1. OPERATIONAL VISUALISATION

A visualisation of the ”AI in Action” is presented in Fig. 7. Our web-based interface (Fig. 8) provides a monitoring space for the previously discussed applications through statistical summaries that support evidence-based conservation decisions, manual validation of spectrogram displays, and potential for alert configurations for unusual patterns. As a foundational component of an evolving marine digital twin ecosystem, this integrated monitoring framework establishes the infrastructure necessary for transitioning from passive surveillance to predictive ecosystem modelling.

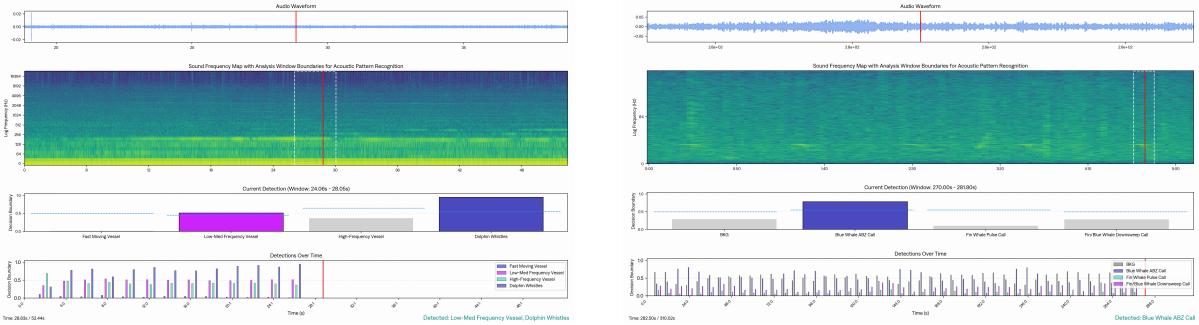


Figure 7: Representation of AI in Action in the Backend

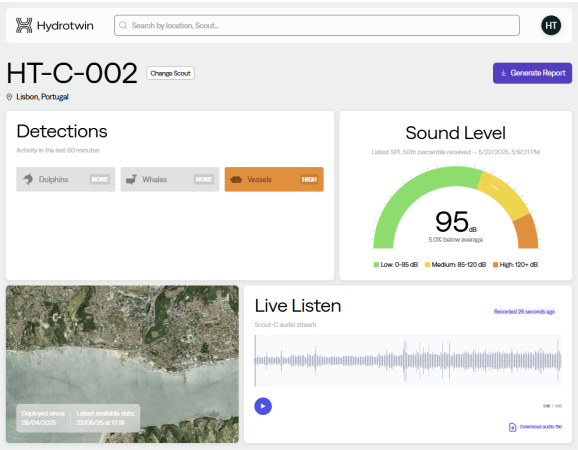


Figure 8: HT-C live listening interface

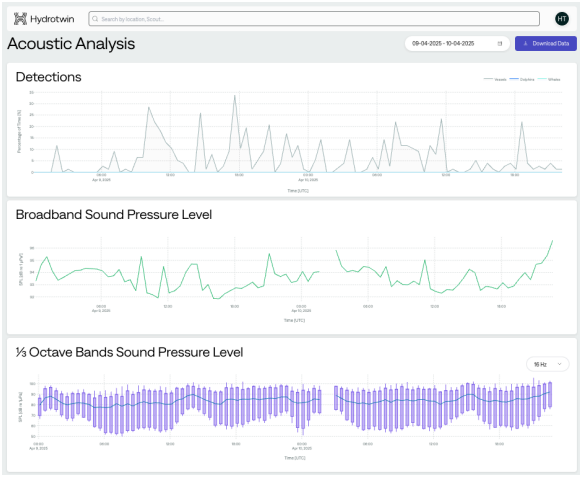


Figure 9: Detection view from HT-C/S

5. DISCUSSION AND FUTURE DIRECTIONS

While Hydrotwin demonstrates significant advances in operational marine acoustic monitoring, several limitations and challenges remain to be addressed. The current system shows uneven performance across detection classes, with particularly low precision for certain call types like downsweep calls (10%). This highlights common challenges in marine bioacoustics: scarcity of balanced training data for rare vocalisations and inherent variability in labelling protocols across practitioners [3, 14]. Through Hydrotwin's ongoing data collection, data cleaning, and labelling efforts, we aim to address these data limitations further.

Environmental and hydrophone variability presents additional challenges, as field validations cannot capture the full range of global acoustic conditions including depth, temperature, and soundscape variations. Edge deployment constraints also force trade-offs between detection accuracy and computational efficiency, relevant in HT-S but mitigated in the continuous energy resource of HT-C in ongoing development effort. Furthermore, the current models are optimised for presence detection rather than precise population counting. However, statistical adjustments based on known precision and recall values can inform population estimates, with ongoing work addressing overlapping call detection and hydrophone calibration challenges.

Our future development focuses on three directions: (1) enhanced detection performance through refined architectures and targeted data augmentation for low-performing classes, (2) advanced anomaly detection for comprehensive environmental monitoring, and (3) evolution toward a complete sensory digital twin ecosystem integrating acoustic monitoring with oceanographic modelling and species distribution forecasting.

As marine environments face increasing pressures, tools like Hydrotwin represent essential components of evidence-based conservation strategies, bridging the gap between research prototypes and operational systems for sustainable ocean management.

6. ACKNOWLEDGEMENTS

We thank Erica Cruz and Carolina Ramos for their labelling efforts with invaluable contribution to the AI models. We also acknowledge the institutional support of Escola do Mar dos Açores, Okeanos, Direção Regional de Políticas Marinhas dos Açores and Porto de Lisboa.

REFERENCES

[1] Dube, K.: "A Comprehensive Review of Climatic Threats and Adaptation of Marine Biodiversity". *Journal of Marine Science and Engineering* **12** (2024), 344.

[2] Ahmed, R. and Tamim, M. T. R.: "Marine and Coastal Environments: Challenges, Impacts, and Strategies for a Sustainable Future". *International Journal of Science Education and Science* **2** (2025), 53–60.

[3] Schall, E. et al.: "Deep Learning in Marine Bioacoustics: A Benchmark for Baleen Whale Detection". *Remote Sensing in Ecology and Conservation* **10** (2024), 642–654.

[4] Domingos, L. C. F. et al.: "A Survey of Underwater Acoustic Data Classification Methods Using Deep Learning for Shoreline Surveillance". *Sensors* **22** (2022), 2181.

[5] Luo, X. et al.: "A Survey of Underwater Acoustic Target Recognition Methods Based on Machine Learning". *Journal of Marine Science and Engineering* **11** (2023), 384.

[6] DCASE: *Supervised Detection of Strongly-Labelled Whale Calls - BioDCASE*. `https://dcase.community/challenge2025/task2`.

[7] Murshed, M. G. S. et al.: "Machine Learning at the Network Edge: A Survey". *ACM Comput. Surv.* **54** (2021), 170:1–170:37.

[8] White, E. L. et al.: "More than a Whistle: Automated Detection of Marine Sound Sources with a Convolutional Neural Network". *Frontiers in Marine Science* **9** (2022).

[9] Nur Korkmaz, B. et al.: "Automated Detection of Dolphin Whistles with Convolutional Networks and Transfer Learning". *Frontiers in Artificial Intelligence* **6** (2023).

[10] Miller, B. S. et al.: "An Open Access Dataset for Developing Automated Detectors of Antarctic Baleen Whale Sounds and Performance Evaluation of Two Commonly Used Detectors". *Scientific Reports* **11** (2021), 806.

[11] Spotter: *Spotter Platfom*. `https://www.sofarocean.com/products/spotter`.

[12] Rademan, M. W., Versfeld, D. J. J., and du Preez, J. A.: "Detecting and Classifying Blue Whale Calls with Wavelet Scattering and Spectral Entropy". *The Journal of the Acoustical Society of America* **157** (2025), 1448–1457.

[13] Team, K.: *Keras Documentation: Keras Applications*. https://keras.io/api/applications/.

[14] Sattar, F.: "A New Acoustical Autonomous Method for Identifying Endangered Whale Calls: A Case Study of Blue Whale and Fin Whale". *Sensors* **23** (2023), 3048.

North Atlantic Right Whale upcall Localization with a Four-Element Acoustic Array on a Slocum Glider

Michael MacGillivray[1], Mae L. Seto[1], S. Bruce Martin[2], and Laird Bolt[2]

[1]Dalhousie University
[2]JASCO Applied Sciences

Michael D MacGillivray, Dalhousie University, 5269 Morris St, Halifax, NS B3J 1B6
Email: michael.macgillivray@dal.ca

Abstract: ***North Atlantic right whales (NARW) are critically endangered, with an estimated 350 individuals remaining. Passive acoustic monitoring offers a means to localize the whales, improving the accuracy of exclusion zones to mitigate vessel strikes and fishing entanglements. This study evaluates multiple time-difference-of-arrival azimuth estimation methods for a compact volumetric array (CVA) integrated onto a Slocum glider. The methods tested include minimum difference estimation, k-means clustering, kernel density estimation (KDE), TDoA maximum likelihood estimation, and azigram-KDE estimation. Simulations in a ROS/Gazebo environment, incorporating realistic underwater acoustic conditions, are used to test each of the methods. Verification was achieved with an in-water trial off the coast of Clam Harbour, Nova Scotia, Canada (2024). The NARW upcall is emulated using a moored source producing linear frequency modulated pulses. The trial revealed that the azigram-KDE estimator performed the best, yielding a mean absolute error of*** 25.6° ***(20.5 % improvement over next best method) and an interquartile range of*** 10.2° ***(40.7 % improvement). Successful localization of NARWs with a glider-mounted CVA advances autonomous marine mammal monitoring and supports ongoing conservation efforts.***

Keywords: *passive acoustic localization, underwater glider, compact volumetric arrays, array processing, signal processing*

1. INTRODUCTION

North Atlantic right whales (NARW) are a critically endangered large whale species with an estimated 350 individuals in existence. To decrease the probability of human-related fatalities, Canada enforces multiple mitigation measures including vessel speed regulations and fishing closures. For instance, upon detection of a NARW, a 2000 km^2 area centred around the point of detection is designated as an exclusion zone for fishing activities [1].

Passive acoustic localization of their vocalizations can more accurately define the exclusion zone around a detected animal compared to centreing around the detectors location. Based on the NARW underwater vocalizations, its relative azimuth and elevation angles with respect to an acoustic receiver can be estimated [7]. To fully define the location of the animal in 3-D space, a range estimate is required. Rough range estimations can be determined from the sonar equation. To calculate this, estimates of vocalization source level (SL), propagation loss (PL), and ambient noise levels (NL) are required. However, the scope of this work pertains only to azimuth estimations of NARW vocalizations.

An emerging and innovative approach uses underwater gliders (UG) as acoustic sensor platforms for passive localization. The Slocum UG (Teledyne Webb Research, Falmouth, MA, USA) is an autonomous underwater vehicle with a variable ballast system to propel itself through the water column in a sawtooth pattern. Integrating UGs with a compact volumetric hydrophone array (CVA) enables passive azimuth estimation of mobile underwater acoustic targets. Using the UG, compared to stationary sonar systems such as sonobouys, enhances localization duration, increases situational awareness for vessel traffic and fishing activity, and makes autonomous mission re-planning meaningful towards longer-duration localization. Prior studies demonstrated the effectiveness of UGs with CVAs in localizing surface vessels and low-frequency acoustic sources operating at depth within the 50–200 Hz band [5]. The application of UGs for marine mammal detection, monitoring, and localization is also documented in the literature [3]. However, there is less development on this for accurate azimuth estimations given irregular and short-duration acoustic sources, such as NARW upcalls.

This study evaluates five methods to estimate the azimuth of NARW upcalls relative to a CVA on an UG. The Slocum G3 UG is integrated with 4 hydrophones, to form a CVA. The azimuth estimation methods include minimum difference estimation, k-means clustering, kernel density estimation (KDE), TDoA maximum likelihood estimation (MLE), and azigram-KDE estimation [6]. Initial algorithmic verification is conducted within a ROS/Gazebo simulation environment [2][8]. Then, the five methods are verified in-water using a stationary underwater acoustic source that emits one-second linear frequency modulated (LFM) pulses to emulate a NARW upcall.

2. PASSIVE ACOUSTIC AZIMUTH ESTIMATION

Passive acoustic azimuth estimation analyzes time, or phase, delays of signals received from multiple hydrophones of known locations (in this case, on the UG) to estimate its azimuth to an underwater source. Assuming the source is in the far-field, plane wave analysis can be applied. The acoustic signal's angle-of-arrival (AoA) to a pair of hydrophones is estimated using equation 1,

$$\theta = \arccos\left(\frac{c \cdot \tau_{12}}{d}\right) \tag{1}$$

where θ is the relative AoA to the hydrophone pair, c is the local speed of sound underwater, τ_{12} is the time-difference-of-arrival (TDoA) estimated from the cross-correlation of the signal arrivals at the two hydrophones, and d is the Euclidean distance between the hydrophones. There is directional ambiguity for a pair of hydrophones, addressed by adding additional hydrophones on a different axis to form a CVA. The next section discussed method to resolve the best azimuth estimate from the CVA.

2.1. AZIMUTH SOLVERS

The performance of five TDoA-based azimuth estimation methods are compared and contrasted for a CVA mounted on an UG. Each method offers a distinct strategy for resolving the directional ambiguity inherent in hydrophone-based measurements. The methods can be broken down into ambiguous azimuth estimation based (first three) and MLE based solvers (last two).

The first method is the minimum-difference solver. The four-element CVA yields 12 ambiguous azimuth estimates. All valid combinations of 3 estimates are generated, excluding pairs from the same hydrophone baseline. Each group's mean angular difference (spread) is computed, and the group with the smallest spread determines the final azimuth.

The second method uses K-Means to cluster the azimuths based on the angular difference. Once completed, the cluster with the minimum angular difference is selected. If the number of values in the cluster is greater than five, the values are sent through the minimum difference solver. If the size is less than 5, the angular mean is taken as the azimuth estimate.

The third method uses a KDE to solve for the optimal azimuth. The probability density function (PDF) of the group of 12 is determined using a von Mises kernel. The von Mises kernel is employed as it is specifically designed for circular data [6]. The peak value of that PDF is the azimuth estimation.

The fourth method implemented is a TDoA-MLE [7]. An array of expected time differences is generated for all possible azimuths, at a specific angular resolution. The optimal azimuth estimation is obtained by using least squares (LS) optimization.

The final method utilizes an azigram-KDE introduced in [6], but adapts the algorithm for CVA azimuth estimation compared to its original use with sonobuoys. This method first constructs an azigram. To do this, short-term Fourier transforms for each signal are computed. A threshold filter is applied to the spectrograms removing uncorrelated noise. Sub-band processing is used, therefore each time-frequency bin is analyzed individually. For each bin, the phase difference is calculated for all possible hydrophone pairs, creating a group of six phase differences. The group is sent to a MLE, which will yield an azimuth estimation. Phase differences are used to estimate the optimal angle because the expected TDoA can corresponds to phase shifts of multiple wavelengths at higher frequencies.

The remaining cells in the azigram are used to generate an azimuth density estimation. Similar to the third method, the von Mises KDE is applied to the azigram results, and the peak of the resulting PDF is the azimuth estimate.

3. NARW LOCALIZATION ROS/GAZEBO SIMULATIONS

The simulation's purpose was to test and validate the five TDoA-based azimuth estimation methods before using them with in-water trial data. The simulation environment was built using

ROS Noetic and Gazebo 11. The Gazebo simulation is built on top of the DAVE project [8]. The UG's mission consisted of pre-defined waypoint tracking. However, the UG is capable of autonomously updating its waypoints based on underwater acoustic localizations, paving the way for future work focused on real-time tracking of marine mammals.

The simulates NARW produces upcalls modelled as LFM pulses from 100 - 200 Hz, with a SL of 165 dB re $1\mu Pa$. According to [7], the Cramér-Rao lower bound azimuth variance does not depend on the frequency modulation of the signal and is determined by the SNR and the minimum and maximum frequencies of the signal spectrum, making the LFM signal an appropriate substitute for simulations. The vocalizations are simulated as time-series data sampled at 16 kHz. Gaussian noise is added to the signals based on SNR calculated using PL estimates from the BELLHOP Acoustic Toolbox [4] and ambient levels obtained from typical sound level distributions measured on a Slocum glider, provided by JASCO Applied Sciences. The localization simulator also includes Gaussian error terms for hydrophone synchronization ($X[\text{s}] \sim \mathcal{N}(0, (\frac{2}{f_s})^2)$ and UG heading ($X[^\circ] \sim \mathcal{N}(0, 5^2)$). The simulation trials consisted of 5,000 localization measurements spanning source-receiver ranges of 100 m to 10 km and azimuths from -180° to 180° relative to the UG. The performance metrics are median error, mean absolute error (MAE), and interquartile range (IQR).

3.1. SIMULATION RESULTS

The results for each azimuth estimation method considered are shown in Table 1. All methods performed well in simulation. Performance-wise, the azigram KDE and MLE methods slightly edged out the others. These results are expected as both methods use the LS-MLE which is optimal for Gaussian error estimations – the case in simulation [7]. Note, real underwater acoustic environments introduce a range of complexities not captured in simulation, which may influence the effectiveness of the algorithms.

azimuth estimation method	**median error** (°)	**MAE** (°)	**IQR** (°)
minimum difference estimator	-0.81	13.1	18.7
k-means clustering estimator	-0.17	14.6	19.6
KDE estimator	-0.65	11.7	17.3
maximum likelihood estimator	-0.20	10.0	15.3
azigram KDE estimator	-0.65	10.2	15.7

Table 1: Simulated azimuth solver results (100–200 Hz). All methods have relatively high accuracy with the MLE and azigram-KDE performing slightly better, as expected.

4. MOORED SOURCE LOCALIZATION IN-WATER TRIALS

In March 2024 JASCO Applied Sciences completed a trial using an UG integrated with a four-element CVA and a JASCO OceanObserver that was performing real-time detection of the signals for telemetry to shore. The hydrophones were mounted on the UG's nose, tail, port side wing, and starboard side wings. The array geometry was defined by the following 2D coordinates (in meters): $[0.0, 0.0]$, $[0.04064, -2.3]$, $[-0.5, -1.6]$, $[0.5, -1.6]$.

The UG travelled both towards, and away from, a moored, calibrated omnidirectional source

projecting acoustic LFM pulses from 700 - 850 Hz and 1250 - 1500 Hz. Higher frequencies were used compared to the simulation due to limitations of the source. The range between the UG and source varied from 125 m to 12 km. The source was at an altitude of 6 metres above the sea floor and at a depth of $\approx$ 70 metres. A repeating sequence of three source signals was projected during the trials: signal A (low-frequency LFM at 165 dB re 1 μPa) followed by a 1-min pause; signal B (low-frequency LFM at 170 dB re 1 μPa) followed by a 1-min pause, and signal C (high-frequency LFM at 170 dB re 1 μPa) followed by a 3-min pause.

4.1. IN-WATER TRIAL RESULTS

The experimental results are presented in Table 2. The first clear trend seen is an increase in azimuth MAE compared to the simulation results. This due to unmodeled uncertainties in the underwater environment, causing a number of outlier measurements. In future work with the simulator, these errors will be included to better reflect real-world conditions.

Among the five methods evaluated, the azigram-KDE estimator demonstrated the best overall performance, achieving a 20.5% reduction in MAE and a 40.7% reduction in IQR compared to the next best method (KDE estimator). Sub-band processing proved effective for accurate azimuth estimation in the presence of in-band noise. In contrast, the performance of the other broadband cross-correlation methods degraded as both correlated and uncorrelated noise levels increased.

azimuth estimation method	**median error** (°)	**MAE** (°)	**IQR** (°)
minimum difference estimator	2.70	34.8	20.7
k-means clustering estimator	-1.32	35.0	21.5
KDE estimator	-2.28	32.2	17.2
maximum likelihood estimator	-1.32	39.8	49.2
azigram KDE estimator	-1.06	25.6	10.2

Table 2: In-water azimuth estimation results. The azigram-KDE estimator outperformed all other methods, while the MLE estimator underperformed due to the Gaussian error assumption being unsuitable.

The MLE performed worse out of all the methods, due to its Gaussian error assumption based on LS optimization. A Gaussian TDoA error is invalid, due to uncertainties in the underwater channel and the underwater acoustic ambient, which introduce outliers that LS optimization is particularly sensitive to. The solver was adapted to use least absolute error (LAE), assuming a Laplacian distribution. Changing the loss function improved the MAE and IQR to 34.8° and 19.9°, respectively. Additionally, the azigram-KDE estimator increase performance when using LAE as its loss function, with a MAE of 22.4° and a IQR of 9.89°.

From Table 3, the higher frequency (HF) signal resulted in a higher azimuth MAE. This is an interesting result, as it is expected the signal with higher frequency and higher bandwidth would provide a more accurate azimuth estimation. The higher MAE can be attributed to increased risk of spatial aliasing and attenuation/signal distortion for the higher frequency signal. However, this is not expected to pose a significant issue for localizing NARW upcalls, which predominantly occur within the 100–200 Hz frequency range. Additionally, it is seen that signal B performs better than the other two signals. This shows clearly that increasing SNR will improve the azimuth estimation accuracy.

signal type	median error (°)	MAE (°)	IQR (°)
Signal A (low-frequency LFM at 165 dB re 1 μPa)	1.21	23.3	9.27
Signal B (low-frequency LFM at 170 dB re 1 μPa)	0.407	17.7	8.03
Signal C (high-frequency LFM at 170 dB re 1 μPa)	-0.797	30.8	9.95

Table 3: Azigram-KDE results for each signal. The LF signals performed better than HF, indicating spatial aliasing or higher attenuation/signal distortion at higher frequencies. Signal B (LF and higher SL) performed best, indicating that higher SNR will improve localization accuracy.

5. CONCLUSION

This work examined five methods to solve for the azimuth angle between a UG and a vocalizing NARW producing upcalls. The azimuth estimation methods were first verified and tested in a simulation environment, where all methods demonstrated reliable and consistent performance. In future work the simulation environment will be used for tracking simulation.

In-water trial data was used to verify the azimuth estimation algorithms. The results show the azigram-KDE estimator outperforms the other broadband TDoA algorithms. While this method can exhibited spatial aliasing at higher frequencies, it is not expected to pose a significant issue when localizing NARW upcalls, which primarily occupy lower frequency bands.

REFERENCES

[1] Government of Canada. Protecting the north atlantic right whale: Speed restriction measures in the gulf of st. lawrence - ssb no.: 10/2024, 2024. Accessed: Apr. 2024.

[2] N. Koenig and A. Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *IEEE/RSJ IROS 2004*, volume 3, pages 2149–2154 vol.3, 2004.

[3] Elizabeth T. Küsel and et al. Marine mammal tracks from two-hydrophone acoustic recordings made with a glider. *Ocean Science*, 13(2):273–288, 2017.

[4] Acoustical Society of America. Ocean acoustics library - oalib, February 2024. Accessed: Dec. 2024.

[5] Dajun Sun and et al. Low frequency three-dimensional doa estimation for underwater gliders using an arbitrary tetrahedral array. *Applied Acoustics*, 214:109707, 2023.

[6] Kamden Thebeau. Localization of north atlantic right whales using a deformable sonobuoy array. Master's thesis, Dalhousie University, December 2024.

[7] Ildar R. Urazghildiiev, Bruce Martin, and David E. Hannay. The accuracy of bearing estimates of wideband signals produced by marine animals. *IEEE Journal of Oceanic Engineering*, 46(3):1057–1067, 2021.

[8] Mabel M. et al. Zhang. Dave aquatic virtual environment: Toward a general underwater robotics simulator. In *IEEE/OES AUV Symposium*, 2022.

Correlation of NARW Vocalizations with Behaviours based on Analysis of Tagged Whale Data towards more Informed Localization

Marina Antipina[1], Mae Seto[1], Jay Kirkham[1], and Sarah Fortune[1]

[1]Dalhousie University, 6299 South St, Halifax, Nova Scotia, B3H 4R2, Canada

Marina Antipina, Dalhousie University, 6299 South St, Halifax, Nova Scotia, B3H 4R2, Canada, 1-902-494-1630, mr549579@dal.ca

Abstract: *The North Atlantic right whale is a critically endangered large whale species due to ship strikes and net entanglement. Insight into their underwater acoustic vocalizations and correlated behaviours was explored through customized animal tracking solutions (CATS) tags. The right whale vocalizations were studied through their call type, sound pressure level (SPL), duration and frequency content whereas their behaviours were interpreted through an assessment of their dive phases and profiles. The acoustic analyses shows that the upcalls have distinct features in their SPL, duration, and frequency content based on their behaviours. The findings show promise in facilitating localization and tracking efforts towards the right whales' conservation.*

Keywords: *NARW, right whale, acoustics, vocalization, CATS, tag, localization, tracking*

1. INTRODUCTION

The North Atlantic right whale (NARW) is one of the most endangered large baleen whales. They are at risk of extinction with a population of approximately 356 individuals remaining as of 2022 [1].

The Gulf of Saint Lawrence (Quebec, Canada) is an active feeding ground for the right whales where 40 % of this population forage from May through October [2]. This area is also a major shipping route to the Great Lakes and is a fixed-gear fishery which overlaps with the right whale presence. The NARWs shift into the Gulf of St. Lawrence therefore led to an increase in their mortality and serious injury [1, 2] from vessel strikes and entanglements in fishing gear.

The study of correlations between right whale behaviours and vocalizations are important for behavioural models such as JASCO's Animal Simulation Model Including Noise Exposure (JASMINE). Specifically, if we know the dive phase and depth associated with each call type, these parameters can serve as inputs to behavioural models. When coupled with tools such as the Robot Operating System and Gazebo, they can generate realistic data sets for the development of detection, localization and tracking algorithms.

2. METHODS

2.1. Whale Tagging Procedure to Collect Data

In-water measurements of NARW vocalizations and movements were captured through Customized Animal Tracking Solutions (CATS) tags on four individuals. In-situ data were collected using suction-cup attached CATS tags deployed on North Atlantic right whales in the Gulf of St. Lawrence in July 2023 by Dr. Fortune's Lab at Dalhousie University and NOAA. CATS tags recorded 3-dimensional accelerations (accelerometer, 200 Hz), attitudes (gyroscope, 50 Hz) and magnetic north for navigation (magnetometer, 50 Hz) as well as depth (pressure sensor), light level (light sensor), and temperature (thermometer) at 50 Hz. Visual imagery (camera, 25 frames/s), acoustic signals (hydrophone, 24 and 96 kHz), and horizontal positioning (Fastloc/Argos GPS, 0.1 Hz) when the tag surfaces were also collected. Four tags (tag numbers are 46, 47, and two 48) were deployed on 4 individuals with an 8 m carbon fibre pole from a small rigid hull inflatable boat. The individuals were identified using the New England Aquarium photo-ID catalogue and assigned an Eubalaena glacialis (EG) number. Individuals known to be active mothers or who were recently entangled or in poor health were not selected [3].

2.2. Tag Data Analysis

The raw CATS tag data were down sampled to 1 Hz, calibrated and transformed into the whales' frame of reference following a standard process [4].

All acoustic data (from the hydrophones) were analyzed with JASCO Applied Sciences' (Halifax, Canada) PAMlab tools which performed automated analyses of the acoustic ambient and marine mammal vocalizations.

A combination of automated detector-classifiers (or automated detectors) and manual review and analysis by experienced analysts was used to determine the presence of NARW calls in the acoustic data. First, a suite of automated detectors was applied to the acoustic data to search for tonals which are narrowband, often frequency-modulated, signals. The automated tonal signal detector identified signals with a high signal-to-noise ratio in frequency-time space spectrograms and classified them against a library of marine mammal calls. All four tags representing the 24 sound (.wav) files (total 24 h of data and each sound file is 60 min duration) were manually analyzed for the presence/absence of NARW calls.

The behaviour was categorised using vertical dive behaviour following well established dive phase metrics. A MATLAB version of the tagtools R package [5] used the CircStat toolbox [6]. This script defines individual dives and classifies them as surfaced, ascent, descent and bottom phases (≥80% of max dive depth) [5]. For purposes of this research, the surface was classified as 0-1 m depth, with dives identified when the individual crossed a 5 m depth threshold. Dive shape was then defined by the proportion of time spent per dive in the bottom phase whereby square-shaped dives have more than 50% of total dive duration in the bottom phase, V-shaped dives have less or equal than 20% of total dive duration in the bottom phase, and U-shaped dives have bottom phase duration between the two [7].

The location of each call collected was acquired from a pseudotrack. The pseudotrack was recreated following a standard process [4] using magnetometer declination angle, heading, in-situ recorded whale positions and tag positions.

Through an examination of the dive profile statistics concurrent with vocalization type statistics it is proposed to explore the strength of correlations between behaviours and vocalizations. In addition to the tag data analysis, the detection range predictions for each individual call were established and compared to each other.

2.3. Detection Range Predictions

JASCO's Detection Range Modeling (DRM) tool estimates the detection range to an acoustic source. The predicted detection range represents the range at which there is a 50% probability of detection for a signal of a given source level, bandwidth characteristic, call depth, and expected ambient noise conditions, along with modelled propagation loss, based on specified environmental parameters (sound speed profile, geoacoustics, bathymetry, and wave height).

Specifically, the source level (SL) and standard deviation (SD) for a right whale upcall were selected as SL = 165 and SD = 3.5 in dB re 1µPa at 1m. The right whale frequency inputs were chosen to span from 100 to 200 Hz. The receiver depth was specified between 22.6 and 42.6 m based on the channel sound axis depth which was calculated as 32.6 m. The detection threshold of the detector was set at 2 dB. Comparison of the predicted detection ranges of calls occurring at different dive phases and depths can inform more accurate NARW localization and tracking efforts.

3. RESULTS

The acoustic analysis confirmed 31 calls from the four tags with 18 calls identified as upcalls, 7 as moans, and 6 were unknowns. To validate that the calls were produced by the tagged individual, a threshold of 130 dB re 1 µPa was imposed to eliminate the distant calls picked up by the tags. If the recorded calls were lower than the threshold, the calls were assumed not to be produced by the tagged whale.

The whale with tag 47 (animal ID 3860, female, age 15) had the highest number of calls as well as the greatest diversity in call type, behaviour, dive phase and profile. The data from tag 47 were therefore chosen as the data set for developing the analysis methods.

There were 20 calls detected in tag 47 where 14 were identified as upcalls, 3 as moans, and 3 were unknowns. However, after applying the threshold, only 4 upcalls and 3 moans were considered valid for the tagged individual. This paper focuses on only the upcalls as it is produced by all ages and both sexes, and thus widely used for detecting right whales.

As discussed, the dive profile was divided into four phases: surfaced, descent, bottom-follow, and ascent (Fig. 1). During the tag 47 deployment the whale first engaged in deep-feeding and then social activities as determined from the dive profile (Fig. 1).

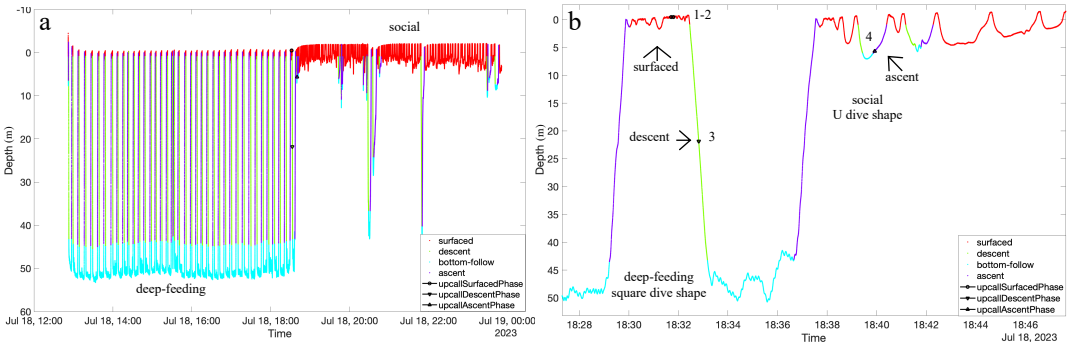


Fig. 1: Dive phase classification: (a) dive profile for deep-feeding and social behaviours and (b) zoomed-in dive profile at transition between both behaviours.

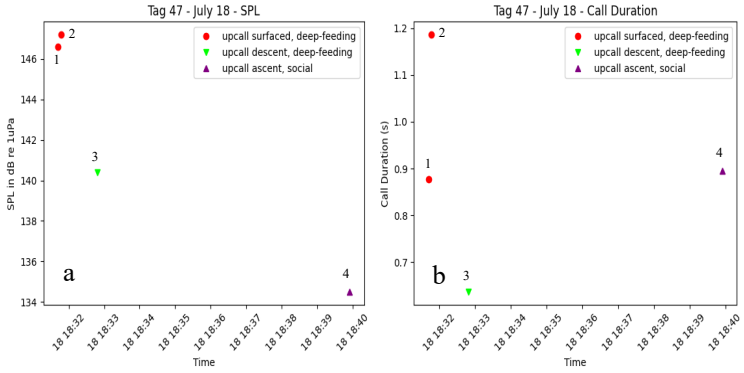
The motion analyses determined at what dive phase the four upcalls occurred in the water column. The first two upcalls (1-2) occurred during the surfaced phase (red) whereas the third upcall (3) occurred during the descent phase (green) (Fig. 1, b). All three calls occurred during the deep-feeding behaviour described with the trapezoidal dive shape. The fourth upcall (4) occurred at the ascent phase (purple) during the social behaviour depicted with the U-dive shape (Fig. 1, b).

The first part of the analyses compares how characteristics of individual calls such as sound pressure, duration, and frequency span vary with the behaviors, dive phases and profiles (Fig. 2).

The highest SPL upcall (2) occurred in the surfaced phase during deep-feeding whereas the lowest SPL upcall (4) happened during the ascent phase during socializing (Fig. 2, a).

During deep-feeding, upcall (2) at the surfaced phase had the highest duration whereas upcall (3) at the descent phase had the lowest duration (Fig. 2, b).

The highest frequency span was upcall (4) in the ascent phase during social behaviour whereas the lowest frequency span was upcall (2) at the surfaced phase during deep-feeding behaviour (Fig. 2, c).



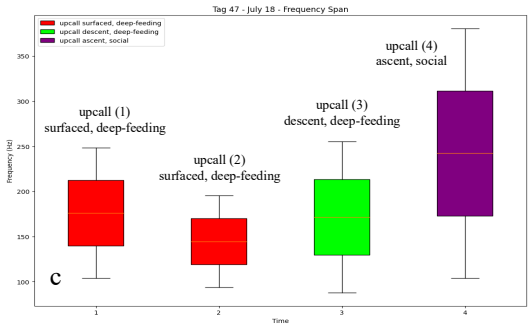


Fig. 2: Characteristics of individual calls such as sound pressure level (a), duration (b), and frequency span (c) varying with the behaviours, dive phases and profiles.

It should be noted that the recorded SPL was impacted by where the tag was attached as well as a result of tag slip. As shown, there is distinct variability in the whale's call durations and frequencies with different behaviours, dive phases and profiles. More data on call characteristics variability within an individual will be valuable for NARW monitoring and conservation.

The conditional detection ranges (Fig. 3) for each call were generated using the 10th, 50th, and 90th percentiles of the empirical noise probability distribution, where the maximum noise level considered is the 90th percentile of the distribution, and the minimum noise level is the 10th percentile of the distribution. On the other hand, the unconditional detection ranges for each call were produced using the full statistical distribution of the noise level (EPDF). In all cases, the source level was treated as a Gaussian distribution with a specified mean and standard deviation discussed earlier.

Since the two upcalls occurred at surfaced phase during deep-feeding at similar depth and location, the DRM was only run for one of them. The detection ranges for upcall (2) at surfaced phase during deep-feeding ranged from up to 5 km on the conditional plot (Fig. 3, a) and up to 11 km on the unconditional plot (Fig. 3, b). The detection ranges for upcall (3) at descent phase during deep-feeding exceed 8 km on the conditional plot (Fig. 3, c). and up to 20 km on the unconditional plot (Fig. 3, d). The detection ranges for the upcall at the at ascent phase during socializing are similar to the detection ranges of the two upcalls at surfaced phase during deep-feeding (Fig. 3, a and b).

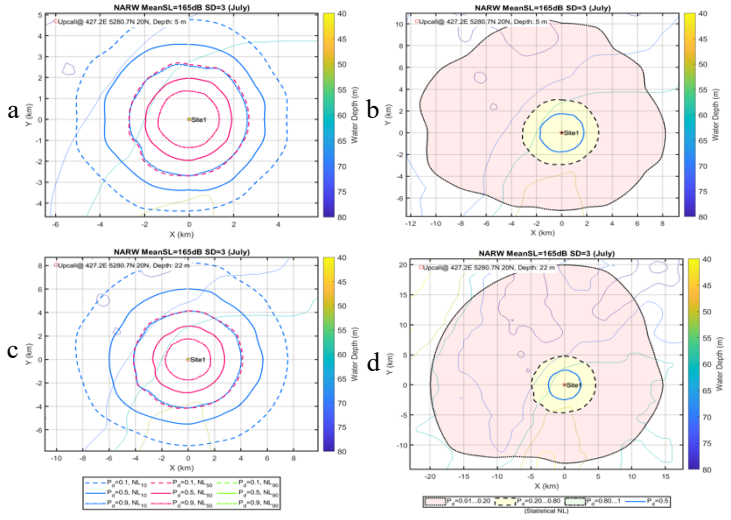


Fig. 3: The conditional detection ranges (a,c) and the unconditional detection ranges (b,d) for the upcall (2) at surfaced phase during deep-feeding (a,b) and for the upcall (3) at descent phase during deep-feeding (c,d) respectively.

4. DISCUSSION

Similar acoustic analysis can be applied for other NARW calls including moans and gunshots. The findings show promise towards a statistical description of individual variations in NARW calls applicable to developing methods for future work with larger data sets. However, more tag data is needed for stronger and statistically significant correlations. The desired research outcome is more accurate determination of NARW activity areas based on detection of their vocalizations. Accurate localizations could facilitate smaller vessel slow-down zones and fishing area closures and more timely alerts to nearby vessels.

5. ACKNOWLEDGEMENTS

We would like to express our deep gratitude to the JASCO members, in particular, Terry Deveau for all his dedicated help and assistance with running DRM, Allison Richardson and Katie Kowarski for their valuable help and insights with the acoustic tag data analysis. A big thank you to the Intelligent Systems Laboratory and Dr. Fortune's Lab at Dalhousie University as well as Mitacs for their support in this research.

REFERENCES

[1] Pettis, H. M., Pace, R.M., & Hamilton, P.K., North Atlantic Right Whale Consortium 2022 annual report card, *Report to the North Atlantic Right Whale Consortium*, pp.1-15, 2023.

[2] **Crowe, L.M., Brown, M.W., Corkeron, P.J., Hamilton, P.K., Ramp, C., Ratelle, S., Vanderlaan, A. S. M., & Cole, T. V. N.**, In plane sight: a mark–recapture analysis of North Atlantic right whales in the Gulf of St. Lawrence, *Endangered Species Research*, 46, pp.227–251, 2021.

[3] **Andrews, R.D., Baird, R.W., Calambokidis, J., Goertz, C.E.C., Gulland, F.M.D., Heide-Jørgensen, M.P., Hooker, S. K., Johnson, M., Mate, B., Mitani, Y., Nowacek, D.P., Owen, K., Quakenbush, L.T., Raverty, S., Robbins, J., Schorr, G. S., Shpak, O.V., Townsend, F.I., Uhart, M., Wells, R. S., & Zerbini, A. N.**, Best practice guidelines for cetacean tagging, *Journal of Cetacean Research and Management*, 20(1), pp.27–66, 2019.

[4] **Cade, D.E., Gough, W.T., Czapanskiy, M.F., Fahlbusch, J.A., Kahane-Rapport, S.R., Linsky, J. M. J., Nichols, R.C., Oestreich, W.K., Wisniewska, D.M., Friedlaender, A.S., & Goldbogen, J.A.**, Tools for integrating inertial sensor data with video bio-loggers, including estimation of animal orientation, motion, and position, *Animal Biotelemetry*, 9, pp.1-21, 2021.

[5] **DeRuiter, S. L., Johnson, M., Sweeney, D., McNamara-Oh, Y. J., Fynewever, S., Tejevbo (Oghenkevwe), R., Marques, T., Wang, Y., & Ogedegbe (Oghenesuvwe), S.**, tagtools: Tools for Working with Data from High-Resolution Biologging Tags, *R package version 0.1.0*, pp.1-91, 2023.

[6] **Berens, P.**, CircStat: A Matlab toolbox for Circular Statistics Toolbox, *Journal of Statistical Software*, 31(10), pp.1–21, 2009.

[7] **Wildlife Computers,** Behavior Log, Dive Classification, *Richmond, WA, USA: Wildlife Computers,* pp. 1-3.

Structured Session
09. Modeling techniques for underwater acoustic scattering and propagation (including 3D effects)

Organised by Boris Katsnelson and Pavel S. Petrov

A Machine Learning Approach for Model Selection in Underwater Acoustic Propagation

Finn P. Henman[1], Joshua J. Wakefield[1], Christopher Ashworth[2], Stewart G. Haslinger[2], Jason Ralph[1], and Duncan P. Williams[3]

[1]Department of Electrical Engineering and Electronics, University of Liverpool
[2]Department of Mathematical Sciences, University of Liverpool
[3]Dstl

Finn Henman, University of Liverpool, Liverpool, L69 7ZF, fhenman1@liverpool.ac.uk

Abstract: Underwater acoustic propagation models use different assumptions and approximations to simplify the wave equation. These assumptions make different models more or less suitable for modelling propagation in different ocean environments, such as a ray tracing, parabolic equation, or normal mode model. Ensuring that one's choice of underwater acoustic propagation model is tailored to a given scenario is important to ensure propagation is modelled as accurately as possible.

Current methods for deciding which model to use are often based on simple thresholds for high or low frequency [1]. For example, one may choose to use a ray tracing model for acoustic frequencies above 1000 Hz, and a parabolic equation model below 1000 Hz. A tool that selects the most suitable model based on multiple environmental parameters would be immensely useful for a variety of underwater acoustics applications. This paper presents a novel method to select the most appropriate model for a given scenario by using a random forest classifier to predict the best model based on the input parameters.

A training set of ocean environments was generated using a variety of values for depth, range, bathymetry, and sound speed profiles. These environments were passed through the RAMSurf (parabolic equation) and BellhopCXX (ray tracing) propagation models and results were compared with Kraken (normal mode), which acted as an arbiter. This provides a basis for model selection supported by empirical data, which takes into account a wider selection of environmental parameters, whilst also allowing for edge cases to be modelled more accurately. The results show a high level of accuracy, with the potential to be further expanded to include a wider range of propagation models and alternative selection metrics.

Keywords: *propagation modelling, machine learning, simulation*

1. INTRODUCTION

In ocean acoustics, propagation modelling is useful for a wide variety of applications, including the study of marine wildlife, underwater navigation, and seabed surveying [2]. For the selection of an underwater acoustic propagation model, it is important to tailor the chosen model to the environment being modelled by taking into account the underlying assumptions and their strengths and weaknesses.

The ray tracing and beam tracing methods are derived from the Helmholtz equation through high-frequency approximations, where the wavelength of the acoustic wave is much smaller than the dimensions of the environment being modelled [3]. These methods simplify complex wave phenomena into more manageable geometrical constructs. This makes ray tracing most suitable for modelling the propagation of high-frequency sound waves in deep water [4].

The parabolic equation (PE) solution of the Helmholtz equation works under the assumption that the wave only propagates in one direction. The Helmholtz equation is factorized into forward and backward propagating components; the backward components are discarded [5]. These models are well suited for modelling the propagation of low-frequency waves in shallow environments [6].

The aim of this research is to provide a framework for the selection of the most suitable candidate from a choice of PE and ray tracing models for specific environmental and frequency parameter spaces. Specific questions are whether the PE and ray tracing models are capable of good performance beyond the standard parameter spaces for which they are usually implemented, and what does the parameter space in which they both show good performance look like. The propagation models used here for this selection competition have been primarily chosen due to their ubiquity in the propagation modelling literature (see for example [4, 6, 7]):

- BellhopCXX - a C implementation [8] of the ray tracing model Bellhop developed by Porter [4].
- RamSurf - a C implementation [9] of the parabolic equation model RAM developed by Collins [5].

In a previous study considering computational cost and convergence criteria [10], RAM and Bellhop were compared. Here, the focus is on selection between the two models for a variety of environmental parameters, rather than considering only computational efficiency. To facilitate these comparisons, a machine learning (ML) approach was implemented, for which an arbiter was required. In practice, this arbiter can be selected for specific environments under consideration, but for this initial demonstration of the approach, another commonly used model was selected, namely Kraken, a normal modes model developed by Porter [11]. The arbiter model is used to determine the "ground truth" against which the two competing models are compared. Bellhop and RAM were chosen as the initial models of interest because of their common use and the fact that they are generally considered to best suited to contrasting conditions, making the transition region between the suitability of the two of particular interest.

The adoption of an ML approach is ideal considering the number of parameters affecting performance; the alternative of using a simple frequency threshold to decide between two models is straightforward, but this approach will fail to find the most suitable model for edge cases, such as low-frequency for a deep environment. ML is well suited to defining the decision boundary in a multi-dimensional parameter space.

2. METHODOLOGY

The training data were generated for a limited selection of generic environments, designed to balance variation and simplification to retain realism whilst avoiding excessive time and computational resources. Every realisation had a flat surface boundary at the air-ocean interface, and a source depth of 25 metres. The environment types were each defined from maximum and minimum values of depth, frequency and surface sound speed, from which each realisation would have a random value chosen from a uniform distribution. Additionally, a random sound speed profile type was chosen. The types of environments are shown in Table 1 along with the range of values for their parameters. This iteration of the modelling framework has been developed to facilitate relatively straightforward expansion, both for the number of environments considered, and their complexity (i.e. defined by the input parameters/labels).

The maximum range of an individual realisation was calculated as a function of frequency, in order to reduce the time spent computing; high-frequency, long-range problems would take a significant amount of time to compute and most of the returned transmission loss plot would be of little use due to signal attenuation. Generally speaking, high-frequency cases are associated with short ranges and low-frequency realisations, with long ranges. This relationship was defined by a linear interpolation of values provided in Table 1 of [10].

Environment	Max Depth (m)	Frequency (kHz)	Surface Sound Speed (ms^{-1})	Sound Speed Profile types
Shallow	60 - 200	0.2 - 5	1480 - 1550	Constant, Upward Refracting
Deep	1500 - 5500	0.05 - 1.5	1480 - 1550	Constant, Munk
Coastal	50 - 60	0.5 - 10	1480 - 1550	Constant, Downward Refracting, Upward Refracting
Sloped	252 - 462	0.4 - 1.4	1480 - 1550	Constant

Table 1: Summary of the different environment archetypes used and their properties, with ranges provided for depth, frequency, surface sound speed and SSP type.

Four different types of sound speed profiles (SSP) were also defined, and these were prescribed as possible options for each environment type appropriately, e.g. in the Shallow environment, the SSP is typically modelled using an upwards refracting or constant profile. The Sloped environment is characterised by an initial segment of flat seabed at depth 200 metres, followed by a shallow downward incline (fixed to $3°$) to a randomly selected depth, followed by another flat stretch out to the maximum range.

To compare the results of the models of interest with the arbiter, a metric was defined. In this case, the modelling output was converted to a single numerical value (mean absolute error), representing the degree to which the results differed from each other. Each propagation model was used to generate a one-dimensional (1-d) array of transmission loss (TL) measurements along the full range of the environment at a fixed receiver depth of 25 metres, with range discretization of one metre. Figure 1 shows an illustrative plot with the three 1-d arrays generated for the same realisation (labelled using values randomly generated from the ranges given in Table 1).

The arrays were pre-processed to remove any outlier values, such as infinities or NaN values which would adversely affect any statistical measure such as the mean. The corresponding points had to also be removed from the arrays generated from the other models to ensure that the same length of range was compared universally. A smoothing filter was then applied to the arrays, again to remove any large outlying values or "jittering" sections. These equalization and smoothing steps were necessary to ensure that the mean absolute error was a meaningful metric for comparing the models. A comparison of a TL array before and after the smoothing filter is shown in Figure 2.

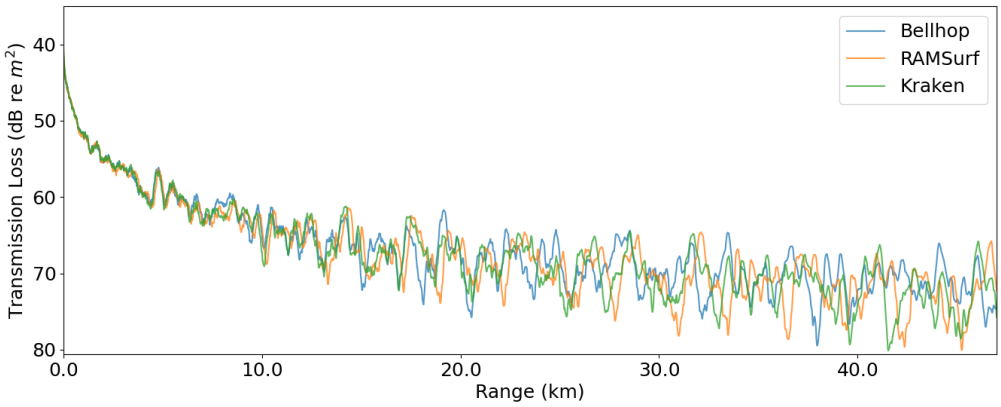


Figure 1: A comparison of the transmission loss at a fixed depth of 25m between all three propagation models. The Shallow environment parameters used were a maximum depth of 72 m, a frequency of 1160 Hz, a maximum range of 48.5 km, and an upwards refracting sound speed profile with a surface speed of 1508 ms^{-1}.

Following this, the mean absolute error (MAE) is calculated using the equation

$$\frac{1}{N}\sum_{i=1}^{N}\left|X_i - X_i^{\mathrm{Kr}}\right|, \tag{1}$$

where X_i is a TL value for either BellhopCXX or RamSurf at range value r_i for maximum range N metres, and X_i^{Kr} is the corresponding TL value for Kraken. This MAE metric is used as the similarity score. The similarity scores are compared and the model with the lowest score (closer to the arbiter) is chosen to be the most appropriate selection for the environmental parameters, and the data y_i are labelled as either BellhopCXX or RamSurf, as indicated in Figure 3.

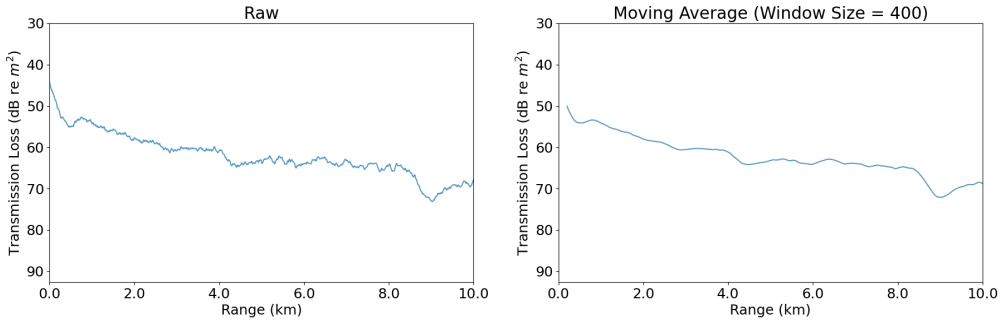


Figure 2: A comparison of a 1D transmission loss array from BellhopCXX before and after applying the smoothing function. The environment used was Shallow, with a maximum depth of 382 m, a frequency of 722 Hz, a maximum range of 10 km, and a constant sound speed of 1511 ms^{-1}.

Four distinct classification models were employed, with each model trained on 70% of the data and evaluated on the remaining 30%, using a random train-test split. In the results included

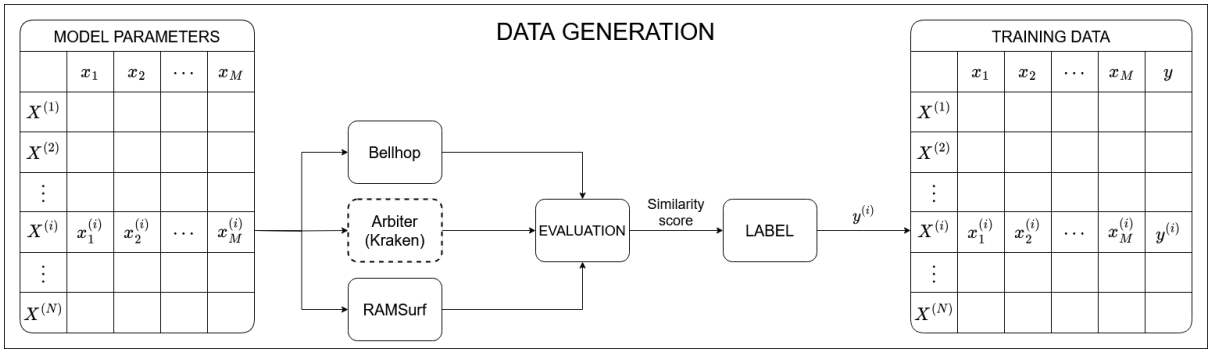


Figure 3: A flowchart showing how environmental parameters are passed through each model to generate training data.

in this paper, 2000 individual realisations, across the four environment types, were used to populate the dataset, with a ratio of 4:7:4:5 for the Shallow, Deep, Coastal and Sloped environment types, respectively. During the data generation and testing process, several different ratio combinations were explored, but this ratio proved to be the most balanced and robust. In future investigations, as additional environment types are included, and more complexity is added to the input parameters, it will be interesting to investigate how this ratio changes. It is important to note that the modelling framework is constructed to support changes to the ratio in a simple manner, through an individual input parameter script.

The ML models were implemented using the scikit learn python library. The same training set was given to each model and they were compared against one another to find which one was the most accurate. A neural network was used initially, however it was deemed unsuitable due to the size of the dataset being too small, but this is something that can be explored in future refinements with application to much larger data generation/collection efforts.

The classifier models evaluated are as follows:

- Random Forest Classifier - An ensemble learning method which builds multiple decision trees, the outputs of which are combined using majority voting to improve accuracy and to control overfitting. It works well with both classification and regression tasks and handles large datasets with higher dimensionality effectively.

- Support Vector Machine - A supervised learning algorithm that finds the optimal hyperplane to separate classes in the feature space. It works well for both linear and non-linear classification, and is particularly effective for high-dimensional spaces.

- Logistic Regression - A linear model for binary classification that estimates the probability of a class using the logistic function. Despite the name, it is a classification algorithm and is widely used due to both its simplicity and efficiency.

- K-Nearest Neighbours - A non-parametric, instance-based algorithm that classifies new data points based on the majority class among its k nearest neighbours in the training data. It is simple and effective, although performance can decrease with high-dimensional data.

3. RESULTS

Table 2 shows the accuracy of the different classifier algorithms in predicting the most accurate model to use, for both a mixed set of environments and per environment. In general, the random forest classifier was the most effective in predicting the correct model, with an accuracy rating of 84.8%, around 4% better than the other three models investigated. Furthermore, the results show the effect that removing the Sloped environment from the dataset has on the overall accuracy of the results. The Sloped environment proved challenging, with the parameter ranges for depth and frequency being restricted due to convergence issues in the solutions for both Kraken and RamSurf. Initially, it was intended to vary the slope angle up to 10°, but numerical instability in the data generation for both RamSurf and the arbiter prevented this. Instead, a parameter study was conducted to amend the ranges of values of depth and frequency for which the Sloped environment produced reasonable results.

The prediction results for this restricted regime appeared to show some bias towards RamSurf, with the possibility that inaccuracies were arising in both models and so RamSurf was often closer to the "truth" than BellhopCXX. The results in Table 2 also indicate that the uncertainty was much higher for this environment than for the other three types. It is important to explore this environment in more detail in future work to understand how the robustness can be improved to match the results for the Shallow, Deep and Coastal environments.

Classifier	**Shallow**	**Deep**	**Coastal**	**Sloped**	**All**
Random Forest	89.1%	90.6%	89.0%	67.9%	84.8%
Support Vector Machine	89.6%	91.9%	76.1%	68.7%	80.6%
Logistic Regression	89.5%	92.3%	74.5%	66.6%	80.6%
K-Nearest Neighbours	89.3%	90.9%	76.8%	65.9%	80.2%

Table 2: The table showing the test accuracy of four different classifiers when trained on datasets made of different environment types

Figure 4 shows the trends in selection for those realisations labelled as more accurate, using BellhopCXX or RamSurf, versus frequency in Hz. In this dataset, the point at which BellhopCXX generally becomes more accurate than RamSurf appears to be around 1.8 kHz, with BellhopCXX being more accurate above that threshold and less accurate below it. This is not unexpected since ray tracing is essentially a high-frequency approximation. This threshold is considerably higher than the 1kHz mentioned in the abstract, as well as the 500Hz threshold defined for high and low frequency in [1]. However, we can see that in many cases, RamSurf obtains a similarly accurate result at frequencies beyond this threshold, and BellhopCXX also performs relatively well at lower frequencies. It should be noted that there are a higher number of low-frequency environments due to the balancing ratio of environment types of 4:7:4:5, and the frequency limits shown in Table 1.

It is also important to note that the selection metric is entirely based on the accuracy of the results compared with the results of the arbiter model, and does not factor in computation time (although these are recorded by the model). Whilst selecting a model based on accuracy is valid and may often be desirable, this is not always the case. A parabolic equation model may provide more accurate results than a ray tracing model in some cases, but the computation time is often significantly longer (particularly for the Deep environment type considered here). Including computation time relative to the other model(s) as a secondary output and using a weighted combination of the two as the selection criteria would be a potentially useful addition to the

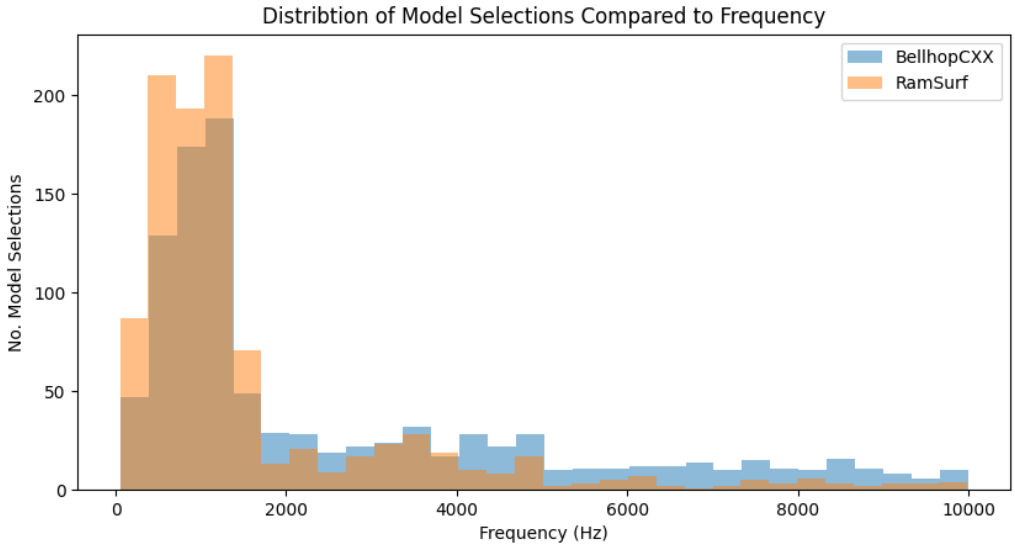


Figure 4: A histogram showing the distribution of model selections against frequency from a dataset of 2000 environments

selection algorithm in future refinements and extensions.

4. CONCLUSION

The model selection algorithm developed here shows very promising results, generating a selection threshold between ray tracing and parabolic equation propagation models at a frequency threshold close to 1.8 kHz for the four generic environment types implemented. This value was obtained for a dataset consisting of 2000 realisations (with 400 for Shallow, 700 for Deep, 400 for Coastal and 500 for Sloped). Changing the size of the dataset, and varying the ratio of the environment types, will lead to different threshold values so it will be important to design ensemble model efforts for specific implementations to obtain statistically more stable results, with reduced uncertainty.

Considering the environment types separately will also produce different threshold values, emphasizing the value of using this ML-based approach, which can be tailored to specific applications. The use of these algorithms provides data-based guidance on multi-parameter-based criteria which may otherwise be too complex to determine manually or with simpler selection processes. The architecture written provides a basis for potential expansion, such as comparing additional models, defining more complex environments and further input parameters, replacing the arbiter for specific applications (for example, Kraken is a good choice as arbiter for environments which can be modelled as waveguides) and including further parameters and metrics to evaluate model performance.

5. ACKNOWLEDGEMENTS

Part of this work was commissioned by Dstl on behalf of the Submarine Delivery Agency as part of the FASTER programme. The authors would like to thank the FASTER team for their support and guidance. The contents include material subject to © Crown copyright (2024), Dstl. This material is licensed under the terms of the Open Government Licence except where otherwise stated. To view this licence, visit `http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3` or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email:
psi@nationalarchives.gov.uk.

REFERENCES

[1] Jensen, Finn B. : "Numerical Models of Sound Propagation in Real Oceans" in *OCEANS 82*, (Washington, D.C, 1982).

[2] Ainslie, Michael, A. : *Principle of Sonar Performance Modelling*, (Springer-Verlag, Berlin, 2010).

[3] Jensen, Finn B., Kuperman, William A., Porter, Michael B., Schmidt, Henrik : *Computational Ocean Acoustics*, (Springer, 2011).

[4] Porter, Michael B. : "Beam tracing for two- and three-dimensional problems in ocean acoustics", *The Journal of the Acoustical Society of America* **146**, 2016-2029 **(2019)**.

[5] Collins, Michael D. : "A split-step Padé solution for the parabolic equation method", *The Journal of the Acoustical Society of America* **93**, 1736-1742 **(1993)**.

[6] Kusel, Elizabeth T., Siderius, Martin : "Comparison of Propagation Models for the Characterization of Sound Pressure Fields", *IEEE Journal of Oceanic Engineering* **44**, 598-610 **(2019)**.

[7] Oliveira, Tiago C. A., Lin, Ying-Tsong, Porter, Michael B. : "Underwater Sound Propagation Modeling in a Complex Shallow Water Environment", *Frontiers in Marine Science* **8 (2021)**.

[8] Pisha, Louis A. et al : "Introducing bellhopcxx/bellhopcuda", *Acoustical Society of America Journal* **153**, A218-A218 **(2023)**.

[9] Paille, Serge S., Levaufre, Arnaud : *RAM PARABOLIC EQUATION*, (`https://github.com/quiet-oceans/ramsurf`).

[10] Hartstra, Iris E., Salomons, Erik, Colin, Mathieu E.G.D., Prior, Mark K. : "ENVIRONMENT DEPENDENT TRANSITION FREQUENCIES FOR RAM AND BELLHOP" in *UACE2019 Conference Proceedings*, edited by John S. Papadakis (Hersonisso, 2019).

[11] Porter, Michael B. : *The KRAKEN normal mode program, NRL/MR/5120-92-6920*, (Naval Research Lab., Washington, DC., 1992).

RAINDROP: A Framework for Real-Time Holistic Underwater Acoustic Maps

Tiago Gomes[1] and Guilherme Beleza Vaz[1]

[1]blueOASIS, Rua do Norte 16, 2655-321 Ericeira, Portugal

Corresponding author: Tiago Gomes, tgomes@blueoasis.pt

Abstract: *RAINDROP is a framework for real-time holistic underwater acoustic maps, targeting applications in acoustic Digital Twins of the Ocean (DTOs). It takes a top-to-bottom approach, starting with input collection and preparation, followed by simulation setup and execution, and finally post-processing and visualization of results. Its framework nature offers a structure that users can use to couple the latest developments at each step of the solution process. All in a performance-oriented approach that includes embedded parallelization to achieve faster than real-time calculations. In the present work, the latest release of RAINDROP is presented and the coupling/automatization of one of its solvers, KRAKEN, based on normal modes theory, is successfuly verified against the 3D penetrable wedge benchmark. Then, RAINDROP is applied in a practical scenario, to assess the impact of shipping noise during the military NATO exercise REPMUS2024 in Tróia and Sesimbra, in Portugal. The environmental inputs are collected from EMODnet and Copernicus Marine Data Store. AIS data from shipping is farmed from aisstream.io and the equivalent acoustic source of each vessel is modeled using the JOMOPANS-ECHO model. A total of 38 days are simulated, focusing on the 63 Hz frequency, achieving calculations up to 25 times faster than real-time. The results are post-processed using the concept of Level of Onset of Biological Effects (LOBE) by the TG Noise under EU's Marine Strategy Framework Directive, demonstrating the potential and capabilities of RAINDROP.*

Keywords: *RAINDROP, Computational Ocean Acoustics, Verification, 3D Wedge, REPMUS2024*

1. INTRODUCTION

Computational ocean acoustics has been the subject of research and development for many decades already. A popular suite of tools concerns the open-source Acoustics Toolbox, distributed under the Ocean Acoustics Library[1] (OALib). It offers a FORTRAN implementation of several acoustic propagation solvers, such as KRAKEN [1] (normal modes theory) and BELLHOP [2] (ray theory). As part of a community effort, the toolbox has advanced significantly

[1]`https://oalib-acoustics.org`

since its inception, including with the addition of MATLAB and Python interfaces. Moreover, new flavors and variations of the solvers have also evolved.

Despite all of this progress, it can be argued that these tools remain dispersed and lack standardization. This leads to steep learning curves that hinder the correct usage of the solvers by less experienced users and limits the complexity of the cases one can develop. Indeed, computational ocean acoustics for practical scenarios is a complex problem, that involves several environmental inputs and approximations that can have a profound impact on the solution and its accuracy.

This consideration is becoming increasingly important as stakeholders from academia, industry and public sector are required to use computational ocean acoustics as part of the solutions to monitor and mitigate excess levels of underwater noise. For example, the European Union's Marine Strategy Framework Directive [3], under Descriptor 11, directs its Member States to keep track and periodically report on this topic following the recommendations of the respective Technical Group (TG). Moreover, the dimension of underwater noise pollution also permeates shipping [4] and other offshore activities, and computational acoustics is becoming indispensable for environmental impact assessment and design of strategies to reduce it. Also, it can be useful to numerically assess how sound propagates underwater, which is important for certain communication and sonar applications, in order to maximize their performance.

Seminal works, such as the EU project JOMOPANS [5], set already some of the foundations for more complex applications of computational ocean acoustics - in this case to model shipping noise. Moreover, in the market, some tools already exist that substantiate this more elaborate and practical side of computational ocean acoustics, such as dbSEA [6] and Qonops [7]. However, these are commercial, closed-source alternatives. While having their own intrinsic value, we feel that the community could greatly benefit from an open-source alternative. Particularly one that facilitates the integration and dissemination of new developments, supports their application to practical scenarios, and promotes the replication of results.

Besides, with the inception of Hydrotwin[2] in 2023, a state-of-the-art acoustic Digital Twin of the Ocean [8], there was a need for software to create the virtual representation of underwater acoustic environments. To our best knowledge, there was no tool that was oriented to such emerging application, which implies a high-level of automation and efficiency. Due to all of these reasons, RAINDROP[3] started its development at around the same time. While still in late stages of development, with a small set of beta testers, it is intended to be eventually made open-source to the community.

2. RAINDROP

RAINDROP is a holistic framework for computational ocean acoustics, oriented for emerging Digital Twins of the Ocean (DTOs) applications, where automatization and efficiency of the virtual models are key. RAINDROP is structured in four major areas that we have identified as critical for its success, which are presented in Figure 1.

The first is related to input data collection and preparation – at the moment, RAINDROP is capable of automatically connecting and processing relevant data from popular open-source databases from the European Union, such as EMODnet[4] (bathymetry) and Copernicus Marine

[2] https://hydrotwin.ai

[3] https://github.com/blueOceanSustainableSolutions/RAINDROP

[4] https://emodnet.ec.europa.eu/en



Figure 1: RAINDROP overview.

Data Store[5] (water temperature and salinity profiles).

The second area is the characterization of acoustic sources, either via pre-defined models or user-provided. RAINDROP already has some embedded models, such as JOMOPANS-ECHO [9] for shipping, and is at the moment finalizing the implementation of in-house models for offshore wind turbines. Nevertheless, the user is also capable of providing its own models/data, either calculated by other numerical tools (e.g. CFD, FEM) or experiments.

The third and most critical area focuses on automating the setup and execution of complex simulations, which starts with the simplification of inputs via a single standard YAML file. RAINDROP is also solver-agnostic, with the goal of offering multiple flavors of acoustic models to take advantage of their unique strengths, and even make them work in tandem. Right now, RAINDROP has integrated (and automated) some of the main open-source solvers of OALib, including KRAKEN, the solver used in this paper - note that the direct usage of these solvers requires a lot of manual set-up for complex cases. There is also on-going work on parallelization, with a successful implementation already available and tested in High-Performance Computing (HPC) platforms.

Finally, RAINDROP also provides modules to systematize and post-process the results, based on best-practices available. One example are the TG Noise recommendations for both continuous [10] and impulsive [11] noise. Some routines are already implemented and can be used in the reporting of underwater noise levels by stakeholders. The results are produced in popular data-exchange formats, such as NetCDF, to facilitate dissemination.

3. WEDGE ANALYTICAL CASE

As part of the on-going Verification&Validation (V&V) of RAINDROP, we recover a popular case in the literature with analytical solution, the 3D penetrable wedge. We adopt the analytical solution obtained in [12], based on the adiabatic mode parabolic equations - valid for small apex angles. It is compared with the numerical simulation by RAINDROP using KRAKEN3D as the solver. A schematic of the 3D wedge and its characteristics are presented in Figure 2 and Table 1, respectively.

The wedge's bathymetry is constructed in RAINDROP as a regular 100 by 100 point grid, in which the normal modes are solved. The maximum phase speed, c_{max}, is set to 2500 m/s, and the vertical discretization set to automatic in KRAKEN3D. In order to obtain the complex pressure field based on the normal modes, RAINDROP also executes FIELD3D on the elements of a Delaunay triangulation of the former grid. One of the options is compared: a standard (STD) run of FIELD3D, based on a series of 2D simulations over the bearing direction; or a gaussian beam tracing (GBT) one, which as the name indicates additionally relies on a series of rays emitted to incorporate horizontal refraction effects, thus better capturing the 3D nature of the

[5] `https://data.marine.copernicus.eu/products`

solution. A comparison of the analytical solution against RAINDROP is presented in Figures 3 and 4.

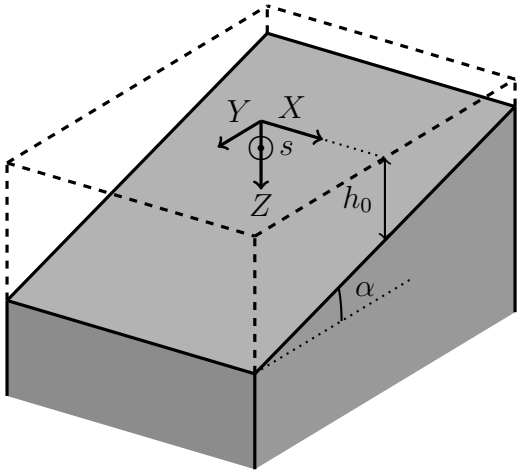


Figure 2: Schematic of 3D wedge case.

Table 1: Variables of 3D wedge case.

	Description	**Value**
f	Source Frequency	50 Hz
z_s	Source Depth	10 m
α	Bottom Slope	5 deg
h_0	Bottom Depth	90 m
c_w	Speed Sound Water	1500 m/s
c_b	Speed Sound Bottom	2000 m/s
ρ_w	Density Water	1 g/cm^3
ρ_b	Density Bottom	2 g/cm^3
α_b	Attenuation Bottom	0.5 dB/λ

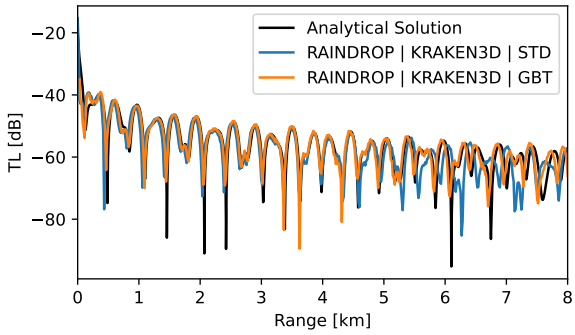


Figure 3: Transmission loss over XZ plane @ $z = 10$ m.

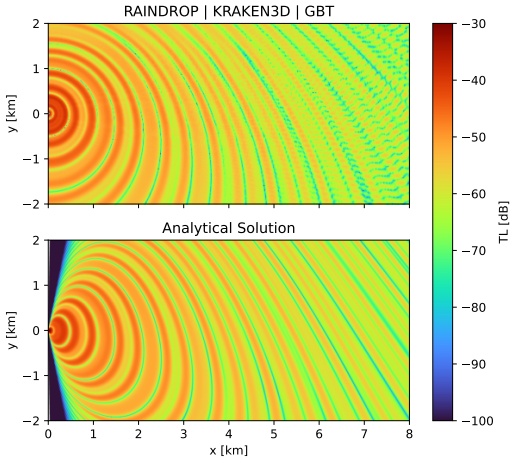


Figure 4: Transmission loss over XY plane @ $z = 10$ m.

It is evident that the STD run fails to capture some of the 3D effects present in this benchmark case. On the other hand, the selection of GBT significantly improves this, and the agreement of both solutions is quite satisfactory. Still, the numerical solution of RAINDROP starts to slowly deviate due to small phase differences that build up. This effect has also been observed by the authors in [13] when applying KRAKEN3D directly to this same 3D wedge case. Thus, with this small exercise, one can verify that the coupling and automation of KRAKEN3D within RAINDROP is correctly implemented.

4. REPMUS2024

The Robotic Experimentation and Prototyping with Maritime Unmanned Systems (REPMUS) is organized annually by the Portuguese Navy, taking place in Tróia and Sesimbra, in

Portugal. It counts with the participation of several NATO countries, and involves a significant number of maritime operations. In the 2024 edition blueOASIS participated for the second time with RAINDROP, which can be used not only to assess the environmental impact of the exercise, but also to estimate how sound propagates and its levels, which is essential for operations that require underwater acoustics for communications or localization - all in real-time.

The exercise had a duration of 19 days, between 09/09/2024 and 27/09/2024. The baseline scenario used to assess the impact consists on the 19 days immediately before, from 19/08/2024 to 06/09/2024. RAINDROP was the tool used to prepare the inputs, execute the simulation and post-process the results. A model of the region was created using bathymetry information from EMODnet: DTM F3 Tile [14] (resolution of $1/16^{th}$ arc-minutes) and some other high-resolution data products available along the coastline (resolution of $1/128^{th}$ arc-minutes) and estuary (resolution of $1/64^{th}$ arc-minutes) by Instituto Hidrográfico. The domain spans from 9.30º W to 8.70º W in the longitude direction (∼53 km) and 38.00º N to 38.55º N in the latitude direction (∼62 km). The final bathymetry can be visualized in Figure 5, whereas in Figure 6 the respective Delaunay triangulation, including the refinement zone in the Sado estuary mouth, can be visualized.

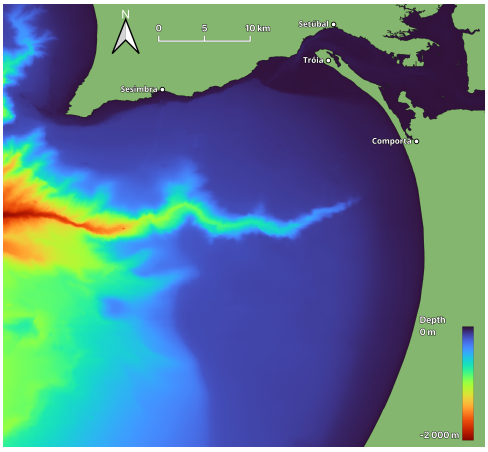


Figure 5: Input bathymetry.

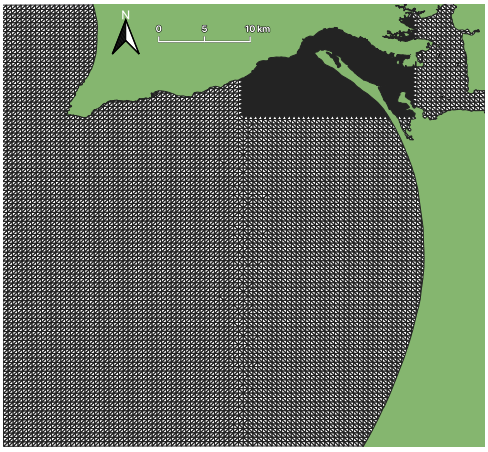


Figure 6: Delaunay triangulation.

While the Sound Speed Profile (SSP) is known to have a considerable impact on how sound propagates, in the present work it is assumed to be constant both spatially and temporally - overall, the SSP shape did not change significantly based on a preliminary analysis performed. The daily SSP was extracted from the Atlantic - Iberian Biscay Irish - Ocean Physics Analysis and Forecast dataset [15] from the Copernicus Marine Data Store at one of the deepest points of the domain, around (38.288º N, 9.300ºW), and averaged over the 38 days under analysis into the SSP in Figure 7.

As for the composition of the seabed, it is quite heterogeneous based on available data by Instituto Hidrográfico under the SEPLAT [16] program. Yet, most of the region of interest is of sand type. Therefore, it was assumed an uniform sandy seafloor, based on the properties described in [17].

Another key input is the AIS data, which was farmed from aisstream.io[6] and stored in a database. It was cleaned *a posteriori* for typical anomalous situations in AIS messages, such as extreme speeds and lengths, or even vessels with nonexistent MMSI which couldn't be objectively characterized. In Figure 8 there is a density map of ship locations via AIS messages

[6] `https://aisstream.io`

before (199 unique vessels) and during (226 unique vessels) REPMUS2024 - it is evident an increase in traffic near Sesimbra during the exercise.

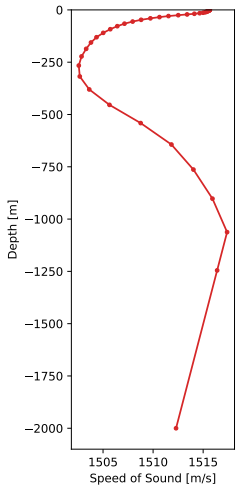


Figure 7: Sound Speed Profile.

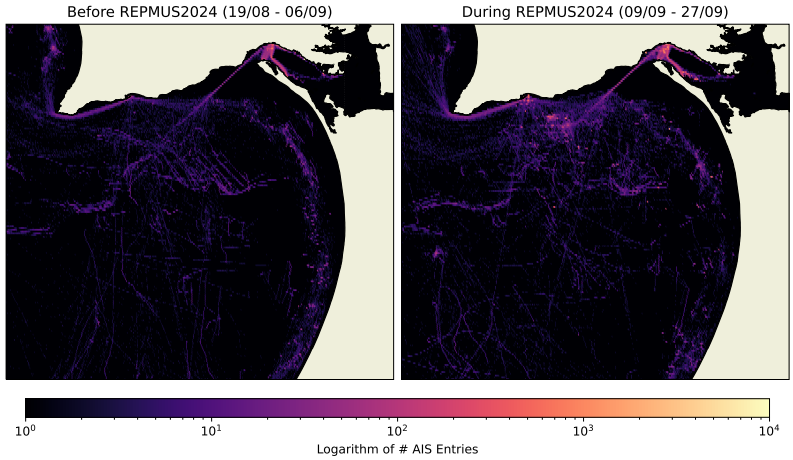


Figure 8: AIS density map.

Once again, RAINDROP used the KRAKEN3D solver, based on the theory of normal modes, to solve the underwater acoustic propagation. Some of the relevant settings are presented in Table 2. For increased performance, FIELD3D was executed with the STD option. Based on the previous 3D wedge case, we know that this option led to some phase differences in relation to the analytical solution. However, the average Transmission Loss values were not significantly impacted. Considering that the end-goal of this case is to analyze Sound Pressure Level (SPL) trends over large periods of time, it is assumed that such phase difference will be averaged out, with minimal impact.

All the configuration is done via a YAML file, and only a few commands are required to start RAINDROP, which fully automatizes the simulation setup and execution. First, the normal modes are obtained using KRAKEN3D, assuming that the environmental conditions do not change. Then, FIELD3D is used to obtain the pressure field of each source/vessel for every timestep. This information is aggregated in periods of 5 minutes, to minimize source flickering when AIS messages are sent at a slow rate. The Sound Level of each vessel is characterized using the JOMOPANS-ECHO model [9] and the respective AIS information. Finally, the acoustic pressure field obtained for each individual source is summed and the resulting field transformed to SPL. Note that it was necessary to assume a value for background ambient noise, since only shipping was modeled: 80 dB re 1μPa, in line with hydrophone data collected around this region in [18].

An example timeframe of one of the simulations is presented in Figure 9. Note that more than 10000 individual frames were calculated to simulate the period before and during REPMUS2024. Taking advantage of RAINDROP's parallelization, a frame was calculated every 12 seconds using 24-cores in a workstation equiped with an AMD Ryzen Threadripper PRO 3975WX 32-cores @ 4.1 GHz - 25 times faster than real-time.

Leveraging some of the post-processing routines implemented in RAINDROP, some of which inspired by the recommendations of TG Noise for continuous noise [10], the concept of Level of Onset of Biologically Adverse Effects (LOBE) is applied. For this exercise, the domain is subdivided into arbitrary habitats/cells of 2 km by 2 km each. For each its computed the percentage of time in which a certain LOBE threshold was exceeded - arbitrarily set to 90 dB re 1μPa in this case. In Figures 10 and 11, the percentage of time the LOBE was exceeded before

and during REPMUS is presented. The difference between both is presented in Figure 12. It is noticeable an increase of this LOBE exceedance metric between the mouth of the estuary, the coast of Sesimbra and far offshore, in part consistent with the higher density of AIS messages received in some of the areas. On the other end, along Portugal's southern coast until Sines, it is actually observed a reduction in LOBE exceedance. However, all differences are small (<5 %) and even the global value of LOBE exceedance is never above 25%, which can be considered an acceptable value.

On the other hand, these differences can still be under the uncertainty introduced by the simplifications assumed. It is also important to bear in mind that REPMUS involves the operation of many other unmanned vehicles, both surface and underwater, which are not taken into account (no AIS) despite contributing to the soundscape. Also, higher-frequencies may be more critical, both in terms of emissions by the vessels in the area and impact on local populations (e.g. bottlenose dolphins).

All in all, a more in-depth analysis is necessary in order to extract more precise conclusions. This includes the improvement of the temporal/spatial accuracy of some of the inputs, like the SSP and seabed properties. Moreover, modeling other acoustic sources, like wind interaction with the free-surface, is expected to reduce the number of approximations and further increase

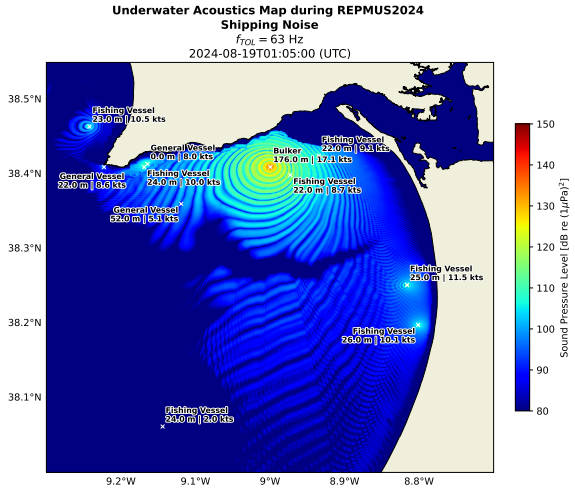


Figure 9: Example acoustic map.

Table 2: Variables of REPMUS2024 case.

	Description	Value
f	Source Frequency	63 Hz
z_s	Source Depth	6 m
c_b	Speed Sound Bottom	1650 m/s
ρ_b	Density Bottom	1.9 g/cm^3
α_b	Attenuation Bottom	0.8 dB/λ
N	Discretization Points	16650
Δt	Timestep	5 min

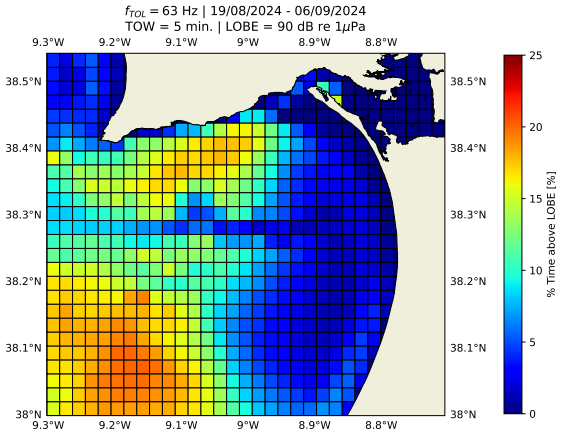


Figure 10: LOBE exceedance: before REPMUS2024.

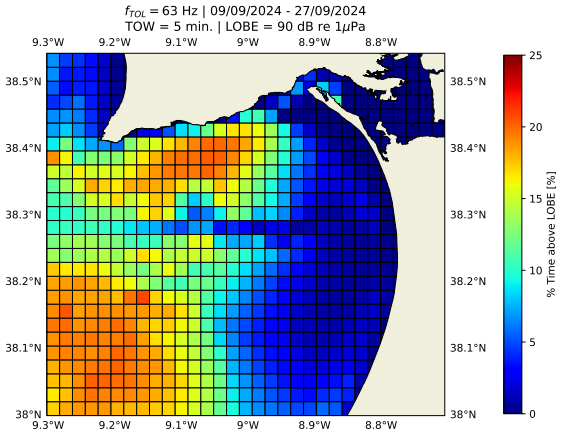


Figure 11: LOBE exceedance: during REPMUS2024.

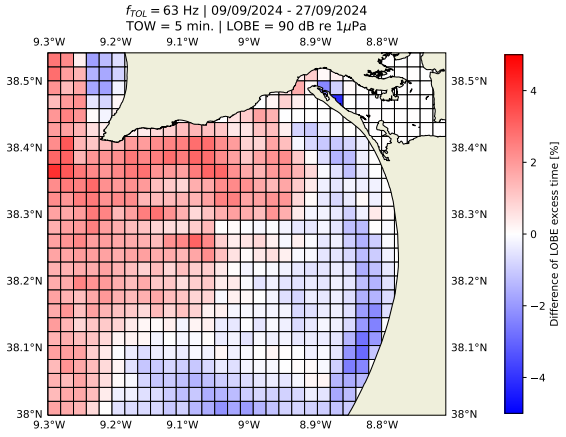


Figure 12: Variation of LOBE exceedance before and during REPMUS2024.

the fidelity of the simulation. Also, a sensitivity and uncertainty analysis could help understand better the relative impact of each component in the final result. This, together with validation against hydrophone data collected during the exercise by blueOASIS, is left as future work.

5. CONCLUSIONS

In this paper, we have presented RAINDROP, a framework for holistic real-time underwater acoustic maps that targets DTO applications. The coupling and automatization of one of its acoustic solvers, KRAKEN3D, is successfully verified using the popular benchmark case of the 3D penetrable wedge. Then, RAINDROP is applied to a practical case, to assess the environmental impact of REPMUS2024 exercise in the region of Tróia and Sesimbra, in Portugal. The capabilities of RAINDROP to collect and processs the input data are demonstrated, and a span of 38 days are simulated, up to 25 times faster than real-time, taking advantage of its parallelization. In the end, a simplified analysis inspired by the metrics set by the TG Noise for continuous noise [10] is presented as proof-of-concept. Assuming a threshold of 90 dB re 1μPa, only a slight increase in LOBE exceedance was obtained during the REPMUS2024 exercise, up to 5% more. Still, globally this threshold is never exceeded more than 25% of the time, so the impact can be considered minimal in the 63 Hz frequency band.

As future work, it is planned a more in-depth V&V study of RAINDROP, with sensitivity analysis and uncertainty quantification, taking into consideration more detailed environmental inputs and effects. Data collected by Hydrotwin's network of smart acoustic recorders in Azores will be leveraged to support this effort. During this time, RAINDROP will continue its development, targeting its open-source release.

6. ACKNOWLEDGEMENTS

This publication has received funding from the European Union's Horizon Research and Innovation Action through the project FLOATFARM (Grant Agreement No. 101136091). The authors acknowledge the support of the Portuguese Navy during the participation on REPMUS2024.

REFERENCES

[1] M. B. Porter, "The KRAKEN normal mode program," Tech. Rep. NRL/MR/5120-92-6920, Naval Research Laboratory, Washington, D.C., May 1992.

[2] M. B. Porter and H. P. Bucker, "Gaussian beam tracing for computing ocean acoustic fields," *The Journal of the Acoustical Society of America*, vol. 82, pp. 1349–1359, Oct. 1987.

[3] "Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive)." Official Journal of the European Union L164/19, June 2008.

[4] International Maritime Organization, "Revised guidelines for the reduction of underwater radiated noise from shipping to address adverse impacts on marine life." MEPC.1/Circ.906, Aug. 2023.

[5] C. de Jong, B. Binnerts, S. Robinson, and L. Wang, "Guidelines for modelling ocean ambient noise," Report of the EU INTERREG Joint Monitoring Programme for Ambient Noise North Sea (JOMOPANS), June 2021.

[6] R. S. Pedersen and M. Keane, "Validation of dBSea, underwater noise prediction software. Pile driving focus," *Journal of Shipping and Ocean Engineering*. In Press/Accepted.

[7] M. Ghezzo, A. Petrizzo, F. Madricardo, T. Folegot, R. Gallou, D. Clorennec, R. Chavanne, E. Hemon, C. Ferrarin, H. Mihanović, K. Pikelj, M. Bastianini, A. Pari, S. Pari, S. Menegon, W. J. McKiver, G. Farella, S. Bosi, A. Barbanti, and M. Picciulin, "Natural and shipping underwater sound distribution in the Northern Adriatic Sea basin and possible application on target areas," *Marine Pollution Bulletin*, vol. 207, p. 116852, Oct. 2024.

[8] T. Gomes, A. Santos, J. Vanneste, A. van Toor, C. Ramos, D. Aromatario, R. Carrico, E. Cruz, and G. Vaz, "Hydrotwin: Capturing the Heartbeat of the Ocean," in *Proceedings of the 2025 7th Experiment@ International Conference (expat'25)*, (Horta, Faial Island, Portugal), Sept. 2025. Accepted, to appear.

[9] A. MacGillivray and C. De Jong, "A reference spectrum model for estimating source levels of marine shipping based on automated identification system data," *Journal of Marine Science and Engineering*, vol. 9, p. 369, Mar. 2021.

[10] J. Borsani, M. Andersson, M. André, A. Azzellino, M. Bou, M. Castellote, L. Ceyrac, D. Dellong, T. Folegot, D. Hedgeland, C. Juretzek, A. Klauson, R. Leaper, F. Le Courtois, A. Liebschner, A. Maglio, A. Mueller, A. Norro, A. Novellino, O. Outinen, A. Popit, A. Prospathopoulos, P. Sigray, F. Thomsen, J. Tougaard, P. Vukadin, and L. Weilgart, "Setting EU threshold values for continuous underwater sound.," Tech. Rep., LU, 2023.

[11] P. Sigray, M. Andersson, M. André, A. Azzellino, J. F. Borsani, M. Bou, and M. Castellote, "Setting EU threshold values for impulsive underwater sound.," Tech. Rep., LU, 2023.

[12] P. S. Petrov and F. Sturm, "An explicit analytical solution for sound propagation in a three-dimensional penetrable wedge with small apex angle," *The Journal of the Acoustical Society of America*, vol. 139, pp. 1343–1352, Mar. 2016.

[13] O. C. Rodriguez, F. Sturm, P. Petrov, and M. Porter, "Three-dimensional model benchmarking for cross-slope wedge propagation," *Proceedings of Meetings on Acoustics*, vol. 30, p. 070004, June 2017. Publisher: Acoustical Society of America.

[14] EMODnet Bathymetry Consortium, "EMODnet Digital Bathymetry (DTM 2024)," 2024. DOI: 10.12770/CF51DF64-56F9-4A99-B1AA-36B8D7B743A1 (Accessed on 05-2025).

[15] E.U. Copernicus Marine Service Information (CMEMS), "Atlantic-Iberian Biscay Irish-ocean physics analysis and forecast." Marine Data Store (MDS), 2017. DOI: 10.48670/MOI-00027 (Accessed on 05-2025).

[16] I. Hidrográfico, "Cobertura dos Sedimentos Superficiais da Plataforma Continental Portuguesa (Corresponde às cartas SEPLAT)," 2025. DOI: 10.71683/f00f3416-2904-47ad-81c1-c51f36449404 (Accessed on 05-2025).

[17] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*. New York, NY: Springer New York, 2011.

[18] E. M. Cruz, *Caracterização do ruído subaquático produzido pelo tráfego marítimo no estuário do Sado e potenciais impactos sobre a população residente de tursiops truncatus (Montagu, 1821)*. Master of Science, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal, 2012.

Measurements of sea-floor backscattering strength of very low frequency acoustic waves in western Mediterranean Sea; a comparison to GABIM [1]

Iannis Bennaceur[1], Nathan Ivkovic[1], Xavier Cristol[1], Adrien Rondeau[1], and Benoit Theckes[1]

[1]Thales Defence Mission Systems, FRANCE

Iannis Bennaceur, 525 route des dolines, 06560 Valbonne, FRANCE
iannis.bennaceur@fr.thalesgroup.com, +33492963291

Abstract: *In the context of anti-submarine warfare, active sonar performances depend on the amount of reverberated energy reaching the receiver after the powerful emission from the source. In the Very Low Frequency (VLF) range, the scattering by the sea-floor provides the largest contribution to the reverberation level that THALES DMS tries to better predict. An at-sea trial took place in the Mediterranean Sea in September* 2024*,* 30 *km off French Riviera coast in a deep-water environment. Two boats were involved, one carrying an acoustic source emitting at* 215 *Hz, immersed at* ≈ 7 *m under the sea-surface, the other conveying a vertical line array composed of* 13 *hydrophones spaced by* 3 *m. The center of the antenna was* 45 *meters below the sea-surface. The boats were spaced from few dozen meters, and* 1*-second Continuous Waves (CW) were transmitted in order to measure the acoustic Reverberated Level (RL) in this quasi-monostatic configuration. The output of classical weighted beamforming according to time and elevation angle allowed to extract RL according to the grazing angle* θ *in a range between* 25 *and* $65°$*. The evolution of the backscattering strength* σ*, deduced from RL, regarding* θ *is in good qualitative agreement with other measures that also took place in the Mediterranean Sea [2], where a strong decrease appears from* $\theta \approx 33°$*. However, our implementation of GABIM [1], ran with a first-guess geoacoustic model for the sea-floor, indicates that scattering from the volume inhomogeneities is the main contribution to backscattering strength, regardless of* θ*. This observation differs from the one made in [2] where scattering by the rough interface between sediments and the harder basement was the main contribution for intermediate grazing angles.*

Keywords: *Sea-floor reverberation, backscattering strength, very low frequency, GABIM*

1. ENVIRONMENT AND REVERBERATED LEVEL

During the trial (see abstract for details), the water temperature from the sea-surface to a depth of $750\,\mathrm{m}$ was measured using a XBT probe. The resulting celerity profile, extended to the water height ($H_w \approx 2388\,\mathrm{m}$) and displayed on the left part of Fig. 1, exhibits a strong thermocline. To construct a geoacoustic model for the sea-floor (right part of Fig. 1), needed for the σ computation, sites 371 and 374 from the Deep Sea Drilling Project were considered (site 372 was excluded because it took place over a rise). The Plio-Quaternary (P-Q) sediment thicknesses estimated from the Seanoe [3] and an in-house databases were found in agreement with the observations of the DSDP, which gave us confidence about using those databases to infer the sediment thickness at the experiment location. A 800-m thick P-Q sediment layer was estimated, that we decided to split into 2 parts; an unconsolidated layer of marls (or sandy muds) of thickness $h = 100\,\mathrm{m}$, and a semi-infinite layer of consolidated sediments. We choose as superficial properties for those marls a porosity of 0.7, a compressional velocity ratio c_p/c_w, with c_w the water sound speed just above the sediment, of 1.013, and a compressional attenuation k_p in dB/m/kHz of 0.102. The evolution of celerity, attenuation and density ρ regarding depth up to h are computed using Hamilton's work. The celerity and density in the consolidated P-Q layer (dots on Fig. 1), comes from the CRUST1.0 model [4]. All the layers are considered fluid (no shear effect). Weighted (Hann's window in time and Hamming's in space) overlapped

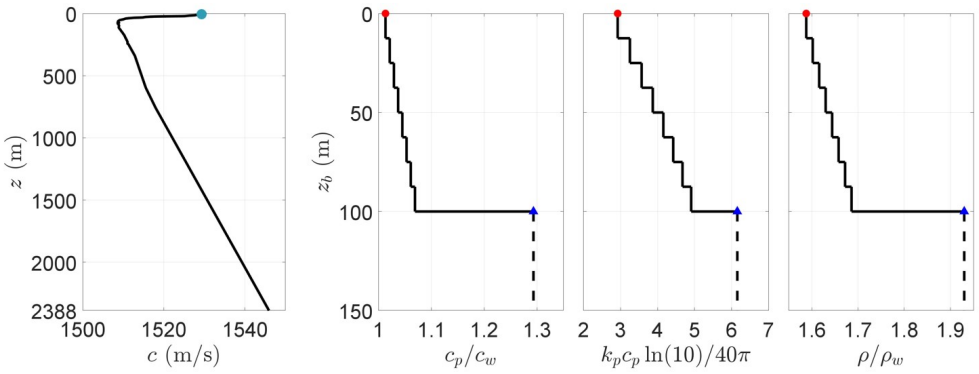


Figure 1: ***Left:*** *Sound-speed volcity profile in water,* ***right:*** *evolution regarding depth of c_p/c_w, normalized k_p (dB/m/kHz), and ρ/ρ_w.*

($50\,\%$) classical beamforming according to time and elevation angle ϕ is performed over the 20 1-s CW emissions. The rapid variation in time of the reverberated field direction of arrival requires short FFT. We chose to perform FFT over 1024 samples (a quarter of the sampling frequency) leading a frequency resolution of $4\,\mathrm{Hz}$. Hence, a 2D cubic interpolation in the (f, ϕ) domain for each temporal bloc is performed to measure RL in dB re $1\,\mu\mathrm{Pa}/\sqrt{\mathrm{Hz}}$ at $215\,\mathrm{Hz}$. Simulation indicates that this operation leads to a bias that do not exceed $+0.25\,\mathrm{dB}$. The 20 outputs of beamforming are then averaged (standard deviations due to ping-to-ping variability around $5\,\mathrm{dB}$) and RL is extracted (white dots on Fig. 2).

2. GABIM IMPLEMENTATION

In the process of implementing GABIM [1], one needs to extract from OASR (one of the software included in the Ocean Acoustics and Seismic Exploration Synthesis (OASES) of H.

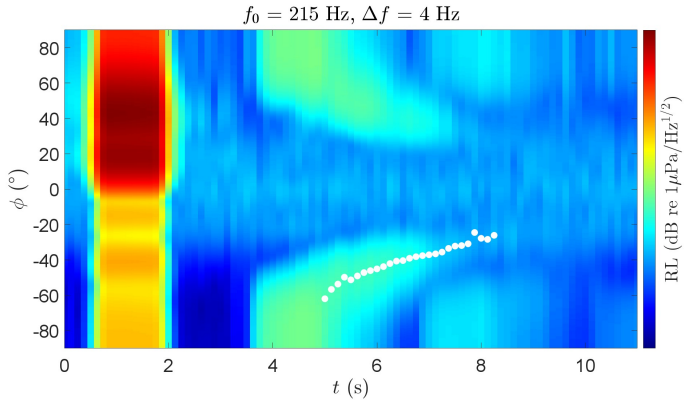


Figure 2: Averaged beamforming output on which the reverberated level is extracted (dots).

Schmidt) the up-going and down-going waves complex amplitudes, A and B respectively, in each layer. We validate this process by successfully comparing our implementation to Fig. 5 of [1] on left part of Fig. 3 that represents the scattering from a very rough rock basement under a smooth sand layer. A second verification consisted in plotting Fig. 11 of [1] with the parameters given in the corrected addendum [5]. The right upper part of Fig. 3 shows strong, unexplained, oscillations for the water-sediment interface roughness and the volume contributions to backscattering strength. A way to reduce the oscillations regarding the water-sediment interface roughness is to consider Eq. (42) of [6] (valid when sediments velocity and density are homogeneous (no gradients)) instead of Eq. (32) of [1]. To reduce the volume contribution oscillations, we have split Eq. (43) of [1] into oscillating-terms proportional to BA^* or to AB^*, with $\cdot^*$ the complex conjugation, and non oscillating-terms proportional to $|A|$ and $|B|$, and only consider the latter for the scattering strength computation. The right lower part of Fig. 3 shows that the oscillations disappeared at the cost of a 5 dB discrepancy in the total scattering strength, for some θ range below 35°. Nevertheless, the overall agreement is reasonable for the use of our implementation of GABIM.

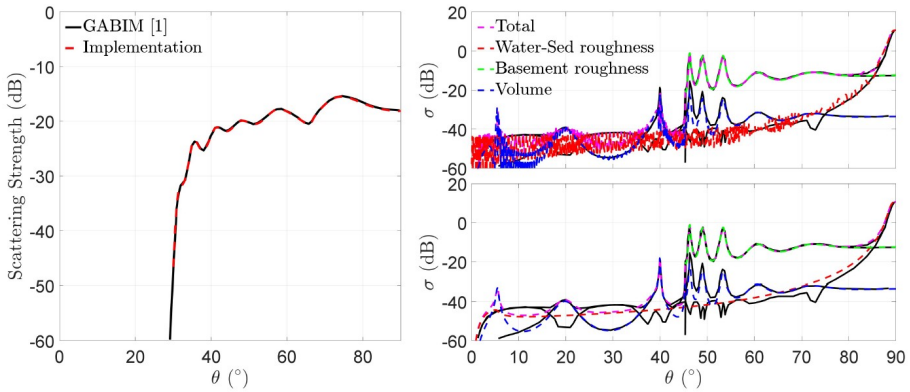


Figure 3: ***Left:*** *Fig. 5 of [1],* ***right:*** *Fig. 1 of [5], without (upper part) and with (lower part) oscillations corrections.*

3. BACKSCATTERING STRENGTH COMPUTATION

The backscattering strength $\sigma(t)$ in dB according to the time t of the RL measurement is computed following:

$$\sigma(t) = \mathrm{RL}(t) - \mathrm{SL} - 10\log_{10}\left(2\pi \int_{R_{\min}(t)}^{R_{\max}(t)} r\mathrm{d}r\mathrm{TL}_g(r)\mathrm{TL}_r(r)D_a(\phi(r),\theta_m(t))\right), \quad (1)$$

where RL is the reverberated level extracted from Fig. 2, SL the source level at 215 Hz in dB re $1\,\mu$Pa/$\sqrt{\text{Hz}}$ @1 m, TL_g and TL_r the transmission losses (re 1 m) from the source to the ensonified area and from the area to the antenna, respectively, and D_a the antenna directivity pattern. $R_{\min}$ and $R_{\max}$ correspond to the inferior and superior radial integration limits, computed numerically with a ray-tracing code, defined such that the go and return time travel $t_0(r)$ is in between $t - \tau/2$ and $t + \tau/2$ where $\tau = 1$ s the CW pulse duration. As the measure of RL at t is the sum of the contributions from all area elements in between $R_{\min}$ and $R_{\max}$, that have been ensonified under a grazing angle from $\theta_{\min}$ to $\theta_{\max}$, one choose to represent the evolution of σ according to a "mean" grazing angle defined as $\theta_m = 0.5(\theta_{\min} + \theta_{\max})$. In Eq. (1), the antenna directivity pattern is taken into account as followed: while pointing in the θ_m direction, the contribution of all area elements is weighted by the directivity pattern evaluated at the elevation angle of arrival ϕ on the antenna, different for each elements. The 2π factor indicates that we already performed the azimuthal integration allowed by the quasi-monostatic configuration of the trial. Finally, all the terms inside the integral in Eq. (1) are expressed on a linear scale.

As the source is immersed near the sea-surface ($\approx$ 7 m) and the emitted wave-length is close ($\approx$ 7.1 m), one needs to be careful when computing the incident pressure field (or TL_g in Eq. (1)) on the ensonified areas. In such configuration, the summation of the complex pressure field (intensity plus phase) associated with each ray should be favored. This produces an emission directivity pattern ("Llyod mirror") that has a strong influence on the incident pressure field at the sea-floor (see left image of Fig. 4). In contrast, the incoherent summation of intensity only produces a incident pressure field on the sea-floor that decays slowly with respect to r (right image of Fig. 4). The computation of σ following Eq. (1) is realized with both $\mathrm{TL}_g^{(C)}$ and

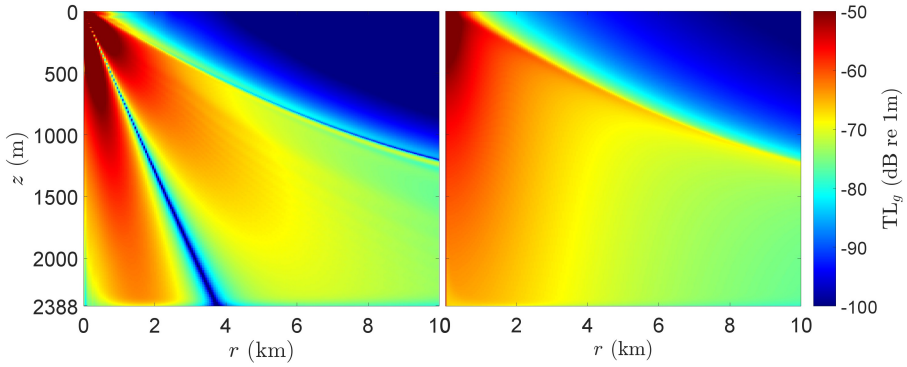


Figure 4: ***Left:*** *Coherent pressure field,* ***right:*** *incoherent pressure field.*

$\mathrm{TL}_g^{(I)}$, the transmission losses from the source to the sea-floor obtained from the coherent and incoherent summation, respectively. The scattering strengths are displayed, along with the measurements of [2], against θ_m on Fig. 5 (left part). Using $\mathrm{TL}_g^{(C)}$ for the calculation of σ produces a strong bump for $\theta_m \approx 32° \approx \tan^{-1}(2388/3800)$ due to the zero in the emission directivity

pattern, whereas σ based on $\mathrm{TL}_g^{(I)}$ is in a better qualitative agreement with past observations. This bump is typically an artifact arising from the classical assumption in Eq. (1) the scattering process occurs at the water-sediment interface [2]. The incident field on the sediment volume and on the sub-bottom interface should be considered instead. Regarding σ based on $\mathrm{TL}_g^{(I)}$, if the SL was not high enough to address the grazing angles lower than $25°$, a strong decrease appears from $\theta_m \approx 33°$. The right part of Fig. 5 represents the three contributions to the total scattering strength issued from our implementation of GABIM calculated with the geoacoustic model described by Fig. 1 (the basement being the sub-bottom interface between the unconsolidated and the consolidated sediments) and the interpolation of the parameters σ_2 and w_2 found in Table I of [1] with respect to $\nu_p = c_p/c_w$. The parameters σ_2 and w_2 are the dimensionless isotropic volume backscattering cross section and the spectral strength of interface roughness, respectively. One sees that both the contributions to σ of the water-sediment and the sub-bottom interfaces are negligible compared to the volume one. The left part of Fig. 6 shows that our im-

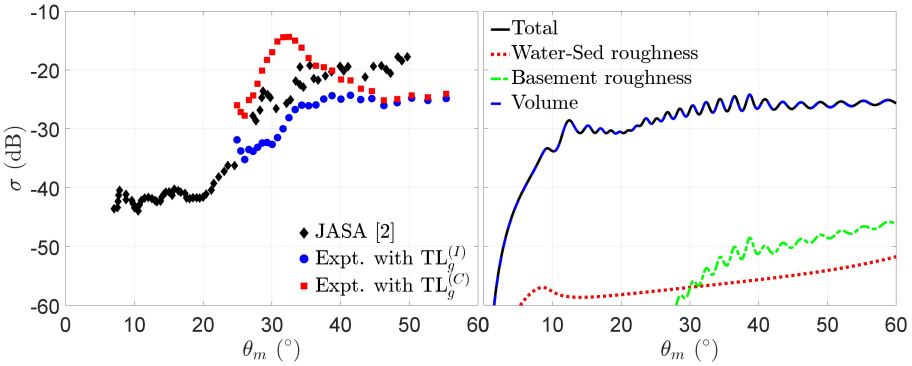


Figure 5: ***Left:*** *Measured backscattering strengths,* ***right:*** *detailed scattering contributions issued from our implementation of GABIM.*

plementation of GABIM, computed with σ_2^G issued from Table I of [1] decently agrees with the measured backscattering strength. A better agreement is achievable if one plays with the evolution of σ_2 w.r.t. z_b. The right part of Fig. 6 shows the volume contribution computed with $\sigma_2(z_b) = \sigma_2^G/4$ for $z_b < 75\,\mathrm{m}$ and $\sigma_2(z_b) = 2\sigma_2^G$ for $z_b \geq 75$. Such an evolution of σ_2 could physically account for upper sediment layers (from $0 < z_b < 75$ meters) less inhomogeneous and deeper layers ($z_b \geq 75$ meters) including turbidites for example. Hence, the volume contribution to backscattering strength of GABIM may account for the strong decrease of σ w.r.t. θ. The significant variation of the modeled σ regarding σ_2 shows that a good agreement between measures and modeling requires a rightful geoacoustic model and a judicious choice for the parameters describing random variations. Nevertheless, the decent agreement obtained on Fig. 6 between measures and modeling based on a first-guess geoacoustic model (Fig. 1) and on the typical values given by Table I of [1] make us confident about using our implementation of GABIM for estimating reverberation level for very low frequency active sonar performances.

4. CONCLUSION

In this paper, a bottom backscattering strength measurement on several 1-s CW at $215\,\mathrm{Hz}$ in the western Mediterranean Sea is described. The acoustic source (immersed at $\approx 7\,\mathrm{m}$) being in the vicinity of the sea-surface, coherent and incoherent incident pressure fields may be

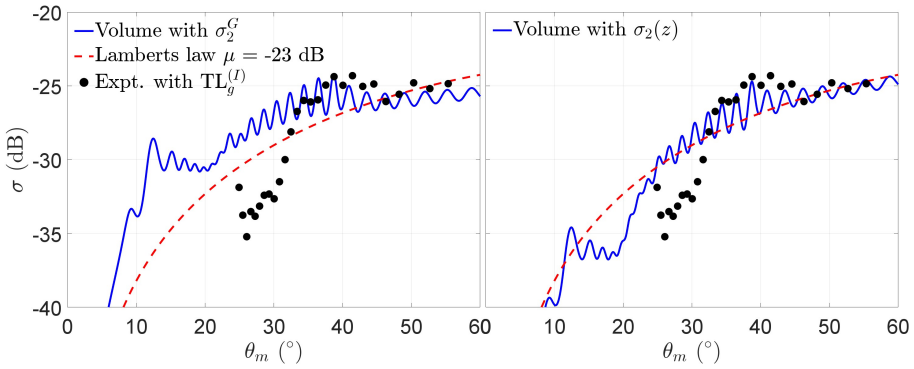


Figure 6: Scattering strength issued from our implementation of GABIM for ***left:*** $\sigma_2 = \sigma_2^G$, ***right:*** $\sigma_2(z)$, *compared to the measured backscattering.*

considered for the measure of σ. While the zero in the emission beam pattern ("Llyod mirror effect") of the coherent field induces a strong bump (artifact according to [2]) in the backscattering strength, the computation of σ upon the incoherent field is in good qualitative agreement with past measurement. Our implementation of GABIM [1], used with a first-guess geoacoustic model and the default parameters describing random variations provided in [1], shows that the isotropic volume inhomogeneities are the main contribution to the overall backscattering strength and allowed decent agreement with the measurements. Moreover, a variation of the volume backscattering strength w.r.t. depth (e.g. to account for deep turbidites layers) may retrieve the strong decrease of the total backscattering strength w.r.t. to the grazing angle.

REFERENCES

[1] D. R. Jackson, R. I. Odom, M. I. Boyd, A. N. Ivakin : "A Geoacoustic Bottom Interaction Model (GABIM)", *IEEE J. Oceanic Eng.* **35**, 603-617, **(2010)**.

[2] C. W. Holland, P. Neumann : "Sub-bottom scattering: A modeling approach", *J. Acoust. Soc. Am.* **104**, 1363-1373, **(1998)**.

[3] M. Bellucci, et al. : "A comprehensive and updated compilation of the seismic stratigraphy markers in the Western Mediterranean Sea", *SEANOE*, **(2021)**.

[4] G. Laske, G. Masters, Z. Ma, M. Pasyanos : "Update on CRUST1.0 - A 1-degree Global Model of Earth's Crust", *Geophys. Res. Abstracts* **15**, Abstract EGU2013-2658, **(2013)**.

[5] D. R. Jackson, R. I. Odom, M. I. Boyd, A. N. Ivakin : "Corrections to "A Geoacoustic Bottom Interaction Model (GABIM)"", *IEEE J. Oceanic Eng.* **36**, 373, **(2011)**.

[6] J. E. Moe, D. R. Jackson : "First-order perturbation solution for rough surface scattering cross section including the effects of gradients", *J. Acoust. Soc. Am.* **96**, 1748-1754, **(1994)**.

TIME-FREQUENCY STRUCTURE OF NORMAL MODES IN RANGE-DEPENDENT WAVEGUIDES: NUMERICAL MODELING AND FIELD EXPERIMENT

Danila Sidorov[a], Devi Bodjona[a], Dmitrii Kovzel[b], Vladimir Bezotvetnykh[b], Andrey Lunkov[a]

[a]Prokhorov General Physics Institute of the Russian Academy of Sciences, 119991, Russia
[b]Ilichev Pacific Oceanological Institute, Far Eastern Branch of the Russian Academy of Sciences, 690041, Russia

Danila Sidorov. Prokhorov General Physics Institute of the Russian Academy of Sciences, 38 Vavilov St., 119991 Moscow, Russia. Fax: +7 499 503 8723.
Email: sidorov@kapella.gpi.ru

Abstract: *Numerical simulation is applied to study a broadband acoustic field formed by a point source at a frequency of 200 to 2000 Hz in range-dependent waveguides with a seasonal thermocline. Two shallow-water waveguide models with varying bathymetry are considered: 1) with a smooth bottom depth variation, 2) with a strong bathymetric feature (barrier). Normal mode theory and parabolic equation simulations are used to study the effect of mode coupling in these scenarios. In the range-dependent waveguide with the strong inhomogeneity, the coupling effect appears in oscillations of modal amplitudes ratio in the frequency domain, and additional modal arrivals in the time domain. These results were confirmed in a field experiment at two 2-km acoustic tracks in the coastal zone of the Japan Sea, where time-frequency structures of normal modes were extracted using correlation processing at a vertical receiving array.*

Keywords: *shallow-water acoustics; broadband signal; inhomogeneous waveguide; parabolic equation; normal modes; mode coupling.*

INTRODUCTION. PROBLEM STATEMENT

Remote sensing of inhomogeneities in the underwater environment is commonly implemented using acoustic signals. This includes the reconstruction of sea bottom properties [1], local relief features, and water layer characteristics. In this study, the focus is made on local inhomogeneities which extent are much smaller than the length of the acoustic track. Previous works have shown that such inhomogeneities can induce a mode coupling effect, leading to oscillations of modal amplitudes in the frequency domain [2]. This effect can be leveraged for acoustic remote sensing in field experiments.

In present study we simulate a broadband acoustic field (200 to 2000 Hz) produced by a point source in shallow-water range-dependent waveguides. Two acoustic tracks with different bathymetries are examined: one with a smooth bottom (Fig. 1a) and another featuring an "underwater barrier" that almost touches the thermocline (Fig. 1b). The objective of this study is to analyse the time-frequency structure of normal modes and detect manifestations of the mode coupling effect in numerical simulations and in a field experiment.

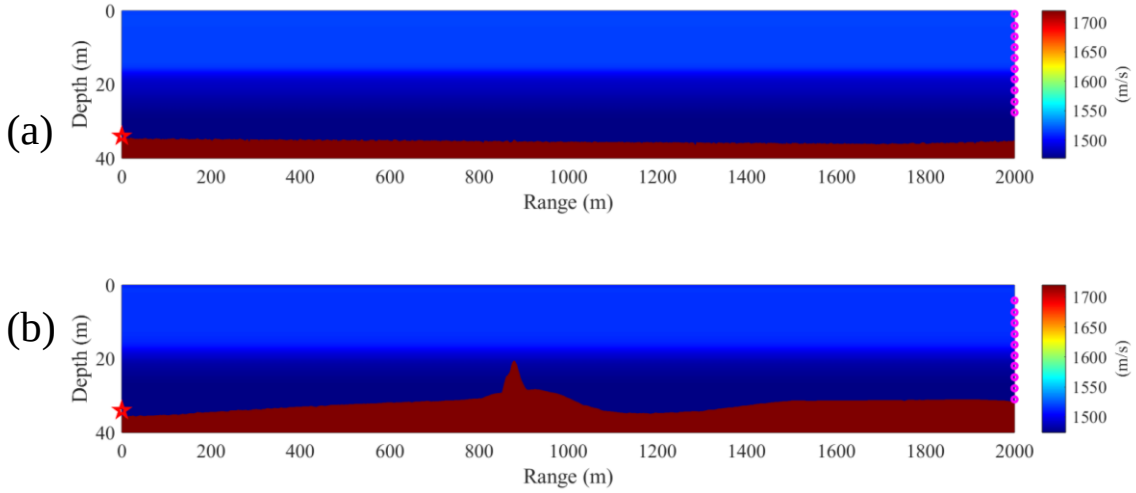


Fig. 1: Waveguide models: (a) with a strong bathymetric feature and (b) with smooth bottom. Red stars mark a point source at depth $z_s = 34$ m. Violet circles at the end of tracks are the receivers of VLA. Sound speed is shown by color. Environmental data were collected in a coastal region of the Japan Sea.

NUMERICAL SIMULATION IN RANGE-DEPENDENT WAVEGUIDES

Pressure field at a given point (x, y, z) with a source frequency f $(\omega = 2\pi f)$ can be represented as a sum of M normal modes [3]:

$$P(\omega, r, z) = \sum_{l=1}^{M(\omega)} A_l(\omega, r)\Psi_l(\omega, r, z)\,, \tag{1}$$

where $A_l(\omega, r)$ is an l-th mode amplitude, $\Psi_l(\omega, r, z)$ is local mode functions. Using parabolic equation method (RAM) [4] the pressure field $P(\omega, r, z)$ is calculated. Extraction of the local mode amplitudes $\widetilde{A_l}(\omega, r)$ at a receiving array is carried out with local eigenfunctions $\Psi_l(\omega, r, z)$ calculated by KRAKENC [5] according to

$$\widetilde{A_l}(\omega, r) = \int_0^H P(\omega, r, z)\Psi_l(\omega, r, z)\, dz. \tag{2}$$

This procedure, Eq. (2), was applied both to simulated and experimental data.

If the sound source spectrum $S_0(\omega)$ is known, the time realization of the signal at the receiver can be obtained using the inverse Fourier transform:

$$p(t,r,z) = 2Re\left(\int_{\omega_1}^{\omega_2} S_0(\omega)P(\omega,r,z)e^{i\omega t}d\omega\right). \quad (3)$$

In our modeling, sound source with the uniform spectrum $S_0(\omega) = 1$ is considered.

FIELD EXPERIMENT

The field experiment was conducted in September 2024, in the coastal region of the Japan Sea. Two acoustic tracks had a length R of 2 km (Fig. 1). Bathymetry was measured with an echosounder. A wideband point source, operating over a frequency range of 200 to 2000 Hz [6], was positioned near the bottom at a depth of $z_s = 34$ m. The linear frequency modulated (LFM) signal lasted for 20 seconds was emitted once per minute. Acoustic signals were recorded at a vertical line array (VLA) of hydrophones [7] that spanned the entire water column, with a spacing of 3 meters between recorders (Fig. 1). In total, 30 cycles were transmitted for each of the two tracks, with the time-frequency structures of the normal modes extracted through correlation processing, mode filtering (2) and taking spectrograms (short-time Fourier transform).

Sound speed profiles in the water column were obtained with CTD. During the experiment the seasonal thermocline was observed, and its depth was ≈ 20 m. The characteristics of bottom in the region are taken from [1] and assumed to be constant: density is $\rho_b = 1800$ kg/m^3, sound speed is $c_b = 1740$ m/s, sound attenuation coefficient is $\beta_\lambda = 0.46$ dB/λ.

RESULTS AND DISCUSSION

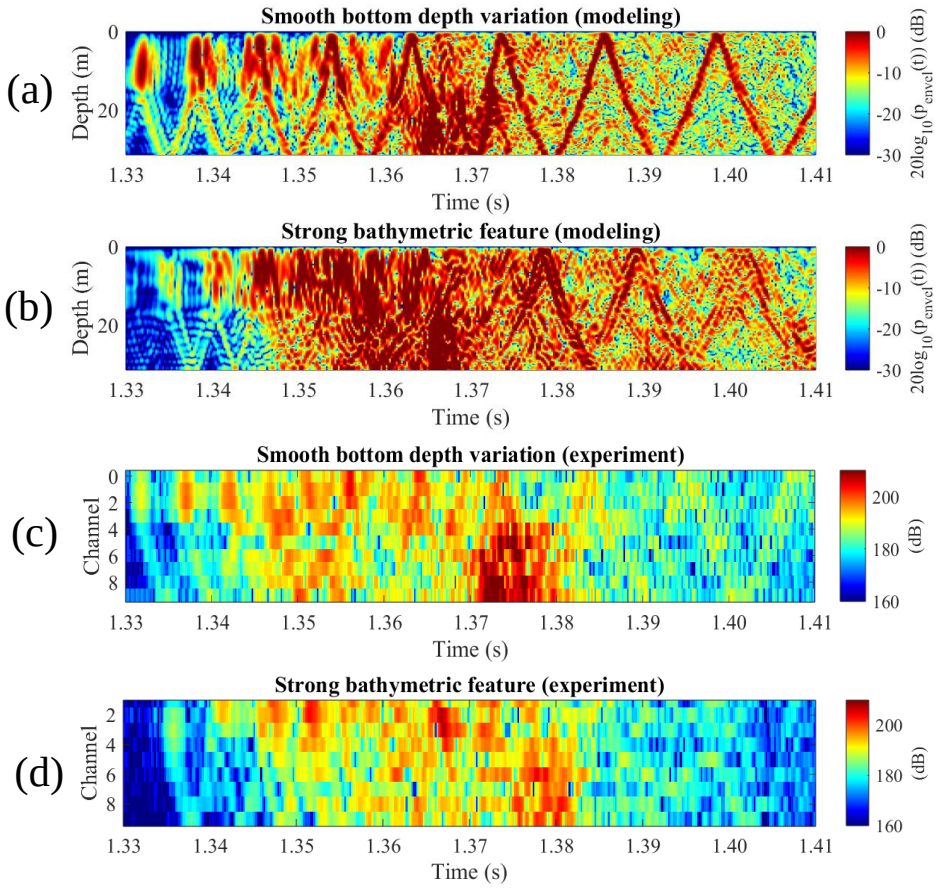


Fig. 2: Spatio-temporal structure of acoustic field $p(t,R,z)$ at the array for 2 tracks shown in Fig. 1; (a) and (b) are the results of numerical simulation. (c) and (d) are the results of the field experiment. The scale is logarithmic.

Figure 2 shows the complicated arrival structure at both tracks. However, first arrivals look similar for two types of waveguides. These arriving rays do not meet the bottom feature and interact with the bottom only near the source. The main part of the received signal is poorly resolved. It is a combination of rays and modes. Analyzing this part of the structure could give information about the presence of the barrier on the track.

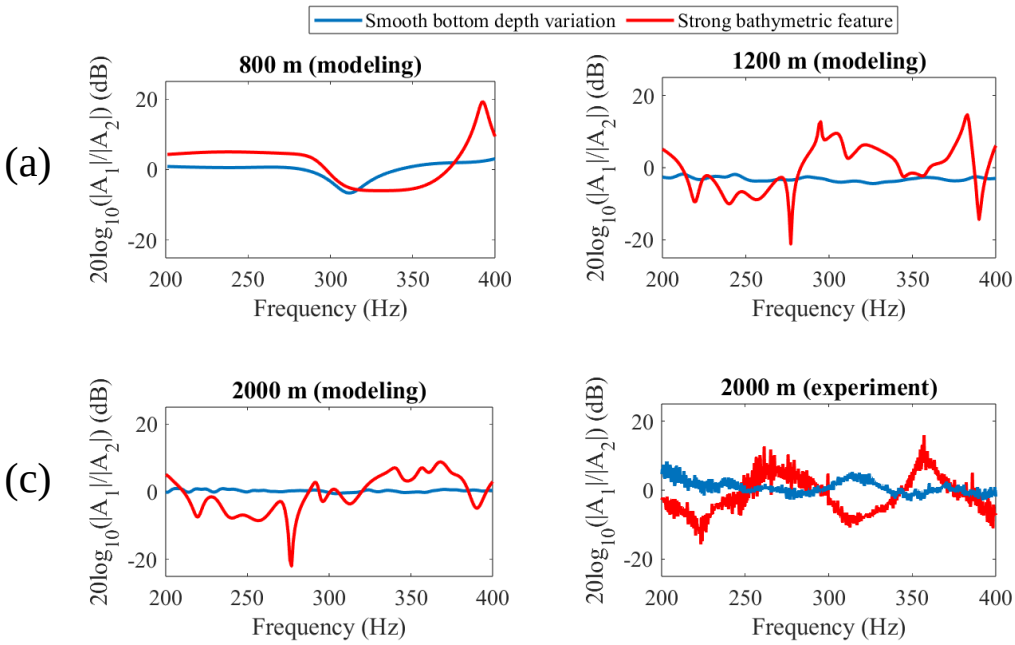


Fig. 3: Ratio of 1-st and 2-nd mode amplitudes on logarithmic scale vs. frequency; (a), (b) and (c) are the numerical simulation results for different ranges from the sound source. (d) is the experimental result for the same range as in (c).

Figure 3 illustrates the amplitude ratio of the first and second modes, $\left|\widetilde{A_1}/\widetilde{A_2}\right|$, at different points along the tracks. Note the oscillations of the ratio for the track with a barrier in Figures 3b and 3c, where the barrier is positioned between the source and the VLA (red curve). Conversely, for smooth bottom scenarios (blue curves in all Fig. 3 and red curve in Fig. 3a), the mode coupling effect is negligible. Results from the field experiment (Fig. 3d) show a pattern similar to that from the numerical simulation.

To investigate this effect further, Figures 4 and 5 present the time-frequency mode structures (spectrograms). Figure 4 depicts extracted modes for the waveguide characterized by a strong bathymetric feature, where mode amplitude oscillations in the frequency domain are linked to energy transfer between modes. Some of the "original" (unconverted) modes generate the "derivative" (converted) modes of another numbers after interaction with an inhomogeneity on the track and its manifestation can be seen in mode spectrograms. Following interaction with the barrier, a portion of energy from, e.g. the second mode shifts to the first mode (see Fig. 4a and 4c). Additionally, the seasonal thermocline alters the modal structure, complicating the dispersion curves. At low frequency mode travel time decreases with increasing mode number, resulting in additional earlier arrivals initiated by mode coupling (Fig. 5a and 5c). The observed time difference $\Delta\tau_{1h}$ between earliest (the "derivative" first mode) and main arrival (the "original" first mode) of the first mode is 16 ms in numerical simulation and 20 ms in the experimental data. We can assume that the "original" first mode propagates with a group speed of $V_{gr,1} = 1465$ m/s (sound speed in the water column under the thermocline) while a higher order modes propagates at a speed of $V_{gr,h} = 1515$ m/s (sound speed in the water near the surface), which coupling leads to an earlier arrival of the first mode. The distance between the source and barrier can be estimated

as $r_s = \Delta\tau_{1h} V_{gr,1} V_{gr,h} / (V_{gr,h} - V_{gr,1})$, with $r_{s,ex} = 888$ m (experiment result) and $r_{s,mod} = 710$ m (simulation result).

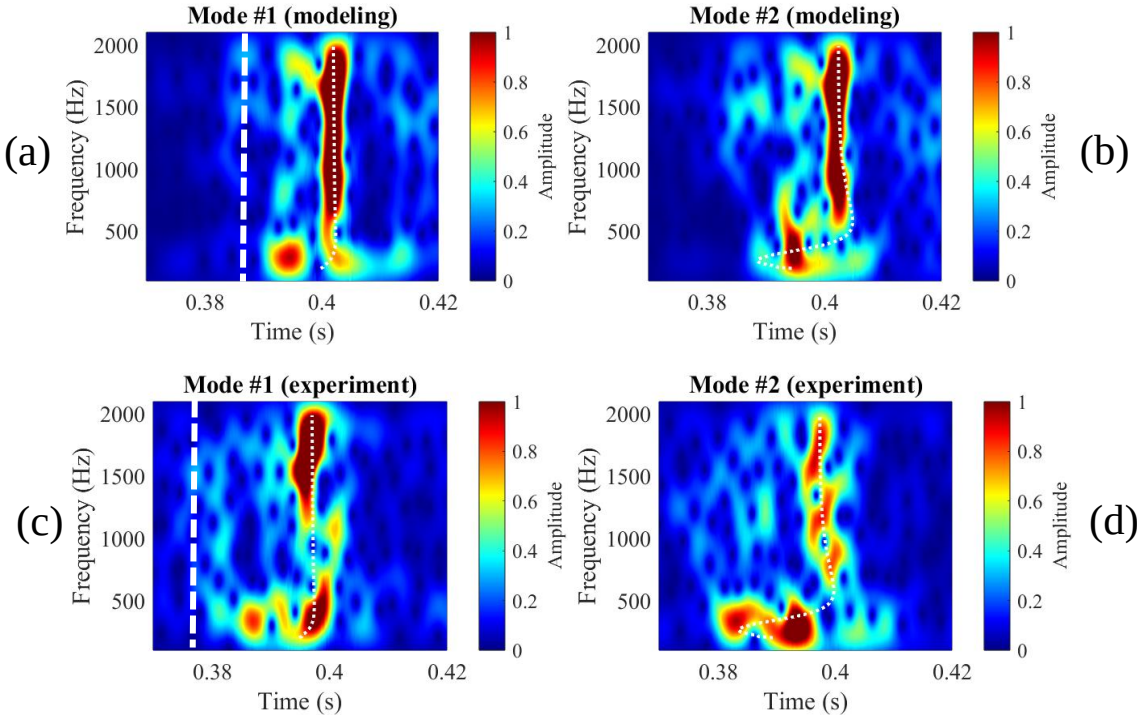


Fig. 4: Time-frequency mode structures for the track with a barrier (Fig. 1b). Top: numerical simulation results for 1-st and 2-nd modes. Bottom: field experiment results for the same modes. Dotted curve is a mode dispersion; dashed line is a mark of the first arrival.

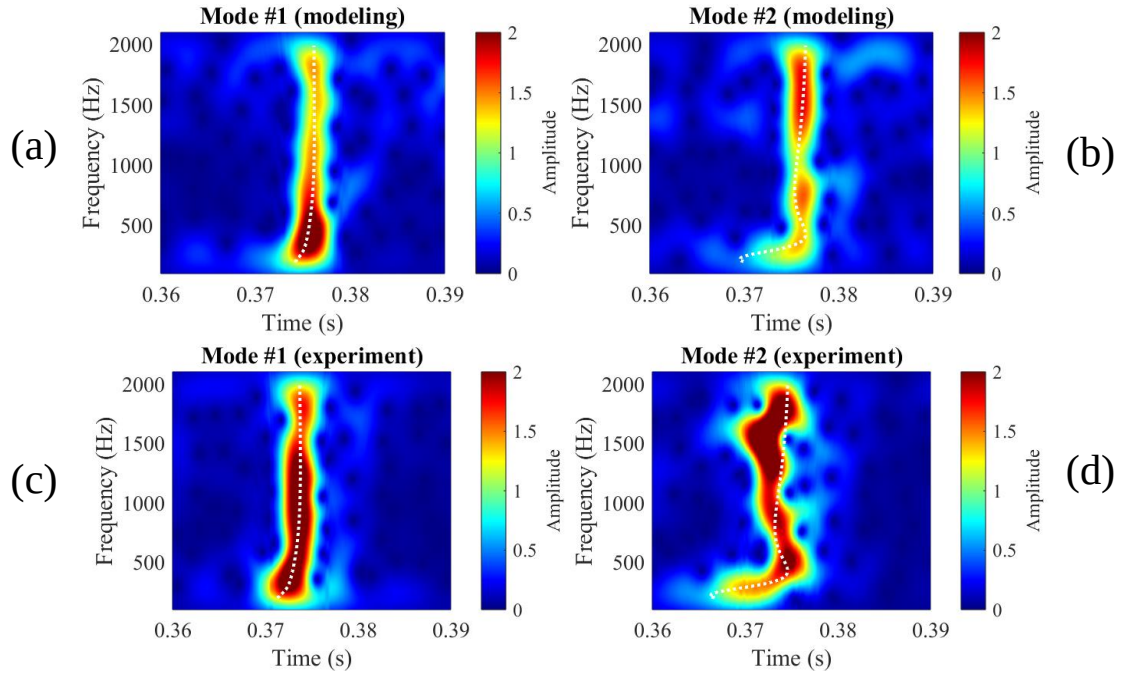


Fig. 5: Time-frequency mode structures for the track with a smooth bathymetry (Fig. 1b). Top: numerical simulation results for 1-st and 2-nd modes. Bottom: field experiment results for the same modes. Dotted curve is a mode dispersion.

CONCLUSION

The simulation of the broadband acoustic field in a range-dependent waveguide with a pronounced bathymetric feature and smoothly varying bottom depths reveals significant effects on acoustic propagation. The presence of a barrier at the acoustic track induces a mode coupling effect. This effect is characterized by oscillations in the modal amplitudes within the frequency domain. Additionally, in the time domain, an increased number of early modal arrivals is observed, which can be used to evaluate the distance between the source and the barrier. These modal features, significant in both frequency and time domains, indicate the complexity introduced by environmental factors like a thermocline and bathymetric variations.

AKNOWLEGMENTS

The work was supported by the Russian Science Foundation, project No. 22-72-10121 (https://rscf.ru/en/project/22-72-10121/).

The measuring instruments used in the experiment were developed at the Pacific Oceanological Institute of the Far Eastern Branch of the Russian Academy of Sciences within the framework of the state assignment "Study of the nature of linear and nonlinear interaction of geospheric fields of transition zones of the World Ocean and their consequences" No. 124022100074-9.

REFERENCES

[1]. **Fershalov M. Y. et al.** Generalization of the method for single-hydrophone geoacoustic inversion: Application to a waveguide with inhomogeneous bottom relief**,** *Underwater Invest. Robot.*, 1, pp. 35, 2021.

[2]. **Lunkov A. A., Shermeneva M. A.,** Mode coupling due to a local inhomogeneity in a shallow-water acoustic waveguide in a broad frequency band, *Acoustical Physics*, 68(5), pp. 467-475, 2022.

[3]. **Katsnelson B., Petnikov V., Lynch J.,** Fundamentals of shallow water acoustics. – Springer Science & Business Media, 2012.

[4]. **M.D. Collins**, Generalization of the split-step Padé solution. *The Journal of the Acoustical Society of America*, 96(1), pp. 382-385, 1994.

[5]. **M.B. Porter,** The KRAKEN normal mode program. *Naval Research Lab Washington DC*, 1992.

[6]. **Akulichev V. A. et al.,** Acoustic tomography of dynamic processes in a sea shelf zone with the use of complex signals, *Acoustical Physics*, 48(1), pp. 1-7, 2002.

[7]. **Rutenko A. N., Kovzel D. G., Gritsenko V. A.,** Self-Contained Vertical Acoustic–Hydrophysical Measuring Systems Mollyusk-19 and Mollyusk-21, *Acoustical Physics*, 70(4), pp. 769-781, 2024.

Sound scattering by nested cylindrically symmetric and linearly invariant anomalies

Sven M. Ivansson[1]

[1]SE-11529 Stockholm, Sweden

Sven M. Ivansson, SE-11529 Stockholm, Sweden; sven.ivansson@gmail.com

Abstract: *Several previous papers have applied the coupled-mode method to assess three-dimensional scattering by cylindrically symmetric anomalies (e.g., seamounts) or anomalies which are invariant in a horizontal direction (e.g., wedges and canyons). Based on a medium discretization with laterally homogeneous rings and strips, an extension to combinations of the two anomaly types has recently been made. For each anomaly, the computations involve recursively determined reflection matrices, which relate the expansion coefficients for incoming and outgoing normal modes. Concerning ring and strip anomalies, the horizontal variation of these expansion coefficients is conveniently described by cylindrical and plane wave functions, respectively. Iterative solution of a linear equation system for the amplitudes of the scattered cylindrical waves from the ring anomalies, involving formulas for transformation between plane and cylindrical waves, allows numerical computation of the field. Within this iterative solution procedure, adaptive numerical integration is used to compute discrete sets of expansion coefficients, resulting from strip-structure interaction, for cylindrical waves incident on the ring anomalies. The present paper develops a further extension, where the ring anomalies need not be disjoint but can be nested within each other and within the strip structure of the medium. Additional translation formulas for cylindrical wave function come into play. Furthermore, the iterates for solution of a pertinent linear equation system for the amplitudes of the scattered cylindrical waves from the ring anomalies are now themselves determined iteratively. The field within a ring anomaly containing other ring anomalies follows by solving a multiple-source problem with the included ring anomalies as (additional) effective sources. Partial waves can be computed to study details of multiple scattering among individual ring and strip anomalies. It can thereby be convenient to use nested ring anomalies with common centre points.*

Keywords: *wave-function translation, reflection-matrix recursion, mode coupling, partial waves*

1. INTRODUCTION

The discrete coupled-mode method, with a horizontal discretization of the medium into laterally homogeneous regions, is extended in [1] to compute the 3D pressure field p in fluid media which include a combination of cylindrically symmetric ring anomalies R and linearly invariant strip anomalies S. The computations are performed iteratively, incorporating successively higher orders of multiple scattering between the ring and strip anomalies. Hence, see [1, Eq. (37)], the successive terms in the field expansion $p = p_R + p_{S,R} + p_{R,S,R} + ..$ are computed. The initial partial field p_R ignores the strip anomalies. Further scattering by the strip anomalies give rise to additional incident waves on the ring anomalies, with discrete sets of cylindrical-wave expansion coefficients, which can be computed efficiently by adaptive wavenumber integration.

For example, see [1, Fig. 4], it is possible to solve the pertinent Helmholtz equation for 3D underwater sound propagation in a strait with two cylindrically symmetric islands. Figure 1 concerns the same example, but for 50 rather than 25 Hz. The two islands, similar with outer and inner radii 1.6 and 1.0 km, respectively, are centred at (x, y) = (1,7) and (1,14) km, respectively, in a Cartesian xy coordinate system for the horizontal plane. Up-sloping strait boundaries appear between x = -1 and -4 km, and between x = 3 and 6 km. Of course, there are now more propagating modes, six rather than three where the water depth is 200 m. The main characteristics of the fields at 50 and 25 Hz are rather similar, however.

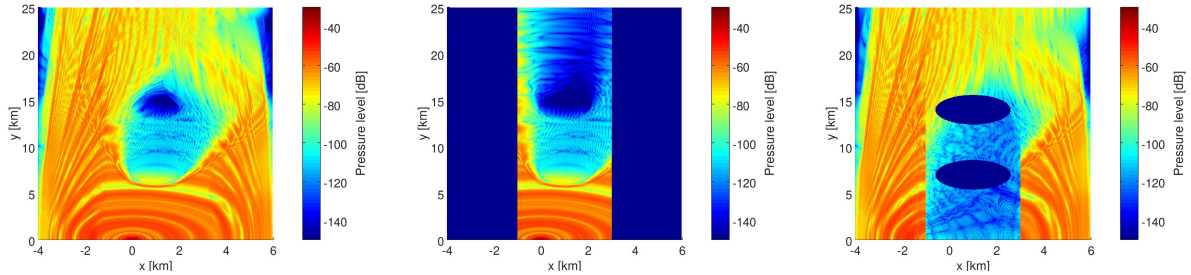


Figure 1: Coupled-mode results, in dB re *1 m, as in [1, Fig. 4(a)-(c)] but for 50 Hz. In particular, the example concerns two islands in a strait. The left panel shows the full field p, while the middle and right panels show partial waves from the iterative solution procedure: p_R (middle), and $p_{S,R}$ (right).*

Returning to 25 Hz, Fig. 2 concerns media with modified water depth for $x <$ -1 km, 200 m for $x <$ -3 km and: constant water depth 200 m for $x <$ -1 km (left), cosine-shaped trench between x = -3 and -1 km with maximum water depth 300 m (middle), and cosine-shaped ridge between x = -3 and -1 km with minimum water depth 100 m (right). Furthermore, the source is moved from (x, y) = (0,0) km to (x, y) = (-3,0) km, but it is still at depth 40 m, and the receivers are still at depth 30 m. The down-slope and up-slope parts of the trench cause reduction and enhancement, respectively, of the field strength. Transmission through the ridge is weak for large y, corresponding to grazing-like incidence with strong reflections back. In the ridge case, the field that is incident on the ring anomaly centred at (1,14) km is thus weak, causing a strong shadow behind it. For the examples in Fig. 2, N×2D approximation results (not shown) are rather good.

As clarified in [2], it is possible to approximate the strip anomalies with very large cylindrically symmetric ring anomalies with a common horizontal centre point far away. The original ring anomalies then appear within the corresponding laterally homogeneous rings with large radii. In effect, a structure appears with nested cylindrically symmetric anomalies. The main

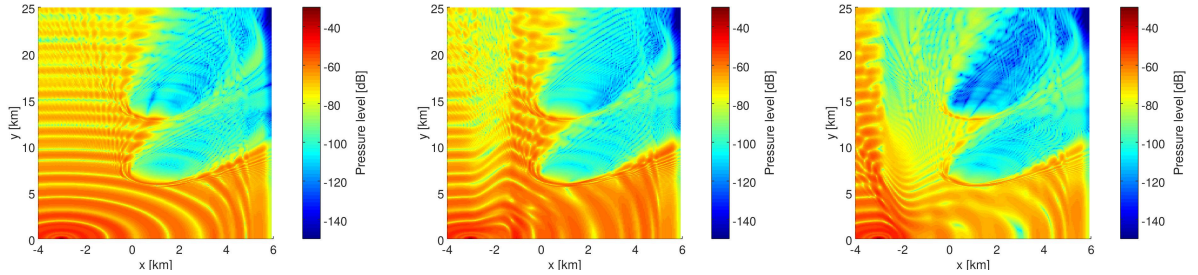


Figure 2: Coupled-mode results at 25 Hz for the full field p, in dB re *1 m, as in [1, Fig. 4(a)] but for modified media as described in the text. In particular, the water depth is 200 m for $x <$ -1 km with the exception of a trench (middle panel) or a ridge (right panel) between $x =$ -3 and -1 km. The source is now at $(x, y) =$ (-3,0) km.*

purpose of the present paper is to generalize [1] to rather general media with nested cylindrically symmetric anomalies within a strip structure. Section 2 provides the basic equations, Sec. 3 proposes an iterative solution method for the main equation (10), and Sec. 4 indicates how to compute partial waves for multiple scattering among individual ring and strip anomalies. There are no further examples, however, since this generalization of [1] has not yet been implemented on the computer.

2. BASIC EQUATIONS

As in [1], consider a fluid medium and a Cartesian xyz coordinate system with depth z. For a symmetric time-harmonic point source with moment-tensor strength M at $\mathbf{x}_s = (x_s, y_s, z_s)$, the problem is to compute the pressure field $p(\mathbf{x})$ at each point $\mathbf{x}$. The typically omitted time factor is $\exp(-\mathrm{i}\omega t)$, where $\omega > 0$ is angular frequency and t is time. The fluid extends between the finite (computational) depth boundaries $z = z_a(x, y)$ and $z = z_b(x, y) > z_a(x, y)$. These exterior boundaries are locally reacting, such that $p(\mathbf{x})$ satisfies [1, Eq. (1)].

In the present paper, the medium is independent of the horizontal y-coordinate except within J cylindrically symmetric ring anomalies with centre points (x_j, y_j) and outer radii $r_{1,j}$, where j=1,2,..,J. In contrast to [1], these ring anomalies need not be disjoint but they can be nested within each other. For computations with the discrete coupled-mode method, the medium is discretized into $N_0 + 1$ strips which are parallel to the y-axis. This discretization takes care of possible y-independent strip anomalies. Except for included ring anomalies, each strip is laterally homogeneous. With N_0 vertical interfaces $x_{1,0} < x_{2,0} < ... < x_{N_0,0}$, strip n covers $x_{n-1,0} < x < x_{n,0}$ for n=2,3,..,N_0, while strips 1 and $N_0 + 1$ include $x < x_{1,0}$ and $x_{N_0,0} < x$, respectively.

As in [1], ring anomaly j is discretized with $N_j + 1$ horizontal rings, numbered from outside and separated by concentric circles $r_j(x, y) = r_{n,j}$. Local polar coordinates $(r_j(x, y), \phi_j(x, y))$ are centred at (x_j, y_j) in the xy-plane. The $r_{n,j}$, $n = 1, 2, .., N_j$, form a decreasing sequence of positive radii. Each ring anomaly j, j=1,2,..,J, can be within a ring $n_e(j) > 1$ of another ring anomaly denoted $j_e(j) = 1, .., J$. Nested chains of ring anomalies can thus be formed, but $n_e(j), j_e(j)$ concern the closest surrounding ring anomaly in a chain. For the outermost ring anomaly j in a chain, $j_e(j) = 0$ and $n_e(j) = 1, 2, .., N_0 + 1$ is the strip in which it is contained. Except for included ring anomalies, and other strips than $n_e(j)$ for ring 1 of a ring anomaly j with $j_e(j) = 0$, each ring is laterally homogeneous. The vertical strip and ring interfaces must not intersect. For n=1,2,..,$N_0 + 1$, it is convenient to define the sets $J_{n,0} = \{j = 1, 2, .., J :$

$n_e(j) = n$ and $j_e(j) = 0\}$. Moreover, for j=1,2,..,J, $J_{n,j} = \{j' = 1, 2, .., J : n_e(j') = n$ and $j_e(j') = j\}$ when n=2,3,..,$N_j + 1$, while $J_{1,j} = \{j' = 1, 2, .., J : n_e(j') = n_e(j)$ and $j_e(j') = j_e(j)$, and $j' \neq j\}$. In [1], where all ring anomalies are disjoint, $J_{n,0}$ is simply denoted as J_n. Figure 3 gives an illustration, cf. [1, Fig. 1].

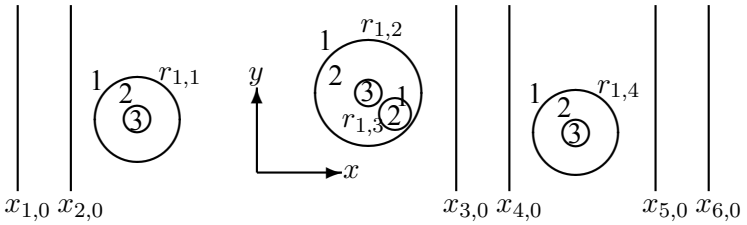


Figure 3: The horizontal xy-plane with a strip structure and four ring anomalies together with their ring radii and ring numbers. In this case, $N_0 = 6$, $J = 4$, $N_1 = N_2 = N_4 = 2$, and $N_3 = 1$. Furthermore, $j_e(1) = j_e(2) = j_e(4) = 0$ and $j_e(3) = 2$, while $n_e(1) = n_e(2) = 3$, $n_e(3) = 2$, and $n_e(4) = 5$. $J_{1,0} = J_{2,0} = J_{4,0} = J_{6,0} = J_{7,0} = \emptyset$, while $J_{3,0} = \{1, 2\}$ and $J_{5,0} = \{4\}$. $J_{2,1} = J_{3,1} = J_{3,2} = J_{1,3} = J_{2,3} = J_{1,4} = J_{2,4} = J_{3,4} = \emptyset$, while $J_{1,1} = \{2\}$, $J_{1,2} = \{1\}$, and $J_{2,2} = \{3\}$. The sound source can be anywhere in the medium, within one of the horizontal ring anomalies or outside all of them.

Densities $\rho_{n,j}(z)$, sound speeds $c_{n,j}(z)$, Lamé parameters $\lambda_{n,j}(z)$, as well as modal horizontal wavenumbers $k_{m,n,j}$ and normalized (n.b.) mode functions $Z_{m,n,j}(z)$, are as in [1]. This is true for the depth boundaries $z_{a;n,j}$,$z_{b;n,j}$, the admittance values $Y_{a;n,j}$,$Y_{b;n,j}$, and the depth intervals $I_{n,j}$ as well. Of course, for each j and n, these quantities are only relevant outside possible interior ring anomalies within the corresponding strip or ring. For $j' \in J_{n,j}$, $\rho_{1,j'}(z) = \rho_{n,j}(z)$, $c_{1,j'}(z) = c_{n,j}(z)$, etc. This is true for all j=0,1,2,..,J.

With the following clarification, j_s and n_s for the source are also as in [1]. Unless $j_s = 0$, j_s and $n_s(> 1)$ concern the smallest ring anomaly containing the source.

For j=0,1,2,..,J and n=1,2,..,$N_j + 1$, the row vectors $\mathbf{\Phi}_{n,j}(\mathbf{x})$ and $\mathbf{\Psi}_{n,j}(\mathbf{x})$ are as in [1]. Of course, ν=0,1,.. still denotes azimuthal order, σ=e,o indicates even,odd parity (for $\nu = 0$ only σ =e), and κ is still the wavenumber parameter pertaining to Fourier transformation with respect to y. For n=1,2,..,N_0+1, the diagonal matrices $\hat{\mathbf{E}}_n$ are also as in [1]. The mode-coupling matrices $\mathbf{F}_{n,j}$,$\mathbf{G}_{n,j}$, $\bar{\mathbf{F}}_{n,j}$,$\bar{\mathbf{G}}_{n,j}$, and $\mathbf{L}_{n,j}$ remain unchanged too.

With $J_{n,0}$ replacing J_n, the field representation [1, Eq. (8)] is still valid. Immediately outside ring anomaly j=1,..,J, [1, Eq. (10)] holds with $n = 1$. Within ring n=2,..,N_j+1 of ring anomaly j=1,..,J (but outside possibly included other ring anomalies there),

$$\begin{aligned} p(\mathbf{x}) \;=\; & \delta_{j_s j}\,\delta_{n n_s} \sum_m a_{m,s} Z_{m,n_s,j_s}(z) \mathrm{H}_0^{(1)}(k_{m,n_s,j_s} r_s(x,y)) + \sum_{j' \in J_{n,j}} \mathbf{\Psi}_{1,j'}(\mathbf{x}) \cdot \mathbf{b}_{1,j'} \\ & + \mathbf{\Phi}_{n,j}(\mathbf{x}) \cdot \left(\bar{\mathbf{a}}_{n,j} + \bar{\mathbf{a}}^L_{n,j} + \bar{\mathbf{a}}^R_{n,j} \right) + \mathbf{\Psi}_{n,j}(\mathbf{x}) \cdot \left(\mathbf{b}_{n,j} + \mathbf{b}^L_{n,j} + \mathbf{b}^R_{n,j} \right) . \end{aligned} \tag{1}$$

Here, $\bar{\mathbf{a}}_{n,j} = \{(\bar{a}_{n,j})_{m,\nu,\sigma}\}$, $\bar{\mathbf{a}}^L_{n,j} = \{(\bar{a}^L_{n,j})_{m,\nu,\sigma}\}$, $\bar{\mathbf{a}}^R_{n,j} = \{(\bar{a}^R_{n,j})_{m,\nu,\sigma}\}$, and $\mathbf{b}_{n,j} = \{(b_{n,j})_{m,\nu,\sigma}\}$, $\mathbf{b}^L_{n,j} = \{(b^L_{n,j})_{m,\nu,\sigma}\}$, $\mathbf{b}^R_{n,j} = \{(b^R_{n,j})_{m,\nu,\sigma}\}$, are column vectors. The coefficient vectors $\bar{\mathbf{a}}^L_{n,j}$ and $\mathbf{b}^L_{n,j}$ represent contributions originating from source and anomalies outside the disk $r_j(x,y) < r_{n-1,j}$. For $n \leq N_j$, $\bar{\mathbf{a}}^R_{n,j}$ and $\mathbf{b}^R_{n,j}$ represent contributions originating from source and ring anomalies inside the disk $r_j(x,y) < r_{n,j}$, while $\bar{\mathbf{a}}^R_{N_j+1,j} = \mathbf{b}^R_{N_j+1,j} = \mathbf{0}$. Apparently, $\bar{\mathbf{a}}^L_{2,j}$ and $\mathbf{b}^L_{2,j}$ are determined by $\mathbf{a}_{1,j}$ from [1, Eq. (10) with n=1]. Moreover, $\mathbf{b}_{N_j+1,j} = \mathbf{b}^L_{N_j+1,j} = \mathbf{0}$ are boundary conditions. As in [1], δ is the Kronecker delta and $r_s(x,y) = |(x,y) - (x_s,y_s)|$.

Next, counterparts of the wave-transformation relations [1, Eqs. (13) and (17)] are formulated. Consider ring anomaly j=1,2,..,J. When $j_e(j) = j' > 0$,

$$\mathbf{\Psi}_{n',j'}(\mathbf{x}) = \mathbf{\Phi}_{1,j}(\mathbf{x}) \cdot \mathbf{Q}_{1,j,n',j'} \quad \text{and} \quad \mathbf{\Phi}_{n',j'}(\mathbf{x}) = \mathbf{\Phi}_{1,j}(\mathbf{x}) \cdot \mathbf{P}_{1,j,n',j'} \tag{2}$$

are valid with $n' = n_e(j) > 1$, where the first relation requires that $|(x,y) - (x_j,y_j)| < |(x_{j'},y_{j'}) - (x_j,y_j)|$. The elements $(Q_{1,j,n',j'})_{m,\nu,\sigma;m',\nu',\sigma'}$ and $P_{1,j,n',j'})_{m,\nu,\sigma;m',\nu',\sigma'}$ of the matrices $\mathbf{Q}_{1,j,n',j'}$ and $\mathbf{P}_{1,j,n',j'}$, respectively, agree with those of $\mathbf{Q}_{j,j'}$, as given in [3, (Eqs. (11)-(12))], with the following changes: 1) replace the modal wavenumbers $k_{m,1}$ by $k_{m,1,j} = k_{m,n',j'}$, 2) concerning $\mathbf{Q}_{1,j,n',j'}$, replace $r_{1,j'}$ by $r_{n',j'}$, and 3) concerning $\mathbf{P}_{1,j,n',j'}$, replace the Hankel functions $\mathrm{H}^{(1)}_{\nu-\nu'}$ and $\mathrm{H}^{(1)}_{\nu+\nu'}$ by the corresponding Bessel functions $\mathrm{J}_{\nu-\nu'}$ and $\mathrm{J}_{\nu+\nu'}$, respectively, and replace division with $\mathrm{H}^{(1)}_{\nu'}(k_{m,1}r_{1,j'})$ by multiplication with $\mathrm{H}^{(1)}_{\nu'}(k_{m,n',j'}r_{n'-1,j'})$. This follows from [1, Eq. (12)] and [4, Eq. (5.10)].

When $n > 1$ and $J_{n,j}$ is nonempty for ring anomaly j=1,2,..,J, consider an included ring anomaly $j' \in J_{n,j}$. Applying [1, Eq. (12)] and [4, Eq. (5.11)], with matrices $\hat{\mathbf{H}}_{n,j}$ as in [3],

$$\mathbf{\Psi}_{1,j'}(\mathbf{x}) = \mathbf{\Phi}_{n,j}(\mathbf{x}) \cdot \hat{\mathbf{H}}^{-1}_{n,j} \cdot \mathbf{Q}_{n,j,1,j'} \quad \text{and} \quad \mathbf{\Psi}_{1,j'}(\mathbf{x}) = \mathbf{\Psi}_{n,j}(\mathbf{x}) \cdot \hat{\mathbf{H}}^{-1}_{n,j} \cdot \mathbf{P}_{n,j,1,j'} \ , \tag{3}$$

are valid for $|(x,y) - (x_j,y_j)| < |(x_{j'},y_{j'}) - (x_j,y_j)|$ and $|(x,y) - (x_j,y_j)| > |(x_{j'},y_{j'}) - (x_j,y_j)|$, respectively. The elements $(Q_{n,j,1,j'})_{m,\nu,\sigma;m',\nu',\sigma'}$ and $P_{n,j,1,j'})_{m,\nu,\sigma;m',\nu',\sigma'}$ of the matrices $\mathbf{Q}_{n,j,1,j'}$ and $\mathbf{P}_{n,j,1,j'}$, respectively, agree with those of $\mathbf{Q}_{j,j'}$, as given in [3, (Eqs. (11)-(12))], with the following changes: 1) replace $k_{m,1}$ by $k_{m,n,j} = k_{m,1,j'}$, 2) concerning $\mathbf{Q}_{n,j,1,j'}$, replace $r_{1,j}$ by $r_{n,j}$, and 3) concerning $\mathbf{P}_{n,j,1,j'}$, replace $\mathrm{H}^{(1)}_{\nu-\nu'}$ and $\mathrm{H}^{(1)}_{\nu+\nu'}$ by $\mathrm{J}_{\nu-\nu'}$ and $\mathrm{J}_{\nu+\nu'}$, respectively, and replace division with $\mathrm{H}^{(1)}_{\nu}(k_{m,1}r_{1,j})$ by multiplication with $\mathrm{H}^{(1)}_{\nu}(k_{m,n,j}r_{n-1,j})$.

With $J_{n,0}$ replacing J_n, [1, Eqs. (20)-(22)] are still valid. Concerning the ring-anomaly computations for each j=1,2,..,J, the reflection matrices $\bar{\mathbf{R}}_{n,j}$, mentioned at the end of [1, Sec. 2.3], are now not only needed for $j = j_s$ and n=1,2,..,n_s but also for n=1,2,..,max$\{n' > 1 : J_{n',j} \neq \emptyset\}$. This is so, since ring anomalies within ring anomaly j act as (additional) effective sources. The counterpart of [1, Eq. (20)] becomes

$$\mathbf{b}^L_{n,j} = \mathbf{R}_{n,j} \cdot \hat{\mathbf{H}}_{n,j} \cdot \bar{\mathbf{a}}^L_{n,j} \ , \ \mathbf{b}_{n,j} = \mathbf{R}_{n,j} \cdot \mathbf{a}_{n,j} \ , \ \bar{\mathbf{a}}^R_{n,j} = \bar{\mathbf{R}}_{n,j} \cdot \hat{\mathbf{H}}_{n,j} \cdot \mathbf{b}^R_{n,j} \ , \ \bar{\mathbf{a}}_{n,j} = \bar{\mathbf{R}}_{n,j} \cdot \bar{\mathbf{b}}_{n,j} \tag{4}$$

for $n = 2, 3, .., N_j+1$. Here, by a generalization of [3, Eqs. (29) and (32)] to the present context, using Eqs. (3),

$$\mathbf{a}_{n,j} = \hat{\mathbf{H}}_{n,j} \cdot \bar{\mathbf{a}}_{n,j} - \delta_{j_s j}\, \delta_{n n_s}\, \frac{\mathrm{i}M\omega^2}{4\,\lambda_{n_s,j_s}(z_s)} \mathbf{\Psi}^{\mathrm{T}}_{n_s,j_s}(\mathbf{x}_s) + \sum_{j' \in J_{n,j}} \mathbf{Q}_{n,j,1,j'} \cdot \mathbf{b}_{1,j'} \tag{5}$$

$$\bar{\mathbf{b}}_{n,j} = \hat{\mathbf{H}}_{n,j} \cdot \mathbf{b}_{n,j} - \delta_{j_s j}\, \delta_{n n_s}\, \frac{\mathrm{i}M\omega^2}{4\,\lambda_{n_s,j_s}(z_s)} \mathbf{\Phi}^{\mathrm{T}}_{n_s,j_s}(\mathbf{x}_s) + \sum_{j' \in J_{n,j}} \mathbf{P}_{n,j,1,j'} \cdot \mathbf{b}_{1,j'} \ . \tag{6}$$

The incoming waves upon ring anomaly j=1,2,..,J appear as $\mathbf{\Phi}_{1,j}(\mathbf{x}) \cdot \mathbf{a}_{1,j}$. With $\mathbf{a}^S_{1,j}$ and $\mathbf{a}^0_{1,j}$ still according to [1, Eqs. (25)-(26)],

$$\mathbf{a}_{1,j} = \delta_{j_s j_e(j)}\, \delta_{n_s n_e(j)}\, \mathbf{a}^0_{1,j} + \sum_{j' \in J_{1,j}} \mathbf{Q}_{j,j'} \cdot \mathbf{b}_{1,j'} + \delta_{0 j_e(j)}\, \mathbf{a}^S_{1,j} + (1 - \delta_{0 j_e(j)})\, \mathbf{a}^R_{1,j} \ , \tag{7}$$

which represents an extension of [1, Eq. (24)], note that $J_{1,j} = \{j' \in J_{n_e(j),j_e(j)} : j' \neq j\}$. Concerning $\mathbf{a}^R_{1,j}$, note the difference to $\bar{\mathbf{a}}^R_{n,j}$ and $\mathbf{b}^R_{n,j}$, with the superscript ^{R} for $\mathbf{a}^R_{1,j}$ indicating ring-anomaly interaction. When $j_e(j) \neq 0$, and hence $n_e(j) > 1$, Eqs. (1) and (2) yield

$$\begin{aligned} \mathbf{a}^R_{1,j} &= \mathbf{P}_{1,j,n_e(j),j_e(j)} \cdot \left(\bar{\mathbf{a}}_{n_e(j),j_e(j)} + \bar{\mathbf{a}}^L_{n_e(j),j_e(j)} + \bar{\mathbf{a}}^R_{n_e(j),j_e(j)} \right) \\ &+ \mathbf{Q}_{1,j,n_e(j),j_e(j)} \cdot \left(\mathbf{b}_{n_e(j),j_e(j)} + \mathbf{b}^L_{n_e(j),j_e(j)} + \mathbf{b}^R_{n_e(j),j_e(j)} \right) \ . \end{aligned} \tag{8}$$

Application of the reflection-matrix operator $\mathbf{R}_{1,j}$ subsequently gives the following slight modification of [1, Eq. (27)]:

$$\mathbf{b}_{1,j} = \mathbf{b}^0_{1,j} + \mathbf{R}_{1,j} \cdot \mathbf{a}_{1,j} \; . \tag{9}$$

Now, $\mathbf{\Psi}_{1,j}(\mathbf{x}) \cdot \mathbf{b}^0_{1,j}$ provides the scattered waves originating not only from the source when $j = j_s \neq 0$, but also from ring anomalies $j' \in J_{n,j}$ for n=2,3,..,$N_j + 1$. In matrix form, Eq. (9) appears as

$$\mathbf{b}_1 = \mathbf{b}^0_1(\mathbf{b}_1) + \mathbf{R}_1 \cdot \left(\mathbf{a}^0_1 + \mathbf{Q} \cdot \mathbf{b}_1 + \mathbf{a}^S_1(\mathbf{b}_1) + \mathbf{a}^R_1(\mathbf{b}_1) \right) \; , \tag{10}$$

Modifications compared to [1, Eq. (28)] are that $\mathbf{Q}$ is now a $J \times J$ block matrix with block j, j' given by $(1-\delta_{jj'})\, \delta_{n_e(j)\, n_e(j')}\, \delta_{j_e(j)\, j_e(j')}\, \mathbf{Q}_{j,j'}$, while $\mathbf{b}^0_1$, $\mathbf{a}^0_1$, $\mathbf{a}^S_1$, and the additional $\mathbf{a}^R_1$ are column vectors with J parts with part j given by $\mathbf{b}^0_{1,j}$, $\delta_{j_s j_e(j)}\, \delta_{n_s n_e(j)}\, \mathbf{a}^0_{1,j}$, $\delta_{0 j_e(j)}\, \mathbf{a}^S_{1,j}$, and $(1-\delta_{0 j_e(j)})\, \mathbf{a}^R_{1,j}$, respectively. As indicated by the notation, all of $\mathbf{b}^0_1$, $\mathbf{a}^S_1$, and $\mathbf{a}^R_1$ now depend on $\mathbf{b}_1$. These dependencies are linear, with possible constant terms arising directly from the source (when $j_s \neq 0$, $j_s = 0$, and $j_s \neq 0$ with $J_{n,j_s} \neq \emptyset$ for some n=2,..,$N_{j_s} + 1$, respectively).

3. NUMERICAL SOLUTION

The amended version of [1, Eq. (29)] becomes

$$\mathbf{b}_1 = (\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \cdot \left[\mathbf{b}^0_1(\mathbf{b}_1) + \mathbf{R}_1 \cdot \left(\mathbf{a}^0_1 + \mathbf{a}^S_1(\mathbf{b}_1) + \mathbf{a}^R_1(\mathbf{b}_1) \right) \right] \; . \tag{11}$$

For trial $\mathbf{b}^0_1$, $\mathbf{a}^S_1$, and $\mathbf{a}^R_1$, it is suitable to compute $(\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \cdot \left[\mathbf{b}^0_1 + \mathbf{R}_1 \cdot \left(\mathbf{a}^0_1 + \mathbf{a}^S_1 + \mathbf{a}^R_1 \right) \right]$ by expanding $(\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1}$ in geometric series, which corresponds precisely to the iterative computation method of [3, Sec. III B]. The computation of $\mathbf{a}^S_1(\mathbf{b}_1)$ for a trial $\mathbf{b}_1$ proceeds according to [1, Sec. 3.1 and Appendix A]. Of course, $J_{n,0}$ thereby replaces J_n in [1, Eqs. (30)-(31)] When $j_s > 0$, $n_{s,0}$ in [1, Sec. 3.1] now equals $n_e(j')$ where j', with $j_e(j') = 0$, is the outermost ring anomaly in the nesting chain containing ring anomaly j_s. Recall that, for each κ, [1, Appendix A] shows how to compute $\bar{\mathbf{a}}^L_{n,0}$ and $\mathbf{b}^R_{n,0}$ recursively by *two* passes of stabilized back-propagation throughout the strip structure to pick up contributions from the left and from the right, respectively. The coefficient vectors $\mathbf{b}^L_{n,0}$ and $\bar{\mathbf{a}}^R_{n,0}$ follow from [1, Eqs. (20)].

To compute $\mathbf{a}^R_1(\mathbf{b}_1)$ for a trial $\mathbf{b}_1$, first note that substitution of the reflection-matrix expressions for $\mathbf{b}_{n,j}$ and $\bar{\mathbf{a}}_{n,j}$ from Eqs. (4) into Eqs. (5)-(6) provides an equation system with the solution, for j=1,2,..,J and n=2,3,..,$N_j + 1$,

$$\begin{aligned}
\mathbf{a}_{n,j} \;=\; & \left(\mathbf{I} - \bar{\mathbf{S}}_{n,j} \cdot \mathbf{S}_{n,j} \right)^{-1} \cdot \Bigg[-\delta_{j_s j}\, \delta_{n n_s}\, \frac{\mathrm{i} M \omega^2}{4\, \lambda_{n_s,j_s}(z_s)} \cdot \Big(\mathbf{\Psi}^{\mathrm{T}}_{n_s,j_s}(\mathbf{x}_s) \\
& + \bar{\mathbf{S}}_{n,j} \cdot \mathbf{\Phi}^{\mathrm{T}}_{n_s,j_s}(\mathbf{x}_s) \Big) + \sum_{j' \in J_{n,j}} \left(\mathbf{Q}_{n,j,1,j'} + \bar{\mathbf{S}}_{n,j} \cdot \mathbf{P}_{n,j,1,j'} \right) \cdot \mathbf{b}_{1,j'} \Bigg]
\end{aligned} \tag{12}$$

$$\begin{aligned}
\bar{\mathbf{b}}_{n,j} \;=\; & \left(\mathbf{I} - \mathbf{S}_{n,j} \cdot \bar{\mathbf{S}}_{n,j} \right)^{-1} \cdot \Bigg[-\delta_{j_s j}\, \delta_{n n_s}\, \frac{\mathrm{i} M \omega^2}{4\, \lambda_{n_s,j_s}(z_s)} \cdot \Big(\mathbf{\Phi}^{\mathrm{T}}_{n_s,j_s}(\mathbf{x}_s) \\
& + \mathbf{S}_{n,j} \cdot \mathbf{\Psi}^{\mathrm{T}}_{n_s,j_s}(\mathbf{x}_s) \Big) + \sum_{j' \in J_{n,j}} \left(\mathbf{P}_{n,j,1,j'} + \mathbf{S}_{n,j} \cdot \mathbf{Q}_{n,j,1,j'} \right) \cdot \mathbf{b}_{1,j'} \Bigg] \; ,
\end{aligned} \tag{13}$$

where $\mathbf{S}_{n,j} = \hat{\mathbf{H}}_{n,j} \cdot \mathbf{R}_{n,j}$ and $\bar{\mathbf{S}}_{n,j} = \hat{\mathbf{H}}_{n,j} \cdot \bar{\mathbf{R}}_{n,j}$. Of course, $\mathbf{b}_{n,j}$ and $\bar{\mathbf{a}}_{n,j}$ subsequently follow from Eqs. (4).

For a ring anomaly j with $j_e(j) = 0$, Eq. (7) provides $\mathbf{a}_{1,j}$ for a trial $\mathbf{b}_1$. Formally defining $\bar{\mathbf{a}}_{1,j} = \bar{\mathbf{a}}^L_{1,j} = \bar{\mathbf{a}}^R_{1,j} = \mathbf{0}$, it is then easy to adapt [1, Appendix A], which concerns the $\bar{\mathbf{a}}^L_{n,0}$ and

$\mathbf{b}^R_{n,0}$, to computation of the coefficient vectors $\bar{\mathbf{a}}^L_{n,j}$ and $\mathbf{b}^R_{n,j}$. The adaptation involves replacement of N_0 by N_j, subscripts $_{,0}$ by $_{,j}$, matrices $\mathbf{K}_n$ by $\mathbf{I}$, and matrices $\hat{\mathbf{E}}_n$ by $\hat{\mathbf{H}}_{n,j}$. Of course, the computations now concern particular ν,σ pairs rather than particular κ, Eqs. (4) are used instead of [1, Eqs. (20)], and the matrices $\mathbf{L}_{n,j}$ are formally obtained from [1, Eq. (A.2)] by setting κ=0 (and replacing the subscripts $_{,0}$ with $_{,j}$). There are two passes of stabilized back-propagation throughout the rings of ring anomaly j to pick up contributions from the source, when $j_s = j$, and from ring anomalies j' with $j_e(j') = j$. The coefficient vectors $\mathbf{b}^L_{n,j}$ and $\bar{\mathbf{a}}^R_{n,j}$ then follow from Eq. (4). For a ring anomaly j' with $j_e(j') = j$, $\mathbf{a}^R_{1,j'}$ and $\mathbf{a}_{1,j'}$ for the trial $\mathbf{b}_1$ subsequently follow from Eqs. (8) and (7). The adapted [1, Appendix A] then provides the coefficient vectors $\bar{\mathbf{a}}^L_{n,j'}$ and $\mathbf{b}^R_{n,j'}$, needed to compute $\mathbf{a}^R_{1,j''}$ and $\mathbf{a}_{1,j''}$ for a ring anomaly j'' with $j_e(j'') = j'$. Continuing in this way, to handle the nesting chains of ring anomalies, $\mathbf{a}^R_{1,j}$ and $\mathbf{a}_{1,j}$ for the trial $\mathbf{b}_1$ follow for all j=1,2,..,J.

The determination of $\mathbf{b}^0_{1,j}$ for a trial $\mathbf{b}_1$ proceeds in a similar way, but independently for each j=1,2,..,J. This time, the computations according to the adapted [1, Appendix A] are initiated with $\mathbf{a}_{1,j}$ temporarily set to zero, and the resulting $\mathbf{b}^R_{1,j}$ provides the desired $\mathbf{b}^0_{1,j}$.

One possibility to extend the iterative solution procedure from [1, Sec. 3.2] to solution of Eq. (11) is as follows. Corresponding to the trial $\mathbf{a}^S_1 = \mathbf{0}$, let $\mathbf{b}^{(0)}_1$ solve $\mathbf{b}^{(0)}_1 = (\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \cdot \left[\mathbf{b}^0_1(\mathbf{b}^{(0)}_1) + \mathbf{R}_1 \cdot \left(\mathbf{a}^0_1 + \mathbf{a}^R_1(\mathbf{b}^{(0)}_1)\right)\right]$. For l=0,1,.., let $\mathbf{b}^{(l+1)}_1$ solve

$$\begin{aligned}\mathbf{b}^{(l+1)}_1 \;=\; & \mathbf{b}^{(0)}_1 + (\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \\ & \cdot \left[\mathbf{b}^{0-}_1(\mathbf{b}^{(l+1)}_1 - \mathbf{b}^{(0)}_1) + \mathbf{R}_1 \cdot \left(\mathbf{a}^S_1(\mathbf{b}^{(l)}_1) + \mathbf{a}^{R-}_1(\mathbf{b}^{(l+1)}_1 - \mathbf{b}^{(0)}_1)\right)\right] \;, \qquad (14)\end{aligned}$$

converging to the sought $\mathbf{b}_1$ as l increases. Here, and below, $\mathbf{b}^{0-}_1$, $\mathbf{a}^{S-}_1$, and $\mathbf{a}^{R-}_1$ denote $\mathbf{b}^0_1$, $\mathbf{a}^S_1$, and $\mathbf{a}^R_1$, respectively, computed without the possible constant term arising directly from the source. Hence, for example, $\mathbf{b}^0_1(\mathbf{b}^{(l+1)}_1) - \mathbf{b}^0_1(\mathbf{b}^{(0)}_1) = \mathbf{b}^{0-}_1(\mathbf{b}^{(l+1)}_1 - \mathbf{b}^{(0)}_1)$. In this case, each $\mathbf{b}^{(l)}_1$, for l=0,1,.., must also be solved iteratively, with iterates $\mathbf{b}^{(l,l')}_1$ for l'=0,1,2,...

Starting from $\mathbf{b}^{(0,0)}_1 = (\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \cdot \left[\mathbf{b}^0_1(\mathbf{0}) + \mathbf{R}_1 \cdot (\mathbf{a}^0_1 + \mathbf{a}^R_1(\mathbf{0}))\right]$, let $\mathbf{b}^{(0,l'+1)}_1 = \mathbf{b}^{(0,0)}_1 + (\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \cdot \left[\mathbf{b}^{0-}_1(\mathbf{b}^{(0,l')}_1) + \mathbf{R}_1 \cdot \mathbf{a}^{R-}_1(\mathbf{b}^{(0,l')}_1)\right] = \mathbf{b}^{(0,l')}_1 + \Delta\mathbf{b}^{(0,l')}_1$ for l'=0,1,2,.., where

$$\Delta\mathbf{b}^{(0,0)}_1 \;=\; (\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \cdot \left[\mathbf{b}^{0-}_1(\mathbf{b}^{(0,0)}_1) + \mathbf{R}_1 \cdot \mathbf{a}^{R-}_1(\mathbf{b}^{(0,0)}_1)\right] \quad \text{and} \qquad (15)$$

$$\Delta\mathbf{b}^{(0,l')}_1 \;=\; (\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \cdot \left[\mathbf{b}^{0-}_1(\Delta\mathbf{b}^{(0,l'-1)}_1) + \mathbf{R}_1 \cdot \mathbf{a}^{R-}_1(\Delta\mathbf{b}^{(0,l'-1)}_1)\right] \quad \text{for } l' \geq 1 \;. \qquad (16)$$

For l=0,1,2,.., with $\mathbf{b}^{(l+1,-1)}_1 = \mathbf{b}^{(l)}_1$ to define the starting starting $\mathbf{b}^{(l+1,0)}_1$, let

$$\begin{aligned}\mathbf{b}^{(l+1,l'+1)}_1 \;=\; & \mathbf{b}^{(0)}_1 + (\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \cdot \left[\mathbf{b}^{0-}_1(\mathbf{b}^{(l+1,l')}_1 - \mathbf{b}^{(0)}_1) + \mathbf{R}_1 \cdot \left(\mathbf{a}^S_1(\mathbf{b}^{(l)}_1)\right.\right. \\ & \left.\left. + \mathbf{a}^{R-}_1(\mathbf{b}^{(l+1,l')}_1 - \mathbf{b}^{(0)}_1)\right)\right] \qquad (17)\end{aligned}$$

for l'=−1,0,1,2,... With $\mathbf{b}^{(l+1,l'+1)}_1 = \mathbf{b}^{(l+1,l')}_1 + \Delta\mathbf{b}^{(l+1,l')}_1$,

$$\Delta\mathbf{b}^{(l+1,0)}_1 \;=\; (\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \cdot \left[\mathbf{b}^{0-}_1(\mathbf{b}^{(l+1,0)}_1 - \mathbf{b}^{(l)}_1) + \mathbf{R}_1 \cdot \mathbf{a}^{R-}_1(\mathbf{b}^{(l+1,0)}_1 - \mathbf{b}^{(l)}_1)\right] \text{ and } (18)$$

$$\Delta\mathbf{b}^{(l+1,l')}_1 \;=\; (\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \cdot \left[\mathbf{b}^{0-}_1(\Delta\mathbf{b}^{(l+1,l'-1)}_1) + \mathbf{R}_1 \cdot \mathbf{a}^{R-}_1(\Delta\mathbf{b}^{(l+1,l'-1)}_1)\right] \qquad (19)$$

for $l' \geq 1$. Note that $\mathbf{b}^{(1,0)}_1 = \mathbf{b}^{(0)}_1 + (\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \cdot \mathbf{R}_1 \cdot \mathbf{a}^S_1(\mathbf{b}^{(0)}_1)$, and that $\mathbf{b}^{(l+1,0)}_1 = \mathbf{b}^{(l)}_1 + (\mathbf{I} - \mathbf{R}_1 \cdot \mathbf{Q})^{-1} \cdot \mathbf{R}_1 \cdot \mathbf{a}^{S-}_1(\mathbf{b}^{(l)}_1 - \mathbf{b}^{(l-1)}_1)$ for l=1,2,.., cf. [1, Eqs. (33)-(35)]. It is here, for Eq. (17) with $l' = -1$, that the typically more time-consuming wavenumber integrations with respect to κ appear.

Computation of the resulting pressure field proceeds according to [1, Sec. 3.3] with minor changes. As already mentioned, [1, Eq. (8)] is still valid (with $J_{n,0}$ replacing J_n), [1, Eq. (10)] holds with $n = 1$ immediately outside ring anomaly j=1,..,J, and Eq. (1) is relevant within ring n=2,..,$N_j + 1$ of ring anomaly j=1,..,J. According to Eq. (7), the incoming waves upon ring anomaly j=1,2,..,J now include $\mathbf{a}_{1,j}^R$ instead of $\mathbf{a}_{1,j}^S$ if $j_e(j) \neq 0$. Hence, computation of the field within the ring anomalies may now involve the adapted [1, Appendix A]. The field representation by [1, Eq. (37)] is still valid, with a natural connection to the iterates of the iterative solution procedure. It is easy to adapt the paragraph in [1] which includes [1, Eq. (47)] to the case with nested ring anomalies.

4. PARTIAL WAVES FOR MULTIPLE SCATTERING AMONG INDIVIDUAL RING AND STRIP ANOMALIES

In principle, it is easy to extend [1, Sec. 5] to the case with nested ring anomalies. An addition is that reflection-matrix recursions including restarts with vanishing matrices may now be needed not only for the strip structure, but also for ring anomalies which contain other ring anomalies. The resulting elementary reflection matrices allow transmission of partial waves to and from the included smaller ring anomalies. Note that reflections from a ring anomaly which is included in another ring anomaly j=1,2,..,J may subsequently be reflected from the outer part (with an elementary $\bar{\mathbf{R}}_{n,j}$-matrix) or from the inner part (with an elementary $\mathbf{R}_{n,j}$-matrix) of the surrounding ring anomaly j.

In [1, Sec. 5], the strips are divided into strip anomalies A, B, C, .., with single connection strips in between. A convenient way to *formally* circumvent corresponding ring division, and the mentioned reflection-matrix recursions including restarts, is to use nested ring anomalies with common centre points. This simplifies the ring-anomaly nomenclature. As an example, the outer and inner parts of the ring anomaly j mentioned in the previous paragraph could be defined as two different ring anomalies.

REFERENCES

[1] S. M. Ivansson, "Multiple sound scattering by combinations of cylindrically symmetric and linearly invariant anomalies," *J. Theor. Comp. Acoust.*, vol. 32, 2450008 (27 pages), 2024.

[2] S. M. Ivansson, "Coupled-mode field computations for underwater canyons and ridges," in *Proceedings of the 7th Underwater Acoustics Conference and Exhibition*, (Kalamata, Greece), pp. 137–144, June 25–30, 2023.

[3] S. M. Ivansson, "Multiple underwater sound scattering by cylindrically symmetric anomalies," *J. Acoust. Soc. Am.*, vol. 147, pp. 1429–1440, 2020.

[4] A. Boström, G. Kristensson, and S. Ström, "Transformation properties of plane, spherical and cylindrical scalar and vector wave functions," in *Field Representations and Introduction to Scattering* (V. V. Varadan, A. Lakhtakia, and V. K. Varadan, eds.), ch. 4, pp. 165–210, North- Holland, 1991.

Structured Session
10. New frontiers in fisheries acoustics: applications, analysis, decision and management tools

Organised by Victor Espinosa

In situ measurements of lateral target strength (TS) of Atlantic bluefin tuna (*Thunnus thynnus*) from purse seiners

Vicente Puig-Pons[1], Sergio Morell-Monzó[1], Héctor Peña[2], Isabel Pérez-Arjona[1] and Victor Espinosa[1]

[1]Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC), Universitat Politècnica de València (UPV), C/Paranimf 1, 46730 Gandia, Spain
[2]Pelagic fish group, Institute of Marine Research P. O. Box 1870 Nordnes, Bergen, Norway

Vicente Puig-Pons, Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC), Universitat Politècnica de València (UPV), C/Paranimf 1, 46730 Gandia, Spain. vipupon@upvnet.upv.es

Abstract: *Atlantic bluefin tuna (Thunnus thynnus) is a species of high economic and environmental value. To ensure its sustainable management, the International Commission for the Conservation of Atlantic Tunas (ICCAT) has established fishing quotas and a limited fishing season from May to July. The purse-seine fleet has only 40 days to catch its allocated quota, making it essential for fishermen to use efficient tools for estimating school biomass. This study proposes the simultaneous use of a calibrated omnidirectional sonar (Furuno FSV-25) and a scientific echosounder (Furuno FCV-38) to enhance biomass estimation. The results present lateral target strength (TS) measurements obtained from the scientific echosounder during the 2024 fishing season. Acoustic data were compared with length estimates derived from a stereoscopic video system used during fish transfer operations. A bimodal length distribution was observed; however, this pattern did not appear in the TS distributions. The use of a clustering algorithm to establish a relationship between acoustic energy and fish size for two distinct fish group is being explored.*

Keywords: *Bluefin tuna, target strength, scientific echosounder, underwater acoustics*

1. INTRODUCTION

Atlantic bluefin tuna is a highly valued species that achieves premium prices in international markets, particularly in Japan. However, during the last decades of the 20th century, bluefin tuna populations were severely impacted by overfishing. In response, the International Commission for the Conservation of Atlantic Tunas (ICCAT) implemented strict fishing quotas and a limited fishing season, confined to the months of May through July. As a result, the purse-seine fleet is allowed only 40 days to catch its allocated quota [1]. To optimize fishing operations and promote sustainable practices, it is essential for fishers to use accurate tools to estimate the biomass of tuna schools. Biomass estimation requires determining the total acoustic energy backscattered by the school, known as volume backscattering strength (Sv). This value is then divided by the acoustic energy backscattered by a single fish, known as the target strength (TS). The TS used is typically the average TS value calculated from all individually detected targets. The number of fish is estimated by dividing Sv by TS, and then multiplying this by the average fish weight and the estimated school volume. The final result, is the school biomass, expressed in tons [2]. The volume backscattering strength (Sv) and the school volume can be obtained from the calibrated omnidirectional sonar [3]. The average fish weight is estimated using a stereoscopic video system, as required by ICCAT regulations [1]. Finally, the mean target strength (TS) is calculated from echosounder data. As omnidirectional sonars beams insonify the fish schools horizontally, is required to have lateral TS of the target fish TS provides information about fish size, while single target detection from echosounder data also enable the assessment of fish orientation, that is, whether the fish are swimming toward or away from the vessel.

In-situ biomass estimation is one of the main objectives of the *ACTTHUN* project (ref. PID2021-127426OB-C21, funded by MICIU/AEI/10.13039/501100011033). To achieve this, we propose the simultaneous use of an omnidirectional sonar and a scientific echosounder. The ACTTHUN project is ongoing, and here we present lateral target strength (TS) measurements obtained from Atlantic bluefin tuna during the 2024 fishing season.

2. MATERIAL AND METHODS

Measurements were carried out during the 2024 bluefin tuna fishing season, between May 26 and June 13. A Furuno FCV-38 scientific echosounder was used to obtain lateral TS measurements. The echosounder operated at 38 kHz, and its transducer was mounted on the starboard side of the boat to ensure overlap between the echosounder and the omnidirectional sonar (Furuno FSV-25) beams.

The echosounder was calibrated in the harbor of L'Ametlla de Mar, prior data collection. The boat was docked, with the water depth at 5 meters. The transducer was positioned mid-water, directed toward the opposite side of the harbor, providing a 60-meter unobstructed acoustic path. Calibration followed the standard target procedure using a 38.1 mm tungsten carbide sphere. A rig was employed to position the sphere in the center of the acoustic beam during calibration [4].



Fig.1: Left: image of echosounder transducer mounted on the starboard side of the boat. Centre: image of rig during sonar calibration. Right: image of rig during echosounder calibration.

The purse-seine vessel carried out five sets between May 26 and June 10. Data were recorded in netCDF4 format. After each catch, the tuna were transferred to a transport cage, which was then towed to a fattening cages anchored near the Spanish coast. During the transfer process, tuna lengths were measured using a stereoscopic video system, following ICCAT guidelines [1].

The data were analysed using two methods (Fig.2). First, a single echo detection (SED) algorithm and a tracking algorithm were applied to process the data. The resulting traces were analysed statistically, and the outcomes were compared with the length data obtained from the stereoscopic video system during the transfer operations. In the second approach, a ping-by-ping analysis was conducted using the SED algorithm. The results from this analysis were also compared with the length measurements provided by the stereoscopic system.

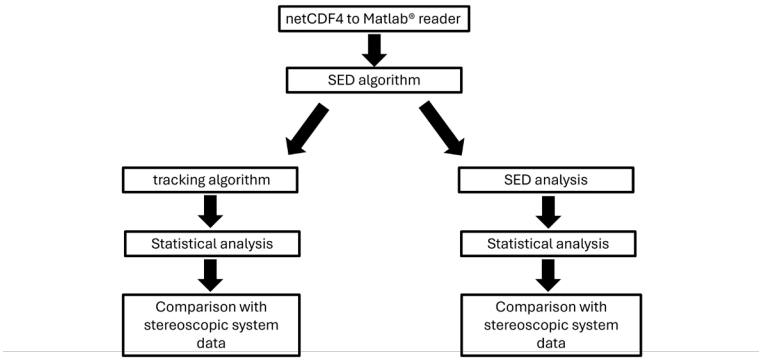


Fig.2: Flowchart for two analysis methods. Tracking method on the left and SED method on the right.

3. RESULTS AND DISCUSSION

The mean length for each catch was calculated using stereoscopic video recordings. Likewise, the mean and maximum target strength (TS) values were obtained through both trace-based and ping-by-ping analyses. The mean swimming tilt angle of the tuna was estimated from isolated traces. The results are summarized in Table 1. Typically, the mean TS is expected to correlate with the mean fish length to establish a TS-versus-length relationship. However, in this study, catches with greater mean lengths exhibited lower TS values, as seen on May 27 and May 29. As a result, it was not possible to establish a reliable TS–length relationship.

Several factors may explain this outcome: the catches on May 27 and 29 showed higher average swimming yaw angles; the number of isolated traces and analyzed pings varied between catches; and the school may have not been fully insonified in all cases.

To better understand this phenomenon, numerical simulations were performed using the Method of Fundamental Solutions (MFS) [5][6]. Being the swimbladder the most important contributor to the backscattering of the BFT, we characterized its anatomical features to develop the geometry considered in the numerical model (Fig. 3). Considering the size and swimming orientation present in the previous fishing seasons, the numerical model estimated a TS vs length relationship, to be compared with the experimental data, to better understand the *insitu* results.

	Date	*May 26*	May 27	May 29	June 1	*June 10*
	Mean length (cm)	205	206	214	178	195
TS from tracking	*Mean TS (dB)*	-19.2	-22.9	-22.2	-18.9	-20.9
	Max TS (dB)	-15.7	-19.5	-19.1	-15.8	-16.3
	Isolated traces	277	59	252	125	940
	Mean tilt angle (°)	93	108	98	96	92
TS from SED	*Mean TS (dB)*	-20.3	-21.3	-23.2	-21.2	-21.4
	Max TS (dB)	-15.6	-16.8	-18.4	-16.5	-16.2
	Pings analysed	328	158	526	449	1542

Table 1: Results from statistical analysis using two analysis methods.

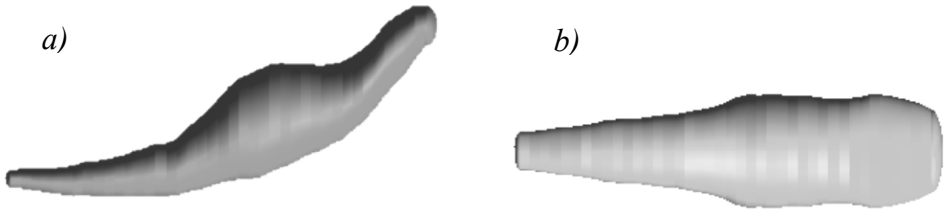


Fig. 3: a) Lateral and b) dorsal view of the BFT swimbladder model.

In Fig. 4, histograms of mean target strength (TS) values for each catch are shown. The TS distributions appear approximately normal, except for May 27 (Fig. 3b), likely due to the lower number of samples obtained. In contrast, the fish length histograms (Fig. 5) show a bimodal pattern across all catches.

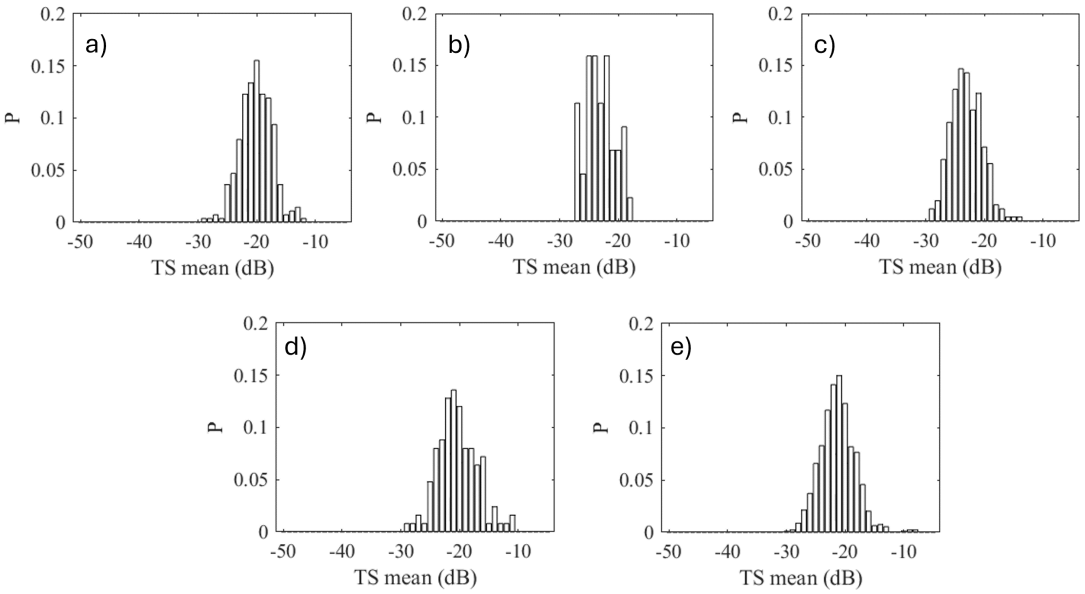


Fig.4: Histograms of mean target strength (TS) values for each catch: (a) May 26, (b) May 27, (c) May 29, (d) June 1, and (e) June 10.

To explore this further, we applied a clustering algorithm to divide the TS distributions into two groups, with the aim of associating each mode of the length distributions with one of these TS clusters. This analysis is ongoing and will be applied to the 2024 data and tested during the 2025 fishing season.

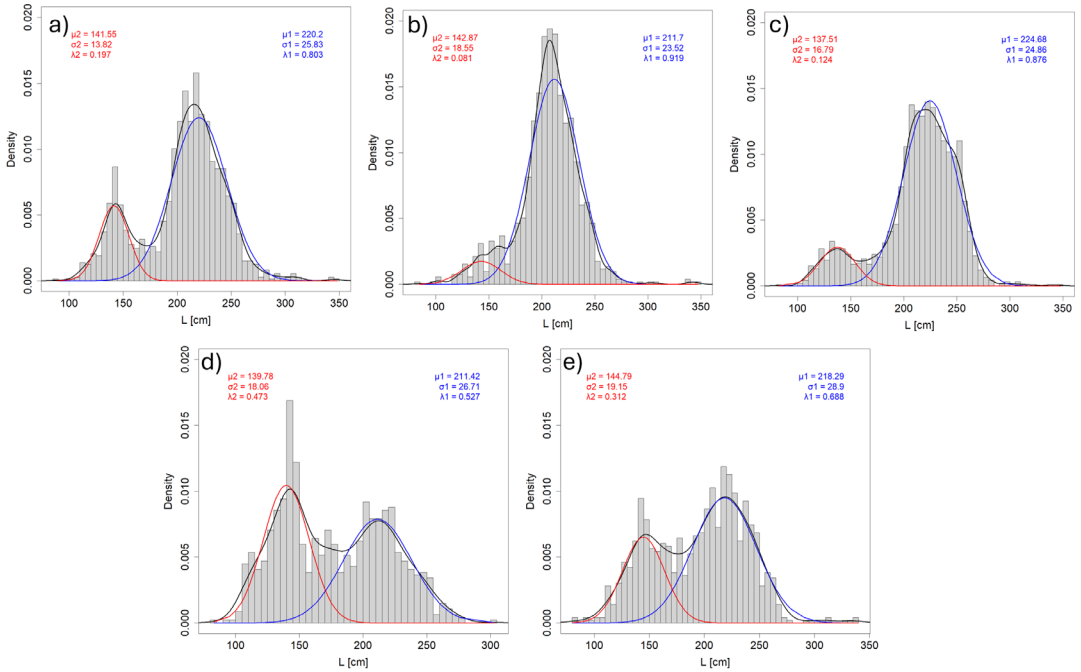


Fig.4: Histograms of fish length values for each catch: (a) May 26, (b) May 27, (c) May 29, (d) June 1, and (e) June 10.

4. CONCLUSIONS

This study presents *in situ* measurements of the lateral target strength (TS) of Atlantic bluefin tuna (*Thunnus thynnus*) obtained during the 2024 fishing season using a scientific echosounder. Two data analysis methods—trace-based tracking and ping-by-ping SED—were applied and compared against fish length measurements from a stereoscopic video system. Despite expectations, a consistent TS–length relationship could not be established, as longer fish exhibited lower TS values. This discrepancy may be attributed to varying swimming tilt angles, differences in sample sizes, and incomplete insonification of the schools. To further investigate these anomalies, numerical simulations were conducted using the Method of Fundamental Solutions (MFS), incorporating anatomical features of the swimbladder. In addition, histograms revealed a mismatch between unimodal TS distributions and bimodal length distributions. A clustering algorithm will be explored to separate TS values into two groups potentially corresponding to the observed length modes. This approach will be further evaluated during the 2025 fishing season.

5. ACKNOWLEDGEMENTS

This work was supported by funding from ACTHHUN project grant ref. PID2021-127426OB-C21 was funded by MICIU/AEI/10.13039/501100011033. The Institute of Marine Research are thanked for financial and technical support to H.P through their internal ICCAT-Atlantic bluefin tuna project 15615.

REFERENCES

[1] **ICCAT.** Recommendation by ICCAT amending the Recommendation 18-02 establishing a multi-annual management plan for Bluefin tuna in the eastern Atlantic and the Mediterranean.2019

[2] **Simonds J, MacLennan D.** Fisheries Acoustics: Theory and Practice Second Edition. Second edi. edited by D. Simonds, J, MacLennan. Blackwell Science. 2007.

[3] **Héctor Peña, Gavin J Macaulay, Egil Ona, Sindre Vatnehol, Arne J Holmin.** Estimating individual fish school biomass using digital omnidirectional sonars, applied to mackerel and herring, *ICES Journal of Marine Science*, Volume 78, Issue 3, July 2021, Pages 940–951, https://doi.org/10.1093/icesjms/fsaa237

[4] **Sindre Vatnehol, Atle Totland, Egil Ona**. Two mechanical rigs for field calibration of multi-beam fishery sonars, *Methods in Oceanography*,Volumes 13–14,2015,Pages 1-12,ISSN 2211-1220, https://doi.org/10.1016/j.mio.2016.02.001

[5] **[5] Pérez Arjona, I.; Godinho, L.; Espinosa Roselló, V.** Numerical simulation of target strength measurements from near to far field of fish using the method of fundamental solutions, Acta Acustica united with Acustica, 2018. 104(1):25-38. https://doi.org/10.3813/AAA.919142

[6] **I. Pérez-Arjona, L. Godinho, V. Espinosa**, Influence of fish backbone model geometrical features on the numerical target strength of swimbladdered fish, *ICES Journal of Marine Science*, Volume 77, Issue 7-8, December 2020, Pages 2870–2881, https://doi.org/10.1093/icesjms/fsaa160

Advances in biomass estimation of Nile tilapia aquaculture with acoustic methods

Luis Lorenzo Carrillo La Rosa[1], Vicente Puig-Pons[2], Sergio Morell-Monzó[2], Isabel Pérez-Arjona[2], and Víctor Espinosa[2]

[1]Departamento Académico de Manejo Pesquero y Medio Ambiente, Universidad Nacional Agraria La Molina, Lima, Perú
[2]Instituto de Investigación para la Gestión Integrada de Zonas Costeras (IGIC), Universitat Politècnica de València, EPS Gandia, Gandia, Spain

Abstract: *Nile tilapia (Oreochromis niloticus) is one of the most farmed fish species in the world with more than 5 million tons per year and is expected to continue to grow in the future. Traditional size monitoring and biomass estimation focuses on catch-and-release sampling methods. Acoustic methods have proven to be a potential tool for aquaculture monitoring. However, scientific echo sounders remain unfeasible in aquaculture due to their high cost. This work shows progress towards monitoring the biomass of Nile tilapia under production-like conditions from lateral target strength (TS) with a low-cost echo sounder. First, a biometric length-weight model was fitted from real specimens. Then, Biosonics DT-X echo-sounder with a single beam transducer operating at ~200 kHz was employed. The performance in estimating the average length through TSmax and TSmean was assessed for 8 size groups (from 14.1 to 43.7 cm) obtaining a* $R^2 > 0.9$*. These results are promising and provide an approach to the use of low-cost methods for biomass monitoring in aquaculture.*

Keywords: *underwater acoustics, lateral target strength, Oreochromis niloticus, single-beam echosounder*

1. INTRODUCTION

Nile tilapia (*Oreochromis niloticus*) is one of the most widely farmed fish species globally due to its fast growth rate, adaptability to various production systems, and high market demand. According to the Food and Agriculture Organization (FAO), global Nile tilapia production reached 5.3 million tons in 2024 [1]. Its rapid growth necessitates regular monitoring of key parameters such as body length, weight, and stocking density to ensure optimal farming performance. In recent years, video-based systems have been increasingly applied to estimate fish length and support a range of aquaculture management tasks. However, these systems often face limitations under poor visibility conditions. In this context, active acoustic systems have emerged as a promising non-invasive alternative for monitoring fish growth in aquaculture.

A growing body of research has explored the use of acoustic techniques for fish monitoring in cages and tanks [2,3,4]. These studies have demonstrated that target strength (TS)—defined as the amount of acoustic energy backscattered by an individual fish [5]—can be correlated with fish length when measured from different orientations (ventral or lateral). The backscattered TS signal depends on multiple factors, including fish size, the presence or absence of a swim bladder, and the orientation of the fish relative to the transducer.

The work [6] showed that a reliable relationship between TS and fish length can be established for Nile tilapia using a scientific echosounder. In their study, tilapia specimens of varying sizes were suspended in a laboratory tank, allowing for controlled measurement conditions. While scientific echosounders can compensate for TS variability due to the fish's position within the acoustic beam, their high cost and complexity limit their operational feasibility for routine use in commercial aquaculture.

In this study, we investigate the relationship between lateral TS, fish length, and weight in Nile tilapia (grey tilapia) using a single-beam echosounder operating at ~200 kHz. Unlike previous studies with restrained fish, our experimental setup involved freely swimming individuals, providing conditions more representative of production environments. The Biosonics DT-X echosounder is qualified as a scientific system, but we limited its performance by using a single beam transducer that was pre-calibrated. This configuration can be replicated with cost-effective quantitative echosounders and offers a significantly lower-cost alternative to scientific systems, thereby enhancing the potential for broader adoption in commercial tilapia farming.

2. DATA AND METHODS

2.1. Biometric measurements and relationships

Biometric data were collected at the Fish Research Center (Centro de Investigación Piscícola, CIMPIS) of Universidad Nacional Agraria La Molina, located in Lima, Peru (12°05′03″ S, 76°56′44.8″ W), over a 13-month period from December 2018 to December 2019. The measurements were conducted in a concrete pond measuring 4.0 × 4.0 × 1.5 m.

A total of Oreochromis niloticus specimens with total lengths (TL) ranging from 13 cm to 44 cm were selected and divided into eight experimental groups. Within each group, the maximum deviation from the mean TL was restricted to ±2 cm to ensure homogeneity in

body size. The number of individuals per group ranged from 6 to 20, depending on fish availability. A detailed summary of the group composition is presented in Table 1. Total length (L) was measured to the nearest millimetre using a fish measuring board (ichthyometer), and individual body weight (W) was determined using a digital precision scale.

Group	Fish number	Mean L [cm]	Min L [cm]	Max L [cm]	Mean W [g]
G1	20	14.1	13.0	15.0	47.0
G2	15	17.9	17.0	19.0	99.4
G3	14	21.5	21.0	22.0	179.9
G4	14	24.6	24.0	25.0	230.1
G5	6	30.3	29.0	31.0	540.0
G6	6	35.1	34.0	36.0	758.3
G7	6	40.8	40.0	41.5	1383.3
G8	6	43.7	43.0	44.0	1445.8

Table 1. Data summary of each size group obtained.

Our previous results presented in [7] showed the relationship between Lt (total length) and Ls (standard length or fork length) and weight for this population of *Oreochromis niloticus* (Figure 1). We define the following biometric models: $W = 0.0265 \cdot Lt^{2.8465}$ and $W = 0.5123 \cdot Ls^{2.81193}$ with determination coefficients (R^2) of 0.982 y 0.977 respectively. Therefore, we decided to use Lt as a measure of length (so in this study L = Lt).

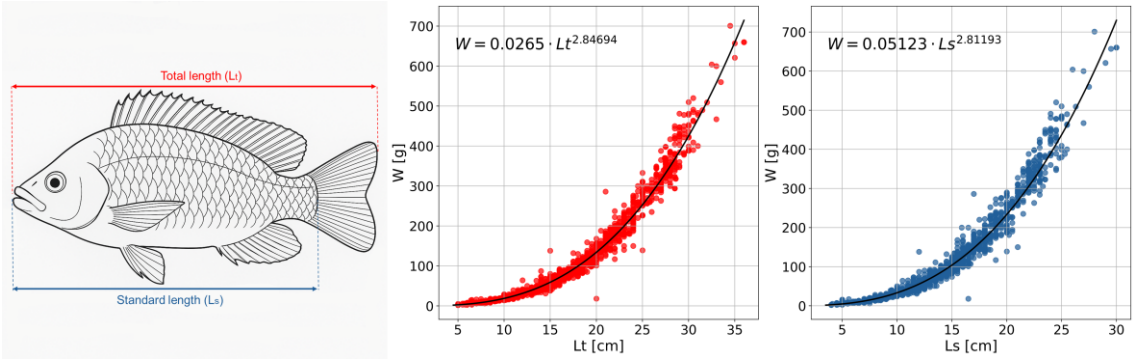


Fig.1: Length-weight relationships for Nile tilapia (Oreochromis niloticus) described in Carrillo La Rosa et al. 2025.

2.2. Acoustic data collection

Acoustic measurements were conducted using a BioSonics DT-X echosounder equipped with a single-beam transducer operating at 201 kHz. The transducer had a nominal beam width of 9.9° and was mounted horizontally at one end of the pond, such that the acoustic beam was directed parallel to the water surface. This configuration allowed for the acquisition of lateral target strength (TS) measurements, providing insight into fish orientation and behaviour under conditions similar to those in production environments.

The acoustic system was configured with a pulse duration of 0.1 ms, a ping rate of 5 pings·s^{-1}, and a source level of 202.3 dB re 1 µPa at 1 m. The receiver sensitivity was –64.9 dB. These settings were selected to optimize the detection of individual fish targets while minimizing interference and multipath reflections. A summary of the system configuration is provided in Table 2.

For each size group, data were collected over 2 to 5 consecutive days, as detailed in Table 1. During acoustic sampling, fish swam freely within the pond, allowing for the observation of natural swimming behaviour and orientation within the acoustic beam. Data acquisition was performed using BioSonics® Visual Acquisition software. Recordings were recorded in 30-minute files to facilitate post-processing and analysis.

Special care was taken to minimize acoustic noise and unwanted reflections. Measurements were conducted under calm water conditions to reduce signal degradation due to surface turbulence or air bubbles. The choice of lateral orientation, although less commonly used than the ventral approach, was made to reflect the typical horizontal posture of tilapia in shallow water environments and to evaluate the feasibility of TS-based length and weight estimation under realistic aquaculture conditions.

2.3. Acoustic data analysis

Acoustic data were processed and analysed using Sonar5-Pro software [8]. Echograms were transformed and filtered using a –95 dB threshold to enhance target image quality and reduce background noise. Given the use of a single-beam transducer, the 2D single echo detector (SED) algorithm was applied to detect individual fish echoes.

The validated TS data from Sonar5-Pro were exported for statistical analysis using Statgraphics Centurion XIX. Relationships were evaluated between both the mean and maximum TS values and the total length (L) of the fish. A logarithmic linear regression model was applied, following the commonly used form:

$$TS = a \cdot \log(L) + b \tag{1}$$

where a represents the slope, b is the intercept, and TL is the total length in centimetres. Although it is standard practice to apply a fixed slope (typically around 20), in [9] it was shown that allowing both a and b to vary provides better fits for TS–L relationships. In this study, both parameters were estimated empirically based on lateral TS measurements and the biometric data collected for each group.

3. RESULTS AND DISCUSSION

The results obtained are summarized in Table 2. For each size group, between 1,402 and 8,696 high-quality echo traces were manually isolated and retained for analysis. The lowest number of traces was recorded in the smallest group which is Group 1 (mean TL=14.1 cm), likely due to weaker echo returns and lower detectability. Conversely, the highest number of traces was obtained for Group 3 (mean TL=21.5 cm), which yielded more consistent and detectable signals.

For each group, the mean, median, and standard deviation were computed from the distributions of both mean TS and maximum TS. Mean TS values ranged from –42.2 dB to –34.8 dB, while maximum TS values ranged from –41.0 dB to –32.0 dB (Table 2).

There is a clear positive correlation between the average TS levels and the average length of the group. Figure 2 shows the scatterplot of the averaged TS (both TS_{mean} and TS_{max}) and the averaged L of each group and the regression analysis. Both regression models showed $R_2 > 0.90$ indicating strong predictive power.

Group	Mean TL [cm]	TS_{mean} (dB)			TS_{max}(dB)			Number of traces
		Average TS_{mean}	Median TS_{mean}	Standard deviation	Average TS_{max}	Median TS_{max}	Standard deviation	
1	14.1	-44.2	-44.0	3.4	-41.3	-41.0	3.5	1402
2	17.9	-43.7	-43.5	3.6	-40.8	-40.6	3.8	2240
3	21.5	-41.7	-40.6	5.7	-39.6	-38.6	6.0	8696
4	24.6	-41.5	-41.4	3.4	-36.9	-37.1	3.7	1698
5	30.3	-38.6	-38.7	4.0	-36.4	-36.4	4.0	5088
6	35.1	-38.3	-38.5	3.4	-35.9	-36.3	3.7	2291
7	40.8	-34.9	-34.4	4.5	-32.5	-31.9	4.6	8612
8	43.7	-34.8	-34.7	3.8	-32.1	-32.0	3.8	5400

Table 2. Averaged TS_{mean}, TS_{max} and its standard deviation for each group size of Nile tilapia

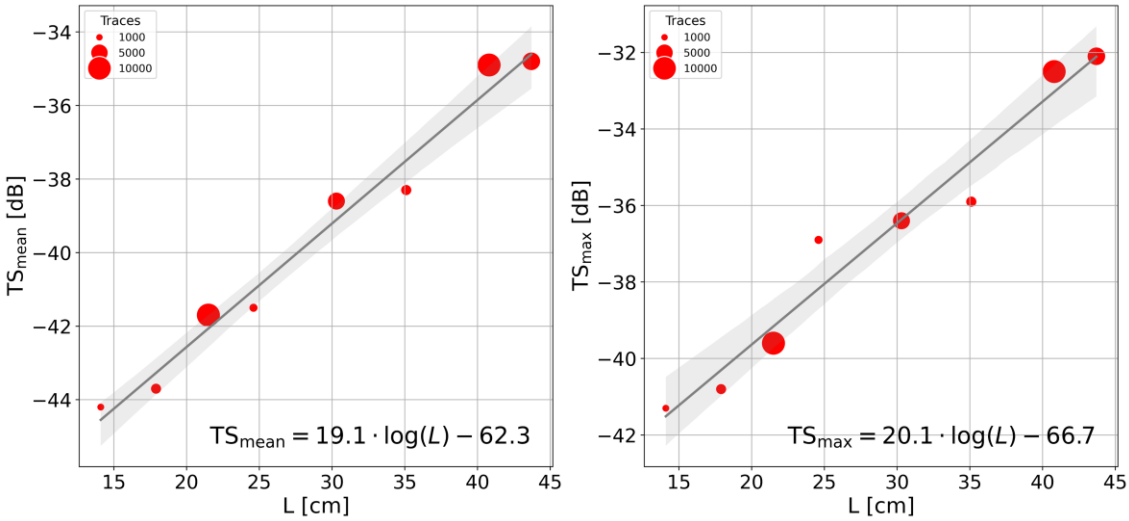


Fig.1: Regression analysis between the averaged TS_{mean} or averaged TS_{max} of each group and the averaged L. The size of the dots is proportionally large to the number of traces in that group. The grey region represents the 95% confidence interval. (left) shows the relationship between TS_{mean} and L with an R^2 of 0.94. (right) shows the relationship between TS_{max} and L with an R^2 of 0.91.

The results obtained are promising and confirm a strong predictive power of the mean length of a Nile tilapia population through the lateral target strength, especially through the TS_{mean}. This study should be complemented by direct estimation of weight through the TS to determine the biomass of the stock. Additionally, the results should be validated with simulations using 3D digital models of *Oreochromis niloticus'* swimbladder. This work provides the first experience in the estimation of TS vs L in free-swimming tilapia, which resembles real production conditions.

4. ACKNOWLEDGEMENTS

This study is part of the ThinkInAzul programme and was supported by MCIN with funding from European Union NextGenerationEU (PRTR-C17.I1) and by Generalitat Valenciana (GVA-THINKINAZUL/2021/009; Principal Investigator: V. Espinosa Roselló, Universitat Politècnica de València).

REFERENCES

[1] **FAO - Food and Agriculture Organization of the United Nations,** *The State of World Fisheries and Aquaculture 2024*, FAO, Rome, pp. n/a, 2024. https://doi.org/10.4060/cd0683en

[2] **Knudsen, F. R., Fosseidengen, J. E., Oppedal, F., Karlsen, Ø., & Ona, E.,** Hydroacoustic monitoring of fish in sea cages: target strength (TS) measurements on Atlantic salmon (*Salmo salar*), *Fisheries Research*, 69(2), pp. 205–209, 2004. https://doi.org/10.1016/j.fishres.2004.05.008

[3] **Rodríguez-Sánchez, V., Rodríguez-Ruiz, A., Pérez-Arjona, I., & Encina-Encina, L.,** Horizontal target strength-size conversion equations for sea bass and gilt-head bream, *Aquaculture*, 490, pp. 178–184, 2018. https://doi.org/10.1016/j.aquaculture.2018.02.034

[4] **[4] Puig-Pons, V., Muñoz-Benavent, P., Pérez-Arjona, I., Ladino, A., Llorens-Escrich, S., Andreu-García, G., Valiente-González, J. M., Atienza-Vanacloig, V., Ordóñez-Cebrián, P., Pastor-Gimeno, J. I., & Espinosa, V.,** Estimation of Bluefin Tuna (*Thunnus thynnus*) mean length in sea cages by acoustical means, *Applied Acoustics*, 197, pp. 108960, 2022. https://doi.org/10.1016/j.apacoust.2022.108960

[5] **Simonds, J. & MacLennan, D.,** *Fisheries Acoustics: Theory and Practice*, 2nd ed., edited by Simonds, J. & MacLennan, D., Blackwell Science, pp. n/a, 2007.

[6] **Liu, J., Setiazi, H., & Borazon, E. Q.,** Hydroacoustic assessment of standing stock of Nile tilapia (*Oreochromis niloticus*) under 120 kHz and 200 kHz split-beam systems in an aquaculture pond, *Aquaculture Research*, 53(3), pp. 820–831, 2021. https://doi.org/10.1111/are.15618

[7] **La Rosa, L. L. C., Morell-Monzó, S., Puig-Pons, V., Pérez-Arjona, I., & Espinosa, V.,** Biometric relationships and condition factor of Nile tilapia (Oreochromis niloticus) grown in concrete ponds with groundwater, Aquaculture International, 33(3), pp. n/a, 2025. https://doi.org/10.1007/s10499-025-01839-7

[8] **Balk, H. & Lindem, T.,** *Sonar4 and Sonar5-Pro Post Processing Systems: Operator Manual Version 6.0.3*, Balk, H. & Lindem, T. (eds.), Oslo, Norway, pp. n/a, 2011.

[9] **[9] McClatchie, S., Macaulay, G. J., & Coombs, R. F.,** A requiem for the use of 20 $\log_{10}$ Length for acoustic target strength with special reference to deep-sea fishes, *ICES Journal of Marine Science*, 60(2), pp. 419–428, 2003. https://doi.org/10.1016/s1054-3139(03)00004-3

Characterisation of aggregated Baltic herring - technique for assessment of possible vessel avoidance reaction

Aleksander Żytko[1], Natalia Gorska[1], Dezhang Chu[2] and Beata Schmidt[3]

[1] Institute of Oceanology PAS, Powstańców Warszawy 55, 81-712 Sopot, Poland,
[2] former NOAA employee
[3] National Marine Fisheries Research Institute, ul. Kołłątaja 1, 81-332 Gdynia, Poland,

Aleksander Żytko, Institute of Oceanology, Polish Academy of Sciences, Powstańców Warszawy 55, 81-712 Sopot, Poland. azytko@iopan.pl

Abstract: *The orientation of herring individuals significantly influences their backscattering properties, which are crucial for accurate hydroacoustic biomass assessments. Vessel-induced avoidance reactions may alter fish orientation distributions during biomass measurements, as observed in many species. However, according to our best knowledge, there is a gap in information regarding the orientation of aggregated Baltic herring as well as its reaction on the approaching vessel during the biomass estimation campaigns.*

To address this gap, we developed an inverse method for assessing the orientation of aggregated Baltic herring. The method compares histograms of herring target strength (TS) obtained numerically for the different fish orientation distributions with TS data collected during routine hydroacoustic biomass assessments in the frame of Baltic International Acoustic Survey (BIAS) cruises. This technique incorporates a novel backscattering model, improving the accuracy of backscattering predictions for Baltic herring. The orientation distributions for herring escaping from the vessel during the herring biomass estimation have been defined. The approach provides a more detailed interpretation of acoustic data, facilitates the characterization of Baltic herring, and enables the assessment of herring responses to passing research vessels.

Keywords: *Target Strength, fish orientation, Baltic Herring.*

1. INTRODUCTION

The realibility of hydroacoustic biomass assessments of Baltic herring (*Clupea harengus*) depends on the accuracy of the used fish target strength (TS) [1], which strongly depends on fish orientation [1 - 3]. Avoidance behavior in response to approaching vessels can alter orientation distributions, potentially biasing biomass estimates [4,5]. Despite its importance, orientation of aggregated Baltic herring is not taken into account during standard acoustic surveys, Baltic International Acoustic Survey (BIAS), carried out every year as the initiative of International Council for the Exploration of the Sea (ICES).

To address this gap, we developed an indirect inverse method to estimate Baltic herring orientation distributions using routine hydroacoustic data [6]. Unlike the existing techniques, this method is low-cost and does not require complex equipment or data collection procedures as well as sophisticated data analysis.

Our approach compares TS histograms from BIAS empirical data with simulated TS distributions generated under various orientation scenarios. In [6] the unimodal herring orientation distribution was considered as the scenario. In this paper the two-modal herring orientation distribution were used and the comparison with the previous approach was done.

2. METHOD

In order to estimate the herring orientation distribution, an algorithm developed in Żytko et al. (2025) [6] was used. The algorithm is described in detail in the publication so below we present only its main idea.

The algorithm is based on comparing the measured TS distributions with the synthetic TS distributions generated for herring aggregations characterized by different orientation distributions. For the TS calculations used in this algorithm, a modified backscattering model was developed that does not have the limitations of the previously used models [6]. The model extended the scope of application of the prolate spheroid resonance model [7] for Baltic herring to the transitional region so as to be in agreement with the Kirchoff approximation. The algorithm is applied to the individual hauls made during the BIAS cruises using biological and TS data collected for each haul.

2.1. Synthetic TS histograms

To estimate the orientation of the herring the numerical TS histograms should be generated for the considered haul. For this, firstly a herring aggregations were generated numerically. In aggregation, each individual was characterized by parameters such as fish length, depth of occurrence, and orientation angle θ (defined as the angle between the fish's longitudinal body axis and the horizontal plane). The herring total length distribution was taken for the selected haul. The fish distribution over the depth was obtained from the echogram corresponding to this haul. For each haul, 100 realizations of the aggregation were generated.

In opposite to the approach presented in [6], two types of herring orientation distributions were considered: unimodal (we use the name “unimodal” for this approach)

and bimodal (bimodal approach is used below to define this approach). The Gaussian unimodal orientation distribution was assumed:

$$\Phi(\theta) = \frac{1}{s_\theta \sqrt{2\pi}} e^{\left(\frac{-(\theta-\bar{\theta})^2}{2{s_\theta}^2}\right)}, \quad (1)$$

where $\bar{\theta}$ – mean orientation angle and s_θ – standard deviation. positive angle indicates the position of the fish with its head up, and negative angle for its head down.
Bimodal orientation distributions, composed of two Gaussian components, were described by five parameters: the mean angles and standard deviations of each component, as well as the proportion of the population belonging to the first component. The mean orientation angles varied within the range from –40 ° to 0°, and the standard deviations ranged from 3° to 15°.

2.2. Comparing measure and numerical TS distributions

Modeled TS distributions were computed for various herring orientation distribution parameters and stored as a multidimensional matrix. For the unimodal approach, the matrix was two-dimensional, with dimensions corresponding to the varying mean orientation angle and standard deviation. In the case of the bimodal approach, the matrix was five-dimensional, accounting for two mean angles, two standard deviations, and the proportion between the two modes.

For each haul, the modeled TS distributions were compared to the measured TS distribution using χ^2 test [8]. The modeled TS distribution with the smallest χ^2 distance was considered the most similar to the measured TS distribution, and the corresponding orientation distribution was therefore assumed to be the most likely.

2.3. Data

The acoustic and biological data used in the algorithm were collected during the annual BIAS cruises, conducted by the National Marine Fisheries Research Institute (NMFRI) in Gdynia, aboard the R/V *Baltica* within the Polish Exclusive Economic Zone. Specifically, we utilized data from the autumn BIAS cruises carried out in 2010 [9].

Acoustic data were collected using a Simrad EK60 echosounder equipped with a 38 kHz split-beam transducer, hull-mounted and downward-facing. Single target strength data as well as target depth data were extracted using SonarData Echoview software (version 4.90.81.19054). To filter out information that is not connected with herring, a cutoff at the -55 dB threshold was applied to the measured TS histogram.

Biological data were obtained from hauls and included species composition and fish length distributions. For the purposes of this analysis, we selected four hauls (z7, z25, z26, z27) with a herring content exceeding 90%, ensuring that the acoustic signals could be confidently attributed to herring aggregations.

3. RESULTS AND DISCUSSION

The algorithm was applied to four hauls. In Fig. 1 a comparison of the results for the unimodal and bimodal approach for the example haul z27 is shown. The bimodal approach yields better matches - the χ^2 value for this case ($\chi^2 = 0.119$) is smaller than for the unimodal orientation distribution case ($\chi^2 = 0.164$). This better fit is also confirmed by comparison of the measured and modeled TS-histograms in Fig.1a and 1b. In the bimodal case (Fig. 1b) the fit is better.

In terms of orientation distributions, the unimodal approach yields for this haul: $\bar{\theta}$ = -20° and s_θ = 5° (Fig. 1c). In the case of the bimodal approach (Fig. 1d) one mode is similar to this mode with mean angle and standard deviation $\overline{\theta_1}$ = -20°, $s_{\theta 1}$ = 3° respectively. The second mode is with a steeper mean angle $\overline{\theta_2}$ = -40° ($s_{\theta 2}$ = 3°). The first mode includes 89% of individuals of the modeled herring aggregation. The second mode can be interpreted as corresponding to fish escaping into deeper water in response to the approaching vessel [4,5].

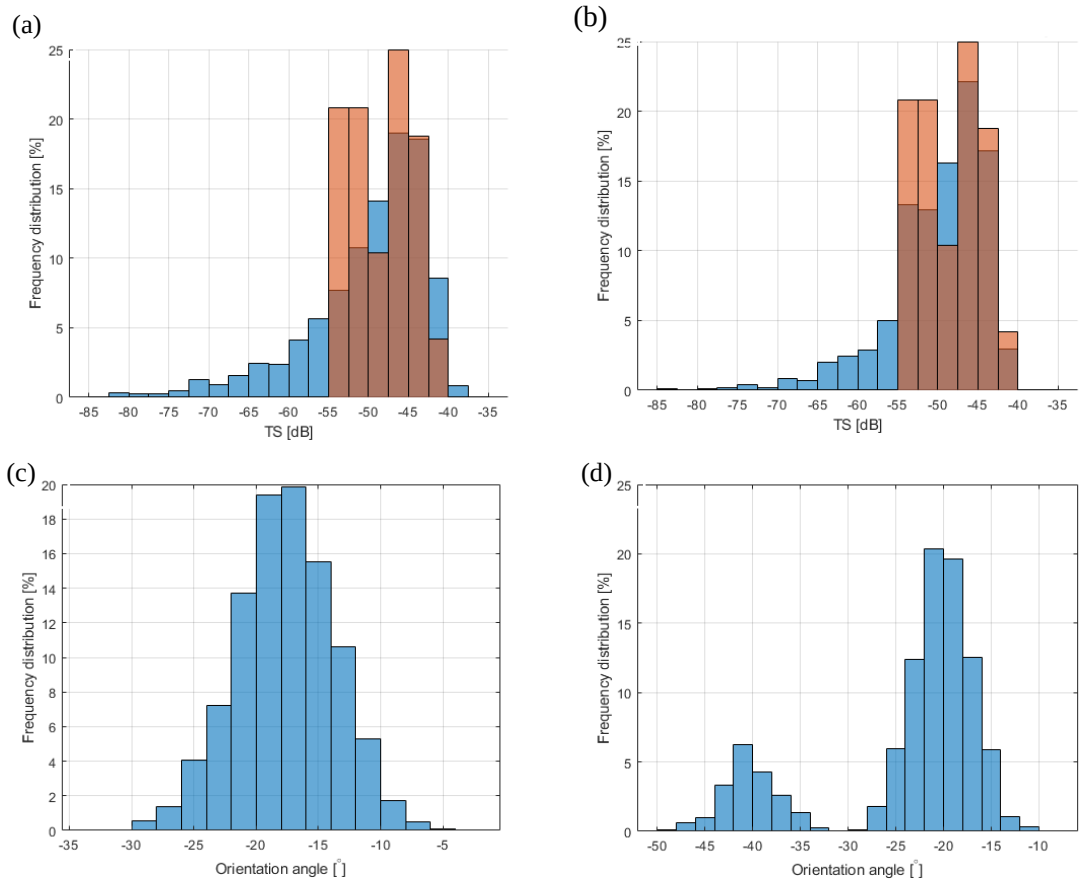


Fig.1: Comparison of distribution of TS measured in the z27 haul (orange) and the closest modelled distribution (blue) for unimodal (a) and bimodal (b) approach. Unimodal (c) and bimodal (d) orientation distributions obtained using the algorithms are presented.

Fig.2 shows the boxplot of the shortest $\chi 2$ distances between the modeled and the measured TS distributions for unimodal (blue lines) and bimodal (red lines) approaches. In

case of the bimodal approach, χ^2 distance are smaller than in case of the unimodal approach for all considered hauls.

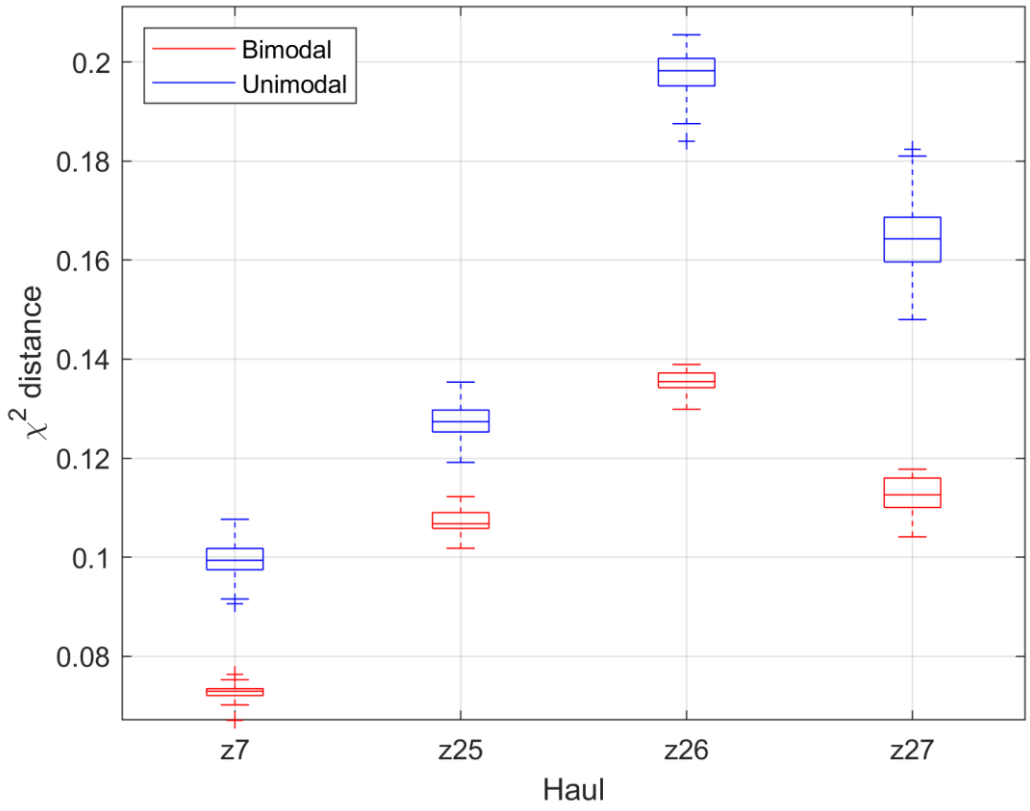


Fig.2: Comparison of the shortest χ^2 distance for unimodal and bimodal approach for four analysed hauls. On each box, the central mark indicates the median and the bottom and top edges indicate the 25th and 75th percentiles, respectively.

The occurrence of complex, multimodal orientation distributions can be explained by the variability of fish avoidance reaction in response to the approaching vessel. It is plausible that aggregation can split into parts, one of which escapes into the depths from the noise of the ship (diving), and the other swims in an unchanged manner or escapes horizontally (lateral escape) [4,5]. Such behavior would naturally result in a broader and more complex distribution of tilt angles.

Alternatively, the data may reflect the presence of multiple aggregations encountered sequentially during the sampling period - a single haul is likely to contain information from more than one aggregation. As a result, the orientation distributions may exhibit multiple modes, each corresponding to different groups of fish or distinct behavioral phases of the same aggregation.

Even if the true orientation distribution is unimodal, it may deviate significantly from a standard Gaussian shape. In such cases, a better fit can be achieved using more flexible unimodal models that account for asymmetry or excess kurtosis. Therefore, the assumption that a single Gaussian distribution can adequately describe the entire dataset may not be valid. In the bimodal approach, the two modes of the orientation distribution can overlap,

forming a more complex distribution shape that better captures the observed variability and addresses this limitation.

4. CONCLUSIONS

Applying the bimodal approach to the previously developed algorithm for the determination of Baltic herring orientation distribution provided a more accurate result than in case of the unimodal approach. The bimodal fish orientation distributions, composed of two modes, give more detailed information than a single broad mode in unimodal distribution. The bimodal approach appears to yield more accurate and plausible results and represents a step forward in the development of the method.

ACKNOWLEDGEMENTS

The data used in this paper were collected by the National Marine Fisheries Research Institute in Gdynia (R/V Baltica) in the frame of the Polish component of the BIAS surveys under the EU Data Collection Framework (CDF).

REFERENCES

[1] **Simmonds, J., MacLennan, D.N**., *Fisheries Acoustics: Theory and Practice. 2nd Edn*., Blackwell Publ., London, 2005.
[2] **Fässler, S.M.M., Gorska, N.,** On the target strength of Baltic clupeids, *ICES J. Mar. Sci.* 66, 1184–1190, 2009, https://doi.org/10.1093/icesjms/fsp005.
[3] **Gorska, N., Idczak, J**., On Baltic herring morphometry and its impact on the backscattering properties. *Oceanologia*, 64(1), 198-211, 2021, https://doi.org/10.1016/j.oceano.2021.10.001.
[4] **De Robertis, A., Handegard, N. O.**, Fish avoidance of research vessels and the efficacy of noise-reduced vessels: a review. *ICES J. Mar. Sci.*, 70(1), 34-45, 2013
[5] **Evans, T. M., Rudstam, L. G., Sethi, S. A., Warner, D. M., Hanson, S. D., Turschak, B., Esselman, P. C.** Fish avoidance of ships during acoustic surveys tested with quiet uncrewed surface vessels. *Fisheries Research*, 267, 106817, 2023.
[6] **Żytko, A., Gorska, N., Chu, D., and Schmidt, B.**, Hydroacoustic technique for determination of the orientation of aggregated Baltic herring, *Oceanologia*, 67(2), 2025.
[7] **Scoulding, B., Chu, D., Ona, E., Fernandes, P.G.**, Target strengths of two abundant mesopelagic fish species. *J. Acoust. Soc. Am.* 137(2), 989–1000, 2015. https://doi.org/10.1121/1.4906177
[8] **Pele, O., Werman, M.,** The quadratic-chi histogram distance family. [in:] Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010 Proceedings, Part II, Springer, Berlin, Heidelberg, 749–762, 2010.
[9] **Grygiel, W., Łączkowski, T., Podolska, M., Wodzinowski, T.**. Research report from the Baltic International Acoustic Survey (BIAS) on board of the Polish r.v. Baltica (20.09–08.10.2010). Working paper on the WGBIFS meeting in Kaliningrad (Russia), 21–25.03.2011. [in:] ICES CM 2011/SSGESST:05, REF. SCICOM, WGISUR, ACOM; Annex 9; 396–429, 2011.

Using open FEM/BEM software for modeling fish target strength

Marek Moszyński[1], Isabel Pérez Arjona[2], Víctor Espinosa Roselló[3], and Anderson Ladino Velásquez[4]

[1]Gdańsk University of Technology, marek.moszynski@pg.edu.pl
[2]Universidad Politécnica de Valencia, iparjona@upvnet.upv.es
[3]Universidad Politécnica de Valencia, vespinos@fis.upv.es
[4]Universidad Politécnica de Valencia, anlave@doctor.upv.es

Marek Moszyński, ul. Gabriela Narutowicza 11/12, 80-233 Gdańsk, Poland, marek.moszynski@pg.edu.pl

Abstract: *The advance in computer modeling techniques observed in last decades allows for creating so-called computer twins in variety of real-world applications. One of the well-developed technique that found to be extremely useful in variety of acoustic areas is based on using Finite Element Method and Boundary Element Method. Both has been already commercialized in the form of advanced software packages that facilitate verification of whole designs. However in many research areas a ready to use commercial software are still not prepared for research calculations. Hence the idea of using programmable methods that allows for the insight into internal workings of the computer model with a usage of a methods dedicated for Finite/Boundary Element approach. Among many of long-developed open-source packages are FreeFEM and Bempp that is characterized by maturity in implementing of sophisticated numerical algorithms and giving higher abstraction level in software programming than modern object oriented languages. To illustrate the proposed idea the modeling of fish target strength pattern is presented. The examples are mainly oriented for using it in statistical calculation of target strength vs. fish length relation.*

Keywords: *Finite Element Method, Boundary Element Method, FreeFEM, Bempp, fish target strength, yellowfin tuna, mackerel*

1. BACKGROUND

The Boundary Element Method (BEM) has been widely used to predict scattering by fish. Among its advantages, is that this method can handle complex geometries and the solution space is one dimension lower than that of the scatterer object since field values over surfaces instead of field values over volumes are required [2]. Several software packages support BEM calculations. Over the past decade, while commercial solutions have become well-established, open-source alternatives have also matured significantly. In the paper two programmable solutions were verified using scattering by fish as a main application area, namely FreeFEM [3] and Bempp [1]. The first one, is interfaced with the boundary element library BEMTool, which is a general purpose header-only C++ library. It also uses modern technique based on representing matrices in the form of compressed hierarchical matrix stored in specialized H-matrix format. The second one is an open-source computational boundary element platform with Python interface, that uses PyOpenCL for just-in-time compilation of BEM kernels. It allows for easy formulation of acoustic and electromagnetic transmission problems, what is especially valuable when applying for fish scattering.

Modelling of the acoustic scattering from a single fish is a problem which has been extensively studied over the past decades. Since the swimbladder plays a fundamental role from the acoustic viewpoint, many models for the evaluation of scattering by individual fish, exclusively consider only swimbladder as a non-penetrable body. However for non-swimbladdered fish the fish body modeled as fluid filled geometry requires the solution of a more complicated transmission problem.

2. THE EXCERPT FROM THEORY OF ACOUSTIC SCATTERING

The target strength of an object is the logarithmic acoustic measure of its size and could be defined as:

$$TS = 20log\frac{p_{scat}(r=1,\theta=\pi)}{p_{inc}}[dB] \tag{1}$$

where p_{inc} and p_{scat} are incident and scattered pressure amplitudes. The scattered pressure is referenced to the normalized distance $r_1 = 1$ meter and its value at backscattering direction ($\theta = \pi$) is used in above definition. Numerical modeling of the target strength value requires considering acoustic scattering phenomena with time-harmonic plane wave incident field. When operated at frequency f_0 the normalized incident pressure could be modelled as $p_{inc} = e^{jk_0(d\cdot x)}$ with $k_0 = 2\pi f_0/c_0$ representing wave number, d its direction and c_0 - sound velocity in medium. The acoustic wave are governed by the Helmholtz equation valid in propagated medium Ω_0 with boundary condition defined on surface Γ [4]:

$$\begin{cases} \nabla^2 p_{tot} + k_0^2 p_{tot} = 0 & \text{in} \quad \Omega_0 \\ \alpha p_{tot} + \beta \partial_n p_{tot} = 0 & \text{on} \quad \Gamma \end{cases} \tag{2}$$

where $p_{tot} = p_{scat} + p_{inc}$ is the total pressure and α, β are coefficients depending on object properties. Moreover, the scattered component satisfies Sommerfeld radiation condition $\lim_{r\to\infty} r(\partial_r p_{scat} - jk_0 p_{scat}) = 0$.

The analytical solution of Eq. 2 can be obtained only for several objects with regular geometry including spherical, cylindrical or ellipsoidal ones. This theoretical solutions are often used in asymptotic methods but have also important application when testing software implementations that aims to solve Eq. 2 for any shaped object by the discretisation of its surface.

The Boundary Element Method solutions use variational form of boundary integral equation as the reformulation of Eq. 2 with surface potentials, which uniquely determine the scattered field. The large number of methods of solving integral equations were proposed in the literature and its numerical verification is a topic of last decades contributions not only in acoustics.

As the first example let us consider a problem of soft scattering that has application in modeling swimbladdered fish target strength. This case assumes $p_{tot} = 0$ ($\alpha = 1$, $\beta = 0$) on object surface what gives boundary integral equation and field representation formula in the form:

$$\begin{cases} S \cdot u = -p_{inc} & \text{on} \quad \Gamma \\ p_{scat} = -S_L \cdot u & \text{in} \quad \Omega_0 \end{cases} \tag{3}$$

where u symbolises the surface solution, while S_L and S are single layer potential operator and single layer boundary operator defined by:

$$\begin{cases} S_L = \int_\Gamma G_k(x-y)\psi(y)d\Gamma(y) \\ S = \int_{\Gamma\times\Gamma} G_k(x-y)\psi(x)\phi(y)d\Gamma(x,y) \end{cases} \tag{4}$$

with Green's function $G_k(x-y) = e^{jk|x-y|}/(4\pi|x-y|)$ and discretisation basis functions ϕ and φ. Solution p_{scat} of Eq. 4 could be obtained by assembling boundary operator matrix S, solving linear equation for surface trace u and multiplying it by potential operator S_L.

The second considered case models non-swimbladder fish for which its body reflects the sounding wave. This case requires solving acoustic transmission problem as the internal part of the fish represents another medium Ω_1 in which the wave could propagate with wave number $k_1 = 2\pi f_0/c_1$. The solution could be represented by single-trace formulation following [5]:

$$\begin{cases} \left[\mathbf{A_0} + \widehat{\mathbf{A_1}}\right] \cdot \begin{bmatrix} u_0 \\ u_1 \end{bmatrix} = \begin{bmatrix} -p_{inc} \\ -\partial_n p_{inc} \end{bmatrix} & \text{on} \quad \Gamma \\ p_{scat} = D_L \cdot u_0 - S_L \cdot u_1 & \text{in} \quad \Omega_0 \end{cases} \tag{5}$$

where $\mathbf{A_0} = \begin{bmatrix} -D_0 & S_0 \\ H_0 & T_0 \end{bmatrix}$ represents Calderon projector in surrounding medium Ω_0 with four assembled matrices from single S, double D, adjoined double T and hypersingular H boundary operators, whereas $\widehat{\mathbf{A_1}} = \begin{bmatrix} -D_1 & \frac{\rho_1}{\rho_0} S_1 \\ \frac{\rho_0}{\rho_1} H_1 & T_1 \end{bmatrix}$ is its scaled version for part in internal medium Ω_1. Reconstruction of scattered field p_{scat} uses introduced earlier single layer potential operator and double layer one:

$$D_L = \int_\Gamma \partial_y G_k(x-y)\psi(y)d\Gamma(y) \tag{6}$$

It is worth to note that from mathematical point of view the first considered case represents the spacial case of the second for which $c_1 \to 0$, $\rho_1 \to 0$ and then block matrix Eq. 5 reduces to ordinary matrix Eq. 3 with one single layer potential operator matrix. This fact allows using Eq. 5 for both presented cases. However, as the Calderon projector contains hypersingular operator the solution procedure for the second case requires applying dedicated matrix assembling and regularisation techniques.

3. SOFTWARE IMPLEMENTATION AND RESULTS

The presented boundary element method equations as applied to meshed objects were implemented with the usage FreeFEM and Bempp software packages. Both were used for implementing the the model of swimbladdered fish scattering pattern, whereas only Bempp was used to model non-swimbladdered fish pattern.

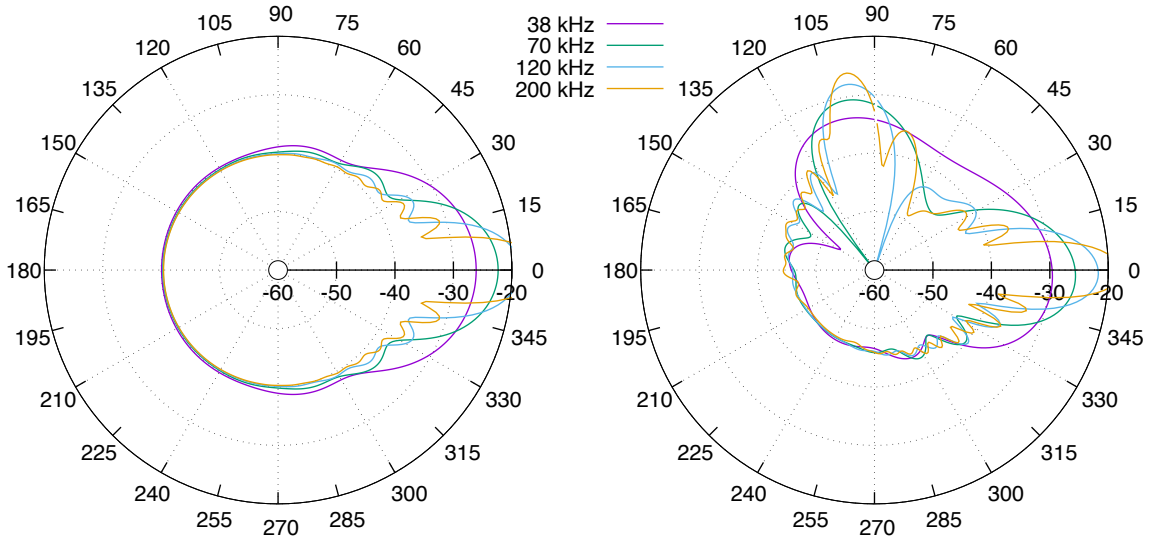


Figure 1: The scattering patterns for soft sphere having diameter equal to 38.1mm (left) and mesh model of 40cm yellowfin tuna swimbladder (right) calculated for sounding frequencies used in fishery acoustics. Backscattering direction is at $\theta = 180°$.

Implementation of soft scattering code was verified by using meshed sphere having radius of 1.905cm with variable number of elements, which in fact imitates spheres used for calibrating underwater equipment. As expected increasing number of elements increases accuracy as compared to theoretical calculations. For both packages the number of source code lines were around 30. However, the FreeFEM code is more difficult in reading as it must define type of variables, what in such specialised problem is not easy to master. The test of speed execution was performed for discretized swimbladder of 40cm yellowfin tuna [6] having mesh with 7502 nodes on M1 MacBook Pro running Sequoia 15.3.2 operating system. In this case FreeFEM v4.14 outperforms Bempp v0.4.2 in calculating scattering pattern having 360 points. The FreeFEM executable runs in 111s as compared to 188s for Bempp_cl implementation.

Fig. 1 presents two figures with patterns prepared for four operated frequencies typically used in fishery acoustics, namely 38kHz, 70kHz, 120kHz and 200kHz. The left pattern presents the results obtained for the model of a sphere with diameter equal to 38.1mm, wheres the right one for a model of 40cm yellowfin tuna.

The second case of fluid-filled targets was implemented in Bempp as it implements operators for transmission problem. Fig. 2 presents 40 lines of Python code that may serve as the minimum working example for solving fish scattering problems using Bempp. To be readable, it is neither parameterised nor modular. It is worth to point out that: (1) the input mesh should be prepared in Gmsh mesh format and should be prepared with resolution appropriate to operating frequency (2) it is assumed that the incident plane wave is arriving from y-axis coordinate (3) the observation circle is defined in x-y plane and contain 1 degree separated points (4) the results of calculation are output to console in the form of two-column table with the angle and absolute value of scattered pressure. The code was executed for 40cm modeled mackerel body having 8582 nodes and its execution in the same scenario as mentioned previously took 17 minutes.

Fig. 3 presents the input meshes for two test cases representing swimbladder of yellowfin tuna and the body of non-swimbladdered mackerel.

```python
# MM 21.5.2025: plane wave scattering from elastic body
import bempp_cl.api
from bempp_cl.api.operators.boundary import helmholtz as boundary
from bempp_cl.api.operators.potential import helmholtz as potential
from numpy import pi, sin, cos, exp, abs, array, concatenate
from scipy.sparse.linalg import gmres

# input parameters
f0, N = 38e3, 360            # frequency, number of output points
c0, rho0, c1, rho1 = 1480, 1024, 1540, 1045 # sea water - fluid
k0, k1 = 2*pi*f0/c0, 2*pi*f0/c1              # wave numbers
d = array([0, -1.0, 0])                      # plane wave direction
r1 = array([[sin(2*pi*i/N),-cos(2*pi*i/N),0] for i in range(N)]).T

# target mesh and its space
grid = bempp_cl.api.import_grid("macXray_carn_L40cm.msh")
Sp = bempp_cl.api.function_space(grid, "P", 1)

# operators
A0 = boundary.multitrace_operator(grid, k0) # external Calderon
A1 = boundary.multitrace_operator(grid, k1) # internal Calderon
SL = potential.single_layer(Sp, r1, k0)
DL = potential.double_layer(Sp, r1, k0)

# incident pressure and its normal derivative
@bempp_cl.api.complex_callable
def pinc(x, n, idx, res): res[0] = exp(1j*k0*d.dot(x))
f = bempp_cl.api.GridFunction(Sp, fun=pinc).coefficients
@bempp_cl.api.complex_callable
def dpinc(x, n, idx, res): res[0] = 1j*k0*exp(1j*k0*d.dot(x))*d.dot(n)
g = bempp_cl.api.GridFunction(Sp, fun=dpinc).coefficients

# solution
A1[0,1] *= rho1/rho0; A1[1,0] *= rho0/rho1  # scaling Calderon
A = (A0 + A1).strong_form()
u,_ = gmres(A * A, A * concatenate([f,g]))  # surface solution
u1 = bempp_cl.api.GridFunction(Sp, coefficients=u[: len(f)])
u2 = bempp_cl.api.GridFunction(Sp, coefficients=u[len(f) :])
pscat = DL * u1 - SL * u2                   # far field solution
print("\n".join(["%d\t%g"%(i,abs(pscat[0][i])) for i in range(N)]))
```

Figure 2: Minimum working example of Bempp code for the calculation of a scattering pattern of meshed object

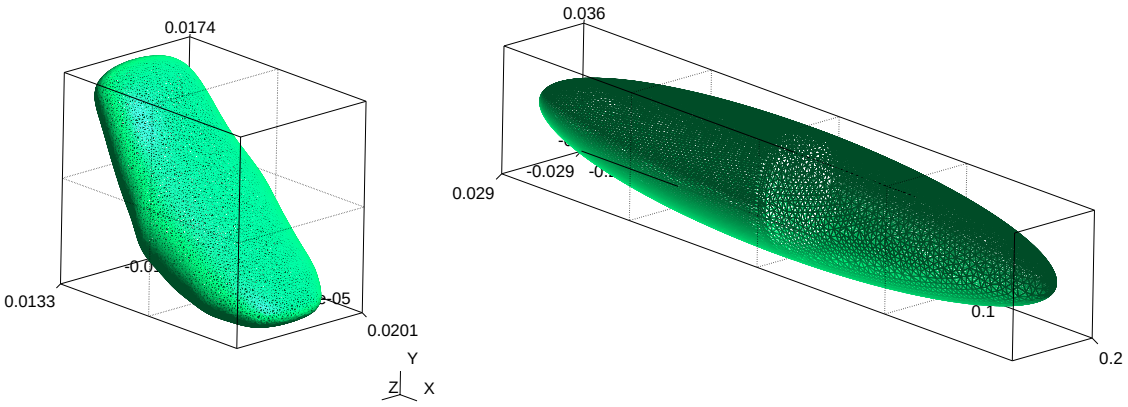


Figure 3: 40cm yellowfin tuna swimbladder mesh model (left) and 40cm mackerel body mesh model (right)

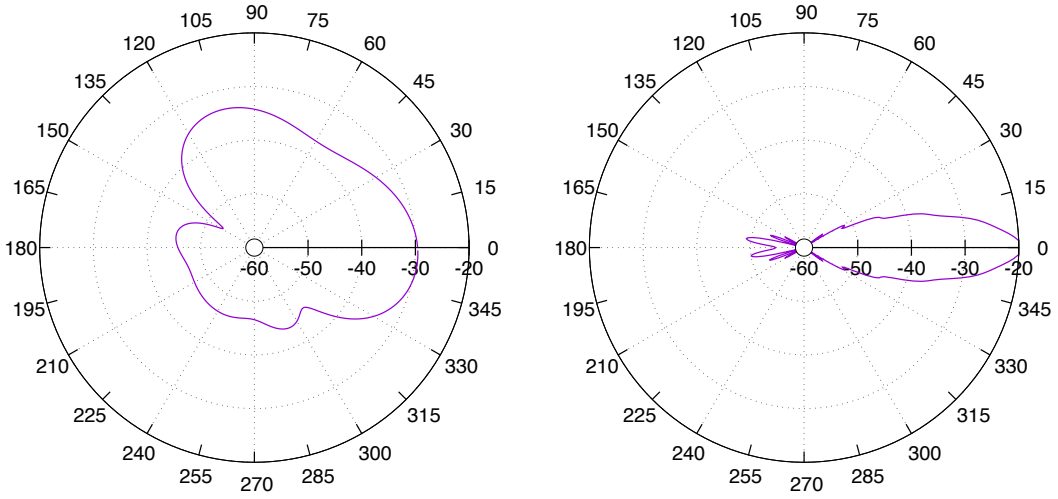


Figure 4: 40cm yellowfin tuna swimbladder scattering pattern (left) and 40cm mackerel body scattering pattern (right). Backscattering direction is at $\theta = 180°$

4. CONCLUSIONS

The paper aimed to verify usefulness of open-source finite and boundary element packages for applications in modeling of fish scattering properties. The two software packages differs by entry level as FreeFEM uses its own language and is dedicated rather for FEM approach. Bempp although is not as fast as FreeFEM allows for clear coding of complicated transmission problems as required for modeling scattering of fish without swimbladder. The execution of this code is far more slower than soft-scattering one.

REFERENCES

[1] Timo Betcke and Matthew W. Scroggs. Bempp-cl: A fast python based just-in-time compiling boundary element library. *Journal of Open Source Software*, 6(59), 2021.

[2] J.D. Gonzalez, E.F. Lavia, S. Blanc, M. Maas, and A. Madirolas. Boundary element method to analyze acoustic scattering from a coupled swimbladder-fish body configuration. *Journal of Sound and Vibration*, 486:115609, 2020.

[3] F. Hecht. New development in freefem++. *J. Numer. Math.*, 20(3-4):251–265, 2012.

[4] Stephen Martin Kirkup. The boundary element method in acoustics: A survey. *Journal of Applied Sciences*, 2019.

[5] Elwin van 't Wout. The boundary element method for acoustic transmission with nonconforming grids. *Journal of Computational and Applied Mathematics*, 445:115838, August 2024.

[6] Anderson Ladino Velásquez. *Acoustic Techniques for Tuna Biomass Estimation*. PhD dissertation, Institut per a la Gestió Integrada de Zones Costaneres, Universitat Politècnica de València, 2024.

Improving Fish School Detection Using Phase Information in Continuous-Wave Underwater Acoustic Signals: A Study Based on Cylindrical Sonar Data

Yohann Gourret[1], Karl Thomas Hjelmervik[2], Tommi Brander[3], and Hector Pena[4]

[1]University of South-Eastern Norway, yohann.gourret@usn.no
[2]University of South-Eastern Norway, karl.t.hjelmervik@usn.no
[3]University of South-Eastern Norway, tommi.brander@usn.no
[4]Institute of Marine Research, hector.pena@hi.no

Abstract: *This paper presents a novel approach to improving fish school detection using underwater acoustics, with a particular focus on the integration of both amplitude and phase information obtained from a cylindrical sonar survey employing Continuous-Wave (CW) pulses. While the inclusion of phase data does not significantly enhance overall classification accuracy, it proves valuable for identifying and discarding potential back lobes. The proposed method leverages Convolutional Neural Networks (CNNs) to analyze the collected sonar data and distinguish between fish schools and non-fish signals. Experimental results demonstrate that incorporating phase information helps reduce false detections associated with back lobes, leading to cleaner classification outputs. Receiver Operating Characteristic (ROC) curves show comparable performance with and without phase data in general, but highlight improved robustness in challenging scenarios where back lobes are present. This study emphasizes the role of phase information in refining underwater acoustic analysis for more reliable fish school detection, particularly in CW-based applications where back lobe artifacts are common.*

Keywords: *Fishery acoustics, detection, Machine Learning.*

1. INTRODUCTION

Accurate identification of fish schools using acoustic methods is crucial for sustainable fisheries management. Traditional methods for fish monitoring rely on labour-intensive and error-prone manual observation and processing. CNNs can automate and enhance the accuracy of fish recognition [1, 2].

Side-looking sonar systems are effective in detecting fish. Nicholas Makris [3] demonstrates large low-frequency systems for detection of entire fish schools over continental-shelf scales. This technology can image vast areas, providing valuable insights into fish population dynamics and behavior. On the other hand, high-frequency sidescan sonar systems detecting individual fish at shorter ranges [4] offer precise detection capabilities and an improved basis for classification of species, but are limited in their range. Sonar systems operating in the low ultrasonic range present a middle ground, offering detection ranges of hundred of meters, and are sufficiently compact for use on smaller vessels, such as unmanned surface vehicles. They balance the need for range and resolution, providing a versatile solution for fisheries management.

We present an approach to fish school recognition using a CNN applied to matched filtered output from a side-looking fishery sonar in the low ultrasonic frequency range. Our method involves conventional detection algorithms on beamformed and matched filtered data, with a time-bearing window of the matched filter output fed into a CNN for binary classification of fish (*Fish School* or *Not Fish School*). We include both amplitude and phase data in the matched filter output, as the phase data helps in recognizing false positives caused by back lobes, a phenomenon resulting from an undersampled array.

The method is demonstrated on 4,760 pings collected by the Institute of marine research with a Simrad SU90 fish finding sonar. The transmission is a 6 ms long continuous wave pulse at 26 kHz. The data used are complex-valued signals sampled at 4 kHz in 64 beams.

2. BACK LOBES

In our data, we can observe back lobes (equivalent to side lobes). Those lobes are around 110° from the main lobe. Looking at the phase, we can observe very interesting patterns, chess patterns (Fig. 1), that make those back lobes easy to identify. We will first present how we visualize the phase and then explain the origin of the patterns.

Visualization of the phase

The original signal has a frequency $f_c = 26$ kHz. We have a complex signal sampled at $f_s = 4$ kHz. We choose to represent the phase information of the complex pressure $p_{i,j}$ at range i and direction j by $|\cos(\arg(p_{i,j}))|$, see Fig. 1.

This representation only works if $\exists n \in \mathbb{N}$ such that $f_c/f_s = \frac{1}{2}n$. In our case $f_c/f_s = 6.5$. To have the same visualization for other frequencies, we can represent the phase information by $|\cos(\arg(p_{i,j}e^{j2\pi f_c t_i}))|$, where t_i is the time at the sample $p_{i,j}$.

Chess pattern

For human eyes, the patterns in the phase information are very easy to identify. Could phase information be beneficial for a neural network in identifying back lobes or, generally, improving classification performance?

Chess patterns in the phase information are due to a variation of the phase measured on the angular axis and on the radial axis. For the angular axis, this originates from the beamformer.

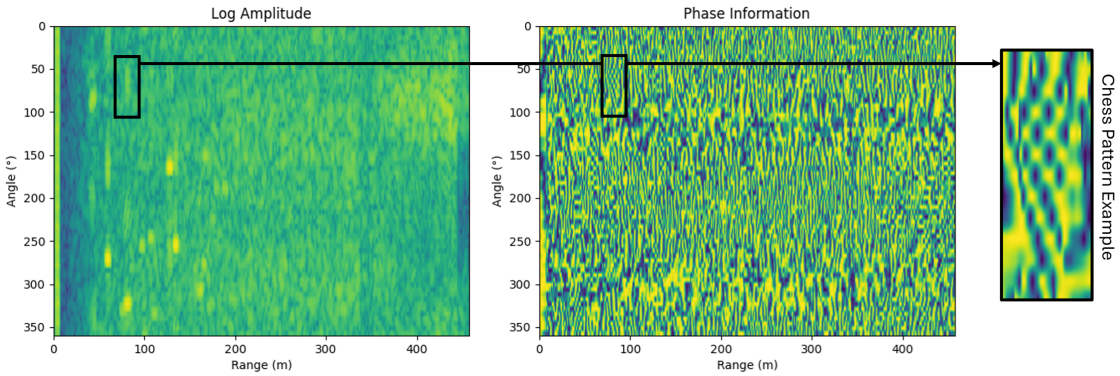


Figure 1: Log Amplitude, Phase Information and Chess Pattern example

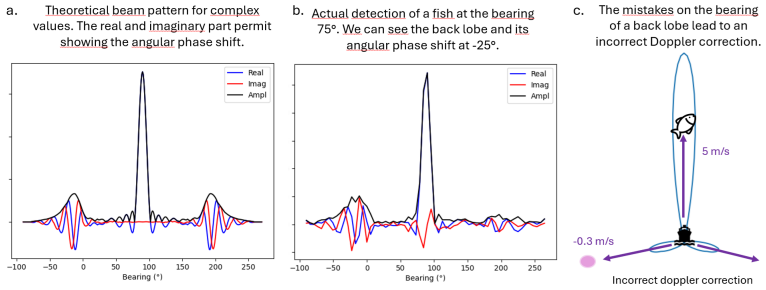


Figure 2: Explanation of the chess pattern effect

The exact beamforming method used by Simrad (Kongsberg Discovery) is unknown, but our guess for the beam pattern is shown in Fig. 2.a. Our analysis shows that the tendency remains similar for different variations of the beamforming method. The figure shows the angular variation of the amplitude and of the phase. We can see that on the back lobes, at around 110°, the phase changes rapidly over angle. The beamforming explains the angular phase shift. Fig. 2.b shows an actual fish school detection and the back lobe. We can see the same type of variation as in the theoretical beam pattern.

For the radial axis, the pattern appears due to the Doppler-corrected matched filtering. As shown in Fig. 2.c, on back lobes, the echo from the fish will be compensated for the wrong Doppler and will shift the phase over time / ”range”. However, for some bearings, for example at 110°/2=55°, the back lobes will be compensated for the correct Doppler by symmetry. Therefore, the pattern over the range is less consistent than the angular pattern, as we observe in Fig. 5 for several detections.

3. METHOD

We use two steps to identify fish schools. The first is to identify a potential fish school by a classical method, and the second is to use a CNN on the potential fish schools to classify them between two classes: *Fish School* or *Not Fish School*.

First step: Classical detection

To detect potential schools of fish, we use a range-dependent threshold. Detection occurs if $p_{i,j} > (G_\sigma * \hat{p})[i] \cdot 10^{T_{\text{dB}}/20}$ with $\hat{p}_i = \frac{1}{64}\sum_{j=1}^{64} p_{i,j}$, G_σ a Gaussian filter with $\sigma = 20$ m and the threshold $T_{\text{dB}} = 12$ dB. There is also a minimum and maximum detection surface and a

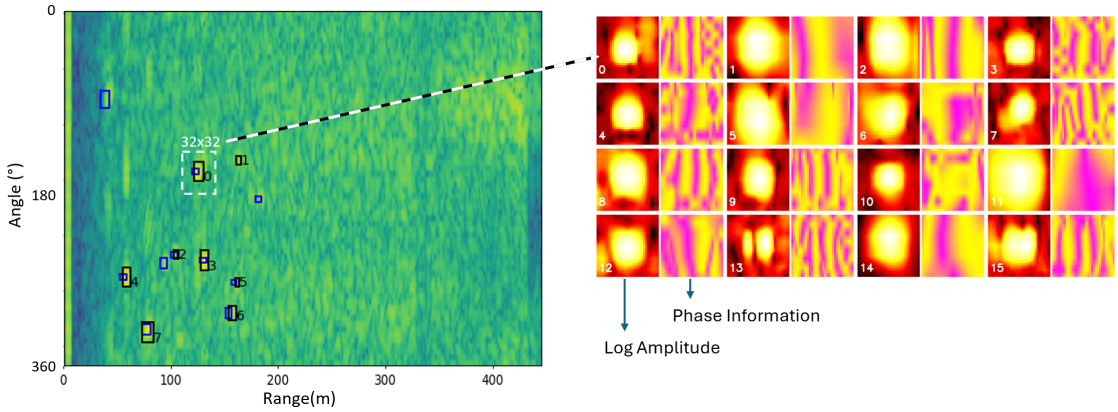


Figure 3: Image extraction from the detections

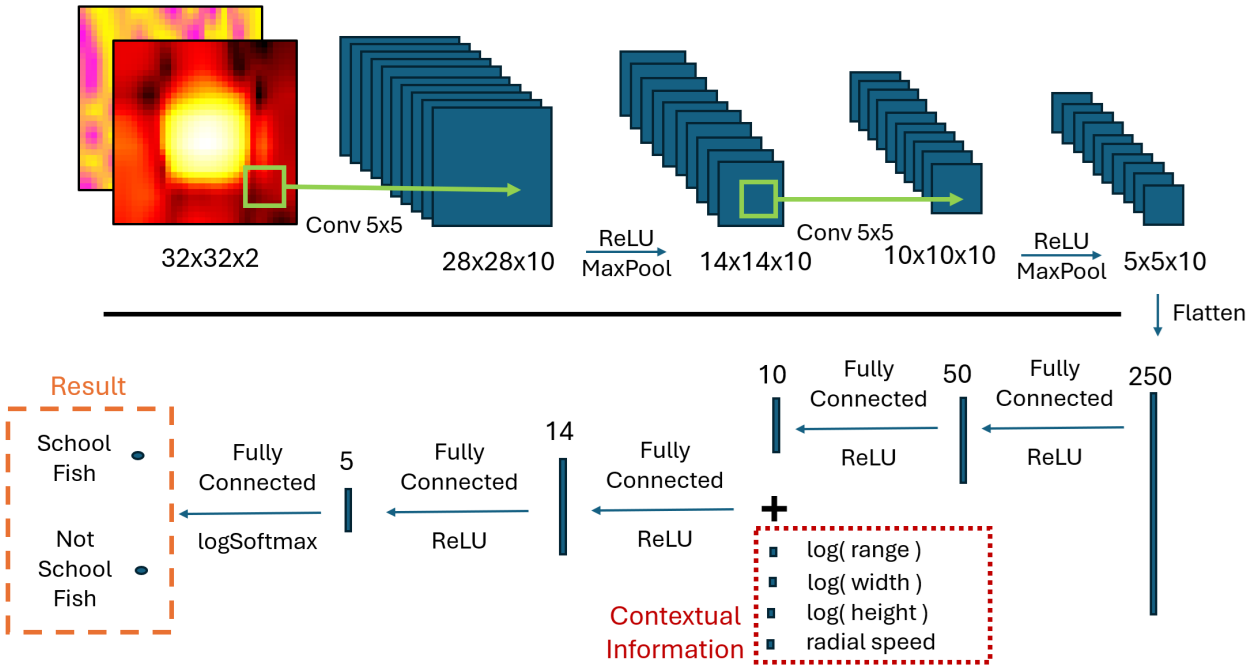


Figure 4: CNN architecture

minimum and maximum range. For each detection box, we generate a 2 channel 32×32 image (Fig. 3), which covers the detection box, is centered on it and has double the side lengths. The first channel of the image is the logarithmic amplitude of the pressure, and the second is the phase information. The resulting image is sent to the CNN for classification.

Second step: CNN

The CNN is presented in Fig. 4. The image of the detection step and four pieces of contextual information are the inputs. Contextual information is introduced later on in the network and contextualizes the image with the logarithm of the range of the detection, the logarithm of the width and the height of the original crop and the radial speed of the boat in relation to the environment. Here, radial speed = boat speed $\times$ cos(bearing).

The dataset is split into a training ($2/3$ of the data) and a testing dataset ($1/3$ of the data).

4. RESULTS AND DISCUSSION

The results of the training for models with and without phase information are shown in Fig. 7. Both models have the same architecture, but for the model without phase information, the phase information is set to 0 at input. The resulting Area Under the Curve on the testing dataset is close to 0.87 for both models. Those results can permit an automatization of the process. However, the phase information does not significantly increase the overall performance.

To identify the relevance of the phase information for the identification of back lobes, we manually identify them to analyze the response of models for those specific cases. This manual work was done for 321 pings at the beginning of the dataset, so they are in the training dataset. This resulted in 207 back lobes, for 3160 detections (6.6%). Some examples of back lobe detections are presented Fig. 5.

Comparison of models by classifying the back lobes between *Fish School* or *Not Fish School* implies choosing a specific threshold. To avoid choosing an arbitrary value, we use the ROC curve of the testing dataset to estimate the confidence of models. Based on the limit threshold needed for the response of the model to be switched between *Fish School* and *Not Fish School*, we can use the ROC curve to estimate the confidence. When this point on the ROC curve is on the left, the Neural Network is confident to classify it as *Fish School*. When it is on the right, the Neural Network is confident to classify it as *Not Fish School*. A visual example of this estimate of the confidence is shown Fig. 6.

Based on this metric, we use the ROC curves of the testing dataset to estimate the confidence of both models, with and without phase information, on the classification of back lobes. The resulting histograms are shown in Fig. 8. Histograms present the confidence of models to classify back lobes as Fish School, but because the back lobes are not fish schools, the closer the confidence is to 0 the better the result is. The average confidence for the model with phase information is 26.4%, which is lower than the 34.1% of the model without the phase information.

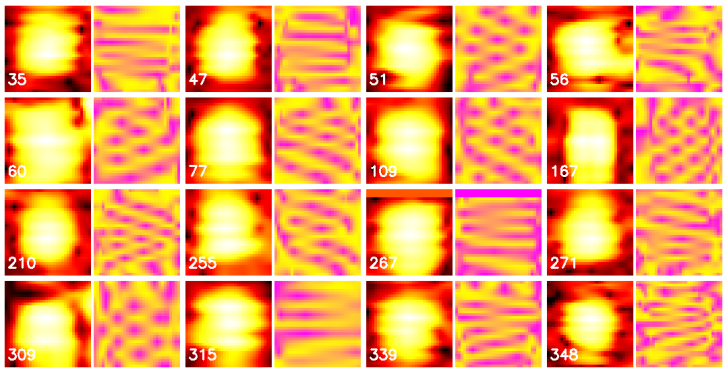


Figure 5: Samples of back lobes detection

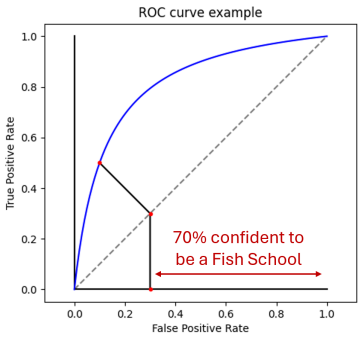


Figure 6: Estimation of the confidence

5. CONCLUSION

The current Area Under of the Curve of 0.87 can permit the automatization of the process, replacing human labor by machine labor. The current algorithm analyzes data ping by ping. Merging information over pings should lead to better results in further works.

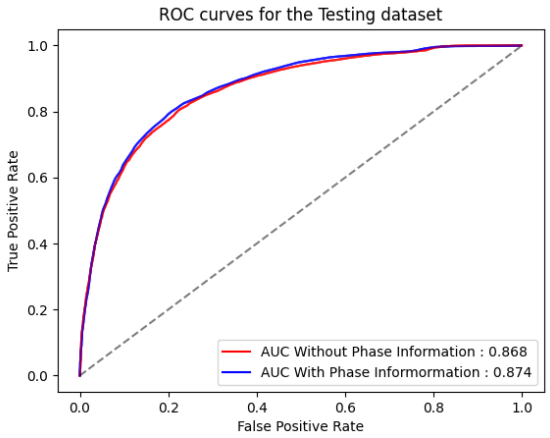


Figure 7: ROC curves for the Testing dataset

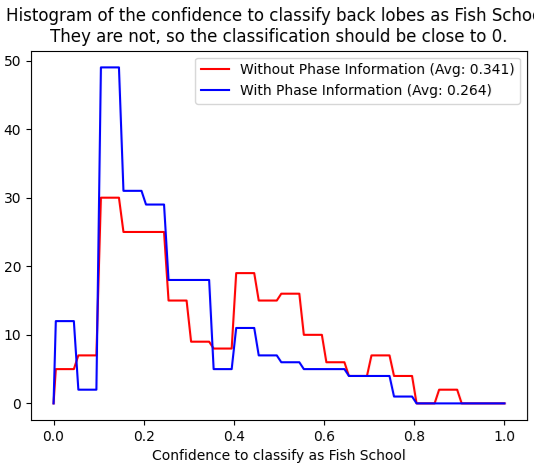


Figure 8: Histogram of confidence

About the phase information, it does significantly increase the performance for back lobes, but it does not significantly increase the performance of the neural network for other detections. Since most back lobes have been filtered away by the threshold of the classical method, the phase information does not increase the overall performance.

ACKNOWLEDGMENTS

This work was supported by the European Union Horizon 2020 research and innovation programme under the SAFARI project [grant number 101147432]. We would like to thank the Institute of Marine Research (IMR) for collecting the data used in this study, and Kongsberg Discovery for providing detailed information on their SU90 sonar system.

REFERENCES

[1] V. Kandimalla, M. Richard, F. Smith, J. Quirion, L. Torgo, and C. Whidden, "Automated detection, classification and counting of fish in fish passages with deep learning," *Frontiers in Marine Science*, vol. 8, Jan. 2022.

[2] W. Shen, M. Liu, Q. Lu, Z. Yin, and J. Zhang, "A fish target identification and counting method based on DIDSON sonar and YOLOv5 model," *Fishes*, vol. 9, p. 346, Sept. 2024.

[3] N. C. Makris, P. Ratilal, D. T. Symonds, S. Jagannathan, S. Lee, and R. W. Nero, "Fish population and behavior revealed by instantaneous continental shelf-scale imaging," *Science*, vol. 311, pp. 660–663, Feb. 2006.

[4] J. L. Ridgway, J. A. Madsen, J. R. Fischer, R. D. Calfee, M. R. Acre, and D. C. Kazyak, "Side-scan sonar as a tool for measuring fish populations: Current state of the science and future directions," *Fisheries*, vol. 49, pp. 454–462, Oct. 2024.

Structured Session
12. Observing the Oceans Acoustically

Organised by Bruce Howe and Kay Gemba

A Submodular Approach to Acoustic Sensor Placement using the MARLIN Digital Twin

Edward Clark[1], Eric Schoof[2], Chris Manzie[2], and Alan Hunter[1]

[1]University of Bath, Bath BA2 7AY, United Kingdom
[2]University of Melbourne, Victoria 3010, Australia

Contact author: Edward Clark, eprc20@bath.ac.uk

Abstract: We present a submodular optimization approach for acoustic sensor placement using the MARLIN digital twin. Our objective is to select sensor locations in a region off the coast of Victoria, Australia, to maximize detection performance across the environment. The sensor placement problem is formulated as a submodular set function, allowing the use of a greedy algorithm that guarantees solutions within a known bound of the optimum. MARLIN provides a high-fidelity simulation environment using open-source oceanographic and bathymetric data. We generate $2,840$ candidate sensor locations and compute acoustic propagation loss using the PyRAM model, incorporating environmental variability. Detection probability is estimated from signal-to-noise ratios, and the joint detectability across the region is evaluated for different sensor configurations. We compare sequential and simultaneous sensor placement strategies, analysing their impact on detection performance and computational efficiency. Results show that the greedy submodular approach achieves near-optimal detection with orders-of-magnitude fewer evaluations than brute-force search, making it practical for large-scale deployments. The study demonstrates the effectiveness of submodular optimization for sensor placement in complex marine environments and discusses trade-offs between placement strategies.

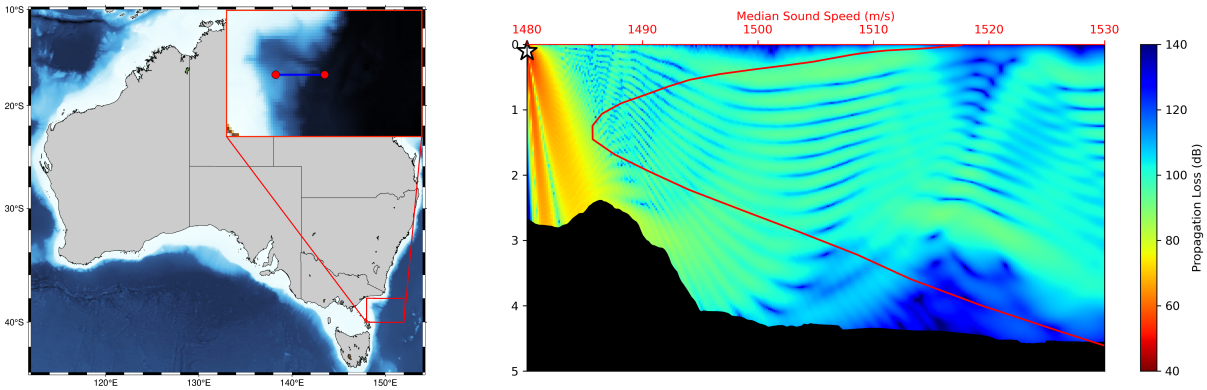
Keywords: *Passive acoustics, Sensor Placement, Submodular Selection*

1. INTRODUCTION

Passive acoustic underwater sensors are key to providing monitoring and deterrence in ocean environments. Their ability for long-range sensing makes them suitable for a range of applications; however, complex acoustic propagation makes optimal placement of them challenging. Temperature, salinity, and pressure gradients all influence the way sound travels through the water column, leading to sound channels where sound can travel hundreds of kilometres, and 'shadow' zones where little can be heard.

Many studies optimise the placement of these sensors using a simple detection range calculation based on water column properties of the surrounding area. This is not valid for regions with highly varying bathymetry and strong sound speed gradients where the propagation is non-uniform across the domain. To predict the sonar performance, acoustic propagation simulations are required to accurately determine how the sound energy travels through the region of interest. Studies have previously done this and extended it to Nx2D simulations to create a 3D placement algorithm but still assume their detection range is only as far until the first time the detection probability falls below a threshold [6]. This is also an assumption that should be challenged, as having gone through the computational effort to calculate the propagation loss a more holistic approach can be used looking at the entire domain.

If one were to consider the efficacy of choosing N sensors from a set $\mathcal{S}$ with $|\mathcal{S}| = R$, generally this would involve evaluating R choose N selections of sensors to decide which selection was the best. This set of sensor selections to evaluate grows with $\mathcal{O}\left(R^N\right)$ for $N << R$, which quickly becomes intractable. To combat this, we leverage the field of submodular optimisation, which gives conditions under which we can perform a greedy selection of sensors in batches of K and still get within a known bound of the optimal solution.



(a) *(b)*

Figure 1: (a) Map of the region of interest off the Victoria coast in Australia. The red line shows the transect used in this study. (b) Example of the propagation loss from a single sensor to a target. The colour bar shows the propagation loss in dB. The red line shows the sound speed profile of the water column.

2. METHODOLOGY

In this paper we select N sensor positions from a set of $|\mathcal{S}| = 2,840$ possible sensor locations from a region of interest off the coast of Victoria, Australia (Fig. 1a), using submodular opti-

mization methods. We further show how choosing sensors in small groups of K can improve proximity to the optimal solution while substantially reducing computation time.

2.1. SUBMODULAR FUNCTIONS

Given a set $\mathcal{S}$ with $|\mathcal{S}| = R$ with subset $S \subseteq \mathcal{S}$, we can define a set function $f(S) : 2^R \to \mathbb{R}$ that maps the selection S from the set $\mathcal{S}$ onto a scalar. We further let the empty set be denoted $\emptyset$. Set functions have several properties that we leverage and are presented here in brief.

Definition 1. *A set function $f(S) : 2^R \to \mathbb{R}$ is monotone increasing if, for all subsets $A \subseteq B \subseteq \mathcal{S}$ it holds that $f(A) \leq f(B)$.*

Definition 2. *A set function $f(S) : 2^R \to \mathbb{R}$ is submodular if, for all subsets $A \subseteq B \subseteq \mathcal{S}$ and $s \notin B$, then*

$$f(A \cup \{s\}) - f(A) \geq f(B \cup \{s\}) - f(B). \tag{1}$$

Definition 2 implies adding an element to a smaller set should result in a larger increase to the set function than adding the same element to a larger set (which contains the smaller set). This is often referred to as the set function exhibiting "diminishing returns".

A canonical example of a monotone increasing submodular function stems from the facility location problem.

Example 1 (Facility Location Problem [3])**.** *Let Ω be a set of clients and $\mathcal{S}$ be a set of facilities. Let $d : \Omega \times \mathcal{S} \to \mathbb{R}$ be the cost of assigning a given client to a given facility. For each client i and subset of facilities $S \subseteq \mathcal{S}$, define $f_i(S) = \max_{s \in S} d(i, s)$. The value of a non-empty S is*

$$f(S) = \sum_{i \in \Omega} f_i(S) = \sum_{i \in \Omega} \max_{s \in S} d(i, s), \tag{2}$$

and $f(\emptyset) = 0$. The objective is to select the subset of facilities S of size $|S| = N$ that maximizes $f(S)$. The set function $f(S)$ is a monotone increasing submodular function.

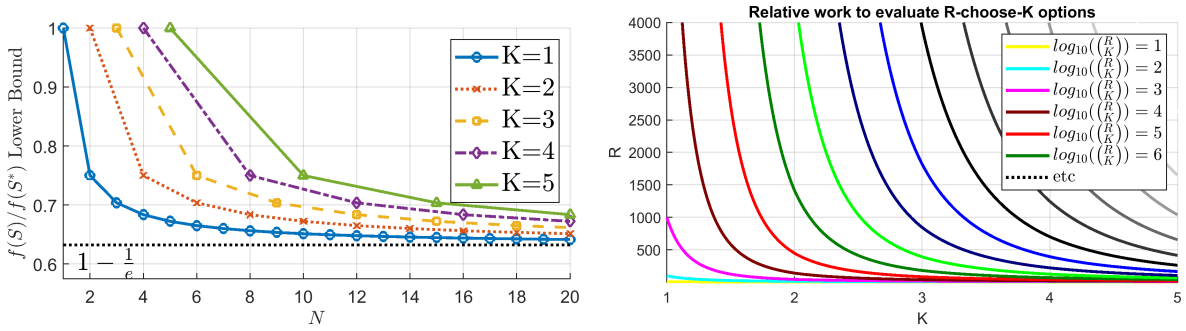
The maximization of a cardinality N set selection problem is NP-hard. However, for the special case of monotone increasing submodular functions, an efficient K-step greedy selection approach performs within 37% of the optimal set selection S^*.

Proposition 1 ([4])**.** *Consider selecting N elements from the set $\mathcal{S}$ which optimizes the value of a monotone increasing submodular cost function $f(S)$ for some $S \subseteq \mathcal{S}$. If we select these elements $K \leq N$ at a time in a greedy fashion where K divides N evenly, then $f(S)$ will satisfy*

$$\frac{f(S) - f(\emptyset)}{f(S^*) - f(\emptyset)} \geq 1 - \left(\frac{q-1}{q}\right)^q \geq 1 - \frac{1}{e} \approx 63\% \text{ as } q \to \infty, \tag{3}$$

where $q = {}^N\!/_K \in \mathbb{Z}^+$.

In practice, the K-step greedy algorithm often performs much better than these bounds. This result also extends to cases where K does not evenly divide N, but we do not consider them in this paper. A graphical interpretation of these results is shown in Fig. 2a. We can also get an intuitive understanding of the computation required to solve these problems optimally by examining Fig. 2b. As an example, it takes approximately 1,000 times more work to evaluate the $K = 3$-step greedy selection for a set of size $R = 850$ than it does to evaluate the $K = 2$-step greedy selection from a set of size $R = 500$.



(a) *(b)*

Figure 2: (a) Theoretical submodularity selection bounds on optimality from Eq. 3. If we want to select a total of N elements in batches of K, this shows us how close the final set will be to the performance of choosing N elements in a single batch of N.
(b) Relative work to evaluate R-choose-K element selections. This represents how much relative computation is required to select a single batch of K elements from a set of size R. Each line represents the same amount of work. Lines are drawn on each order of magnitude of increasing work.

2.2. PROBLEM STATEMENT

A set of $2,840$ possible sensor locations is generated using a uniform distribution across the region of interest from a depth of 0-2000m. From this, we use the MARLIN digital twin (developed on from work in [1]) to simulate the propagation loss using the PyRAM propagation model from every possible sensor location across the domain. PyRAM is a range-dependent acoustic propagation model that uses the split-step Pade method to calculate the propagation loss from a source across 2D environment [2]. As propagation loss can be assumed to be reciprocal, that is the propagation loss from a sensor to a target is the same as from the target to the sensor, we can calculate the propagation loss from every possible sensor location to every point in the region of interest. We take environmental data from the Copernicus Marine Service, which provides water column properties such as temperature, salinity, and pressure. Bathymetry data is taken from the GEBCO dataset, which provides global bathymetry data [5]. The probability of detection is then calculated using the signal-to-noise ratio (SNR) and a logistic function to map the SNR to a probability of detection.

Let Ω be the set of points of interest in the region and $\mathcal{S}$ be the set of possible location. For each point of interest i and sensor s, let the probability of detection be $d(i,s)$. Selecting a subset of N sensors $S \subseteq \mathcal{S}$ that maximizes the average detectability $f(S) = {}^1/_{|\Omega|} \sum_{i\in\Omega} \max_{s\in S} d(i,s)$, is an instance of the Facility Location Problem from Example 1. This implies that the average detectability $f(S)$ of a set of sensors is monotone increasing submodular function, so a K-step greedy selection of N sensors will be near optimal according to the bounds in Eq. 3.

3. RESULTS

In this section, we will examine the output of a submodular sensor selection problem to a transect off the Victorian coast. If we choose the one sensor that maximizes the average detectability, we can see the output in Fig. 5a. If we instead want to choose two sensors out

of the $2,840$ possible sensor locations (see Fig. 3), we would need to evaluate 2840 choose 2, which is approximately 4×10^6 possible pairs. We can see the output of the best selection in Fig. 5d, with an average detectability of $\mu = 0.69$. We leveraged our submodularity results (Eq. 3) to instead choose two sensors sequentially knowing that in the worst case we will get within 37% of the optimal value (see Fig. 2a). Choosing sensors this way will result in evaluating only 5679 sensor placements which is almost 700x fewer evaluations compared to the optimal selection. Figure 5c shows that our average detectability using this method is $\mu = 0.68$, which is within 1.5% of the true optimal we obtained from a brute-force search seen in Fig. 5d. Figure 5b also shows the theoretical lower bounds for $K < N$ from Eq. 3.

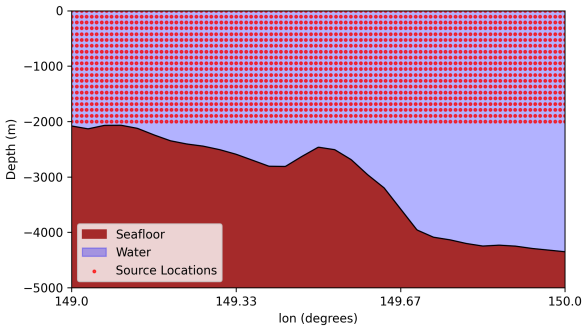
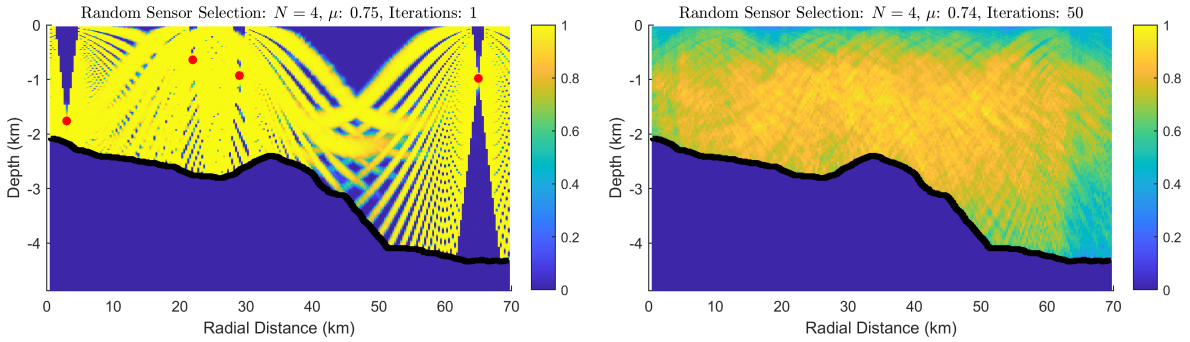


Figure 3: Sensor locations: uniformly distributed between 0-2km depth along the entire domain.

We also evaluated selecting three sensors simultaneously, which is 2,840 choose 3 or approximately 3.8×10^9 possible triples. Using the submodular approach selecting individual sensors sequentially we need to evaluate $8,517$ sensor placements, which is more than 447,000x faster. One might be tempted to perform a random sensor selection since computing the optimal value for large selections is difficult. We can see in Fig. 4 that selecting $N = 4$ sensors randomly resulted in an average detection probability of $\mu = 0.74$ compared to the average detection probability $\mu = 0.88$ of the greedy selection $K = 1$. In Fig. 5b, we can see how the average detectability increases with the number of sensors selected N. We also see that we do better as the sensors are chosen in larger groups K, up to the globally optimal value of $N = K$.



(a) *(b)*

Figure 4: (a) Single random selection of $N = 4$ sensors with $\mu = 0.75$. (b) Average detectability of a random selection of $N = 4$ sensors over 50 iterations with $\mu = 0.74$.

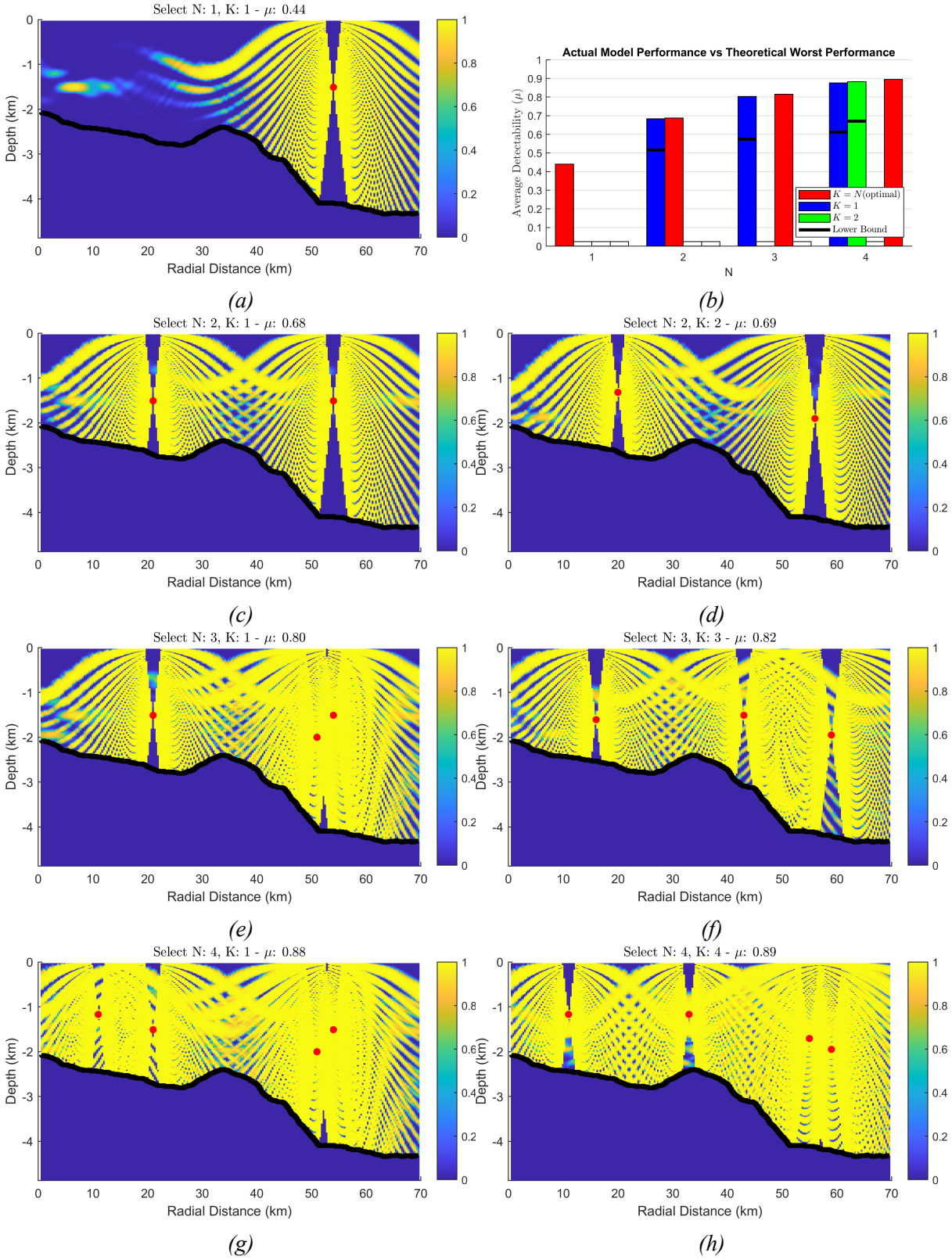


Figure 5: Selecting multiple sensors to maximise average detectability. (a) $N = K = 1$, (c) $N = 1, K = 2$, (d) $N = K = 2$, (e) $N = 1, K = 3$, (f) $N = K = 3$, (g) $N = 1, K = 4$, and (h) $N = K = 4$. The compute resources for case (h) was 1357x more than for (g), yet the average detectability of (g) was within 1%. (b) Average detectability for $N = K$ and $N < K$.

4. DISCUSSION

Sequential sensor placement impacts detection performance differently than simultaneous placement, but practical constraints also play a role. For rapid deployment, placing the first sensor at the location with the highest average detectability and then selecting subsequent sensors to maximize the average detectability of the remaining locations enables high detection performance at the start of a deployment. However, this approach does not achieve the same overall detection performance as placing sensors using a simultaneous approach. For long-term deployments, simultaneous placement yields higher detection performance across the region of interest, as shown in Fig. 5b. Placing the sensors simultaneously allows for a more holistic view of the region, as the sensors can be placed in locations that complement each other, rather than just maximizing the average detectability of the remaining locations.

Computation time is a key factor in the sensor placement problem. This study has used 2,840 simulated sensor locations which takes up to a week to compute on a standard desktop computer. This could be significantly reduced by using a high-performance computing cluster to run in parallel, but this is not always available. If one then had to evaluate the performance of every possible sensor placement, this would take an intractable amount of time. Submodular optimization allows us to reduce the number of evaluations required to find a near-optimal solution. Taking the example of selecting two sensors from 2,840 possible locations, we can reduce the number of evaluations from approximately 4×10^6 to 5,679, which is a significant reduction in computation time. To see how quickly the computation gets completely intractable, we can consider selecting 10 sensors from 2,840 possible locations, which results in approximately 9.2×10^{27} comparisons in the optimal case, but only 28,355 comparisons with a batch size of $K = 1$. This reduction in computation time is crucial for practical applications, as it allows for rapid deployment of sensors in the field.

5. CONCLUSION

This study has shown that submodular optimization can be used to select sensor locations in a region of interest to maximize the detection performance of the sensors. Using a submodular approach allows us to reduce the number of evaluations required to find a near-optimal solution, which is crucial for practical applications. In real terms, we can reduce computation time of evaluating sensor locations by a factor of 700x for two sensors and 447,000x for three sensors. This makes the near-optimal deployment of modern multi-sensor deployments feasible in a reasonable time frame. An understanding of the trade-offs between sequential and simultaneous sensor placement is also important, as it can impact the detection performance of the sensors.

6. ACKNOWLEDGEMENTS

This study has been conducted using E.U. Copernicus Marine Service Information (https://doi.org/10.48670/moi-00016). Edward Clark is grateful to the UKRI Centre for Doctoral Training in Accountable, Responsible, and Transparent AI (ART-AI) and SEA for supporting his PhD.

REFERENCES

[1] Edward Clark, Alan Hunter, Olga Isupova, and Marcus Donnelly. Optimising sensor path planning with reinforcement learning and passive sonar modelling. In *7th Underwater Acoustics Conference and Exhibition, UACE 2023*, pages 411–418, 2023.

[2] Marcus Donnelly. Python adaptation of the range-dependent acoustic model (ram)., June 2022.

[3] Andreas Krause and Daniel Golovin. *Submodular Function Maximization*, pages 71–104. Cambridge University Press, 2 2014.

[4] G. L. Nemhauser and L. A. Wolsey. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, pages 265–294, 1978.

[5] Pauline Weatherall, Sheila Caceres Ferreras, Sara DA Cardigos, Natalie Cornish, Sam R Davidson, Boris Dorschel, Hayley Drennon, Vicki Ferrini, Hugh A Harper, Tea Isler, et al. The gebco_2024 grid-a continuous terrain model of the global oceans and land. 2024.

[6] Xiaohan Zhu, Ye Wang, Zeyu Fang, Lei Cheng, and Jianlong Li. Strategic deployment in the deep: Principled underwater sensor placement optimization with three-dimensional acoustic map. *The Journal of the Acoustical Society of America*, 156(4):2668–2685, October 2024.

Structured Session
13. Pile driving noise -Radiated noise from offshore wind construction

Organised by Jonas von Pein and Vincent E. Premus

Evaluating the Applicability of Scaling Laws for Mitigated Pile Driving Noise to Highly Frequency Dependent Mitigation Systems

Jonas von Pein[1] and Matthias Ram[1]

[1]DHI Wasy GmbH

Jonas von Pein
DHI Wasy GmbH, Grimm 8, 20457 Hamburg, Germany
jovp@dhigroup.com

Abstract: *The installation of offshore wind turbines usually requires the driving of piles into the seabed. The pile driving activity leads to underwater noise with the potential of negative consequences for the marine fauna, such as behavioral changes or hearing impairments. To address these concerns, many countries have established noise limits for pile driving activities, necessitating effective noise mitigation measures. To assess the sound levels and to optimize noise mitigation measures several modelling approaches have been developed in the last decade. These include numerical and empirical models. A semi-empirical approach for the estimation of the sound levels are the scaling laws for unmitigated and mitigated impulsive pile driving scenarios. It has been shown that these lead to reliable estimates for several different scenarios.*

The recently published scaling laws for mitigated pile driving noise consider the influence of the strike energy, pile diameter, ram weight, water depth, and mitigation system such as double big bubble curtain (DBBC) and the combination of a noise mitigation screen and a DBBC. However, the noise mitigation combination that has been most widely used in Germany is the combination of a Hydro Sound Damper (HSD) system directly at the pile and a DBBC. The insertion loss of the HSD, the DBBC and a combination thereof is strongly frequency dependent. Within this paper, publicly available measurement data is compared to the trends derived for mitigated pile driving to determine whether the scaling laws can also be applied to the popular HSD/DBBC combination.

Keywords: *pile driving, scaling laws, noise mitigation*

1. INTRODUCTION

The installation of offshore wind farms usually requires the driving of piles into the seabed. A common method to drive the pile is impulsive pile driving, which generates high impulsive underwater noise levels. The emitted noise can impact the marine fauna. These impacts may include behavioral changes or hearing impairment. Therefore, the associated noise levels and their impacts need to be assessed and noise limits need to be obeyed in many countries. Usually noise mitigation systems are necessary to fulfill the noise limits. These noise mitigation systems are often based on a system in the direct vicinity of the pile and one or two bubble curtains further away.

For the assessment of the underwater noise levels, several different modeling approaches have been developed. These include detailed numerical models [1] as well as empirical [2] and semi-empirical models. Scaling laws have been developed for unmitigated [3] and mitigated pile driving [4]. These enable to scale the influence of the strike energy, pile diameter, ram weight, and water depth from one scenario to another. The scaling laws for mitigated pile driving include scenarios with one big bubble curtain, two big bubble curtains (DBBC), a fully absorbing noise mitigation screen directly at the pile and the combination of the noise mitigation screen with a DBBC. However, the noise mitigation combination that has been most widely used in Germany is the combination of a Hydro Sound Damper (HSD) [5] system directly at the pile and a DBBC. The insertion loss of an HSD/DBBC combination is strongly frequency dependent and not simple to apply within a numerical model. Therefore, the scaling laws derived for the estimation of the sound exposure levels (SEL) with the application of a DBBC are tested with publicly available measurement data from six wind farms in order to identify, if they are applicable for the highly frequency dependent mitigation system combination HSD/DBBC.

2. SCALING LAWS FOR PILE DRIVING SCENARIOS WITH A DBBC

It is important to note that the scaling laws derived in the following are based on generalizations, assumptions and numerical model runs that come along with uncertainties, as shown in [4]. Aspects such as the geo-acoustic profile, specific pile design, the interaction between pile head and hammer parts as well as the bubble curtain layout such as radii and air supply are not incorporated within the scaling laws. However, given a general comparability between the two considered scenarios, scaling laws have shown to capture the trends of the four considered parameters.

Scaling laws for an HSD/DBBC combination have not been developed.

In the following the scaling laws for the SEL that were developed for the application for scenarios with a DBBC are repeated and applied. This is done due to the similarity in the shape of the frequency dependent insertion loss with a DBBC and the HSD/DBBC combination as shown in [2]. The definition of the SEL follows [6].

The influence of the strike energy E is scaled with

$$\Delta \mathrm{SEL}_E = 10 \log_{10} \left(\frac{E_i}{E_0} \right) . \tag{1}$$

The best-fit approximation for the influence of the SEL on the pile diameter d led to a scaling law of

$$\Delta \mathrm{SEL}_d = 15.7 \log_{10}\left(\frac{d_i}{d_0}\right) . \tag{2}$$

The influence of the hammer type is usually taken into account in detailed numerical models by computing the interaction between hammer and pile [1]. A simplified approach is the scaling of the hammer influence by considering the ram weight. The influence of the ram weight m_r on the SEL is scaled by

$$\Delta \mathrm{SEL}_{m_r} = -8.5 \log_{10}\left(\frac{m_{r,i}}{m_{r,0}}\right) . \tag{3}$$

The influence of the water depth h is summarized by

$$\Delta \mathrm{SEL}_h = 0.13 \left(h_i - h_0\right) . \tag{4}$$

meaning that the SEL is expected to increase by 1.3 dB per 10 m increase in water depth.

3. APPLICATION OF SCALING LAWS WITH TO HSD/DBBC SCENARIOS

In the following, the scaling laws derived for a DBBC are applied to measurement data from six different wind farms at which pile driving was conducted with the application of a HSD/DBBC combination.

3.1. AVAILABLE MEASUREMENT DATA

The evaluation is based on publicly available data sets, primarily sourced from the German MarinEARS portal [7]. These include measurements from six offshore wind farms: Arkona in the Baltic Sea, and Deutsche Bucht, Kaskasi, Sandbank, Trianel, and Veja Mate in the North Sea. Additionally, data from the Vineyard Wind project in the United States were incorporated, covering both monopile and skirt pile installations for the converter platform [8].

The ranges of the considered parameters such as strike energy, pile diameter, water depth, and ram weight is summarized in Table 1.

To ensure consistency, only measurements taken at distances between 670 m and 800 m from the pile were considered. All sound exposure levels (SELs) were scaled to a nominal distance of 750 m using the damped cylindrical spreading model described in [9]. Water depth values

Table 1: Details of the piling configurations used for the validation.

wind farm unit	data points [-]	strike energy [kJ]	pile diameter [m]	water depth [m]	ram weight [t]
Arkona	56	579 - 3220	7.5	22.6 - 36.9	200
Deutsche Bucht	27	1349 - 2146	8.0	39.6 - 40.5	200
Kaskasi	31	1341 - 2656	6.5	21.1 - 27.0	150
Sandbank	60	1516 - 1935	6.8	25.1 - 34.1	175
Trianel	14	1369 - 1892	8.0	28.6 - 33.9	200
Veja Mate	55	1509 - 2252	7.8	39.0 - 40.6	200
Vineyard	12	2488 - 4091	9.6	39.5 - 43.7	200
Vineyard conv	1	1216	2.45	38.6	100

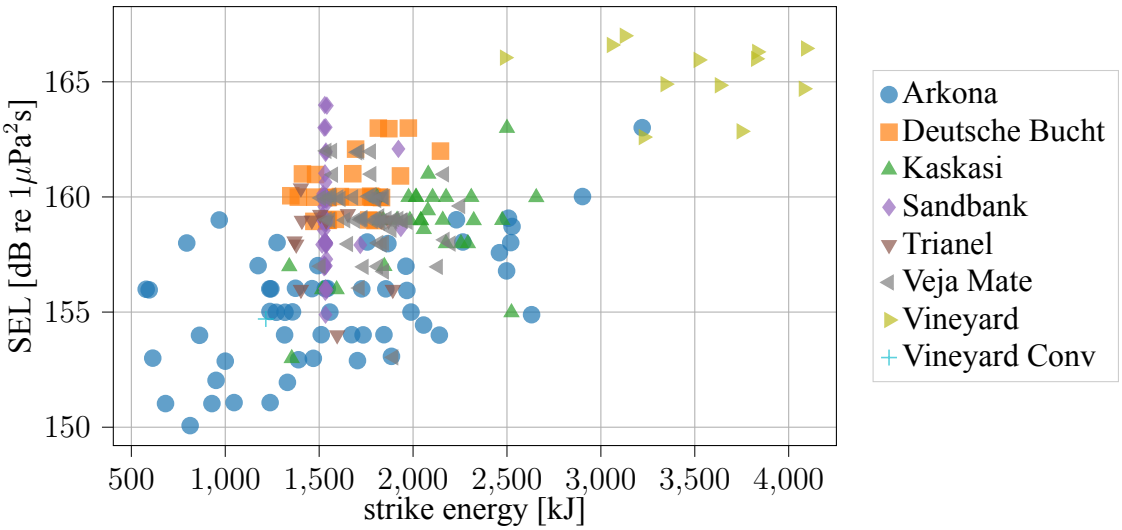


Figure 1: All measured SEL over the maximum strike energy.

for each site were obtained from EMODnet [10]. For the Vineyard Wind project, SELs were available from two directions around 750 m. Therefore, the arithmetic mean of the SEL_{05} values was used in the analysis.

Furthermore, all data points are displayed within Figure 1 over the maximum of the applied strike energy. Therein, the spread of the data from every individual wind farm is visible.

3.2. COMPARISON OF TRENDS

In the following, the scalability of the measurements is evaluated. If not stated differently, all measured data points are scaled to a strike energy of 3000 kJ, a ram weight of 200 t, a water depth of 40 m and a pile diameter of 8 m.

3.2.1. WATER DEPTH

The validation of the influence of the water depth on the SEL with the application of a HSD/DBBC combination is challenging. On the one hand, the water depth variation within a wind farm is usually relatively small and at the same time the water depth influence is expected to be around 1.3 dB per 10 m change in water depth, which is in the range of measurement uncertainties.

Furthermore, the data provided within MarinEARS is rounded and pile designs are also influenced by the water depths.

The influence of the water depth can be evaluated for each wind farm individually or by considering all data points. Only data of two wind farms with changing bathymetry were available. These are Arkona and Sandbank. The measured SELs of these two wind farms are scaled to the same strike energy, ram weight, and pile diameter and displayed within Figure 2 over the water depth. A linear best-fit approximation is conducted that shows a positive relation between the SEL and the water depth. However, quite a large variation around the trend-lines remains. This can partly be linked to the influence of the water depth on the pile design, associated penetration depth, operational parameters of the DBBC, or the site-specific geo-acoustic conditions. Since

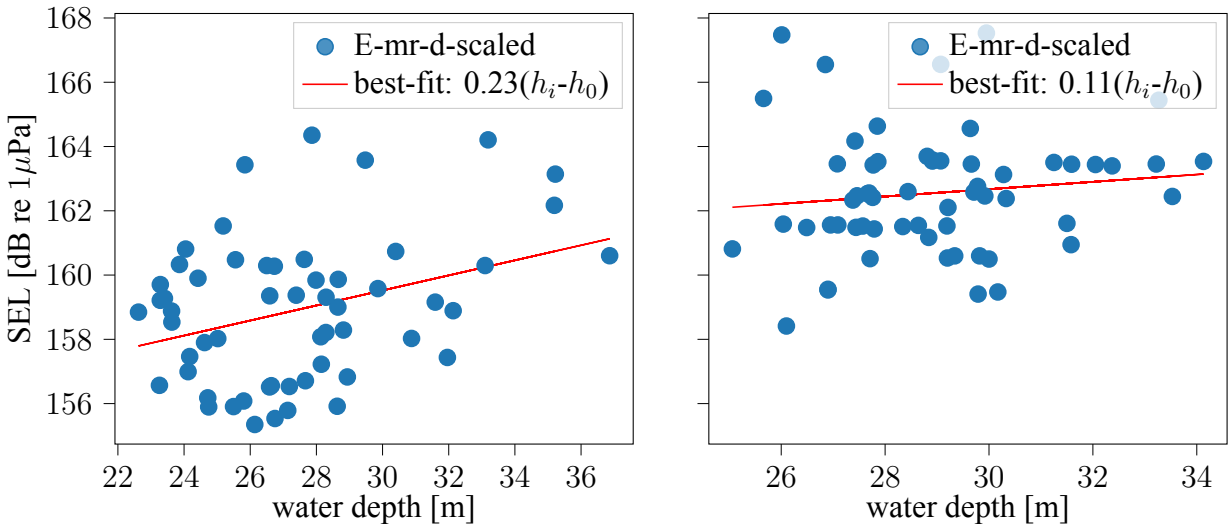


Figure 2: Two examples of energy, ram weight, and diameter scaled SEL for the Arkona (left) and Sandbank (right) offshore wind farm in dependence on the water depth combined with a best-fit approximation of a linear trend-line.

only a relative small water depth is covered by the two examples, the individual data points of all six wind farms are displayed within Figure 3 over the water depth.

Therein, a spread between the different wind farms can be observed. However, a best-fit approximation of a linear trend-line is leading to a 0.13 dB$(h_i - h_0)$-trend line, which is between the two individual results derived for Arkona and Sandbank and the same trend as derived for the DBBC only, cf. Equation (4). Therefore, the 1.3dB/10m scaling law is also considered in the following evaluations.

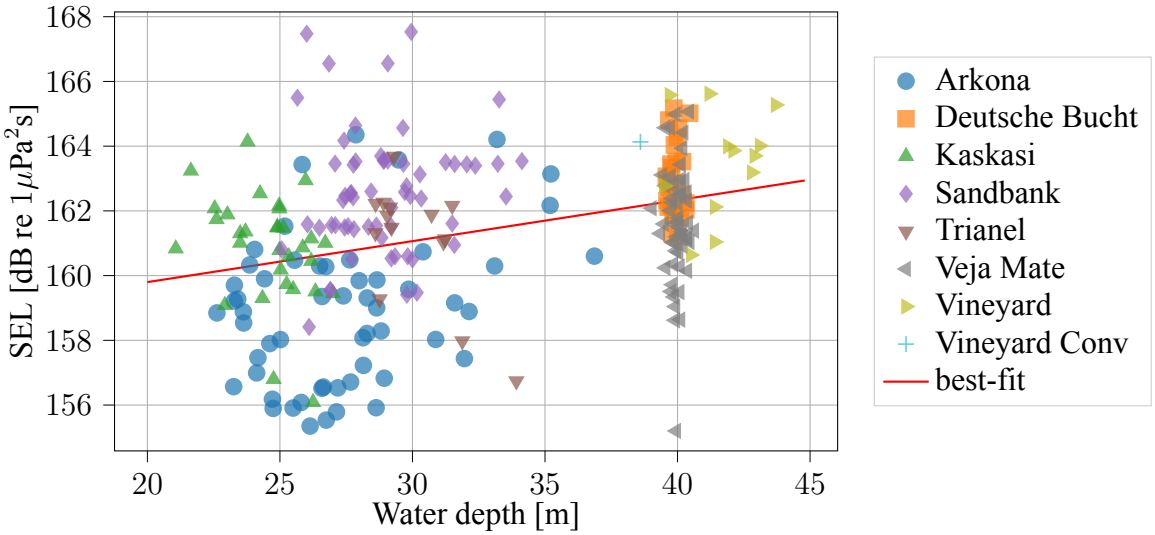


Figure 3: All strike energy, ram weight, and pile diameter scaled SEL over the water depth.

3.2.2. PILE DIAMETER

The comparison of the dependence of the SEL on the pile diameter is conducted by comparing statistical evaluations of the individual wind farms after scaling all data points to the same strike energy, ram weight, and water depth. The resulting median and the 25th and 75th percentile of the results are displayed within Figure 4. The percentiles are shown as the min and max errorbars. The data sets of the German wind farms have pile diameters between 6.5 m and 8 m. The comparison of the German wind farms shows a good agreement in the SEL. The only Baltic Sea data set from Arkona had generally the lowest SELs. Only the converter station of Vineyard Wind and the Vineyard Wind monopiles have considerably deviating diameters. However, the scaled German North Sea scenarios show a good agreement with both the Vineyard monopiles and the skirt pile. Furthermore, a good match between the converter station and the monopile of Vineyard can be seen.

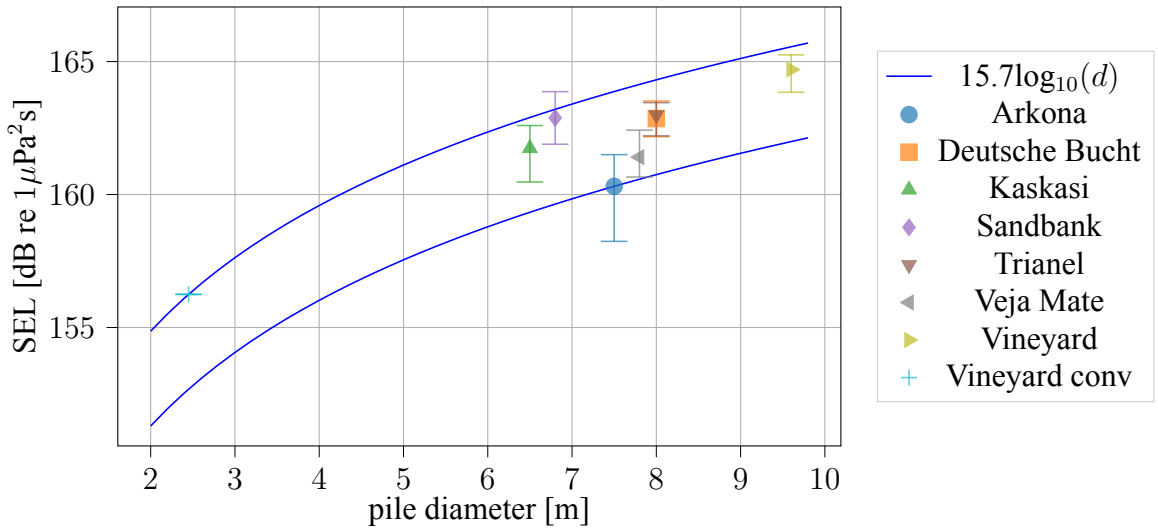


Figure 4: The median and the 25th and 75th percentile represented by the errorbars of every wind farm after scaling to the same strike energy, ram weight, and water depth over the pile diameter in combination with the pile diameter scaling law.

3.2.3. DISCUSSION

The statistical evaluation of the application of all four scaling laws to the measurements is displayed within Figure 5. All resulting median of German wind farms from the North Sea are within a range of 2.5 dB. Arkona is the only wind farm in the Baltic Sea and generally shows slightly lower results and quite a high variation from turbine to turbine. Furthermore, the results from Vineyard wind are very well comparable to the SELs measured in Germany.

The application of the scaling laws initially developed for scenarios with the application of a DBBC shows that their application to HSD/DBBC scenarios of different wind farms leads to reasonable agreement.

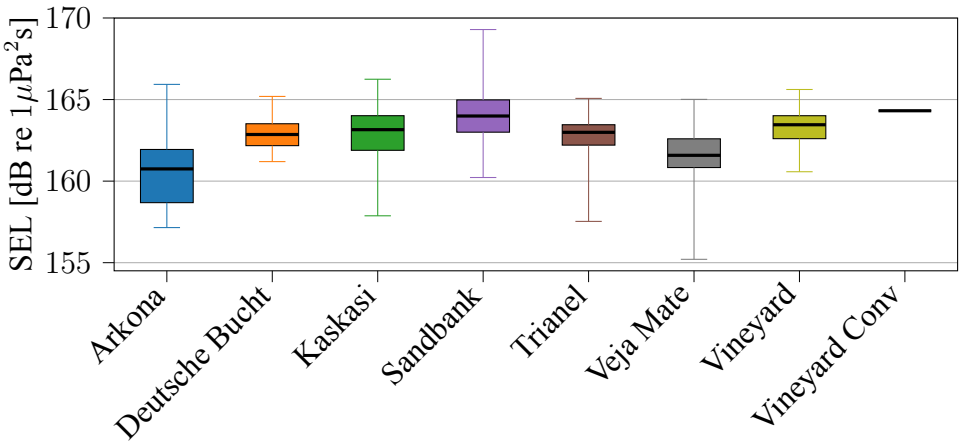


Figure 5: The statistical evaluation of the scaled SEL of every considered wind farm. All measured SEL are scaled to a strike energy of 3000 kJ, a pile diameter of 8 m, a ram weight of 200 t, and a water depth of 40 m. The box represents the 25th and 75th percentile and the errorbars represent the minimal and maximal scaled SEL.

4. COMPARISON OF MEASURED AND SCALED INSERTION LOSS

Scaling laws have shown to be a sufficient tool to get estimations of the SEL for mitigated and unmitigated pile driving. The MarinEARS data base also provides SEL levels for unmitigated pile driving for the six German wind farms considered in Table 1. These can be used to derive the insertion loss achieved with the HSD/DBBC combination. When comparing the unmitigated and mitigated values, they need to be adjusted such that they are scaled to the same strike energy. The available data sets were evaluated this way. If more than one reference measurement was available, they were averaged. The achieved broadband insertion losses are displayed within the left part of Figure 6.

The same evaluation was done but the unmitigated SEL was scaled with the data set provided in [3] combined with a few newly published data sets. The mean of the scaled results were considered as the reference SEL. This leads to a more condensed representation of the insertion loss, cf. right plot in Figure 6. This indicates that the consideration of scaled unmitigated SEL levels can help reduce the influence of potential local outliers.

Both results are well in line with the insertion loss stated by Bellmann *et al.* [2] who provide a range between 18 and 19 dB.

A comparison of the measured and scaled SEL of the unmitigated case is provided in Table 2. The comparison shows that the measured reference SELs of Arkona and Sandbank were relatively high, suggesting on the other hand a relatively high insertion loss.

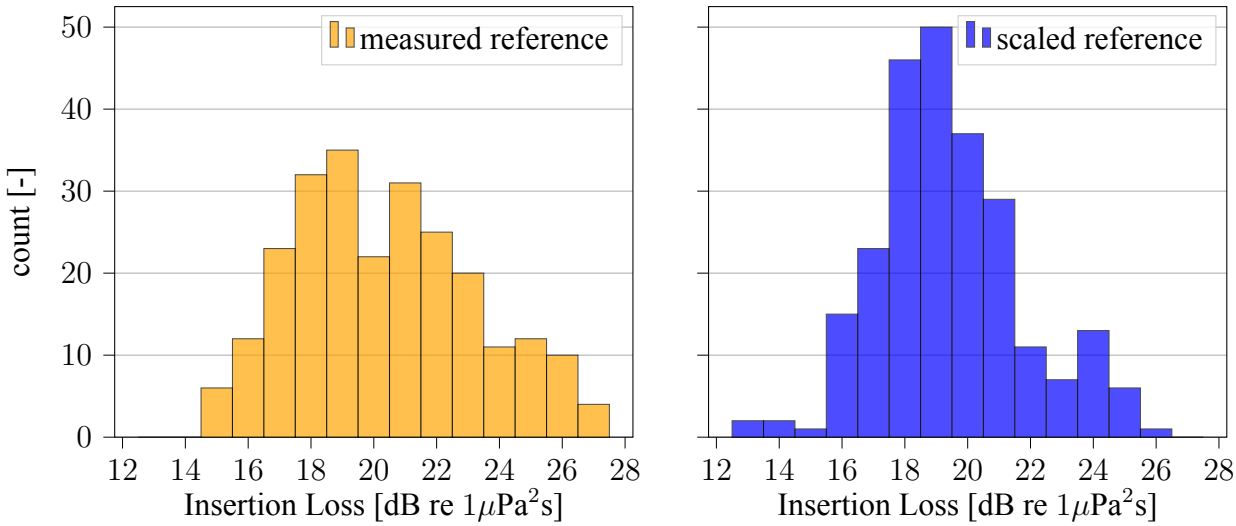


Figure 6: The distribution of the insertion loss derived by comparing the measured unmitigated SEL to the measured HSD/DBBC results (left) and a scaled unmitigated SEL to the measured HSD/DBBC results (right).

Table 2: Comparison of the measured and scaled SEL [dB re 1μPa^{2}s] for unmitigated pile driving in German wind farms.

wind farm	measured SEL	scaled SEL
Arkona	179.0	177.1
Deutsche Bucht	177.8	178.3
Kaskasi	178.5	178.4
Sandbank	180.5	177.1
Trianel	176.1	177.8
Veja Mate	177.9	178.4

5. CONCLUSIONS AND OUTLOOK

The shown comparisons indicate that the scaling laws originally developed for mitigated pile driving with the application of a DBBC can similarly be applied to scenarios where the HSD/DBBC combination is planned to be applied. While the scaled medians showed very good agreement across different sites, significant variability in the measured SELs was observed. This variation was present within each wind farm and is likely linked to differences in DBBC operational parameters, site-specific soil conditions, pile designs, and measurement uncertainties.

As always, great care needs to be taken when choosing the baseline scenario in order to ensure meaningful comparisons between the baseline and evaluated piling scenarios.

The evaluation of the insertion loss demonstrated that applying a scaled reference value reduces variability in the results.

A next step is the investigation of other highly frequency dependent mitigation solutions such as the AdBm system.

REFERENCES

[1] J. von Pein, S. Lippert and O. von Estorff, ”Validation of a finite element modelling approach for mitigated and unmitigated pile driving noise prognosis”, *The Journal of the Acoustical Society of America* **149**(3), 1737-1748 (2021)

[2] M. A. Bellmann, A. May, T. Wendt, S. Gerlach and P. Remmers, ”Underwater noise during the impulse pile-driving procedure: Influencing factors on pile-driving noise and technical possibilities to comply with noise mitigation values”, (2020).

[3] J. von Pein, T. Lippert, S. Lippert and O. von Estorff, ”Scaling laws for unmitigated pile driving: Dependence of underwater noise on strike energy, pile diameter, ram weight, and water depth”, *Applied Acoustics* **198**, (2022).

[4] J. von Pein, T. Lippert, S. Lippert and O. von Estorff, ”Scaling laws for mitigated pile driving: Dependence of underwater noise on strike energy, pile diameter, ram weight, water depth, and mitigation system”, *The Journal of the Acoustical Society of America* **156**, (2024).

[5] B. Bruns, C. Kuhn, P. Stein, J. Gattermann, K.-H. Elmer, ”The new noise mitigation system Hydro Sound Dampers: history of development with several hydro sound and vibration measurements”, Proceedings of Inter-Noise 2014, (2014).

[6] International Organization for Standardization, ”DIN ISO18406:2018-08 - Underwater acoustics - Measurement of radiated underwater sound from percussive pile driving (ISO 18406:2017)”, Geneva, Switzerland, (2018).

[7] MarinEARS Marine Explorer and Registry of Sound specialist information system for underwater noise and national noise-register for the notification of impulsive noise events in the German EEZ of the North- and Baltic Sea to the EU according to the MSF (2025).

[8] E.T. Küsel, C. Graupe, T. J. Stephen, C. Lawrence, M. P. Cotter, and D.G. Zeddies, ”Underwater Sound Field Verification: Vineyard Wind 1 Final Report”. Document 03233, Version 1.0. Technical report by JASCO Applied Sciences for DEME Group, (2024).

[9] T. Lippert, M. A. Ainslie and O. von Estorff, ”Pile driving acoustics made simple: Damped cylindrical spreading model”, *The Journal of the Acoustical Society of America* **143**(1), 310-317 (2018).

[10] European Marine Observation and Data Network (EMODNet), (2025).

Underwater Noise in Offshore Impact Piling Including Pile-Soil Contact Relaxation

Canny, Khairina A.[1], Peng, Yaxi[1], Tsetas, Athanasios[1], and Tsouvalas, Apostolos[1]

[1]Department of Civil Engineering and Geosciences, Delft University of Technology

Corresponding author's email:

Abstract: *As the trend shifts toward the installation of larger foundation piles for offshore wind farms, which are associated with lower frequency excitations, accurately predicting the resulting sound and vibrations requires a precise characterization of soil behaviour and pile-soil interaction. In addition to noise emissions caused by pile installation, substrate-borne vibrations are particularly perceptible to various marine biota. Both seabed vibrations and underwater noise raise concerns about ecological impacts, emphasizing the need for predictive models that accurately represent the interactions between pile, soil, and seawater. This paper examines the effects of the inclusion of the pile-soil contact mechanism during impact pile driving both in the underwater soundand the seabed vibrations. The pile-soil mechanism condition is modelled by the introduction of linear springs at thepile-soil interface allowing for relative displacement to develop between the soil and the pile. A case study is conducted to explore the implications of the contact mechanism, focusing on the two key outputs: the noise levels in thesurrounding fluid and particle motion within the substrate. Sensitivity analysis is performed to evaluate how variations in contact conditions during impact piling influence these critical metrics.*

Keywords: *underwater noise, offshore pile driving, fluid-soil-structure interaction, contact stress relaxation*

1. INTRODUCTION

Offshore wind energy is scaling rapidly, with turbine sizes increasing and foundation systems requiring larger and deeper-installed piles. These large-diameter monopiles introduce complex dynamic behavior, particularly in the low-frequency range, where soil-pile interaction becomes a dominant factor in the response. Growing evidence confirms that low-frequency substrate vibrations affect fish and aquatic invertebrates, as shown in recent studies [1, 2]. To better predict these environmental impacts, a more realistic fluid-soil-pile interaction model should account for the contact mechanism. Seminal works by Tsouvalas and Metrikine [3] and Peng et

al. [4] used semi-analytical vibroacoustic models to account for elastic properties of the marine sediment, emphasizing the role of interface waves. Tsetas et al. [5] extended this approach to vibratory pile driving, introducing non-linear contact mechanisms between pile and soil (i.e. history dependent friction). Building on this, Molenkamp et al. [6] employed a two-step boundary element approach to couple pile dynamics with acoustic propagation, finding that contact mechanism reduces Scholte wave amplitudes but enhances near-resonant pile motion. Bohne et al. [7] explored these dynamics numerically for smaller impact-driven piles, while Di et al. [8] addressed soil-pile behaviour in the interface in the context of nondestructive testing. This paper extends the semi-analytical model of [3] by explicitly incorporating pile-soil contact mechanism into impact pile driving scenarios. This sections are organized as follows. The next section introduces the fundamental geometrical configuration and material properties of the model. In Section 2, we derive the governing equations for the cylindrically symmetric case, providing the analytical foundation of the proposed method. Section 3 presents a case study to analyze the effects of soil contact stress relaxation. Finally, Section 4 summarizes the main findings and discusses potential directions for future research.

2. CYLINDRICALLY SYMMETRIC CASE

In this section, we derive the complete set of governing equations for the cylindrically symmetric configuration, following the framework established in [3, 4]. Our formulation extends previous models by incorporating a contact mechanism at the pile-soil interface, modeled through a viscoelastic constitutive relation.

2.1. MODEL DESCRIPTION AND GOVERNING EQUATIONS

The near-source model of impact pile driving is illustrated in Figure 1. In this model, the pile is represented as a shell occupying the domain $0 \leq z \leq L$. Surrounding the pile, the fluid domain is modeled as a three-dimensional, inviscid, and compressible medium, characterized by a compressional wave speed c_f and mass density ρ_f. The fluid occupies the region $z_0 \leq z \leq z_1$, where z_0 and z_1 correspond to the sea surface and seabed levels, respectively. Below the seabed, the soil is described as a three-dimensional elastic medium with density ρ_s, and compressional and shear wave velocities c_p and c_s, respectively. In the near-source approach, the soil extends downward from the seabed ($z = z_1$) to the bedrock level ($z = z_2$). The soil-pile interaction is bounded at a lower depth $z = L$ by a rigid boundary, which has been shown to exert a negligible influence on the dynamic response of the pile and, by extension, on the mechanisms of underwater noise generation. The governing equations for this coupled system are derived in line with the formulation presented in [4], and are expressed in Equations (1)–(2), which describe the dynamic behavior of the pile, fluid, and soil domains under impact loading conditions.

$$\mathbf{L}\mathbf{u}_{pl} + \mathbf{I}_m \ddot{\mathbf{u}}_{pl} = -[H(z - z_1) - H(z - L)]\mathbf{t}_s + [H(z - z_0) - H(z - z_1)]\mathbf{p}_f + \mathbf{f} \quad (1)$$

$$\nabla^2 p_f(r, z, t) - \frac{1}{c_f^2}\ddot{p}_f(r, z, t) = 0 \quad \text{and} \quad (\lambda + 2\mu)\nabla(\nabla \cdot \mathbf{u}_s) - \mu\nabla \times (\nabla \times \mathbf{u}_s) = \rho_s \ddot{\mathbf{u}}_s \quad (2)$$

The notation $\mathbf{L}$ and $\mathbf{I}_m$ represent the stiffness and inertial matrices, respectively following the formulation provided in [5]. The pile displacement vector is denoted by $\mathbf{u}_{pl}(z, t) =$

$[u_{pl,r}, u_{pl,z}]^{\mathrm{T}}$, where $u_{pl,r}$ and $u_{pl,z}$ correspond to the radial and vertical displacements in the axisymmetric case. These components describe the dynamic deformation of the pile under external and internal forces. The external force acting on the pile surface is represented as $\mathbf{f}(z,t) = [f_{rr}, f_{rz}]^{\mathrm{T}}$, accounting for the radial and shear tractions imposed by the surrounding medium. The function $H(z-z_i)$ denotes the Heaviside step function, which is used to activate the material interfaces at specific depths within the fluid and soil domains.

The governing equations that characterize the dynamic response of the fluid and the elastic soil media are described in Equation (2). A Helmholtz decomposition is employed to facilitate their analytical solution. Specifically, the velocity field in the fluid is represented as $\mathbf{v}_f = \nabla\phi_f$. The displacement field in the elastic soil medium is expressed as $\mathbf{u}_s = \nabla\phi_s + \nabla\times\left(0, -\frac{\partial\psi_s}{\partial r}, 0\right)$. These scalar and vector potentials represent compressional and shear wave components, respectively, and allow to decouple the elastic wave propagation modes. Applying the Fourier transform, the elastic wave equation can be written in terms of the potential functions, capturing both compressional and shear behavior in the soil medium. These potentials are written as (3)

$$\begin{aligned} \nabla^2\tilde{\phi}_f(r,z,\omega) &= -\frac{\omega^2}{c_f^2}\tilde{\phi}_f(r,z,\omega) \\ \nabla^2\tilde{\phi}_s(r,z,\omega) &= -\frac{\omega^2}{c_p^2}\tilde{\phi}_s(r,z,\omega); \quad \nabla^2\tilde{\psi}_s(r,z,\omega) = -\frac{\omega^2}{c_s^2}\tilde{\psi}_s(r,z,\omega) \end{aligned} \tag{3}$$

The notations ϕ_f, ϕ_s, and ψ_s denotes the fluid potential, soil compressional potential, and soil shear potential, respectively.

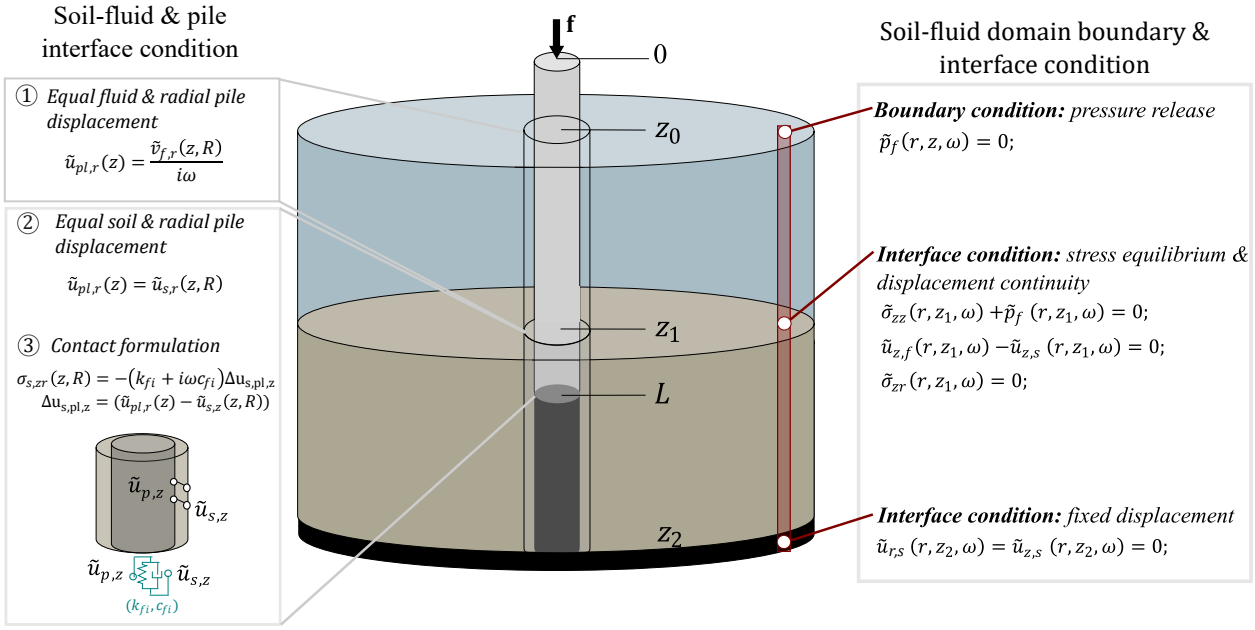


Figure 1: Illustration of the model description for fluid-soil-pile interaction and the contact interface representation using elastic spring-dashpot.

2.2. SOLUTION TO THE COUPLED VIBROACOUSTIC PROBLEM

The modal matching technique employed in the near field model builds upon the framework established in the study by Tsouvalas [3] and Peng [4], and the modal representation used to model the shell structure uses the approach developed in [5]. Applying the work of Tsetas et al. [5], the semi-analytical finite element (SAFE) model is applied for thin shells, which is based on the Love-Timoshenko shell theory; consequently, the shell displacement incorporating both radial and axial motion is expressed as a modal summation in Equations (4), where each mode satisfies the governing equations and boundary conditions of the shell structure.

$$\tilde{u}_{pl,r}(z,\omega) = \sum_{m=1}^{\infty} A_{0m} U_{pl,r0m}(z) \quad \text{and} \quad \tilde{u}_{pl,z}(z,\omega) = \sum_{m=1}^{\infty} A_{0m} U_{pl,z0m}(z) \tag{4}$$

The unknown modal constants, A_{0m}, will be determined through the coupled system, with the subscript m denoting the axial mode number. In this context, the terms $U_{pl,r0m}$ and $U_{pl,z0m}$ represent the pile eigenvectors in the radial and vertical directions, respectively.

As for the fluid potentials, the solution satisfying the interface and boundary conditions shown in Figure 1 is expressed as a summation of modes, as given by Equations (5):

$$\begin{aligned} \tilde{v}_{f,z}(r,z,\omega) &= \sum_{p=1}^{\infty} C_p H_0^{(2)}(k_p r)\tilde{v}_{f,z,p}(z); \quad \tilde{v}_{f,r}(r,z,\omega) = \sum_{p=1}^{\infty} C_p H_1^{(2)}(k_p r)\tilde{v}_{f,r,p}(z) \\ \tilde{p}_f(r,z,\omega) &= \sum_{p=1}^{\infty} C_p H_0^{(2)}(k_p r)\tilde{p}_{f,p}(z) \end{aligned} \tag{5}$$

Likewise, the soil potentials satisfy the interface and boundary conditions, as shown in Figure 1). Therefore, they are expressed as a summation of modes, as written in Equation (6)

$$\begin{aligned} \tilde{u}_{s,z}(r,z,\omega) &= \sum_{p=1}^{\infty} C_p H_0^{(2)}(k_p r)\tilde{u}_{s,z,p}(z); \quad \tilde{u}_{s,r}(r,z,\omega) = \sum_{p=1}^{\infty} C_p H_0^{(2)}(k_p r)\tilde{u}_{s,r,p}(z) \\ \tilde{\sigma}_{s,zz}(r,z,\omega) &= \sum_{p=1}^{\infty} C_p H_0^{(2)}(k_p r)\sigma_{s,zz,p}(z); \quad \tilde{\sigma}_{s,zr}(r,z,\omega) = \sum_{p=1}^{\infty} C_p H_1^{(2)}(k_p r)\sigma_{s,zr,p}(z) \\ \tilde{\sigma}_{s,rr}(r,z,\omega) &= \sum_{p=1}^{\infty} C_p \left(H_0^{(2)}\tilde{\sigma}_{s,rr,p}^{H_0}(z) + \frac{1}{r} H_1^{(2)}\tilde{\sigma}_{s,rr,p}^{H_1}(z) \right) \end{aligned} \tag{6}$$

The soil displacement in the radial and vertical directions is represented as $\tilde{u}_{s,r}$ and $\tilde{u}_{s,z}$, respectively. The stresses in the soil, expressed in cylindrical coordinates, are denoted by $\sigma_{s,zr}$, $\sigma_{s,rr}$, and $\sigma_{s,zz}$. Additionally, C_p is a complex-valued coefficient determined by solving the force response of the coupled system. The functions $H_0^{(2)}$ and $H_1^{(2)}$ are the Hankel functions of the second kind, which describe the propagation of cylindrical waves in an unbounded medium.

Interface condition 1: The radial velocity at the fluid level must match the radial velocity at the pile interface. This is expressed as $i\omega \cdot \tilde{u}_{pl,r}(z,\omega) = \tilde{v}_{f,r}(R,z,\omega)$ within the fluid domain $z_0 < z < z_1$ and at the pile interface $r = R$. The detailed expression for this condition is substituted from Equations (5) and (6). This interface condition is represented as Equation (7):

$$i\omega \sum_{m=1}^{\infty} A_{0m} \cdot U_{pl,r,0m}(z) = \sum_{p=1}^{\infty} C_p \cdot H_1^{(2)}(k_p R)\tilde{v}_{f,r,p}(z) \tag{7}$$

Interface condition 2: The displacement in the radial direction is continuous across the pile and soil interface. This condition is written as $\tilde{u}_{pl,r}(z,\omega) = \tilde{u}_{s,r}(R,z,\omega)$ within the soil domain $z_1 < z < L$ and at the pile interface $r = R$. The expression for this condition is substituted from Equations (5) and (6). The corresponding interface condition is given by Equation (8):

$$\sum_{m=1}^{\infty} A_{0m} \cdot U_{pl,r,0m}(z) = \sum_{p=1}^{\infty} C_p \cdot H_1^{(2)}(k_p R)\tilde{u}_{s,r,p}(z) \tag{8}$$

Interface condition 3: A constitutive relation for the contact in the z-direction is introduced, incorporating an elastic spring and dashpot at the pile contact interface, denoted as k_{fi} and c_{fi}, respectively, as shown in Figure 1. This relation is expressed as $\tilde{\sigma}_{s,zr}(z,R,\omega) = -(k_{fi} + i\omega c_{fi}) \cdot \big(\tilde{u}_z(z) - \tilde{u}_{s,z}(z)\big)$ within the soil domain $z_1 < z < L$ and at the pile interface $r = R$. The corresponding form, derived from Equations (5) and (6), is given in Equation (9):

$$\tilde{\sigma}_{s,zr}(z,R,\omega) = -(k_{fi} + i\omega c_{fi}) \left(\sum_{m=1}^{\infty} A_{0m} \cdot U_{pl,z,0m}(z) - \sum_{p=1}^{\infty} C_p \cdot H_0^{(2)}(k_p R)\tilde{u}_{s,z,p}(z) \right) \tag{9}$$

In order to ensure the stability of Equation (9), we divide both sides of the Equation by $(k_{fi} + i\omega c_{fi})$. This results in the following equation (10):

$$-\frac{\sum_{p=1}^{\infty} C_p H_1^{(2)}(k_p R)\sigma_{s,zr,p}(z)}{(k_{fi} + i\omega c_{fi})} + \sum_{p=1}^{\infty} C_p \cdot H_0^{(2)}(k_p R)\tilde{u}_{s,z,p}(z) = \sum_{m=1}^{\infty} A_{0m} \cdot U_{z,0m}(z) \tag{10}$$

Meanwhile, the full contact limit ($\lim_{k_{fi}\to\infty}$) effectively cancels out the first term in equation (10), resulting in $\sum_{m=1}^{\infty} A_{0m}\, U_{z,0m}(z) = \sum_{p=1}^{\infty} C_p\, H_0^{(2)}(k_p R)\, \tilde{u}_{s,z,p}(z)$. To proceed, the operations on Equations (7), (8), and (10) are carried out to obtain the soil-fluid orthogonality. Afterwards, the modal matching method is applied to solve the pile-soil-fluid interaction. These steps follow the same framework as outlined in reference [3]. However, the main modification in the subsequent analysis will be in Equation (10), where the introduction of the spring-dashpot contact model alters the interaction terms between the soil and pile.

3. CASE STUDY

In this section, we examine the impact of contact mechanism on the interaction between the pile and the surrounding soil. Through a sensitivity analysis, we explore how changes in the contact parameters influence the system, with a particular focus on large pile diameters. The case study, as depicted in Figure 2, highlights the key features and setup for this analysis.

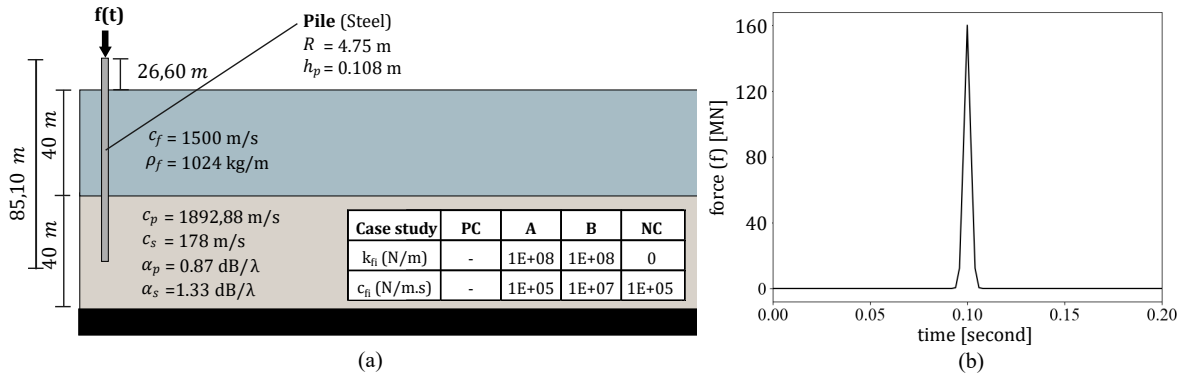


Figure 2: The case study properties: (a) the basic dimension and the sensitivity analysis for multiple contact parameters; (b) the hammer force in time domain.

The overall fluid pressure in the time domain, as shown in Figure 3, along with the snapshots of the pressure field in Figure 4, both reveal that the fluid pressure oscilation increase as the contact at the pile-soil interface is relaxed.

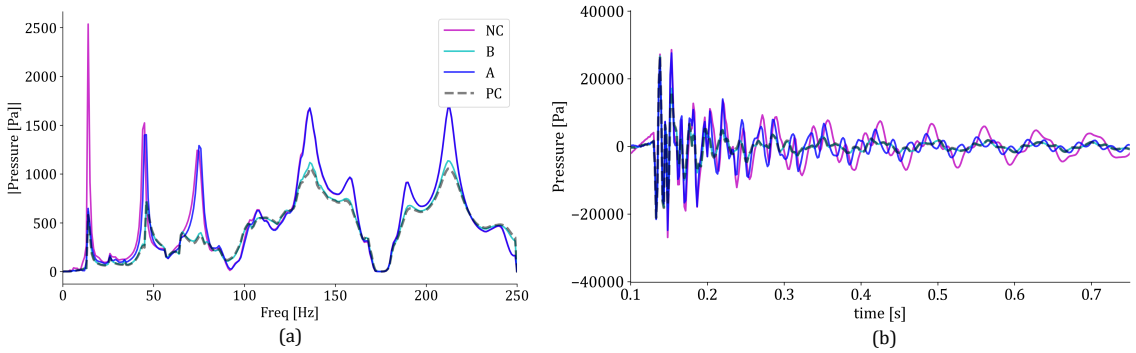


Figure 3: The fluid pressure at r =40m and z=20m: (a) Frequency response of the fluid pressure under 1MN constant amplitude frequency sweep and (b) time-domain fluid pressure.

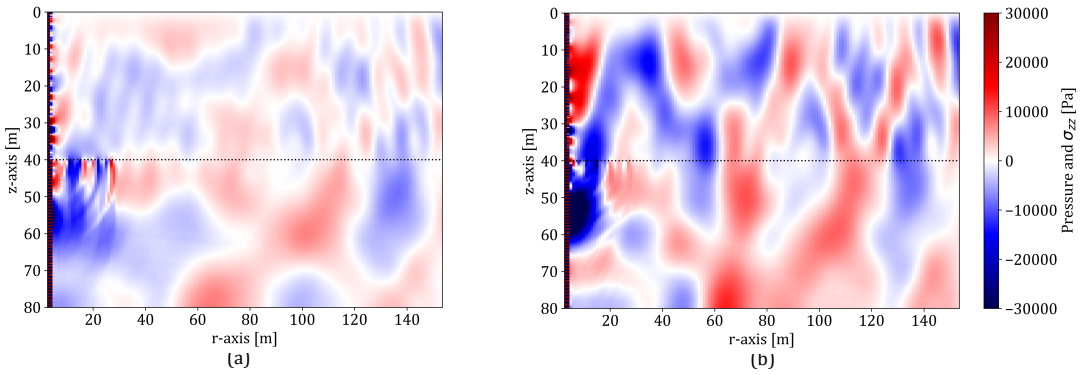


Figure 4: The fluid pressure and soil stress field snapshot at t=0.25 second: (a) perfect contact (PC), and (b) no contact (NC).

In contrast to the fluid pressure, Figures 5 and 6 illustrate that seabed soil vibrations decrease as the contact condition is relaxed compared to the perfect contact case. This holds true for

both the radial and vertical directions. These trends align with the frequency-domain analysis presented in Figures 5a and 6a. Moreover, increasing the damping constant (Case B) leads to a reduction in the pile dynamics, which, in turn, results in a decrease in fluid pressure oscillation and an increase in soil vibration, as seen in Figure 3,5, and 6. As the contact damping increases, the system's response gradually approaches the one observed under perfect contact conditions.

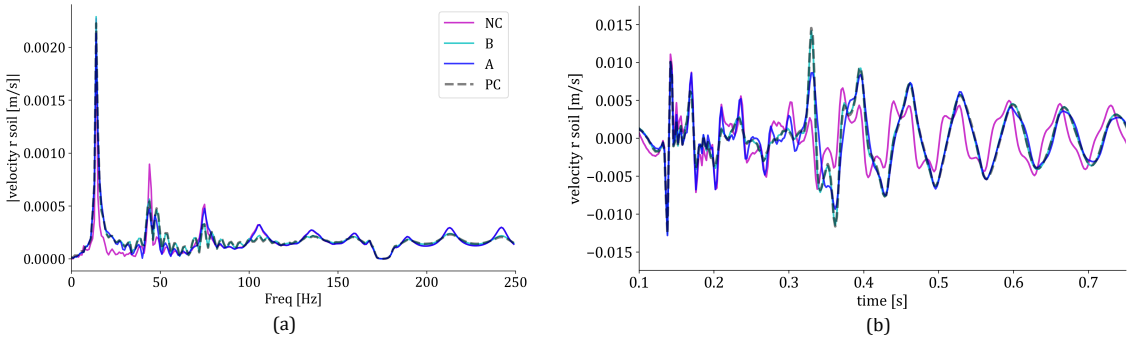


Figure 5: The soil velocities at the interface, r=40 m : (a) Frequency response of the radial velocity under 1MN constant-amplitude frequency sweep, and (b) Time-domain representation of the radial velocity.

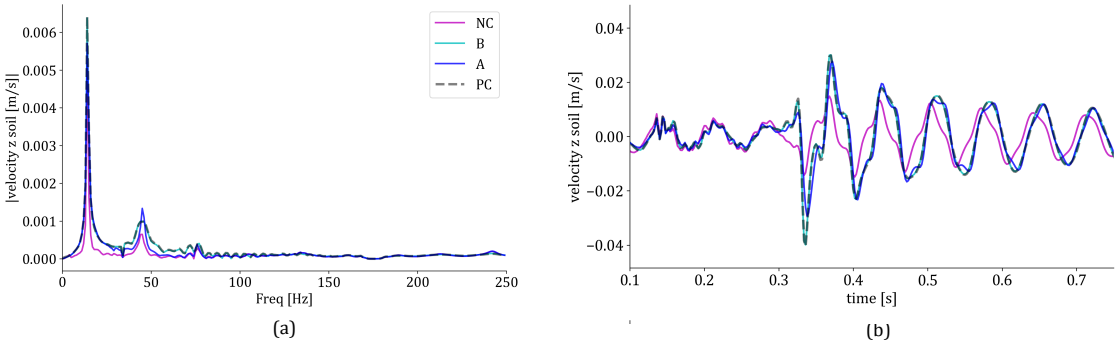


Figure 6: The soil velocities at the interface at r =40m and z=20m: (a) Frequency response of the vertical velocity under 1MN constant-amplitude frequency sweep, and (b) Time-domain representation of the vertical velocity

4. CONCLUSION AND FUTURE WORK

This study concludes that incorporating contact interface formulation enhances pile dynamics by relaxing the perfect contact constraint at the soil-pile interface. The results for impact pile driving are consistent with the vibro-pile prediction model proposed by [6], particularly in demonstrating a reduction in wave amplitudes at the soil-fluid interface. Furthermore, fluid pressure oscillations increase as the pile-soil interface becomes more relaxed. Given that fluid pressure and particle motion near the seabed can significantly affect marine life especially benthic species such as invertebrates and bottom-dwelling fish, these findings underscore the importance of accurately modeling interface behavior.

The future work will focus on improving the modeling approach by treating the contact mechanism parameter as a calibration variable for soil characterization, validated against experimental data. Another area of development will be to incorporate more advanced models for contact mechanism that reflect media specific behavior, like Maxwell-type spring-dashpot models for mud-like seabed conditions. The model will also be expanded to include depth and frequency-dependent soil properties, accounting for variations in density, contact behavior, and compressional and shear wave velocities.

5. ACKNOWLEDGEMENT

The authors gratefully acknowledge SEASOUNDS for providing financial support for this research project. SEASOUNDS has received funding from the European Union's Horizon Programme under the Marie Skłodowska-Curie actions HORIZON-MSCA-2022-DN-01 (Grant agreement ID: 101119769

REFERENCES

[1] M. Solé, K. Kaifu, T. A. Mooney, S. L. Nedelec, F. Olivier, A. N. Radford, M. Vazzana, M. A. Wale, J. M. Semmens, S. D. Simpson, G. Buscaino, A. Hawkins, N. A. de Soto, T. Akamatsu, L. Chauvaud, R. D. Day, Q. Fitzgibbon, R. D. McCauley, M. André: "Marine invertebrates and noise", *Frontiers in Marine Science* **10**, (2023).

[2] A. D. Hawkins, R. A. Hazelwood, A. N. Popper, P. C. Macey: "Substrate vibrations and their potential effects upon fishes and invertebrates", *Journal of the Acoustical Society of America* **149**(4), 2782–2790 **(2021)**.

[3] A. Tsouvalas, A. V. Metrikine: "A three-dimensional vibroacoustic model for the prediction of underwater noise from offshore pile driving", *Journal of Sound and Vibration* **333**, 2283-2311 **(2014)**.

[4] Y. Peng, A. Tsouvalas, T. Stampoultzoglou, A. Metrikine: "A fast computational model for near- and far-field noise prediction due to offshore pile driving", *The Journal of the Acoustical Society of America* **149**(3), 1772-1790 **(2021)**.

[5] A. Tsetas, A. Tsouvalas, A. V. Metrikine: "A non-linear three-dimensional pile-soil model for vibratory pile installation in layered media", *International Journal of Solids and Structures* **269**, (2023).

[6] T. Molenkamp, A. Tsetas, A. Tsouvalas, A. Metrikine: "Underwater noise from vibratory pile driving with non-linear frictional pile-soil interaction", Journal of Sound and Vibration 576, (2024).

[7] T. Bohne, T. Grießmann, R. Rolfes: "Comprehensive analysis of the seismic wave fields generated by offshore pile driving: A case study at the BARD Offshore 1 offshore wind farm", *The Journal of the Acoustical Society of America* **155**, 1856-1867 **(2024)**.

[8] T. Di, W. Wu, Y. Zhang, Z. Wang, X. Liu, M. Xu, G. Mei: "Semi-analytical approach for nondestructive test for offshore pipe piles considering soil-pile slippage induced by heavy hammer impact", *Ocean Engineering* **260**, 112080 **(2022)**.

Acoustic emissions from offshore monopile installation using jetting technology

Matej Simurda and Areti Ioanna Tsigkinopoulou[1]

[1]Ørsted Wind Power A/S, Fredericia Denmark

Matej Simurda, matsi@orsted.com, Orsted Wind Power A/S, Kraftvaerksvej 53, 7000 Fredericia Denmark.

Abstract: *This study presents comprehensive underwater noise recordings collected from first offshore installation of large monopile foundations using a novel low-noise jetting technology in Ørsted's German project Gode Wind 3. Variability of the measured sound levels is assessed across different installation locations, distances from the foundation and phases of the installation. Measurements are analysed against standard acoustic metrics such as peak pressure level, sound pressure level and cumulative sound exposure level. Additionally, the measurements are compared against typical previously measured acoustic emissions from impact piling. Both sets of data are integrated with a sound propagation model to derive corresponding sound source levels and simulate propagation to extended distances. The results are subsequently discussed in the context of various underwater noise regulations from different markets. It is demonstrated how the low-noise jetting technology produces significantly lower sound levels by more than 30 dB. Consequently, this also yields lower distances to environmental thresholds.*

Keywords: *Pile Driving, Underwater Noise, Noise mitigation, Acoustic measurements.*

1. INTRODUCTION

The rapid expansion of offshore wind energy projects has brought significant attention to the environmental impacts associated with the installation of wind turbine foundations. One of the primary concerns is the underwater noise generated during the installation process, particularly from traditional methods such as impact piling. These high noise levels can have detrimental effects on marine life, including behavioural disturbances and physical harm to marine mammals and fish [1, 2].

In response to these environmental challenges, there is a growing need for innovative installation techniques that minimize underwater noise emissions. The industry has initially focused on reduction of acoustic emissions from impact piling at the source. These solutions are typically referred to as noise abatement systems (NAS). Examples of NAS are big bubble curtains, Hydro-sound damper or AdBm [3]. These systems are considered mature and capable of removing majority of acoustic emissions in the water column. However, NAS alone are often these days not sufficient to comply with regulatory requirements due to the growing size of foundations and increasingly more stringent underwater acoustic regulations. Primary focus over the past years has therefore been on reduction of sound emissions of the source itself. This either means modifying the traditional impact piling method or replacing it with another technology. Recent examples of such incentives are announcements of intended GBM trials [4], deployment of CAPE Holland Vibro Lifting technology [5] or further development of the IQIP's EQ-Piling method [6].

Ørsted has pioneered the use of a novel low-noise jetting technology in its German offshore wind project, Gode Wind 3 demonstrating a significant advancement in reducing the acoustic footprint of offshore installations.

This study presents a comprehensive analysis of underwater noise recordings collected during the first offshore installation of large monopile foundations using the jetting technology. Three monopiles of maximal outer diameter of 9 meters were installed to their final penetration depths. The recordings are evaluated using standard acoustic metrics including peak pressure level, sound pressure level, and cumulative sound exposure level [7]. Variability of sound levels is assessed across different installation locations, distances from the foundation, and phases of the installation process. The measurements are further compared against typical acoustic emissions from conventional impact piling methods [3] and discussed in the context of various underwater noise regulations from different markets. The findings demonstrate that the low-noise jetting technology produces significantly lower sound levels, achieving reductions of more than 30 dB compared to traditional methods. This substantial decrease highlights the potential for this technology to mitigate the environmental impact of offshore wind farm installations.

2. MEASUREMENT METHODOLOGY

All underwater noise measurements were conducted using Soundtrap ST600 [9] underwater noise measuring device with sampling rate and processing setup to capture frequency range from 10 Hz to 16 kHz. The hydrophones were calibrated using a pressure chamber method, ensuring compliance with the requirements of DIN 45653 [10] and ISO 18406 [11] standards.

The measurement layout consisted of fixed positions set at three distinct distances from the installation site: 500 meters, 750 meters, and 1500 meters. Two measurement positions were established at both 500 meters and 750 meters, while a single position was set at 1500 meters.

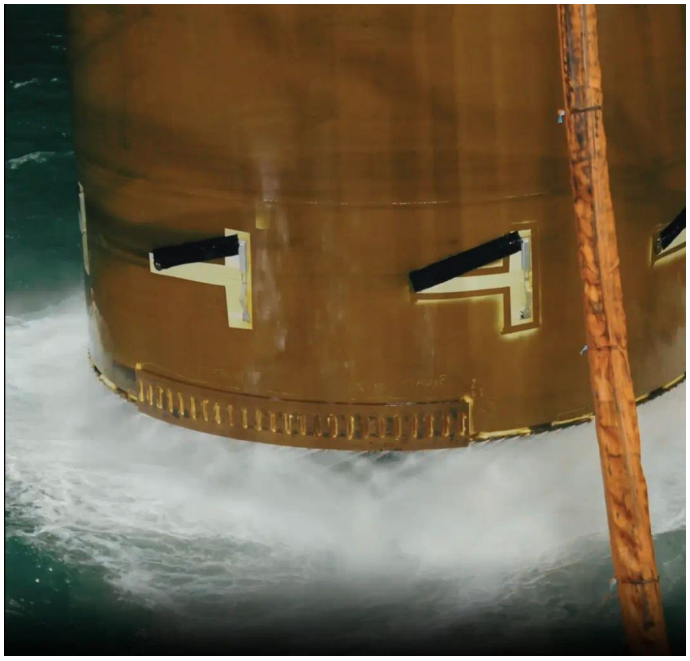


Figure 1: Lower-noise-technology used to install monopile foundations at Gode Wind 3 in Germany [8].

These positions were arranged along two transects oriented 180 degrees from each other: one in an east/north-east direction and the other in a west/south-west direction.

The acoustic emissions from the jetting process can be categorized as a continuous and non-impulsive sound source. The pressure time series measurements were, therefore, analyzed in terms of sound pressure level (SPL) with averaging period of 5 seconds and peak pressure level ($L_{p,pk}$), as defined by ISO 18405 ([7])

3. INSTALLATION METHOD

The tested low-noise installation method utilizes a jetting system mounted at the toe of the pile around its entire annulus. The system injects high-pressure water into the surrounding soil breaking structure-soil interaction and allowing the monopile to descend toward its final penetration depth. Specific phases of the installation procedure could in most cases be identified in the recordings with their acoustic signatures as discussed in the next section.

4. UNDERWATER NOISE MEASUREMENT RESULTS

Example of measured SPL over the duration of the jetting installation is presented in Figure 2. The example shows recordings sound levels at a distance of 750 m from the installation location. Chronologically stepping through the different phases of the installation, a pressure test was observed first with sound levels clearly elevated above ambient noise. The pressure test involves flushing of jetting nozzles. The toe of the pile is in approximately middle of the water column and pumps are operating at the medium pressure. This was followed by a long interval corresponding to the positioning of the installation vessel, deployment of the

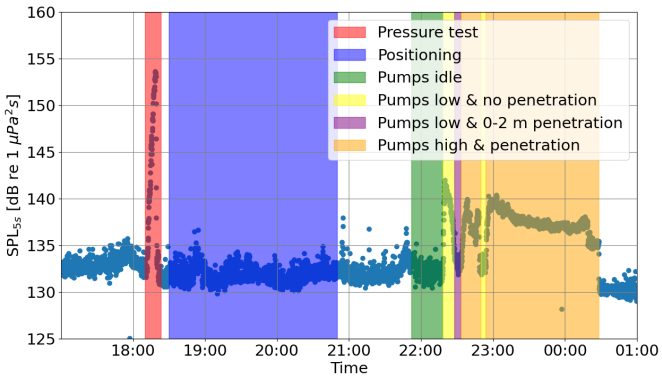


Figure 2: Example of SPL *time series at distance of 750 m from foundation with different installation phases highlighted.*

reference frame and setting the monopile into a pile gripper. No significant increase of sound levels is observed during this interval. The same can be said about the subsequent phase where pumps are turned on but operating on idle pressure. The sound levels start increasing once the operating pressure increases and the toe of the pile is still within the water column. The same pressure on pumps is maintained as the pile is lowered towards the seabed and the pile starts penetrating the first meters of the sediment which results in a decrease in the measured sound levels back to the ambient values. The pumps' operating pressure afterwards increases again resulting in elevation of sound levels close to the values observed during the first meters of penetration. A temporal decrease in sound levels is also observed during this time interval which corresponds to a temporal decrease of the pumps' operating pressure. The pressure is then lowered down for a few minutes and sound levels drop down to ambient values. Finally, the installation using the high pressure is resumed and continues until final penetration depth is reached. The sound levels initially slowly decrease by approximately 3 dB over the duration of 1 hour and 20 minutes. They then suddenly drop by another 2 dB which is most likely caused by lowering the operating pressure slightly.

Equivalent features can be observed for $L_{p,pk}$ time series . There is however, greater variability in the measured $L_{p,pk}$ values and lower signal-to-noise ratio which is a common observation since the peak pressure level quantifies instantaneous maximal amplitude in pressure measurements. Given the values tend to be very low (typically well below 160 dB of $L_{p,pk}$) this acoustic metric is not discussed in greater detail in the rest of this work as it is highly unlikely it will cause any significant environmental impacts/produce large distances to environmental thresholds currently applicable in underwater noise regulations across markets.

Similarly, the pressure test is not a requirement for future deployments and pressure testing can be done onshore / on board the vessel. The rest of this work, therefore focuses on discussing acoustic measurements and the corresponding implications to environmental impacts from the jetting phases only.

The following key findings applicable to all installed foundations can be derived from the measurements: 1) Positioning cannot be reliably detected from the recordings and does not significantly contribute to the ambient sound levels. 2) Pumps operating at idle pressure do not results in elevation of measured sound levels. 3) Pumps operating at low pressure result in elevated sound levels only if the toe of the pile is not embedded into the sediment. 4) Sound levels during the first meters of penetration drop toward ambient sound levels when the pumps operate at low pressure. 5) Jetting with high pressure raises sound levels above ambient values

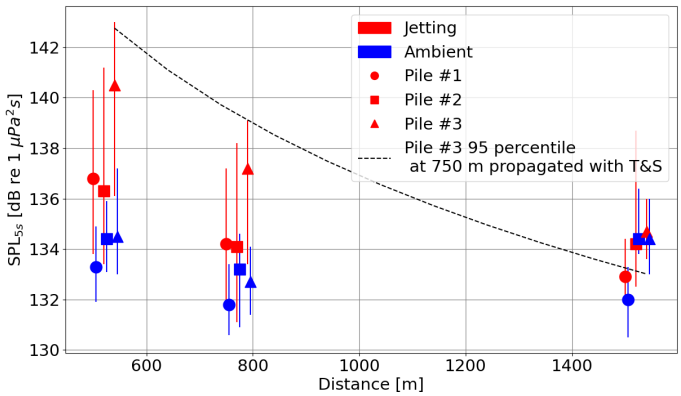


Figure 3: SPL *sound levels vs. distance from the installation location. Dots denote median (50 percentile) values, vertical lines denote spread between 95 and 5 percentile exceedance values. The dashed line depicts 95 percentile value from 750 m propagated with Thiele and Schellstede [12] transmission loss used in the environmental impact assessment section. Measurement points are artificially shifted from the actual measurement positions for better readability.*

regardless of the penetration depth. 6) Elevation of sound levels is likely primarily caused by jetting nozzles if the toe of the pile is in the water column. 7) Elevation of sound levels is likely primarily caused by pumps if the operating pressure is high, and toe of the pile is sufficiently embedded into sediment.

Aggregated overview of measured SPL across different distances from the foundation is presented in Figure 3. The plot depicts basic statistical metrics of SPL, namely 50 (median), 95 and 05 percentile exceedances for intervals when pumps were operating at low/high pressure (jetting) and for the time interval right before penetration when pumps were idle (ambient).

The following key findings applicable to all installed foundations can be derived from the measurements: 1) Sound levels from jetting are significantly elevated above ambient level at the distances of 500 m and 750 m for two (piles # 1 and # 3) out of three foundations. 2) Sound levels from jetting cannot be reliably distinguished from ambient levels at the distance of 1500 m for all 3 foundations. 3) Ambient sound levels from jetting trials are considered high (median values of 132-135 dB SPL) compared to other regions.

An example of measured 1/3 octave SPL spectra at distance of 750 m from the foundation # 3 for jetting and ambient periods is presented in Figure 4. This example depicts the case where signals from jetting could be separated from ambient levels across largest frequency range, namely from 200 Hz onwards. The threshold frequency for # 1 was 400 Hz and 630 for #2. Overall, the jetting emissions appear to be of a more broadband nature without any distinct harmonics compared to a typical impact piling spectrum which is also depicted in the figure. The piling spectrum was adopted from ITAP empirical measurements ([3], Figure 14 right plot). The spectrum was scaled to correspond to a broadband sound level for an equally sized foundation (9 meter diameter monopile) and hammer energy of 4000 kJ. The value was obtained by considering logarithmic trend of sound levels scaling with diameter as depicted in Figure 12 in [3]. Additionally, ITAP's prognosis submitted as part of the Revolution Wind Underwater Acoustic Analysis [15] estimates sound levels of 184 dB and 185 dB for 12 meter and 15 meter diameter monopile respectively for hammer energy of 4000 kJ. This translates to a broadband sound level of 182.7 dB for the 9 m monopile.

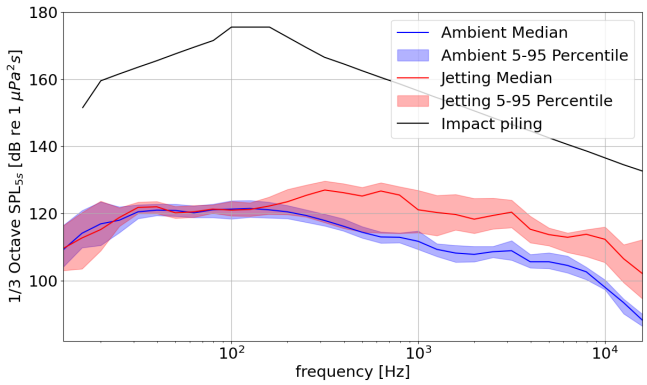


Figure 4: 1/3 Octave SPL *sound levels at distance of 750 m from foundation for ambient and jetting periods compared to a typical impact piling spectrum [3].*

5. ENVIRONMENTAL IMPACT ASSESSMENT

The above presented measurements were used to calculated distances to various environmental thresholds typically used for assessing impact of installation activities on marine fauna. The measurements from pile # 3 were used as a reference as these exhibited highest sound levels and cleanest separation from background noise at the distance of 500 and 750 m. The limited number of distances and their close proximity, however, still prohibits to calculate any reliable transmission loss. An alternative approach was, therefore, taken where measured 750 m sound levels were propagated using Thiele and Schellstede (T&S) [12] frequency dependent transmission loss model. This gives a theoretical estimate of what the true acoustic contribution from jetting is at further distances (e.g. beyond 1500 m where installation measurements were already masked by the background noise) and still agrees well with the near-source measurements as depicted in Figure 3. Same exercise was conducted for traditional impact piling using the previously discussed spectrum. A set of permanent threshold shift (PTS) criteria commonly used in the U.S. [13] was selected for this study. These criteria or its subset are often also adapted in other regulatory markets such as the United Kingdom, Denmark or Poland. All the criteria utilize cumulative sound exposure level (SEL_{cum}) as the acoustic metric. This often corresponds to total acoustic exposure produced from an installation of a single foundation (monopile) and the 750 m sound levels have to be adjusted accordingly. A correction factor of 38.6 dB which corresponds to 2 hours of continuous installation was added to 95 percentile of measured jetting sound levels (139.1 + 38.6 = 177.7 dB SEL_{cum}). Similarly, a correction factor of 37 dB was added to the impact piling sound level (182.7 + 37 = 219.7 dB SEL_{cum}). This correction was derived from publicly available underwater noise assessments and corresponds to 5000 hammer blows at 4000 kJ which is a representative level of total driving energy assumed in such modelling exercises [14, 15]. Results of calculated distances to threshold are presented in Table 1 for both low noise installation method and the impact piling with various levels of broadband attenuation. The PTS criteria distinguish between impulsive and non-impulsive sounds with different threshold values for both types. While the low-noise method produces non-impulsive noise, distances to both sets of thresholds are presented for this method to offer more direct comparison against the piling ranges. The low-noise consistently yields lower distances to thresholds, often one or two orders of magnitude lower than the traditional impact piling. The differences become smaller when a broadband reduction of 20 dB is considered for

Hearing group	Threshold impulsive (non-impulsive) [dB SELcum]	LN impulsive [m]	LN non-impulsive [m]	IP unabated [m]	IP 10 dB reduction [m]	IP 20 dB reduction [m]
LF	183 (197)	366	91	23972	9958	3279
HF	193 (201)	50	27	286	129	58
VHF	159 (181)	523	100	2664	1299	615
PW	183 (195)	233	81	5374	1771	578
OW	185 (199)	203	59	2769	1028	382

Table 1: Distances to various PTS acoustic threshold calculated for low-noise (LN) installation method and impact piling (IP)

impact piling. One exception is the range to the low-frequency (LF) hearing group that remains an order of magnitude higher even for the maximal considered sound reduction.

6. DISCUSSION

The low-noise jetting method offers an alternative to traditional pile driving, substantially reducing acoustic impacts in marine environments both in terms of absolute sound level values as well as producing less intrusive different type of noise, namely continuous and non-impulsive. The presented study discusses aggregated findings from the first set of real offshore application measurements. Additional measurements collected in quieter environments would provide further insights into true acoustic emissions at broader frequency range and at further distances.

REFERENCES

[1] A. N. Popper et al., “Sound Exposure Guidelines for Fishes and Sea Turtles: A Technical Report prepared by ANSI-Accredited Standards Committee S3/SC1 and registered with ANSI,” ASA S3/SC1.4 TR-2014, 2014. Available at: *https://doi.org/10.1007/978-3-319-06659-2*

[2] B. L. Southall et al., “Marine Mammal Noise Exposure Criteria: Initial Scientific Recommendations,” *Aquatic Mammals*, vol. 33, no. 4, pp. 411?522, 2007. Available at: *https://doi.org/10.1578/AM.33.4.2007.411*

[3] M. Bellmann et al., *Underwater noise during percussive pile driving: Influencing factors on pile-driving noise and technical possibilities to comply with noise mitigation values*. Final report for the Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU), commissioned and managed by the Federal Maritime and Hydrographic Agency (BSH). 2000 Available at: *https://www.itap.de/media/experience_report_underwater_era-report.pdf*

[4] A. Buljan, “Ecowende to Install Three Monopiles at Dutch Offshore Wind Farm in ’Silent Mode’,” *Offshore Wind*, April 12, 2024. Available at:

https://www.offshorewind.biz/2024/04/12/ecowende-to-install-three-monopiles-at-dutch-offshore-wind-farm-in-silent-mode/

[5] A. Buljan, "Van Oord to Use CAPE Holland Tool to Build 'Most Ecological Offshore Wind Farm Yet'," *Offshore Wind*, June 11, 2024. Available at: *https://www.offshorewind.biz/2024/06/11/van-oord-to-use-cape-holland-tool-to-build-most-ecological-offshore-wind-farm-yet/*

[6] A. Memija, "IQIP Launches New Piling Technique for Noise-Compliant Installation of Largest Monopiles," *Offshore Wind*, September 6, 2024. Available at: *https://www.offshorewind.biz/2024/09/06/iqip-launches-new-piling-technique-for-noise-compliant-installation-of-largest-monopiles/*

[7] International Organization for Standardization, *Underwater Acoustics ? Terminology*, ISO 18405:2017. Available at: *https://www.iso.org/standard/62406.html*

[8] Ørsted, *Ørsted successfully pilots new technology that further optimises offshore wind monopile installation*, Ørsted, 2024, *https://orsted.com/en/media/news/2024/07/orsted-successfully-pilots-new-technology-that-fur-13959650*.

[9] Ocean Instruments, "Soundtrap ST600 STD Long Term Recorder," 2025. [Online]. Available: *https://www.oceaninstruments.co.nz/product/soundtrap-st600-std-long-term-recorder/*. [Accessed: 20-May-2025].

[10] *DIN 4553:1986-11 ? Office furniture; concepts*, Deutsches Institut f□r Normung e.V. (DIN), November 1986. *https://www.din.de/en/getting-involved/standards-committees/office-furniture-standards-committee-4553*.

[11] International Organization for Standardization (ISO), *ISO 18406:2017 ? Underwater acoustics ? Measurement of radiated underwater sound from percussive pile driving*, 2017, *https://www.iso.org/standard/62407.html*.

[12] Thiele, J., & Schellstede, H. (1980). Standardwerte zur Ausbreitungsd□mpfung in der Nordsee. FWG-Bericht 1980-7, Forschungsanstalt der Bundeswehr f□r Wasserschall und Geophysik.

[13] National Marine Fisheries Service. (2024). *2024 Update to: Technical Guidance for Assessing the Effects of Anthropogenic Sound on Marine Mammal Hearing (Version 3.0): Underwater and In-Air Criteria for Onset of Auditory Injury and Temporary Threshold Shifts*. NOAA Technical Memorandum NMFS-OPR-71. Retrieved from *https://repository.library.noaa.gov/view/noaa/66184*

[14] National Marine Fisheries Service. (2022). *Sunrise Wind Acoustic Modeling Report*. NOAA Technical Report OPR-1. Retrieved from *https://media.fisheries.noaa.gov/2022-06/SunriseWind_2022App_AcousticModelingReport_OPR1_0.pdf*

[15] Bureau of Ocean Energy Management. (2020). *Underwater Acoustic Modeling Report for the [Project Name] Offshore Wind Energy Project*. Retrieved from *https://www.boem.gov/sites/default/files/documents/renewable-energy/state-activities/App-P3-Underwater-Acoustic-Modeling-Report.pdf*

THE IMPACT OF PILE DRIVING NOISE ON NORTH ATLANTIC RIGHT WHALE DETECTION USING COHERENTLY BEAMFORMED HYDROPHONE ARRAYS

V. Premus, P. Abbot, T. Abbot, E. Illich, J. Browning, V. Kmelnitsky, J. Freise, and A. Logan

Ocean Acoustical Services and Instrumentation Systems, Inc., A Wholly-Owned Subsidiary of ThayerMahan, Inc., 175 Cabot Street, Ste. 400, Lowell, MA, 01854

Corresponding author: V. Premus, Ocean Acoustic Services and Instrumentation Systems, Inc., 175 Cabot Street, Ste. 400, Lowell, MA, 01854, Phone: 781.862.8339, Fax: 781.862.0572, Email: premus@oasislex.com, vpremus@thayermahan.com

Abstract:
Measurements of radiated noise from offshore wind construction activity were collected on four bottom-mounted hydrophone arrays during May-August 2024 on the southern New England continental shelf. Calibrated source operations using a J-13 acoustic projector were conducted after completion of construction in September 2024 to quantify North Atlantic right whale (NARW) upcall detector-classifier performance and measure transmission loss at the site. Comparison of 32-channel array and omni-directional hydrophone detection performance reveals a nearly four-fold detection range advantage for the array in this 17logR TL environment. The calibrated source results were then used to validate the hypothesis that NARW upcall detection is a narrowband, rather than broadband, detection problem—the effective noise bandwidth (ENBW) employed in the sonar equation is linked to the mean instantaneous bandwidth of the upcall, empirically determined to be 12 Hz for a recognition differential of 5.5 dB. This finding has important implications for extrapolation of upcall detection performance models to new noise environments. Measurements of pile driving noise power spectra and array gain in third octave bands, as well as beam noise bearing dependence during pile driving, will be presented. These noise measurements, coupled with the narrowband ENBW definition, are then employed to predict upcall detection range for an array in the presence of pile driving.

Keywords: *Pile driving noise, offshore wind, hydrophone array, beamforming, passive sonar, performance metrics,*

INTRODUCTION

Analysis is presented of bottom-mounted hydrophone array measurements of pile driving radiated noise collected in May through August 2024 during Orsted's Revolution Wind construction project. The site and array laydown, located in US continental shelf waters off southeastern Massachusetts, is shown in Figure 1. This is an ecologically sensitive location and seasonal habitat for the endangered North Atlantic right whale. Acoustic data was collected on a ThayerMahan SeaPicket system[1] instrumented with a 32-channel hydrophone array and embedded digital signal processor. This paper decomposes radiated noise analysis of pile driving into "during strike" and "between strike" intervals and describes detection performance modelling for the NARW upcall that emphasizes the worst case "during strike" noise levels. It is well known that pile driving noise is quite loud. For example, even with bubble curtain noise mitigation,[2] median omni-directional hydrophone noise spectrum levels in the third octave band centered at 125 Hz measured more than 120 dB re 1μPa//Hz during the hammer strike at a range of 7 km from the noise source. The single hydrophone exposed to such a noise level is left completely masked and unable to detect all but the nearest vocalizing baleen whales. By comparison, median beam noise spectrum levels at the beamformer output for the same third octave band during the hammer strike were observed to be lower than 90 dB re 1μPa//Hz, illustrating that array gain, or spatial noise rejection, of more than 30 dB is achievable in such a highly anisotropic noise field. The analysis will show that coherent beamforming restores the desired upcall detection performance, i.e., detection ranges in excess of 10 km, for large angular sectors steered away from the line-of-sight (LOS) to the monopile in noise conditions that leave the hydrophone blanked. This is of particular importance to passive acoustic monitoring (PAM) during conditions of low visibility or at nighttime, when visual observation is typically impaired.

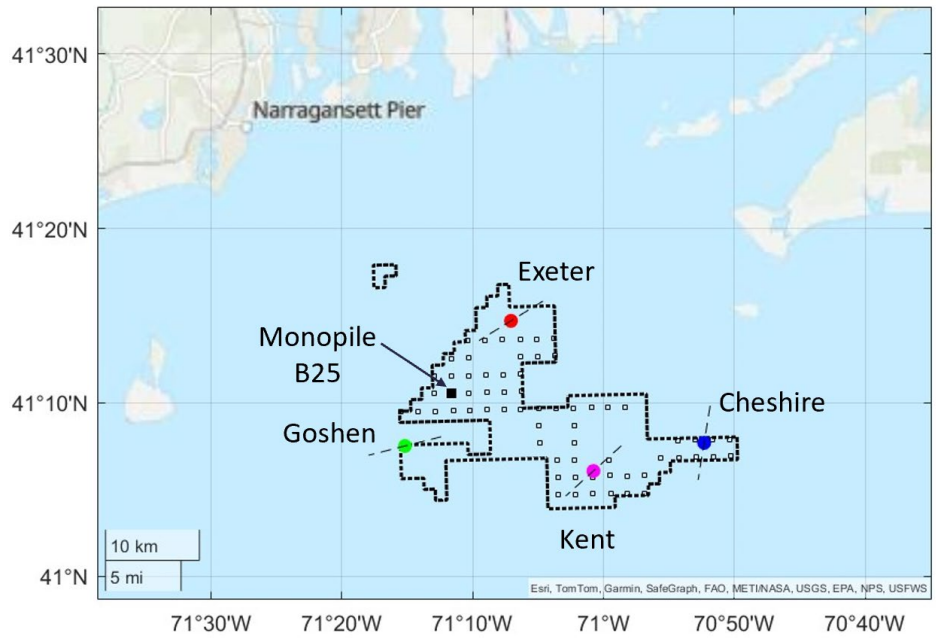


Figure 1 Revolution Wind lease area off southeastern Massachusetts. Lease area boundary denoted in black. Monopile locations denoted with black squares. Positions of the four bottom-mounted SeaPicket arrays deployed from May-August 2024 are indicated: Goshen (green), Exeter (red), Cheshire (blue), and Kent (magenta). For each array, orientation is denoted by dashed blue line.

METHOD

The passive acoustic sensing system used to support this work was the ThayerMahan SeaPicket system[1] comprised of a bottom-mounted hydrophone array tethered to a surface buoy instrumented with an embedded digital signal processor (DSP) and BGAN satellite communications modem. Figure 2(a) shows the SeaPicket surface expression after deployment. Fig. 2(b) depicts the sensor, a 32-channel, low-power, hydrophone array built by Raytheon Missiles and Defense (RMD) in Portsmouth, RI. Hydrophones are omni-directional and uniformly spaced at one half-wavelength for a design frequency of 625 Hz, or 1.2 m spacing, and 37.2 m total aperture length. Element data is digitized with 24-bit precision at a sample rate of 2.5 kHz.

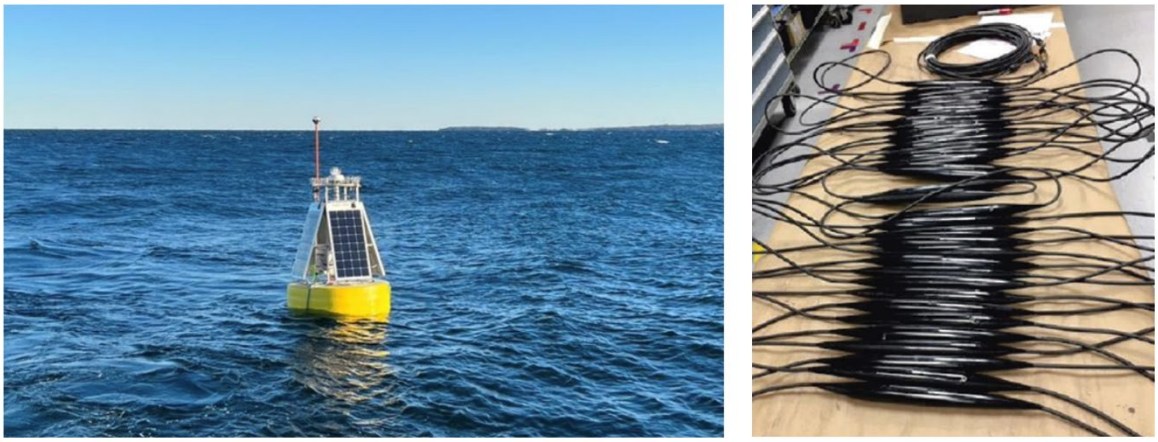


Figure 2 SeaPicket system: (a) Surface buoy with satellite comms and solar panels, and (b) 32-channel hydrophone array

The real-time passive sonar processing architecture consists of a fast Fourier transform (FFT) followed by a frequency domain conventional beamformer.[3] Beam response is normalized to ensure distortionless response to an incident plane wave signal. Element data are Hanning shaded for reduced sidelobe response. All hydrophones are calibrated, and hydrophone sensitivity and preamp gain are used to support real-time reporting of ambient noise levels and received levels of signals of interest in absolute levels. Following the beamformer, detection and classification are performed using a "binarized" implementation of a spectrogram correlator that compares a candidate spectrogram feature to members of a "kernel library" to produce a confidence or similarity score.[4] The NARW upcall classifier operating point was programmed for a false alarm rate (FAR) of 1 per hour. The binarized spectrogram correlator runs in real-time, processing 32 beams concurrently at a frequency resolution of 3.9 Hz and an update rate of 64 milliseconds. For more detail on the sensor and processing, the reader is referred to [3].

RESULTS

To quantify the effect of pile driving noise on the performance of the North Atlantic right whale detector-classifier using a passive sonar equation reconciliation, it is best to process the array data through the lens of the real-time, spectrogram-correlator algorithm. This

means using a FFT length, percentage overlap, and update rate that support a temporal granularity capable of resolving short duration (~100 ms) broadband transient signals. Thus, the DSP processing parameters used for this analysis are: df = 3.938 Hz (window corrected bin width of 5.89 Hz), OL = 75%, and dT = .0634 s. Representative noise power spectrum measurements for one of the arrays (Goshen) deployed at a range of 7 km to monopile B25 installed on 27 July 2024 are shown in Figure 3. The results are reported for the median frequency bin and median beam in each of 12 third-octave bands (although not in third octave band levels), partitioned into "during strike" and "between strike" intervals. To facilitate comparison to Wenz historical observations, the median noise levels are corrected to noise spectrum level (i.e., units of dB re 1μPa^2//Hz).

The data in Fig. 3 confirms what is already well known, radiated noise from the hammer strike is very loud. Even with the bubble curtain noise mitigation in place, at a range of 7 km—roughly 65 dB transmission loss (TL) in this 17logR environment—peak median noise spectrum levels approach 130 dB re 1μPa//Hz at 100 Hz, the bottom of the NARW upcall support band. Strike noise is also seen to exhibit a strong frequency dependence, with levels decaying rapidly above 100 Hz—this has important implications for PAM as much of the frequency support for the North Atlantic right whale upcall lies in the band 100-250 Hz.

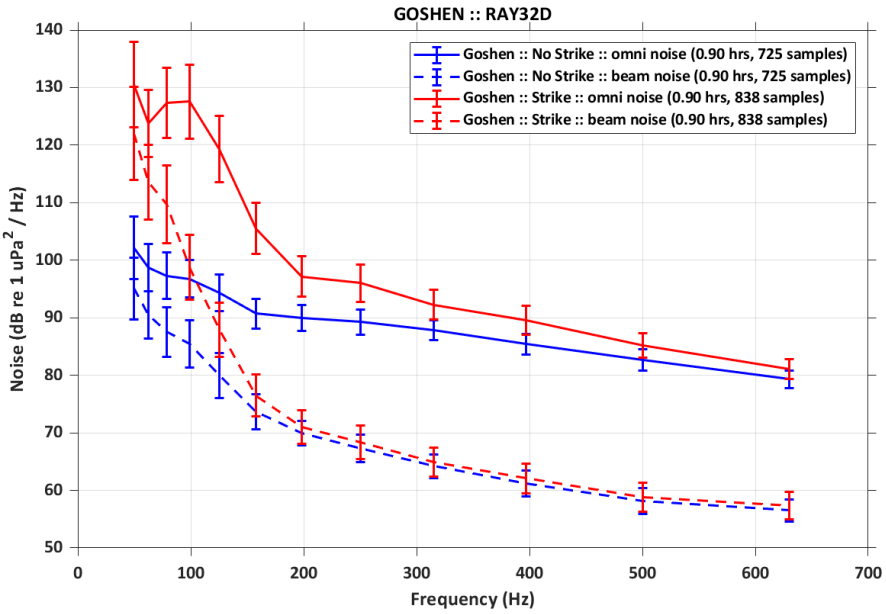


Figure 3 Noise power spectrum observed at array Goshen on 27 July 2024. Goshen is 7 km distance from the monopile, B25. Results are reported for the median frequency bin in third octave band, and median beam for the case of the beamformer output. Observations are partitioned into "strike" and "between strike" intervals. Sample size N ≅ 800.

At the center of the upcall support band, at a frequency of 150 Hz, the median omni noise spectrum level measures 105-115 dB re 1μPa//Hz, compared to a median beam noise spectrum level of 76-84 dB re 1μPa//Hz. It will be shown below that a beam noise spectrum level in the mid-80s dB//Hz, after the appropriate correction to in-band noise level, is enough to support a detection range more than 10 km in this 17logR TL environment.

Figure 4 shows the corresponding measured array gain (AG) on all four arrays during the same monopile event, B25, on 27 July 2024. Of particular note is the array gain during the hammer strike, which measures more than 25 dB over most of the array operating band at

Goshen, which is the closest of all four arrays to the monopile. In fact, measured array gain on all four arrays is observed to be well in excess of directivity index (gray line in Fig. 4), or $10\log(2L/\lambda)$ where L is the aperture length and λ is acoustic wavelength. This attests to the degree of anisotropy in this construction noise environment.

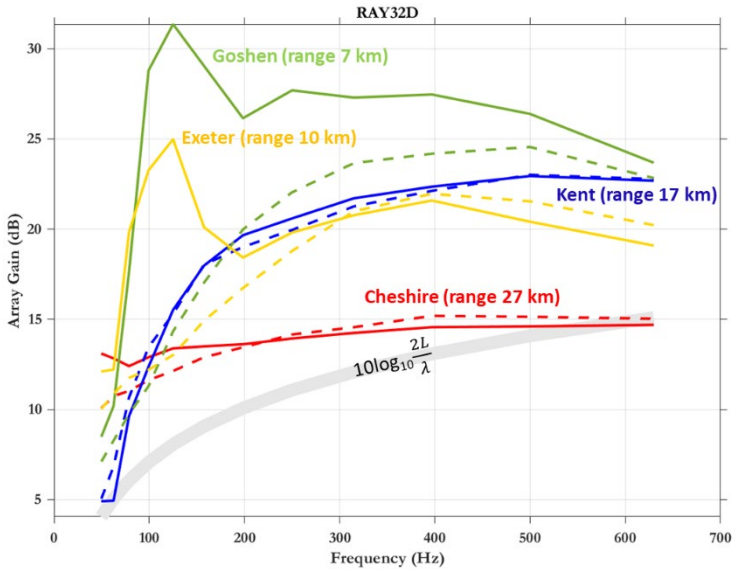


Figure 4 Measured array gain observed for monopile B25 on 27 July 2024 for all four arrays depicted in Fig. 1. Ranges of the pile to Goshen (green), Exeter (yellow), Kent (blue), and Cheshire (red), are 7 km, 10 km, 17 km, and 27 km, respectively.

To examine the detection performance impact of beamformer response to pile driving noise, it is necessary to first address the issue of in-band noise definition and the question of whether NARW upcall detection, and perhaps baleen whale detection generally, is a broadband or narrowband detection problem. On September 4, 2024, calibrated source operations using a J-13 acoustic projector were conducted at the Revolution Wind site after the completion of construction operations to quantify right whale upcall detector-classifier performance and measure *in-situ* transmission loss. Source operations were conducted at range offsets of 1, 2, 4, 6, and 8 NM at a source level of 160 dBrms re 1μPa @ 1 m.

Source testing revealed two important findings. First, comparison of the 32-channel array and omni-directional hydrophone detection performance revealed a 3.6x detection range advantage for the array, consistent with a roughly 10 dB of array gain observed in this 17logR TL environment. Second, the results showed that accurate interpretation of detector performance must occur at the instantaneous Fast Fourier Transform (FFT) frame level, recognizing that the effective noise bandwidth (ENBW) driving the detection decision is defined by the instantaneous upcall bandwidth.[5] Analysis of over 1500 upcall detection events conducted that day, as well as 4,090 exemplars from the DCLDE 2013 St. Andrews database,[6] showed that the instantaneous bandwidth of the upcall can vary from 8-30 Hz depending on SNR. The instantaneous bandwidth corresponding to a recognition differential of 5.5 dB (see discussion below) was determined to be approximately 12 Hz. Passive sonar equation analysis was used to show that the narrowband noise model based on an ENBW of 12 Hz accurately predicted the measured detection performance of the spectrogram correlator, while the broadband noise model (40-400 Hz) overestimated the applicable in-band noise level and thus severely underpredicted performance.[5] This finding has important implications for extrapolation of upcall detection performance models to new noise environments such as pile driving and is employed in the noise analysis that follows.

Measurements of in-band (12 Hz) beam noise during monopile installation were made on all four arrays throughout the 4-month construction period. Figure 5 shows one such example for monopile B39 which was installed on 1 July 2024—this monopile was chosen due to its central location. The lefthand panel depicts a representative snapshot of in-band beam noise contours for Goshen (green), Exeter (red), Kent (magenta), and Cheshire (blue) during a hammer strike. The noise level reticule dynamic range is 75-115 dB re μPa in 10 dB steps.

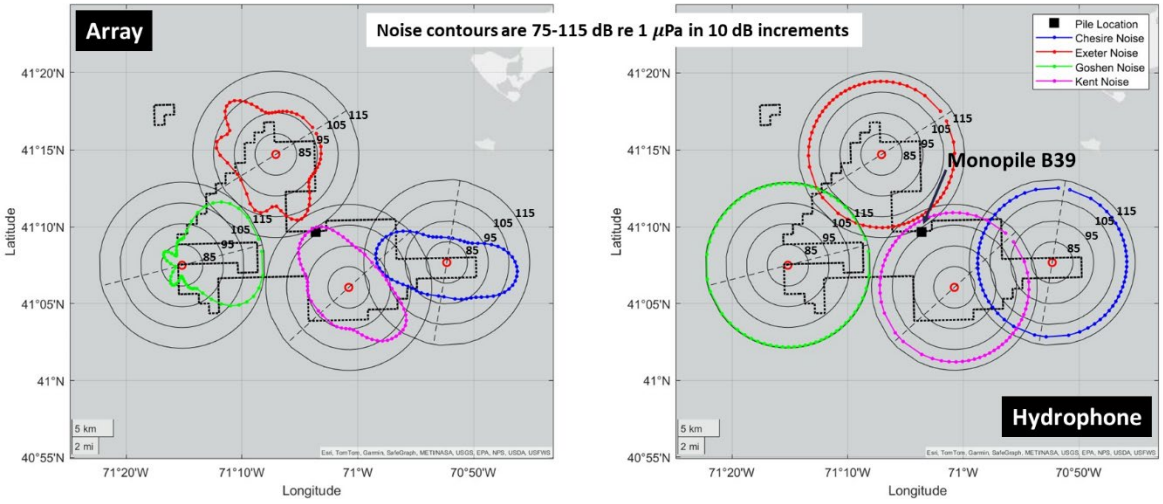


Figure 5 Measured in-band (12 Hz) beam and omni noise distributions during the hammer strike for monopile event B39 at 1500 GMT on 1 July 2024.

In-band noise level was calculated using an incoherent average of noise level in a 5.9 Hz window-corrected frequency bin over the support band 98-248 Hz. Empirical analysis has shown this to be the band for which the spectrogram correlator derives most of its positive performance. The incoherent average noise level in a 5.9 Hz band was then corrected for the 12 Hz ENBW of the spectrogram correlator detection decision. The righthand panel of Fig. 5 shows the corresponding omni-directional hydrophone noise contours taken from channel 16 of the array. Notice that the peak beam noise response for each array is equal to that of the omni-directional hydrophone and is closely aligned with the line-of-sight (LOS) to the monopile. For directions away from the monopile LOS however, the noise level is significantly reduced. This is the performance benefit of array gain. More precisely, array gain is the ratio of the area subtended by the hydrophone noise response to that of the beam noise response. The more anisotropic the noise distribution, the greater the benefit of the coherently beamformed array for acoustic detection.

Following Cox,[6] the passive sonar equation, given in (1), is typically expressed in terms of the figure of merit (FOM), which is defined as the maximum transmission loss that can be tolerated by a system and still maintain the desired receiver operating point:

$$FOM = TL = SL - (NL - AG) - NRD \quad (1)$$

where SL is the source level, NL is the in-band noise level at the hydrophone output, AG is the array gain, and NRD is the recognition differential, respectively. As reported previously,[3] the NRD of the ThayerMahan spectrogram correlator implementation has been empirically determined to be 5.5-6 dB for 50% Pd at a false alarm rate of 1 per hour.

Substituting the measured noise power spectrum results of Fig. 3 into eq. (1), it is straightforward to show that array gain against the hammer strike can yield a detection range of 10 km in this 17logR TL spreading loss environment. At an upcall source level of 170 dBrms re 1μPa @ 1 m, a coherently beamformed hydrophone array needs to deliver a beam noise spectrum level in the mid-80s dB re 1μPa//Hz, as shown in Fig. 3, or an in-band beam noise level of mid-90s re 1μPa in a 12 Hz band, in order to realize a detection range in excess of 10 km (e.g., 170 – 96 – 6 = 68 and $10^{68/17}$ = 10 km).

Further, used in conjunction with a transmission loss model, bearing-dependent measurements of beam noise can be substituted for the (*NL-AG*) term in (1) to extrapolate NARW upcall detection performance to the case of a realistic pile driving noise environment. Figure 6 depicts modelled detection performance contours for monopile B39 on 1 July 2024 for the four arrays corresponding to the measured beam noise contours shown in Fig. 5. This is not a uniformly worst-case hammer strike FFT frame, but a representative one. Here the reticule dynamic range is 0-10 km in 2.5 km steps.

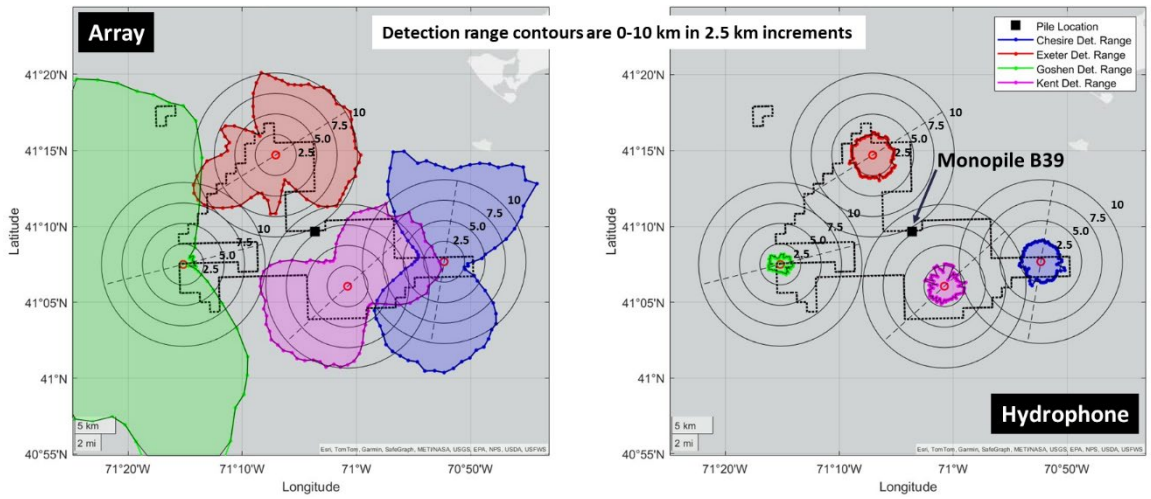


Figure 6 Modelled detection contours for a representative FFT frame corresponding to beam and omni-directional noise contours of Fig. 5 during the hammer strike for monopile B39 at 150000 GMT on 1 July 2024. Minor degree of bearing dependence in the hydrophone contours is due to bearing dependence of the modelled transmission loss.

Apart from the dramatic difference in total area coverage for the array relative to the single hydrophone, a couple of observations stand out. First, the nulls in the detection response at each array are aligned with the LOS to the monopile. The width of the null for the arrays at beam aspect (broadside) to the monopile, Kent, Exeter, and Cheshire, is much narrower than that of the array at endfire aspect, e.g., Goshen. This observation prompts the question as to the optimal orientation of the array relative to the pile driving—is it better to orient the monopile to endfire (as in the case of Goshen, green) to protect the most sensitive detection bearings at beam aspect, or to broadside (as in the case of Kent, magenta) to minimize the bearing extent of the impact. This question is still under consideration. Array orientation is primarily driven by wind conditions and currents at the time of deployment. Also, movement of the arrays as construction activity migrates from pile to pile incurs additional risk, so optimization on an individual monopile basis is challenging. Second, while the arrays closest to the monopile, in this case Exeter (red) and Kent (magenta) exhibit the most pronounced impact to performance, all four arrays show predicted detection coverage to 10

km or more for one-half to two-thirds of beamspace. The single hydrophone detection coverage during the strike is, of course, uniformly degraded at all bearings and struggles to reach 2.5 km range. In the case of Goshen at this time instant, where the in-band noise level approaches 115 dB re 1μPa, the detection range is projected to be less than 1 km.

CONCLUSION

This work reviewed analysis of pile driving radiated noise collected on four bottom-mounted hydrophone arrays during Orsted's Revolution Wind monopile installation from May-August 2024. The measurements show that median omni-directional hydrophone noise spectrum levels during the hammer strike can exceed 120 dB re 1μPa//Hz at a range of 7 km from the noise source. However, analysis of beam noise results demonstrated spatial noise rejection, or array gain, of up to 30 dB, during intervals corresponding to the hammer strike. The capacity of arrays to spatially filter such a strongly anisotropic noise distribution means that it is possible to deliver detection ranges more than 10 km *even during the hammer strike* for bearing sectors steered away from monopile LOS, while the single hydrophone is completely masked. Next steps are to conduct calibrated, ground truthed source operations during construction operations to validate the detection performance projections in the presence of pile driving presented herein.

ACKNOWLEDGEMENTS

This work was supported under IR&D investment from ThayerMahan Inc., at the direction of VADM (Ret.) M. Connor, CEO, and R. Hine, ThayerMahan co-founder and President, ThayerMahan Offshore.

REFERENCES

[1] M. Connor and R. Hine, "Continuous unmanned airborne and underwater monitoring platform," U.S. Patent 11,105,662 B2 (August 31, 2021).
[2] B. Würsig, C.R. Greene, T.A. Jefferson, "Development of an air bubble curtain to reduce underwater noise of percussive piling," *Mar. Env. Res.*, **49**, 79-93 (2000)
[3] V. Premus, P. Abbot, V. Kmelnitsky, C. Gedney, and T. Abbot, "A wave glider-based, towed hydrophone array system for autonomous, real-time, passive acoustic marine mammal monitoring," *J. Acoust. Soc. Am.*, **152**, 1814-1828 (2022).
[4] T. Abbot, V. Premus, and P. Abbot, "A real-time method for autonomous passive acoustic detection-classification of humpback whales," *J. Acoust. Soc. Am*., **127**, 2894–2903 (2010).
[5] V. Premus, P. Abbot, E. Illich, T. Abbot, J. Browning, V. Kmelnitsky, "North Atlantic right whale detection range performance quantification on a bottom-mounted linear hydrophone array using a calibrated acoustic source," *J. Acoust. Soc. Am.*, (2025) (submitted).
[6] D. Gillespie, *DCLDE 2013 workshop dataset*, University of St Andrews Research Portal (2019)
[7] A. W. Cox, *Sonar and Underwater Sound, 2nd ed.*, Lexington Books, Lexington, MA (1982).

Structured Session
15. Seabed Acoustics

Organised by Megan Ballard, Kevin Lee and Nick Chotiros

Scattering and reflection from soft gassy sediments: Geoacoustic modeling approach

Anatoliy Ivakin[1]

[1]*Applied Physics Laboratory, University of Washington, USA*

Contact author: Anatoliy Ivakin, APL-UW, 1013 NE 40th Street, Seattle Washington 98105, USA; E-mail: aniv@uw.edu; Tel: 1-206-616-4808

Abstract: *Presence of even a small volume fraction of gas bubbles can significantly change acoustic properties of underwater sediments, particularly, their effective sound speed and attenuation. In this paper, a geoacoustic modeling approach which, after corrections in input parameters for gas content, is used for estimating the reflection loss and scattering strength, major characteristics describing geoacoustic bottom interactions. One of these corrections is discussed in more detail and shown to be significant in a strong volume scattering regime, which seems a reasonable assumption for gassy sediments with bubbles and voids or with stronger fluctuations of sound speed caused by spatially varying concentration of bubbles. Results of numerical simulations of angular dependences of reflection loss and backscattering strength for soft gassy sediments are presented.*

Keywords: *Reflection and scattering coefficients, methane bubbles, rough interfaces, volume heterogeneity, intrinsic and scattering attenuation.*

1. INTRODUCTION

There is extensive literature on geoacoustic modeling of underwater sediments aimed at predicting reflection loss and scattering strength, major characteristics describing sound-bottom interactions, see e.g. [1,2] and references therein. Soft water saturated sediments in these models are usually treated as effective fluid media whose input parameters are density, sound speed and attenuation. For instance, GABIM [2], Geoacoustic Bottom Interaction Model, presents results for canonical classes of granular sediment, each with a representative set of these parameters defined by their empirical relationships with the mean grain size [1].

One of the limitations of such approach is that this modeling relies on accuracy of the mentioned empirical relationships, which however have been well established only for gasless sediments [1]. This limitation can be essential as it is known that even a small volume fraction of gas bubbles may significantly change sound speed and attenuation in the sediment. There are various models to quantify these changes and provide theoretical relationships of sound speed and attenuation with gas content, see e.g. [1,3-5]. In this paper, Wood's model is used simply to provide a representative example of such relationships.

Various approaches to consider reflection loss and scattering strength of gassy sediments were discussed in literature [5-12]. For instance, the effect of methane bubbles on bottom reflection was observed in Lake Kinneret and its treatment was given using an effective model of gassy sediments [5-8]. Effects of gas bubbles on scattering strength of soft sediments was considered in [1,2,9-12]. In this paper, we suggest that geoacoustic modeling approach developed for gasless environments, such as that used in GABIM, is applicable to estimating the reflection loss and scattering strength of gassy sediments as well once the values of input parameters are corrected to account for these changes. One of these corrections is discussed in more detail and shown to be significant in a strong volume scattering regime, which can be a reasonable assumption for sediments with such strong discrete scatterers as gas bubbles and voids or with stronger fluctuations of sound speed cased by continuous spatial variations of bubble concentration.

2. GEOACOUSTIC MODEL INPUT PARAMETERS

GABIM [2], Geoacoustic Bottom Interaction Model, presents comparison of results for 10 "canonical" sediment classes, defined by their different mean grain size, with a set of geoacoustic input parameters whose values are determined empirically as representative for each class. In this paper, we consider only soft sediments with inputs for two classes, Fine Sand and Clay, given in Table 1, defined in [2] and used here in calculations.

Sediment type	a_ρ Density Ratio	a_c Sound Speed Ratio	δ Loss Tangent	q Volume Scattering Ratio	Roughness Spectral Exponent	Roughness Spectral Strengths
Fine Sand	1.936	1.1536	0.01669	0.002	3.25	0.000016
Clay	1.410	0.9892	0.00149	0.001	3.25	0.000016

Table 1. Representative canonical parameters taken from [2] and used as inputs in calculations for gasless sediments.

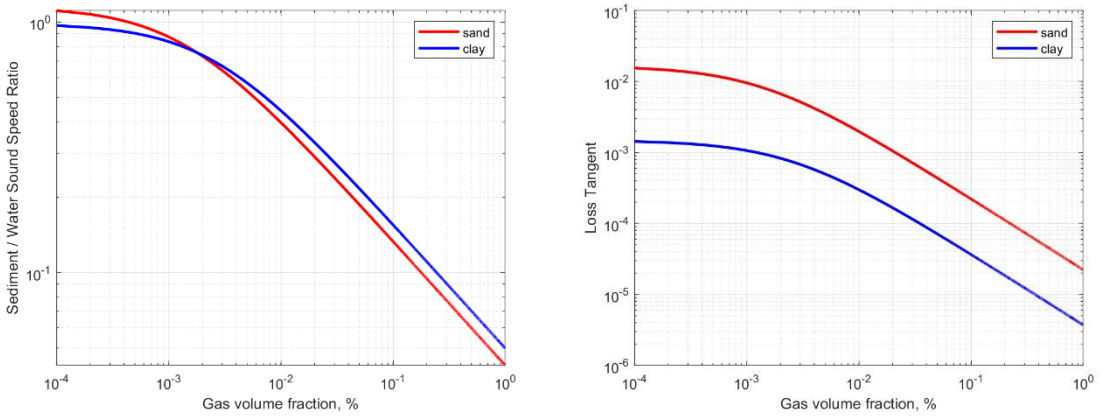
The inputs include sediment/water density and sound speed ratios, a_ρ and a_c, and loss tangent (attenuation parameter) δ. This is sufficient for predictions of the reflection coefficient, V, and bottom reflection loss, $20\log|V|$, for soft sediments. The model also includes 3 additional parameters, volume scattering ratio q (a ratio of volume scattering and attenuation coefficients), and roughness spectral strength and exponent, which are needed for predictions of the so-called equivalent surface scattering coefficient,

$$M_S = M_{Sr} + M_{Sv} \tag{1}$$

where M_{Sr} and M_{Sv} are its roughness and volume scattering components. Bottom scattering strength is defined as $10\log M_S$.

It is known that even a small volume fraction of gas bubbles may significantly change sound speed and attenuation, and therefore at least two of these inputs, a_c and δ, should be corrected in the case of gassy sediments. There are various theoretical relationships of effective sound speed and attenuation with gas content and bubble size (see, e.g., [1, 3-5]). Here, in this paper, for a demonstration of the geoacoustic approach, a Wood's model is chosen. The model accounts for additivity of density and compressibility in the sediment that includes two components, gas bubbles inclusions and the host gasless surrounding sediment, to derive the effective sound speed of the mixture. This requires adding only one parameter, volume fraction of gas bubbles, α, which may provide significant corrections for the sediment sound speed and attenuation but practically does not affect density (at small α typical for gassy sediments). The model is known to be sufficiently accurate at low frequencies or small enough bubbles (microbubbles) where resonance effects controlled by bubble size are negligible. Including additional parameters, such as the bubble size and shape, is similar, and, although allows accurate accounting for resonance and other effects, does not change the approach in general.

In Figure 1, Wood's model results are given for effective sound speed ratio and loss tangent of sand and clay sediments with methane bubbles as functions of gas volume fraction assuming that values of methane density and sound speed are 0.717 g/l and 330 m/s respectively. This provides Wood's model-based values to input parameters, summarized in Table 2, for numerical simulations presented in following sections of this paper.



(a) (b)

Fig.1: Sound speed ratio (a) and attenuation parameter (b) for gassy sand and clay sediments as functions of gas volume fraction according to Wood's model.

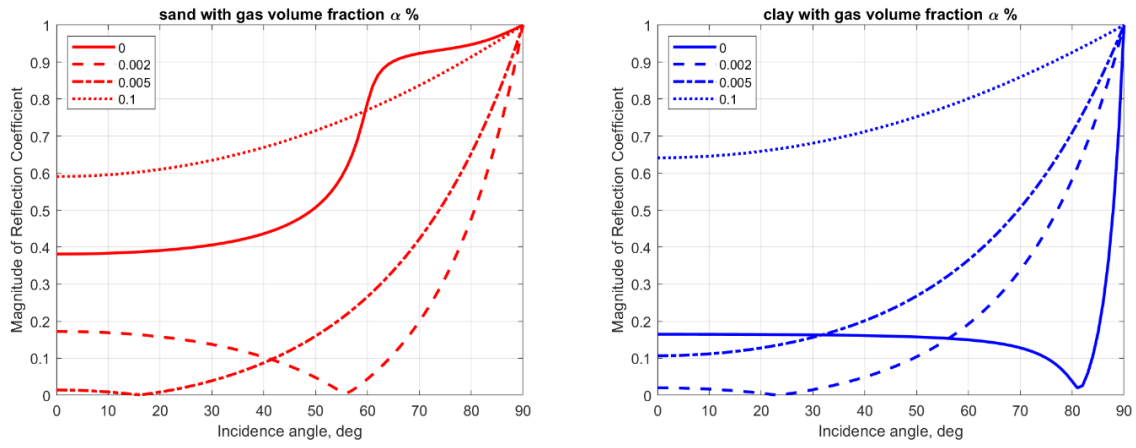
α, %	0	0.002	0.005	0.1
a_c , Fine Sand	1.1536	0.7319	0.5315	0.1331
a_c , Clay	0.9892	0.7383	0.5723	0.1550
δ , Fine Sand	0.01669	0.0067	0.0035	0.00022
δ , Clay	0.00149	0.00083	0.00050	0.000037

Table 2. Wood's model-based values for sound speed ratio, a_c , and attenuation parameter, δ, given different gas volume fractions α (in %) used in calculations for fine sand and clay sediments.

3. BOTTOM REFLECTION LOSS AND ROUGHNESS SCATTERING STRENGTH

The following calculations are aimed at illustrating quantitatively significant effects of gas bubbles (particularly microbubbles) in the sediment, which particularly change contrasts of parameters (effective density, compressibility, sound speed, attenuation, and impedance) at interfaces, affecting bottom reflection and rough interface scattering.

Figure 2 shows the magnitude of reflection coefficient vs incidence angle for gassy sand (a), and gassy clay (b) sediments using inputs given in Tables 1 and 2 at different gas volume fractions according to Wood's model. The deep minima of the magnitude of reflection coefficient are caused by reduction of sound speed and corresponds to angles at which the water-sediment interface is nearly transparent, that is it has a minimal contrast of angular dependent impedances (zero if attenuation is ignored).

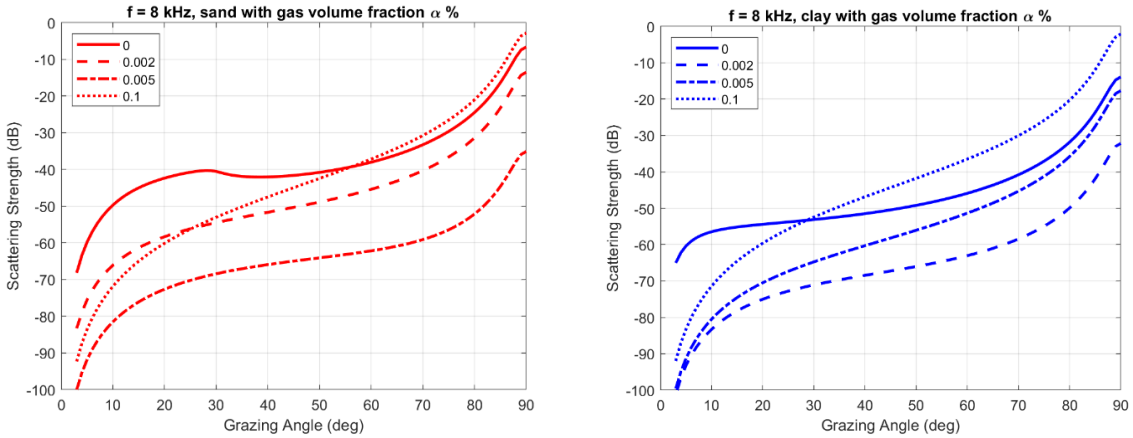


(a) (b)

Fig.2: Reflection coefficient vs incidence angle for gassy sand (a), and gassy clay (b) sediments at different gas volume fractions.

In [9], numerical examples are given for angular dependences of rough bottom scattering strength to show the effect of a reduced sound speed typical for a small presence of gas in the sediment. Here, in this paper, we apply the same approach and same expressions for the

scattering coefficient M_{Sr} as in [1,9] but use different inputs, gasless sediment canonical parameters and Wood's model-based parameter corrections given in Tables 1 and 2. As in [9], we assume that roughness parameters are the same as in the gasless case. Results are given in Figure 3 which shows the roughness backscattering strength as a function of grazing angle.



(a) (b)

Fig.3: Roughness backscattering strength of gassy sand (a) and clay (b).

4. VOLUME SCATTERING MODEL

Presence of gas bubbles and voids may significantly affect both mechanisms of volume scattering in the sediment, its continuous and discrete heterogeneity [10]. For discrete scattering, this particularly results from changing the contrast between parameters of the bubble and surrounding effective host medium, and, for continuous scattering, this is due to enhancing fluctuations of sediment effective parameters (caused by spatial variations of bubble volume concentration).

Consider the volume component of equivalent surface scattering coefficient, see Eq.(1), which for backward directions in the case of the sediment half-space, using expressions given in [1,2,9], can be presented as follows:

$$M_{Sv} = \frac{|W|^4 \mathrm{Real}(\gamma)}{2a_\rho^2}\, q \tag{2}$$

$$\gamma = \sqrt{(1+i\delta)^2 - a_c^2 \sin^2\theta} \tag{3}$$

$$W = 1 + V = \frac{2a_\rho a_c \cos\theta}{a_\rho a_c \cos\theta + \gamma} \tag{4}$$

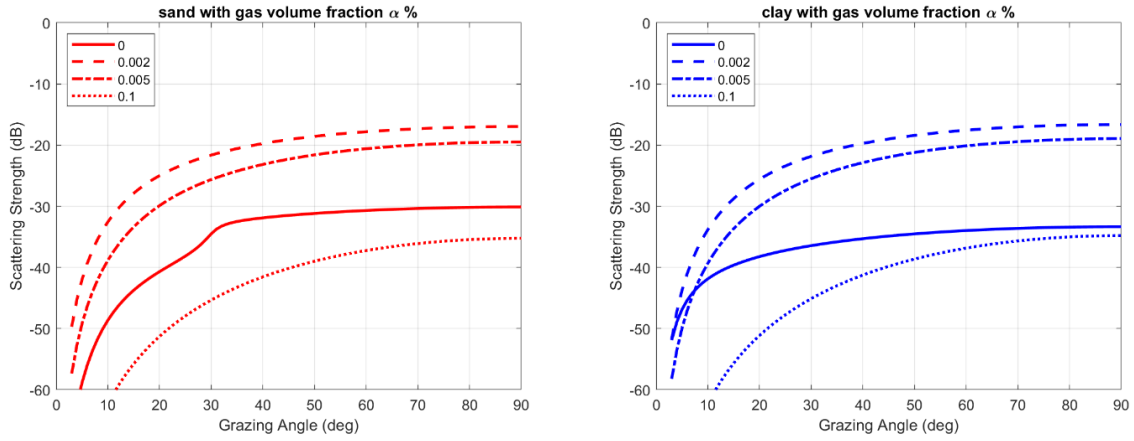
where $V(\theta)$ and $W(\theta)$ are the reflection and transmission coefficients at the water-sediment interface as functions of incidence angle θ, and other parameters are defined earlier and given in Tables 1 and 2. In particular, $q = m_v/\beta$ is the ratio of volume backscattering coefficient per unit volume, m_v, and the attenuation coefficient in the sediment, β. The attenuation parameter includes two components, intrinsic loss and total volume scattering loss (integrated over all directions of scattering)

$$\beta = \beta_0 + \beta_v\,, \qquad \beta_v = \oint m_v d\Omega \tag{5}$$

The result is significantly dependent on the ratio of these two components, μ, and the scattering indicatrix, ψ,

$$q = \frac{\mu}{1+\mu}\psi, \quad \mu = \beta_v/\beta_0\,, \quad \psi = m_v/\beta_v \tag{6}$$

Strong scattering regime corresponds to the case where scattering loss mechanism is dominating compared to intrinsic loss, $\mu \gg 1$ and $q = \psi$. Note, that according to Wood's model, intrinsic loss of the gassy sediments is significantly lower than for gasless ones, see Fig.1 and Table 2. For isotropic scattering, $\psi = (4\pi)^{-1}$, and therefore, in the case of strong isotropic scattering in the sediment, equation (2) can be used only with $q = (4\pi)^{-1}$. Results are shown in Figure 4.



(a) (b)

Fig.4: Volume backscattering strength of gassy sand (a) and clay (b).

Note that in Table 1 parameter q is fixed and very small, $q \ll (4\pi)^{-1}$. According to first equation in (6), this is possible in two different cases. First, this may correspond to the case where scattering is isotropic but very weak with $\mu = (4\pi)q$, i.e. is μ = 0.025 and 0.0125 for

sand and clay respectively. Otherwise, or in the second case of strong scattering ($\mu \gg 1$), one should assume $\psi \ll 1$, that is scattering is very anisotropic, which may be not applicable to small discrete inclusions, like microbubbles in gassy sediments. The case of anisotropic scattering may need an additional consideration due to possible effects of non-spherical shape of large enough bubbles which usually present in soft sediments [5].

5. ACKNOWLEDGEMENTS

This work was supported by the United States – Israel Binational Science Foundation (BSF), grant 2018150. The results were presented at UACE2025, 15-20 June 2025, Halkidiki, Macedonia, Greece.

REFERENCES

[1] **D.R. Jackson and M. Richardson**, *High-Frequency Seafloor Acoustics*, Springer Science, NY, (2007).

[2] **D.R. Jackson, R.I. Odom, M.L. Boyd, and A.N. Ivakin**, A Geoacoustic Bottom Interaction Model (GABIM), *IEEE J. Ocean. Eng.*, 35(3), 603-617, (2010).

[3] **A. L. Anderson and L. D. Hampton**, Acoustics of gas-bearing sediments. I. Background, *J. Acoust. Soc. Am.* 67(6), 1865–1889 (1980).

[4] **A.L. Anderson and L. D. Hampton**, Acoustics of gas-bearing sediment. II. Measurements and models, *J. Acoust. Soc. Am.* 67(6), 1890–1903 (1980).

[5] **R. Katsman, A. Lunkov, E. Uzhansky, B. Katsnelson**. Effective model of gassy sediments and acoustical approach for its verification, *Proceedings of Meetings on Acoustics*, Vol. 44, 005001 (2021).

[6] **E. Uzhansky, B. Katsnelson, A. Lunkov, R. Katsman, A. Ivakin**, Estimating bubble size in Lake Kinneret through reflection of wideband sound signals from gas rich sediment environment, *Proceedings of Meetings on Acoustics*, Vol. 50, 005001 (2023).

[7] **E. Uzhansky, R. Katsman, A. Lunkov, B. Katsnelson**, Integrated methodology for gas content assessment and prediction in shallow muddy lake sediments: acoustic mapping and correlation analysis, *MethodsX* , 13, 102799, (2024).

[8] **B. Katsnelson, R. Katsman, A. Lunkov, and I. Ostrovsky**. Acoustical methodology for determination of gas content in aquatic sediments, with application to Lake Kinneret, Israel, as a case study. *Limnology and Oceanography: Methods*, 15(6), 531-541, (2017).

[9] **A.N. Ivakin**, Models of scattering for remote acoustic sensing of the seafloor, in *Acoustical Oceanography,* T.G. Leighton, G.J. Heald, H.D. Griffiths, G. Griffiths (eds.), *Proceedings of the Institute of Acoustics*, v.23: Part 2, pp. 268-275, (2001).

[10] **A.N. Ivakin**, Modeling of acoustic backscatter for remote detection and characterization of oil and gas in marine environment, in, *Proceedings of the Institute of Acoustics*, v.34: Part 3, pp.1202-1209, (2012).

[11] **L. Fonseca, L. Mayer, D. Orange, and N. Driscoll**, The high-frequency backscattering angular response of gassy sediments: Model/data comparison from the Eel River Margin, California, *J. Acoust. Soc. Am.* 111(6), 2621–2631 (2002).

[12] **F. Boyle and N. Chotiros**, A model for high-frequency acoustic backscatter from gas bubbles in sandy sediments at shallow grazing angles, *J. Acoust. Soc. Am.* 98(6), 531-541 (2001).

Structured Session
16. Sonar performance modeling and verification: Active and passive sonar

Organised by Mathieu Colin and Kevin Heaney

The future evolution of mid-frequency active sonar performance in a changing ocean

Xavier Cristol[1], Bruno Chalindar[1], Nicolas Fisichella[1], and Herald Rabeson[1]

[1]Thales Defence Mission Systems, Sophia-Antipolis, France

Xavier Cristol, Thales DMS, 525 route des Dolines, 06903 Sophia-Antipolis, France
xavier.cristol@fr.thalesgroup.com

Abstract: *Climate change and global political-economic developments affect the future performance of underwater acoustic instruments through several interrelated channels: changes in sound velocity profile and volume attenuation in a physically evolving ocean, combined with a likely increase in ambient noise from shipping traffic and route changes. This multifactorial problem has already been the subject of several published analyses, with controversial conclusions. Climate change is primarily driven by changing atmospheric CO2 concentration; for this parameter, we have adopted the IPCC's Shared Socio-Economic Trajectory SSP5-8.5, which is the most pessimistic, but apparently the most probable curve. The variations of physical ocean parameters over the 21st century have been the subject of numerous models, the results of which are available online: 120 models were compared in the Coupled Model Intercomparison Project - Phase 6 (CMIP6). We have chosen the predictions of the ocean climate model CNRM-ESM2, developed by the CNRM (Toulouse, France), which not only has a good reputation, but also lies in the middle of the very different trends proposed by the 120 models. In this article, we present preliminary global conclusions based on a planetary-scale analysis of the evolution of the various terms of the Sonar Equation, along with the tactical behavior of sonar use. We restrict ourselves to a generic fictitious active VDS (Variable Depth Sonar) sonar operating in the mid-frequency range (at 2kHz). Global 2°latitude x2° longitude maps of maximum detection ranges are presented. Apart from some local variations, surprisingly simple trends emerge for the period 2019-2100: in both the northern and southern hemispheres, detection ranges remain stable in winter and tend to increase with time in summer, as a result of the contradictory interplay between changes in the shape of velocity profiles (accentuation of permanent and seasonal thermoclines) and reduction in volume absorption.*

Keywords: *climate change, active sonar, variable depth sonar, mid-frequency sonar, sonar performance, sonar equation, performance prediction*

1. SENSITIVITY OF THE SONAR EQUATION TERMS TO ENVIRONMENTAL CONFIGURATION

One traditional way to quantify the performance of an active sonar system is to evaluate the signal excess (SE), which is the difference between the signal-to-noise ratio (SNR) at the end of the full processing sequence and a detection threshold (DT). This is the well-known active sonar equation:

$$SE = SL–2TL + TS - ((NL–AG)(+)RL) + PG - DT \quad (1)$$

The detection threshold (DT) is an arbitrarily parameter tuned by the sonar operator according to his or her tolerance of false alarms. Source level (SL), self noise, array gain (AG) and time processing gain (PG) are essentially dependent on the sonar technology and cinematics, whereas the target strength (TS) is a characteristic of the target. The remaining terms—transmission loss (TL), ambient noise level, and reverberated level (RL)—are characteristic features of the environment and geometrical configuration (i.e., source and sonar locations and orientation). The propagation conditions may affect the directionality of ambient noise and, consequently, the array gain.

Figure 1 displays a flowchart summarizing the causal relationships between long-term, variable environmental features (on the left) and the associated acoustic properties of the ocean (in the center). These properties, in turn, are connected to the terms of the sonar equation, which they influence (on the right). The sound channel directly impacts the propagation of sonar pings, ambient noise, and reverberation. In other words, it influences TL, NL, and RL. The sound channel also affects the directionality of ambient noise and plays a role in array gain. Furthermore, the shape of the sound speed profile is of primary importance in optimizing the performance of variable depth sonars (VDS) for target detection. The target adopts a depth that minimizes its detectability. Thus, the shape of the sound channel directly influences sonar operating tactics and target depth. The sound channel is primarily characterized by the shape of the sound speed profile and volume attenuation. These characteristics depend mainly on ”hydrospheric” parameters, such as the temperature, salinity, and pH of water masses.

Ocean-atmosphere interactions impact TL, NL, RL, and AG by modifying the shape of the sound speed profile through the formation of shallow mixing layers (acoustic surface channels). These interactions are also responsible for ambient noise sources (e.g., wind, wave breaking, and rain) and surface loss due to wind-generated surface waves. The impact of ice cover on propagation, noise, and reverberation, which is limited to polar regions, will be disregarded in this article. Similarly, the impact of traffic noise and its evolution is ignored, which is reasonable apart from the traffic lines themselves, at 2 kHz. The biosphere plays an important role in ambient noise and as a source of reverberation. Human behavior is a major contributor to changes in sonar performance. Humans appear to be the main source of low-frequency noise, particularly close to shipping lanes and fishing zones, which are sensitive to historical and economic changes. We will firstly ignore the technological advance, even if there should be no stop in the increase of active sonar efficiency through new transmitted frequencies and waveforms, improvements to array technology, and advances in time processing. At the same time, more efficient anechoic coatings will likely improve target discretion against active sonar. All ”technological” terms in the sonar equation are affected, including the detection threshold, which could be lowered by improved echo-false alarm discrimination made possible by AI.

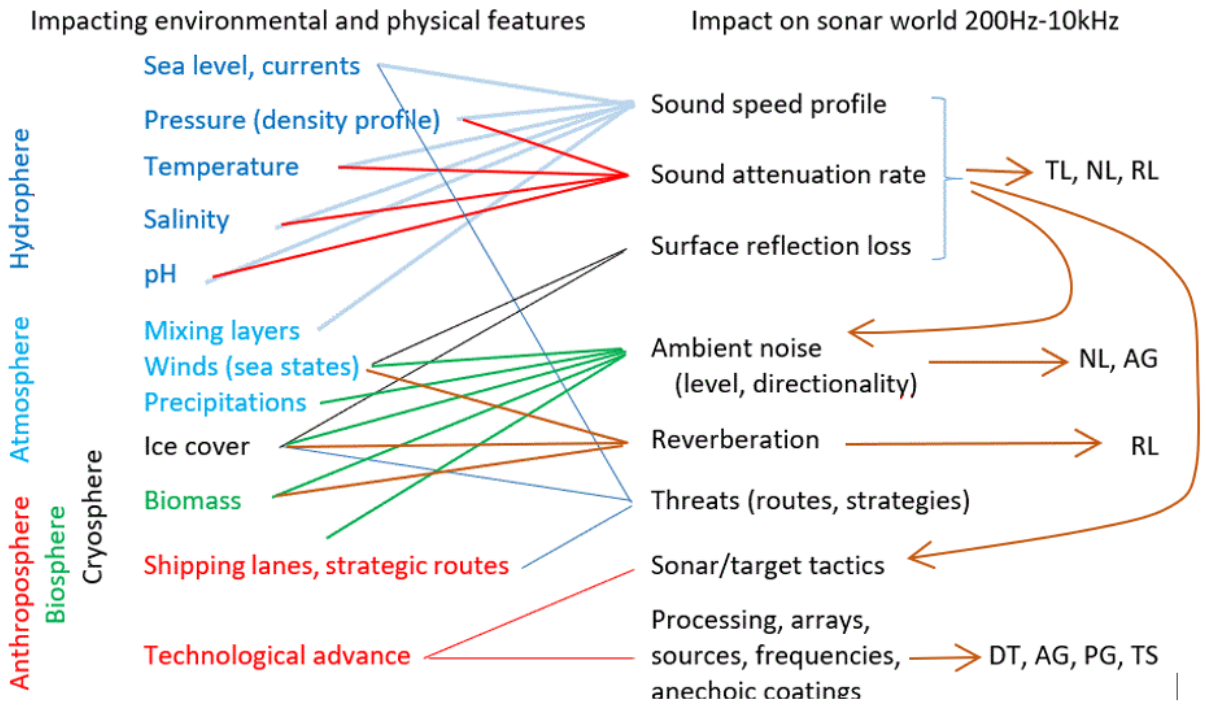


Figure 1: Connections between physical, environmental and acoustical features of the ocean and the various terms of the Sonar Equation.

2. METHODOLOGY FOR PREDICTING THE FUTURE EVOLUTION OF SONAR EQUATION TERMS

All natural and human-origin features impacting the sonar equation terms, which are listed in the previous section, have undergone significant observed modifications in the past and will likely continue to do so in the future. For example, considering the properties of water, the observed past surface temperature has increased at a rate of 0.1°C per decade over the last 50 years, and the pH has decreased by 0.025 per decade, reducing its basicity (”acidification”). The biomass was strongly reduced (factor 1/2 for predator fishes at trophic level 2-3). The shipping almost never stopped increasing since the end of World War II, resulting in an increase of more than 3 dB per decade of low frequency ambient noise. The future trajectory of sonar performance is contingent upon the future trends of these phenomena and others.

Regarding the physical features of the ocean and atmosphere, climate change is the main cause of change. It is primarily driven by the future trajectory of atmospheric CO2 concentration. For this parameter, we adopted the IPCC’s Shared Socioeconomic Pathway (SSP) 5-8.5, the most pessimistic and apparently most probable curve. The variations of physical ocean parameters over the 21st century have been the subject of numerous models, the results of which are available online: 120 models were compared in the Coupled Model Intercomparison Project - Phase 6 (CMIP6). We chose the predictions of the CNRM-ESM2 ocean climate model, developed by the CNRM in Toulouse, France. This model has a good reputation and provides trends that fall between the wide range of trends proposed by the 120 models. This ”median” trend is ideal for our initial investigation. The NOAA website provides the sampled vertical variations of ocean parameters by 2100 according to the CNRM-ESM2 model down to a depth of 500

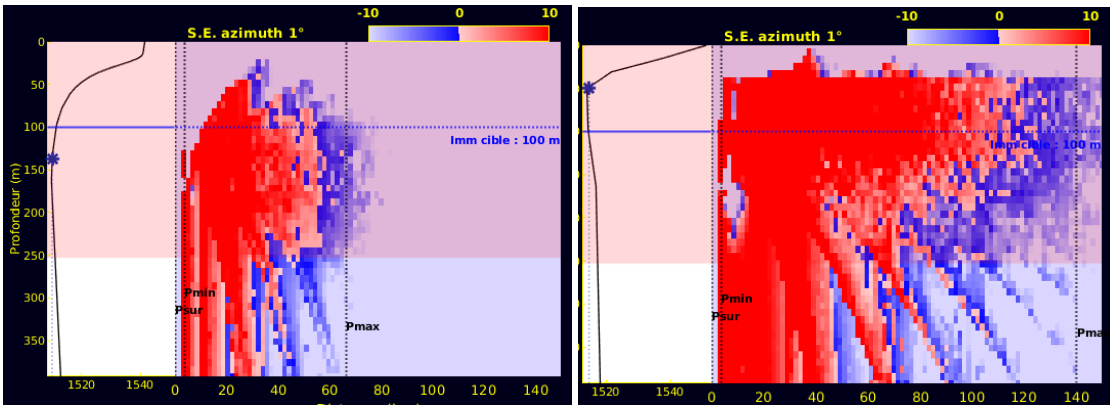


Figure 2: Vertical views of Signal Excess in August for years 2019 (left) and 2100 (right) at a same Mediterranean location (ranges in km); the sonar depth (asterisk) depends on the shape of the sound speed profile.

meters. We extrapolate at greater depths using an exponential decay inspired by past variability.

In this article, we examine the evolution of the performance of one mid-frequency active variable depth sonar over the 21st century. The properties of this VDS are listed below:

Transmitted waveform: Central frequency 2 kHz, Bandwidth 250 Hz, Pulse duration 4 s

Source level: 216 dB (omnidirectional)

Towing ship velocity: 12 knt (impact on self-noise)

Line array's Directivity Index: 20 dB (broadside)

Target Strength: 10 dB

Local mean seastates of considered seasons (impact on ambient noise and surface loss)

The specific feature of a VDS is its ability to be towed at variable depths and adapt to the local sound-speed profile according to specific rules, maximizing its detection efficiency. Similarly, the target chooses depths that minimize its indiscretion by avoiding surface channels and local sound speed minima. Our evaluations of the Signal Excess take this tactical game into account by changing the VDS depth according to the local sound speed profile for the given season and year. Figure 2 illustrates the radical impact that depth adaptation has on the resulting SE: the future shaping of the surface thermocline will result in a sound speed minimum at a depth that is very different from what it is now. Then, the VDS will not be towed at the same depth.

With the help of the sonar performance tool proprietary Thales software TARANIS, we computed the three-dimensional (3D) distributions of SE along bearing angle, range, and depth for the VDS located at the meshes of a 2° latitude x 2° longitude grid over the world ocean; examples of such results out from our computation are the vertical cuts, at given bearing angle, of the SE displayed by Figure 2.

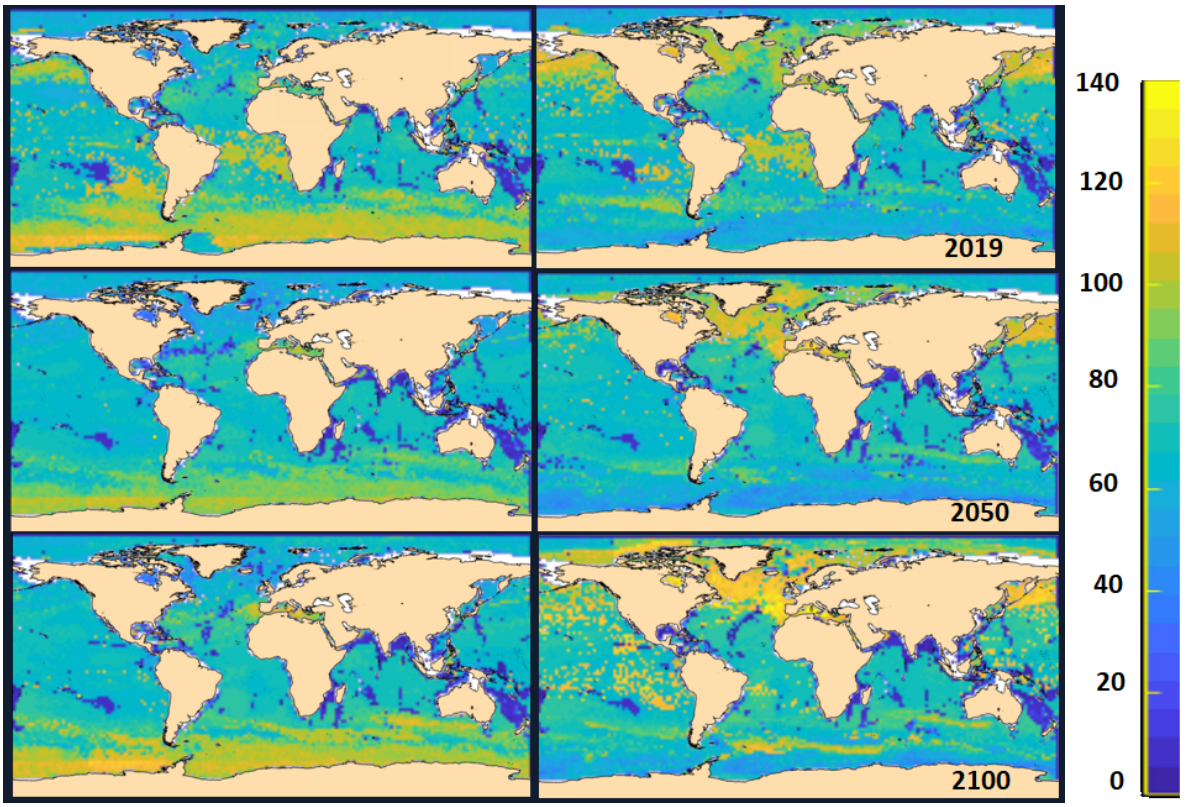


Figure 3: World maps of maximum detection range, in kyd, in February (left column) and August (right column) for years 2019, 2050 and 2100.

3. PRELIMINARY RESULTS

To get a direct sense of the planetary scale, it is necessary to extract and map some scalar moments of the full distributions of SE. Perhaps the most striking characteristic is the maximum detection range, which is evaluated at each sonar location by inspecting all bearing angles and possible target depths. Global 2° latitude x 2° longitude maps of maximum detection ranges are presented by Figure 3. Apart from some local variations, surprisingly simple trends emerge for the period 2019-2100. In the northern hemispheres, and particularly in Northern Atlantic and North-Western Pacific, detection ranges remain stable in winter and tend to increase with time in summer, as a result of the contradictory interplay between changes in the shape of velocity profiles (accentuation of permanent and seasonal thermoclines) and reduction in volume absorption due to water acidification. It seems that detection in the second Convergence Zone will become less uncommon in the future. Similar but less marked trends can be observed in the southern hemisphere. Certain areas of limited extension are currently unfavorable to the use of VDS and are likely to remain so until the end of the century: Arabian Sea, Bay of Bengal, Indo-Pacific and South-Western Pacific.

4. CONCLUSION AND PERSPECTIVES

The multifactorial problem of the impact of future ocean changes on sonar performance has already been the subject of many published analyses, as demonstrated by the extensive bibliography in reference [4]. No global conclusion seems to have emerged, and the results of these analyses may sound controversial. Many studies are limited to specific basins or regions (e.g. references [1], [2] [4]); there is a particular interest in Arctic and northern waters. In many respects, our analysis aligns with reference [3].

Future work should include the following extensions to this very preliminary article:

- Other metrics of sonar performance than just maximum detection range (detection volume, range of 1st contact loss, etc.);
- Additional environmental data (a set of other predictive models of ocean warming than CNRM-ESM2 should be selected);
- Additional modeling of global sonar performance (trying to give some figures in dB to the technological advances, trends towards lower frequencies);
- Additional modeling of threat evolution (trying to include realistic trends in anecoic coating technology);
- Evolution of performances for other sonar types (dipping sonar, hull mounted sonar, passive sonar);
- Evolution of optimal active sonar frequencies [0.1, 10]kHz, by basin and season;
- Impact trends on multistatism, fixed networks and acoustic communications.

REFERENCES

[1] A. Farkas : “The Effects of Climate Change on Oceanic Sonar Use in the Upper European Continental Shelf”, https:doi.org/10.26443/msurj.v19i1.203 **(2024)**.

[2] I. Possenti, G-J. Reichart, L. de Nooijer, F-P. Lam, C. de Jong, M. Colin, B. Binnerts, A. Boot, A. von der Heydt : “Predicting the contribution of climate change on North Atlantic underwater sound propagation”, *PeerJ 11:e16208 DOI 10.7717/peerj.16208* **(2023)**.

[3] A. Gilli, M. Gilli, A. Ricchi, A. Russo, S. Carniel : “Climate Change and Military Power: Hunting for Submarines in the Warming Ocean” in *TNSR Journal, Vol.7 Issue 2*, (2024).

[4] V.O. Oppeneer, N. Risser, R. van Vossen, K. Pelekanakis, G. Giorli, A. Russo : “The Impact of Climate Change on Sonar Performance: A Case Study in the High North” in *Proceedings of the Institute of Acoustics* (2024).

Comparisons Between Code- and Time-Division-Multiplexing for MIMO Sonar Signal Processing

Michaela Murzyniec[1], Yannik Steiniger[1], and Sven Schröder[2]

[1]German Aerospace Center (DLR) Institute for the Protection of Maritime Infrastructures, Fischkai 1 Bremerhaven 27572, Germany
[2]ATLAS ELEKTRONIK GmbH, Sebaldsbrücker Heerstr. 235 Bremen 28309, Germany

Contact author: Michaela Murzyniec, Fischkai 1, 27572 Bremerhaven, Germany, michaela.murzyniec@dlr.de

Abstract: *Multiple-input multiple-output (MIMO) sonar systems enhance angular resolution by using travel-time information from all transmitter–receiver pairs compared to conventional single-input multiple-output (SIMO) systems. When dealing with multiple transmitters in a MIMO system it is crucial to separate pulses properly at the receiver side, which is possible by using multiplexing methods such as code division multiplexing (CDM), time division multiplexing (TDM), or frequency division multiplexing (FDM). The latter is not suitable for MIMO systems due to spatial ambiguities and image artifacts it introduces and is therefore not considered within this study. CDM enables simultaneous pulse transmission, however, it suffers from crosstalk caused by a non-zero cross-correlation of the pulses, which limits its performance. On the other hand, TDM allows for perfect pulse separation and thus a reduction of the crosstalk. However, during a prolonged overall ping period changes in the acoustic channel may occur which lead to artifacts after coherent processing. This work investigates the performance of MIMO sonar systems which implement CDM, TDM, and their combination with a focus on the reduction of crosstalk in a short range application. The pulse separation methods are studied through simulations and experimental validation in a harbour environment using various pulse types and number of transmitters. The results show that the generalised integrated sidelobe level (GISL) as a measure for the crosstalk can be reduced by the integration of TDM.*

Keywords: *multiple-input multiple-output sonar, MIMO, pulse separation, code division multiplexing, time division multiplexing*

1. INTRODUCTION

Multiple-input multiple-output (MIMO) systems are widely investigated in modern radar [1] and recently more often in sonar systems [2, 3, 4]. Among other configurations such as single-input single-output (SISO), single-input multiple-output (SIMO), and multiple-input single-output (MISO), MIMO stands out by improving angular resolution through the use of a virtual array that exceeds the aperture of the physical array. This is possible when information from every transmitter–receiver pair is utilized. The main challenge when applying the MIMO principle lies in accurate separation of transmitter pulses at the receiver. This can be achieved using code division multiplexing (CDM), time division multiplexing (TDM) or frequency division multiplexing (FDM) [3]. CDM enables simultaneous transmissions within the same frequency band and time window by assigning distinct coding schemes to each transmitter. In this way, the ping rate is comparable with SIMO configuration and supports long-range applications. However, imperfect pulse separation can cause inter-channel interference known as crosstalk, which decreases detection performance and may introduce ghost target artifacts [4]. In TDM non-overlapping time slots are assigned to each transmitter while maintaining the same frequency band. This approach to separation reduces crosstalk and enables pulse distinction. However, this way of transmission reduces the overall ping rate of the MIMO system. Additionally, during pulse transmission temporal changes in the acoustic channel may appear, limiting its use to short- and medium-range applications. FDM separates pulses by allocating different frequency bands to each transmitter within the same time window. While this method allows for perfect pulse separation, it reduces the range resolution, introduces spatial ambiguities in the beampattern and is sensible to frequency-dependent propagation effects [5]. Therefore, FDM is unsuitable for coherent MIMO signal processing and is not considered here.

In this work, we present a comparison between CDM, TDM, and propose a combined approach for signal processing in MIMO sonar systems. Combining CDM and TDM should provide a high overall ping rate of the sonar system while at the same time reduce the crosstalk. Multiple scenarios with targets at different distances, various number of transmitter and different pulse types are simulated. More specifically, we investigate the pulse separation methods with linear frequency modulated (LFM), LFM sequence [2], LFM chain [6] and orthogonal frequency-division multiplexing (OFDM) signals for two and four transmitters. We analyse the performance of each method based on their point-spread function (PSF) and the generalised integrated sidelobe level (GISL). The simulation results are further validated through additional experiments conducted in a harbour environment.

2. SIMULATION SETUP

The simulation covers six scenarios of using CDM and TDM as shown in Figure 1. The first scenario resembles a pure CDM case where the two signals overlap completely in time. The second scenario involves a 10 ms delay between the two signals. In the third scenario, the second signal starts immediately after the first one finishes. The fourth, fifth, and sixth scenarios explore increasing gaps of 10 ms, 20 ms and 30 ms between the signals, respectively. For a considered processing range of 30 m the final two scenarios represent fully separated signals in time. To present the concept clearly, two LFM pulses are shown in Figure 1. However, all six scenarios consider various pairs of pulses, all using the same configuration of time delays. The pair of LFM pulses is chosen purely as a representative example.

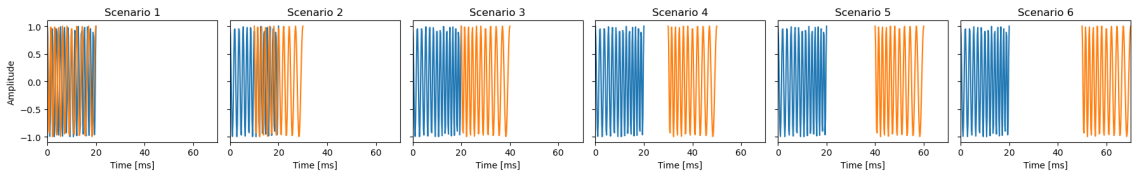


Figure 1: CDM-TDM scenarios considered in the simulation and experiment. Scenario 1 is pure CDM while the subsequent scenarios add a further time delay of 10 ms between the pulses. Blue line represents LFM-Up and orange line LFM-Down pulse.

The MIMO sonar system is simulated to match the experimental hardware specifications, including a sampling frequency f_s of 192 kHz, a center frequency f_c of 50 kHz and a bandwidth B of 20 kHz. The receiving antenna is a linear array with 32 channels. We simulate the scenarios with two and with four transmitters with a spacing of half a wavelength and centered at the origin of the receiver. All scenarios are simulated with LFM sequence (LFMS), LFM chain (LFMC) and OFDM pulses, which are designed based on previous work [4, 7]. In addition, the two transmitter setup also uses a LFM-Up pulse for the first and a LFM-Down pulse for the second transmitter. All pulse length T are set to 20 ms.

A point target is simulated at 2 m, 5 m and 10 m distance at an azimuth direction of 0°. The simulation also accounts for bottom and surface reverberation. Ambient noise is modeled as normally distributed, incoherent white noise.

3. MEASUREMENT SETUP

The measurement takes place in the fishing harbour in Bremerhaven, Germany nearby the Institute for the Protection of Maritime Infrastructures of the German Aerospace Center (DLR). The basin has depth of approximately 6 m and a width of 50 m. The sonar system, consisting of a 32-channel receiver antenna and one transmitter at each end of the receiver, was installed on the quay wall at the depth of 2 m. The estimated speed of sound c in the water is 1480 m/s. A air-filled sphere was placed as a target in front of the sonar at ranges R of 2 m, 5 m and 10 m at a depth of 2 m. The same six CDM-TDM scenarios with all four pulse types as in the simulation are considered in the experiment.

4. RESULTS

All scenarios are qualitatively analysed based on the output of the beamformer. From this we further calculate the GISL of the target beam signal $s_{TB}(r)$ as

$$\mathrm{GISL}^{(p)}(s_{TB}(r)) = \left(\frac{\int_{\Omega_{SL}} |s_{TB}(r)|^p dr}{\int_{\Omega_{ML}} |s_{TB}(r)|^p dr} \right)^{\frac{p}{2}} \tag{1}$$

to compare the scenarios quantitatively. With p set to 2, this corresponds to the integrated sidelobe level, which is commonly used for sonar pulse form optimization [8]. The area of the mainlobe Ω_{ML} is defined as

$$\Omega_{ML} = \left[R - \frac{c}{2B}, R + \frac{c}{2B} \right] \tag{2}$$

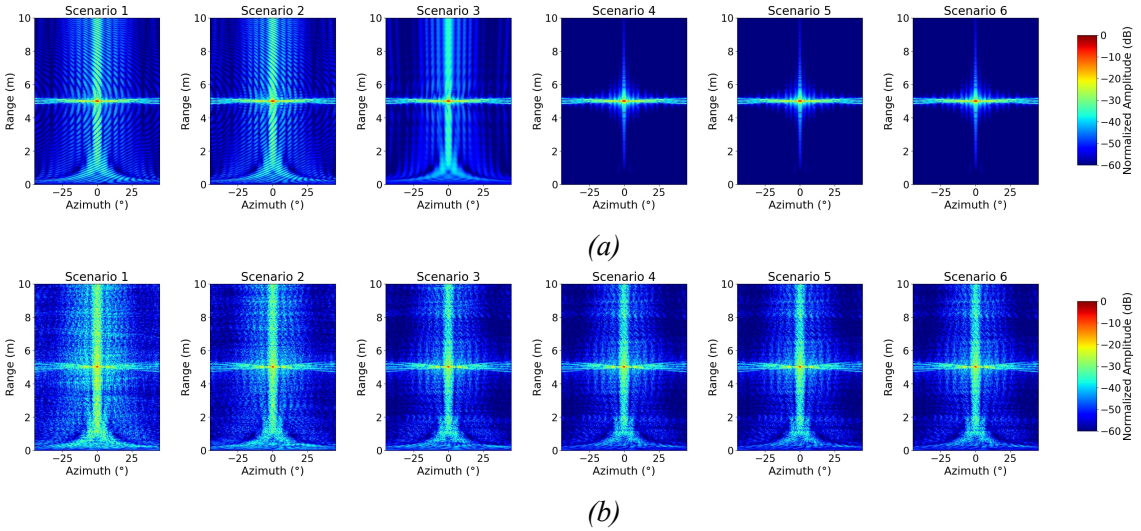


Figure 2: Simulated PSF for the six CDM-TDM scenarios obtained for configuration with (a) two transmitters with LFM-Up and LFM-Down pulses and (b) four transmitters with LFMC pulses. Point target located at 5 m and 0°.

and the area of the sidelobe Ω_{SL} as

$$\Omega_{SL} = \left[R - \frac{cT}{2}, R + \frac{cT}{2}\right] \setminus \Omega_{ML}. \tag{3}$$

Since due to movements of the target in the water its exact position is unknown in the measurement, we estimate it by first searching for the beam with the highest intensity. Afterwards, the position of the maximum within this beam in a range of ± 0.5 m around the setup range (2 m, 5 m or 10 m) is taken as the final target position.

Figure 2 compares the simulated PSF of all six scenarios for a target at 5 m for a MIMO sonar with two transmitters using the LFM-Up and LFM-Down pulses and a MIMO sonar with four transmitters using the LFMC pulses. In both cases it can be seen that for a larger time delay between the signals the crosstalk in the area around the target beam is reduced.

Figure 3 shows the simulated PSF of all considered pulse types for the scenarios 1, 4 and 6 with a MIMO sonar with two transmitters. The selected scenarios represent the case of pure CDM, a mixture of CDM and TDM without complete separation in time and complete separation in time. Especially the LFMS and LFMC pulses suffer from high crosstalk. Adding time delay between these pulses can efficiently reduce the crosstalk. As expected the best result is achieved for the complete separation of the signals in time. Furthermore, the overall noise background can be reduced by combining CDM with TDM.

Figure 4 displays the results of the corresponding experimental measurements for the same setup and scenarios as in Figure 3. For all setups the target is present in the direction in front of the MIMO sonar. However, only for the LFMC pulses the point target is clearly visible. Here the positive effect of combining CDM with TDM can be seen. When considering a not complete separation in time the crosstalk is reduced. Additionally, due to the short distance to the target effects from temporal changes in the acoustic channel are limited.

Table 1 and 2 list the GISL obtained during simulation studies and practical experiments for all investigated scenarios and pulse types for a MIMO sonar with two transmitters. The simulation shows an improvement in GISL for all pulse types when TDM is combined with CDM. In many cases a time shift of 20 ms between the pulses (scenario 3) is already sufficient to

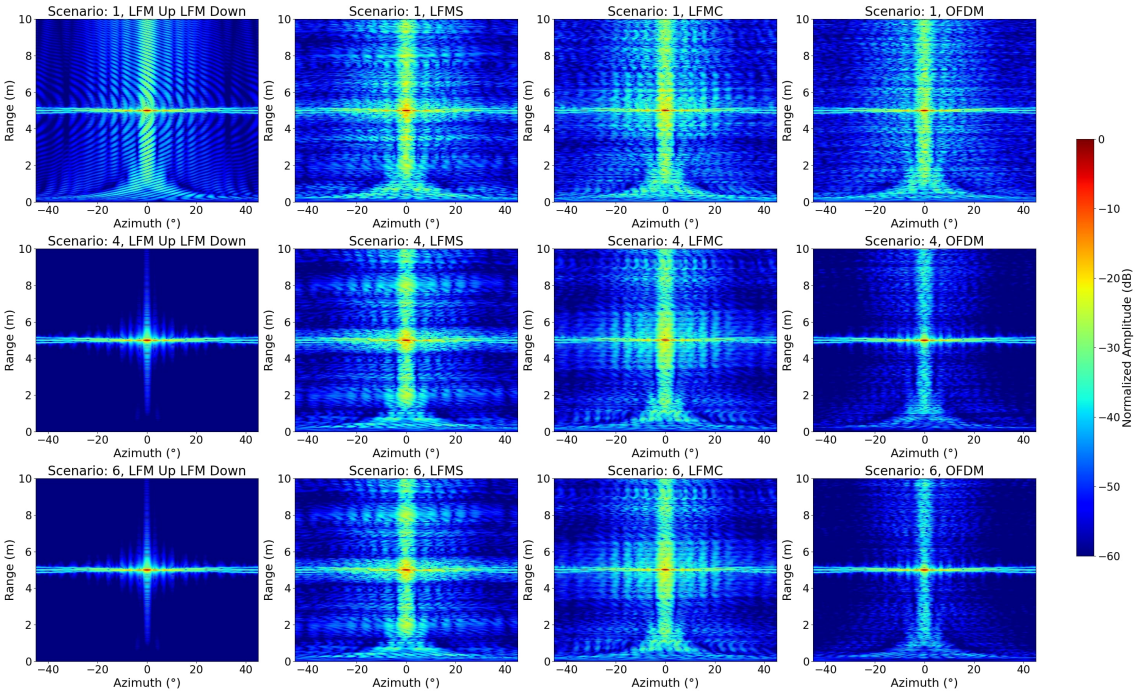


Figure 3: Simulated PSF for all considered pulse types obtained for a configuration of two transmitters the three CDM-TDM scenarios 1, 4 and 6. Point target located at 5 m and 0°.

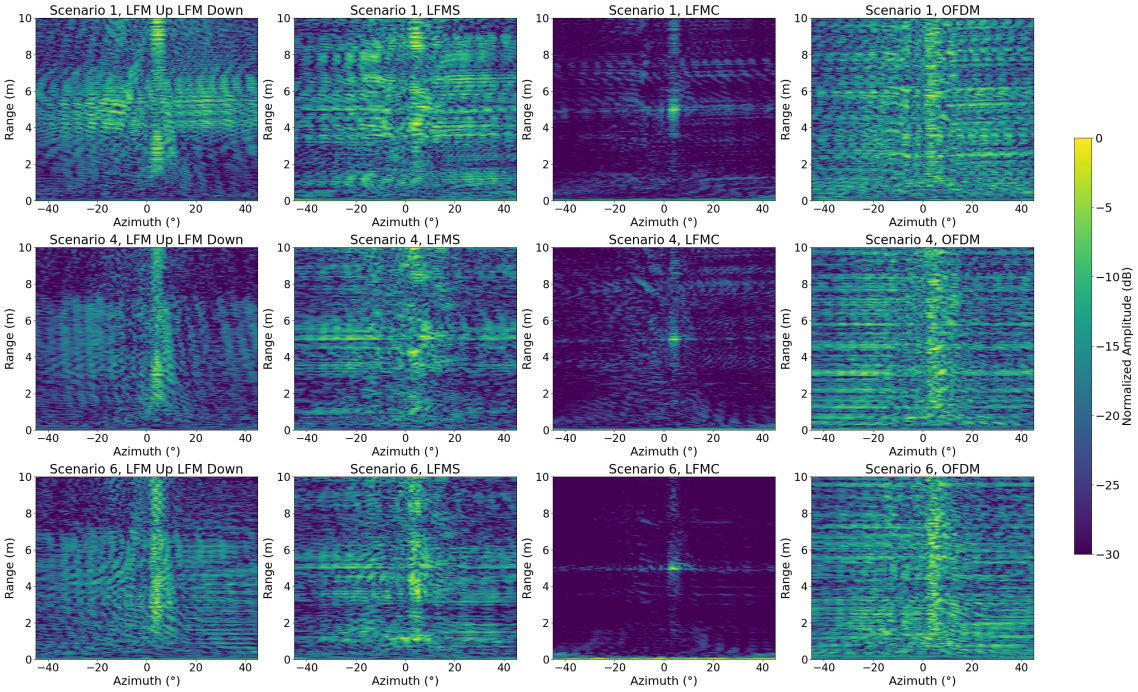


Figure 4: Beamformer output in the experiment for all considered pulse types obtained for a configuration of two transmitters the three CDM-TDM scenarios 1, 4 and 6. Point target located at 5 m and 0°.

Pulse type	Target range in m	GISL in dB for scenario					
		1	2	3	4	5	6
LFM-Up & LFM-Down	2	-28.1	-28.8	-29.9	-31.4	-32.5	-32.5
	5	-30.2	-30.8	-32.1	-33.8	-35.3	-35.3
	10	-31.4	-32.1	-33.4	-35.4	-37.4	-37.4
LFMS	2	-24.7	-24.7	-25.7	-26.4	-26.5	-26.5
	5	-26.7	-26.5	-27.9	-28.6	-28.7	-28.7
	10	-27.9	-28.1	-29.1	-29.8	-29.9	-29.9
LFMC	2	-24.8	-24.9	-26.1	-26.7	-26.8	-26.8
	5	-24.4	-24.5	-25.0	-25.5	-25.6	-25.6
	10	-26.5	-26.4	-27.6	-28.4	-28.5	-28.5
OFDM	2	-26.3	-26.9	-28.7	-29.8	-30.5	-30.5
	5	-28.3	-29.5	-30.8	-32.4	-32.8	-32.8
	10	-30.3	-30.1	-32.3	-33.4	-34.6	-33.9

Table 1: Performance of the CDM-TDM scenarios based on the GISL from the simulations.

Pulse type	Target range in m	GISL in dB for scenario					
		1	2	3	4	5	6
LFM-Up & LFM-Down	2	-4.1	1.2	-5.4	-1.0	-6.3	-7.4
	5	1.4	-1.6	2.5	-3.4	-1.0	-3.3
	10	0.0	-2.7	-2.4	1.4	-4.1	-1.9
LFMS	2	-1.8	-4.2	-2.4	-5.7	-8.2	-5.4
	5	-2.7	0.6	-3.0	-0.2	-4.1	-4.8
	10	-4.3	-1.1	-5.5	-3.4	-4.6	-2.3
LFMC	2	-15.0	-16.2	-15.9	-17.3	-20.1	-19.4
	5	-19.7	-14.1	-20.5	-16.8	-20.6	-22.5
	10	-4.4	-5.9	-2.8	-3.0	-5.7	-6.6
OFDM	2	-3.2	-1.1	-1.8	-3.5	-4.7	-8.8
	5	-4.9	-6.4	-2.2	-4.5	-6.5	-6.1
	10	-8.1	-6.1	-4.0	-5.3	-3.6	-8.5

Table 2: Performance of the CDM-TDM scenarios based on the GISL from the experiments.

achieve the same performance as with complete pulse separation. Compared to the simulations, the experiments show higher values for the GISL. The reason for this is the idealised nature of the simulation, which does not take all possible acoustic phenomena and noise sources into account. In general, the experimental results does not show the same clear behaviour of the GISL for the different scenarios. However, especially for the case of LFMC, for which the target is clearly visible, the performance improves with adding a larger time delay between the pulses.

5. CONCLUSIONS

This work has analysed the effect of combining TDM and CDM for pulse separation in a MIMO sonar system. A broad range of simulated setups with varying extend of time division, array configuration and pulse types have been conducted. Furthermore, these setups have

been investigated through experiments in a harbour environment. For short range applications, adding a time delay between transmitted signals can reduce the crosstalk of the MIMO system. This positive effect already applies for short time delays that do not involve complete separation of the signals in time and thus allow for higher ping rates compared to pure TDM. Further work needs to extend the simulation in order to include more acoustic phenomena and allow for temporal changes in the acoustic channel.

REFERENCES

[1] J. Li, P. Stoica, *MIMO Radar Signal Processing*, (John Wiley & Sons, 2009).

[2] Y. Pailhas, J. Houssineau, E. D. Delande, Y. Petillot, D. E. Clark: "Tracking underwater objects using large MIMO sonar systems" in *2nd Underwater Acoustics Conference and Exhibition*, (Rhodes, 2014).

[3] Y. Pailhas, Y. Petillot: "Wideband CDMA Waveforms for Large MIMO Sonar Systems" in *2015 Sensor Signal Processing for Defence (SSPD)*, (Edinburgh, 2015).

[4] S. Schröder, J. Reermann, S. Barnes, D. Kraus, A. Kummert: "Study on pulse form design for monostatic MIMO sonar systems" in *OCEANS 2023 - Limerick*, (Limerick, 2023).

[5] D. Cohen, D. Cohen, Y. C. Eldar: "High Resolution FDMA MIMO Radar", *IEEE Transactions on Aerospace and Electronic Systems* **56(4)**, 2806–2822 **(2020)**.

[6] S. Schröder, J. Reermann, M. Stephan, D. Kraus, A. Kummert: "Experimental demonstration of the angular resolution enhancement of a monostatic MIMO sonar" in *6th Underwater Acoustics Conference and Exhibition*, (virtual, 2021).

[7] S. Schröder, J. Reermann, S. Barnes, D. Kraus, A. Kummert: "Pulse form optimization for MIMO sonar systems" in *International Conference on Underwater Acoustics 2024*, (Bath, 2024).

[8] D. A. Hague: "Adaptive Transmit Waveform Design Using Multitone Sinusoidal Frequency Modulation", *IEEE Transactions on Aerospace and Electronic Systems* **57(2)**, 1274-1287 **(2021)**.

Decoupling Environmental Complexity: Direct Analysis in Sonar Performance Space

Jonas Halse Rygh[1] and Karl Thomas Hjelmervik[2]

[1]Norwegian Defence Research Establishment, Nedre vei 8, 3183 Horten, Norway
[2]University of South-Eastern Norway, Raveien 215, 3184 Borre, Norway

Contact author: J. H Rygh, Norwegian Defence Research Establishment, Nedre vei 8, 3183 Horten, Norway, e-mail: jonas-halse.rygh@ffi.no

Abstract: *The impact of environmental variability on sonar performance variability is crucial to understanding how to effectively manage sonar resources. Complexities arise due to the fact that it is the interplay between different environmental parameters that determine the sonar performance. It may not only be the sound speed profile, but also the bottom depth and wind speed. Consequently, there is need for an algorithm that can capture the combined effects of all included environmental parameters. We propose a method that uses empirical orthogonal functions and clustering on computed signal excesseses, in sonar performance space. This bypasses all environmental parameters altogether. The resulting clusters are then analysed and compared with the clusters one would get by only focusing on one environmental parameter, the sound speed profile in our case.*

Keywords: *Sonar performance modelling, Environment, Sonar operation planning aids.*

1. INTRODUCTION

Sonar performance modeling is crucial for planning sonar operations. Environmental data—such as bathymetry, sediment types, and oceanographic conditions—inform acoustic propagation models that estimate transmission loss and, for active sonar, expected reverberation levels [1]. The sonar equation allows the estimation of the signal excess, a parameter commonly used for the planning and execution of sonar missions [2].

In UACE 2023 [3] we proposed a method of organising the sonar environment into clusters of similar sound speed profiles, and further estimating the sonar performance based on representative sound speed profiles for each cluster. While this method is fast and programmatically simple, it fails to encompass other environmental parameters, such as bottom depths, sediment types, and ambient noise levels. For instance, two identical Munk profiles [4] with different bottom depths would be placed in the same cluster if a naive clustering algorithm on sound speed alone was used. In reality these two situations lead to fundamentally different sonar performance; one with convergence zones and one without depending on the bottom depths.

We propose an alternative method that bypasses all environmental parameters by compressing and clustering a pre-computed subset of signal excesses (or other 2D measures). This method captures the combined effects of the supplied environmental parameters, which can be arbitrarily many, at the cost of a significant increase in the computation time.

Here we apply the method on an area around the Vøring Plateau [5], with environmental data from Gebco [6] and Copernicus Marine Services [7].

2. METHOD

We select (lat, lon) in the extent $[\text{lat}_0, \text{lat}_1, \text{lon}_0, \text{lon}_1]$ such that the distance along any point is approximately Δ, which results in N points. The resulting grid is thus not uniform, but adjusted such that points near the poles are sampled less. Then we compute the signal excess for each point, using the acoustic model Lybin [8], with predefined sonar settings and environmental parameters using the supplied environmental data. The method is currently limited to range-independent environmental settings.

The signal excesses are then cropped to a selected depth interval $(z_{\text{min}}, z_{\text{max}})$, due to both numerical constraints (arrays must have the same dimensionality) and to limit the method to consider relevant depths (which should be determined by the application). The cropped signal excesses are then flattened to a vector, with dimensions $N_d = N_z \times N_r$, where N_z and N_r are the depth and range dimensionalities, respectively. The remaining EOF-analysis and clustering then follows the steps of [3]. We then form the matrix of signal excess vectors

$$\mathbf{S} = \begin{pmatrix} \vec{s}_1 & \vec{s}_2 & \dots & \vec{s}_N \end{pmatrix}^T, \tag{1}$$

where $\vec{s}_i$ is the flattened signal excess vector for $(\text{lat}_i, \text{lon}_i)$. We furthermore define the matrix

$$\mathbf{G} = \frac{\mathbf{S}_{\text{linear}} - \mu\left(\mathbf{S}_{\text{linear}}\right)}{\sigma\left(\mathbf{S}_{\text{linear}}\right)}, \tag{2}$$

where $\mu(x)$ and $\sigma(x)$ is the average and standard deviation of the signal excess matrix in linear scale along the N dimension with resulting dimensionality N_d. Next we use Empirical orthogonal function (EOF) analysis [9] and calculate the covariance matrix $\mathbf{X} = \mathbf{G}^T\mathbf{G}$, with

rank$(\mathbf{X}) = \min(N, N_d)$. For a given N_d, we must there for at least have $N \geq N_d$ samples. The EOFs are the eigenvectors $\vec{u}$ of the covariance matrix

$$\mathbf{X}\vec{u} = \lambda\vec{u}, \tag{3}$$

and the EOF coefficients are given by

$$c_{ij} = \vec{g}_i \cdot \vec{u}_j, \tag{4}$$

where i runs over all N, and j runs over a subset of $N_{\text{pca}} < N_d$. The N_{pca} is selected such that most of the variance is contained in the reduced dataset, often set to a certain threshold. It is worth mentioning that this step may be crucial, as clustering algorithms struggle with large vector sizes. In our case, $N_d = N_z \times N_r$ can be quite large. This was not an isse in [3] since the dimensionality was of order $N_{\text{model-z}}$, where $N_{\text{model-z}}$ was the number of depths in the ocean model data. Lastly, a clustering algorithm, in our case K-means [10] is used on the coefficient matrix c_{ij}, with a predefined number of clusters.

3. DATA

The area of interest is given by the extent $[67°\text{N}, 70°\text{N}, 5°\text{W}, 5°\text{E}]$, which covers an area of about $A \approx 136,000\,\text{km}^2$. We gather bathymetry data from Gebco, and temperature and salinity profiles from MET [11] (compiled from Copernicus Marine Service) at time 14 September 2024, 20:00. Figure 1 presents the environmental parameters of the area.

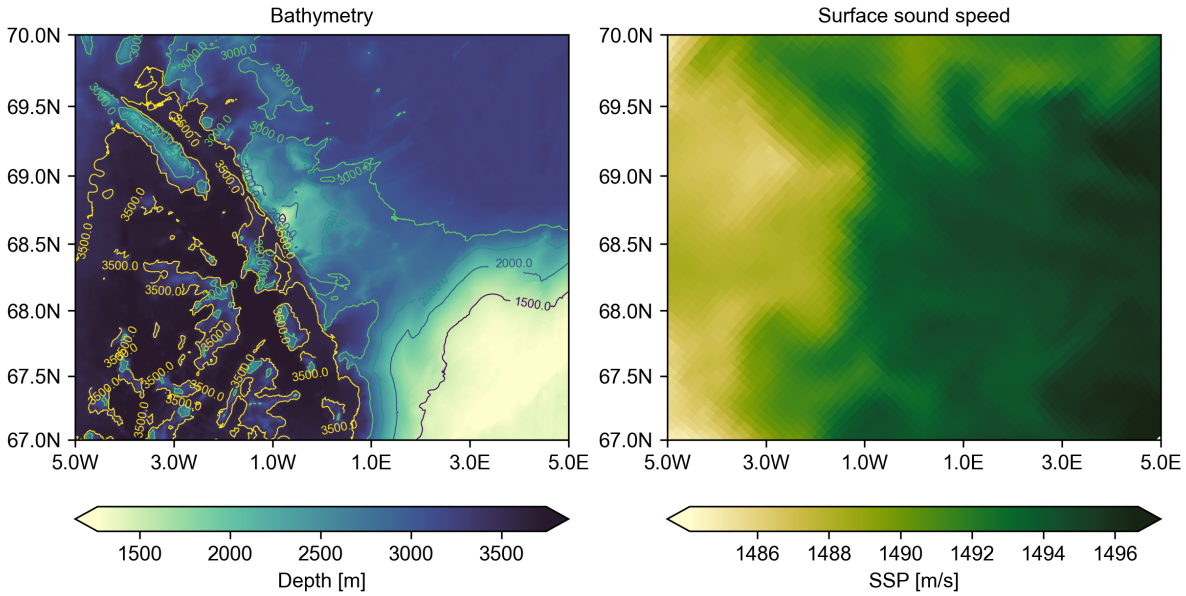


Figure 1: Plot of bathymetry data (left) and surface sound speed profile (right).

This area contains the Vøring plateau, which is seen in the bottom right corner, with bottom depths ranging from $1500\,\text{m}$ to $3500\,\text{m}$. The surface sound speed shows a variation of $10\,\text{m/s}$, from $\simeq 1486\,\text{m/s}$ in the west to $\simeq 1496\,\text{m/s}$ in the east. The reason for choosing this area is that it may or may not allow convergence zones.

4. RESULTS

The grid of latitudes and longitudes was generated using a spacing of $\Delta = 4.5\,\text{km}$, resulting in $N = 6791$ points. The depth interval was chosen to be $z_{\text{min}} = 0\,\text{m}$ and $z_{\text{max}} = 400\,\text{m}$, which results in a flattened dimensionality of $N_d = N_z \times N_r = 40 \times 140 = 5600 < N$, using $\Delta z = 10\,\text{m}$ and $\Delta r = 500\,\text{m}$ with a maximum range of $70\,\text{km}$. The transducer settings were for a generic long range sonar placed at depth $z = 50\,\text{m}$. The signal excesses were computed in 30 minutes, using a laptop computer.

The principal components were extracted from the averaged flattened signal excesses, using the PCA module from scikit-learn [12], using 320 components, chosen such that the explained variance was at 90%. The number of clusters were then visually determined to be $N_{\text{clusters}} = 7$, by observing when additional clusters were deemed negligible in size. The resulting clusters are shown in figure 3 and compared to the clusters from the method used in [3].

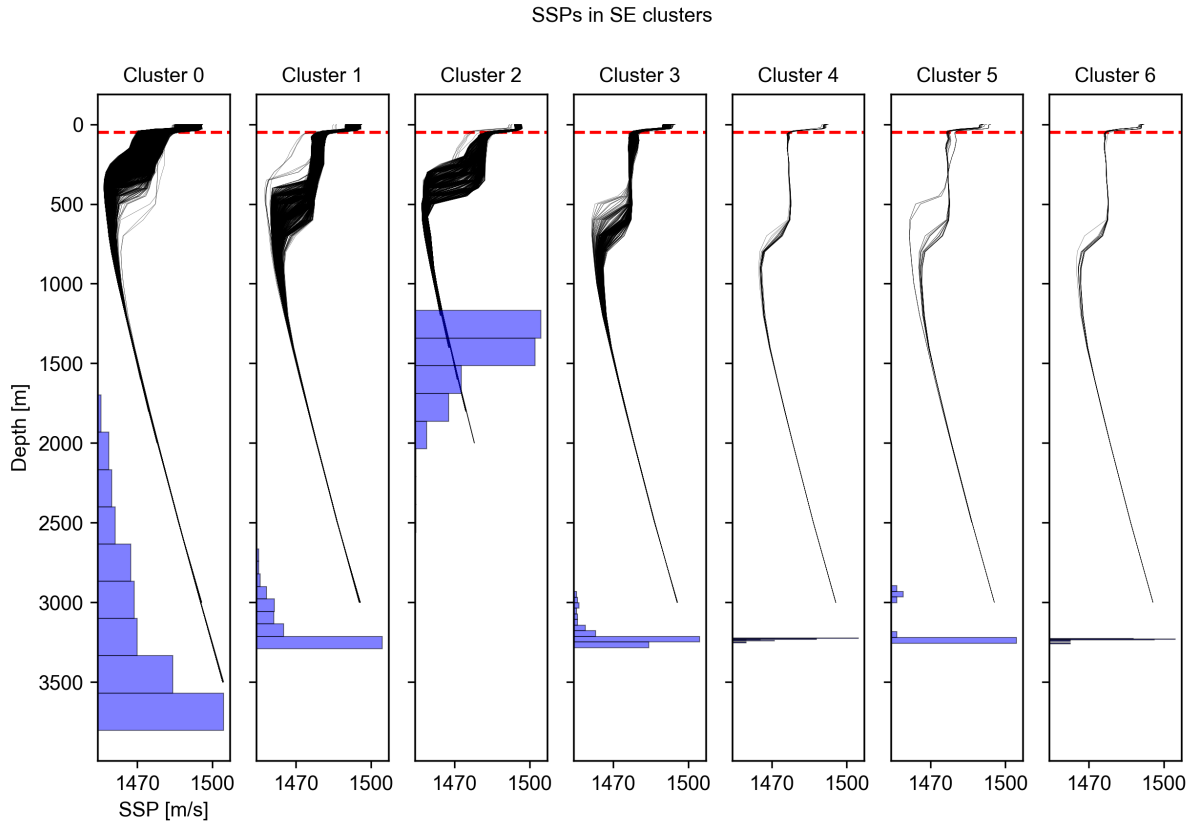


Figure 2: Sound speed profiles and bottom depth distributions for each cluster

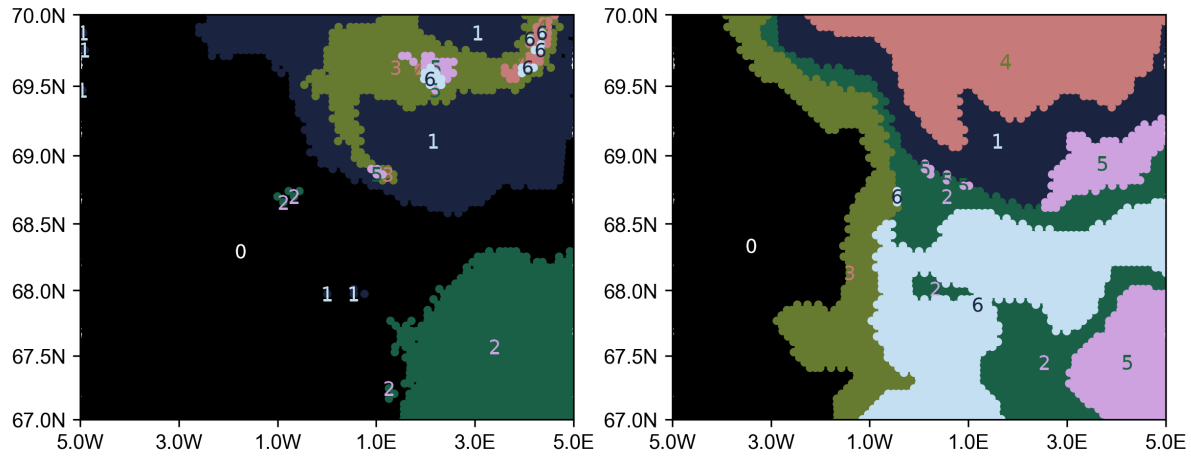


Figure 3: Clusters obtained using current method (left) and clusters obtained using the method in [3] (right)

Figure 4 shows the average and the 90th percentile (based on distances from cluster centroid) of the signal excess for each cluster, denoted by $\langle n \rangle$ and δn.

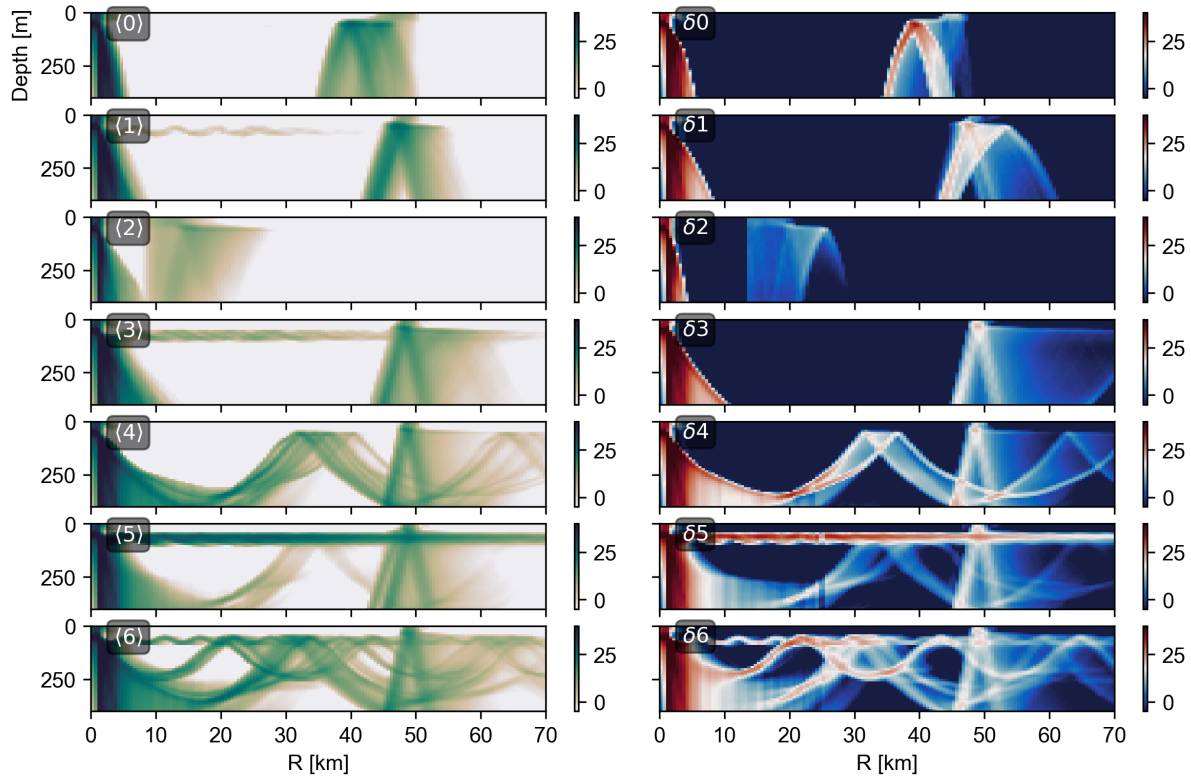


Figure 4: Average (left) and 90th percentile (right) of the signal excesses for each cluster, shown in logarithmic scale.

5. DISCUSSION

Consider figure 3, cluster 0 from both methods resemble each other to about 1°W, where the 0 cluster obtained from the proposed method continues eastward in a narrow passage between

cluster 1 and 2. Comparing with figure 2 we see that the overlapping area is very deep ($>$ $3000\,\mathrm{m}$) and contains sound speed surface values that are in the lower range of the surface values. We also observe that the clusters obtained from only considering sound speed profiles seem to capture areas with large gradients in surface sound speed values, and not bottom depths.

The most important result is found in the South-Easternmost area, where the bathymetry is around $1500\,\mathrm{m}$. The current method puts this area into one cluster, namely cluster 2. This cluster is not repeated elsewhere (except around $(69°\mathrm{N}, 1°\mathrm{W})$ where the depth is also around $1500\,\mathrm{m}$). This means that the cluster mostly contains shallow sound speed profiles and bottom depths ($1500\,\mathrm{m}$). Comparing this to the clusters found using the former method, we observe that the corresponding clusters, $2'$ and $5'$ (n' denotes clusters using the other method) contain sound speed profiles that are both shallow and deep ($3000\,\mathrm{m}$).

Now consider figure 4. We observe that the signal excess plots mainly describe two different situations: convergence zones in clusters $n_{\mathrm{cz}} \in \{0, 1, 3, 4, 5, 6\}$ and bottom reflections in clusters $n_{\mathrm{br}} \in \{2\}$. We also see that the 90th percentile and the average are almost similar for each cluster, which suggests that the average is representative for individual signal excesses in the covered area. The exception is for clusters 1 and 2, where some of the profiles contain a channel. They are expected to be allocated to a separate cluster, but this has not been achieved even by manipulating the number of clusters. Therefore, more work is to be done on this issue. However, the current method is an improvement since the former method results in single clusters containing both convergence zones and bottom reflections, which would be a poor representation of the sonar performance for an area. Thus, the current method can supply decision makers with maps that account for convergence zones.

From figure 4 we also observe differences within clusters describing the same situation. In n_{cz}, we see that the clusters $(3, 5, 6)$ have a variety of sound speed channels and are geographically close to each other; see figure 2. Cluster 4 is barely visible, but is located in the North-Easternmost area, in-between cluster 6, and almost contains a sound speed channel. Clusters $n_{\mathrm{cz}} = 0$ and $n_{\mathrm{cz}} = 1$ have pure convergence zones without sound speed channels. The convergence zone is also located at different ranges for different clusters, *e.g.* $40\,\mathrm{km}$ for cluster 0 and around $50\,\mathrm{km}$ for cluster 1. This is expected, as cluster 0 contains deeper depths $(3000\,\mathrm{m})$, allowing steeper angles to complete the loop before hitting the bottom.

The method seems successful in creating clusters that are physically relevant. Although this method seems to work as intended, there are currently several challenges. Firstly, the clusters are valid for only one set of sonar parameters. This may be fine for hull-mounted sonars, where the depth parameter is constant, but may be non-trivial for variable depth sonars. Secondly, the current method relies on range-independent environmental parameters. There are currently no obvious way of implementing range-dependent environmental parameters into a clustering algorithm. Finally, the new approach is significantly slower than the method described in [3].

6. SUMMARY

A clustering algorithm for geographical mapping of sonar performance was implemented and it's performance compared to an existing algorithm [3]. The new clustering approach decouples environmental complexities by clustering directly in sonar performance space. The sonar performance was estimated using the acoustic model Lybin [2] with input from both realistic oceanographic and bathymetric data. The resulting clusters were shown to be physically relevant and could be divided into clusters influenced by convergence zones and bottom reflection. Lastly, the method demonstrated that correct predictions of sonar performance relies on

the interplay between several different environmental parameters, and that naively focusing on one alone is not sufficient.

REFERENCES

[1] J. M. Hovem, *Marine Acoustics - The Physics of Sound in Underwater Environments*. Peninsula Publishing, 2012.

[2] E. M. Bøhler, P. Østenstad, and K. T. Hjelmervik, "A Monte Carlo approach for capturing the uncertainty in sonar performance modelling," p. 8.

[3] J. H. Rygh, K. T. Hjelmervik, and T. Jenserud, "Sonar performance modelling for clustered watermasses," in *Proceedings of the Underwater Acoustics Conference and Exhibition*, (Kalamata, Greece), pp. 693–700, Institute of Applied and Computational Mathematics (I.A.C.M.), Foundation for Research and Technology – Hellas (FORTH), June 2023.

[4] W. H. Munk, "Sound channel in an exponentially stratified ocean, with application to sofar," *The Journal of the Acoustical Society of America*, vol. 55, pp. 220–226, 02 1974.

[5] D. Myer, S. Constable, and K. Key, "Magnetotelluric evidence for layered mafic intrusions beneath the vøring and exmouth rifted margins," *Physics of the Earth and Planetary Interiors*, vol. 220, p. 1–10, 07 2013.

[6] G. B. C. G. 2023, "The gebco_2023 grid - a continuous terrain model of the global oceans and land.," Apr 2023.

[7] Copernicus Marine Service, "Copernicus marine environment monitoring service (cmems)." `https://marine.copernicus.eu`. Accessed: 2024-04-01.

[8] E. Dombestein and T. Jenserud, "Improving Underwater Surveillance: LYBIN Sonar performance Prediction," in *MAST Americas 2010*, (Washington DC, USA), 2010.

[9] K. T. Hjelmervik and K. Hjelmervik, "Improved estimation of oceanographic climatology using empirical orthogonal functions and clustering," in *2013 MTS/IEEE OCEANS - Bergen*, pp. 1–5, June 2013.

[10] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[11] Norwegian Meteorological Institute, "Arctic ocean physics analysis and forecast, 6.25 km 1-hourly frequency." `https://thredds.met.no/thredds/catalog/fou-hi/topaz5-arc-1hr_files/catalog.html`.

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau,

M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

Mapping the Impact of Climate Change on Indicators of Acoustic Propagation in the Atlantic Ocean

V.O. Oppeneer[1], N. Risser[1], A. Russo[2], G. Giorli[2], R. van Vossen[1]

[1]TNO, The Hague, The Netherlands
[2]CMRE, La Spezia, Italy

victor.oppeneer@tno.nl

Abstract: *Climate change is causing a significant warming of the seas and oceans, leading to changes in temperature and salinity distributions within the water column. The sound-speed-profile (SSP) depends on these distributions, and its vertical shape affects the acoustic propagation inside the water column. However, making inferences on the impact of climate change on acoustic propagation is not straightforward. In order to increase our understanding of this complex relationship, this paper presents a case study on propagation in the Atlantic Ocean. It considers the effect of climate change on indicators of acoustic propagation such as the sonic layer depth, sound channel axis, and convergence zones. Based on (horizontally) high-resolution climate modelling projections, maps are created that reveal how the sonic layer depth and channel axis might evolve over time up to 2050. By coupling the climate data with ocean acoustic propagation models, similar maps for convergence zone distances can be created. Maps of these properties provide insights into latitudes and longitudes where the effects of climate change on acoustic propagation are the strongest and in which regions a limited effect is expected. An outlook is given on the validity of historical databases.*

Keywords: *Climate change, sonic layer depth, sound channel axis, convergence zone, Atlantic Ocean*

1. INTRODUCTION

Climate change is causing a significant warming of the seas and oceans [1], and leading to changes in temperature and salinity distributions within the water column. The underwater propagation of sound depends on the sound speed profile of the water column, as well as on the sea surface and sediment conditions [2] [3] [4]. The propagation of sound is important for signal modelling, noise mapping, and sonar performance. Therefore, it is necessary to know and understand if the changing climate influences the acoustic propagation and to what degree. In addition, it tells how valid historical databases such as the World Ocean Atlas [5] are.

Previous studies that investigated the effect of climate change on underwater acoustics [6] [7] [8] focused on case studies that were relatively localized in space. It showed that in specific environments there is an impact of the changing climate on the underwater acoustics; such as the appearance of new ducts, changes in long range propagation in the arctic, and that the ambient noise may change even without significant changes in sound speed profiles. However, the propagation is influenced by not only the environment, but also the chosen source and receiver depths, which complicates its study and analysis.

In this paper, we present a case study on the effect of the climate change on indicators of acoustic propagation in the Atlantic using environmental data from climate models. The acoustic properties investigated are the sonic layer depth, the channel axis depth, and the convergence zone lengths.

First background information is given on the chosen indicators, followed by an explanation on the area of interest, the used climate model data, and how the indicators can be computed from the climate data. Then the results are displayed and commented on. Finally, a conclusion is given on the presented data and figures.

2. BACKGROUND INDICATORS OF ACOUSTIC PROPAGATION

The sonic-layer-depth (SLD) is the vertical distance from the ocean surface to the depth of the sound speed maximum [9]. This generally coincides with the mixing layer depth, a region for which the seawater mixes and the temperature and salinity are almost constant. The sound speed is a function of temperature, salinity, and pressure. Hence, the end of the mixing layer generally coincides with a (local) sound speed maximum. The SLD is of interest, as it characterizes acoustic surface ducts [9].

Below the mixing layer depth there is generally a sharp decline in the ocean's temperature resulting in decreasing sound speed. At some point, the water column has a constant temperature all the way down to the bottom. This part of the water column is called the deep isothermal layer [10]. At the top of the isothermal layer, there is the sound speed minimum, usually referred to as the sound channel axis (SCA). The SCA is of interest, as it allows long-range propagation of sound.

The convergence zone distance is the horizontal distance between a source and a receiver point (at the same depth) for which high levels of sound occur. A necessary requirement for a convergence zone to exist is that rays may not interact with the sediment, as this would break up caustics [11]. This requirement is achieved by the existence of a so-called critical depth [10], a depth deeper than the SCA, whose sound-speed equals the sound-speed maximum for depths smaller than the SCA.

3. METHOD

3.1. Area of interest and climate models

The chosen area of interest is the Atlantic Ocean between -75 to 15 degrees of longitude and 10 to 68 degrees latitude in steps of 4 degrees in latitude, and 5 degrees in longitude. Points with water depth shallower than 1000 m were ignored, as well as points in the Mediterranean Sea. Two months of interested are chosen: March and September. These are assumed to be the coldest and warmest months respectively.

In this study, temperature and salinity profiles were used at the latitude-longitude position of each grid cell. These profiles were combined with the depth and an equation [13] to get the sound-speed profile (SSP) at each grid cell. The profiles come from climate model data belonging to CMIP6 [14] [15] and following the Shared Socio-economic Pathway 5-8.5 scenario [12]. Three models were used: CNRM-CM6-1 [16], EC Earth3P [17], and HadGEM3-GC3 [18]. Monthly averaged climate model data was selected in the time range from 1971 to 2050.

3.2. Calculating the Indicators of Acoustic Propagation

Sound channel axis (SCA)

After having determined the SSP, the sound channel axis is determined by identifying the depth of minimum sound-speed. In some cases, especially close to the arctic or Greenland, the sound-speed minimum occurs at, or close to, the air-water interface.

Sonic layer depth (SLD)

The sonic layer depth is the depth for which the sound-speed profile reaches a local maximum (due to a mixing creating a more-or-less constant temperature and salinity in the top layer of the water column) [10]. This local maximum is the first local maximum between the sea-surface and the channel axis, assuming that there is a well-mixed top layer.

Convergence zone length (CZ length)

It is not trivial to choose a source depth in computations to determine the distance at which a convergence zone happens. However, if a convergence zone is present, it is known that there is a critical depth, at which the rays must travel (almost) horizontally. This effect can be exploited to find the convergence zone distance as follows:

1. Determine the SCA depth.
2. Determine the maximum SSP above the SCA depth (uSSPmax).
3. Check whether there exists a critical depth.
    - If there is none, then there can be no convergence zone.
4. Place a source at the critical depth and simulate some acoustic rays that leave the source almost horizontally. Trace these rays through the water column.
5. Check at what distance the rays return to this depth and orientation and save this distance.

3.3. Combining the results

The SCA, SLD, and CZ length, are all computed using the three climate models. Furthermore, as specific years could be outliers, it is important to consider the results of a

group of years together. Therefore, statistics is applied to the datapoints. For each month, the three models give different results for the SCA depth, SLD, and CZ lengths. It was chosen to pick the median value of the three, assuming that Nan's (Nan=Not a number) are ignored (for example a Nan can occur when there is no convergence zone). The next step is combining the results of a set of years. In this work the results of 20 years are combined, by taking the mean. The occurrences of the finite values are recorded on the model level. If only one out of the three models result in a convergence zone of length L, then L is reported as the convergence zone length for that position. However, the occurrence for that position and year is only equal to 1/3.

4. RESULTS AND DISCUSSION

4.1. Sound channel axis depth

For the month of March, it can be observed that the SCA does not vary much in the period of 1971-1990 and 2031-2050 for most latitudes below 45 degrees, as can be seen in Fig. 1. For this part of the ocean, the SCA depth is between 800 and 1600 m which is deeper than the areas most affected by climate change [19]. Above the 45 degrees latitude, there are large changes in the second period with changes up to 50% in positive and negative directions compared to 1971-1990, with the largest changes between the coast of Canada up to the Reykjanes Ridge and Greenland. This area is where the ocean gets an influx of cold fresher water due to melting of ice.

Although it is not shown, similar results can be obtained for other months such as September. The major difference between the two months is the direction of the percentual changes.

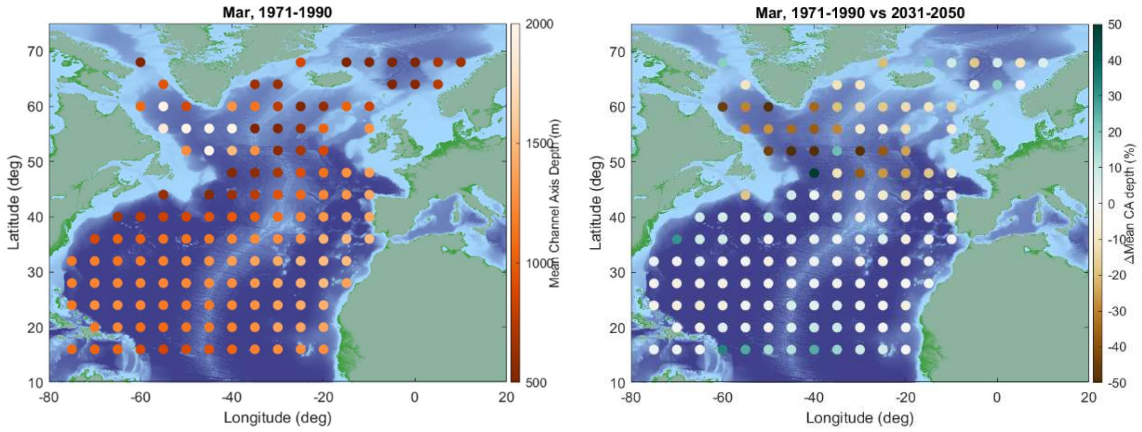


Fig.1: Map of the channel axis depth in the Atlantic Ocean for the month of March. Left: average from 1971 to 1990. Right: percentual variation for the time bracket of 2031-2050 compared to 1971-1990.

4.2. Sonic layer depth

The sonic layer depth can only exist if there is a good mixed surface layer, where the sound can get trapped. As can be seen in Fig 2, the SLD decreases over time for most of the Atlantic in March. Even though the SLD decreased for the most part of the Atlantic Ocean, there is an exception South of Greenland, where there is a region with a strong increase in the SLD, between 10% and 50%.

During the summer, it can happen that the increase in sea surface temperature leads to a (small) gradient of warmer water near the surface to colder water near the end of the mixing layer. This then leads to a negative gradient in the sound-speed-profile. The result is that there are months in the year for which there is no sonic layer. In our study, the sonic layer depth completely disappears for only two locations off the coast of west Africa in March. However, in September there is almost nowhere a SLD because the sea surface temperature is too high.

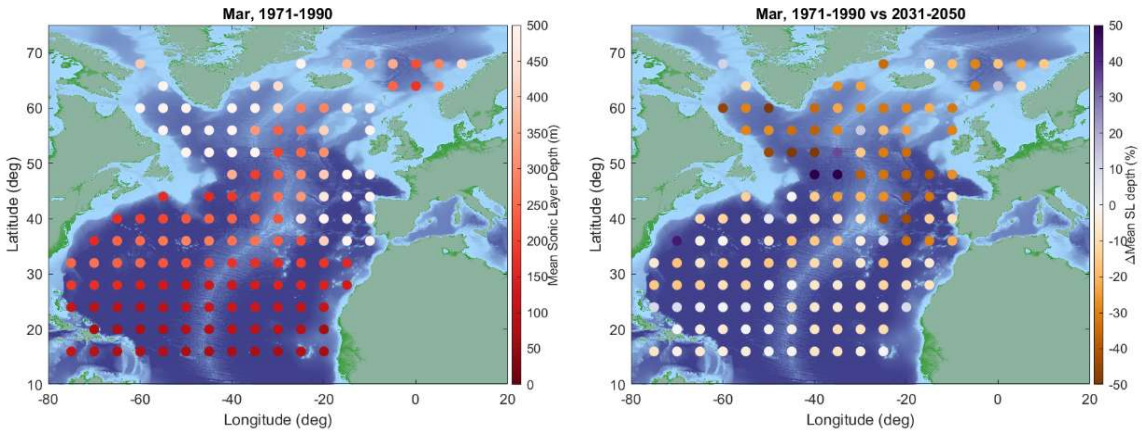


Fig.2: Map of the sonic layer depth in the Atlantic in March. Left: average from 1971 to 1990. Right: percentual variation for the time bracket 2031-2050 compared to 1971-1990.

4.3. Convergence zone length

Fig. 3 shows a map of the CZ lengths for both the month of March and September, and their percentual changes from 1971-1990 to 2031-2050. The first observation is that during March there is a higher variability in the CZ lengths than in September in the period 1971-1990. The percentual changes are also larger in March than in September. Furthermore, it can be observed that along the Atlantic ridge and around Greenland and Iceland (both regions where the ocean is shallower), there are more instances in which a CZ cannot exist in September. This effect will be even stronger for the period 2031-2050 and is caused by the surface temperature increasing too much. The critical depth does not exist anymore, and the sound interacts with the bottom.

For the locations that have a convergence zone in the future, there is no clear pattern for the increase or decrease in its length in the month of March. For September, there is an overall slight increase in the CZ length.

Fig. 4 shows the average CZ length (left) for all the considered data points as well as the total occurrence of a CZ (right) and it also shows these variables for only the datapoints at a latitude above or equal to 45 degrees. For the Atlantic Ocean as a whole, only a minor trend is visible where the average lengths seem to converge irrespective of the month. The opposite is happening for the higher latitude area, where the CZ lengths seem to diverge between the months; for March the lengths become shorter and for September the lengths become larger (Fig. 4, left panel).

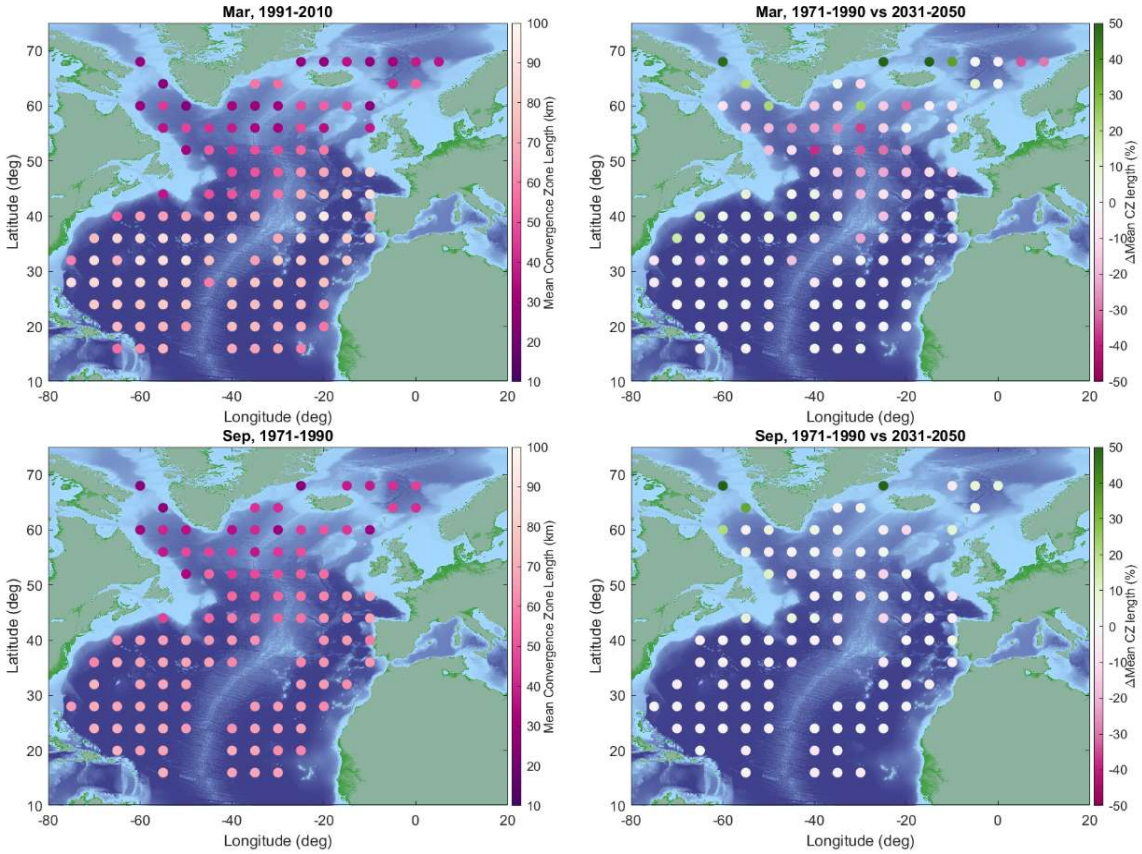


Fig.3: Map of the convergence zone lengths in the North Atlantic. Left: average from 1971 to 1990. Right: percentual variation for the time bracket 2031-2050 compared to 1971-1990. Top: March. Bottom: September.

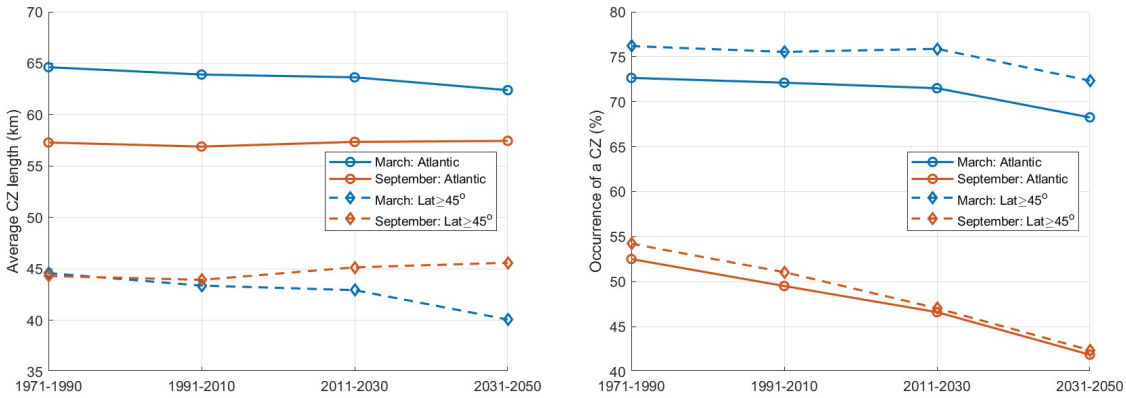


Fig.4: Left: Average of the Convergence Zones lengths over the Atlantic and for latitudes over 45 degrees. Error bars are relatively large and are not plotted here. Right: Occurrence (on a model-based level) of convergence zones in the Atlantic and for latitudes over 45 degrees.

It is also observed that the occurrence of convergence zones decreases over time, meaning that convergence zones will happen less often (Fig. 4, right panel). This phenomenon happens for both months and is not restricted to the higher latitudes. This is

attributed to a warming of the sea surface resulting in a critical depth occurring less frequently.

These results for the CZ lengths have implications for the use of historical databases. Databases such as the WOA are 20 years monthly and seasonal averages. Depending on what part of the Atlantic Ocean one is interested in, the validity of such a database may no longer apply in modern or future scenarios. This is especially true for the higher latitudes.

5. CONCLUSIONS

This study shows how it is possible to use temperature and salinity profiles from climate models to compute indicators of acoustic propagation, such as the sound channel axis depth, sonic layer depth, and convergence zone lengths.

Analyses based on three climate models show that the sound-channel axis depth varies mostly for latitudes higher than 45 degrees between the period of 1971-1990 and 2031-2050. This is because the sound channel axis is generally shallower at these latitudes, and the heat still penetrates to these depths.

The sonic layer depth changes across the whole Atlantic with a general decrease in its depth in March, which has implications for acoustic propagation in the surface duct.

Finally, the convergence zones were analysed. The effects of climate change for the month of March are stronger than the month of September. In September there is a marginal increase in the convergence zone length between the two periods, whereas in March no general statements could be made for the Atlantic. Analysis showed that for latitudes above 45 degrees there is a trend in the length of the convergence zone, where it shortens in March and increases in September. If the length variation differs depending on the month, the occurrence of convergence zones decreases over time, for both March and September, due to increasing sea surface temperatures.

The implications for historical databases depend on the application. For latitudes below 45 degrees, historical databases can still be used to compute the sound channel axis. Concerning the latitudes above 45 degrees, for the sonic layer depth or the convergence zone length, one will make errors if uses historical database too outdated.

6. ACKNOWLEDGEMENTS

The authors acknowledge the support and funding from the Netherlands Ministry of defence, from the NATO Office of the Chief Scientist and of the Richard Lounsbery Foundation.

REFERENCES

[1] **H. Pörtner, D. Roberts, V. Masson-Delmotte, P. Zhai, E. Poloczanska, K. Mintenbeck, M. Tignor, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama and N. Weyer**, *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate - Tehnical Summary*, IPCC, 2019.

[2] **L. Brekhovskikh and Y. Lysanov**, *Fundamentas of Ocean Acoustics*, Moscow: AIP Press, Springer, 2003.

[3] **M. Ainslie**, *Principles of Sonar Performance Modeling*, Heidelberg: Springer, 2010

[4] **F. Jensen, W. Kuperman, M. Porter and H. Schmidt**, *Computational Ocean Acoustics*, Springer, 2011.
[5] **J. Reagan, T. Boyer, H. García, R. Locarnini, O. Baranova, C. Bouchard, S. Cross, A. Mishonov, C. Paver, D. Seidov, Z. Wang and D. Dukhovskoy**, *World Ocean Atlas 2023*, NOAA National Centers for Enviornmental Information, 2024.
[6] **L. Possenti, G. d. N. L. Reichart, F. Lam, C. de Jong, M. Colin, B. Binnerts, A. Boot and A. von der Heydt**, *Predicting the contribution of climate change on North Atlantic underwater sound propagation*, PeerJ, pp. 1-26, 2023.
[7] **K. Pelekanakis, A. Kondi, A. Russo and S. Carniel**, *Simulation of Climate Change Impact on Underwater Acoustic Transmission Loss*, in IEEE/MTS OCEANS, Singapore, 2024.
[8] **V.O. Oppeneer, N. Risser, R. van Vossen, K. Pelekanakis, G. Giroli and A. Russo**, *The Impact of Climate Change on Sonar Performance: A case study in the high north*, in ICUA, Bath, UK, 2024.
[9] **R. Helber, C. Barron, M. Carnes and R. Zingarelli**, *Evaluating the sonic layer depth relative to the mixed layer depth*, Journal of Geophysical Research: Oceans, 113 (C7), pp. 1-14, 2008.
[10] **P. Etter**, *Underwater Acoustic Modeling and Simulation*, Boca Raton, FL: CRC Press, 2018.
[11] **R. Urick**, *Principles of underwater sound*, 3rd ed., New York: McGraw-Hill, 1983.
[12] **R. Keywan et al**, *The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview*, Global Environmental Change, 42, pp. 153-168, 2017.
[13] **K. Mackenzie**, *Nine-term Equation for Sound Speed in the Oceans*, Journal of the Acoustical Society of America, 70, pp. 807-812, 1981
[14] **R. j. Haarsma et al**., *High resolution model intercomparison project (highresmip v1.0) for CMIP6*, Geoscientific Model Development, 9 (11), pp. 4185-4208, 2016.
[15] **V. Eyring et al**., *Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization*, Geoscientific Model Development, 9 (5), pp. 1937-1958, 2016.
[16] **A. Voldoire et al**, *Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1*, Journal of Advances in Modeling Earth Systems, 11, pp. 2177-2213, 2019.
[17] **R. Döscher et al**, *The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6*, Geosci. Model Dev., 15, pp. 2973-3020, 2022.
[18] **K. Williams et al**, *The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations*, Journal of Advances in Modeling Earth Systems, 10, pp. 357-380, 2017.
[19] **J. Abraham, M. Baringer, N. Bindoff, T. Boyer, L. Cheng, J. Church and J. Conroy**, *A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change*, Rev. Geophys, 51 (3), pp. 450-483, 2013.

J-divergence detection currency after thresholding a Rician signal in Gaussian noise

Douglas A. Abraham

Applied Physics Laboratory
University of Washington

Douglas A. Abraham
Applied Physics Laboratory
University of Washington
1013 NE 40th Street
Box 355640
Seattle, WA, 98105-6698
USA
abrahad@uw.edu

Abstract: *Sonar performance modeling traditionally entails the use of single-measurement probabilities of detection and false alarm as metrics. Extension to systems utilizing multiple measurements is often complicated enough that the scenarios are simplified to fit the single-measurement tools, thereby approximating system-level performance. By allowing an approximate representation of single-measurement detection performance in the form of J-divergence detection currency (JDC), a more accurate assessment of system-level performance can be obtained across multiple measurements. As a measure of detection performance potential, JDC can be evaluated at different points in the signal and information processing chain. After coherent detection processing, a common step in multiple-measurement systems is the thresholding of individual measurements before they are combined in a final detection decision. For example, this typically occurs when combining measurements across waveforms in active sonar or across sensors in distributed systems. The focus of this paper is on techniques for evaluating JDC after thresholding for an integrated-intensity detector operating on the standard sonar signal models (Gaussian-fluctuating, deterministic, and the more general Rician model) in Gaussian noise.*

Keywords: *detection performance, J-divergence, Rician signal*

1. INTRODUCTION AND BACKGROUND

Traditional techniques for modeling sonar performance typically exploit the sonar equation to obtain the signal-to-noise power ratio (SNR) and then relate SNR to single-measurement probabilities of detection and false alarm (i.e., P_d and P_f). Applications employing multiple measurements are often simplified to fit within the context of the single-measurement tools, potentially leading to inaccurate predictions of system performance. An alternative approach using the J-divergence as a general representation of *detection currency* was presented in [1]. At the expense of an approximate relationship to the detection probabilities, multiple independent measurements are easily incorporated by summing their J-divergences. Owing to a data processing inequality (processing of data can only maintain or decrease J-divergence, but not increase it), the technique allows evaluation at different points in the signal and information processing chain, with an interpretation as representing detection performance potential. The focus of this paper is on carrying the analysis of J-divergence through the thresholding process commonly found in multiple-measurement systems. For example, echoes from multiple waveforms in active sensing or measurements from different sensors in a distributed system typically require the individual measurements to exceed a local threshold before they are combined to form a final detection decision.

The J-divergence (J) is a one-to-one non-linear mapping from the SNR obtained through the sonar equation to a scalar measure of detection performance. For example, from [1], an instantaneous intensity sample containing a Gaussian-fluctuating signal in Gaussian noise results in $J = s^2/(1+s)$ where s [unitless] is the linear-quantity SNR. This implies detection performance is quadratically related to SNR when $s \ll 1$ and tends to be linear when $s \gg 1$. Accounting for thresholding has little effect when the SNR is high enough for P_d to be close to one, but will significantly reduce J when the individual measurement is unlikely to be detected. Carrying this effect into the analysis of multiple-measurement detection systems produces a more accurate prediction of system-level performance, particularly at low SNR.

The primary contributions of this paper are techniques for the evaluation of J-divergence after thresholding for an integrated-intensity detector (e.g., after quadrature matched filtering or in an energy detector) observing one of the standard sonar signal models (Gaussian-fluctuating, deterministic, or Rician) in Gaussian noise. The J-divergence after thresholding is introduced and examined using the Gaussian-fluctuating signal in Sect. 2 and related to the Kullback-Leibler divergences of the excess over the threshold. Extending the results to the Rician signal model (which contains the deterministic and Gaussian-fluctuating signals) requires an approximation. In Sect. 3, the integrated-intensity decision statistic is modeled by a generalized gamma (GG) distribution with the parameters obtained by matching moments with the Rician model. To avoid simultaneous solution of the moment equations, an approximation to the GG power parameter is presented, which simplifies the problem to the solution of a single non-linear moment equation. The J-divergence after thresholding is then derived for GG distributions in Sect. 4 and used in Sect. 5 to demonstrate that the approximation error for the Rician model is adequately low over a wide range of parameters.

2. J-DIVERGENCE AFTER THRESHOLDING

The J-divergence [2] is a distributional distance measure useful in representing detection performance when applied to the probability density functions (PDFs) of a decision statistic

under the noise-only (H_0) and signal-present (H_1) hypotheses. If T is the detector decision statistic and $f_0(t)$ and $f_1(t)$ are the PDFs under the two hypotheses, the J-divergence is

$$J = \int_{-\infty}^{\infty} [f_1(t) - f_0(t)] \log\left[\frac{f_1(t)}{f_0(t)}\right] \, \mathrm{d}t. \tag{1}$$

As described in [1], this can be converted to a logarithmic quantity, termed the J-divergence detection currency (JDC), through JDC $= 5\log_{10}(J)$ [units: dB], as if it were a ratio of squared intensities. The J-divergence is combined over multiple independent measurements by summing the individual linear quantities, with the result representative of an optimal processor (e.g., using a likelihood-ratio detector).

In many multiple-measurement applications, individual measurements are thresholded prior to being combined. By treating the thresholding process as producing either a null state (threshold not exceeded) or providing a measurement representative of the threshold exceedance, the J-divergence after thresholding can be shown to be

$$J = [P_d - P_f] \log\left[\frac{1 - P_f}{1 - P_d}\right] + \int_{h}^{\infty} [f_1(t) - f_0(t)] \log\left[\frac{f_1(t)}{f_0(t)}\right] \, \mathrm{d}t, \tag{2}$$

where $P_d = \Pr\{T \geq h | H_1\}$, $P_f = \Pr\{T \geq h | H_0\}$, and h is the local threshold. This form is useful for analytical evaluation when the data PDFs and the log likelihood ratio combine in the integrand to result in simple forms. It is also a convenient form for numerical evaluation.

As an example, consider a decision statistic formed by integrating (summing) M independent and identically distributed intensities, as might occur in an energy detector in passive sensing or after post-matched-filter integration in active sensing. If each intensity contains a Gaussian-fluctuating signal in Gaussian noise with a linear-quantity SNR of s [unitless], the decision statistic is gamma distributed with shape parameter M and scale parameter $1 + s$. Thresholding the decision statistic at h yields J-divergence

$$\begin{aligned} J = {} & [P_d - P_f] \log\left[\frac{1 - P_f}{(1 - P_d)(1 + s)^M}\right] + Ms\left[P_d - \frac{P_f}{1 + s}\right] \\ & + \frac{h^M e^{-h}}{\Gamma(M)} \frac{s}{(1 + s)} \left[\frac{e^{hs/(1+s)}}{(1 + s)^{M-1}} - 1\right], \end{aligned} \tag{3}$$

where $P_f = 1 - \tilde{\gamma}(h; M)$, $P_d = 1 - \tilde{\gamma}(h/[1 + s]; M)$, and $\tilde{\gamma}(t; \alpha)$ is the incomplete gamma function normalized by $\Gamma(\alpha)$ [4, pg. 293]. Removing the thresholding by letting $h \to 0$ so P_d and $P_f \to 1$ yields $J = Ms^2/(1 + s)$ from the second term in (3), which produces the result found in [1, eq. 6] when $M = 1$ (representing a single intensity sample).

The simplicity of this result makes it quite useful. However, signals exhibiting greater consistency from observation to observation are better represented by a deterministic or Rician signal. Similar to evaluating J-divergence before thresholding, the Bessel functions found in the intensity PDFs of these signals in Gaussian noise complicate a direct solution. However, as presented in Sect. 3, an accurate approximation that is straightforward to evaluate can be obtained using a generalized gamma distribution.

2.1. RELATIONSHIP TO KULLBACK-LEIBLER DIVERGENCES

For some PDFs, the distribution of the excess over the threshold, $Z = T - h$ given T exceeds h can be described simply. For example, when T is exponentially distributed with

mean λ, $Z = T - h$ is also exponentially distributed with mean λ when conditioning on $T \geq h$. This can facilitate evaluation of the J-divergence after thresholding in (2) by describing the integral using the Kullback-Leibler divergences (KLDs) of Z in conjunction with P_f and P_d.

First note that the PDF of Z (given $T \geq h$) under H_0 and H_1 is easily described by the PDF of the unthresholded data and the appropriate exceedance probability,

$$\tilde{f}_0(z) = f_0(z+h)/P_f \text{ and } \tilde{f}_1(z) = f_1(z+h)/P_d \text{ for } z \geq 0. \tag{4}$$

This allows rephrasing the integral in (2) as $[P_d - P_f]\log[P_d/P_f] + P_d\tilde{I}_{1:0} + P_f\tilde{I}_{0:1}$, where $\tilde{I}_{i:j} = \int_{-\infty}^{\infty} \tilde{f}_i(z)\log\left[\tilde{f}_i(z)/\tilde{f}_j(z)\right]\mathrm{d}z$ is the KLD [3] formed using the distributions of the threshold exceedance. The J-divergence after thresholding,

$$J = [P_d - P_f]\log\left[\frac{P_d[1-P_f]}{P_f[1-P_d]}\right] + P_d\tilde{I}_{1:0} + P_f\tilde{I}_{0:1}, \tag{5}$$

can then be characterized as comprising a binary component and contributions accounting for the excess over the threshold after moderation by the probability the event occurs.

3. APPROXIMATING A RICIAN INTEGRATED INTENSITY WITH A GENERALIZED GAMMA DISTRIBUTION

The most common signal models employed in sonar analysis are the deterministic and Gaussian-fluctuating signals. Spanning the two is the Rician signal [4, Sect. 7.5.3], for which the integrated instantaneous intensity (T) is proportional to a non-central-chi-squared-distributed random variable with $2M$ degrees of freedom,

$$\frac{2}{1+\rho s}T \sim \chi^2_{2M,\delta} \tag{6}$$

where M and s are as described in Sect. 2, $\rho \in [0,1]$ is the ratio of the Gaussian-random signal power to the total signal power, and $\delta = 2M(1-\rho)s/(1+\rho s)$ is the non-centrality parameter. Setting $\rho = 0$ produces a deterministic signal and a Gaussian-fluctuating signal is obtained when $\rho = 1$ (with simplification to the gamma distribution described in Sect. 2).

The PDF of the non-central-chi-squared distribution contains a Bessel function that complicates a direct analysis of the J-divergence. For the purposes of evaluating J-divergence after thresholding, the Rician signal model can be approximated with a generalized gamma (GG) distribution. If X is gamma distributed with shape α and unit scale, then $T = \beta X^{1/c}$ follows the GG distribution and has PDF

$$f_T(t) = \frac{ct^{c\alpha-1}e^{-(t/\beta)^c}}{\Gamma(\alpha)\beta^{c\alpha}} \text{ for } t \geq 0,\ \alpha > 0,\ \beta > 0, \text{ and } c > 0. \tag{7}$$

When the GG power parameter $c = 1$, T is gamma distributed with shape parameter α and scale parameter β. The Weibull distribution can also be seen to arise when $\alpha = 1$.

The difficulty in exploiting the GG model lies in obtaining the parameters (α, β, c) when a signal is present. Choosing the parameters that match the first three moments of T (i.e., $\mu'_k = E[T^k]$ for $k = 1$, 2, & 3) between the Rician and GG models results in the moment

equations

$$\mu_1' = M(1+s) = \frac{\beta\,\Gamma(\alpha+1/c)}{\Gamma(\alpha)}, \tag{8}$$

$$\mu_2' = M^2(1+s)^2 + M(1+\rho s)(1-\rho s+2s) = \frac{\beta^2\,\Gamma(\alpha+2/c)}{\Gamma(\alpha)}, \text{ and} \tag{9}$$

$$\mu_3' = M(M+1)(M+2)(1+\rho s)^3\left[1+3d+\frac{3Md^2}{M+1}+\frac{M^2d^3}{(M+1)(M+2)}\right] = \frac{\beta^3\,\Gamma(\alpha+3/c)}{\Gamma(\alpha)},$$

where $d = (1-\rho)s/(1+\rho s)$. These equations need to be solved for α, β, and c as functions of ρ, s, and M. For a Gaussian-fluctuating signal in Gaussian noise (i.e., when $\rho = 1$), the decision statistic is gamma distributed so $c = 1$, $\alpha = M$, and $\beta = 1 + s$. Setting $s = 0$ then produces the parameters under H_0 (i.e., β reduces to one). However, with both α and c trapped inside gamma functions, the solution for signals with $\rho < 1$ must be obtained iteratively. As this is prohibitive in many applications, an approximation only requiring the solution of a single non-linear equation is presented in Sect. 3.1.

Once the parameters of the GG model are obtained under the two hypotheses, the J-divergence after thresholding can be evaluated as described in Sect. 4.

3.1. APPROXIMATE MOMENT MATCHING

After evaluating c though precise moment matching for a wide range of (ρ, s, M), it was found that using the following approximation was adequate for evaluating J-divergence after thresholding,

$$c \approx 1 + \exp\left\{-0.5 - \exp\left\{-1.3 - \frac{1}{2M} + \rho\left(2.8 + \frac{0.6}{M}\right)\right\}\right\}\Phi(S_{\text{dB}}/5)\,, \tag{10}$$

where $\Phi(z)$ is the standard normal cumulative distribution function. The SNR used in (10) is the logarithmic quantity, $S_{\text{db}} = 10\log_{10}(s)$ [units: dB].

Given c, the moment ratio $\mu_2'/[\mu_1']^2$ is straightforward to solve numerically for α. When a root-finding routine is not available, a Newton-Raphson iteration can be implemented as follows. Let $A = \log(\alpha)$ so $\alpha = e^A$ and initialize A to $\log M$. The solution (in terms of A) can then be obtained by the iteration

$$A \leftarrow A - \frac{\log\left(\frac{\mu_2'}{[\mu_1']^2}\right) + 2\log\Gamma\left(e^A+\frac{1}{c}\right) - \log\Gamma\left(e^A+\frac{2}{c}\right) - \log\Gamma\left(e^A\right)}{e^A\left[2\psi\left(e^A+\frac{1}{c}\right) - \psi\left(e^A+\frac{2}{c}\right) - \psi(e^A)\right]}, \tag{11}$$

where $\psi(\cdot)$ is the digamma function and $\mu_2'/[\mu_1']^2 = 1+(1+\rho s)(1-\rho s+2s)/[M(1+s)^2]$ for the Rician model. The number of iterations required to achieve a specified error tolerance was seen to be decreasing in M and ρ and increasing in SNR (i.e., the combination of $M = 1$, $\rho = 0$ and high SNR requires the most iterations). Except at the highest SNRs (e.g., total SNR in excess of 30 dB), fewer than ten iterations were necessary to drive the relative error below 10^{-6}. Given α and c, the scale parameter β is obtained from (8) through $\beta = M(1+s)\Gamma(\alpha)/\Gamma(\alpha+1/c)$.

4. J-DIVERGENCE OF THRESHOLDED GENERALIZED GAMMA DISTRIBUTIONS

The J-divergence after thresholding can be obtained using the general form described in (2). Using the GG PDFs under H_0 and H_1 results in terms that can be described as partial moments and partial log moments. The upper partial pth-power moment for the GG distribution is

$$\int_h^\infty t^p f_T(t)\,\mathrm{d}t = \beta^p \int_{h'}^\infty \frac{z^{\frac{p}{c}+\alpha-1}e^{-z}}{\Gamma(\alpha)}\,\mathrm{d}z = \beta^p \Gamma(\alpha + p/c, h')/\Gamma(\alpha) \tag{12}$$

where $h' = (h/\beta)^c$ and $\Gamma(\alpha, x) = \int_x^\infty z^{\alpha-1}e^{-z}\,\mathrm{d}z$ is the complementary incomplete gamma function. The upper partial log moment of the GG distribution can be simplified to

$$\int_h^\infty \log(t) f_T(t)\,\mathrm{d}t = \log(\beta)[1 - F_T(h)] + \frac{g(h';\alpha)}{c}, \tag{13}$$

where

$$g(h;\alpha) = \int_h^\infty \frac{\log(z) z^{\alpha-1}e^{-z}}{\Gamma(\alpha)}\,\mathrm{d}z \tag{14}$$

is the upper partial log moment of a gamma distribution with shape parameter α and unit scale. As shown in Sect. 4.1, this can be evaluated by truncating an infinite summation.

Before proceeding to the J-divergence after thresholding, define the normalized complementary incomplete gamma function as $\tilde{\Gamma}(z;\alpha) = \Gamma(\alpha, z)/\Gamma(\alpha) = 1 - \tilde{\gamma}(z;\alpha)$, noting that it is simply one minus the normalized incomplete gamma function. Using the transformed threshold $h_i' = [h/\beta_i]^{c_i}$, for $i = 0$ and 1, then allows describing the single-measurement detection probabilities as

$$P_d = \tilde{\Gamma}(h_1';\alpha_1) \quad \text{and} \quad P_f = \tilde{\Gamma}(h_0';\alpha_0)\,, \tag{15}$$

where the subscript on each parameter defines the hypothesis under consideration. The J-divergence after thresholding for the GG model can then be described as

$$\begin{aligned} J =& (P_d - P_f)\log\left[\frac{c_1(1-P_f)\Gamma(\alpha_0)}{c_0(1-P_d)\Gamma(\alpha_1)}\right] + (c_0\alpha_0 P_d - c_1\alpha_1 P_f)\log(\beta_0/\beta_1) \\ &- \alpha_1\tilde{\Gamma}(h_1';\alpha_1+1) - \alpha_0\tilde{\Gamma}(h_0';\alpha_0+1) + (c_1\alpha_1 - c_0\alpha_0)\left[\frac{g(h_1';\alpha_1)}{c_1} - \frac{g(h_0';\alpha_0)}{c_0}\right] \\ &+ \tilde{\Gamma}(h_1';\alpha_1 + c_0/c_1)\,\frac{\Gamma(\alpha_1 + c_0/c_1)}{\Gamma(\alpha_1)}\left[\frac{\beta_1}{\beta_0}\right]^{c_0} + \tilde{\Gamma}(h_0';\alpha_0 + c_1/c_0)\,\frac{\Gamma(\alpha_0 + c_1/c_0)}{\Gamma(\alpha_0)}\left[\frac{\beta_0}{\beta_1}\right]^{c_1}. \end{aligned} \tag{16}$$

Despite the length of this result, each term is straightforward to evaluate.

4.1. NUMERICAL EVALUATION OF THE GAMMA PARTIAL LOG MOMENT

The upper partial log moment of the gamma distribution from (14), $g(h;\alpha)$, is required to evaluate the J-divergence after thresholding of the gamma, Weibull, and generalized gamma distributions. A numerical solution in the form of a truncated infinite summation can be obtained by noting that $\log(z)z^\alpha$ is equal to the derivative of $z^\alpha = e^{\alpha\log z}$ with respect to α. This allows describing $g(h;\alpha)$ according to

$$g(h;\alpha) = \frac{1}{\Gamma(\alpha)}\left[\frac{\partial}{\partial\alpha}\int_h^\infty z^{\alpha-1}e^{-z}\,\mathrm{d}z\right] = \frac{1}{\Gamma(\alpha)}\frac{\partial\Gamma(\alpha,h)}{\partial\alpha}, \tag{17}$$

which is one over $\Gamma(\alpha)$ times the derivative of the complementary incomplete gamma function with respect to its order. Taking the partial derivative of the series expansion of $\Gamma(\alpha, h)$ found in [5, pg. 178, eq. 8.7.3] with respect to α and using it in (17) results in

$$g(h;\alpha) = \psi(\alpha)[1 - \tilde{\gamma}(h;\alpha)] - h^{\alpha}e^{-h}\sum_{k=0}^{\infty}\frac{h^k[\log(h) - \psi(\alpha+k+1)]}{\Gamma(\alpha+k+1)}, \tag{18}$$

where $\psi(\cdot)$ is the digamma function. When $h \to 0$, it is clear that the partial log-moment tends to the full log moment, which is simply $\psi(\alpha)$. Truncating the infinite summation at $K_{\max} = \lceil 5 + 9h^{0.6} \rceil$ was found to be sufficient for the cases considered here.

5. ERRORS IN APPROXIMATING J-DIVERGENCE FOR A RICIAN SIGNAL

The parameters defining the J-divergence after thresholding for a Rician signal model include: the SNR (s), the fraction of random signal power (ρ), the number of independent intensities being integrated (M) and the decision threshold (h). As described in Sect. 3.1, these are used to obtain GG parameters (α_1, β_1, c_1) when signal is present, with the noise-only parameters set to $\alpha_0 = M$, $\beta_0 = 1$, and $c_0 = 1$. The J-divergence after thresholding is then evaluated as described in Sect. 4. A precise value was obtained for comparison through numerical integration of (2).

The error in the logarithmic quantity (i.e., JDC $= 10\log_{10}(J)$ [units: dB]) was evaluated for values of SNR and decision threshold producing $P_f \in [10^{-16}, 0.1)$ and $P_d \in [0.1, 0.9999]$, with $\rho \in [0, 1]$ and $M \in [1, 100]$. The maximum absolute error is shown in Fig. 1 for $M = 1$ (solid lines) as a function of ρ and as a function of $1/M$ (dotted lines), with the maximum taken over the other parameters. Using approximate moment-matching from Sect. 3.1 in the GG model (reddish-brown lines in Fig. 1) provides essentially the same worst-case accuracy (< 0.1 dB) as using exact moments (blue lines), with significantly less computational effort. Although the Weibull model (gold lines) is better for a single intensity ($M = 1$) and nearly deterministic signal, it fails when integrating intensities (i.e., when $M > 1$).

6. CONCLUSIONS

Techniques for evaluating the J-divergence detection currency (JDC) after thresholding for the basic sonar signal models in Gaussian noise were presented for an integrated-intensity detector (e.g., after matched filtering or in an energy detector). Although the Gaussian-fluctuating signal yielded exact theoretical results, the more general Rician model required an approximation exploiting the generalized gamma (GG) distribution. This was facilitated by an empirical approximation to the GG power parameter to produce a straightforward evaluation of the other model parameters and, subsequently, the JDC after thresholding.

These results will be useful in modeling sonar performance in applications where multiple measurements are subjected to a thresholding operation prior to being combined in a final detection decision.

7. ACKNOWLEDGEMENTS

Sponsored by the Office of Naval Research; Code 32, Undersea Signal Processing.

REFERENCES

[1] D. A. Abraham. Modeling sonar performance using J-divergence. In *Proceedings of the Underwater Acoustics Conference & Exhibition*, Kalamata, Greece, 2023.

[2] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London, Series A*, 186:453–461, 1946.

[3] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[4] D. A. Abraham. *Underwater Acoustic Signal Processing: Modeling, Detection, and Estimation*. Springer, 2019.

[5] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

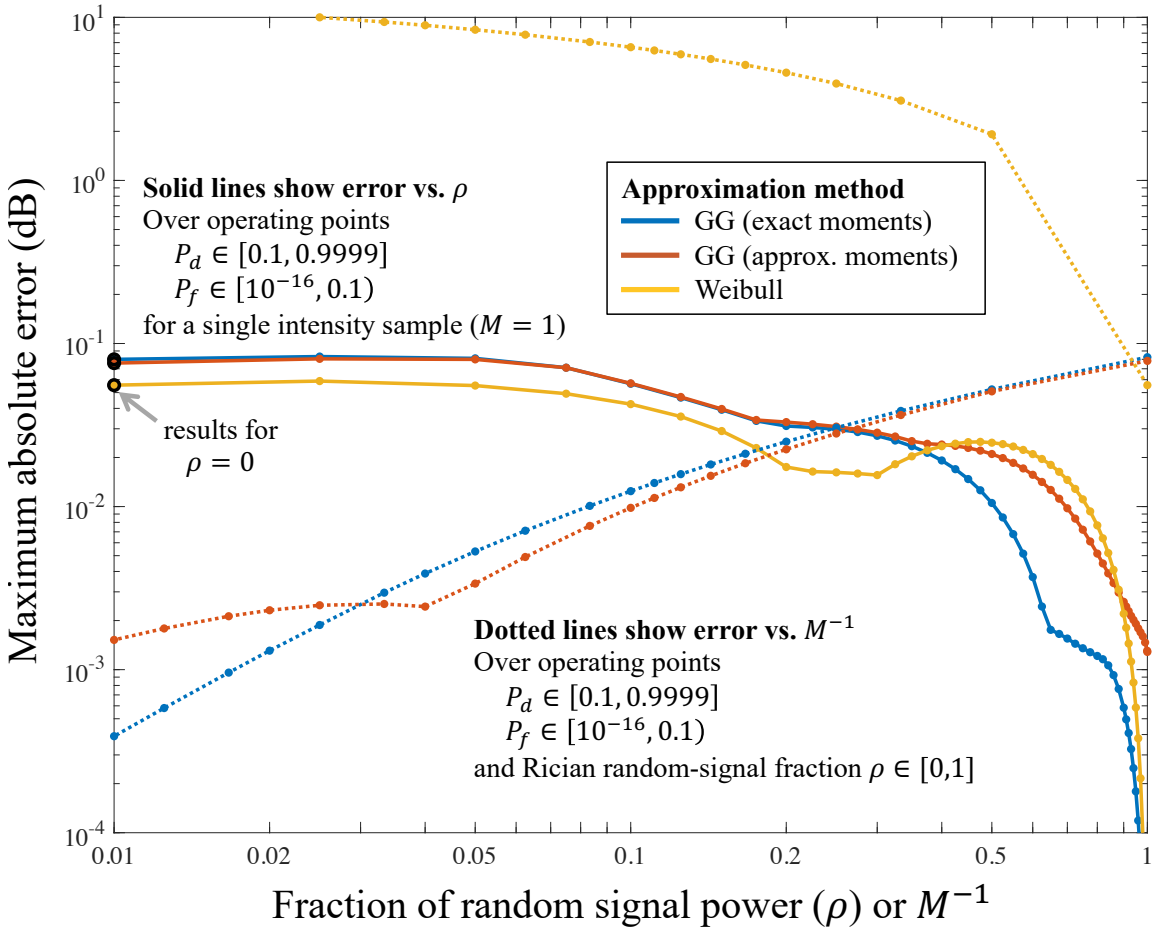


Figure 1: Error in JDC after thresholding for various approximation methods.

Structured Session
17. Target Echo Strength – Measurements and Modelling

Organised by David Nunn

Target echo strength of layered media with an active surface

Kevin Hostombe[1] and Delf Sachau[1]

[1]Helmut-Schmidt-University - Chair of Mechatronics, Holstenhofweg 85, 22043 Germany

Kevin Hostombe, Holstenhofweg 85, 22043 Hamburg, Germany, Phone: +49 40 6541 2611, hostombk@hsu-hh.de

Abstract: *Layered media, in form of both fluid layers and solid structures, play an important role in underwater acoustics. They occur as sonar windows, sonar reflectors, anechoic coatings, tank-linings or underwater structures and vehicles. However, the acoustically either desired total absorption, total reflection or total transmission is not achieved in practice. The acoustic behaviour of the layers play a particularly important role in the resulting target echo strength of an object. Active surfaces, for example made of piezo-ceramics, increase the possibilities of influencing the target echo strength beyond the geometry, material properties and material composition of the passive layers. This study compares the target strength behaviour of different layered media including active surfaces for underwater application. An analytical model is used to investigate the vibro-acoustic behaviour of the solid structures, the influence of the active surface and the resulting reflection and transmission behaviour. When the reflection of a surface is minimized by a single active layer, the transmission is not minimized and in some cases even increased. This is a particular problem in layered media, as the transmitted sound waves can cause subsequent layers and structures to vibrate and thus increase the total target echo strength. Various countermeasures are being investigated to prevent this.*

Keywords: *layered media, active surface, active layer, piezoelectric layer, reflection, transmission, analytical model, Echo Reduction, Transmission Loss*

1. INTRODUCTION

Layered media in underwater acoustics are often either chosen specifically for acoustic purposes or are unavoidable for other reasons. The layers are often desired to either totally reflect, transmit or absorb an incident acoustic wave (e.g. reflection: sonar reflectors; transmission: sonar-windows; absorption: anechoic coatings or tank linings). However, the total effect is almost never achieved in practical underwater applications, although achieving good reflection or transmission is much easier than preventing it, respectively achieving good absorption. Active surfaces, for example made of piezoceramics, increase the possibilities of influencing the reflection/transmission/absorption characteristics of layered media, and thus increase the possibilities of influencing the resulting Target Strength (TS) beyond the geometry, material properties and material composition of the passive layers. There have been many theoretical and practical approaches on investigating the effect of active surfaces / active coatings towards reflection, beginning with [1] over to [2], and [3], and most recently [4] and [5]. In most cases, the interaction between the incident wave and the active layer is investigated without considering possible other layers behind the active coating, as well as their significance for an optimal actuator and control design, and without considering oblique incidence. In this study, an approach to layered media with an active surface is shown, taking into account oblique sound incidence and its interaction with infinite solid elastic layers. Furthermore, different configurations of layered media behind the active layer and placement of the active layer are investigated. Finally, conclusions are drawn for the ideal configuration of the active layer and the control system to reduce Target Strength in layered media.

2. ACOUSTIC MODELING OF LAYERED MEDIA

A transfer matrix approach is used to model layered media. The systematic of order and numbering of the layers is based on [7] and shown in Figure 1 (a). Each layer i is assigned its specific acoustic properties depending on the material, as shown in Table 1. The transfer matrix model for investigating the influence of up to two active layers on reflection and transmission is adapted from [3] Equation (23). For modeling the elastic layer, different approaches can be found in [6] and the assumptions, simplifications and accuracy limits for the different models of plates and elastic layers are discussed there in detail. In order to get the best accuracy with focus on vibroacoustic, the approach from [6] Equation (27) is selected, which takes into account antisymmetric as well as symmetric motion. This makes it possible to take a further step away from idealized 1D transmission models and to investigate the influence of possible vibroacoustic effects on the performance of the active layer. For considering oblique sound incidence, the effective proportion of the wavenumber k and of the characteristic Impedance Z_c in normal direction are expressed as function of the incident angle θ_i in each layer i: $k_{eff,i} = cos(\theta_i)k_i = cos(\theta_i)\frac{2\pi f}{c_l}$, $Z_{c,eff,i} = \frac{Z_{c,i}}{cos(\theta_i)} = \frac{\rho_i c_{l,i}}{cos(\theta_i)}$, with the frequency f, longitudinal wave speed $c_{l,i}$ and density ρ_i. θ_i is calculated from Snell's law with $\theta_i k_i = \theta_{i+1} k_{i+1}$.

2.1. CONTROL ERRORS AND TARGET STRENGTH

When considering oblique sound incidence, the influence of the width of the individual piezoelectric actuators must also be taken into account. This is because for oblique sound in-

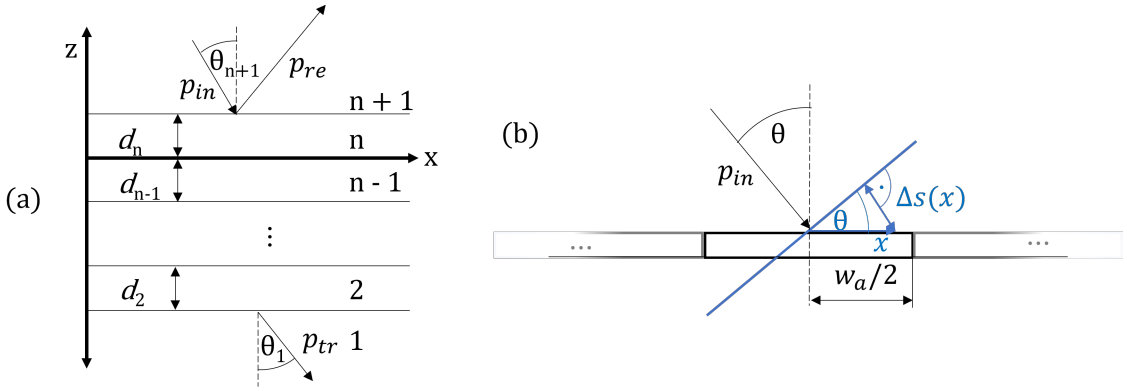


Figure 1: (a) Geometry of layered media, adapted from [7] (b) pressure error of a finite sized piezo strip for oblique incidence

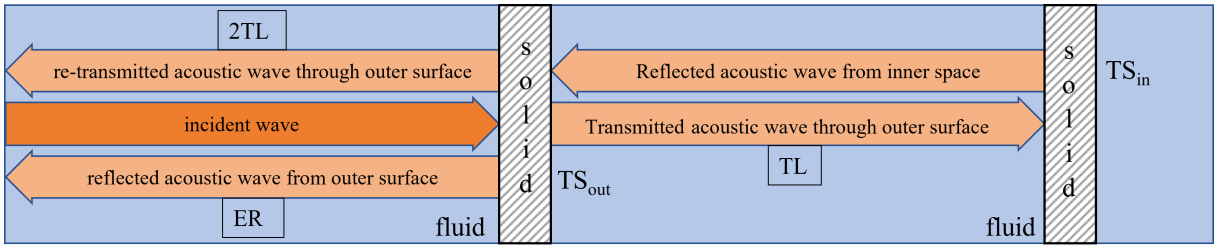


Figure 2: Acoustic waves incident, reflected and transmitted into an underwater structure and ER and TL, adapted from [8]

cidence, only the velocity of an infinitesimally wide part of the actuator corresponds perfectly to the velocity needed for influencing the incident sound wave in the desired way (for example $p_{re} = 0$). The difference between the remaining part of the actuator and the sound wave therefore leads to a deviation from the desired value of the reflection respectively transmission. Therefore, the actuator layer is assumed as finite sized strips with width w_a that are lined up horizontally next to each other, as shown in Figure 1 (b). The error $\epsilon(\theta, w_a) = \frac{\Delta p_{error}}{p_{in} - p_{re,passive}}$ for the control of the reflected and transmitted pressure for a finite sized actuator strip with width w_a under oblique incidence θ can be expressed as:

$$\epsilon(\theta, w_a) = (w_a - \int_{-\frac{w_a}{2}}^{\frac{w_a}{2}} cos(k\Delta s(x))dx)\frac{1}{w_a} = (1 - \frac{2sin(k\frac{w_a}{2}sin\theta)}{k\frac{w_a}{2}sin\theta}) \tag{1}$$

$$|R^*(\theta, w_a)| = |R(\theta)_{active}| + (|R(\theta)_{active}| - |R(\theta)_{passive}|) \cdot \epsilon(\theta, w_a) \tag{2}$$

$$|T^*(\theta, w_a)| = |T(\theta)_{active}| + (|T(\theta)_{active}| - |T(\theta)_{passiv}|) \cdot \epsilon(\theta, w_a) \tag{3}$$

For objects with an outer and an inner solid layer/structure, the acoustic wave that is transmitted from the outer to the inner structure will excitate the inner structure. This problem is described and modeled in [8] and shown in Figure 2. Emission Reduction (ER) and inner Transmission Loss (TL) can be calculated from R, respectively from T:

$$ER = 10\log|\frac{1}{R^2}|, TL = 10\log|\frac{1}{T^2}| \tag{4}$$

When reducing ER and TL of the outer structure, the total TS can be calculated as seen in [8], Eq. (35):

$$TS_{total} = 10\log(10^{\frac{TS_{out}-ER}{10}} + 10^{\frac{TS_{in}-2TL}{10}}) \tag{5}$$

material properties	Piezo (P)	Steel (S)	Water (W)	Air (A)
ρ in kg/m^3	7600	7850	1000	1.2
c_l in m/s	4108	5920	1500	343
g_{33} in F/m	24.535×10^{-3}	-	-	-
e_{33} in N/(Vm)	15.25	-	-	-
E in N/m^2	-	210×10^9	-	-
ν	-	0.3	-	-

Table 1: Material properties

3. MODELING RESULTS

To evaluate the influence of the passive layers on the controled system with one acutator, four different passive layer configurations are selected as Benchmarks. In a second step the influence of the actuator placement and the number of actuators is investigated by using three additional configurations. Figure 3 shows the sequence and dimensions of the layers of each configuration. The material properties used for each layer are listed in Table 1.

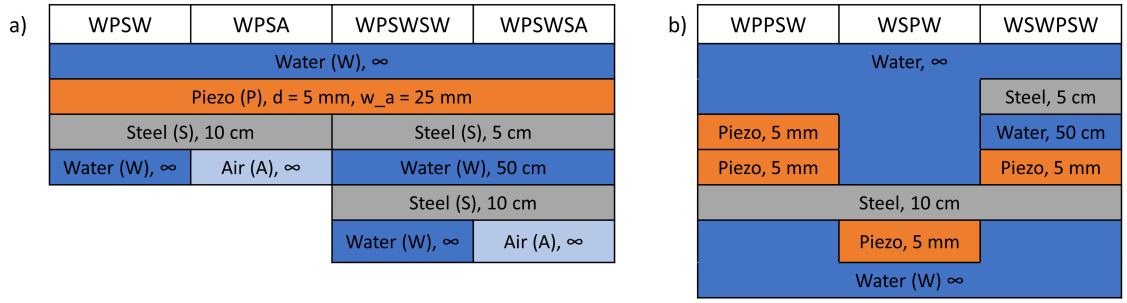


Figure 3: a) Configuration of passive layers with single actuator, b) Configuration of actuator placement and number

3.1. INFLUENCE OF PASSIVE CONFIGURATION

The corrected Echo Reduction ER* is shown for $0.1\,\text{kHz} \leq f \leq 20\,\text{kHz}$ and $0° \leq \theta_{n+1} \leq 85°$ in Figure 4.

There are two main acoustic effects that can be clearly identified especially for passive condition: The coincidence frequency for configurations WPSW and WPSA, as well as the $\lambda/2$ waveguide effect of the $0.5\,\text{m}$ thick water layer for WPSWSA and WPSWSW. The results for active control are shown within the same limits as in the passive case, but related to the Helmholtz number $k \cdot w_a/2$ instead of the frequency, as the error ϵ directly depends on the actuator width w_a. The displayed results are calculated for $w_a = 0.05\,\text{m}$. As the acoustic effects on ER^* can still be seen with control on, the influence of the angle θ and $k \cdot w_a/2$ can be seen in

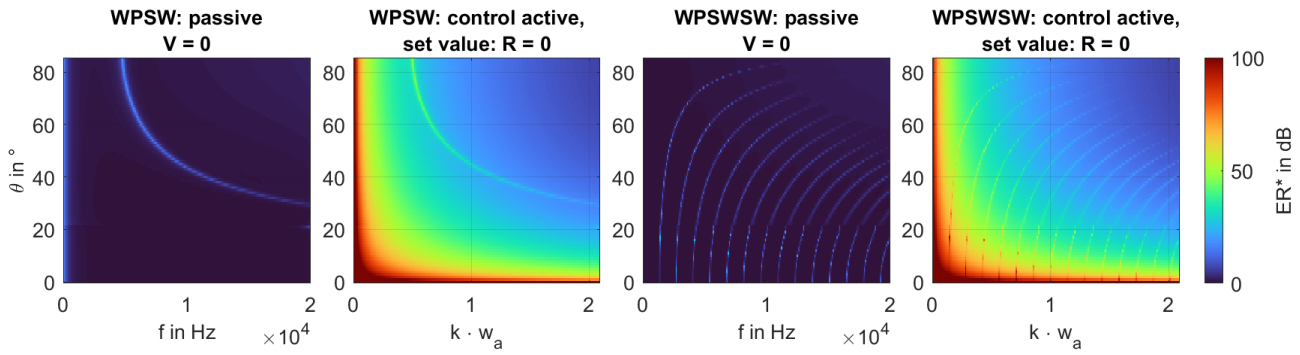


Figure 4: Comparison of acoustic effects of one and two steel layers (WPSW, WPSWSW) with control off and control on.

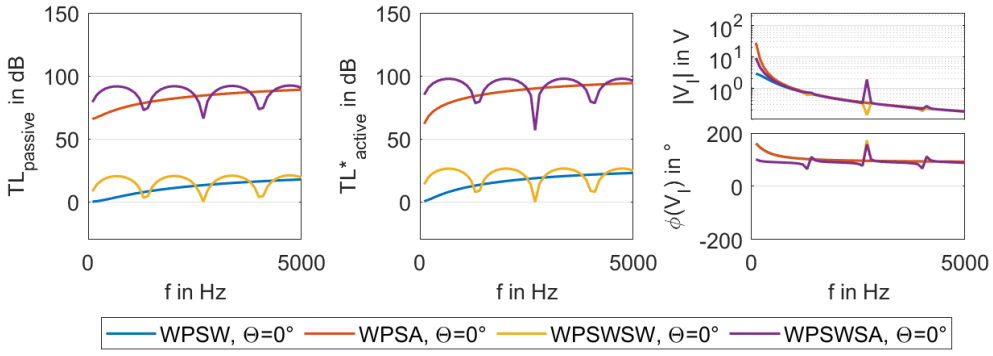


Figure 5: TL in active and passive mode and V_I for layer WPSW, WPSA, WPSWSW and WPSWSA.

the picture. As the Echo Reduction for most of the frequency band is 0 dB without control, the minimum gain in ER^* with active control is about 10 dB for the most unfavorable constellation of θ and f in the shown limits. The results for control on are based on the assumption of ideal linear actuator behavior and an ideal control system, in particular the performance and linearity limits of the actuator are not taken into account here, but the interaction of those limits with the overall reduction performance could be easily implemented in the future if the corresponding data is available.

Because active control is usually used in the lower frequency band, the upper frequency limit is decreased to 5 kHz. As Figure 4 already shows the dependency of acoustic effects from θ, for further comparison of the first four configurations, only values for $\theta = 0°$ will be shown. Comparing TL in passive and active condition as well as Voltage V_I in Figure 5, the $\lambda/2$ waveguide effects can be seen as well as the nearly soundhard boundary delivered by air. The interesting aspect is, that for active control the average of TL slightly increases in all configurations, but some local minima further decrease, which leads to more acoustic energy reaching the bottom layer. As the minima of TL WPSWSW and WPSWSA can also be observed as a peak in the Voltage, obviously more electric energy is needed to compensate the waveguide effect and thus leads to increased Transmission. Summing up the findings from the results shown in this plot, a single active layer on top of a multilayer configuration that controls ER will not lead to reduced TL. For system consisting of outer and inner structures, especially in the lower frequency range there is still much acoustic energy getting to the inner structure with its own acoustic response TS_{in}.

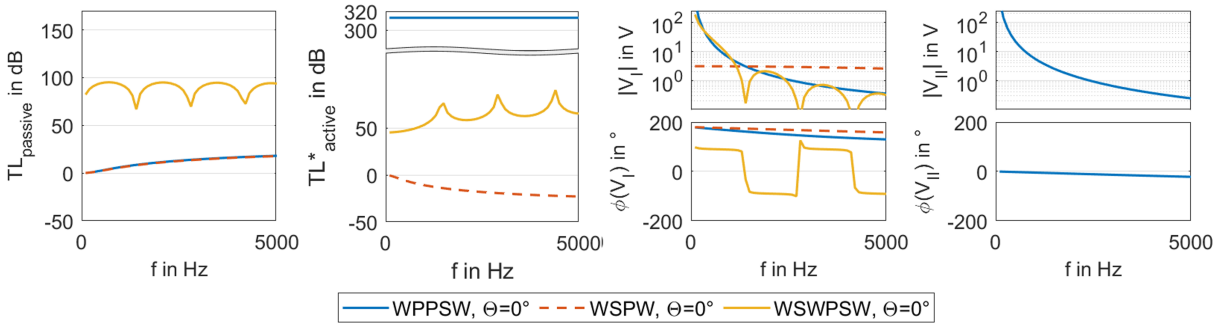


Figure 6: Influence of actuator placement and number (configurations from Figure 3 b))

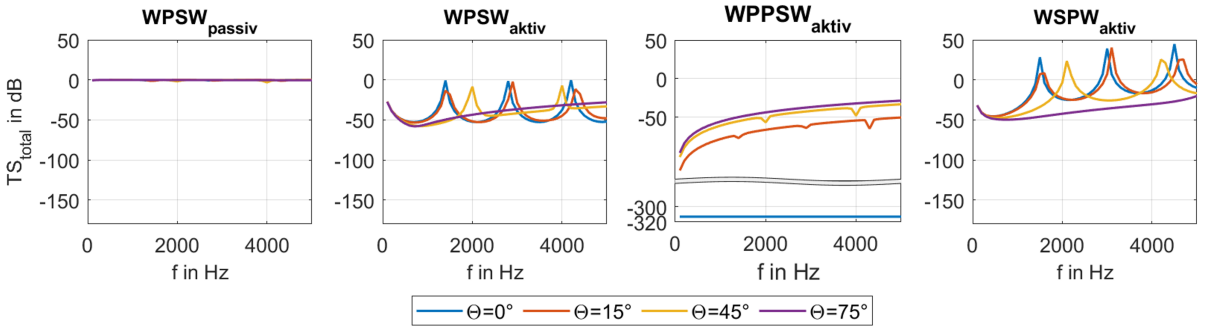


Figure 7: Comparison of TS_{total} for WPSW, WPPSW and WSPW with $TS_{out} = TS_{in} = 0$ dB

3.2. INFLUENCE OF ACTUATOR PLACEMENT AND NUMBER

For increasing TL, different measures are investigated by changing actuator placement and number, as shown in in Fig. 3 (b): First, using a double-layer actor, with the second layer being controlled for T=0 (WPPSW). Second, shifting the single active layer to another position inside the layered media (WSPW and WSWPSW). The results are shown in Figure 6. The results of TL^*_{active} show a very strong increase for the dual layer, while for the single layer with different positions it clearly decreases in comparison to the passive mode. The strong performance of the dual layer regarding TL comes at the cost of a much higher voltage (for each active layer) especially for $f < 2000$ Hz.

3.3. IMPACT ON TARGET STRENGTH

The impact of the resulting ER and TL on the total TS of a multilayer object can be evaluated using Eq. 5. Therefore, TS_{in} and TS_{out} are set to arbitrary fixed values, for simple evaluation and comparison 0 dB is chosen for both, and ER and TL are chosen from WPSW, WPPSW and WSPW as these configurations represent the outer layer. The results are shown in Figure 7. The passive layer with $ER \approx$TL≈ 0 dB shows a TS_{total} of nearly 0 dB, while the layer with the dual actuator shows the lowest TS, especially for normal incidence. WPSW in active mode shows an overall reduction of TS_{total}, in contrast configuration WSPW at the frequencies near the $\lambda/2$ waveguide effect target strength is increased, possibly by energy from the active system being added to the incidend acoustic energy.

4. CONCLUSION

In this study, an extended analytical approach for calculating transmission and reflection of layered media with an active coating has been used to investigate different layer configurations. For coincidence- and waveguide effects a clear influence on the frequency response of the required control voltage is determined. A single active layer can only be used to control either transmission or reflection, but not both at the same time. Because of that, for example the transmission of an incident wave will not be suppressed when actively reducing reflection and in some cases even increased. This leads to the question which measures can be found to reduce transmission and reflection at the same time. Therefore different positions of the actuator layer have been investigated and their effect on the total TS has been evaluated. The dual layer actuator shows by far the best results especially for $f < 2000\,\mathrm{Hz}$ and also with oblique incidence up to $85°$, although it needs much higher voltages compared to other configurations that were evaluated. The sometimes heavy phaseshifts and amplitude peaks of the ideal control voltage and with the necessary control voltages for low frequencies ask for a robust, adaptive system with low-frequency-optimized actuators to successfully increase ER or TL of plane, multilayered objects for oblique incidence.

REFERENCES

[1] Beatty, L. G.: “Acoustic Impedance in a Rigid-Walled Cylindrical Sound Channel Terminated at Both Ends with Active Transducers,”, *The Journal of the Acoustical Society of America* **36**, 1081-1089 **(1964)**.

[2] Howarth, T. R., Varadan, V. K., Bao, X. & Varadan, V. V.: “Piezocomposite coating for active underwater sound reduction”, *The Journal of the Acoustical Society of America* **91**, 823-831 **(1992)**.

[3] Cai, C., Liu, G. R. & Lam, K. Y.: “A Transfer Matrix Approach for Acoustic Analysis of a Multilayered Active Acoustic Coating”, *Journal of Sound and Vibration* **248**, 71-89 **(2001)**.

[4] Pyun, J. Y. et al.: “Design and Analysis of an Active Reflection Controller That Can Reduce Acoustic Signal Refer to the Angle of Incidence”, *Sensors* **21**, 5793 **(2021)**.

[5] Tang, J., Bai, Y., Yan, L. & Wang, W.: “GMA phased array for active echo control of underwater target”, *Applied Acoustics* **190**, 108646 **(2022)**.

[6] Arasan, U. et al.: “On the accuracy limits of plate theories for vibro-acoustic predictions”, *Journal of Sound and Vibration* **493**, 115848 **(2021)**.

[7] Folds, D. L. & Loggins, C. D.: “Transmission and reflection of ultrasonic waves in layered media”, *The Journal of the Acoustical Society of America* **62**, 1102?1109 **(1977)**.

[8] Kim, J. et al.: “Simplified Target Strength Analysis Procedure of an Underwater Vehicle Considering Target Strength Absorbing Materials”, *Journal of Marine Science and Engineering* **13**, 62 **(2025)**.

TES Modelling of TESSMEX objects M32 and M32T

Elizabeth Chilvers

QinetiQ Winfrith, Brownsea House, Dorset Innovation Park, Dorchester, DT2 8XJ.

Elizabeth Chilvers Tel: +44(0)1305 365436 Email: emchilvers@QinetiQ.com

Abstract: *Target Echo Strength (TES) measurements of TESSMEX object M32 were made over a wide range of frequencies at Loch Goil as part of International collaboration during March 2025, building on related measurements of M32T and M32C made in 2022. To contribute to the understanding of scattering from these objects, a number of modelling studies are being undertaken. The model predictions help identify features of interest in the measurements, aid analysis and provide mutual validation between the modelling and measurements. The models use several techniques to examine scatter features of the objects across the frequency spectrum. Ray Acoustics Modelling is applied at high frequencies to the acoustic scatter (including multiple bounces) from M32 enabling the identification of the causes of significant scatter features. Finite Element techniques tackle modelling M32 and M32T at lower frequencies, modelling not only absorption, transmission and reflection but also coupled vibration.*

Keywords: *Target Echo Strength, TESSMEX*

1. INTRODUCTION

The TESSMEX M32 object [7] is a physical 1:10 scale model of a structure originally defined in the Benchmark Target Strength Simulation (BeTSSi) mathematical modelling workshops [2][3][4], while M32T [7] is a truncated version of M32. To contribute to the understanding of scattering from these objects, a number of modelling studies have been undertaken, implemented within COMSOL Multiphysics® [1]. The results of these models are compared to measurements undertaken by QinetiQ in March 2022 and 2025.

Ray Acoustic modelling of M32 has been undertaken at high frequencies. This technique models sound as a large number of packets of sound energy that travel along rays and interact with surfaces in a probabilistic manner. The complex internal scattering within the objects may be modelled this way, but not diffraction effects nor vibrations associated with coupling into the structure.

Finite Element Modelling was used to model M32 and M32T at lower frequencies. This allows fuller consideration of the behaviour of the surfaces of the objects: modelling not only reflection, transmission and absorption, but also coupled vibration.

2. GEOMETRY

The M32 object is shown on the left of *Fig. 1*. The outer hull consists of a tapered cylinder with faceted hemi-spherical ends. The inner pressure hull consists of a cylinder with a "cat's-eye" inset sphere (highlighted in blue) at the bow and a "cross-plate" (highlighted in green) at the stern. While they were not part of the theoretical BeTSSi model 32, five support ribs (highlighted in pink) are needed to hold the inner pressure hull in place and a lifting eye (highlighted in orange) is required to suspend the object in the water. Reflections between the cylindrical inner pressure hull and these flat support ribs of M32 are expected to cause additional TES features not predicted for the BeTSSi model 32. The inner pressure hull is air-backed and constructed of 10mm thick steel whereas the water-backed outer-hull is made of 2mm steel.

The M32T object is similar to the stern of M32 as shown on the right of *Fig. 1*. Its outer hull consists of a tapered cylinder with a disk at the bow and a smooth hemi-spherical dome at the stern. The inner pressure hull of M32T is held in pace with two support ribs (highlighted in pink) and consists of a cylinder with a "cross-plate" (highlighted in green) at the stern and a corner inset (highlighted in blue) at the bow.

For the modelling in this paper the source insonifies the object from a fixed distance (22m) from the centre of the object: this was the arrangement of measurements of M32 and M32T undertaken in March 2025 and 2022 respectively at Loch Goil, Scotland. The objects are rotated about their respective centres with an aspect angle of 0° defined as "looking" at bow (the smaller end of M32 and the flat end of M32T). We assume that the object is being perfectly insonified by sound at a pure frequency.

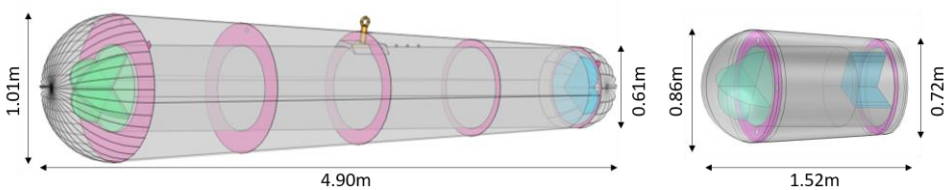


Fig. 1: M32 (left) and M32T (right) geometry.

3. RAY ACOUSTIC MODELLING

This technique models sound as a large number of packets of sound energy that travel along rays and interact with surfaces in a probabilistic manner. As described in [1] "Ray acoustics is valid in the high-frequency limit where the acoustic wavelength is much smaller than the characteristic geometric features." M32 is rotated about the central lifting eye and for each aspect angle considered, rays are released from the centre of a square receiver in a cone towards the object. When each individual ray interacts with the surface the probabilities that the rays are reflected, transmitted or absorbed is determined from the angle of incidence when the rays hit the surface, the properties of the material and its thickness, as well as the medium behind it. The rays that are scattered back towards the source will hit the receiver, where the strength of the acoustic impact of these rays is counted up and the times at which they occur recorded. Individual rays can also be tracked and their interaction with the object examined.

Range vs aspect "waterfall" plots can be created by considering the time at which the sound particles (rays) are captured by the receiver plate and the aspect of rotation of M32, about a vertical axis through its central lifting eye. Choosing finer aspect resolution, or releasing additional rays, produces more detailed waterfalls, but the run time of the models increases linearly. The limit on how a fine resolution can be produced is associated with computational time, not the methodology: investigations of specific areas of interest could be undertaken at a finer resolution.

Fig. 2 shows the Ray Acoustic predicted waterfall plot at 30kHz using one million rays. Many, but not all, of the dark (high value) lines running through the plots can be identified as reflections from the inner pressure hull and support ribs. Other notable features are from the inset sphere and cross-plate of the inner pressure hull.

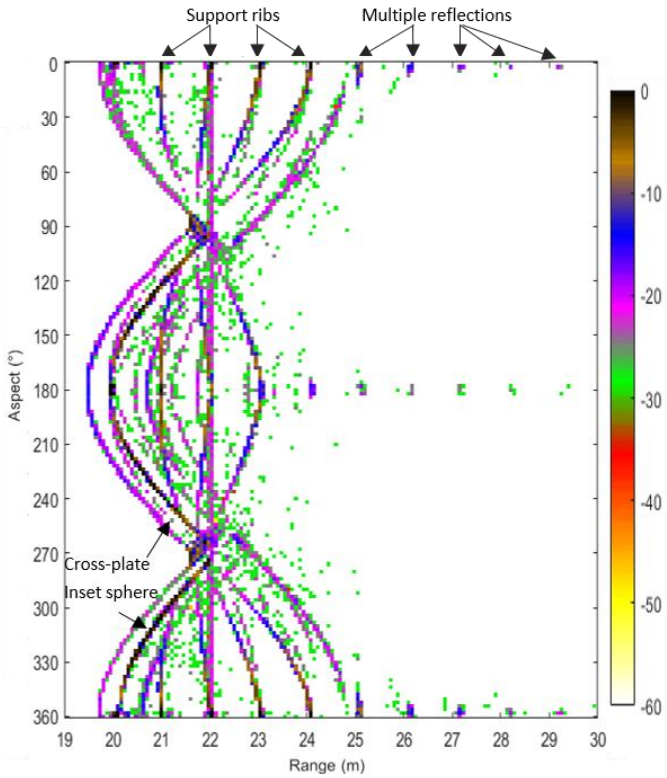


Fig. 2: Range vs aspect waterfall plots of M32 at 30kHz. Ray Acoustics Prediction.

Fig. 3 shows the waterfall plot produced from the measurement of M32 undertaken in March 2025 also at 30kHz for a short pulse. It should be noted that the measurements were not undertaken at exactly horizontal elevation and that in reality the domed ends of M32 are faceted, they are not the perfect hemi-spheres modelled. Ray Acoustics modelling requires significant run time to predict low TES levels therefore, for ease of comparison with the model output, the measurement on the right of *Fig. 3* has a lower threshold of approximately -30dB applied and contamination removed.

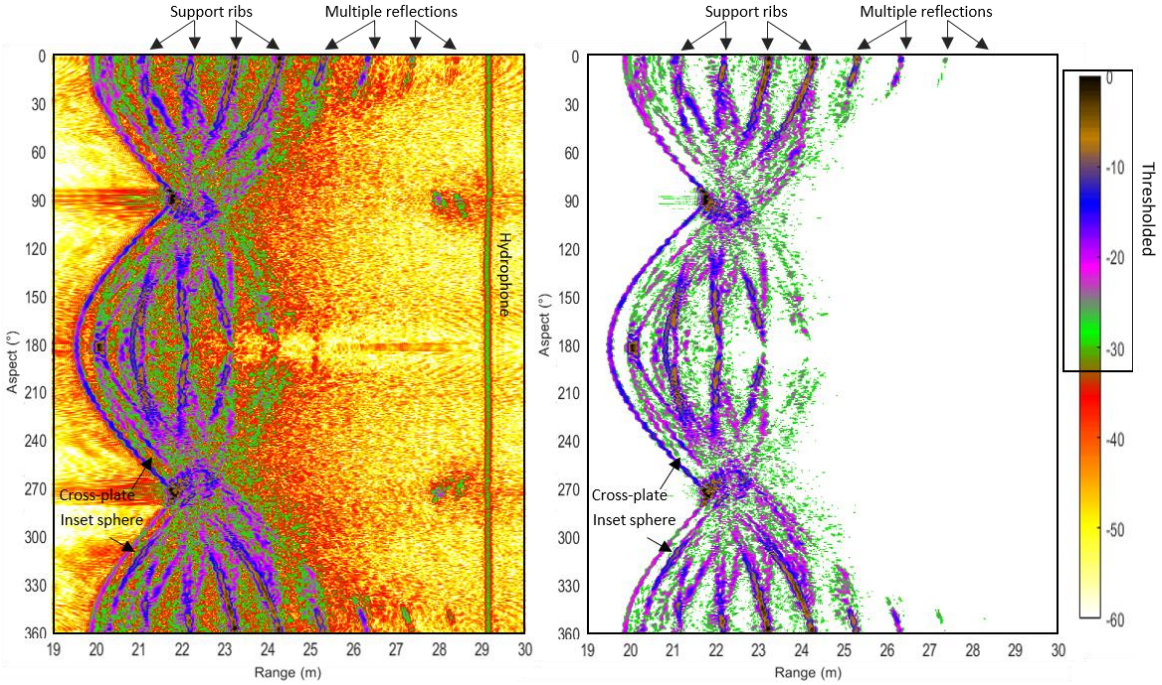


Fig. 3: Range vs aspect waterfall plots of M32 at 30kHz. Full Measurement (left) and with a threshold applied (right).

Most of the features are in the same place and a good comparison can be seen, but there are some notable differences. Features can be seen in the predicted waterfall around 0°/360° and 180° extending beyond the rest of the waterfall. These are caused by multiple reflections between the support ribs and show returns with an apparent range beyond the actual target. These multiple reflections are also seen in the measurement at 0°/360° but those at 180° are only seen at higher frequencies. The response of the inset sphere and cross-plate is also slightly higher in the model. These could be due to damping effects not yet modelled.

It is worth noting that in the preceding discussion TES has been treated as monostatic; with the source and receiver as points in the same location. However, the modelled results, like the measurements, will include small bistatic elements. For the Ray Acoustic model, rather than the prediction at a single point, rays that impact a square receiver are captured. The size of this receivers is important. A larger receiver is more sensitive as it captures more rays, however it is less monostatic. Yet by simply reducing the receiver size many of the low-level green and purple details are lost (there is effectively a low-level cut-off). Thus a larger number of initial rays must be modelled for the smaller receiver. For the measurements the source and receiver are different transducer arrays, so they also record TES with a small bistatic angle

4. FINITE ELEMENT MODELLING

This technique allows fuller consideration of the behaviour of the surfaces of the object: modelling not only reflection, transmission and absorption but also coupled vibration. The object is set inside of a sphere of water. The outer layer of this sphere simulates an infinite domain using a Perfectly Matched Layer. For this finite element model to be accurate the mesh of the object, as well as that of the surrounding water, must have elements no larger than one fifth of a wavelength. Thus, as the frequency increases, the computational demand increases dramatically.

The Sound Pressure Level (SPL) can be calculated and visualised as a colour over a horizontal slice through the modelled volume. *Fig. 4* shows this in the water viewed from above at 3kHz where M32 and M32T are insonified at an angle of 240° aspect as an example. In the dark blue "shadow zone" incident sound has been partially blocked by the body. Interference between the incident sound and that reflected from the object cause the alternating bands of SPL. The interference bands are also present within the water filled regions between the outer-hull and inner pressure hull of the object.

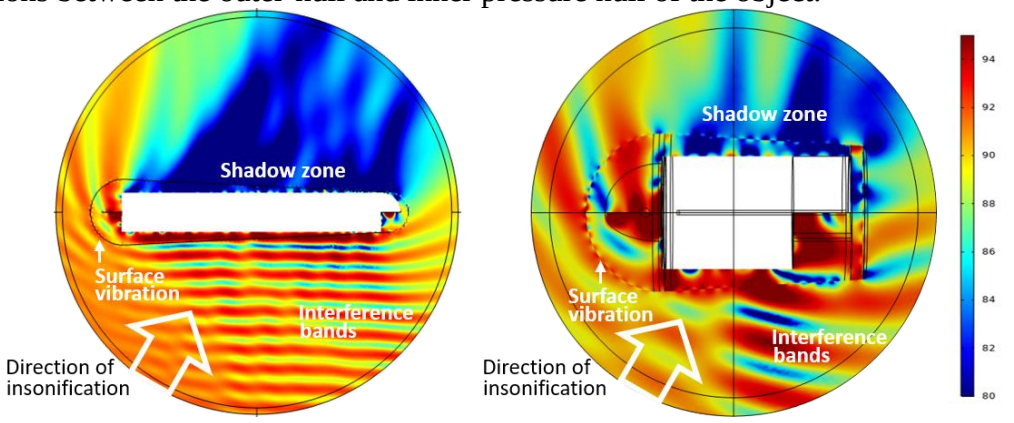


Fig. 4: FEM predicted sound pressure level in water at 3kHz. M32 (left), M32T (right). Note: M32 is set inside a larger sphere of water than the smaller M32T.

Finite Element Modelling (FEM) also predicts the coupled vibration of the steel inner and outer hulls of the objects. Oscillations in colour, linked to the surface vibration excited, can be seen in *Fig. 4* around the outer-hull of the objects. The sound vibrations in the inner flooded regions will also impact on the walls of the pressure hull, causing them to vibrate and set up sound vibration in the air filled volume, having its own SPL although this is markedly less than in the surrounding water.

The frequency domain Integrated Target Strength (ITS) predicted by FEM may be extrapolated from this near field prediction. The predicted ITS at 22m can be compared with that calculated from the measured response to insonification by a Linear Frequency Modulation (LFM) chirp and to a continuous wave (CW) pulse at the same distance.

Fig. 5 shows the FEM predicted ITS of M32T at 22m in black compared with the ITS from the measurements undertaken in 2022 at 3kHz and 5kHz. However it should be noted that the FEM predictions have assumed that the object is orientated perfectly; that it is not tilted, whereas the 2022 measurements contained an alignment offset of about 1.5°. For both frequencies a short 5-cycle pulse and a longer 13ms pulse (39-cycle for 3kHz and 65-cycle for 5kHz) have been extracted from the LFM chirp. At 5kHz there were also measurements of the response to a short 5-cycle CW pulse and a longer 13ms CW pulse. The longer pulses exhibit a sharper variability – this is more like the model predictions which relate to an infinitely long pulse. Good agreement can be seen, particularly at 5kHz.

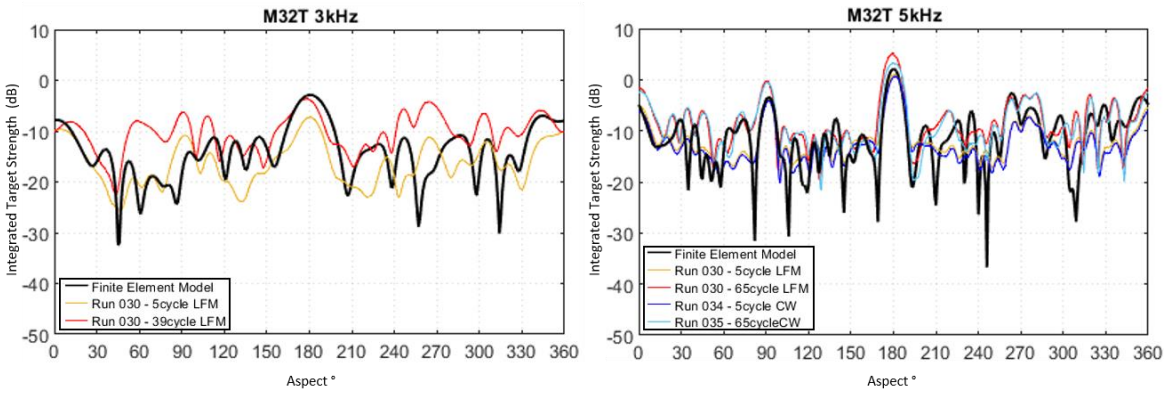


Fig. 5: FEM predicted Integrated Target Strength of M32T (in black) compared with measurements (colour) *at 3kHz (left) and at 5kHz (right).*

5. CONCLUSIONS AND FUTURE WORK

A number of modelling studies have been undertaken to contribute to the understanding of acoustic scattering from the TESSMEX objects M32 and M32T. Ray Acoustic and Finite Element Modelling have been used to examine scatter features from the objects at various frequencies. These models compare well with measurements undertaken so provide mutual validation between the modelling and measurements.

The modelling conducted thus far can be extended and built upon. This may include additional frequencies, to examine areas of interest more closely and sensitivity studies of the predicted TES to a number of factors such as material properties and alignment offset. The modelling could also be extended to other objects, such as TESSMEX M32C.

It should be noted that the modelling undertaken assumes that the object is insonified at pure frequencies. Short pulses might also be investigated using Fourier techniques. However, this would increase by orders of magnitude an already computationally demanding process.

REFERENCES

[1] **Comsol MultiPhysics,** *COMSOL Acoustics Module User's Guide*, v6.2, 2023.
[2] **Gilroy, C. De Jong, B. Nolte, and Schäfer,** *BeTSSi II - Benchmark Target Strength Simulation*, WTD 71/FWG, 2013.
[3] **B. Nolte, I. Schäfer, C. de Jong and L. Gilro**y, BeTSSi II Benchmark on Target Strength Simulation, *In Forum Acusticum proceedings*, Krakow Poland, 2014.
[4] **Ehrlich, Fillinger, Gilroy, Nijhof, Schäfer,** *BeTSSi IIB Workshop Task Specification V1.pdf* 2016.
[5] **F. Ihlenburg**, *Finite Element Analysis of Acoustic Scattering*, Springer-Verlag, 1998.
[6] **Atalla, Sgard**, *Finite Element and Boundary Element Methods in Structural Acoustics and Vibration*, CRC Press, 2015.
[7] **Bundeswehr Technical Center For Ships and Naval Weapons, Maritime Technology and Research**, *Engineering drawings of M32*, WTD 71 110-001181-1483/00-00-000**,** 01-00-000, 02-00-000, 02-01-000, 02-02-000, May 2017.
[8] **TNO,** *M32T_surfaces_AsBuild_rev1_20211028.pdf*, October 2021.

Porting a ray tracing method for determining the target echo strength (TES) of underwater objects to GPUs - a field report

Ralf Burgschweiger[1], Ingo Schäfer[2], Delf Sachau[1] and Jan Ehrlich[2]

[1] Helmut Schmidt University, University of the Federal Armed Forces Hamburg, Faculty of Mechanical Engineering, Chair of Mechatronics, Hamburg Germany
[2] Bundeswehr Technical Center for Ships and Naval Weapons, Maritime Technology and Research (WTD 71), Acoustic Modelling, Eckernförde, Germany

Ralf Burgschweiger, Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg
Fakultät für Maschinenbau, Institut für Mechatronik
Holstenhofweg 85, 22043 Hamburg, Germany
Phone: ++49-40-65 41-44 94
Email: burgschr@hsu-hh.de

Abstract: *Investigations of a test implementation of the "simple" Kirchhoff approximation method (KIA) for GPU graphics adapters (using OpenCL) have shown that, depending on the GPU hardware available, significant performance improvements can be achieved. Therefore, as part of the "Computational Acoustics" research project, the BEAM ray tracing method, originally implemented in C++, was ported to address the required acoustic problems (bistatic, monostatic calculations and their combination). This article discusses the challenges encountered during this process, shares the practical experiences implementing the complex parallelized algorithm in OpenCL and shows the results achieved. It also highlights the speed advantages compared to conventional high-performance PC workstations using examples of underwater objects.*

Keywords: *Target echo strength, ray tracing, GPU*

1. INTRODUCTION

To estimate the backscattering behaviour of objects under water, the so-called target echo strength (TES) is used, which results from the ratio of incident to reflected sound intensity calculated back to a distance of one meter from the object. Various numerical methods (boundary and finite element methods, approximate methods, etc.) are available for this purpose, which were implemented in high-performance applications as part of the joined research project “Computational Acoustics” at the Helmut-Schmidt-University, Hamburg, and the Bundeswehr Technical Center for Ships and Naval Weapons, Maritime Technology and Research (WTD 71), Eckernförde, Germany.

In recent years, the ray tracing-based solver BEAM [1] has been developed, which is able to calculate the TES of underwater structures consisting of surfaces with appropriate boundary conditions and combined with fluid fillings. Due to its high computational speed, it is also very well suited for sweeps over a given frequency range.

In many practical applications for calculations (image processing, artificial intelligence, games, etc.), the use of graphics processors (GPUs) has proven to be very powerful. Therefore, a sub-task of the project is to implement the BEAM algorithm on GPUs and to determine which performance improvements can be achieved compared to the current “standard CPU code”.

This article describes the challenges and practical experience of porting the BEAM method from standard C++ code to OpenCL, gives an overview of the advantages and disadvantages and presents the defining characteristic of a GPU for performance.

The results achieved, also for different hardware, are given using various problem types.

2. BASICS OF THE METHODS / RELEVANT DIFFERENCES

Two years ago at UACE 2023, the advantages of using AVX2-CPU instructions were demonstrated using the approximate Kirchhoff method [2] as an example, for full details see [3]. In the meantime, further work has been done on this and a migration of the KIA process to OpenCL has been implemented for test purposes.

Table 1 shows the new solution times (including the GPU version) for the calculation example at that time (BeTSSi hull, consisting of 178,000 elements, TES values calculated on a spherical section of 0°...180°, elevation ±20°, step width 0.2°, 181,101 evaluation points, averaging over 11 frequencies).

method / variation	solver time [s]
BEM	≈ 300,000
MKIA (AVX2)	≈ 31
MKIA (AVX512)	≈ 11
MKIA (GPU, RTX4090)	**2.9**

Table 1: MKIA solving times

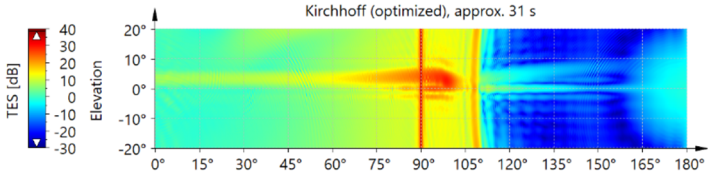


Fig. 1: TES MKIA example (181,101 eval. Points)

Due to the performance advantage of the GPU variant by a factor of 3.8, it was decided to implement an adaptation of the BEAM ray tracing code for GPUs.

2.1. Differences between the methods

The following list provides a brief overview of the differences between the “simple” Kirchhoff method and the “complex” BEAM algorithm in terms of application possibilities and resulting workload that are relevant to this article.

Kirchhoff approximation		BEAM ray tracer
● supports only convex surfaces	⇔	● supports convex and concave surfaces
● supports only hulls with predefined reflection factors	⇔	● supports internal structures with mixed boundary conditions
● the computational load depends only on the number of surface elements (one integration per element)	⇔	● the computational load depends on the maximum number of traced child beams per tree (one tree per start beam, see 2.2) - The number of summations doubles per level, i.e., for a maximum tracing level of $N_{b,max} = 7$, up to 255 beams must be considered - A small change in the direction of sound incidence may lead to completely different beam paths

2.2. Ray tracing algorithm

Fig. 2 and 3 show an example of the structural setup of a “beam tree” resulting from ray tracing up to a tracking level of 5 (in 2D, using a spherical shell with an internal rigid angled plate) to illustrate the difference to the linear load distribution in the KIA algorithm.

Starting from a given sound source and direction, a start beam (with level 0) is generated and tracked further by testing for hits on the structure(s).

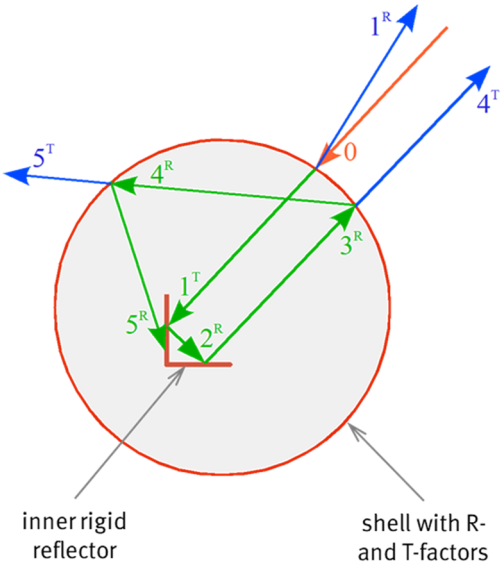


Fig. 2: 2D ray tracing, stop level 5

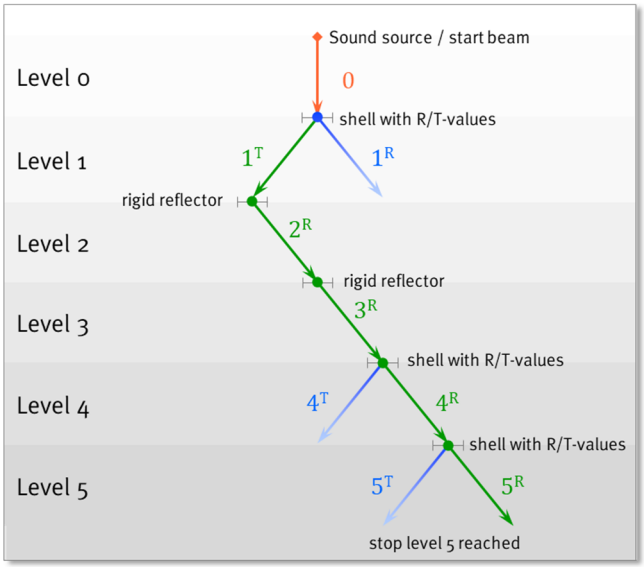


Fig. 3: resulting “beam tree”

Each colored dot ● in fig. 3 corresponds to the “hit” of an element of the surfaces. It is easy to see that, depending on the boundary conditions given, each hit of a beam leads to a decision as to whether a transmitted ($\#^T$) and/or reflected child beam ($\#^R$) is generated and whether this (or these) must be tracked further.

The BEAM method can be parallelized well when subsets of all required "start beams” are assigned to the available CPU (or GPU) cores, since the tracing of one single beam is independent of all others.

After the complete construction of a beam tree, the complex pressure components per partial beam are added up with regard to the evaluation points in a post-processing calculation.

2.3. Relevant hardware differences

To illustrate the challenges involved in implementing the algorithm for a GPU, the relevant differences between a CPU and a GPU are briefly listed below.

“conventional” CPU		Graphics processor (GPU)
● consists of independent cores (up to 192 physical cores per CPU chip or 384 logical cores with hyperthreading)	⇔	● consists of so-called stream processors (**SPs**), organized in compute unit groups (**CUs**), up to 21,760 SPs / 170 CUs per GPU (≙ 128 SPs per CU)
● large and fast cache memory (up to 80 KB L1- and 2 MB L2-cache per core)	⇔	● high speed access to very limited private storage per CU (up to 128 KB)
● fast access to the entire host main memory	⇔	● “slower” access to GPU global memory and need to copy relevant data between host and GPU
● independent instruction execution per core	⇔	● synchronized instruction execution per compute unit (SIMD, single instruction multiple data)

The consequences of these differences for performance will be discussed in more detail later. Table 2 provides an overview of the available and used GPUs:

GPU model	key data					benchmark	price
	CUs	SPs	GPU family	RAM	year	OpenCL	GPU
nVidia GeForce RTX 4090	128	16,384	AD102	24 GB GDDR6x	2022	317,945	1,950 €
nVidia GeForce RTX 4070	56	7,168	AD104	12 GB GDDR6x	2024	222,849	1,000 €
AMD Radeon RX 7900	96	6,144	Navi 31	24 GB GDDR6	2022	212,200	980 €

Table 2: available and tested GPUs

Table 3 provides an overview of the used “conventional” workstations:

CPU model	key data						price
	cores	threads	Gen.	RAM	year	Core freq.	CPU
AMD ThreadRipper Pro 7975	32	64	Zen 4	256 GB DDR5	10/2023	4.0/5.3 GHz	5,000 €
AMD EPYC 9655	96	192	Zen 5	768 GB DDR5	10/2024	2.6/4.5 GHz	7,000 €

Table 3: available workstations

3. OPENCL

OpenCL (Open Computing Language, [4]) is a framework for writing programs that execute across heterogeneous platforms consisting of central processing units, graphics processing units, digital signal processors (DSPs), field-programmable gate arrays (FPGAs) and other processors or hardware accelerators [5].

OpenCL was chosen to implement the adaptation, as it was available on all computers and operating systems used in the project at the time of implementation and is still supported by the hardware manufacturers.

OpenCL has several advantages over “standard” CPU code:

- Hardware independency
  The C-based instruction code used for OpenCL is mostly hardware independent and may run on CPUs, GPUS or FPGAs. When “new” hardware is available, it may be used without code modification, if OpenCL is supported by the driver
- On-the-fly compilation
  All required functions are compiled by the OpenCL driver at runtime in a short time (usually < 5 seconds). This also allows code adjustments to be made depending on the current calculation before it is executed, e.g. defining the number of frequencies for dimensioning the required arrays.
- OS independent
  OpenCL code can run under Windows or LINUX without changes.

3.1. Challenges using OpenCL

But there are also a number of challenges, some of which only became apparent during implementation:

- no dynamic memory support
  The size of all required arrays etc. must be known in advance and the data must be provided in one large memory block.
- no C++ object classes / no inheritance
  All required data must be transferred to special “GPU”-compatible structures that are processed with separate function calls.
- no support for complex numbers
  All functions requiring complex data had to be rewritten.
- no recursive function calls
  The hit-check and ray tracing routines had to be re-implemented without use of recursion.
- very few local memory
  The private memory / register set of a single streaming processor is very small (approx. 1 ... 4 KB), so the size of all local variables had to be minimized as far as possible in order to avoid outsourcing this data to the much slower global memory of the GPU device.
- low double precision performance
  As the performance of OpenCL on GPUs is usually at least twice as high for single as for double precision, the code was only implemented for simple accuracy to reduce the amount of data and speed up data exchange between host and GPU.
  It must be ensured that the size ratios between the evaluation distance and element “size” (edge length etc.) match when using single precision; as a rule of thumb, a ratio of approx. 1:1000 (1 m structure size for 1000 m evaluation distance) should be considered.

3.2. Main “Disadvantage” of using GPUs for raytracing

The biggest challenge while implementing the algorithm was the so-called synchronized instruction execution per computing unit (SIMD, Single Instruction Multiple Data). This means that identical code execution takes place on all streaming processors (SPs) of one compute group (CU). This behaviour has the following consequences:

- Different program paths due to decisions or branches force the SPs to wait until the end of the different code in each case.

- Looking at ray tracing, it becomes clear that no reliable prediction can be made about the number of “child” rays traced based on the “hits” of the elements (and thus the number of decisions in the code).
- A few “complex” beam trees within one CU lead to significant delays for all “simple” trees. If, for example, only 16 complex trees occur within a CU with 128 SPs, all remaining 112 SPs must “wait” until their processing is complete.
- For the most efficient utilization of the GPU, the number of decision paths in the code must therefore be as low as possible.

4. RESULTS

All results of the test cases shown were determined using a BeTSSi submarine model (length approx. 62 m) with different numbers of elements. It contains internal structures to which various boundary conditions (**BCs**, elastic shells, R/T factors, rigid) have been assigned.

All calculations were carried out for the monostatic case (sound source position = evaluation position) over the specified number of frequencies using start beams with a predefined “width” or one start beam per element. The figures showing the TES results are intended only to clarify the respective test case.

The deviations between the results of the CPUs and GPUs were within the scope of the calculation accuracy (single precision). The factors given indicate the speed advantage of the GPU compared to the CPU “reference solution”. Due to the limited space available, the results are not discussed in detail here but evaluated in the conclusions.

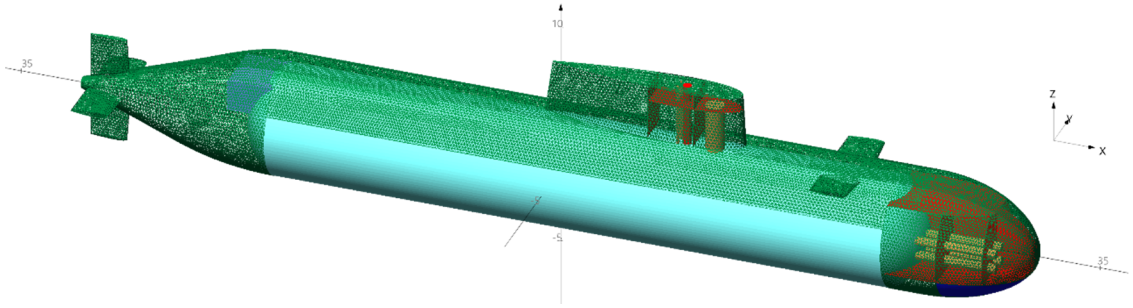
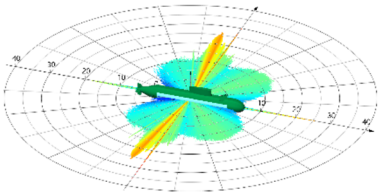


Fig. 4: BeTSSi model with inner structures (116,126 elements, only for illustration)

Test case 1: 1,8M elements, XY plane, R/T conditions

beam width: 0.2 m (max. start beam count $N_{sb,max}$: 17,160)
elements: 1,858,016
frequencies: 21
BCs: R = 0.5, T = 0.5
evaluation: XY-plane, 72,001 points, ±180°, 0.005° steps

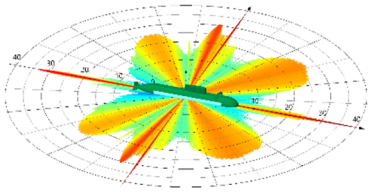


System / Hardware	Solver / mode	Cores / SPs	Threads / CUs	Solver time [s]	Factor
AMD Ryzen Threadripper Pro	BEAM (AVX2)	32	64	1,491.8	1.00
AMD EPYC 9655	"	96	192	924.0	1.61
AMD Radeon RX 7900 XTX	BEAM-OCL/1D	6,144	32	520.0	2.81
nVidia GeForce RTX 4070 SUPER	"	7,168	16	457.0	3.26
nVidia GeForce RTX 4090	"	16,384	16	235.3	6.34
nVidia GeForce RTX 4090	BEAM-OCL/2D	16,834	32/1	179.8	**8.30**

Table 4: Solution times for test case 1

Test case 2: 464K elements, XY plane, mixed boundary conditions

beam width: one per element
elements: 464,504
frequencies: 21
BCs: mixed, realistic (shells, R/T-factors, rigid)
evaluation: XY-plane, 36,001 points, ±180°, 0.01° steps

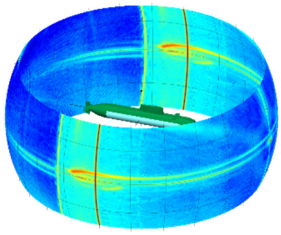


System / Hardware	Solver / mode	Cores / SPs	Threads / CUs	Solver time [s]	Factor
AMD Ryzen Threadripper Pro	BEAM (AVX2)	32	64	5,434.6	1.00
AMD EPYC 9655	"	96	192	3,142.0	1.73
AMD Radeon RX 7900 XTX	BEAM-OCL/1D	6,144	32	1,278.1	4.25
nVidia GeForce RTX 4070 SUPER	"	7,168	8	2,324.4	2.34
nVidia GeForce RTX 4090	"	16,384	8	1,504.9	3.61
"	BEAM-OCL/2D	"	64/1	563.6	**9.64**

Table 5: Solution times for test case 2

Test case 3: 1.8M elements, spherical section, R/T conditions

beam width: 0.2 m (max. start beam count $N_{sb,max}$: 17,160)
elements: 1,858,016
frequencies: 21
BCs: R = 0.5, T = 0.5
evaluation: spherical section, 72,821 points, ±180° / ±25°, 0.5° steps

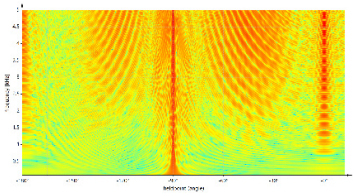


System / Hardware	Solver / mode	Cores / SPs	Threads / CUs	Solver time [s]	Factor
AMD Ryzen Threadripper Pro	BEAM (AVX2)	32	64	3,110.0	1.00
AMD EPYC 9655	"	96	192	1,656.6	1.88
AMD Radeon RX 7900 XTX	BEAM-OCL/2D	6,144	32	2,558.8	1.22
"	BEAM-OCL/2D/R4	"	"	2,310.0	1.35
nVidia GeForce RTX 4090	BEAM-OCL/2D	16,834	128/1	1,203.1	2.58
"	BEAM-OCL/2D/R4	"	"	1,094.0	**2.84**

Table 6: Solution times for test case 3

Test case 4: 464K elements, XY plane, "waterfall" (multiple frequencies)

beam width: one per element
elements: 464,504
frequencies: 491 (vertical axis)
BCs: mixed, realistic (shells, R/T-factors, rigid)
evaluation: XY-plane, 3,601 points, ±180°, 0.1° steps,



System / Hardware	Solver / mode	Cores / SPs	Threads / CUs	Solver time [s]	Factor
AMD Ryzen Threadripper Pro	BEAM (AVX2)	32	64	11,429.0	1.00
AMD EPYC 9655	"	96	192	7,280.0	1.57
AMD Radeon RX 7900 XTX	BEAM-OCL/2D, 5 FB	6,144	226/1	3,912.0	2.92
nVidia GeForce RTX 4090	BEAM-OCL/2D, 10 FB	16,834	226/1	1,850.3	6.18
"	BEAM-OCL/2D, 5 FB	"	"	1,210.9	**9.44**

Table 7: Solution times for test case 4

5. CONCLUSIONS AND OUTLOOK

The implementation for the GPU leads to the following conclusions:

- The achievable acceleration of a process on GPUs basically depends on the "similarity" of the task to be solved. Performing identical tasks with different data is the optimum.

  This means for ray tracing:
  The smaller the "differences" between the beam paths, e.g. due to smaller distances between the evaluation points, the greater the speed advantage. Resorting the evaluation points according to their distance (case 3, spherical section) led to a speed advantage of ≈ 10%.
- The use of so-called two-dimensional kernel functions (recognizable in the result tables by the addition "/2D" in the Solver column), where one call is made for each start/evaluation point combination, leads to a significantly better utilization of the GPU cores.
  With the one-dimensional kernel functions tested initially, one call is made for each evaluation point, which led to a significantly poorer utilization of the GPU, especially if there are fewer evaluation points than cores.
- Overall, depending on the application, the best GPU achieved a speed advantage by a factor of 2.5 … 10 in relation to the CPU reference.
- Considering the price of the hardware, a corresponding GPU offers a better price-performance ratio, as no high-performance CPU is required in this case.

However, high-performance CPUs continue to offer advantages if the required memory space, e.g. for matrix-based methods such as BEM, exceeds the memory space available on the GPU and are generally also more flexible in use.

Surprisingly, the speed factor of the CPU with 96 cores is only ≈ 1.6 … 1.9, whereas at least 3 was expected, also due to the 5th core generation. Therefore, it is planned to explore whether the insights gained from the GPU implementation can be transferred to the CPU code version.

6. PROJECT PARTNER

This project is funded by the Bundeswehr Technical Center for Ships and Naval Weapons, Maritime Technology and Research (WTD 71), Eckernförde, Germany.

REFERENCES

[1] **R. Burgschweiger, I. Schäfer, M. Ochmann and B. Nolte**, "Results of the ray-tracing based solver BEAM for the approximate determination of acoustic backscattering from thin-walled objects", in *Internoise 2014*, Melbourne, Australia, pp. 1 … 10, 2014.

[2] **I. Schäfer and B. Nolte,** "Berechnung der akustischen Rückstreustärke von Unterwasserobjekten mit Hilfe der Randelementmethode und der Kirchhoffschen Hochfrequenznäherung" in *DAGA 2008*, Dresden, Germany, 2008.

[3] **R. Burgschweiger, I. Schäfer, D. Sachau and J. Ehrlich**, "High-performance calculation of the acoustic backscattering strength based on the Kirchhoff high-frequency approximation", in *Underwater Acoustic Conference and Exhibition (UACE 2023)*, Kalamata, Greece, pp. 303 … 310, 2023.

[4] **Khronos Group**, "OpenCL for Parallel Programming of Heterogeneous Systems", https://www.khronos.org/opencl/

[5] **Wikipedia**, "OpenCL", https://en.wikipedia.org/wiki/OpenCL

Structured Session
18. Towards Automatic Target Recognition. Detection, Classification and Modelling

Organised by Johannes Groen, Yan Pailhas, Roy Edgar Hansen and Narada Warakagoda

Development of an airborne circular sonar for research and teaching

Blair Bonnett[1], Kaushikk V. N.[1], Sudhanshu Apte[1], and Thomas Fickenscher[1]

[1] Chair for Radio Frequency Engineering, Helmut Schmidt University, 22043 Hamburg, Germany

Corresponding author: *Blair Bonnett, Lehrstuhl für Hochfrequenztechnik, Fakultät für Elektrotechnik, Helmut-Schmidt-Universität, 22039 Hamburg, Germany; blair.bonnett@hsu-hh.de*

Abstract: *Sonar is largely considered to be an underwater technique due to the predominance of other measurement methods for above-water applications. However, in-air echolocation is used by various animals (bats being the obvious example). It has also been utilised in some electronic sensors, for example basic range finders and fluid level sensors. Recent publications have shown that synthetic aperture sonar (SAS) images can be formed from data collected with standard audio speakers and microphones. This allows complex scenes to be imaged in a controlled laboratory environment for validating processing and analysis techniques, although differences in the acoustic response of the targets in air versus in water must be considered. Such a system can also be a valuable tool for teaching and demonstrating sonar principles.*

This paper introduces an airborne circular SAS system under development. It uses a Peerless OC16SC00-04 speaker (16 mm diameter) and Syntiant SPG08P4HM4H-1 MEMs microphones (4 mm by 2 mm). An AMD Zynq system-on-chip comprising a dual-core ARM Cortex-A9 processor and AMD 7 series FPGA is used to control the system and synchronise the components. The system is operated over a standard Ethernet connection. The small size of the microphones and the ability of the FPGA to handle many microphones simultaneously (62 in the current setup) allows for receiver arrays analogous to those used by underwater systems. This creates an opportunity to implement array-based processing techniques such as interferometry or motion compensation.

As with underwater sonar, precise calibration of the system and knowledge of the environment (in particular, the speed of sound) are required to obtain good results. We show the current status of the system and its calibration along with some applications. Future development plans, including for alternative hardware designs, will be outlined. All custom hardware designs and software for using the system will be released under open source licenses.

Keywords: *synthetic aperture sonar, airborne sonar, sonar teaching*

1. INTRODUCTION

Sonar is largely considered to be an underwater technique. Echolocation was first investigated for obstacle detection in the wake of the *Titanic* sinking and was then developed (initially under the name ASDIC) for anti-submarine detection. Subsequent research and usage have mostly focused on underwater applications. Although other methods (optical, radar, lidar etc.) are more prevalent for above-water applications, the principles of sonar also work in air. In nature, many species of bat use echolocation for navigation and hunting as do some other mammals and birds, while some prey species have evolved countermeasures [1]. In-air sonars have been used for a variety of applications such as aids for blind people [2] and to develop and test general sonar processing techniques [3] while ultrasonic distance sensors are commonly used in robotics and fluid level monitoring.

Recent publications have shown that synthetic aperture sonar (SAS) processing can be used to form images from data collected with standard audio speakers and microphones [4, 5]. This can be used to image complex scenes in controlled environments to generate datasets for validating processing and analysis techniques [6]. Although the acoustic response of specific targets will differ in air compared to water, they exhibit the same range of complex scattering effects (occlusion, multiple scattering, elastic scattering etc) that is observed in underwater sonar.

In this paper, we will introduce an airborne circular SAS we are developing. This is based around micro-electromechanical systems (MEMS) microphones which are sufficiently small to form receiver arrays, similar to those commonly used in underwater SAS systems. We start by describing the electronics used in the system and the platform it is mounted to. This is followed by details of the firmware used to control the system and capture the microphone outputs, and the subsequent processing to form SAS images from this data. Some example images captured by the airborne sonar are shown before we conclude with plans for improving and extending this system.

2. ELECTRONICS

A MEMS microphone uses a thin membrane suspended above a fixed backplate as the sensor. The surrounding case is sealed apart from one hole known as the audio port. When a pressure wave enters the audio port, it deflects the membrane. This change in distance between the backplate and the membrane can then be detected as a change in capacitance. They are small and widely used in consumer devices such as smartphones, wearable electronics and televisions.

For this project, the Syntiant (formerly Knowles) SPG08P4HM4H-1 MEMS microphone was selected. This comes in a 4 mm × 2 mm package with an acoustic port diameter of 0.65 mm, an overload point[1] of 120 dB_{SPL}, an SNR of 64 dB with a 1 Pa pressure wave and a total harmonic distortion of less than 0.1 % for the same 1 Pa input. Although designed for use in the audible range, the ultrasonic response plot in the datasheet suggests a −3 dB cutoff at approximately 44 kHz. The output of the microphone is a single-bit pulse-density modulation (PDM) signal (i.e., the output of a sigma-delta modulator) referenced to a clock signal of up to 4.8 MHz supplied to the microphone. Each microphone can be configured to output its data on either the rising or falling edge of the clock signal, allowing two microphones to share the same data line (often utilised for stereo pairs in consumer devices).

[1] Note that all sound pressure levels mentioned in this paper are relative to the 20 µPa reference level used in airborne audio (the limit of human hearing) instead of the 1 µPa reference normally used in underwater acoustics.

A printed circuit board (PCB) was designed to hold eight of these microphones with a spacing of 8.5 mm (approximately half a wavelength at a frequency of 20 kHz). A 12-wire ribbon cable provides power and the sampling clock, and returns the eight data lines. Solder bridges are used to configure whether the microphones output on the rising or falling edge of the clock. To avoid excess load on the input clock signal, a Texas Instruments LMK1C1108 clock buffer is used to distribute a copy of it to each microphone.

A Peerless OC16SC00-04 tweeter is used as the transmitter. This has a circular face with a diameter of 16 mm, a sensitivity of 81.8 dB_{SPL} at 1 m for a 1 W input signal and a power rating of 100 W (which should then correspond to 87.8 dB_{SPL} at 1 m). Its nominal frequency range is 2.5 kHz to 20 kHz; the datasheet indicates good on-axis performance up to 40 kHz although the off-axis strength decreases before this. An STMicro TDA7498E class D audio amplifier in a manufacturer-supplied evaluation kit rated for 160 W is used to drive the speaker.

The target to be imaged is mounted on a rotating stage (technically, this forms an inverse circular SAS as the target moves while the imaging system is stationary). A Phidgets RB-Phi-267 stepper motor is used to turn the stage. The motor has a step angle of 1.8° and is connected to a 1:99.51 planetary gearbox for a final step angle of 0.0181°. The Pololu Tic 36v4 stepper motor driver is used to control the motor.

To control the system, a Digilent Zybo Z7-20 development board is used. This contains an AMD (formerly Xilinx) Zynq-7020 system-on-chip (SOC) which integrates a dual-core ARM Cortex-A9 processor and an FPGA. This combination allows custom hardware interfaces to be designed in a hardware description language and connected to the processor via standard communications buses. The development board contains a number of peripherals, of which the gigabit Ethernet connector, audio codec and expansion connectors are used in this project. Two PCBs have been built to interface to this development board: one for communication with the motor controller and another with connectors for up to eight of the microphone boards. Six of these connectors can handle all eight microphones from a PCB while the final pair have one connection unavailable as the corresponding output from the development board is needed for the clock signal. This allows up to 62 microphones to be connected to the system.

3. PLATFORM

The system (shown in Figure 1a) is mounted onto a 1.2 m long by 0.6 m wide wheeled trolley so that it can be easily moved for teaching, demonstrations etc. The rotating stage consists of an 80 cm diameter disc of 6 mm thick perspex with the motor mounted from underneath. A series of captive ball bearings provide support to the edge of the stage. For some experiments, a 10 mm thick and 20 mm high wooden border is mounted around the perimeter of the stage to contain loose material; Figure 1a shows the stage filled with gravel. The PCBs are mounted at the other end of the trolley from the rotating stage while a lower shelf holds the power supplies.

The speaker and microphones are mounted in a 3D-printed holder allowing their arrangement to be easily customised for a given scenario. In the results presented here, an interferometric pair of 16-element receiver arrays with a baseline of 51 mm is used with the transmitter located equidistantly between and aligned with the centre of the arrays. This setup is shown in Figure 1b. The centre of the speaker is located 750 mm laterally from the centre of the rotating stage. The head is tilted down with the height of the transmitter set so that its boresight is pointing at the centre of the stage, i.e., the height depends on the tilt angle.



(a) *(b)*

Figure 1: The airborne sonar system: (a) rotating stage, sound head and PCBs (power supplies are on a shelf below the main table) and (b) a close-up of the sound head. The speaker is in the centre, and two horizontal linear arrays of MEMS microphones (each made up of two of the microphone PCBs) can be seen, one above and one below the speaker.

4. FIRMWARE

The VHDL language is used to implement the firmware for the FPGA side of the Zynq-7020. The main sonar peripheral has a number of responsibilities:

- Configure the audio codec to act as a digital-to-analogue converter to drive the amplifier.
- Load signals to transmit from the processor.
- Handle the signals returned from all microphones, including separating the signals from the rising and falling edges of each input, and allowing the processor to specify reordering of the input channels for flexibility.
- When requested, perform a ping. This involves sending the samples of the previously loaded signal to the audio codec for transmission, synchronising the receiver to the transmitter, and recording a set number of samples from the microphones.

The recording utilises direct memory access to directly place the samples in RAM without any processor overhead. As this is performed by the FPGA, all the microphones inputs are captured simultaneously; the accuracy in relative timing between different microphones, and between the speaker and the microphones, is determined by the electronics.

Currently only a single core of the processor is used. This is programmed in C utilising the support libraries provided by the vendor along with a number of custom libraries. It is designed to act as a simple server over the Ethernet connection provided on the development board. A client can connect to the server, configure the sonar and then start a capture. The recorded samples are returned to the client formatted in a simple packet.

5. PROCESSING

The processing is performed in the Python programming language with a number of common libraries, namely NumPy, SciPy and Xarray. The processing code communicates with the sonar over the network interface described previously. As recorded samples are received, they are placed in a structured dataset and subsequently stored using the netCDF binary format.

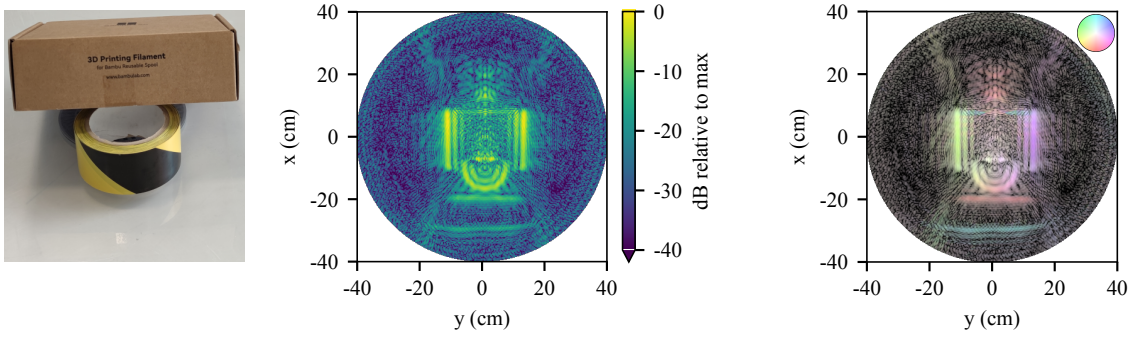
As mentioned in the electronics section, the output of the microphones is a single-bit PDM signal; we use the maximum 4.8 MHz sampling rate. This needs filtering and downsampling to provide a higher-resolution signal at lower data rate. When performed in real-time, this requires careful design of the filters to ensure both computational efficiency and sufficient attenuation of the high-frequency noise in the original signal. In this case, we have no real-time requirements and so standard finite impulse response (FIR) filters can be used. SciPy's `signal.decimate` function is used to downsample the signal in two stages, first by a factor of eight after applying a low-pass FIR filter with 160 taps and then by a factor of six after applying another low-pass FIR filter with 120 taps. In both stages the group delay of the filters is compensated for to avoid introducing phase errors. The resulting signal is sampled at 100 kHz.

The microphone outputs have a low-frequency drift. To remove this and any other out-of-band noise, a FIR bandpass filter with transition frequencies 5 kHz either side of the transmitted spectrum is applied. Pulse-compression is then performed by correlating the received echo with a replica of the transmitted signal. This correlation is performed in the Fourier domain with a guard band to avoid circular effects. The pulse-compressed signal is converted to its complex analytic representation and a range-varying gain applied before delay-and-sum backprojection is used to form an image.

6. EXAMPLE IMAGES

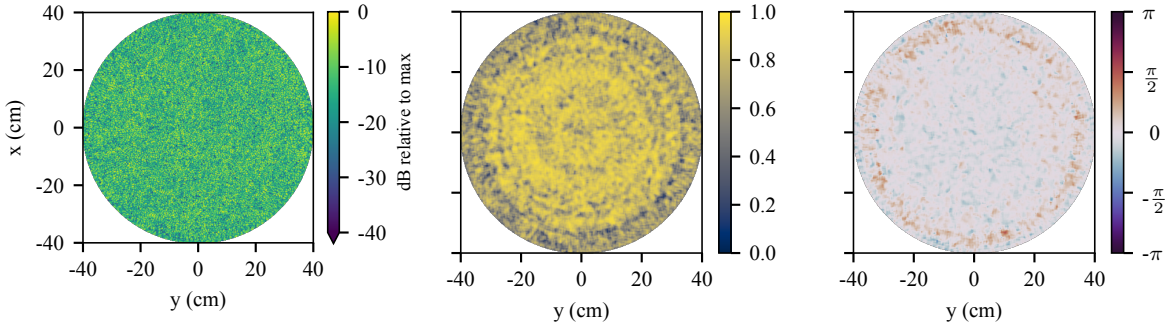
In initial testing, it was observed that the received energy began to decrease as the frequency exceeded 20 kHz. As both the speaker and microphones should be able to operate up to 40 kHz, it is assumed that the amplifier is the cause of this limitation. To accommodate this, a 10 kHz to 30 kHz hyperbolic frequency-modulated chirp of 2 ms duration was used in the data collection for the images presented here. This modulation ensures a higher proportion of the chirp uses the lower part of the spectrum. The stage was rotated 3° between pings which results in an approximately 50 % overlap of the phase centre arrays from ping to ping. The images presented here have been reconstructed at a 2 mm resolution with areas outside the rotating stage removed.

The first scene is a cardboard box (205 mm × 220 mm × 70 mm) tilted up on a roll of tape (108 mm diameter, 50 mm high), shown in Figure 2a. The SAS image from the upper array (Figure 2b) shows that the edges of the targets reflect strongly. Some multipath echoes can be observed as can some resonance patterns. The background is mostly noise as the rotating stage is highly specular. The repeating pattern around the outer edge of the circle is due to a stationary reflector outside the stage which is smeared around the entire circle. Figure 2c shows a colour-by-aspect (CBA) version of the reconstructed image. Following the approach of [7], this encodes the direction from which each pixel appears brightest as the hue, the inverse entropy of the pixels as the saturation and the intensity as the value. This means pixels which were observed from all directions tend to greyscale while pixels which were observed primarily from one direction have the corresponding colour. The embedded colour wheel shows the mapping of colour to view direction. From this, it can be seen that, as expected, the sides and edges of the targets were mostly imaged from the corresponding side of the stage.



(a) Scene *(b) SAS image from upper array* *(c) CBA image from upper array*

Figure 2: Results from imaging a box tilted up on a roll of tape with a 30° grazing angle.



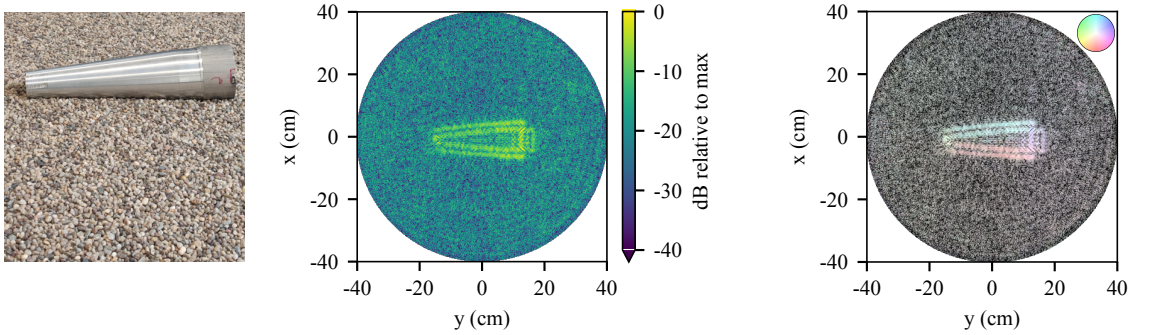
(a) SAS image from upper array *(b) Coherence between arrays* *(c) Interferometric phase*

Figure 3: Results from imaging a bed of gravel with a 30° grazing angle.

For the remaining results, the rotating stage was filled with gravel to a depth of 20 mm. The nominal size range of the gravel is 3 mm to 5 mm although both larger and smaller pieces are present in small quantities. Figure 3a shows the SAS image reconstructed from the upper array with no extra target present. This has the speckle nature expected of a rough background. The coherence between images from the two arrays (Figure 3b) is high in the middle of the stage but decreases near the edges. It is hypothesised that the edges are not insonified on every ping (i.e., are outside the full-view area) at the high end of the spectrum due to the narrow beamwidth of the transmitter. These areas of the image would then have a lower SNR leading to reduced coherence. The interferometric phase between the images (Figure 3c) is low in the centre of the stage, corresponding to a surface with no significant height variations. The higher values around the edge correspond to areas of low coherence and as such can be ignored.

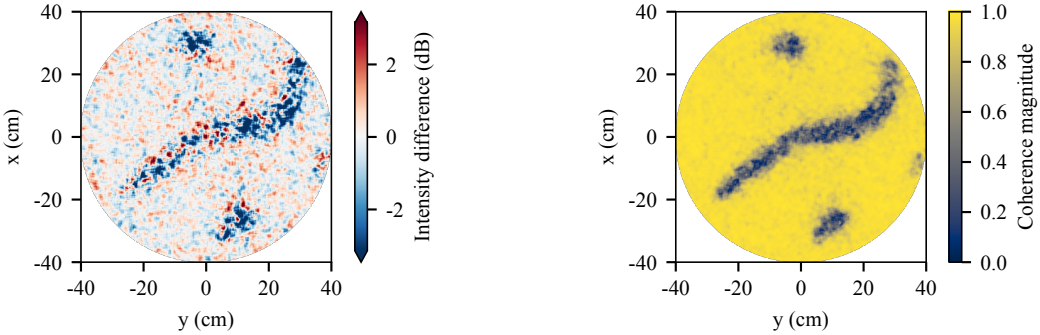
A shell-like shape machined from aluminium was then placed on the gravel. This is 290 mm long and tapers from 80 mm to 40 mm in diameter with an untapered portion at the large end. This scene is shown in Figure 4a and the SAS image from the upper array in Figure 4b. A double echo can be seen; it is not known whether this is a multipath reflection or an elastic scattering effect. The CBA image of Figure 4c shows that the echoes from the target were observed directionally while the speckle pattern from the gravel was captured from most angles.

Finally, the suitability of the system for change detection was investigated. Two images of the gravel were captured. Between these captures, some trenches were created in the gravel with a spade. As shown in Figure 5, both incoherent and coherent change detection can be successfully performed on these images.



(a) Scene *(b) SAS image from upper array* *(c) CBA image from upper array*

Figure 4: Shell-like target imaged at a 15° grazing angle.



(a) Incoherent change detection *(b) Coherent change detection*

Figure 5: Change detection between two images of a bed of gravel. A 7 by 7 pixel window (1.4 cm by 1.4cm) was used to smooth the results in the incoherent case. The same size window was used when estimating the coherence.

7. CONCLUSION

The initial results presented here show that MEMS microphones are suitable for collecting airborne sonar data which can be coherently processed to form SAS images. Furthermore, their small size allows them to be formed into arrays which offers the potential for testing standard sonar algorithms which require such arrays, for example bathymetric estimation from interferometry. The successful application of change detection suggests that the system produces repeatable results (the only significant changes were those deliberately introduced to the scene).

There are improvements to the current system we wish to implement. As mentioned above, the transmitter amplifier appears to be limiting the maximum frequency to a point below that which the speaker and microphones should be capable of. Both the digital-to-analogue conversion and the amplifier also introduce delays between a ping being requested and the chirp starting. Although this can be calibrated to a point where any receiver timing error due to these delays do not cause significant problems to image formation, it would be preferable to eliminate them. A class D or similar amplifier driven directly from the FPGA would allow full control over the transmitter and could be designed to work with higher frequencies. Full calibration of the transmitter source levels, receiver sensitivity and beampatterns of both transmitter and receiver is also planned.

The images presented here were processed assuming a sound speed of 343 m/s. Some work has been performed towards measuring the sound speed with the system. In the future, we intend to add temperature sensors to the system to also allow standard temperature-based sound speed models to be used. Pressure and humidity sensors can also be added. These parameters typically have minimal impact on the sound speed, but may be more significant for modelling the attenuation of the sound.

A linear variant of this system is planned in the future. The ability to add non-ideal motion, for example yaw or sway, to the capture geometry would allow the collection of data more representative of an underwater system. Interfacing a MEMS receiver array to a single-board computer such as a Raspberry Pi instead of requiring the use of a CPU + FPGA system would make the system more accessible for others to implement, albeit with capacity for fewer microphones and potentially with less timing accuracy.

The designs for this system have been released under open-source licenses and can be found at `https://github.com/airsonar`. The circuit designs are available under the CERN Open Hardware Licence and source for the firmware under the BSD 2-clause plus patent license.

ACKNOWLEDGEMENTS

The authors would like to thank the technical staff of Helmut Schmidt University for their assistance in building the airborne sonar and the experimental targets. We would also like to thank Dr. Holger Schmaljohann of WTD 71 for valuable feedback and suggestions regarding this project.

REFERENCES

[1] W. E. Conner and A. J. Corcoran. “Sound Strategies: The 65-Million-Year-Old Battle Between Bats and Insects”. *Annual Review of Entomology*, vol. 57, 2012, pp. 21–39.

[2] L. Kay. “A sonar aid to enhance spatial perception of the blind: engineering design and evaluation”. *Radio and Electronic Engineer*, vol. 44, no. 11, 1974, pp. 605–627.

[3] P. T. Gough, D. de Roos, and M. J. Cusdin. “Continuous transmission FM sonar with one octave bandwidth and no blind time”. *IEE Proceedings F (Communications, Radar and Signal Processing)*, vol. 131, no. 3, 1984, pp. 270–274.

[4] T. E. Blanford, J. D. McKay, D. C. Brown, J. D. Park, and S. F. Johnson. “Development of an in-air circular synthetic aperture sonar system as an educational tool”. *Proceedings of Meetings on Acoustics*, vol. 36, no. 1, Aug. 2019, p. 070002.

[5] T. E. Blanford, L. Garrett, J. D. Park, and D. C. Brown. “Leveraging audio hardware for underwater acoustics experiments”. *Proceedings of Meetings on Acoustics*, vol. 46, no. 1, Nov. 2022, p. 030002.

[6] T. E. Blanford, D. P. Williams, J. D. Park, B. T. Reinhardt, K. S. Dalton, S. F. Johnson, and D. C. Brown. “An in-air synthetic aperture sonar dataset of target scattering in environments of varying complexity”. *Scientific Data*, vol. 11, no. 1, Nov. 2024, p. 1196.

[7] D. S. Plotnick and T. M. Marston. “Utilization of Aspect Angle Information in Synthetic Aperture Images”. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, 2018, pp. 5424–5432.

Current state of the openSTB sonar simulator

Blair Bonnett[1], Holger Schmaljohann[2], Sudhanshu Apte[1], and Thomas Fickenscher[1]

[1] Chair for Radio Frequency Engineering, Helmut Schmidt University, 22043 Hamburg, Germany
[2] Bundeswehr Technical Center for Ships and Naval Weapons, Maritime Technology and Research (WTD 71), 24340 Eckenförde, Germany

Corresponding author: *Blair Bonnett, Lehrstuhl für Hochfrequenztechnik, Fakultät für Elektrotechnik, Helmut-Schmidt-Universität, 22039 Hamburg, Germany; blair.bonnett@hsu-hh.de*

Abstract: *The development and implementation of algorithms for synthetic aperture sonar (SAS) processing requires data with ground truth for validation. In most instances, it is difficult or even impossible to collect real data with sufficiently accurate ground truth. Instead, simulated data is commonly used for this purpose. The requirements for these simulations may range from a simple setup with a few point targets to large models of realistic scenes requiring significant computational resources.*

We previously outlined plans to create an open-source sonar simulation framework to facilitate community-driven development and sharing of simulation techniques. This project, subsequently named the openSTB (open sonar toolboxes) simulator, is now under active development. It is written in Python utilising widely available numerical processing libraries such as NumPy, and has support for high-performance computing (HPC) environments built in. It is designed in a modular structure with each plugin having a specified interface. A simulation is then created by combining the desired plugins. As well as the actual simulation methods, this modular structure extends to the loading of inputs and saving of results, allowing the simulator to be integrated with existing data formats and other processing tools. Many common implementations of these plugins are included, while the defined interface allows users to write their own to customise the simulations to their needs.

In this paper we will present the current state of the openSTB simulator. This will include examples of different simulations that can be performed, including simulations utilising an HPC cluster. Future plans for further included plugins and simulation techniques will also be outlined.

Keywords: *synthetic aperture sonar, simulation*

1. INTRODUCTION

Developing, implementing and improving any processing algorithm requires data with *ground truth*, i.e., for which correct results is known, in order to evaluate its performance. It is often difficult or impossible to collect such data suitable for verifying sonar algorithms. For example, tracking the position of a vehicle while it is underwater typically relies on an intertial navigation system. This is not sufficiently accurate to generate a well-focused synthetic aperture sonar (SAS) image. Techniques such as the displaced phase-centre antenna (DPCA) algorithm are used to correct the navigation data based on the collected sonar data. By definition, real data with this problem does not have ground truth and so using it to analyse the performance of the correction algorithm can only be done with secondary statistics such as the sharpness of an image generated using the corrected navigation. Simulated data is therefore vital for evaluating the performance of such algorithms.

The complexity of the required simulations will vary based on application. Simulating a small number of ideal point targets can be sufficient for simple cases such as checking the point response of a processing chain. Bigger scenes can be modelled with a large number of point targets. Some algorithms may required more advanced techniques such as incorporating elastic scattering effects or modelling shadows cast by objects in the scene. Each increase in simulation complexity results in a corresponding increase in both the effort required to write and validate the simulation code and the computational resources required to perform the simulation in a timely manner.

In a presentation at the International Conference of Underwater Acoustics in 2024, we postulated that a significant majority of sonar researchers had at some time written a simulator of some form. We then proposed the creation of an open-source sonar simulation framework to refocus such individual development efforts into a single tool that would benefit the whole community. Access to a high-quality simulation tool would also reduce the cost of entry for new researchers in the community who need data to work with. A standard tool to create controlled, reproducible data could also be useful in evaluating and comparing different algorithms for a particular task.

This simulation framework, named the openSTB (open sonar toolboxes) simulator, is now under development using the BSD-2-Clause Plus Patent license. In this paper we will first list the general design principles used for the simulator and then the software stack it uses. An overview of how the simulator can be accessed and used along with its current features are then given, followed by some example simulations. We conclude with an overview of our ideas for future improvements to the simulator.

2. DESIGN PRINCIPLES

A general-purpose simulator needs to be readily adaptable to the needs of each user. To ensure this flexibility, the following design principles have been adopted for the openSTB simulator:

- **Everything should be a plugin.** All operations, from loading a configuration through modelling a physical process to scheduling computation of each piece of the simulation and saving the results in the desired format should be performed by plugins. A simulation is then defined as the set of plugins that should be combined to generate the output.

- **Plugins should be easy to write.** In order to customise a simulation, the user may wish to write their own plugins. The interface for each kind of plugin needs to be well-defined, and the expected inputs and outputs clearly documented.
- **Plugins should be easy to install and use.** Having written a plugin, the user needs be able to easily install it so that the simulator can find it, or be able to simply point the simulator to the file containing the plugin.
- **Batteries should be included.** Many operations will be common across a wide range of simulations. Plugins for such operations should be provided with the simulator.

3. SOFTWARE STACK

The Python programming language has been chosen as it is itself open-source and widely available, and has a large numerical and scientific programming ecosystem. We also believe that a large proportion of the target audience of the project will have some degree of familiarity with both Python and the scientific Python ecosystem. Although the intention is for the simulator to contain sufficient plugins for many use-cases, using a language with which users are likely to be familiar lowers the barrier to customising it with their own plugins.

As of the time of writing the external libraries the simulator depends on include the following:

- NumPy (`https://numpy.org/`) for array handling and numerical computation
- SciPy (`https://scipy.org/`) for various scientific computations
- Dask (`https://www.dask.org/`) and its distributed task scheduler (`https://distributed.dask.org/`) for parallel and distributed computing
- Zarr (`https://zarr.dev/`) for storing results
- Quaternionic (`https://quaternionic.readthedocs.io/`) to implement quaternion algebra for rotations
- Click (`https://click.palletsprojects.com/`) to generate a command-line interface to use the simulator

It is worth noting that the Numba just-in-time compiler (`https://numba.pydata.org/`) is included as a sub-dependency of these libraries. It is envisaged that in the future some parts of the simulator may utilise Numba for performance improvements.

Large simulations are computationally expensive, but the process is typically straightforward to parallelise as in many cases the final result for a collection of targets can be broken into a sum of the results obtained from subsets of those targets. The Dask library used by the simulator provides support for parallel computing by maintaining a graph of tasks to be performed and scheduling their execution. The standard Dask scheduler works on a single machine, utilising all the CPU cores (or a portion of the cores if desired) to perform a parallel computation. To utilise a high-performance computing (HPC) environment, the Dask distributed scheduler can be used. This distributes the computation across multiple nodes in a cluster, and can be used with many common HPC resource managers. The simulator includes support for both the single-machine and distributed schedulers.

It is intended to follow the scientific Python community's SPEC0 (Minimum Supported Dependencies) guideline [1] to determine which versions of Python and the libraries to support. Broadly, this corresponds to versions of Python released in the previous three years and versions of the libraries released in the previous two years. Exceptions may be made if required, for example to avoid a critical bug in a library.

4. USING THE SIMULATOR

The openSTB website at `https://openstb.dev` hosts documentation and other details about the project. Development of the simulator occurs on GitHub at `https://github.com/openSTB/simulator`. Users who wish to help with future development, or who want to test new features, can install it from that Git repository. Releases are made on the Python package index (PyPI) under the name `openstb-simulator` so that the simulator can be installed with the user's preferred Python package manager.

It is envisaged that many users will want a simple interface which they can supply with a configuration file and receive the results of the simulation. The simulator includes a simple command-line interface to provide this functionality. By default, it reads a configuration file in TOML format but, in keeping with the design principles, support for other configuration formats can be added as plugins (this could include support for loading the configuration from a network source). The configuration file defines the set of plugins, including all relevant parameters for those plugins, which make up the simulation.

Other users may wish to use the simulator as a library, for example to perform multiple simulations with varying parameters, to integrate it into an existing interface or simply to use the functionality of some plugins. The code is structured so as to facilitate such use, whether from a simple script or from a more interactive source such as a Jupyter notebook. It is commented, has full docstrings to describe the behaviour of each function and includes static typing information.

4.1. CURRENTLY INCLUDED PLUGINS

As mentioned in the design guidelines, the components of a simulation are implemented as plugins. In this section, we provide a non-exhaustive overview of the plugins that are included with the simulator at the time of writing. This is intended both to show the sorts of simulation that are possible 'out of the box', and to give an idea as to how simulations can be customised by replacing the appropriate components.

Ultimately, a sonar system can be described to the simulator as a transmitter, one or more receivers and the signal used by the transmitter. A transducer plugin gives details of the position, orientation and distorting factors (such as the beampattern) of both the transmitter and the receivers. The signal plugin is responsible for sampling the transmitted signal for use in the simulation. These samples can be windowed by an attached window plugin; currently a set of common windows are included. These plugins can be separately defined or, for convenience, a system plugin can be used to provide all details of a particular system.

The trajectory followed by the system during the simulation is specified by a plugin which reports the position, orientation and velocity of the system at a requested time. Plugins implementing idealised trajectories are included, and it is straightforward to write a custom plugin which loads this information from an external dataset. Similarly, the times at which a ping is started is provided by a plugin; implementations for a constant time interval between pings and a constant distance travelled between pings are included.

After a ping is transmitted, the simulator needs to calculate how long the signal takes to reach a particular target and return to the sonar. This is done by a travel time calculator plugin; implementations making the stop-and-hop assumption and iteratively calculating the travel time to include intra-ping motion (described in more detail in the example simulation section) are both included.

Distortions, either amplitude-only or frequency-dependent, model how the signal changes as it propagates. Currently included plugins include acoustic attenuation based on the model in [2], Doppler distortion of the signal due to the movement of the sonar, geometric spreading and beampattern effects of the transducers.

Higher-level operations are also designed as plugins. This includes how to load simulation configuration from a file (support for a TOML configuration file is included), how to utilise the other plugins to perform a simulation (a point target simulation is currently implemented) and how to store the results (support for both NumPy and MATLAB files are included).

5. EXAMPLE SIMULATIONS

The following simulations use idealised point targets to model a scene. Each of these targets has a reflectivity factor which is the fraction of incident energy that is scattered back to the sonar; note that there is no aspect dependence. This is a simple model of the behaviour of a target, but can suffice for many applications, and is also a useful approach to validate the general design of the simulator framework.

The simulation is performed in the frequency domain. This allows the time delay corresponding to the travel time to be applied without interpolation, and also enables frequency-dependent effects to be modelled. The steps taken to calculate the echo observed by a receiver from a target are:

1. Evaluate the spectrum of the transmitted signal at the frequencies corresponding to the desired length and sampling interval of the final simulated result.
2. Modify the spectrum with any transmit effects, for example the beampattern of the transmitter, spreading losses or acoustic attenuation.
3. Compute the two-way travel time it takes the transmitted pulse to reach the target and echo back to the receiver.
4. Apply this delay and the target reflectivity to the spectrum.
5. Apply any receive distortions to the spectrum.
6. Use an inverse Fourier transform to obtain the time-domain echo signal from the target.

Summing the echoes from each of the targets in the scene gives the final signal recorded by the receiver. Note that in practice this summation is performed in the frequency domain before taking the inverse Fourier transform. For large scenes, each worker can independently compute the echo from a subset of the targets with the final result then being the sum of all these intermediate results.

The system used for the examples consists of a 3 cm wide and 5 cm high transmitter and a 35-element linear array of 3 cm by 3 cm receivers with a 3.5 cm spacing between elements. A 10 ms long linear frequency-modulated chirp from 70 kHz to 90 kHz was transmitted. Spherical spreading, acoustic attenuation and the beampatterns of the transducers were included in the simulation, though it is worth noting that our processing toolchain compensates for these effects when reconstructing an image. Configuration files for all the following simulations are included as examples with the simulator.

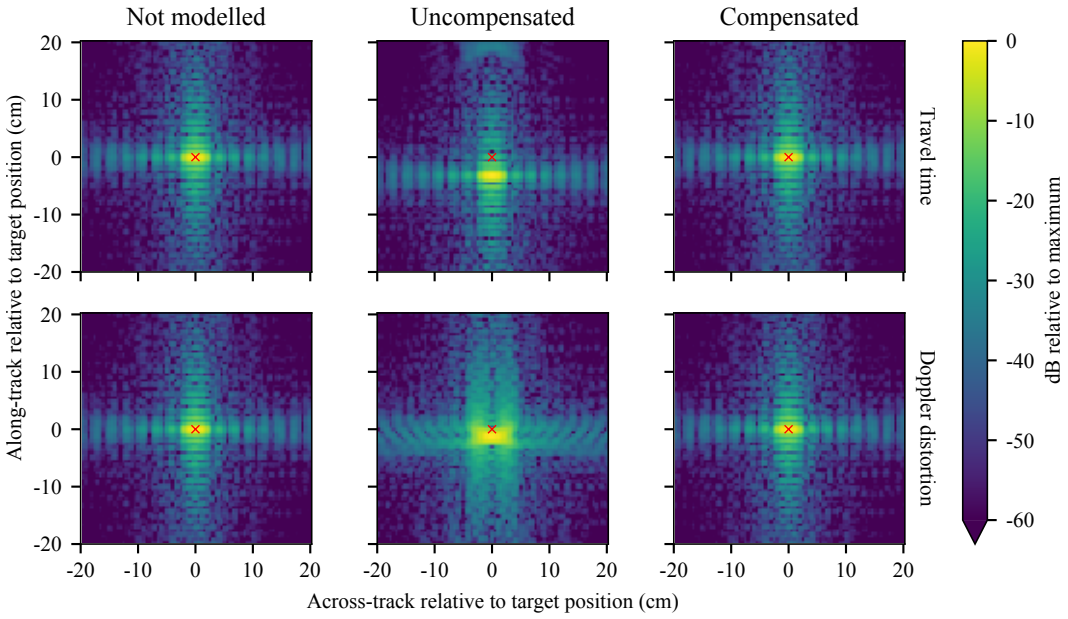


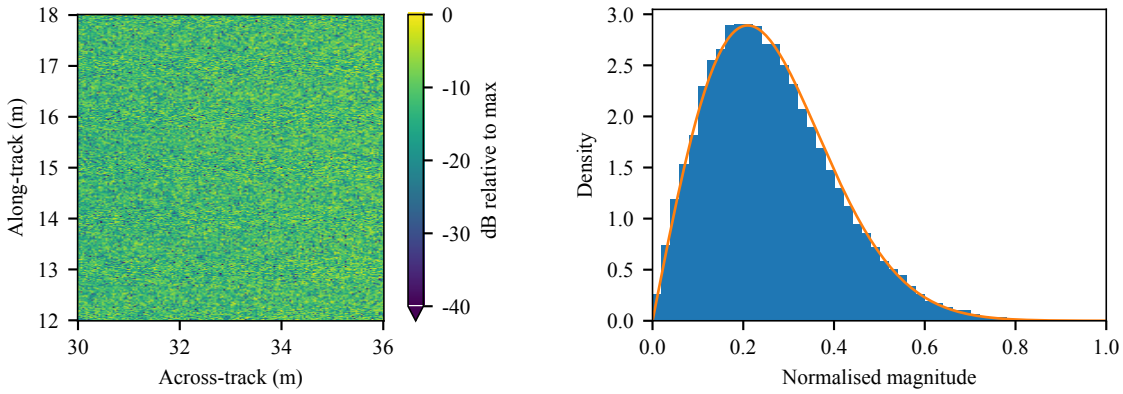
Figure 1: Point-spread function of images reconstructed from simulations of a single point target. The top row models the impact of the vehicle motion on the two-way travel time while the bottom row models the Doppler distortion of the signal. The left column has no motion effects, the middle column shows the impact of the uncompensated motion, and the right column applies compensation during reconstruction. The axes show position relative to the target location which is marked with a red cross.

5.1. INTRA-PING MOTION

It is not uncommon to use the stop-and-hop assumption that the vehicle is stationary for the duration of a ping and then instantaneously moves to the position of the next ping. From a simulation viewpoint, this both simplifies the computation of the two-way travel time of an acoustic pulse, and removes a source of distortion which may otherwise complicate the analysis of the results if the intra-ping motion is not relevant. However, in some cases it is desirable to model the continuous movement of the vehicle during ping to give a more realistic result.

One of the included travel time plugins supports finding the two-way travel time taking motion into account. The velocity at the time of transmission is used to approximate the receive position, and then the time of reception is iteratively adjusted until the time taken for the sonar to move to the proposed receive position matches (within some tolerance) the time the sound would take to move along the corresponding path to the target and back to that receive position. The upper row of Figure 1 shows (from left to right) the point-spread function (PSF) of a single target with the stop-and-hop approximation, the PSF with this iterative travel-time plugin in use and is not compensated during imaging, and the PSF if the motion is compensated during imaging. The intra-ping motion shifts the target, and the compensation corrects the shift.

The movement of the sonar also causes a Doppler shift to the signal. One of the plugins models this following the wide-band approach of [3] to calculate a scale factor η and modify the spectrum of the echo from each target accordingly. The lower row of Figure 1 shows a similar set of PSFs (Doppler not modelled, uncompensated and compensated). The distortion consists of both a shift and a blurring of the PSF, and can be compensated during imaging.



(a) *(b)*

Figure 2: (a) The intensity image and (b) the magnitude histogram of a piece of the flat, bland seafloor modelled by idealised point targets. The overlaid line in (b) is a Rayleigh distribution fit to the data.

5.2. BLAND SEAFLOOR

It is important to ensure that the available resources are being fully utilised during a simulation. To check this, a patch of bland seafloor was modelled by a rectangle measuring 30 m in along-track and 50 m in across-track containing uniformly distributed point targets at a density of 10 000 per m^2 or 3.44 per square wavelength at the centre frequency. The system followed a 30 m long trajectory at a 10 m height over ground along the edge of the rectangle. 66 pings were simulated, spaced 45 cm apart and each recording a trace of duration 0.25 s.

Five nodes of the HSUper cluster at Helmut Schmidt University were used to run the simulation. Each node had two Intel Xeon Platinum 8360Y processors and 256 GB of RAM, yielding a total of 360 CPU cores. The simulation, consisting of 2310 traces from each of the 15 million targets, was completed in 1 day and 20 hours of wall time using almost 690 days of CPU time. This successfully saturated the available cores, indicating the simulation controller and Dask scheduler were not causing bottlenecks. Figure 2 shows a section of an image reconstructed from the simulated data. This image exhibits the expected fully-developed speckle with the histogram of the magnitudes having the corresponding Rayleigh distribution.

6. FUTURE PLANS

Although point target modelling is useful in many cases, it is intended to add a facet-based simulation method. This would allow scattering models based on the target properties and the incident and scattering angles to be attached to each target in a scene. The shadows cast by objects in the scene could also be modelled with such a simulator, for example by following the methods in [4]. Further extensions, such as to model double-scattering within the scene, could also be investigated. Other simulation methods, for example the recently-published Fourier-domain wavefield rendering [5], could be considered in the future.

At this stage, the simulations only use CPU processing. It is desired to extend it to use available GPUs. This would include investigating how to optimally use a mixture of GPU and CPU resources, such as in a single workstation or in a cluster with many CPU nodes and few GPU nodes.

Some investigation into how the simulator behaves with very large scenes is required. This may necessitate some changes in how a simulation is divided into tasks to avoid memory problems. On a related note, the limits of the Dask task scheduler have not yet been explored. Each task requires a scheduler overhead of around 200 µs to handle [6]. Above a certain number of cores (dependent on the duration of the tasks), the scheduler will be saturated and cause a bottleneck. This upper limit is likely to be in the thousands of cores. It is suspected that this will not be a problem for most use-cases. Increasing the number of usable cores would require either increasing the size of each task (which has its own limits, for example, the available memory on a node) or moving to a custom scheduling solution, such as direct MPI communication.

It is hoped that plugins will be contributed by the community. Plugins which are likely to be widely useful and straightforward to maintain may be accepted into the main repository. It is intended to create an organisation to store and list user-contributed plugins that are not added to the main project (for those familiar with the Sphinx documentation generator, similar to the `sphinx-contrib` group for its community-provided addons).

The overall project (open sonar toolboxes) has been designed to be able to expanded to projects other than simulation. For example, a toolbox implementing common image reconstruction methods could be created, or one which performs post-processing operations such as autofocus. These would not be required to be in Python, and could serve both as useful starting points for new research and in benchmarks for comparing results between different systems or algorithms. Community involvement in this, as well as improvements to the simulation framework, would be greatly welcome.

ACKNOWLEDGEMENTS

Computational resources (HPC cluster HSUper) have been provided by the project hpc.bw, funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr. dtec.bw is funded by the European Union – NextGenerationEU.

REFERENCES

[1] Scientific Python. *SPEC 0 — Minimum Supported Dependencies*. URL: `https://scientific-python.org/specs/spec-0000/`.

[2] M. A. Ainslie and J. G. McColm. “A simplified formula for viscous and chemical absorption in sea water”. *The Journal of the Acoustical Society of America*, vol. 103, no. 3, 1998, pp. 1671–1672.

[3] D. W. Hawkins and P. T. Gough. “Temporal Doppler effects in SAS”. *Proceedings of the Institute of Acoustics: Sonar Signal Processing*, vol. 26, no. 5, 2004.

[4] B. Thomas, C. Sanford, and A. J. Hunter. “Occlusion Modeling for Coherent Echo Data Simulation: A Comparison Between Ray-Tracing and Convex-Hull Occlusion Methods”. *IEEE Journal of Oceanic Engineering*, vol. 49, no. 3, 2024, pp. 944–962.

[5] C. J. Sanford, B. W. Thomas, and A. J. Hunter. “Fourier-Domain Wavefield Rendering for Rapid Simulation of Synthetic Aperture Sonar Data”. *IEEE Journal of Oceanic Engineering*, vol. 49, no. 4, 2024, pp. 1501–1515.

[6] Dask. *FAQ - Dask documentation*. URL: `https://docs.dask.org/en/stable/faq.html` (visited on 05/27/2025).

A First Comparison of GANs and Diffusion Models for Generating Sidescan Sonar Images

Yannik Steiniger[1] and Benjamin Lehmann[2]

[1]German Aerospace Center, Institute for the Protection of Maritime Infrastructures, 27572 Bremerhaven, Germany
[2]City University of Applied Sciences Bremen, 28199 Bremen, Germany

Contact author: Yannik Steiniger, Fischkai 1, 27572 Bremerhaven, Germany, yannik.steiniger@dlr.de

Abstract: *We investigate generative deep learning models, namely a R3GAN and a diffusion model (DM), for the generation of synthetic sidescan sonar images. Both models are able to generate snippets with realistic object and shadow geometries as well as pixel distributions. Since sonar data from specific objects is scarce, the synthetic images are used to augment a training dataset for object classification. This augmentation can increase the performance of a shallow convolutional neural network (CNN) as well as a state-of-the-art vision transformer. In our analysis, images from R3GAN show a higher variability but images from DM lead to a better performance of the classification models improving the balanced accuracy of the CNN by 12%.*

Keywords: *Deep Learning, Generative Models, Diffusion Models, Synthetic Sonar Data, Sonar Imagery*

1. INTRODUCTION

With the analysis of sonar images being increasingly shifted towards the use of deep learning models the need for sufficient training datasets grows. Compared to other computer vision applications, like autonomous driving, there exists no large, publicly available dataset of sonar images for a broad range of object classes. Furthermore, gathering and annotating real world data with a high degree of variability is cumbersome and expensive. It requires specialized equipment as well as trained staff. Thus, research on generating synthetic sonar images gained more attention in recent years. Applications range from highly realistic but computational slow models based on the wave equation and ray tracing [1] over Fourier-domain wavefield rendering approaches [2] to faster but less transparent deep learning based generative models, such as generative adversarial networks (GAN) [3, 4]. In the past few years diffusion models (DM) [5] have been established as another standard in the computer vision domain for generating synthetic images, especially to generate natural RGB images. This alternative approach to GANs does not suffer from training instabilities, like mode collapse, and can produce more detailed images, but typically requires more training data. DMs have recently been applied to enhance the resolution of sonar images which led to an improved classification performance [6].

In this work we, for the first time, train a GAN as well as a DM to generate realistic sidescan sonar images of multiple object classes. We consider the classes *Tire*, *Cylinder* and *Wreck*. For these object classes less than 30 training samples are available. The aim of these synthetic images is to augment a dataset of real sonar images for the training of a classifier. We use a basic convolutional neural network (CNN) as well as a modern vision transformer (ViT) as classifiers. In our study we compare the images generated by the GAN and DM based on qualitative measures, e.g., realistic object and shadow shapes, and on their ability to increase the performance of the classifiers when being used to augment the training dataset. Furthermore, we investigate the variability of the methods by successively increasing the portion of synthetic images in the training dataset and using a clustering based approach.

2. METHODS AND MATERIALS

2.1. SIDESCAN SONAR DATASET

The sidescan sonar images used to train and evaluate the generative models as well as the classification models were collected during several measurement campaigns between 2019 and 2020. An *Edgetech 2205* system operating at 850 kHz mounted on an *ATLAS ELEKTRONIK SeaCat* AUV was used to gather this data. The sonar data was processed with an experimental processing chain to form grayscale intensity images. Snippets of relevant objects were manually extracted to form the dataset. Similar to previous work [7] this dataset consists of the classes *Tire*, *Rock*, *Cylinder*, *Wreck* and *Background*. In Table 1 the number of samples per class in the training and test dataset are listed. As it can be seen from the number of snippets, the classes *Tire*, *Cylinder* and *Wreck* are underrepresented. These will be the objects for which the generative models should generate additional synthetic data. Since multiple snippets of the same object exist, the training and test split was setup such that snippets from the same object are only in one dataset. Due to the limited amount of data, snippets from the three underrepresented classes were split nearly 50/50 while adhering to the aforementioned constrain.

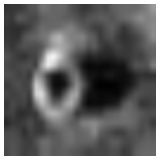
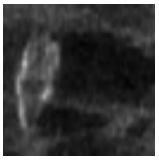
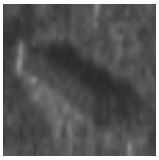
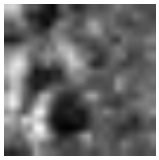
Object class	*Tire*	*Rock*	*Cylinder*	*Wreck*	*Background*
Number of training snippets	18	500	22	12	500
Number of test snippets	17	719	19	12	719
Example image					

Table 1: Overview of the sidescan sonar dataset.

2.2. GAN MODEL

A GAN is a deep learning model which can be used to generate synthetic images based on a provided training dataset. It consists of two sub-models, the generator G and the discriminator D, which are typically implemented as a CNN and trained in an adversarial manner. On the one hand, the discriminator learns to distinguish between images from the dataset and those from the generator. On the other hand, the generator is trained to transform a noise vector z into images similar to the training samples x in order to fool the discriminator. In general, this training procedure can be described by the loss function

$$\mathcal{L} = \mathrm{E}_{z\sim p_z} - \log\left(1 + \exp\left(-D(G(z))\right)\right) + \mathrm{E}_{x\sim p_{data}} - \log\left(1 + \exp\left(-D(x)\right)\right), \tag{1}$$

which the discriminator tries to maximize and the generator to minimize. Here p_z and p_{data} are the distributions of the noise and training data, respectively. A common problem when training a GAN is the so-called mode collapse. This is a state of the GAN in which the generator outputs the same image regardless of the input noise. Different loss functions, like the Wasserstein loss [8], or other training strategies like specific pre-training [9] have been developed to deal with the training difficulties of GANs.

In this work we use the state-of-the-art R3GAN [10] for generating synthetic sonar images. R3GAN uses a regularized relativistic GAN loss to over-come the training difficulties. This loss combines a relativistic pairing loss [11]

$$\mathcal{L} = \mathrm{E}_{z\sim p_z, x\sim p_{data}} - \log\left(1 + \exp\left(D(G(z)) - D(x)\right)\right) \tag{2}$$

with the zero-centred gradient penalties

$$R_1 = \frac{\gamma}{2}\mathrm{E}_{x\sim p_{data}} \left\|\Delta_x D\right\|^2 \tag{3}$$

and

$$R_2 = \frac{\gamma}{2}\mathrm{E}_{x\sim p_G} \left\|\Delta_x D\right\|^2 \tag{4}$$

where γ is a hyperparameter controlling the degree of regularization. The relativistic paring loss directly couples the real data $x \sim p_{data}$ and synthetic images $x \sim p_G$ in the training process while R_1 and R_2 regularize the gradient norm of D on the real and synthetic images, respectively. Both regularization term should ensure the convergence of the model. Furthermore, the

networks of D and G are designed based on a modern ResNet architecture. Both networks consist of residual blocks and are setup symmetrically meaning that the generator upsamples the input through the resolutions 4×4, 8×8, 16×16 and 32×32 while the discriminator downsamples the image through the reverse order of resolutions. For a detailed overview about the architecture see [10]. We train R3GAN with its default parameters for 7813 iterations, which is equivalent to 250,000 images.

2.3. DENOISING DIFFUSION PROBABILISTIC MODELS

Denoising Diffusion Probabilistic Models (DM) are a class of generative models to synthesize data, given a two-staged process. In a forward diffusion process a *clean* input image x_0 taken from the training dataset is gradually disturbed by adding Gaussian noise. After completing consecutive noising steps, also referred to as timesteps, the image content approaches an isotropic Gaussian noise while the information is lost in noise completely. In the reverse process, a neural network (typically a U-Net architecture) is trained to remove the added noise to unveil the original data or at least data that look similar to the original one. With this approach the network was supposed to learn the underlying pixel distribution in the datasets. The term pixel distribution refers to the spatial dependencies among neighbouring pixels (up to the complete image domain), rather than the statistics of individual pixels in isolation. The loss function to be minimized is the mean squared error (MSE) between the true noise ϵ and the predicted noise ϵ_θ added at timestep t

$$\mathcal{L} = \sum_t \mathrm{E}_{\epsilon \sim \mathcal{N}(0,I)} \|\epsilon - \epsilon_\theta(x_t, t)\|^2, \tag{5}$$

where I is the identity matrix, x_t is the disturbed image and t is uniformly sampled from 1 to T, the maximum number of steps. The generation of synthetic images starts with isotropic Gaussian noise and all pixels not obeying the learned pixel distribution are removed respectively denoised.

The standard DM loss functions treat all timesteps equally, which may lead to bias towards early learning steps and tend to overfit to high signal-to-noise (SNR) examples, which is of special interest for sonar images which might suffer from noise. Therefore the P2-weighting was introduced [12, 13], which formulates a loss re-weighting function $w(t)$ based on the signal-to-noise ratio. It was proven to neglect early (less noisy) steps and promotes learning in later (more noisy) steps. Following widely adopted implementations the explicit weighting scheme

$$\mathcal{L}_{P2} = \sum_t w(t) \mathrm{E}_{\epsilon \sim \mathcal{N}(0,I)} \|\epsilon - \epsilon_\theta(x_t, t)\|^2, \tag{6}$$

with

$$w(t) = \left(\frac{1}{1 + \mathrm{SNR}(t)} \right)^\alpha \tag{7}$$

is applied, where α is a hyperparameter, commonly set as $\alpha \in [1.0, 2.0]$. In all our experiments the training was started from scratch, where for each object class $\in \{$*Tire*, *Cylinder*, *Wreck*$\}$ one individual network was trained and used afterwards to generate synthetic object images.

Exp.	Tire		Cylinder		Wreck		Rock		Background	
	real	synth.	real	synth.	real	synth.	real	synth.	real	synth.
1	18	0	22	0	12	0	36	0	36	0
2	18	18	22	14	12	24	36	0	36	0
3	18	32	22	28	12	38	50	0	50	0
4	18	82	22	78	12	88	100	0	100	0
5	18	482	22	478	12	488	500	0	500	0

Table 2: Number of real and synthetic snippets per class in each experiment.

3. EXPERIMENTAL SETUP

The main purpose of generating synthetic sonar snippets is to augment a training dataset for a classification model at hand. Thus, the quality of the generated images will be measured indirectly in terms of performance improvements of the classification model. Especially, we are interested in how many synthetic snippets can be generated and added to the training dataset to still see noticeable improvements. In Table 2 the five experiments conducted in this work are listed, with increasing number of training snippets, where *Experiment 1* serves as the baseline without any synthetic data added. Based on previous work [9] we reduce the number of rock and background snippets to be 36 in order to balance the dataset. In *Experiment 2* we add synthetic snippets to the classes *Tire*, *Cylinder* and *Wreck* so that every class has 36 training snippets in total. Since this number is still very low compared to conventional deep learning dataset, we successively increase the number of real rock and background snippets as well as synthetic tire, cylinder and wreck snippets through *Experiments 3 – 5*. The total number of snippets per class is 50, 100 and 500, respectively.

We train two different classification models on the datasets described in Table 2, namely a shallow CNN used in previous work [7] and a state-of-the-art vision transformer (ViT) [14]. Our shallow CNN consists of three convolutional layers with 8, 16 and 32 kernel of size 3×3. The output of a convolutional layer is passed through a ReLU activation function, batch normalization and 2×2 max pooling. Features from the last convolutional layer are compressed using a fully connected layer with 100 neurons prior to the final output layer with five neurons. Dropout is added before both fully connected layers. All input images are scaled to match the input size of the network, which is 64×64 pixel. We train this CNN without any pre-training for 50 epochs with the Adam optimizer and a learning rate of 10^{-4}, which is reduced by a factor of 0.1 after every 20 epochs.

As an alternative method to verify the gain in performance we employ the original vision transformer with little modifications to account for sonar images and training data used. The core innovation in using ViT architectures is the replacement of traditional CNN architectures with transformer encoder architectures, well known from natural language processing. This milestone introduced a new performance area for image classification tasks. The image under consideration is split into fixed-size patches, flattened and embedded linearly into a vector. A token is added to represent the image's overall classification and processed by a transformer encoder stack, consisting of multi-head self-attention and feedforward layers, to be forwarded to a multi-layer perceptron head for final classification. In our experiments we made use of the ViT base model (ViT-B) with patch sizes of 16×16 pixels, with pre-trained weights from the ImageNet-1K dataset. For the different sonar datasets with varying cardinality in our experiments, we achieved comparable performances by employing 5 to 25 epochs, using the Adam

optimizer and a learning rate of 10^{-5}. In the following we report the performance for the model trained for 20 epochs. In addition to that we reduced number of neurons in the classification MLP to 100 and the number of classification classes to 5 to match the sonar dataset.

We evaluate the classification performance of both models in all five experiments using the balanced accuracy

$$ACC_{\text{bal}} = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \text{recall}(c) \tag{8}$$

and the macro F1-score

$$F1_{\text{macro}} = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} F1(c) = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \frac{2 \cdot \text{precision}(c) \cdot \text{recall}(c)}{\text{precision}(c) + \text{recall}(c)}. \tag{9}$$

Both metrics take an average over the different classes $\mathcal{C} = \{$*Tire*, *Rock*, *Cylinder*, *Wreck*, *Background*$\}$ and are thus suitable for the unbalanced test dataset. Note that neither during training of the generative models nor during training of the classification models images from this test dataset have been used.

To further study the data generated by the GAN and the DM we investigate the diversity in the synthetic images. For this we use a clustering based approach [7]. The clustering approach starts by selecting one generated image as the center of a cluster. Images whose pixel-wise Euclidean distance to this initial image is below a given threshold will be grouped to this cluster. In the next iteration an image that has not been assigned to a cluster serves as the center point of the next cluster and the process is repeated with all unassigned images. Finally, the number of clusters determines the diversity of the synthetic images, with a large number of clusters relating to a high diversity.

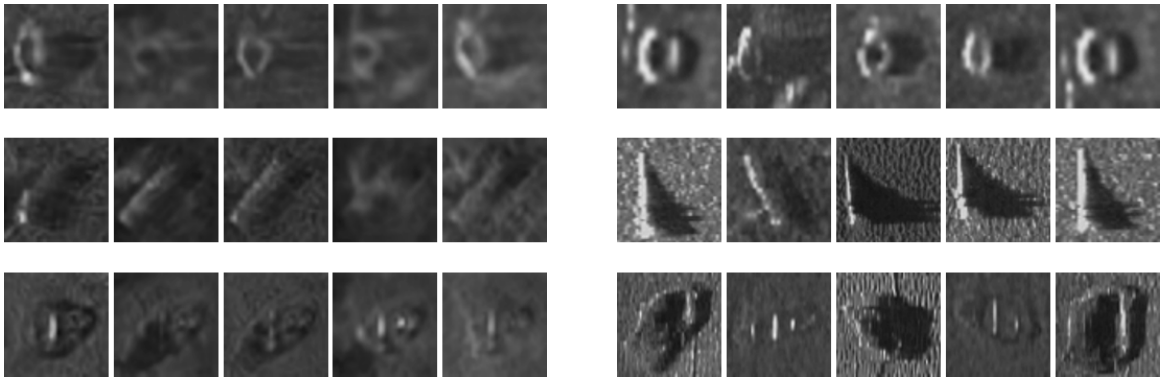
4. RESULTS

4.1. GENERATED IMAGES

Before investigating the influence on the classification performance we analyse the synthetic images generated by R3GAN and DM quantitatively. Figure 1 shows five random examples per class from both generative models. Both methods succeed in generating the required object shapes and realistic shadow geometries. The contrast in the images generated by DM is higher than the one in the images from R3GAN. Both models generate different snippets and do not suffer from any form of mode collapse. All cylinders generated by R3GAN in Figure 1 (a) have the same orientation, which is also over-represented in the remaining generated images. The training dataset, however, also contains examples with a different orientation (see Table 1). Thus, R3GAN seems not to be able to cover this variability. The images generated by DM show a strong similarity to the training images, e.g., the first generated wreck is similar to the example shown in Table 1.

4.2. CLASSIFICATION PERFORMANCE

Table 3 lists the classification performance measured in balanced accuracy and macro F1-score for the five experiments and images generated by R3GAN and DM. Note that since the



(a) *(b)*

Figure 1: Synthetic snippets generated with (a) R3GAN and (b) DM.

Exp.	Generator	CNN		ViT	
		ACC_{bal}	$F1_{\text{macro}}$	ACC_{bal}	$F1_{\text{macro}}$
1	-	0.493±0.033	0.423±0.015	0.287	0.348
2	R3GAN	0.506±0.044	0.440±0.018	0.249	0.216
3	R3GAN	0.531±0.031	0.457±0.023	0.322	0.249
4	R3GAN	0.528±0.032	0.461±0.021	0.402	0.331
5	R3GAN	0.443±0.047	0.395±0.016	0.408	**0.465**
2	DDPM	**0.559**±0.055	0.445±0.029	0.290	0.320
3	DDPM	0.534±0.034	0.460±0.022	0.293	0.493
4	DDPM	0.552±0.029	**0.469**±0.017	0.289	0.277
5	DDPM	0.481±0.050	0.430±0.033	**0.442**	0.440

Table 3: Classification performance of the CNN and ViT in the five experiments with increasing number of synthetic snippets. Best value marked in boldface.

capacity of the shallow CNN is too small to be pre-trained on ImageNet-1K it is initialized with random weights. Thus, we initialize and train the shallow CNN ten times and report the mean and standard deviation of the metrics in Table 3. *Experiment 1* serves as the baseline performance without any synthetic data. Through *Experiment 2* to *5* the number of synthetic images is successively increased (see Table 2). Except for *Experiment 5*, which adds over 470 synthetic samples per class, augmenting the training dataset leads to an improvement in balanced accuracy and macro F1-score for the shallow CNN. With images generated by R3GAN the performance even decreases in *Experiment 5* compared to the baseline. In contrast to this, the ViT required a large training dataset and only shows a clear improvement in balanced accuracy and macro F1-score in *Experiment 5*. For the shallow CNN images from DM lead to a better performance in the individual experiments than images from R3GAN. Also, for most experiments the images from DM lead to a better performance of the ViT. The best overall performance is achieved with the shallow CNN in *Experiment 4* with 100 samples per class and synthetic images generated by the DM. Adding the synthetic snippets increases the balanced accuracy by 12% and the macro F1-score by 11%.

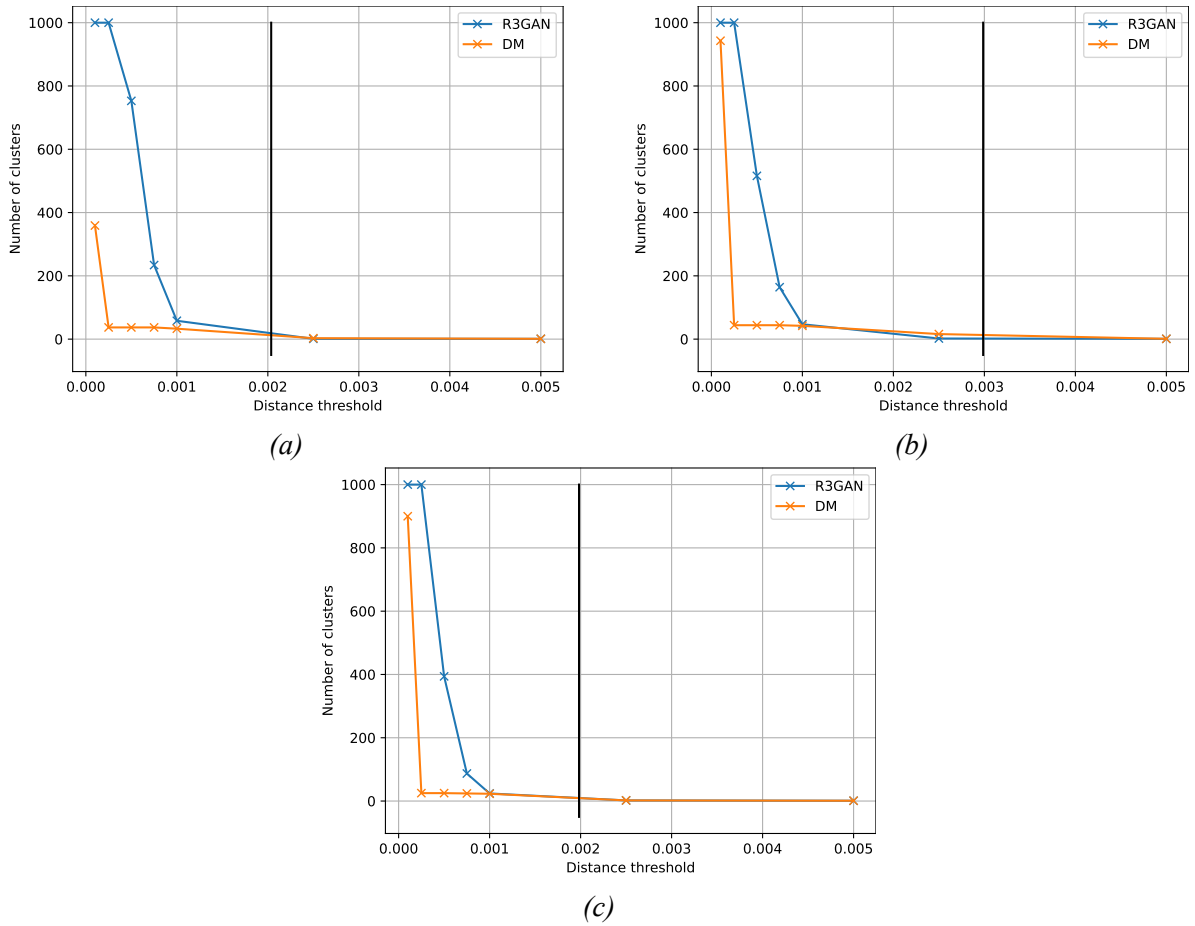


Figure 2: Number of clusters of the synthetic snippets for class (a) Tire, (b) Cylinder and (c) Wreck.

4.3. DIVERSITY ANALYSIS

As described at the end of Section 2 we investigate the variability of the generative models by clustering. We generate 1,000 images of the classes *Tire*, *Cylinder* and *Wreck* with both models and perform the clustering separately for each class. Figure 2 shows the number of clusters for a varying distance threshold. Naturally, as the threshold decreases the number of clusters grows. The black line indicates the mean distance across all 1,000 images for each class as a reference. Based on the number of clusters R3GAN shows a higher variability than DM. Already for a low threshold of $2.5 \cdot 10^{-4}$ the number of clusters is below 50 while for R3GAN each image is still its own cluster. For the class *Tire* and a threshold of 0.001 R3GAN and DM produce 58 and 33 clusters, respectively, which improves over previous results of a transfer-learned GAN with 21 clusters [9].

5. SUMMARY

In this work we have, for the first time, compared a GAN and a DM for the generation of synthetic sidescan sonar images. Both methods are able to create realistic highlight and shadow geometries of different object classes. Augmenting a training dataset for classification with the synthetic snippets leads to a performance improvement of a shallow CNN as well as a ViT.

The better performance is achieved with images from DM which can increase both balanced accuracy and macro F1-score by more than 10% for the CNN and more than 25% for the ViT. We anticipate an improvement, though not as significant, with a more comprehensive dataset that is more representative for a classification problem. We have also analysed the variability of the generated images, which is higher for R3GAN. The better quality and contrast of the images generated by the DM can be a reason for the better results in the classification analysis.

REFERENCES

[1] J. M. Bell, L. M. Linnett: "Simulation and analysis of synthetic sidescan sonar images", *IEE Proceedings - Radar, Sonar and Navigation* **144(4)**, **(1997)**.

[2] C. J. Sanford, B. W. Thomas, A. J. Hunter: "Fourier-domain wavefield rendering for rapid simulation of synthetic aperture sonar data", *IEEE Journal of Oceanic Engineering* **49(4)**, **(2024)**.

[3] Y. Steiniger, J. Stoppe, T. Meisen, D. Kraus: "Dealing with highly unbalanced sidescan sonar image datasets for deep learning classification tasks" in *Global OCEANS 2020 MTS/IEEE Singapore - U.S. Gulf Coast*, (virtual, 2020).

[4] M. Jegorova, A. I. Karjalainen, J. Vazquez, T. M. Hospedales: "Unlimited Resolution Image Generation with R2D2-GANs" in *Global OCEANS 2020 MTS/IEEE Singapore - U.S. Gulf Coast*, (virtual, 2020).

[5] J. Ho, A. Jain, P. Abbeel: "Denoising Diffusion Probabilistic Models" in *Advances in Neural Information Processing Systems*, **33**, **(2020)**.

[6] O. Bryan, T. Berthomier, B. D'Ales, T. Furfaro, T. S. F. Haines, Y. Pailhas, A. Hunter: "A diffusion-based super resolution model for enhancing sonar images" *Journal of the Acoustical Society of America* **157(1)**, **(2025)**.

[7] Y. Steiniger, J. Stoppe, D. Kraus, T. Meisen: "On the detection and classification of objects in scarce sidescan sonar image dataset with deep learning methods" in *7th Underwater Acoustic Conference and Exhibition*, (Kalamata, 2023).

[8] M. Arjovsky, S. Chintala, L. Bottou: "Wasserstein generative adversarial networks" in *Proceedings of the 34th International Conference on Machine Learning*, (Sydney, 2017).

[9] Y. Steiniger, D. Kraus, T. Meisen: "Generating synthetic sidescan sonar snippets using transfer-learning in generative adversarial networks", *Journal of Marine Science and Engineering* **9(3)**, **(2021)**.

[10] Y. Huang, A. Gokaslan, V. Kuleshow, J. Tompkin: "The GAN is dead; long live the GAN! A modern baseline GAN" in *The 38 Annual Conference on Neural Information Processing Systems*, (Vancouver, 2024).

[11] A. Jolicoeur-Martineau: "The relativistic discriminator: a key element missing from standard GAN" in *International Conference on Learning Representations*, (New Orleans, 2019).

[12] J. Choi, J. Lee, C. Shin, S. Kim, H. Kim, S. Yoon: "Perception Prioritized Training of Diffusion Models" in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (New Orleans, 2022).

[13] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S. Ghasemipour, B. Ayan, S. Mahdavi, R. Lopes, T. Salimans, J. Ho, D. Fleet, M. Norouzi: “Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding”, (arXiv:2205.11487). arXiv. https://doi.org/10.48550/arXiv.2205.11487, **(2022)**.

[14] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby: “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”, (arXiv:2010.11929). arXiv. https://doi.org/10.48550/arXiv.2205.11487, **(2010)**.

Transformers and Ensembles for Object Detection in Sidescan Sonar Images

Yannik Steiniger[1]

[1]German Aerospace Center, Institute for the Protection of Maritime Infrastructures, 27572 Bremerhaven, Germany

Contact author: Yannik Steiniger, Fischkai 1, 27572 Bremerhaven, Germany, yannik.steiniger@dlr.de

Abstract: *Transformer models have evolved as the state-of-the-art for computer vision tasks, such as image classification and object detection. For sonar images, they recently have been applied for classifying snippets of synthetic aperture sonar images. In this work, deep learning models for object detection based on standard convolutional neural networks (CNN) and on transformers are investigated and compared. The Retina-SWIN transformer model achieves a true-positive rate of up to 99.5%. However, all investigated deep learning detectors suffer from a high number of false alarms. Thus, the models are incorporated into two-step pipelines and ensembles. With both setups the number of false alarms is reduced. However, the detection ensembles show a lower true-positive rate than the two-step pipeline. The best overall model is Retina-SWIN, which is trained for localisation of objects, combined with a custom CNN, which classifies the extracted snippets. It reaches a maximum true-positive rate of 94.8% at 33.4 false alarms, which is a reduction of over 80% compared to the baseline Retina-SWIN.*

Keywords: *Deep Learning, Automatic Target Recognition, Sonar Imagery, Sidescan Sonar, Transformer*

1. INTRODUCTION

Being able to observe a scene under water is crucial for the security of maritime infrastructures. Due to physical limitations of optical systems, imaging sonar systems, such as sidescan or synthetic aperture sonar, which utilise acoustic instead of electro-magnetic waves, are commonly used to survey large seafloor areas. These sensors are typically mounted on an autonomous underwater vehicle, which follows a predefined path to scan the requested area. Analysing the captured data automatically is essential to increase the autonomy of a holistic surveying system. Nowadays, deep learning methods like convolutional neural networks (CNN) are primarily used for computer vision tasks relevant for the analysis of sonar images, such as classification, detection and segmentation [1, 2, 3, 4]. Recently, vision transformers have emerged as the state-of-the-art in many computer vision tasks and their application to the classification of sonar images has been investigated as well [5, 6]. However, transformer-based models for detecting objects in sonar images have not been analysed yet.

In this work, CNNs and transformer models are compared for the detection of objects in sidescan sonar images. As representatives for the CNN methods one-stage as well as two-stage detectors are selected. The transformer-based methods investigated here are Retina-SWIN [7] and Deformable DETR (DDETR) [8], which use the attention mechanism in the network backbone or detection head, respectively. In addition, all detectors are used in two different two-step pipelines, which further process predictions with another CNN to reduce the number of false alarms [9]. Furthermore, the combination of multiple different object detectors into an ensemble is studied. For classification, such ensembles of CNNs have shown to improve the overall performance [1]. The results show that the transformer-based method Retina-SWIN can reach a true-positive rate of 99.5% but at the cost of a high false alarm rate. Processing the detections with a CNN in the two-step pipeline significantly reduces the number of false alarms per image but also negatively impacts the true-positive rate. The best trade-off is achieved with a Retina-SWIN model, which is trained for localisation and a subsequent CNN for classification, reaching a maximum true-positive rate of around 95% at slightly over 30 false alarms per image. An ensemble of different object detectors can increase the detection performance but at the cost of an increased number of false alarms.

The remainder of the paper is organised as follows. First, Section 2 provides a brief introduction into CNN- and transformer-based object detection. Afterwards, the experimental setup is described in Section 3. In Section 4 the results of the experiments are presented and discussed. Finally, the paper closes with a summary in Section 5.

2. BACKGROUND

2.1. CNN-BASED DETECTORS

Generally, CNN-based detectors can be divided into two categories: one-stage and two-stage detectors. A two-stage detector, like CenterNet2 [10], predicts in the first stage a large number of region of interests (ROI) and filters, classifies and refines these ROIs in the second stage. In contrast to this, a one-stage detector, like YOLOv8 [11], directly predicts bounding boxes and object classes from an input image without intermediate ROIs. This typically makes one-stage detectors faster than two-stage detectors but also less accurate.

2.2. TRANSFORMER-BASED DETECTORS

A transformer is a deep learning architecture that has originated from natural language processing and that translates an input sequence into an output sequence via an encoder-decoder architecture and an attention mechanism [12]. The components of the attention mechanism are a set of key-value pairs $\mathbf{K} \in \mathcal{R}^{N_k \times N_l}$ and $\mathbf{V} \in \mathcal{R}^{N_v \times N_l}$ as well as queries $\mathbf{q} \in \mathcal{R}^{N_k \times 1}$, where N_l is the number of key-value pairs, e.g., the number of words in a sentence and N_k and N_v are the number of features of the keys, queries and values, respectively. $\mathbf{K}$, $\mathbf{V}$, and $\mathbf{q}$ are realized through a linear combination of the inputs to the attention module with learnable weight matrices and the attention is calculated as

$$\text{attention}(\mathbf{K}, \mathbf{V}, \mathbf{q}) = \text{softmax}\left(\frac{\mathbf{q}^{\mathrm{T}}\mathbf{K}}{\sqrt{N_k}}\right)\mathbf{V}^{\mathrm{T}}. \tag{1}$$

In computer vision models, this attention mechanism can either be used in the backbone of the model, which learns to extract general features from the input image, or in the head of the model, which performs the specific task based on these features, e.g., the prediction of bounding boxes and object classes. In the backbone, in order to be applied to an image, the input is split into patches, which form the input sequence of the encoder. The encoder applies the attention mechanism and outputs learned image features. When used in the detection head, attention is used in an encoder and decoder, where the encoder further transforms features from a backbone model and the decoder maps an input query sequence into predictions using these features.

3. EXPERIMENTAL SETUP

3.1. SIDESCAN SONAR DATASET

To train the object detectors as well as the classification models for the two-step approach sidescan sonar images from several sea trials are used [13, 9]. In these images objects from the four classes *Tire*, *Rock*, *Cylinder* and *Wreck* were annotated with manually defined bounding boxes tightly enclosing the object highlight and acoustic shadow. The detection dataset consists of the whole sidescan sonar images as input and the bounding box coordinates and class labels as target variables. For the classification dataset the snippets defined by the bounding boxes are extracted from the sonar images. In addition, random background snippets are selected to form a fifth *Background* class. The final classification dataset consists of the extracted snippets as input and the class labels as the target variable.

Images from the classes *Tire*, *Cylinder* and *Wreck* are very limited. Thus, the training and test split for both datasets is done such that these classes are roughly split 50:50. However, multiple images of the same objects, e.g., from different viewing angels, exist in the dataset. Thus, the data is split such that images from the same object are either in the training or test set. Additionally, if multiple objects are in the same sidescan sonar image, their associated snippets are all assigned to either the training or test set. This ensures that the same objects are used for training and testing the detectors and the classifier. These restrictions result in the number of samples in the training and test set for each class reported in Table 1.

Table 1: Overview about the datasets. Sidescan images are used for detection. Snippets from the classes Tire, Rock, Cylinder, Wreck and Background are used for classification.

Dataset	Number of sidescan images	tires	rocks	cylinders	wrecks	background
Training	769	24	2288	15	10	1390
Test	128	12	167	22	10	719

3.2. DETECTION PIPELINES

Besides the direct application of the deep learning detectors two additional pipelines previously introduced in [9] are used and shown in Figure 1. The purpose of these two-step pipeline is to reduce the number of false alarms. In the first two-step approach the deep learning detection models are used for pure localisation of the objects. To train these models all objects are considered as one general class *Target*. The classification is carried out in the second step by an additional CNN. In the second two-step pipeline the detections from the deep learning models are filtered by another CNN. This classifies the extracted snippets either into *Target* or *Background*. Detections whose corresponding snippets are classified as background will be filtered.

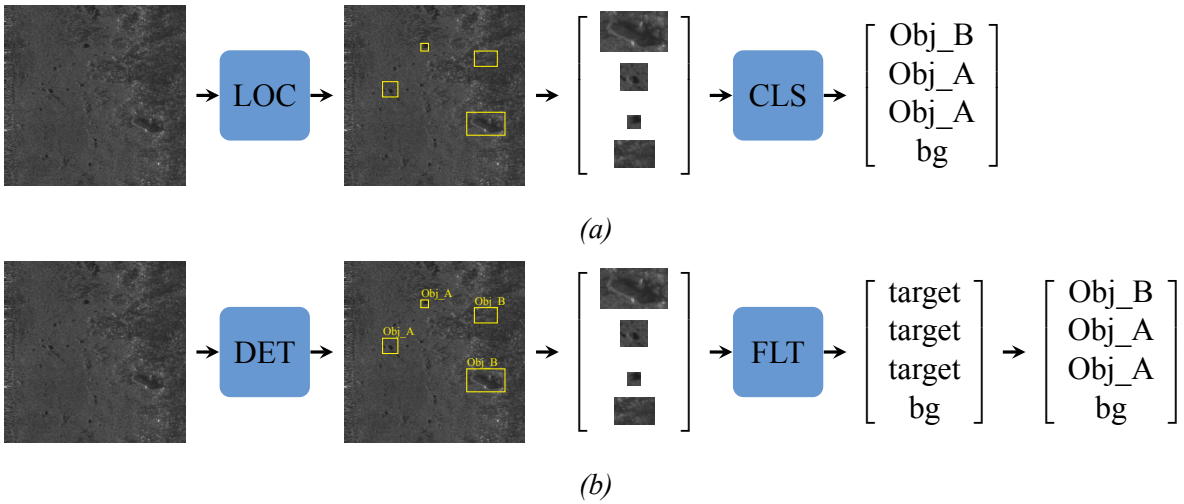


Figure 1: Two-step detection pipelines. (a) The detector is only used for localization (LOC) while a CNN classifies the extracted snippets in a subsequent step (CLS). (c) The detector (DET) predicts bounding boxes and object classes and another CNN carries out a binary classification (FLT) to reduce the number of false alarms by filtering background snippets.

In this study, the deep learning models YOLOv3 [14], YOLOv8 [11], CenterNet2 [10], Retina-SWIN [7] and DDETR [8] are considered. The models are initialized with pre-trained weights based on the MS COCO dataset. They are trained on the sonar data with their standard configuration using the MMDetection toolbox [15]. All models are trained for 100 epochs except Retina-SWIN which is trained for only 5 epochs to prevent performance degradation. As classification and filtering networks the same shallow architecture with three convolutional layers and training parameter as in [9] is used.

3.3. DETECTOR ENSEMBLING

Combining the output of multiple models into one ensemble for a final decision has shown to be beneficial for sonar image classification [1]. In this work an ensemble is build by combining the output of multiple different detectors. Based on the previous sections we consider three types of ensembles DET, LOC+CLS and DET+FLT, meaning that only outputs from within one detection pipeline are combined. In addition, the detections of the DET ensemble after ensembling are filtered with the CNN used in the filtering step of the DET+FLT pipeline. The combination of the individual predictions of the models in an ensemble is based on the number of detectors that predict the same object class at a similar position. Two detections are considered to belong to the same object if the predicted class matches and the intersection over union (IoU) of the bounding boxes exceeds a value of 0.5. If a pre-defined number of models n_{det} detects the same object it is counted as a detection of the ensemble. For the final prediction of the ensemble the bounding box with the highest prediction score is considered.

4. RESULTS

The detection performance in all setups is compared based on receiver operating characteristic (ROC)-like curves. In contrast to standard ROC curves for classification performances, here the average number of false alarms per image is given on the abscissa. In addition to the IoU, the pixel-wise Euclidean distance d between the center pixel of the true and predicted bounding box is considered to determine true-positive detections. Figure 2 displays the ROC-like curves of the five models for the three detection pipelines. As expected the two-step pipelines produce less false alarms but also reduce the number true-positive detections. Based on the IoU for determining true-positives, DDETR shows the worst performance of the five deep learning detectors. The performance is better if d is considered, which shows that the shape of the bounding boxes predicted by DDETR does not match the true ones. This is also the reason for the bad performance of DDETR in the DET+FLT pipeline, where snippets are extracted from the sonar image and feed to the filtering CNN based on the predicted bounding boxes. With badly defined bounding boxes parts of the object are missing in the snippet which leads to a wrong classification result. Retina-SWIN shows the best performance in this comparison achieving a maximum true-positive rate of 99.5%.

For a further quantitative analysis, Table 2 lists the maximum true-positive rate TPR_{max}, the false alarm rate corresponding to this value as well as the false alarm rate for true-positive rates of 0.9 and 0.5 of the individual detectors and pipelines. The deep learning model trained for localisation in the LOC+CLS pipeline is indicated with an -L. The CNN-based models show a higher TPR_{max} in the DET+FLT pipeline while the both transformer-based models perform better in the LOC+CLS setup. Retina-SWIN for localization and the CNN for classification achieves the best trade-off of a high TPR and a low number of false alarms with a TPR_{max} of 94.8% and 33.4 false alarms per image. If a low false alarm rate is preferable over a high TPR, YOLOv8 for localization and the CNN for classification gives the best result with only 0.45 false alarms per image at a TPR of 50%.

With the five deep learning models available, five different ensembles E1-E5 are investigated by successively combining more detectors. Based on their individual performance in the previous analysis, starting with only Retina-SWIN in E1, YOLOv3, YOLOv8, CenterNet2 and DDETR are subsequently added to form E2-E5. Figure 3 shows the performance metrics ex-

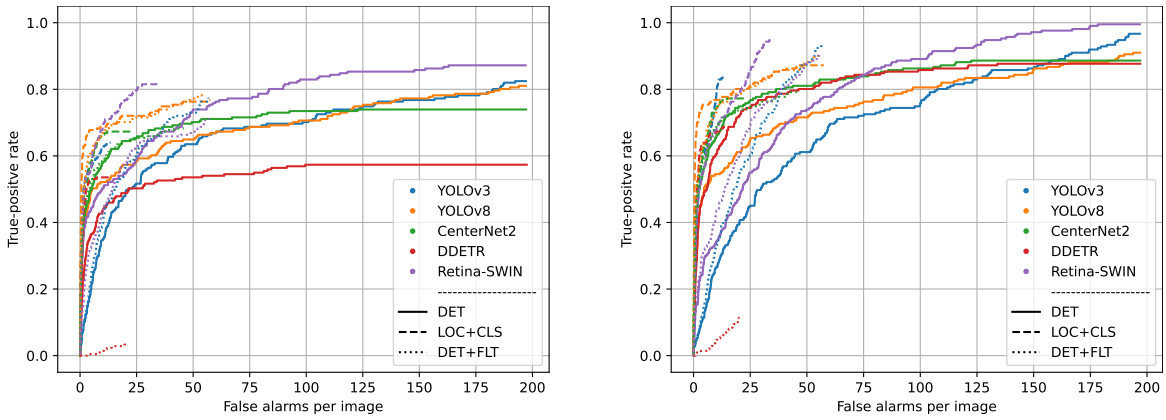


Figure 2: ROC-like curves of the detectors with (a) IoU=0.5 and (b) d=10 px.

tracted from the ROC-like curves for the ensembles. For each ensemble the minimum number of detectors that have to predict a specific detection in order to count as a detection of the ensemble n_{det} is varied from one to four. For example, with E3 and $n_{det} = 2$ at least two models of Retina-SWIN, YOLOv3 and YOLOv8 have to predict an object of the same class at the same position to generate a detection of this ensemble. The ROC-like curves with $d = 10$ px are used to extract the metrics in Figure 3. However, the plot for IoU=0.5 show the same behaviour. Figure 3 additionally shows the performance metrics of Retina-SWIN-L+CNN as a reference.

Increasing n_{det} reduces the maximum true-positive rate but also the number of false alarms. Enlarging the ensemble set while keeping n_{det} fixed increases the maximum true-positive rate but also the number of false alarms. Since Retina-SWIN already has a high TPR_{max}, increasing the set of models to form the ensemble while counting a detection if only one model outputs a prediction only increases the number of false alarms. An additional filtering of the detections from the ensemble slightly reduces the TPR_{max} but also drastically the number of false alarms. Creating ensembles from the LOC+CLS detection pipeline leads to a worse performance compared to the Retina-SWIN-L+CNN baseline. The number of false alarms can be reduced in this ensemble from 33.4 to 6.3 but the TPR_{max} drops below 75%. The same behaviour can be seen for the DET+FLT pipeline. However, here the best TPR_{max} for an ensemble with at least two matched detections is 78.2% at a false alarm rate of 13. Compared to the baseline, no ensemble configuration can increase the maximum true-positive rate and at the same time reduce the number of false alarms. The best ensemble is E3 with the DET+FLT pipeline and $n_{det} = 3$ which achieves a TPR_{max} of 87.7% at 31.4 false alarms per image.

5. SUMMARY

This work has compared CNN- and transformer-based models for object detection in side-scan sonar images. Retina-SWIN, which uses the attention mechanism in the backbone model to extract general features, achieves the highest true-positive rate of 99.5%. A two-step detection pipeline in which the detector is trained for localization only and the classification is carried out by an additional CNN significantly reduced the number of false alarms per image from 178 to 33 while keeping a high maximum true-positive rate of 94.8%. Furthermore, ensembles of the detectors are constructed by combining the predictions of different models. These ensem-

Method	TPR_{max}	**False alarms**		
		@ TPR_{max}	@ TPR=0.9	@ TPR=0.5
YOLOv3	0.967	192.016	163.891	29.698
YOLOv8	0.910	193.837	187.651	4.636
CenterNet2	0.886	122.450	-	2.504
DDETR	0.877	133.961	-	5.860
Retina-SWIN	**0.995**	178.295	103.504	21.643
YOLOv3-L+CNN	0.839	12.884	-	2.240
YOLOv8-L+CNN	0.872	45.636	-	**0.450**
CenterNet2-L+CNN	0.773	14.674	-	1.519
DDETR-L+CNN	0.687	**13.054**	-	1.589
Retina-SWIN-L+CNN	0.948	33.403	**27.876**	2.550
YOLOv3+CNN	0.934	56.411	52.860	17.767
YOLOv8+CNN	0.900	53.295	53.295	1.264
CenterNet2+CNN	0.791	42.860	-	2.085
DDETR+CNN	0.114	19.899	-	-
Retina-SWIN+CNN	0.900	55.000	55.000	13.953

Table 2: Maximum true-positive rate and number of false alarms for the detection methods with d=10 px. Best values marked in bold.

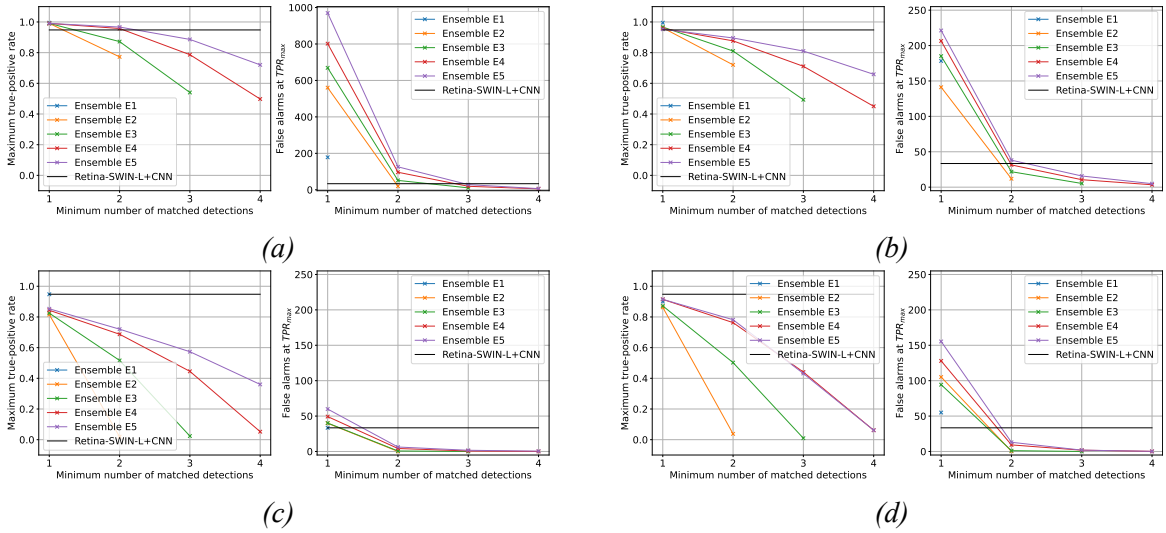


Figure 3: Maximum true-positive rate and number of false alarms of the ensembles with d=10 px. Ensembles are based on the detection pipeline (a) DET, (b) DET with a filtering after ensembeling, (c) LOC+CLS and (e) DET+FLT. Note that the y-axis scale in the right sub-figure in (a) differs from the others.

bles can also reduce the number of false alarms but reduces the maximum true-positive rate too much. The best ensemble combines detections from Retina-SWIN, YOLOv3 and YOLOv8 and achieves a maximum true-positive rate of 87.7% and reduces the false alarms to 31.

REFERENCES

[1] D. P. Williams: "Demystifying Deep Convolutional Neural Networks for Sonar Image Classification" in *Proceedings of the 4th Underwater Acoustics Conference and Exhibition (UACE)*, (Skiathos, 2017).

[2] L. Li, Y. Li, C. Yue, G. Xu, H. Wang, X. Feng: "Real-time underwater target detection for AUV using side scan sonar images based on deep learning", *Applied Ocean Research* **138**, **(2023)**.

[3] I. D. Gerg, V. Monga: "Deep Multi-Look Sequence Processing for Synthetic Aperture Sonar Image Segmentation", *IEEE Transactions on Geoscience and Remote Sensing* **61**, **(2023)**.

[4] Y. Steiniger, D. Kraus, T. Meisen: "Survey on deep learning based computer vision for sonar imagery", *Engineering Applications of Artificial Intelligence* **114**, **(2022)**.

[5] B. W. Sheffield, J. Ellen, B. Witmore: "On vision transformers for classification tasks in side-scan sonar imagery" in *International Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar 2023* (Lerici, 2023).

[6] N. Waraiagoda, Ø. Midtgaard: "Vision Transformers for Sonar Image Classifaction" in *International Conference on Underwater Acoustics 2024* (Bath, 2024).

[7] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo: "Swin transformer: Hierarchical vision transformer using shifted windows" in *2021 IEEE/CVF International Conference on Computer Vision* (virtual, 2021).

[8] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, J. Dai: "Deformable DETR: Deformable Transformers for End-to-End Object Detection" in *9th International Conference on Learning Representations (ICLR)* (virtual, 2021).

[9] Y. Steiniger, J. Stoppe, D. Kraus, T. Meisen: "On the detection and classification of objects in scarce sidescan sonar image dataset with deep learning methods" in *7th Underwater Acoustic Conference and Exhibition (UACE)* (Kalamata, 2023).

[10] X. Zhou, V. Koltun, P. Krähenbühl: "Probabilistic two-stage detection", *arXiv:2103.07461*, 2018, [online] Available: https://arxiv.org/abs/2103.07461.

[11] G. Jocher, A. Chaurasia, J. Qiu: "YOLO by Ultralytics" (8.0.0), 2023, [Software] https://github.com/ultralytics/ultralytics.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin: "Attention Is All You Need" in *Advances in Neural Information Processing Systems 30* (Long Beach, 2017).

[13] Y. Steiniger, J. Groen, J. Stoppe, D. Kraus, T. Meisen: "A study on modern deep learning detection algorithms for automatic target recognition in sidescan sonar images", in *6th Underwater Acoustics Conference and Exhibition* (virtual, 2021).

[14] J. Redmon, A. Farhadi: "YOLOv3: An Incremental Improvement", *arXiv:1804.02767*, 2018, [online] Available: https://arxiv.org/abs/1804.02767.

[15] K. Chen *et al.*: "MMDetection", *arXiv:1906.07155*, 2019, [online] Available: https://arxiv.org/abs/1906.07155.

Structured Session
19. Uncertainty quantification and machine learning in signal processing

Organised by Angeliki Xenaki and Zoi-Heleni Michalopoulou

Source Localization in Shallow Water Using a Neural Network Based on Range-Dependent Sound Speed Profile Model

Jing Guo[1, 2], Juan Zeng[1], Li Ma[1]

[1] *Key Laboratory of Underwater Acoustic Environment, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China*

[2] *University of Chinese Academy of Sciences, Beijing 100049, China*

Abstract **—** In range-dependent shallow-water environments, the complex spatiotemporal variations of the environmental sound speed profile (SSP) lead to mismatches between the SSP model and the actual environment, posing challenges in passive source localization. Leveraging the nonlinear feature extraction and pattern recognition capabilities of deep learning (DL) approaches, this paper investigates the range-dependent SSP model-based neural network for source localization. The performance of the network trained on a linear equivalent SSP model is evaluated, where the environmental SSP is parameterized as a linear function of range. The test results indicate that the neural network, after training on the dataset generated from the equivalent SSP model, demonstrates the ability to localize sources in range-dependent shallow-water environments. By applying transfer learning with experimental data from the Yellow Sea, the range-dependent network better fits the environmental model, leading to reduced errors and improved localization accuracy.

Key Words **—** source localization, deep learning, sound speed profile, range dependence, shallow water

1 Introduction

Passive source localization remains a critical topic in underwater acoustics. Numerous advanced methods based on matched-field processing (MFP) [1], such as matched-mode processing (MMP) [2] and matched-beam processing (MBP) [3], have been extensively developed. However, these methods are constrained in complex environments due to their dependence on the precise representation of environmental model. Especially shallow-water regions, where internal waves frequently induce strong spatiotemporal variability in the sound speed profile (SSP), present significant challenges. Such variability often leads to substantial mismatch between assumed algorithmic models and the actual environment conditions, resulting in notable performance degradation in MFP localization methods [4]~[7]. To address environmental uncertainties, Collins *et al.* [8],[9] introduced the focalization method, which jointly estimates both environmental parameters and source location by incorporating SSP model parameters into the optimization process. Nevertheless, conventional multi-parameter optimization algorithms, such as genetic algorithms (GA), suffer from slow convergence due to the high dimensionality and complexity of the parameter space. Additionally, the computational burden of generating replica fields constitutes an algorithmic constraint in practical MFP applications.

Recently, deep learning (DL) approaches [10] have emerged as promising tools for source localization, owing to their powerful nonlinear feature extraction and pattern recognition capabilities. In contrast to model-based MFP approaches, DL methods follow a data-driven paradigm, wherein neural networks autonomously learn latent features from acoustic data. Prior studies [11]~[13] have demonstrated the feasibility of using neural networks to directly map hydrophone signals to source locations, such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. As DL approaches bypass the online acoustic field calculation, a properly offline trained neural network can reduce online computational costs relative to MFP approaches. Despite these advantages, DL localization approaches face two fundamental limitations:

(1) the lack of physical interpretability, with performance strongly reliant on data quality and quantity; and (2) the limited availability and high cost of acquiring real-world underwater acoustic data, which hampers the generation of high-quality training data and restricts network generalization and performance.

In this paper, a hybrid method is proposed that integrates MFP with DL approach for passive source localization in shallow-water environments. Specifically, training data are generated using a range-dependent equivalent SSP model that characterizes the primary variability of the environmental SSP, thereby enabling effective training of the neural network. The training set consists of all acoustic samples generated within a given parameter space and can be regarded as a set of precomputed replica fields based on the equivalent SSP model. A linear equivalent SSP model that characterizes the range dependence of the environmental SSP influenced by internal waves is employed. The neural network is able to learn from the precomputed replica samples of the training set and utilize the learned features to identify the inputs from unknown environments. This process is functionally equivalent to matching measured samples with precomputed replica samples through a neural network, while preserving partial physical relevance through model-based feature construction. The pretrained network demonstrates effective source localization performance on the validation set, confirming the feasibility of the proposed method. Following pretraining, transfer learning (TL) [15] ~ [19] is employed to further improve the network's localization accuracy and generalization in real-world conditions. A limited number of measured acoustic samples are used to fine-tune the pretrained network, enabling better adaptation to actual environmental. The transfer-learned network performs well on the experimental test set, further validating the effectiveness of the proposed hybrid method based on the range-dependent equivalent SSP model.

2 Localization algorithms

2.1 Range-dependent equivalent SSP model

In conventional MFP localization methods, empirical orthogonal functions (EOFs) are commonly used as basis functions to construct a range-independent equivalent SSP, which serves to mitigate model mismatch caused by temporal variations in the SSP [8][20]. The range-independent equivalent SSP model can be expressed as

$$C(z) \simeq C_0(z) + \sum_{m=1}^{M} \varphi_C^m A_m \,. \tag{1}$$

Here φ_C^m and A_m denote the mth basis function and its corresponding coefficient, respectively. The background SSP (C_0) is typically obtained by averaging historical SSP measurements and the basis functions (*e.g.*, EOFs) are also extracted from historical data, thereby leveraging environmental prior information [20]. This range-independent equivalent SSP model effectively captures the vertical structure of the environmental SSP. When the range dependence of the environmental SSP is weak, the model yields satisfactory localization performance. However, its inability to account for range-dependent variations limits its applicability in shallow-water environments influenced by internal waves.

By incorporating the range-dependent variation into the previously mentioned range-independent equivalent SSP model, the resulting model acquires the ability to capture the range dependence of the environmental SSP. Within short range intervals, the environmental SSP is assumed to vary slowly and can be approximated as a linear function of range [14]. Dividing the environmental SSP into L range segments, and the range-dependent SSP is then approximated by linear profiles over these segments. The linear equivalent SSP model can be expressed as

$$\widehat{\boldsymbol{C}} \simeq \left\{ \left[\hat{C}_1 \sim \hat{C}_2\right]_1, \cdots, \left[\hat{C}_l \sim \hat{C}_{l+1}\right]_l, \cdots, \left[\hat{C}_L \sim \hat{C}_{L+1}\right]_L \right\}, \tag{2}$$

$$\hat{C}_l(z) \simeq C_0(z) + \sum_{m=1}^{M} \varphi_m^l A_m^l \,, \tag{3}$$

where $\left[\hat{C}_l \sim \hat{C}_{l+1}\right]_l$ denotes that the SSP within the lth segment varies linearly from $\hat{C}_l$ to $\hat{C}_{l+1}$ over range. Under the linear equivalence condition, the linear equivalent SSP model can represent the primary spatial variations of the environmental SSP.

However, the selection of segment intervals largely depends on empirical experience and environmental prior information.

In shallow-water environments, internal wave activity is typically dominated by a primary tidal constituent, inducing pronounced sound speed fluctuations that exhibit strong tidal periodicity. In shallow regions of the Yellow Sea, where internal waves are prominent, the semi-diurnal tidal constituent (M2) dominates the dynamics [21]. Under such conditions, the vertical structure of environmental SSP, such as the thermocline, can be effectively parametrized by a background SSP C_0 and basis function φ_C, both derived from 12-hour SSP measurements covering a full semi-diurnal tidal cycle. Given the background SSP and the basis functions, the linear equivalent SSP model can then be parameterized by a coefficient matrix β corresponding to the basis functions.

$$\beta = \begin{bmatrix} A_{1,1} & \cdots & A_{1,M} \\ \vdots & \ddots & \vdots \\ A_{L+1,1} & \cdots & A_{L+1,M} \end{bmatrix}. \tag{4}$$

The coefficient matrix β contains a total of $(L+1) \times M$elements, where L denotes the number of linear segments used to model the range dependence of the SSP, and M represents the number of basis functions. By incorporating the linear equivalent SSP model to MFP framework, the broadband source localization problem is reformulated to a nonlinear multi-parameter optimization problem with the Bartlett processor [1][4]

$$(\alpha, \beta, r, z) = max\{F(\alpha, \beta, r, z)\}, \tag{5}$$

$$F(\alpha, \beta, r, z) = \frac{1}{N_f}\sum_{j=1}^{N_f} |\boldsymbol{V}_{\alpha,\beta,r,z}^{H}\boldsymbol{X}|^2. \tag{6}$$

Here r denotes source range,z denotes source depth, and α represents additional unknown environmental parameters. N_f is the number of frequency points used in MFP. The objective of the optimization is to maximize the correlation between the measured field vector $\boldsymbol{X}$, and the replica field vector $\boldsymbol{V}_{\alpha,\beta,r,z}$.

Numerical simulation is performed using the MFP algorithm based on a linear equivalent SSP model to assess its performance. The simulation environment is constructed using SSP measurements collected during the 2011 Yellow Sea experiment, as illustrated in Fig. 1(a), with a constant water depth of 40 m and a semi-infinite bottom. The linear equivalent SSP model is divided into three segments ($L = 3$), each spanning

a range interval of 4000 m in the simulation. Only the first EOF ($M = 1$), derived from the simulated environmental SSP, is used as the basis function. Environmental parameters, including seabed properties, the vertical line array (VLA) configuration, and the source location, are summarized in Table 1. The acoustic field is computed using the range-dependent acoustic model (RAM) with a maximum range of 12 km and a range step of 10 m. The genetic algorithm (GA) used in the MFP optimization adopts a generation gap of 0.9, a crossover probability of 0.9, a mutation probability of 0.05, and parallel processing with 40 individuals over 50 generations, resulting in 2000 parameter evaluations. Table 2 presents the parameter search ranges and the localization results obtained by the MFP algorithm based on the linear equivalent SSP model.

Tab. 1. Simulation environment parameter settings.

Seabed			Vertical line array			Source		
sound speed	density	attenuation	element count	element spacing	depth coverage	range	depth	signal frequency
m/s	g/cm^3	dB/λ	\	m	m	km	m	Hz
1600	1.5	0.25	16	1.5	12.5~35	10	25	650 700 750 800 850

Tab. 2. Parameter search ranges and MFP result in simulation.

	Equivalent SSP model				Source	
Parameter search ranges	A_1	A_2	A_3	A_4	range (km)	depth (m)
	[-20,30]	[-20,30]	[-20,30]	[-20,30]	[8,12]	[5,40]
MFP result	-0.27	22.27	-19.32	2.32	9.9	25

For shallow-water source localization, numerical simulation demonstrates that the linear equivalent SSP model effectively captures primary features of range-dependent environmental SSP and enables accurate source localization. Figure 1(b) shows the reconstructed equivalent SSP, while Fig. 1(c) presents the estimated localization ambiguity surface. It can be observed that the equivalent SSP model adapts to range-dependent environmental variations by adjusting the model parameters $A_1 \sim A_4$, and the optimal parameters are obtained during the optimization process, yielding an accurate

source location estimate (9.9 km / 25 m). The simulation results indicate that the linear equivalent SSP model successfully characterizes the range-dependence of the environmental SSP and can be effectively integrated into localization algorithms under shallow-water internal wave conditions. More detailed studies of the range-dependent MFP methods will be presented in other separate works by the authors.

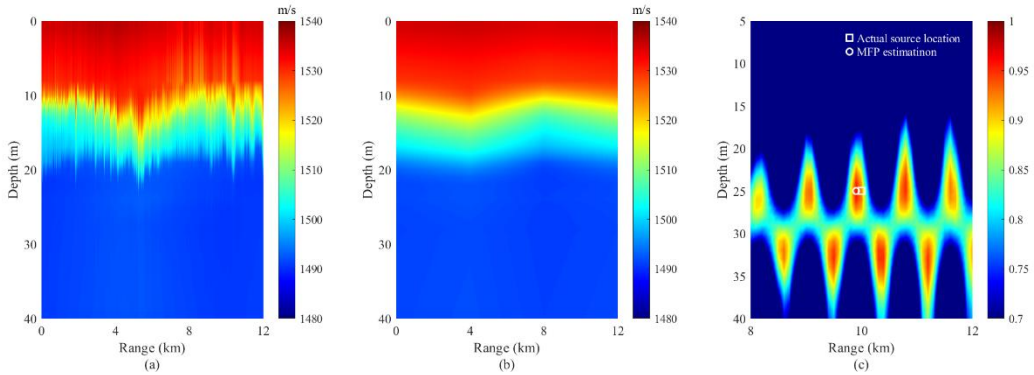


Fig.1 Range-dependent MFP localization.

(a) Simulated environmental SSP; (b) Reconstructed equivalent SSP;

(c) Localization ambiguity surface.

2.2 Model-based deep learning

By rewriting Eq. (6) in vector form, the MDP can be interpreted as a weighted correlation measure. Specifically, the replica vector $\boldsymbol{P}_r$, received by an array with N elements located at depths z_N, is transformed into a covariance matrix $\boldsymbol{R}_{Pr}$, which is then used to perform a weighted correlation with the measured vector $\boldsymbol{P}$, where (r, z) denotes the actual source location. The localization ambiguity function $A(\boldsymbol{P}, \boldsymbol{P}_r)$ is evaluated over a search grid to form the ambiguity surface:

$$A(\boldsymbol{P}, \boldsymbol{P}_r) = \frac{\boldsymbol{P}\boldsymbol{R}_{Pr}\boldsymbol{P}^*}{\|\boldsymbol{P}\|_2^2 \|\boldsymbol{P}_r\|_2^2}, \tag{7}$$

$$\boldsymbol{R}_{Pr} = \boldsymbol{P}_r^* \boldsymbol{P}_r, \tag{8}$$

$$\boldsymbol{P} = [P(r, \mathrm{z}, z_1) \quad \cdots \quad P(\mathrm{r}, \mathrm{z}, z_N)], \tag{9}$$

$$\boldsymbol{P}_r = [P_r(r', z', z_1) \quad \cdots \quad P_r(r', z', z_N)]. \tag{10}$$

Each point on the localization ambiguity surface corresponds to a candidate source location (r', z') in the replica model, along with the replica covariance matrix $\boldsymbol{R}_{Pr}$. However, the optimal model parameters are obtained through the iteratively and

randomly searching within the parameter space. Due to the need to traverse nearly the entire parameter space, conventional nonlinear multi-parameter optimization algorithm (*e.g.*, the genetic algorithm) suffers from high computational cost and require repeated searching for each localization task, which results in substantial online computational burden. From an algorithmic perspective, offline precomputation of extensive replica samples ($\boldsymbol{R}_{Pr}$) allows the online stage to bypass parameter searching and directly identify the measured samples against the precomputed database, resulting in substantial reductions in computational load and processing time. Data-driven DL approaches exhibit potential for this identification task, leveraging their inherent capabilities in nonlinear feature extraction and pattern recognition. Notably, prior study has shown that neural networks can effectively emulate MFP [22][23].

This section discusses a range-dependent neural network for source localization in shallow-water environments influenced by internal waves, integrating MFP with DL approaches. A range-dependent equivalent SSP model—such as the linear equivalent SSP model mentioned in Sec. 2.1—is employed to generate training data. This approach is equivalent to precomputing a large set of replica samples offline and replacing conventional parameter optimization and matching processes with a neural network. By training the network with model-generated data, the reliance on extensive real-world acoustic data is reduced. The range-dependent network extracts generalizable features from the replica samples for source localization, and higher localization accuracy can be further achieved through transfer learning using limited amount of real-world measured data. The network is adapted from the multi-task learning U-Net described in Ref. [23], with the preprocessing procedure and overall network architecture briefly illustrated in Figs. 2 and 3. The source localization algorithm based on the range-dependent neural network is implemented as follows:

(1) Using the linear equivalent SSP model, which has been validated for source localization in range-dependent environment, numerous SSPs are generated over a given model parameter grid. Based on these SSPs, the corresponding replica acoustic fields are computed as replica samples, which are then converted into feature maps through a preprocessing procedure to form the training set. Details of the training,

validation, and test sets are provided in Sec. 2.3.

(2) After pretraining, the neural network approximates the source localization system based on the linear equivalent SSP model. The performance of the pretrained network on the validation set is presented in Sec. 3.1. A limited amount of experimental data is used to fine-tune the pretrained network via transfer learning, completing the full source localization workflow. The performance of both the pretrained and transfer-learned networks on the experimental test set is discussed in Sec. 3.2.

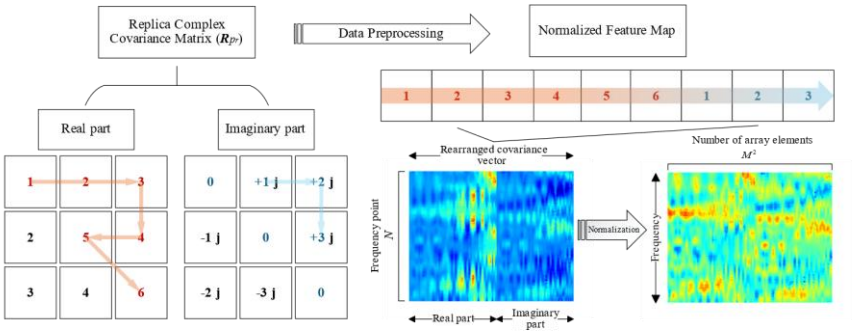


Fig.2 Data preprocessing of the replica samples.

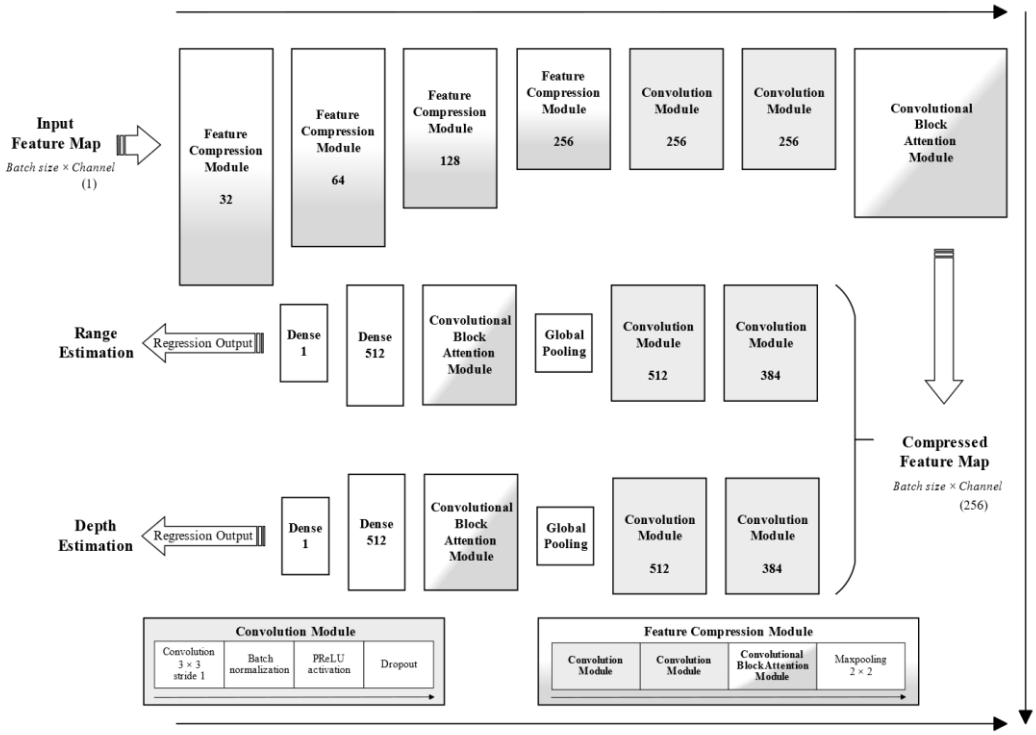


Fig.3 Multi-task learning U-Net framework for source localization.

2.3 Datasets generation

In the summer of 2011, the Institute of Acoustics, Chinese Academy of Sciences, conducted an underwater acoustic propagation experiment in the shallow region of the Yellow Sea, where significant internal wave activity was present. The experimental deployment is shown in Fig. 4. The water depth is 40 m, with the source located at a depth of 22 m and a range of 10 km from the receiving VLA. The source emitted linear frequency modulated (LFM) signals ranging from 650 to 850 Hz, with a modulation duration of 3 s and a repetition interval of 45 s. The receiving VLA consisted of 16 hydrophones, with the uppermost hydrophone positioned at 12.5 m depth and the lowermost at 35 m, spaced at 1.5 m intervals. A temperature-depth (TD) sensor array near the receiving VLA recorded internal wave activity dominated by the semidiurnal tide. Figure 5(b) shows the 36-hour SSP measurements in the experiment. The first three EOFs [Fig. 5(a)] extracted from the 12-hour SSP measurements indicated in Fig. 5(b)—corresponding to the semidiurnal tidal period—serve as the basis functions for the linear SSP model.

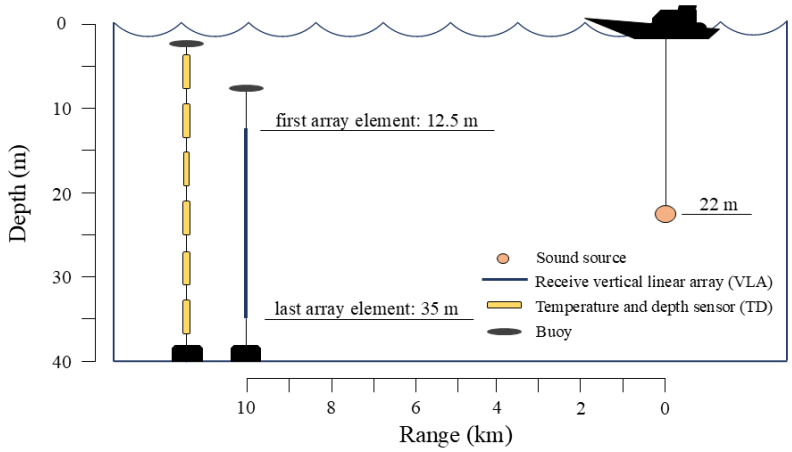


Fig.4 The deployment of the 2011 summer experiment in the Yellow Sea.

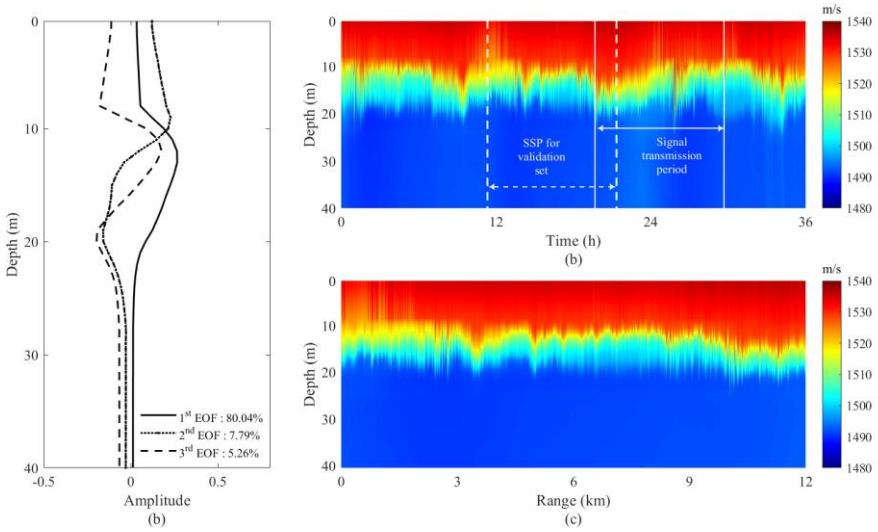


Fig. 5 Information of environmental SSP characteristics.

(a) The first three EOFs; (b) 36-hour real-world SSP measurements;

(c) 12-km range-dependent SSP for validation set generation.

Environment-representative linear equivalent SSPs are constructed to generate the training set, while the maximum computation range for acoustic field is set to 12 km. When the first-order EOF is used as the basis function ($M = 1$), the linear equivalent SSP model requires only the given model parameters

$$Paralist = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_N \end{bmatrix} = \begin{bmatrix} A_1^1 & \cdots & A_{L+1}^1 \\ \vdots & \ddots & \vdots \\ A_1^N & \cdots & A_{L+1}^N \end{bmatrix}. \tag{6}$$

Consistent with the MFP simulation in Section 2.1, set $L = 3$ (range interval of 4000 m) to construct equivalent SSPs. Here, N represents the total number of equivalent SSPs to be constructed. The EOF coefficient A varies from -20 to 30 in increments of 10, producing a total of $N = 1296$ SSPs. Table 3 summarizes the environmental parameter settings used for replica samples generation. The complete training set consists of 265,680 replica samples, generated from all combinations of: 1296 SSPs, 41 source ranges, 5 source depths, and 3 sea depths. Meanwhile, the validation set for evaluating the range-dependent network's localization performance is generated using the 12-km SSP shown in Fig. 5(c). This range-dependent SSP is derived from SSP measurements and extended according to an internal wave propagation speed of 0.33 m/s [21]. As specified in Table 3, the validation set comprises 123 simulated samples (41

source ranges × 3 source depths).

Tab. 3 Configuration of grid parameters for datasets generation

	Sea depth m	Source range km	Range interval km	Source depth m	Depth interval m	Signal frequency Hz	Frequency interval Hz
Training Set	38 40 42	8 ~ 12	0.1	10 ~ 30	5	650 ~ 850	1
Validation Set	40	8 ~ 12	0.1	10 ~ 30	10	650 ~ 850	1

3 Simulation and experimental data evaluation

3.1 Validation set performance

The training set incorporates extensive synthetical range-dependent features of the linear equivalent SSP model, while the neural network learns these characteristics through feature extraction to enhance source localization performance. During network training, the learning rate is set to 10^{-3} and batch size to 4 for 5 epochs. Other network parameters not explicitly mentioned are set according to Ref. [23]. The training is conducted on a platform equipped with an *Intel(R) Core(TM) Ultra 5 125H* CPU and an *Nvidia GeForce RTX 4060 Laptop GPU*, with a total training time of approximately 12 hours. The training loss curve and mean absolute errors (MAEs) for range and depth estimation are shown in Fig. 6. Both the loss curve and MAE trends indicate that the network effectively converges.

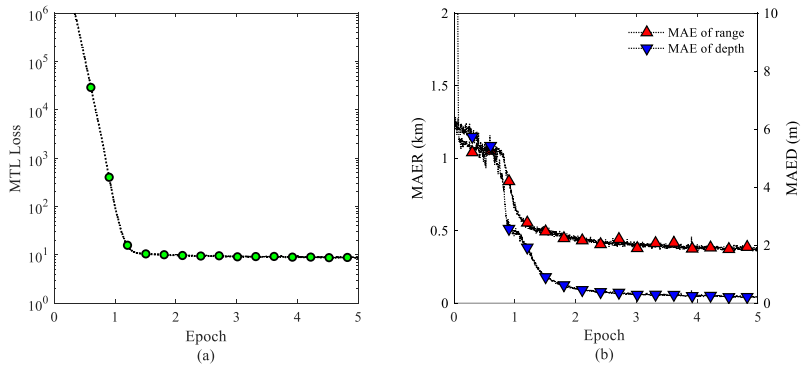


Fig. 6 Training process of range-dependent network.

(a) MTL loss curve; (b) MAEs of range and depth estimation.

The validation set consists of replica samples with a fixed sea depth of 40 m and source depths of 10 m, 20 m, and 30 m. Figure 7(a) presents the source ranges estimated by the range-dependent network on the validation set, while Fig. 7(b) shows the MAEs of range and depth estimates under the three source depth conditions. The network successfully captures variations in source range, indicating effective learning of range-dependent features from the extensive replica samples generated by the linear SSP model. Consequently, the estimated source range curves on the validation set closely follow the trend of the actual range variation. The network yields relatively accurate estimates for sources at 20 m depth. While discrepancies between the linear equivalent SSP model and the environmental SSP model result in localization errors (MAE < 1.5 km in range and MAE < 7 m in depth), the errors remain within acceptable bounds. These results support the feasibility of replacing conventional parameter optimization (e.g., genetic algorithm) in MFP localization algorithm with DL approach. In particular, the neural network is capable of learning key features from the linear equivalent SSP model, achieving practically acceptable localization performance in range-dependent shallow-water environments.

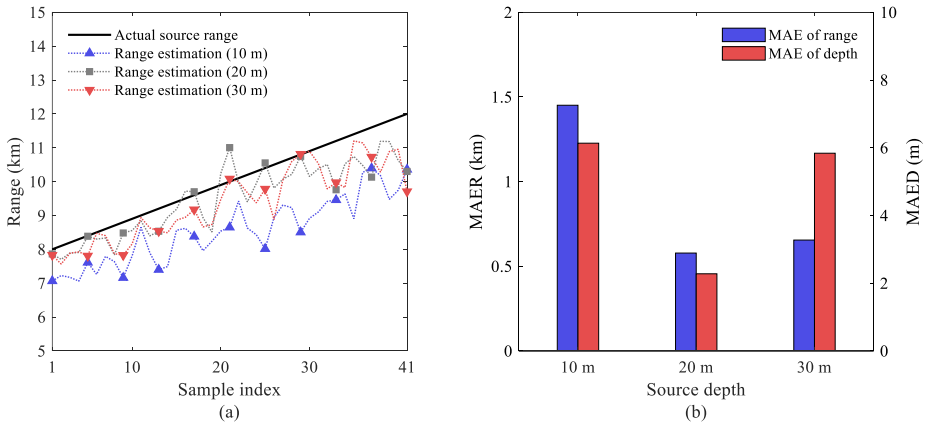


Fig. 7 Network performance on the validation set.

(a) Estimation curves of source location; (b) MAEs of source localization.

3.2 Transfer learning and generalization testing

Although the pretrained network exhibits preliminary localization capability, its performance in complex real-world environments requires further enhancement via

transfer learning. The feature compression module remains frozen to preserve learned representations, while the downstream output layers are fine-tuned using experimental data. This strategy theoretically maintains the network's performance in the source domain (the equivalent SSP model) while improving its generalization ability in the target domain (environmental SSP model). The experimental test set includes 300 received signals from the 2011 Yellow Sea experiment, with a signal-to-noise ratio (SNR) of approximately 16 dB estimated via spectral analysis. Another 200 received signals from the experiment are used as the transfer set. For the transfer learning, learning rate of 10^{-4} with batch size 1 across 10 epochs are employed. The pretrained and transfer-learned networks are evaluated on the experimental test set. Due to the unaffordable computational demands of the linear model-based MFP method (about 300-hours processing time with convergence-ensuring iterations in the real-world environment), the range-independent normal MFP is employed for comparison. The normal MFP utilizes a single experimental SSP measured at the receiving VLA.

Figure 8 shows the estimated results of the pretrained network, the transfer-learned network, and the normal MFP method on the experimental test set, while Fig. 9 presents the corresponding MAEs of source localization. The pretrained network's range estimation curve confirms preliminary source localization capability in experimental range-dependent environment. After pretraining, the range-dependent network significantly outperforms the normal MFP method on the experimental test set, with a range estimation MAE of 0.91 km, much smaller than the 2.93 km MAE of the normal MFP. However, the pretrained network's depth estimation performance is comparable to that of the normal MFP method, with depth estimation MAEs of 10.8 m and 11.8 m, respectively. This implies that range-dependence of the SSP may not be the decisive factor for source depth estimation. Depth estimation accuracy remain constrained by other parameter mismatches, including receiving array configuration, SSP vertical structure, and source depth itself (as indicated by the different MAE levels at different source depths shown in Fig. 7), along with other possible factors. By further learning from experimental data, the transfer-learned network yields substantial improvement in depth estimation and modest range accuracy enhancement. The depth estimation MAE

of the transfer-learned network is reduced to 4.1 m, which is less than half that of the pretrained network. The range estimation MAE of the transfer-learned network decreases to 0.84 km. The enhancement can be reasonably attributed to the transfer set providing the network with supplementary real-world environmental information, thereby significantly improving both depth estimation accuracy and the generalization capability of the range-dependent network.

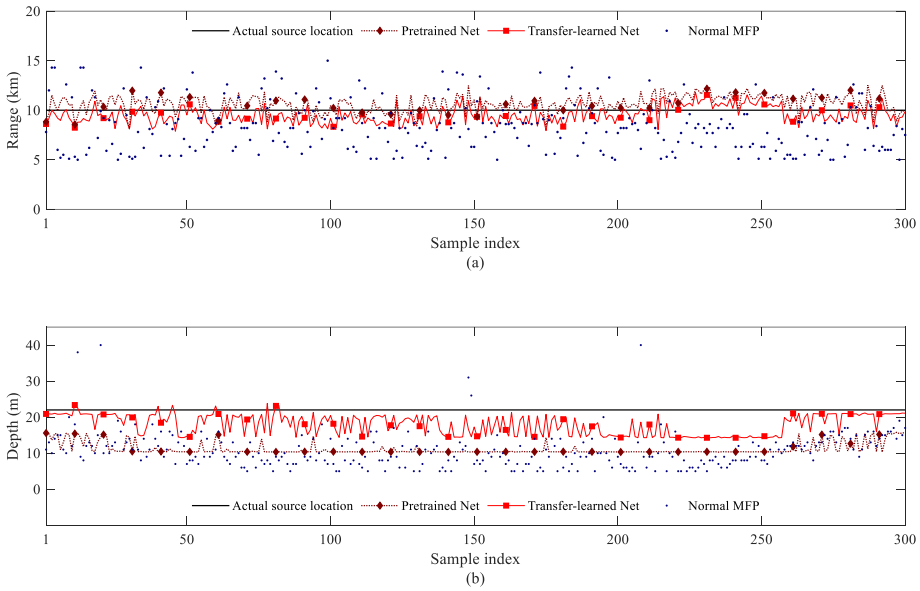


Fig. 8 Algorithms performance on the experimental test set.

(a) Estimation of source range; (b) Estimation of source depth.

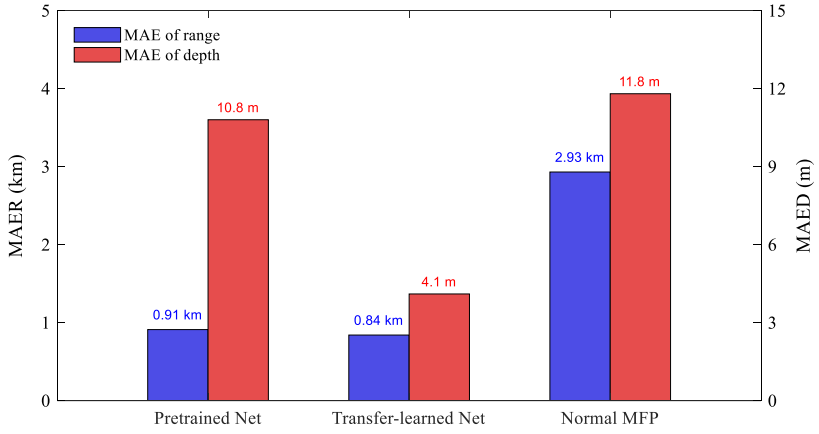


Fig. 9 MAEs of source localization algorithms.

4 Conclusion

For the shallow-water internal wave environment in the Yellow Sea, this study employs a linear equivalent SSP model to generate replica samples. The extensive replica samples are subsequently utilized to train the neural network for source localization in range-dependent environments. Transfer learning based on limited experimental data further improves the network's adaptation to the environmental model and enhances its generalization capability. From the MFP perspective, this hybrid method effectively replaces conventional parameter optimization algorithms (*e.g.*, genetic algorithms) in MFP methods with neural networks. The offline trained network utilizes learned replica samples to identify and estimate unknown measured inputs, thereby reducing the online computational burden of MFP methods. Taking the platform and experimental test set in this study as an example, linear model-based MFP method would require approximately 300 hours of online computation to completes estimation of the test set. In contrast, after 12 hours of offline network training, the range-dependent neural network completes estimation of the test set in merely 40 seconds of online computation. From the DL perspective, the range-dependent equivalent SSP model constrains the feature distribution of the training set. The neural network extracts and learns sufficient synthetic sound speed variation patterns from the training set, effectively mitigating the shortage of measured data for DL methods. Specifically, the proposed range-dependent network is trained primarily on model-generated data, with transfer learning performed using only 200 experimental received signals.

The performance of the proposed hybrid method on both the validation set and experimental test set confirms its effectiveness for source localization in range-dependent shallow-water environments. Future researches on the range-dependent network may include: (1) development of improved range-dependent equivalent SSP models tailored to diverse acoustic environments, and (2) optimization of network architectures to enhance localization performance.

References

[1] Homer P. Bucker, “Use of calculated sound fields and matched-field detection to locate sound sources in shallow water,” *J. Acoust. Soc. Am.*, vol. 59, no. 2, pp. 368–373, 1976, doi: 10.1121/1.380872.

[2] Yoo K and Yang TC, “Broadband source localization in shallow-water in the presence of internal waves,” *J. Acoust. Soc. Am.*, vol. 106, pp. 3255–3269, 1999, doi: 10.1121/1.428179.

[3] Yang TC and Yates T, “Matched-beam processing: application to a horizontal line array in shallow-water,” *J. Acoust. Soc. Am.*, vol. 104, pp. 1316–1330, 1998, doi: 10.1121/ 1.424341.

[4] Rachel M. Hamson and Richard M. Heitmeyer, “Environmental and system effects on source localization in shallow water by the matched-field processing of a vertical array,” *J. Acoust. Soc. Am.*, vol. 86, pp. 1950-1959, 1989, doi: 10.1121/1.398573.

[5] Cristiano Soares, Martin Siderius, and Sergio M. Jesus, “Source localization in a time-varying ocean waveguide,” *J. Acoust. Soc. Am.*, vol. 112, pp. 1879-1889, 2002, doi: 10.1121/1.1508786.

[6] Peter Gerstoft and Donald F. Gingras, “Parameter estimation using multifrequency range-dependent acoustic data in shallow water,” *J. Acoust. Soc. Am.*, vol. 99, pp. 2839-2850, 1996, doi: 10.1121/1.414818.

[7] A.G. Sazontov and I.P. Smirnov, “Source Localization in an Acoustic Waveguide with Inaccurately Known Parameters Using Matched Processing in the Mode Space,” *Acoust. Phys*., vol. 65, pp. 450-459, 2019, doi: 10.1134/S1063771019040171.

[8] Michael D. Collins and W. A. Kuperman, “Focalization: Environmental focusing and source localization,” *J. Acoust. Soc. Am.*, vol. 90, pp. 1410-1422, 1991, doi: 10.1121/1.401933.

[9] Ralph N. Baer and Michael D. Collins, “Source localization in the presence of internal waves,” *J. Acoust. Soc. Am.*, vol. 118, pp. 3117-3121, 2005, doi: 10.1121/1.2041331.

[10] LeCun Y, Bengio Y and Hinton G, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015, doi: 10.1038/nature14539.

[11] S. K. R. Prasad and S. Gurugopinath, “Deep Learning Techniques for Detection of Underwater Acoustic Sources,” *2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON)*, Jaipur, India, pp. 1-6, 2023, doi: 10.1109/IEMECON56962.2023.10092324.

[12] Li X, Song W, Gao D, Gao W and Wang H, “Training a U-Net based on a random mode-coupling matrix model to recover acoustic interference striations,” *J. Acoust. Soc. Am.*, vol. 147, pp. EL363–369, 2020, doi: 10.1121/10.0001125.

[13] Liu Y, Niu H, and Li Z, “A multi-task learning convolutional neural network for source localization in deep ocean,” *J. Acoust. Soc. Am.*, vol. 148, pp. 873–883, 2020, doi: 10.1121/10.0001762.

[14] Harry A and DeFerrari, “Effects of horizontally varying internal wavefields on multipath interference for propagation through the deep sound channel,” *J. Acoust. Soc. Am.*, vol. 56, no. 1, pp. 40–46, 1974, doi: 10.1121/1.1903230.

[15] Niu H, Reeves E, and Gerstoft P, “Source localization in an ocean waveguide using supervised machine learning,” *J. Acoust. Soc. Am.*, vol. 142, pp.1176–1188, 2017, doi: 10.1121/1.5000165.

[16] Huang Z, Xu J, Gong Z, Wang H and Yan Y, “Source localization using deep neural networks in a shallow-water environment,” *J. Acoust. Soc. Am.*, vol. 143, pp. 2922–2932, 2018, doi: 10.1121/1.5036725.

[17] Yoon S, Yang H and Seong W, “Deep learning-based high-frequency source depth estimation using a single sensor,” *J. Acoust. Soc. Am.*, vol. 149, pp. 1454–1465, 2021, doi: 10.1121/10.0003603.

[18] Liu Y, Niu H, and Li Z, “Source ranging using ensemble convolutional networks in the direct zone of deep water,” *Chinese Phys. Lett.*, vol. 36, no. 4, pp. 044302, 2019, doi: 10.1088/ 0256-307X/36/4/044302.

[19] Wang W, Ni H, Su L, Hu T, Ren Q, Gerstoft P, et al, “Deep transfer learning for source ranging: deep-sea experiment results,” *J. Acoust. Soc. Am.*, vol. 146, pp. EL317–322, 2019, doi: 10.1121/1.5126923.

[20] Lester R. LeBlanc and Forster H. Middleton, “An underwater acoustic sound velocity data model,” *J. Acoust. Soc. Am.*, vol. 67, pp. 2055-2062, 1980, doi: 10.1121/1.384448.

[21] Tao Hu, Zhen Wang, S. M. Guo, Li Ma, “Inversion of the internal wave velocity using the normal-mode amplitude fluctuation of an underwater sound field,” *Journal of Harbin Engineering University*, vol. 41, no. 10, pp. 1518-1523, 2020, doi: 10.11990/jheu.202007063.

[22] M.D. Liu, H. Q. Niu, and Z. L. Li, “Implementation of Bartlett matched-field processing using interpretable complex convolutional neural network,” *JASA Express Lett.*, vol. 3, no. 2, pp. 026003, 2023, doi: 10.1121/10.0017320.

[23] Peng Qian, W. M. Gan, H. Q. Niu, G. h. Ji, Z. L. Li, and G. J. Li, “A feature-compressed multi-task learning U-Net for shallow-water source localization in the presence of internal waves,” *Applied Acoustics*, vol. 211, no. 109530, 2023, doi: 10.1016/j.apacoust.2023.109530.

Structured Session
20. Underwater acoustic calibration, measurement, and standards

Organised by Stephen Robinson

ANALYSES OF SIGNAL PROCESSING METHODS APPLIED IN WIDE BAND FREE-FIELD CALIBRATIONS OF HYDROPHONES

Guanghui Jia[1,2], Yi Chen, [1,2] and Liuqing Yang[1,2]

[1] Hangzhou Applied Acoustics Research Institute, Hangzhou, 311400, China
[2] Zhejiang Key Laboratory of Acoustical Testing Technology

Guanghui Jia, Hangzhou Applied Acoustics Research Institute, Fuyang District, Hangzhou, Zhejiang Province 311400, China, 13732264446@163.com

Abstract: *In free-field underwater acoustic measurements, wide band free-field calibrations of hydrophones are limited by the reflecting acoustic waves from boundaries. The free-field calibration model containing the directing acoustic wave and reflecting acoustic waves are established, and signal processing methods can be used to eliminate the reflections from boundaries, which are widely applied in wide band hydrophone calibrations and expanding their frequency limitations. The characteristics of the directing acoustic wave and reflecting acoustic waves of free-field calibrations are discussed based on a linear response system. The applications of signal processing in time domain and frequency domain which enable wide band free-field calibration to be undertaken in a limited space and decrease the reflections from boundaries are described. The wide band experiments of hydrophones were respectively performed using noise signals and chirp signals in several laboratory underwater tanks, and the wide band sensitivities were obtained. The responses of different signal processing methods were analysed which can give instructions of wide band hydrophone calibrations in laboratory underwater tanks.*

Keywords: *calibration, wide band, signal processing, free-field*

1. INTRODUCTION

Hydrophones are widely used in underwater measurements to detect the pressure variations of acoustic waves in water and produce an output voltage proportional to the acoustic pressure [1], [2]. Hydrophones are usually calibrated in a laboratory water tank with limited volume. Reflecting acoustic waves from boundaries and water surface will interferer the free-field measurement acoustic environment. To perform hydrophone calibrations in these water tanks, gated sinusoidal are utilized in discrete acoustic frequencies. These gated sinusoidal signals can be used to eliminate the reflecting acoustic waves in time domain through the arriving time interval of the directing acoustic wave and reflecting acoustic waves. In that condition, the receiving signals of hydrophones need reach the steady-state conditions before the reflecting acoustic waves interfere the directing acoustic wave [3]. However, some hydrophones are widely applied in underwater acoustic monitors, which receive wide band signals, such as noise signals and chirp signals, and they are preferred to be calibrated by wide band signals as well. However, wide band calibration cannot be performed using gated sinusoidal signals because they will last long time, and the receiving signals by a hydrophone is the combination of the directing acoustic wave and reflecting acoustic waves, which is hardly distinguishable in time domain.

There are two ways to perform wide band calibration in limited water tanks. One is to perform in a larger water region, and the reflecting acoustic waves will be attenuated by spherical transmitting and boundaries. Some primary calibrations were performed in large water regions [4]. However, the large water region is not an ideal place for wide band calibrations, because the calibration environments are hard to control, and the background noise will influence the calibration processing, especially for wide band noise signals. The noise signals and background noises can hardly be distinguished by the method of coherent average. The other way is to perform signal processing in the calibration. The complex moving weighted average (CMWA) technique can obtain the wide band transform impedance of the projector and the receiver and eliminate the reflecting acoustic waves [5]. The frequency domain filter processing (FDFP) is another signal processing technique based on the space relationship between the projector and the receiver [6], [7]. Chirp signals are used in the method of CMWA, and the moving average process is performed in time domain, where the dependence of time and frequency is known. Random signals can be used in the method of FDFP, where the relationship between the frequency and time is hard to obtain. The processes are performed in frequency domain.

The CMWA and FDFP are used in free-field primary calibrations and optical calibrations. The frequency response of transfer function is simulated and analysed in this paper, which is crucial to realise wide band calibrations. The differences between the CMWA and FDFP is also discussed which is performed in different water tanks.

2. PRINCIPLE OF WIDE BAND CALIBRATIONS

For a wide band free-field calibration, the system consisting of a projector and a hydrophone can be considered as a linear time invariance system, the frequency response of the transform impedance in the reverberation condition can be expressed as follows

$$Z(f) = \frac{S(f)H(f)M(f)}{r} \tag{1}$$

where, $Z(f)$ is the transform impedance, $H(f)$ is the frequency response of the water tank transfer function, $S(f)$ is the transmitting current response of the projector, $M(f)$ is the sensitivity of the hydrophone, r is the distance between the projector and the hydrophone.

If the measurement is performed in a free-field acoustic environment without reflecting acoustic waves, the circumstance can be considered as a spherical wave radiated by a source, where the water tank transfer function can be expressed as $H(f)=1$. When the receiving acoustic wave is the superposition of the directing acoustic wave and reflecting acoustic waves, that water tank transfer function can be expressed as

$$H(\Delta d_n, f)=\frac{p_0(d,f)+\sum_{n=1}^{N}p_n(d,f)}{p_0(d,f)}=1+\sum_{n=1}^{N}A_n e^{-\frac{j\omega\Delta d_n}{c}} \tag{2}$$

where, $p_0(d,f)$ is the directing acoustic wave, $p_n(d,f)$ is the nth reflecting acoustic waves, A_n is the ratio of nth reflecting acoustic wave to the direct acoustic wave, Δd_n is the acoustic propagation difference between the reflecting acoustic waves and the directing acoustic wave, $\omega=2\pi f$.

The transform impedance can be expressed as

$$Z(f)=\frac{S(f)M(f)}{r}\left(1+\sum_{n=1}^{N}A_n e^{-\frac{j2\pi f\Delta d_n}{c}}\right) \tag{3}$$

In the processing of CMWA, the frequency response of the transfer impedance at frequency f_0 is $Z(f_0)$, which can be calculated as

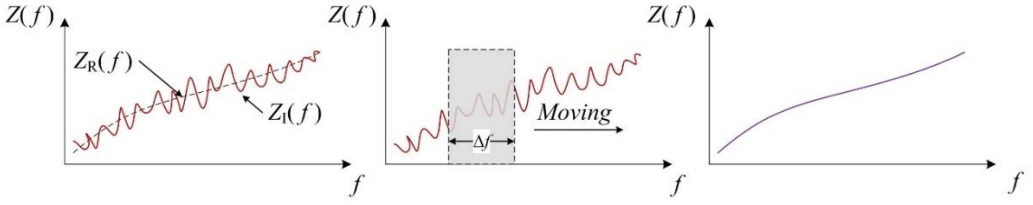
$$\bar{Z}(f_0)=\frac{1}{\Delta f}\int_{f_0+\frac{\Delta f}{2}}^{f_0+\frac{\Delta f}{2}}Z(f)\,df \tag{4}$$

where, $\bar{Z}(f_0)$ is the transform impedance at frequency f_0 after CMWA, Δf is the frequency interval. In equation (4), the transmitting current response $S(f)$ is symmetrical, and the sensitivity $M(f)$ is flat. This processing can be shown in Figure 1 (a), the frequency response of $Z_R(f)$ is the free-field transform impedance without interference, and the frequency response of $Z_I(f)$ is the transform impedance with interferences.

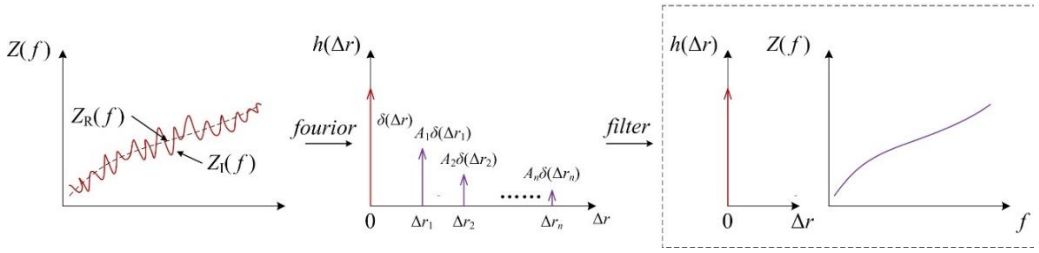
By using Fourier transform, the transfer function $H(f)$ can be expressed in space function $h(\Delta r)$ in the spatial domain as

$$\begin{aligned}h(\Delta r)&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(1+\sum_{n=1}^{N}A_n e^{-jk\Delta d_n}\right)e^{jk\Delta r}d\Delta r=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{jk\Delta r}d\Delta r+\sum_{n=1}^{N}\frac{A_n}{2\pi}\int_{-\infty}^{\infty}e^{-jk(\Delta r_n-\Delta r)}d\Delta r\\&=\delta(\Delta r)+\sum_{n=1}^{N}A_n\delta(\Delta r_n-\Delta r)\end{aligned} \tag{5}$$

where, the first term $\delta(\Delta r)$ is directing acoustic wave, and $A_n\delta(\Delta r_n-\Delta r)$ in the second term is the nth reflecting acoustic waves. Since the acoustic propagation path difference Δr_n is a discrete value, it is possible to design a lowpass spatial filter to reduce the acoustic reflecting waves with propagation path differences Δr_n. The equation (5) and the low-pass filter can be simplified shown in Figure 1 (b). To realize the linear phase response, the finite-duration impulse response filter is used to reduce the reflecting acoustic waves.



(a)



(b)

Fig.1: The processing of CMWA and FDFP, (a) the processing of CMWA, (b) the processing of FDFP.

3. COMPARISON BETWEEN DIFFERENT METHODS

3.1 Simulations and analyses of reflections in a water tank

According to the principle of the wide band calibration theory, the frequency response of the transform impedance is determined by the frequency response of water tank transfer function $H(f)$. The frequency response of the water tank transfer function is dependent on the time delay, spatial position, and reflections. However, the frequency response of water tank transfer function cannot be obtained directly. To analyse these methods, a simulation of water tank with reflections was established, and the water tank transfer function can be analysed.

A small water tank always has high strength of reflections, and the simulation is performed in this kind of water tank condition. A projector and receiver are deployed in the centre of a small water tank with size of 1.8 meters length, 1.2 meters depth, and 1.0 meters width. The distance between the projector and the receiver is 0.3 meter, and the acoustic propagation distance between the direction and the first-reflection of different boundaries is approximately to 1.24 meters, 1.04 meters, and 1.8 meters. The strength of the multiple reflections is small which can be ignored. To ensure the simulation is efficient, absolute reflections are hypothesised.

The CMWA is only available to wide band signals that the information of time and frequency is known, and the water tank transfer function can be obtained through the chirp signals in time domain. FDFP can attenuate reflections in spatial domain, and it is available to any wide band signals without the information of time and frequency. To compare these two methods, the chirp signal is used in this simulation. The original and processed water tank transfer function is shown in Fig.2. For amplitude frequency response of the water tank transfer function, it converges to 1 in theory. Fig.2 (a) and Fig.2 (b) show that these two methods are efficient in wide band calibration of hydrophones. The maximum deviation of these two methods is less than 2%. For phase frequency response of the water tank transfer function, the maximum deviation is 1 °, which is approximate to theorical 0 ° . The method of CMWA has high resolution than FDFP, however, the resolution of FDFP is also enough to analyse the frequency response for hydrophone calibrations. In addition, the projector-receiver system is a reasonable, which has low frequency limitation after processing. Thus, the low frequency limitation of this simulation is 2500 Hz.

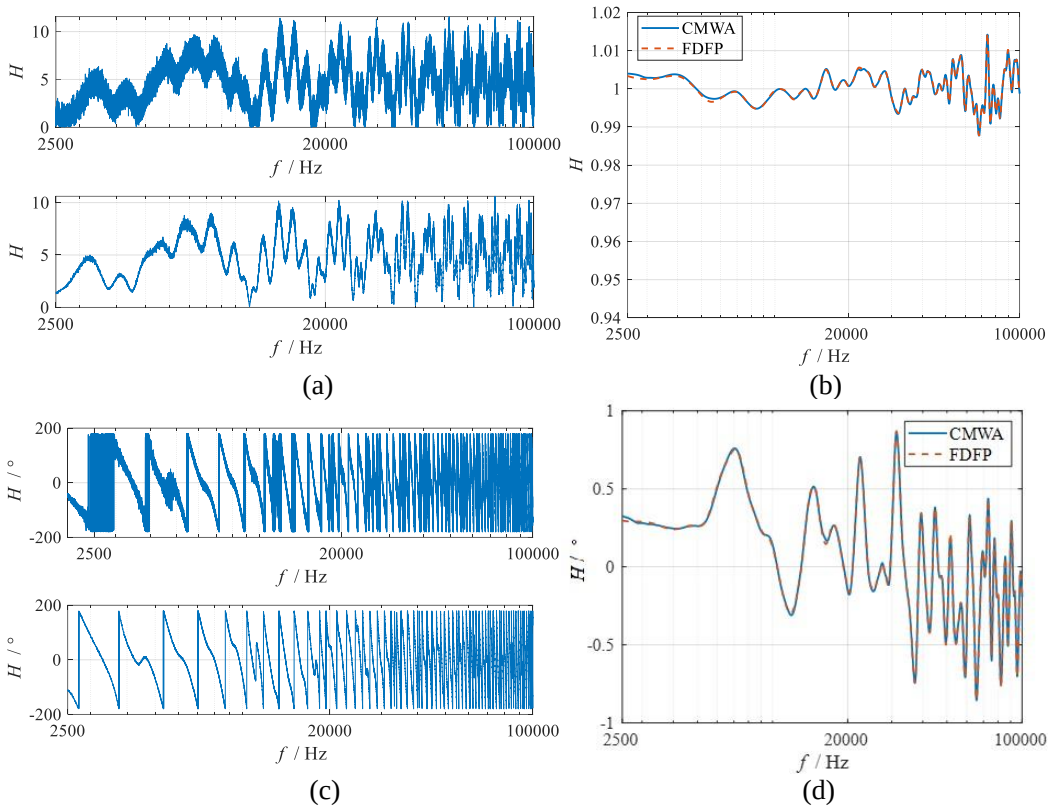


Fig.2: Water tank transfer function H(f), (a) the amplitude response of the original water tank transfer function, (b) the amplitude response after processing, (c) the phase response of the original water tank transfer function, (b) the phase response after processing.

3.2 Wide band calibration results of hydrophones

Several wide band hydrophone experiments are performed in different water tanks. The hydrophone of B&K 8103 was calibrated in a water vessel of 1.8 meters length and 1.2 meters radius through an optical method, and the method of FDFP and CMWA are utilized respectively. In these experiments, the frequency resolution of CMWA is calculated through the relationship of time and sweep rate. The frequency resolution of FDFP is calculated through Fourier transform. Although the frequency resolution is different in these experiments, the frequency responses will be obtained correctly in both methods. The sensitivity results show that the method of CMWA and FDFP are consistent with each other, and the maximum deviation between these two methods is approximately to 0.3 dB. The same hydrophone is also calibrated by a primary method in a larger water tank, and the discrete frequency tone burst signals are used in the calibration. These results are shown in Figure 3.

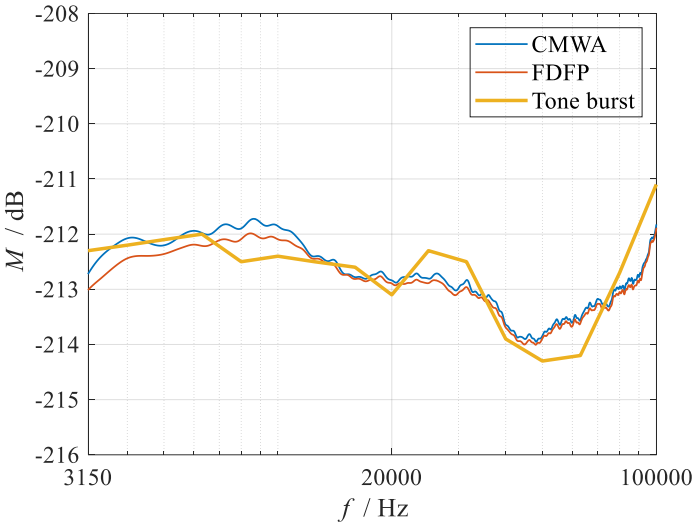


Fig.3: Sensitivity of the B&K 8103 hydrophone calibrated by different methods in a vessel.

A hydrophone of B&K 8106 was wide band calibrated using random noise signals in a large water tank by a primary calibration method, and the calibration range was 250 Hz to 10 kHz. The wide band frequency response of the sensitivity can be obtained directly, and the reflections were eliminated by the method of FDFP. The discrete frequency tone burst signals were also used in the free-field calibration as well, and this calibration can only be performed in the frequency range 1 kHz to 10 kHz. The calibration results are shown in Fig.4, and the maximum deviation between the tone burst signal and FDFP is less than 0.7 dB, which validates that the wide band calibration by FDFP is accurate.

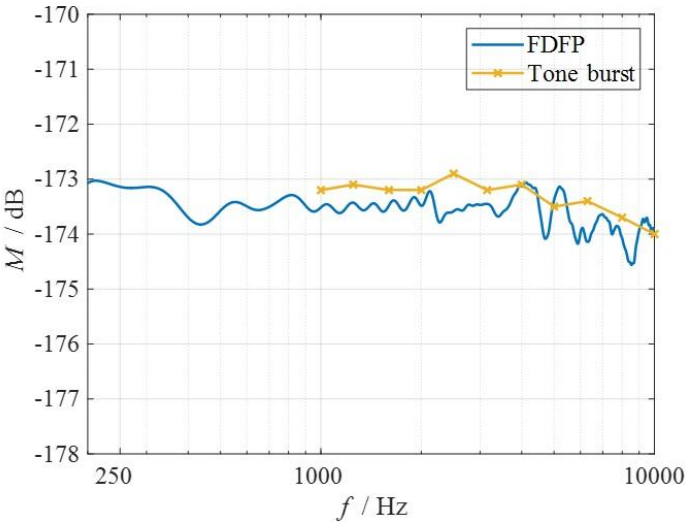


Fig.4: Sensitivity of the B&K 8106 hydrophone calibrated by different methods in a large water tank.

4. CONCLUSION

Wide band sensitivity is crucial for hydrophones, and, it is a big challenge to calibrate in free-fields. For discrete frequency tone burst signals, which is widely used in the free-field

calibration of hydrophones, it is difficult to distinguish the directing acoustic wave to reflecting acoustic waves, and the acoustic interference will influence the calibration accuracy. To increase the wide band calibration accuracy and eliminate the reflecting acoustic waves from boundaries and water interface, the signal processing technique is applied in free-field calibrations.

Two well-known signal processing methods applied in free-field calibrations are introduced, and the characteristics of these methods are analysed through the simulation and experiments. Although different processing is performed and analysed, the frequency response of transfer impedances can be obtained accurately. The amplitude and phase response are approximate to theorical quantity. The commercial hydrophones type of 8103 and 8106 manufactured by B&K were calibrated in different water tanks using wide band method of CMWA and FDFP respectively. These hydrophones were also measured by traditional primary method using tone burst signals. The methods of CMWA and FDFP are in a good agreement with each other in wide band frequency range. In addition, the sensitivities calibrated by wide band calibration method is approximate to the results calibrated by primary method in 1/3 octave frequency points. Therefore, the wide band calibration methods of CMWA and FDFP can be accurately applied to hydrophone calibrations

5. ACKNOWLEDGEMENTS

This work is supported by the National Key R&D Program of China (Grant No. 2022YFF0607503).

REFERENCES

[1] **R. J. Bobber**, Underwater Electroacoustic Measurements, Peninsula Press, Los Altos CA, pp. 27–31, 1988.

[2] IEC 60565-1: 2020, Underwater acoustics – Hydrophones – Calibration of hydrophones – Part 1: Procedures for free-field calibration of hydrophones, International Electrotechnical Commission, Geneva, Switzerland, 2020.

[3] **S. P. Robinson, G. Hayman, P. M. Harris, and G. A. Beamiss**, Signal modelling methods applied to the free-field calibration of hydrophones and projectors in laboratory test tanks, Meas. Sci. Technol. 29, 085001, 2018.

[4] **S. P. Robinson, P. M. Harris, B. Ford, J. Ablitt, Y. Chen, A. Biber, A. C. Corakci, R. Costa-Felix, S. Crocker, A. E. Isaev, V. Tyalimpi, and G. Latha**, CCAUV.W-K2 final report – Key comparison CCAUV.W–K2: Calibration of hydrophones in the frequency range from 250 Hz to 500 kHz, Metrologia, 59, Number 1A 09003, 2022.

[5] **A. E. Isaev and A. N. Matveev**, Calibration of hydrophones in a field with continuous radiation in a reverberating pool, Acoustical Physics 55, 762–770 , 2009.

[6] **Yi Chen, Liuqing Yang, Lisheng Zhou, Xiaofeng Jin, Guanghui Jia**, Calibration of hydrophones using a frequency domain filter processing method: theory and experiment, Meas. Sci. Technol. 32 035012, 2021.

[7] **Guanghui. Jia, Yi. Chen, Juan. Tu, and Dong Zhang**, A primary wideband calibration method for noise hydrophones based on three-transducer spherical wave reciprocity in the frequency domain, J. Acoust. Soc. Am. 156, 1633–1645, 2024.

Calibration and its correction for the sensitivities of hydrophones in the air chamber

Liuqing Yang [1, 2,3], Yi Chen[1,3], Boyue Wu[1,3]

[1] Hangzhou Applied Acoustics Research Institute, Hangzhou, Zhejiang, 311400, China
[2] Qingdao Innovation and Development Base, Harbin engineering university, Qingdao, Shandong, 266000, China
[3] Zhejiang Key Laboratory of Acoustical Testing Technology, Hangzhou, Zhejiang, 311400, China

Liuqing Yang, 82th Guihuaxi Road, Fuyang district, Hangzhou city, Zhejiang province, China. yanglqathz@163.com

Abstract: *The free field represents the sound field most closely aligned with the application scenarios of underwater acoustic transducers. Calibration of underwater acoustic parameters in the kilohertz frequency range is generally conducted under controlled laboratory conditions to ensure accuracy. Compared to water, sound waves in air exhibit shorter wavelengths, which facilitates the formation of a free field in confined spaces and results in a lower low-frequency limit. A method for calibrating the sensitivity of hydrophones within an air chamber has been proposed to accurately determine their free field sensitivity at lower frequencies. The acoustic field characteristics of hydrophones operating under free field conditions were investigated, leading to the establishment of a correction coefficient for sensitivity applicable in both air and water. This coefficient utilizes the proportional relationship between the average receiving force and the sensitivity of hydrophones. Sensitivity measurements were performed on spherical hydrophones with diameters of 30 mm within the air chamber, spanning frequencies from 500 Hz to 6.3 kHz. Concurrently, the sensitivity of hydrophones within the same frequency range was calibrated using the vibrating liquid column method and the three-transducer spherical-wave free*

field reciprocity method in water, yielding results that closely align with the corrected air chamber measurements. The correction coefficient for hydrophone sensitivity can be employed to assess the sensitivity of underwater hydrophones.

Keywords: *spherical hydrophone, air chamber, correction coefficient, sensitivity.*

INTRODUCTION

The low-frequency extension of free-field calibration for hydrophones in finite water environments has long been a research focus and technical challenge in underwater acoustic measurements. To reduce the lower frequency limit for calibrating underwater acoustic transducers in confined water spaces, Beatty et al. [1]proposed the Prony spectral estimation method, which models the transient signals of transducers using pole-zero representations to estimate the phase and amplitude of steady-state acoustic waves from transient signals. Building upon this, a multipath signal modelling technique based on exponential signal forms was developed, representing the multipath signals received by the transducer as a superposition of exponential functions[2]. By applying a least-squares error algorithm, the parameters of these exponential signals are estimated to derive the phase and amplitude of the steady-state acoustic signal. Compared to the Prony spectral estimation method, the multipath modelling technique further reduces the lower frequency limit for calibrating underwater acoustic transducers in finite water environments[3].

For linear frequency-modulated (LFM) acoustic signals generated by a sound source, there exists a fixed frequency difference between the direct and reflected signals received by the transducer. Accordingly, Heyser[4] applied the Time Delay Spectrometry (TDS) method to transducer calibration, effectively suppressing the influence of reflected waves by narrowband filtering of the received broadband signals. Based on TDS, Chen[5] et al. proposed the Frequency Domain Filtering in Position (FDFP) technique, which constructs spatial filters according to the geometric relationships between transducers, effectively eliminating the effects of boundary reflections. However, for free-field calibration of hydrophones at frequencies of several hundred hertz, there is currently no established laboratory calibration method.

The wavelength of sound waves in air is approximately one-fifth that in water. For a measurement space of the same size, the lower frequency limit for free-field conditions in air is also one-fifth that in water. More importantly, while the operational lower frequency limit of an anechoic water tank is typically above 1 kHz, an anechoic chamber in air can achieve a free-field lower frequency limit as low as 40 Hz. Calibrating hydrophone sensitivity in air could significantly extend the lower frequency limit of free-field calibration and improve measurement accuracy. However, the characteristic impedance of air and water differs drastically, and hydrophone sensitivity is closely related to the propagation medium. This paper addresses the critical challenge of correcting hydrophone sensitivity between air and water media.

CALIBRATION AND CORRECTED METHOD OF THE HYDROPHONE SENSITIVITY

When excited by acoustic waves, the sensing element of a piezoelectric hydrophone undergoes deformation, thereby generating an electric current. The magnitude of deformation is proportional to the hydrophone's output, allowing for sensitivity correction by comparing the forces received by the hydrophone in different media.

1.1. Corrected method of the hydrophone sensitivity

The force received by a hydrophone's surface is primarily determined by the total acoustic field in which it is situated. In either air or water, the characteristic impedance of the hydrophone surface is significantly greater than that of the surrounding medium, allowing the surface to be treated as a rigid boundary. Consequently, the acoustic pressure at the hydrophone surface can be expressed as:

$$p = \frac{p_0}{(k_0 a)^2} \sum_{m=0}^{\infty} (-j)^{m+1} (2m+1) \frac{P_m(\cos\theta)}{\frac{d}{d(k_0 a)} h_m^{(1)}(k_0 a)} e^{-j\omega t} \tag{1}$$

where, k_0 is the wavenumber in the medium, a is the hydrophone's radius, p_0 represents the amplitude of the incident plane wave, $P_m(\bullet)$ corresponds to the m-th order Legendre function, and $h_m^{(1)}(\bullet)$ signifies the m-th order spherical Hankel function of the first kind.

The force acting on the hydrophone surface is

$$F = \iint_S p ds = 4\pi a^2 p_0 \frac{e^{jka}}{1+jka} \tag{2}$$

Since the wavenumbers of acoustic waves differ between water and air media, the ratio of received forces in different media can be expressed as:

$$v = \frac{1+jk_1 a}{1+jk_2 a} \bullet \frac{e^{jk_2 a}}{e^{jk_1 a}} \tag{3}$$

where, k_1 denotes the wavenumber in water, and k_2 represents the air. According to IEC 60565[6], the ratio of a hydrophone's open-circuit voltage to the sound pressure at the point prior to hydrophone insertion is defined as its sensitivity. Given that F is proportional to the hydrophone's output voltage, the sensitivity measured in air and water can be expressed as

$$M_{\text{h-corrected}} = vM_{\text{h}} \tag{4}$$

where, M_{h} denotes the hydrophone sensitivity measured in air, and $M_{\text{h-corrected}}$ represents the corrected sensitivity in water. The ratio v of received forces between air and water can be defined as the air-to-water sensitivity correction

factor. The frequency dependence of v is shown in Fig. 1, revealing that:(1) Below 1 kHz, the measured sensitivities in air and water show agreement. (2) The correction factor increases with frequency. (3) For spherical hydrophones, larger dimensions yield greater correction factors at identical frequencies. (4) A 15 mm radius spherical hydrophone exhibits a correction factor of 2.5 at 10 kHz.

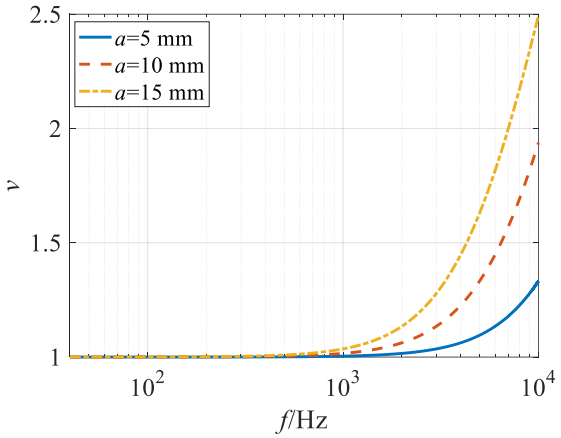


Fig.1: Variation of the correction factor with frequency. The solid line represents the correction factor of the hydrophone with a radius of 5 mm. The dashed line represents the correction factor of the hydrophone with a radius of 10 mm. The dashed-dot line represents the correction factor of the hydrophone with a radius of 15 mm

1.2. calibration principle

In the anechoic chamber, the hydrophone sensitivity was calibrated using the comparison method, with the acoustic field arrangement illustrated in Fig. 2. The distances between the omnidirectional sound source and the hydrophone (d_{h}) and between the source and the microphone (d_{m}) are as shown in the figure. The hydrophone sensitivity in air is then given by:

$$M_{\mathrm{h}} = \frac{d_{\mathrm{h}}}{d_{\mathrm{m}}} \bullet \frac{U_{\mathrm{h}}}{U_{\mathrm{m}}} \bullet M_{\mathrm{m}} \tag{5}$$

where, U_{h} is the hydrophone's open-circuit output voltage, U_{m} denotes the microphone's open-circuit output voltage, and M_{m} represents the microphone sensitivity. Combining with Eq. (4), the hydrophone sensitivity in water can be expressed as:

$$M_{\text{h-corrected}} = v\frac{d_{\mathrm{h}}}{d_{\mathrm{m}}} \bullet \frac{U_{\mathrm{h}}}{U_{\mathrm{m}}} \bullet M_{\mathrm{m}} \tag{6}$$

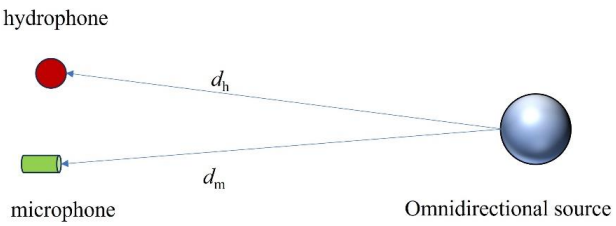


Fig.2: Geometry of the calibration setup for the hydrophone, microphone, and source.

MEASUREMENT EXPERIMENT

As shown in Fig.3. the measurement system comprised a signal generator, power amplifier, sound source (B&K 4295), microphone (B&K 4955), and data acquisition unit, with a 30-mm spherical hydrophone positioned 2.2 m from the source and the microphone placed 3.45 m away, where tone-burst signals across the operational frequency range were generated and the open-circuit voltages from both microphone and hydrophone were acquired via digitizer.

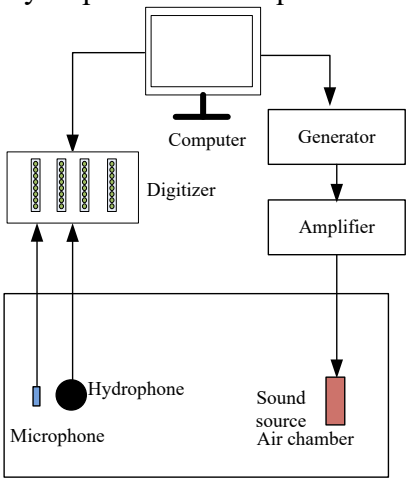


Fig.3: Experiment setup.

RESULTS AND DISCUSSION

To validate the accuracy of the hydrophone sensitivity measurement and correction methodology, the RHS-30 hydrophone's underwater sensitivity was calibrated using both the vibrating liquid column method and the three-hydrophone spherical transducer free-field reciprocity technique. Figure 4 compares the measured sensitivity of the RHS-30 spherical hydrophone in both the anechoic chamber and underwater environments with the corrected sensitivity values across the 500 Hz to 6.3 kHz frequency range. The results demonstrate that the discrepancy between the air chamber measurements and underwater measurements increases with frequency, reaching a maximum sensitivity level difference of 6 dB at 6.3 kHz. After applying the correction formula (4) to the air chamber measurements, the corrected sensitivity shows excellent agreement with the directly measured underwater sensitivity the maximum deviation of approximately 1.6 dB occurs at 2.5 kHz.

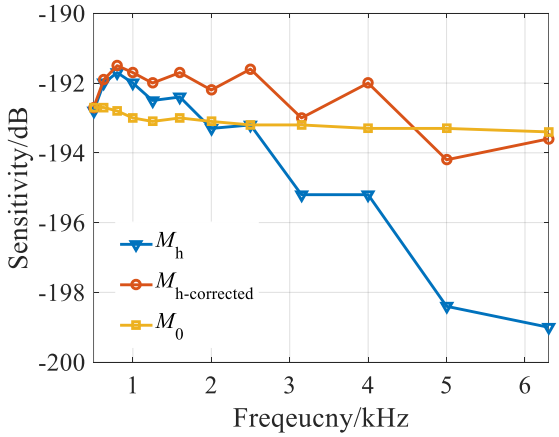


Fig.4: Comparison of measured sensitivity for the RHS-30 hydrophone. The triangular markers represent the sensitivity level measured in air, the circular markers denote the corrected sensitivity level after correction, and the square markers indicate the directly measured sensitivity in water.

CONCLUSIONS

After correcting the sensitivity across different media using the force received at the hydrophone surface, the hydrophone sensitivity can be calibrated in an anechoic chamber. Since the low-frequency cutoff of the free field in an anechoic chamber is significantly lower than that in an anechoic water tank, this approach enables low-frequency calibration of hydrophones and their arrays. Furthermore, this method can also be applied to evaluate the sensitivity of buried hydrophones or those embedded in ice.

REFERENCES

[1] BEATTY L G, GEORGE J D, ROBINSON A Z. Use of the complex exponential expansion as a signal representation for underwater acoustic calibration [J]. The Journal of the Acoustical Society of America, 1978, 63(6): 1782-94.

[2] TRIVETT D, ROBINSON A Z. Modified Prony method approach to echo - reduction measurements [J]. The Journal of the Acoustical Society of America, 1981, 70(4): 1166-75.

[3] ROBINSON S, HARRIS P. Modelling acoustic signals in the calibration of underwater electroacoustic transducers in reverberant laboratory tanks [J]. 1999.

[4] HEYSER R C. Acoustical measurements by time delay spectrometry [J]. Journal of the Audio Engineering Society, 1967, 15(4): 370-82.

[5] CHEN Y, YANG L, ZHOU L, et al. Calibration of hydrophones using a frequency domain filter processing method: theory and experiment [J]. Measurement Science and Technology, 2020, 32(3): 035012.

[6] IEC60565. IEC 60565-1: 2020. Underwater acoustics-Hydrophones-Calibration of hydrophones-Part 1: Procedures for free-field calibration of hydrophones. Edition 1: [S]. 2020.

THE EFFECT OF TEMPERATURE ON ACOUSTIC TRANSDUCER SENSITIVITY AND TRANSMITTING RESPONSE

Caitlin Kean[1] and William Slater[2]

[1]Naval Undersea Warfare Center 1176 Howell St, Newport, RI 02841
[2]Naval Undersea Warfare Center 1176 Howell St, Newport, RI 02841

Abstract: *This study examines the effect of temperature on the receiving sensitivity and transmitting response of several underwater acoustic transducers and hydrophones: the Underwater Sound Reference Division (USRD) Type F30, F42A/D, H48, H52, and H56. The data were collected at the USRD over a five-year period, covering a frequency range of 1 Hz to 250 kHz and temperatures between 3°C and 35°C. The results reveal a notable inverse relationship between temperature and sensitivity below 1 kHz. For hydrophones with a preamplifier, the extent of the inverse correlation varies considerably at low frequencies where the preamplifier gain decays. Above 1 kHz the relationship is irregular, alternating between regions of no clear relationship, inverse relationship, and nonlinear relationship. The differences in sensitivity and transmitting response due to temperature follow similar trends. The findings of this study highlight the importance of calibrating transducers at the temperature they will be used for best accuracy. When temperature-matched calibration data is not available, the findings suggest extra uncertainty is present beyond that reported in calibration certificates.*

Keywords: *Received Sensitivity, Transmitting Response, Calibration, Temperature, Reciprocity, Comparison Calibration, USRD, Transducer, Hydrophone, Projector, NIST, APTF, LOFAC, Uncertainty, Piezoelectric, Coupling Gain, Electronic Impedance.*

1. INTRODUCTION

The Underwater Sound Reference Division (USRD) at the Naval Undersea Warfare Center (NUWC) Division in Newport, RI, provides underwater acoustic measurement and calibration services to government laboratories, industry, and academia both in the United States and internationally. It also operates a standards leasing program, disseminating calibrated acoustic transducers to U.S. and global customers. As the U.S. standardizing authority for underwater sound, the USRD offers measurement services traceable to the International System of Units (SI), with the National Institute of Standards and Technology (NIST) designated as the institute for sound in water [1]. USRD facilities include the Open Tank Facility (OTF), Low-Frequency Facility (LOFAC), and the Acoustic Pressure Tank Facility (APTF). These facilities and standards are supported by a research and development program focused on advancing underwater acoustic measurement techniques.

Previous research has investigated the effect of temperature on transducer sensitivity. Van Buren reported that the impact of temperature on sensitivity was a complex, frequency-dependent function. Ford et al expanded this work to 13 different transducers, showing that more complex designs exhibited larger temperature dependent differences than simple ones. Burghouwt expanded the data set to lower frequencies, showing temperature dependence was significant, particularly near the corner frequency of a built-in preamplifier. All studies have shown cases where the hydrophone's sensitivity is significantly affected. This study expanded the frequency range further, includes projectors, and investigates the contribution of preamplifiers at low frequency.

1.1. TRANSDUCERS

Understanding the design and material properties of the transducers is crucial for interpreting their temperature-dependent behavior. Key characteristics of the transducers analyzed in this research are summarized in Table 1, including housing material, piezoelectric element, preamplifier presence, oil fill, and frequency range. Images and more detailed descriptions are available in the USRD catalog [5].

Transducer	Rubber	Active Element	Preamp (Y/N)	Oil Fill	Frequency Range
H52	Butyl	Li_2SO_4	Y	Castor	20 Hz – 150 kHz
H56	Butyl	PZT-4	Y	Castor	10 Hz – 65 kHz
H48	Butyl	PZT-4	Y	Castor	1 Hz – 20 kHz
F42A	Polyurethane (PRC 1538)	PZT-4	N	N/A	1 kHz – 35 kHz
F42D	Polyurethane (PRC 1538)	PZT-4	N	N/A	6 kHz – 160 kHz
F30	Neoprene	Li_2SO_4	N	Castor	10 kHz – 150 kHz

Table 1: Summary of Transducer Materials.

1.2. PIEZOELECTRIC PROPERTIES

PZT-4, used in Type H48, F42, and H56 transducers, and lithium sulfate, used in Type H52 and F30 transducers, exhibit different temperature sensitivities.

1.2.1. PZT-4

The equation for sensitivity in PZT-4 for a sphere is displayed in equation 1 [7]:

$$\frac{V}{P_0} = \frac{b}{c^2 + c + 1}\left[g_{33}\left(\frac{c^2 + c - 2}{2}\right) - g_{31}\left(\frac{c^2 + c + 4}{2}\right)\right] \quad (1)$$

where V is the open-circuit voltage, P_O is the external pressure, c is the ratio of inner to outer radii, and g_{33} and g_{31} are the electromechanical constants. These constants can be expressed in terms of other material properties (Equations 2) [9].

$$g_{31} = \beta_{33}^T d_{31} = \frac{d_{31}}{\varepsilon_{33}^T}, \qquad g_{33} = \beta_{33}^T d_{33} = \frac{d_{33}}{\varepsilon_{33}^T}. \quad (2)$$

The piezoelectric charge constants d_{31} and d_{33} represent the polarization generated per unit of mechanical stress and ε_{33}^T is the dielectric permittivity. Previous research indicates that, from 0°C to 40°C, d_{31} and d_{33} vary minimally, with $d_{31} \approx -2d_{33}$ [10]. However, ε_{33}^T increases approximately linearly with temperature between 20°C and 40°C [11]. Given these relationships, the sensitivity ($\frac{V}{P_0}$) is approximately inversely proportional to ε_{33}^T, exhibiting a $\frac{1}{x}$ relationship with temperature for PZT-4.

1.2.2. LITHIUM SULFATE

Lithium sulfate used in Type H52 and F30 is also dependent on temperature for sensitivity. For frequencies well below the fundamental resonance of the crystals, the result is a constant equal to -0.0146 dB/°C [2]. This means that the sensitivity is decreasing constantly in the logarithmic scale as temperature increases.

1.3. CASING

All hydrophones in this study are encased in an elastomer to protect the active element from water, as shown in Table 1. The viscoelastic properties of these materials, particularly butyl rubber, are critical to understanding their performance, as these properties are influenced by both temperature and frequency. For example, historical USRD data [6] has demonstrated that temperature variations affect the sensitivity of butyl rubber-encased hydrophones in the 100-150 kHz range due to changes in the material's impedance. Viscoelastic materials exhibit increased molecular mobility at higher temperatures and lower frequencies, resulting in more elastic, rubber-like behavior. Conversely, they become stiffer and glass-like at lower temperatures and higher frequencies, as the reduced time for molecular relaxation restricts movement [12].

2. METHODS

The temperature sensitivity of underwater acoustic transducers was assessed through measurements at two facilities: LOFAC and APTF. LOFAC utilized both the coupler reciprocity method (for H48) and System K (for H52 and H56), while APTF conducted annual reciprocity calibrations. The lower-frequency measurements were performed at 3-5°C, 20°C, and 35-40°C, while the higher-frequency measurements were performed at 3°C, ambient (18°C-20°C), and 30°C. Table 2 summarizes the data collected, outlining the transducer type, facility, method, measured parameter (Free-Field Voltage Sensitivity (FFVS) or Transmit Voltage Response (TVR)), frequency range, number of unique serial numbers tested, total number of repetitions, and years of data collection. The F42A and F30 data represent repeated measurements of the same two units used as projector and reciprocal transducers in APTF. F30 data were only collected in 2022-2023, as it replaced a previous unit.

Unit	Facility	Method	Measurement	Frequency Range	# Units	Total # of Repetitions	Year(s) Collected
H48	LOFAC (Coupler)	Reciprocity	FFVS	1 Hz – 2000 Hz	3	3	2018-2021
H52	LOFAC (System K)	Comparison	FFVS	3 Hz – 2000 Hz	5	5	2019
H56	LOFAC (System K)	Comparison	FFVS	1 Hz – 1600 Hz	5	5	2023
H56	N/A	Coupling Gain	N/A	1 kHz – 80 kHz	1	1	2023
H52	APTF	Reciprocity	FFVS	1 kHz – 250 kHz	4	11	2018-2023
F42D	APTF	Reciprocity	FFVS	1 kHz – 250 kHz	3	9	2018-2023
F42A	APTF	Reciprocity	TVR	1 kHz – 40 kHz	2	38	2018-2023
F30	APTF	Reciprocity	TVR	40 kHz – 250 kHz	2	14	2022-2023
F42A	APTF	Electrical Impedance	N/A	1 kHz – 40 kHz	1	21	2018-2023

Table 2: Summarizing Testing Method for the Different Units.

3. RESULTS

These results show how the sensitivity of a hydrophone or transmitting response of a projector changes for a given change in temperature. The plots display the difference in sensitivity on the vertical axis for a corresponding change from ambient temperature (18-20°C) indicated in the legends. For low-frequency data (below 2 kHz), data are presented for three hydrophones (H48, H52, and H56). High-frequency results (above 2 kHz) were separated into projectors and hydrophones. The hydrophones included H52 and F42D, while the projectors comprised F42A and F30. As H52 had the largest population size, further evaluation was conducted to assess consistency for an individual serial number.

Multiple serial numbers of each model were tested, and the mean differences in sensitivity (or transmitting response) are indicated by the points in Figures 1-6. The standard deviation of the difference across the population of each model is indicated by the error bars. The standard deviation also accounts for measurement uncertainty from random sources. Measurement

uncertainty from systematic sources cancels in the subtraction to calculate sensitivity and transmitting response differences.

3.1. LOW FREQUENCY

The low-frequency results consisted of data collected using the reciprocity coupler (H48) and System K (H52 and H56). Figure 1 illustrates the change in sensitivities for H48, H52, and H56 corresponding to a change from ambient temperature (20°C) to a hotter or colder extreme between 1 Hz and 2000 Hz.

The most notable temperature-dependent sensitivity change occurs in the H52 data from 1 to 60 Hz, as shown in Figure 1. When operating below the H52's rated lower limit of 20 Hz, the sensitivity variation with temperature is substantial and inconsistent across different H52 units. The mean sensitivity changes by 1 dB to 2.5 dB with temperature, and the variation between individual H52s can reach 4 dB. This variability suggests that some H52s exhibit minimal temperature sensitivity at lower frequencies, while others show significant changes. This spread coincides with the preamplifier roll-off, indicating that components controlling the time constant at the preamplifier input are susceptible to temperature variations. Given this correlation with the preamplifier, degradation of preamplifiers in some serial numbers may be a contributing factor. The H48 and H56 also exhibit this trend, albeit to a lesser extent, with the H56 showing 0.5 dB to 1 dB more variation than the H48. These data emphasize the importance of temperature-dependent calibration at low frequencies. Furthermore, a temperature correction derived for one serial number may not be applicable to another, especially below 60 Hz.

From 60 to 800 Hz (Figure 1), all units exhibit a consistent inverse relationship between temperature and sensitivity, with increasing temperature corresponding to decreasing sensitivity. While temperature increases result in negative sensitivity differences (red), temperature decreases result in positive sensitivity differences (gray and black). However, this effect is nonlinear, as the sensitivity difference is not directly proportional to the temperature change. For example, a 10°C temperature reduction (gray) produces a similar sensitivity difference to a 17°C temperature reduction (black) for the H52. These differences are consistent across different serial numbers and trials, as indicated by compact error bars that are less than 0.2 dB in many cases. Unlike the preamplifier roll-off effects below 60 Hz, the consistent temperature effects between 60 and 800 Hz suggest that a single temperature correction may be applicable across different serial numbers within this range. However, this observed relationship deviates from predictions based on the material properties of Li_2SO_4 (for H52) and PZT-4 (for H48 and H56) in section 1.2, which predicted different rates of decrease. It is important to note that the PZT-4 relationship for ε_{33}^{T} is reported for temperatures above 20°C, and the relationship may change below this temperature. Further research is needed to quantitatively connect these relationships.

At the top of Figure 1, the H48 temperature effects remain consistent up to 2000 Hz. The H48 was tested in the reciprocity coupler, designed to operate up to 2 kHz with limited noise. The H52 data only goes up to 800 Hz, as Burghouwt indicated the noise increases above that frequency, making it harder to discern meaningful temperature trends [4]. This system limitation is evident in the H56, where the data from 800 kHz to 1600 kHz appears more scattered, making it difficult to draw conclusions about temperature effects, which are similar in magnitude to the noise.

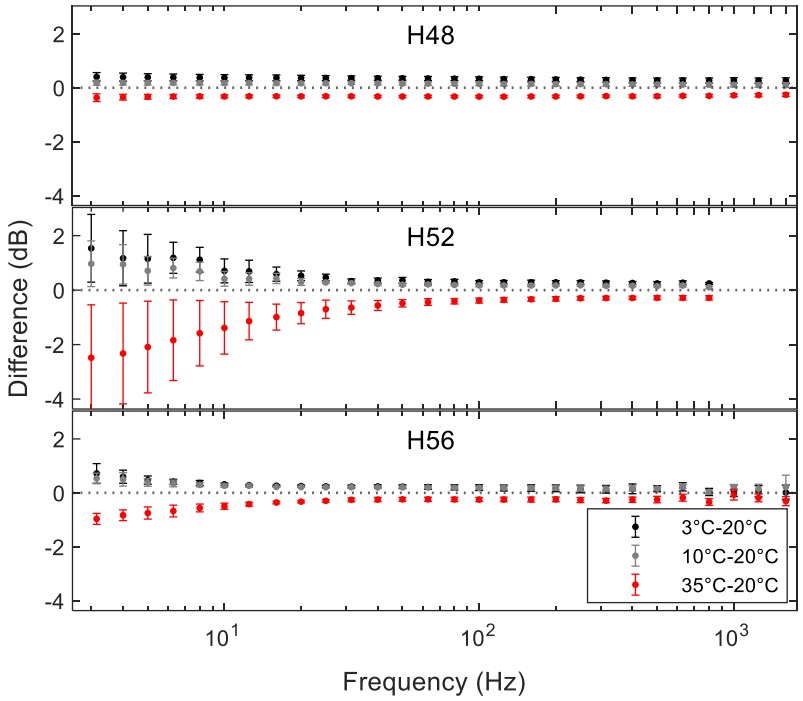


Figure 1: Comparing Mean Sensitivity Differences for Three Temperatures Compared to Ambient for H48, H52, and H56 at Low Frequency.

Figure 2 compares the change sensitivity and coupling gain across the indicated temperature differences for the Type H56. In the coupling gain measurement, an electrical signal is injected directly into the hydrophone's preamplifier and across the sensitive element using an input dedicated to this purpose. The ratio of this input signal over the resulting hydrophone output characterizes the gain of the circuit. Where the electrical properties of the preamplifier and sensitive element are controlling factors, the change in the coupling gain should predict the change in the sensitivity. The coupling gain under-predicts the sensitivity change by about 0.25 dB in the preamplifier roll-off region, and by 0.3 dB to 0.5 dB above that region. Where temperature dependent calibration is unavailable, coupling gain is a valuable prediction tool to reduce what would otherwise be a much larger error in the roll-off region.

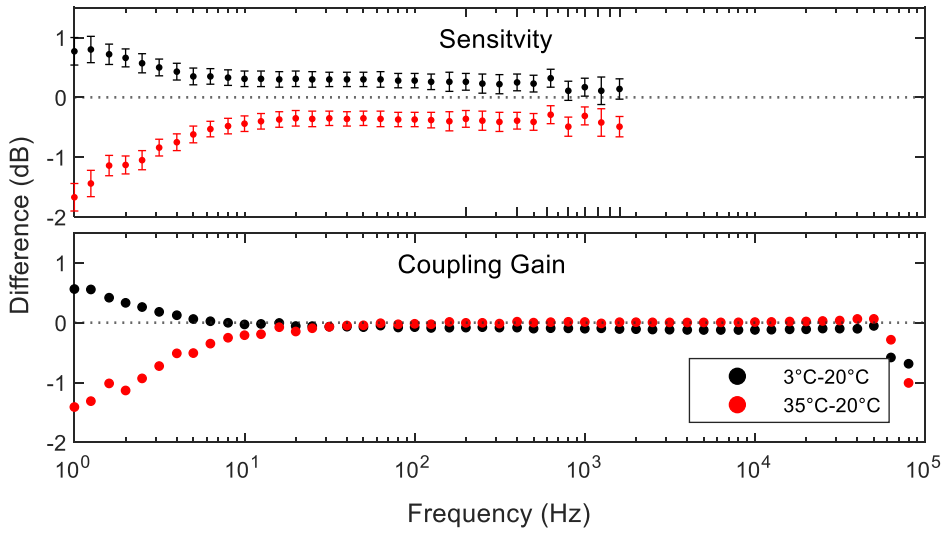


Figure 2: Comparing Sensitivity and Coupling Gain Difference for Two Temperatures Compared to Ambient for H56 SN 27

3.2. HIGH FREQUENCY

The consistent relationship between temperature and sensitivity observed at lower frequencies does not hold across the entire frequency range, as illustrated in Figure 3 for H52 and F42D devices between 1 kHz and 125 kHz. For both H52 and F42D, sensitivity and temperature exhibit an inverse relationship below 7 kHz, similar to that seen in the low frequency data above 60 Hz. This relationship disappears between 7 kHz and 35 kHz, where the overlapping red and black data points (and error bars) indicate no clear correlation. Inverse proportionality re-emerges between 35 kHz and 70 kHz. Above 70 kHz, the relationship becomes increasingly nonlinear, with sensitivity decreasing under both positive and negative temperature changes from ambient[1].

Consequently, to minimize calibration uncertainty, each unit should ideally be calibrated at its operating temperature, as a simple temperature correction formula is insufficient. Below 100 kHz, the temperature-induced sensitivity changes are within 0.5 dB, comparable to the measurement uncertainty of a calibration in this frequency range. However, the non-zero means calculated across multiple trials and units suggest a small, temperature-dependent change in sensitivity. Using these hydrophones at a temperature other than the calibration temperature would introduce an additional uncertainty of approximately 0.2 dB to 0.5 dB. Above 100 kHz, the effect is more pronounced, potentially adding up to 3.5 dB of uncertainty when operating at a different temperature. This unpredictable behavior above 100 kHz may be attributed to the temperature-dependent properties of the rubber housing materials, as discussed in Section 1.3, highlighting the need for further research on the combined effects of high frequencies and temperature on these materials.

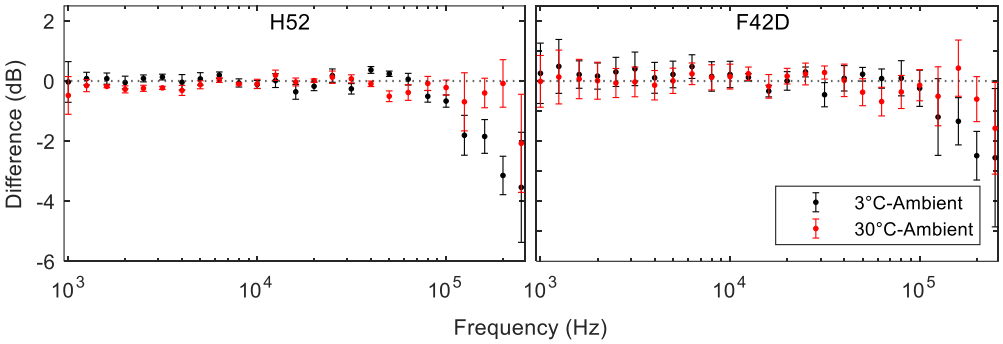


Figure 3: Comparing Mean Sensitivity Difference for Two Temperatures Compared to Ambient for H52 and F42D at High Frequency.

Analysis of the H52 dataset, comprising 11 calibrations spanning 5 years and 4 serial numbers, demonstrates the consistency of the sensitivity-temperature relationship across the H52 model range. Figure 4 compares serial number 47 (right graph) to the aggregated results of the other three serial numbers (left graph). Isolating other serial numbers and comparing them with the remaining three shows a similar trend. The similarity between these graphs below 100 kHz, in both pattern and error bar size, suggests a consistent sensitivity-temperature variation across H52s. This implies that the results from a single unit can be extrapolated to other H52 units, potentially simplifying testing procedures and reducing the need for extensive testing of multiple units.

[1] In these datasets ambient temperature is 18°C or 20°C.

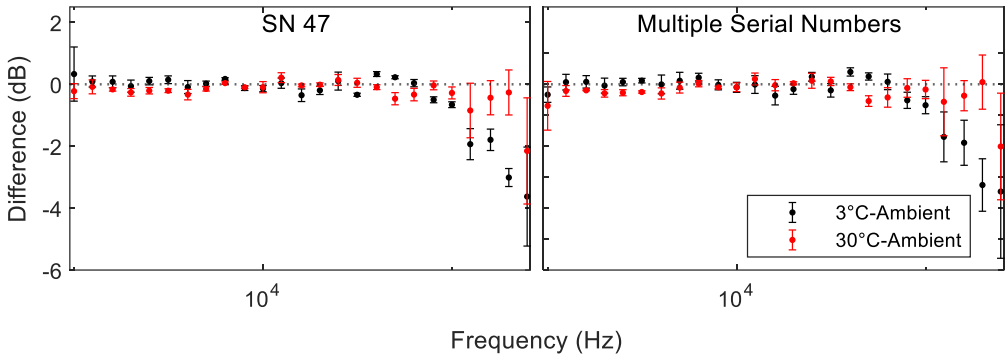


Figure 4: Comparing Mean Sensitivity Difference for Two Temperatures Compared to Ambient for One Serial Number (H52 SN 47) to Multiple Serial Numbers.

Figure 5 presents the results for the difference in transmit response compared to ambient conditions at various temperatures. The Type F42A (left graph) results exhibit a similar pattern to the sensitivity differences observed for the similarly constructed Type F42D hydrophone in this frequency range. In contrast, the Type F30 projector exhibits a larger temperature difference at 40 kHz, with the temperature difference decreasing from 40 to 63 kHz. Above 70 kHz, the uncertainty increases significantly. Separating the serial numbers (middle and right graph) reveals inverted behaviors, where one unit displays higher sensitivity at hotter temperatures, while the other displays lower sensitivity compared to ambient. This contrasts with the H52, where each serial number displayed the same pattern, suggesting that, for the Type F30 transducer, temperature effects cannot be extrapolated from one unit to another. Therefore, each unit requires individual testing to characterize its temperature dependence.

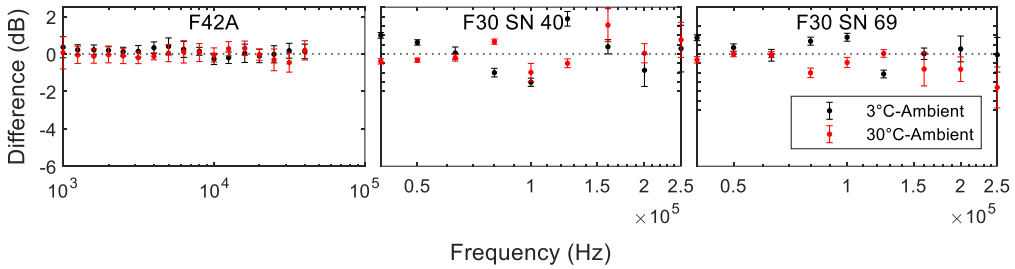


Figure 5: Comparing Mean Transmitting Response Difference for Two Temperatures Compared to Ambient for Projectors F42A (left), F30 SN 40 (middle) and SN 69 (right).

Since electrical impedance can indicate unit stability, the temperature-induced deviation in decibels was analyzed for the electronic impedance of the Type F42A projector, encompassing 21 calibrations spanning 5 years. Figure 6 compares the difference in transmitting response for the Type F42A with the difference in electrical impedance across the indicated temperatures. The difference, reported in decibels, remains relatively constant from 1 to 30 kHz, with 3°C exhibiting an electrical impedance 0.1 dB greater than ambient and 30°C exhibiting an electrical impedance 0.3 dB less than ambient. Similar to the coupling measurement, electrical impedance appears to be one of the variables influencing temperature variation, but not the only factor influencing change in transmitting response. The differences in transmitting response and electronic impedance are closely correlated up to approximately 10 kHz, above which the temperature dependence of the transmitting response inverts until around 20 kHz. Thus, while electrical impedance measurements can provide insights into the effects of temperature on a unit, they cannot replace the need for calibrating specific units at their intended operating temperature.

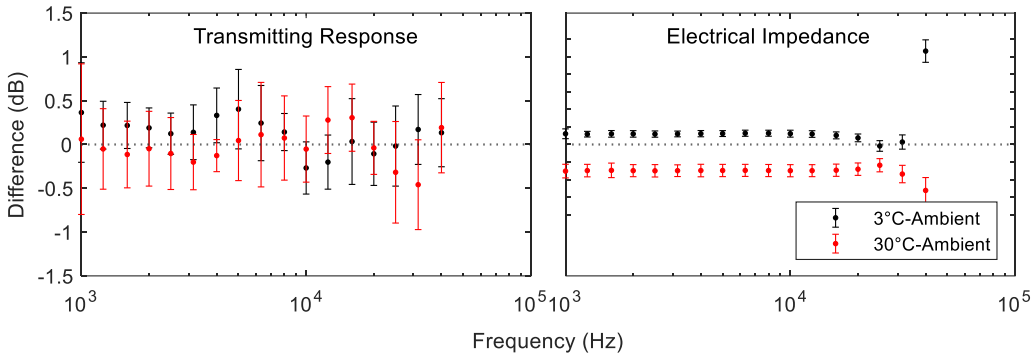


Figure 6: Comparing Mean Electrical Impedance Difference to Transmitting Response at Two Temperatures Compared to Ambient for F42A SN 141.

4. CONCLUSION

This study demonstrates the complex and often unpredictable effects of temperature on the sensitivity and transmitting response of USRD underwater acoustic transducers. The relationship between temperature and sensitivity is inverse below 1 kHz. Above 1 kHz there are regions of no clear relationship, inverse relationship, and nonlinear relationship, suggesting the contribution of factors beyond the temperature dependence of the lithium sulfate and PZT-4 piezoelectric properties. The findings reveal that the temperature dependence varies with transducer type and material, frequency, and the presence of preamplifiers, precluding the use of simple temperature corrections.

A key conclusion is that accurate underwater acoustic measurements require careful temperature management and, ideally, calibration at the intended operating temperature. Particularly below 20 Hz (with preamplifier-equipped hydrophones) and above 100 kHz, relying solely on nominal calibration data can introduce 1 dB to 3.5 dB of calibration error, significantly greater than the calibration uncertainty provided on USRD calibration certificates. At low frequencies when preamplifier roll-off is a controlling factor of the hydrophone sensitivity, a coupling gain style measurement can characterize all but about 0.25 dB of this error. At high frequencies, impedance measurements can predict some of this error.

Ultimately, this research reinforces the need for caution when interpreting underwater acoustic data acquired at temperatures different from the calibration temperature. Further research in methods for accounting for temperature effects on sensitivity and transmitting response could reduce the need for time-consuming and expensive testing, at least for some consistent and well-understood hydrophones, increasing the availability of accurate underwater acoustic measurements across a wide range of applications.

REFERENCES

[1] S. E. Crocker, E.G. Westerholm, and A.R. Garceau, “Uncertainties for Underwater Acoustic Transducer Calibrations Performed in Simulated Oceanic Environments in the Frequency Range of 1 kHz to 250 kHz,” Naval Undersea Warfare Center Division Newport, Technical Document 12316, 2019.

[2] A.L. Van Buren, R.M. Drake, and A.E. Paolero: “Temperature dependence of the sensitivity of hydrophone standards used in international comparisons”, *Metrologia* 36, 281-285 (1999).

[3] B. Ford, S. Robinson, and J. Ablitt: “A Study of the stability exhibited by hydrophones when exposed to variations in temperature and hydrostatic pressure”, *Proceedings of Meetings on Acoustics* 44, 070024 (2021).

[4] J.A. Burghouwt: “The effect of pressure and temperature on the sensitivity of a type H52 hydrophone” in PennState Theses and Papers in the Graduate Program in Acoustics (2019).

[5] Ivey, L. E. (1991, April). USRD Transducer Catalog. Orlando, FL, USA: Naval Research Laboratory.

[6] USRD. (1982, May). Underwater Electroacoustic Standard Transducers Catalogue. Orlando, FL, USA: Naval Research Laboratory.

[7] Ivey, L. E., “NRL-USRD Series F42 Omnidirectional Standard Transducers”, Orlando, FL, USA: Naval Research Laboratory, NRL Memorandum Report 3969, 1979.

[8] Ivey, L. E., “High-Pressure Piezoelectric Ceramic Hydrophone for Infrasonic and Audio Frequencies”, Orlando, FL, USA: Naval Research Laboratory, NRL Report 7260, 1971.

[9] C.H. Sherman and J.L. Butler, Transducers and arrays for underwater sound, (Springer, 2007).

[10] M.W. Hooker, “Properties of PZT-Based Piezoelectric Ceramics Between -150 and 250°C”, Lockheed-Martin Engineering & Science Co., NASA/CR-1998-208708, 1998.

[11] L.Tang and W. Cao, “Temperature dependence of self-consistent full matrix material constants of lead zirconate titanate ceramics”, *Applied Physics Letters* 106, 052902 (2015).

[12] R.N. Capps, “Dynamic Young’s moduli of some commercially available polyurethanes”, *The Journal of the Acoustical Society of America* 73, 2000 (1983).

Application of 2E-2U Method for Free-Field Underwater Acoustic Calibration of Hydrophones and Projectors in a Reverberant Laboratory Test Tank

Ata Can ÇORAKÇI[1], Alper BİBER[2], Talha SAYDAM[3], Osman Said BİŞKİN[3], Serkan AKSOY[3]

[1] TÜBİTAK Marmara Research Center, Underwater Acoustic Laboratory Kocaeli, TÜRKİYE
[2] Ulus Elektronik, Kocaeli, TÜRKİYE
[3] Department of Electronics Engineering, Gebze Technical University, Kocaeli, TÜRKİYE

Ata Can ÇORAKÇI, TÜBİTAK Marmara Research Center, Underwater Acoustic Laboratory Kocaeli, TÜRKİYE, cancorakci@gmail.com – Current affiliation: CTBTO, Wien, Austria

Abstract: *In this study, an implementation of a Two-Equation Two-Unknown (2E-2U) method for calibrating hydrophones and projectors below 1 kHz in a reverberant laboratory test tank is described. At low frequencies, performing necessary amplitude and phase measurements for sensor calibration in a test tank is a challenging job due to insufficient echo-free time of the tank. To overcome this challenge, the 2E-2U method has been applied to the signals received during the calibration measurements that are windowed in time. Thus, calibration measurements have been made possible down to 250 Hz. Initially, received sensitivity for the hydrophone is calculated and validated through comparisons with pressure calibration in a pressured coupling chamber. Secondly, a transmitting voltage response (TVR) of a segmented Class V Flextensional projector is calculated and compared with the calibration data obtained from a method defined in relevant standards.*

Keywords: *Underwater acoustic hydrophone and projector calibration, low frequencies, 2E-2U method.*

1. INTRODUCTION

In underwater acoustic applications, calibrations of hydrophones (receivers) and projectors (transmitters) are extremely important for accurate measurements, described in relevant standards [1-3]. Comparison and reciprocity free-field calibration methods are commonly used for this aim [4]. These methods may require open-water test facilities depending on frequency requirements, especially at low frequencies where free-field conditions are no longer valid due to boundary reflections of the test tank facility. However, the open-water test facilities pose disadvantages in terms of costs, maintenance, and control. To overcome these disadvantages, it is preferred to conduct measurements in the laboratory test tanks, considering the limits of low frequencies. In practice, if the frequency of interest meets the free-field acoustic conditions, the free-field calibration method can be applied in the laboratory test tanks. However, generally at low frequencies below 1 kHz, an initial transient of the receiver and signals reflected by the boundaries of the test tank surfaces overlap the useful part of the received signal, preventing accurate measurement due to an insufficiently clear signal. Although various methods and algorithms have been proposed to overcome this issue [5-10], their implementation is not practical due to complex mathematical operations required. Therefore, in this study, an application of a fundamental algorithm, known as the Two-Equation Two-Unknowns (2E-2U) method, which is used originally in solving low-frequency electromagnetic problems, has been proposed as an efficient alternative.

2. TWO-EQUATIONS TWO-UNKNOWN (2E-2U) METHOD

The 2E-2U method is based on solving two-equations with two-unknowns to calculate the amplitude and phase of a single-frequency signal. This method allows for the calculation of amplitude and phase information for a signal with a quarter period within a stable section after the transient regimes have ended and before the reflections from the boundaries. For instance, in the case where the investigated part of signal is less than a full period, the Fourier method cannot be used because it requires at least one full period to be run after the transient regime for amplitude and phase calculations. According to the 2E-2U method, the acoustic pressure values measured at times t_1 and t_2 denoted as u_1 and u_2 respectively, are written as follows

$$u_1 = A\sin(\omega t_1 + \theta) \tag{1}$$

$$u_2 = \mathrm{A}\sin(\omega t_2 + \theta) \tag{2}$$

where A represents the amplitude, ω the angular frequency and θ the phase of the fields [11-13]. Since A and θ are two unknowns in two-equations given in equation (1) and (2), the method is referred to as the Two-Equations Two-Unknows (2E-2U) method. The solution of equation (1) and (2) for the two different time values $t_1$and t_2 are

$$A = \left|\frac{u_1}{\sin(\omega t_1 + \theta)}\right| \tag{3}$$

$$\theta = \arctan\left(\frac{u_2\sin(\omega t_1) - u_1\sin(\omega t_2)}{u_1\cos(\omega t_2) - u_2\cos(\omega t_1)}\right). \tag{4}$$

Here, a clear received signal is essential for accurately extracting A and θ. To achieve this, any DC offset present in the measured signals must be removed following the filter process.

3. REVERBERANT TEST TANK MEASUREMENTS

The 2E-2U method is applied to calibrate the hydrophone within the frequency range of 250 Hz to 1.4 kHz, and the projector within the range of 250 Hz to 11 kHz at the TÜBİTAK Marmara Research Center Underwater Acoustic Laboratory. The calibration measurements are conducted at 15 discrete frequencies for the hydrophone and 24 discrete frequencies for the projector. The reverberant laboratory test tank used for these measurements has dimensions of 15 meters in length, 10 meters in width, and 7.5 meters in depth. The water temperature in the tank is approximately 20 °C.

In the measurements, a time-gated signal (T_{TG}) is defined using a direct arrival time (τ_{da}), an initial transient duration (T_{it}) and a tank echo-free time (τ_{ef}) as follows

$$T_{TG} = \tau_{ef} - (\tau_{da} + T_{it}) \tag{5}$$

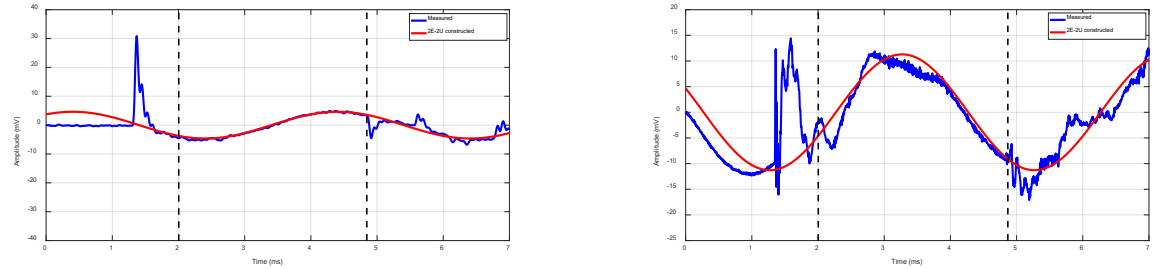
where applying the 2E-2U method to the signals measured during the calculated T_{TG}, amplitude and phase values at low frequencies can be determined in the reverberant test tank. The general view of the reverberant test tank and the pressure coupling chamber [14] used for comparison measurements is shown in Figure 1. The general views of the HBK 8104 hydrophone element, which are used for calibration measurements, and the newly developed segmented Class V Flextensional projector element are presented in Figure 2. The transmitter and the receiver are located at the position of 6.5 m in length, 5 m in width, 3.5 m in depth and 8.5 m in length, 5 m in width, 3.5 m in depth, respectively. This setup corresponds to a reverberant test tank with the echo-free time of approximately $\tau_{ef} = 4.8$ ms and the direct arrival time is $\tau_{da} = 1.3$ ms. On the other hand, T_{it} is related to the quality factor of the sensors. Although their exact values may not be known initially, its effect can be observed in the measurement results. As part of this study, the measurement results at four different frequencies for the hydrophone and the projector (shown in blue), along with the compared results obtained by applying the 2E-2U method (shown in red), are presented in Figure 3 [15]. In the graphs, the signal's direct arrival at the receiver τ_{da} and the tank echo-free time τ_{ef} are indicated with dashed lines. Within the first dashed line, the initial transient effects of the sensors can be clearly observed. The interval between the two dashed lines corresponds to the steady-state regime of the signal. In principle, the 2E-2U method can be applied using the pressure values at any two-time instances within this interval. In this study, to minimize potential undesired interference effects from the left and right parts of the signal, the 2E-2U method is applied using the pressure values taken from two-time instances near the center of the signal.



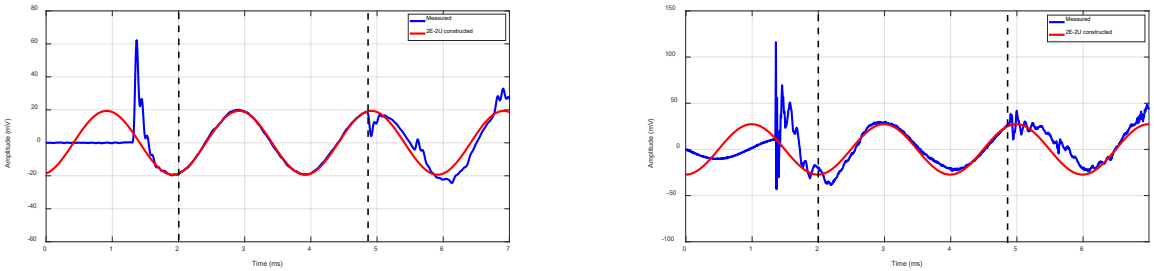
Fig 1: Reverberant test tank (left) and pressure coupling chamber (right).



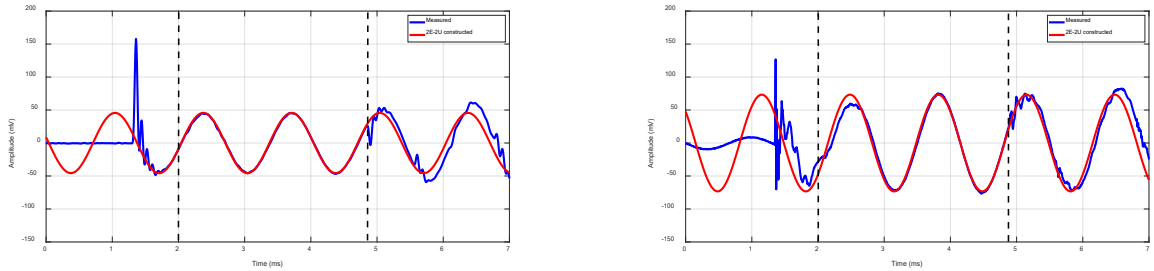
Fig 2: HBK 8104 hydrophone (left) and segmented Class V Flextensional projector (right).



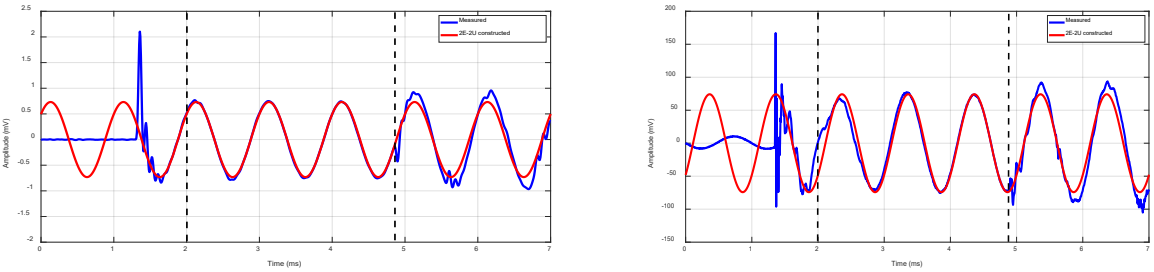
250 Hz



500 Hz



750 Hz



1000 Hz

Fig 3: Constructed signals with the 2E-2U method (in red) and the measurement results (in blue) at four discrete frequencies for the hydrophone element (left) and the projector element (right).

4. CALIBRATION OF TRANDUCERS

The calibration calculations are performed by the methods described in the relevant standards [1-3]. For the hydrophone, the calibration results are calculated with the following equation

$$M_H = \frac{V_H}{V_R} \times M_R \tag{6}$$

where M_H and M_R are the free-field receive sensitivity of the unknown and calibrated sensor, respectively. V_H and V_R are the open-circuit output voltages of the unknown and calibrated sensor, respectively. For the projector, the calibration results are calculated with the following formula

$$S_{V,P} = \frac{d \times V_{PH}}{M_H \times V_P} \tag{7}$$

where $S_{V,P}$ is the transmitting voltage response (TVR) of the projector, d is the distance between the projector and the calibrated hydrophone, $V_{P,H}$ is the open circuit voltage at calibrated hydrophone from a projector, M_H is the free-field receive sensitivity of calibrated hydrophone and V_P is the projector electrical drive voltage.

Figure 4 (left) compares the calibration results of the HBK 8104 hydrophone obtained from the measurements in the test tank and the pressure coupling chamber. The plot shows the hydrophone sensitivity levels derived from the 2E-2U reconstructed signals in the test tank (circles) and from the pressure calibration method in the pressure coupling chamber (squares). Figure 4 (right) shows also a comparison of the calibration results for the segmented Class V Flextensional projector obtained using the 2E-2U method and the reference calibration measurements conducted in the test tank. The plot illustrates the TVR of the projector, derived from the 2E-2U reconstructed signals (circles) and from direct measurements in the test tank (squares). The error bars in the figures represent the uncertainties for a cover factor of $k = 2$ [16].

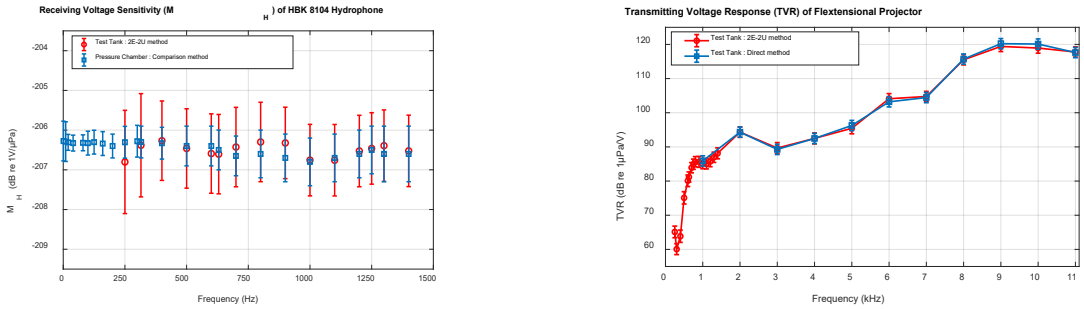


Fig 4: The 2E-2U calculated M_H's of the hydrophone in the tank (circles) and the pressure calibration method (squares) in the chamber (left). The 2E-2U calculated TVR's of the projector in the tank (circles) and the calibration with the direct measurement in the tank (squares) (right).

In this context, the uncertainties arise from three main sources: i) the distance between the transmitter and the receiver, ii) the failure to achieve steady-state conditions, and iii) a low signal-to-noise ratio caused by undesired internal and external noise. The first source of uncertainty is minimized by carefully positioning the transmitter and receiver. Applying the 2E-2U method helps

to reduce the second source of uncertainty. Therefore, the most critical source of uncertainty is unwanted noise. The internal noise is minimized using low-noise equipments. The external noise sources can be classified as broadband noise and noise at specific frequencies. Coherent averaging reduces the impact of the broadband noise. For this reason, the measured signals are averaged at least 50 times before being used as the input for the 2E-2U method. No external noise is encountered at the frequencies corresponding to the source signal frequency in the measurement setup. Additionally, since effective electromagnetic compatibility (EMC) filters are used in the laboratory power supply lines of the experimental setup, no noise issues are observed due to unwanted harmonics of the power supply frequency or other sources of electrical interference.

5. CONCLUSION

This study presents the application of the 2E-2U method for calibrating underwater acoustic transducers in the reverberant laboratory test tanks at frequencies below 1 kHz. The results obtained using the 2E-2U method are compared with the existing methods in the literature, showing good agreement for both the hydrophone and the projector. The main advantage of the 2E-2U method, enabling researchers to perform underwater acoustic calibration measurements below 1 kHz in laboratory test tank, lies in its simplicity and minimal requirements for memory and computation. This method is particularly significant as it eliminates the need for costly and logistically demanding facilities such as the open-water test setups and the anechoic test tanks. As a future study, the application of a NE-NU (N Equations, N Unknowns) method may be considered for broadband calibration in reverberant laboratory test tanks. The NE-NU method can be applied to multiple source frequencies simultaneously. Thereby, it significantly reduces the time required for broadband underwater acoustic calibration measurements.

REFERENCES

[1] Procedures for Calibration of Underwater Electroacoustic Transducers, ANSI/ASA S1.20:2012. New York: American National Standards Institute, 2012.
[2] Underwater Acoustics - Hydrophones - Calibration of Hydrophones - Part 1: Procedures for Free-Field Calibration of Hydrophones, IEC 60565-1:2020 Geneva: International Electrotechnical Commission, 2020.
[3] Underwater Acoustics - Hydrophones - Calibration of Hydrophones - Part 2: Procedures for Low Frequency Pressure Calibration, IEC 60565-2:2019 Geneva: International Electrotechnical Commission, 2019.
[4] **Bobber R. J.,** *Underwater Electroacoustic Measurements*, Peninsula Publishing, 1990.
[5] **Piquette J. C.,** Applications of the method for transducer transient suppression to various transducer types, *J. Acoust. Soc. Am.*, 94-2, 646-651, 1993.
[6] **Ludwig G., Brendel K.,** Calibration of hydrophones based on reciprocity and time-delay spectrometry, *IEEE Transactions on Ultrason. Ferroelectr. Freq. Control*, 35-2, 168-174, 1988.
[7] **Cochard N., Lacoume J. L., Arzelies P., Gabillet Y.,** Underwater acoustic noise measurement in test tanks, *IEEE Journal of Ocean Eng.*, 25, 516–522, 2000.
[8] **Zhang Y-M., Tang R., Li Q., Shang D-J.,** The low frequency sound power measuring technique for an underwater source in a non-anechoic tank, *Meas. Sci. and Tech.*, 29, 035101, 2018.

[9] **Isaev A. E., Nikolaenko A. S., Chernikov I. V.,** Suppression of reverberation distortions of a receiver signal using the water tank transfer function, *Acoust. Phys.*, 63, 175–84, 2007.
[10] **Robinson S. P., Hayman G., Harris P. M., Beamiss G. A.,** Signal-modelling methods applied to the free-field calibration of hydrophones and projectors in laboratory test tanks, *Meas. Sci. and Tech.*, 29, 085001, 2018.
[11] **Furse C. M.,** Faster than Fourier: Ultra-efficient time-to-frequency domain conversions for FDTD, *IEEE Antennas Propag. Mag.*, 42–6, 24–33, 2000.
[12] **Saydam T., Aksoy S.,** Error analyses of the two-equations two-unknowns method for low-frequency FDTD problems, *Journal of Comp. Elect.*, 19-4, 1573-1578, 2020.
[13] **Saydam T., Aksoy S.,** Reliability analysis of N-equations N-unknowns method for the solution of the finite-difference time-domain (FDTD) problems, *Frequenz*, 76-5/6, 331-335, 2022.
[14] **Biber A., Golyk O., Çorakçı A. C.,** Pressure coupling chamber for hydrophone calibration, Patent No: 11,953,640 B2, *United States Patent,* 2023.
[15] **Çorakçı A. C., Biber A., Saydam T., Aksoy S.,** Application of 2E-2U method for free-field underwater calibrations of hydrophones and projectors in a reverberant laboratory test tank, *Meas. Sci. and Tech.,* 35, 085002, 2024.
[16] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML, Evaluation of Measurement Data: Guide to the Expression of Uncertainty in Measurement (Joint Committee for Guides in Metrology, JCGM 100), Paris, France: Bureau International des Poids et Mesures – BIPM, 2008.

Commercial Broad Range Calibration by Multi-Method Synthesis

Nicolas Wood[1], Emma Carline[2]

[1] Ocean Sonics, 110 Parkway Drive, Truro Heights, NS, Canada
[2] Ocean Sonics, 110 Parkway Drive, Truro Heights, NS, Canada

Nicolas Wood
Ocean Sonics
110 Parkway Drive
Truro Heights, B3L1N8
Canada
Nicolas.wood@oceansonics.com

Abstract: *There are a variety of challenges that come with calibrating hydrophones. One of those is that it is not possible to calibrate their full frequency range using a single test fixture. In-tank calibrations can be done down to the low-kHz range (tank size permitting) but there are many applications for which there are still several decades of frequency range of interest below that point. Open water calibrations can be done for broader frequency ranges but cost limits their application in a commercial environment. This paper will discuss experimentation with combining additional lab-based calibration methods to produce a synthesized calibration ranging from 1 Hz up to 200 kHz in a production setting. The work done compares tank-based calibrations by comparison and low frequency pistonphone measurements to comparison calibrations done using a vibrating water column apparatus. Coiled waveguides were tested previously, thought reliable results were not achieved. Favourable results were produced in terms of achieving good agreement between the calibration fixtures in their respective frequency ranges.. The vibrating water column fixture worked in a limited range though an improved fixture may permit increased useable frequency range.*

Keywords: *Calibration, Low frequency*

1. INTRODUCTION

Ocean Sonics designs and builds the icListen, a broadband digital ethernet-based hydrophone with edge computing capabilities. Accurate estimates of the hydrophone's sensitivity are essential both in-house for quality control and for the end user. The calibration certificate provides the hydrophone sensitivity at select frequencies across the frequency range of the hydrophone. There are some limitations with the current calibration method however, with regards to the range of frequencies that can be calibrated.

The current calibrations use a tank-based method of calibration by comparison. A hydrophone unit under test (UUT) is compared with a reference unit with known sensitivity. An uncalibrated projector, the Ocean Sonics-made icTalk, is placed in the centre of the tank with the two hydrophones equally spaced on either side of the projector. The projector plays a sequence of short-duration tones at a series of frequencies between 10 and 200 kHz for icListen HF hydrophones and between 5 and 12.5 kHz for the icListen AF hydrophones. The layout for tank calibrations is shown in Figure 1.

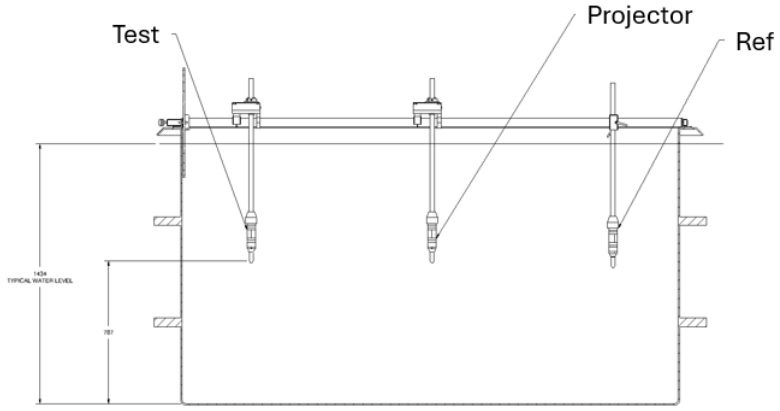


Figure 1: Calibration tank layout with Test and Reference hydrophones and icTalk projector.

To calculate the Receive Voltage Sensitivity (RVS) of the UUT, we assume that the sound pressures measured at the UUT and the reference unit are the same. Therefore, we get the following equality:

$$RVS_{UUT}(f) = RVS_{REF}(f) - V_{REF}(f) + V_{UUT}(f) \quad (1)$$

Where $V_{REF}(f)$ is rhe received voltage level at the reference hydrophone at frequency f in dBV, $RVS_{REF}(f)$ is the sensitivity in dBV/µPa for the reference unit at frequency f, and $V_{UUT}(f)$ and $RVS_{UUT}(f)$ are the corresponding quantities for the UUT.

The goal of the calibration is to calculate $RVS_{UUT}(f)$ for all frequencies f in the range of target frequencies. The calculation is straightforward, but the challenge is obtaining an accurate estimate of $V_{REF}(f)$ and $V_{UUT}(f)$ through in-tank measurements. The standard range calibrations at Ocean Sonics as mentioned previously follow the guidelines provided in the IEC 60565 standard [1] and can measure these parameters to a satisfactory level of precision.

In-tank calibrations are limited in useful frequency range by the tank dimensions. This is due to requirements on the data window. According to the IEC60565 standard, the data window should generally include at least one full period of the waveform in the steady-state period of the waveform while excluding the portion of the signal containing reflections from the tank surfaces. The Ocean Sonics calibration tank is 2.74m long, 1.52

m wide, and the water depth is typically 1.41 m. The hydrophones are each spaced 1.00 m from the projector and typically 0.81 m deep. Therefore, when the projector emits a tone, the time between the first arrival and the arrival of the first reflection is 380 µs. However, the first part of the signal contains the transient so the minimum frequency that can be measured while meeting the data window requirements is approximately 4 kHz. While a standard pistonphone can provide a point at 250 Hz or similar, there is a desire to calibrate across the full lower frequency range.

Effective low frequency calibration methods exist but we have the additional practical requirements that the calibration methods use our own equipment to facilitate timing precision with the test hydrophones and that they permit frequent calibrations with good ease of use. Open water techniques or the laser pistonphone method and coupler reciprocity method described in [2] are not appropriate for the above reasons.

A method using a novel coil-waveguide method was previously investigated for low frequency calibrations at Ocean Sonics [3] with a reasonable form factor of the coil-shaped wave guide. In principle the calibrator worked but it was not possible to perform calibrations with the precision of the in-tank calibrations.

The primary objective of the work at hand was to investigate additional calibration methods to expand the frequency range of Ocean Sonics calibrations. When selecting low frequency calibration methods, consideration was given to the ease of use of moving a UUT hydrophone between calibration fixtures and the ability to use the same icTalk projector electronics for the excitation signals. Using the icTalk enables easy timing synchronization with the hydrophones and maintains the same methods of programmatic control of all the devices in the different calibration techniques.

Two methods were considered that complied with these requirements: a vibrating column test fixture and a modification of the in-tank method to allow lower frequencies. Together it was hypothesized that these methods could extend calibrations from a minimum of 5 kHz, down to 1 Hz.

A vibrating column calibration fixture uses a vertically moving container to vary the pressure experienced by the UUT. This can be done with a physics-based approach with an actuator with positional feedback with pressure levels calculated from the actuator displacement, or with the use of a reference hydrophone.

For the modifications to the in-tank measurements, two approaches were considered which are essentially relaxations of the data window requirements.

The first modification was to reduce the size of the data window, allowing partial periods of the waveforms. In calculating the rms of the received voltage, the current method repeats each tone frequency twice where the second tone is 90-degree phase shifted. The signals are then squared and point-wise summed, giving a DC-signal with amplitude equal to that of the tones. This is possible due to the identity $cos^2(\theta) + sin^2(\theta) = 1$. In theory, this should give the same amplitude even if the data window length is substantially less than a full period.

The second modification was to allow the sampling window to extend into the reflection period of the signal. While at first this appears ill-advised, there are a couple of reasons that this may be viable for low frequency tones in our setup. The first of these is that for high frequencies the reflections can appear highly random which is likely due to small (mm range) variations in hydrophone or projector positioning equating to large variations in the phase difference of the reflected signals. However, for very low frequencies the relative phase shift between the two positions in the tank is very small so the received reflections interfere with the generated waveform in near-identical manners. This was already observed in previous tests. A remaining issue was the presence of phase inversions due to reflections as well as stepwise transients that are present at the beginning and end of the

windowed drive signal. The proposed solution was to apply heavy bandpass filtering. While this greatly modifies the signal amplitude, the fact that the UUT signal is a scalar multiple of the reference signal suggests that filtering would preserve this relationship and the sensitivity calculation in equation (1) would still hold. It seemed that it would be at least worth trying the method since it is easy to do with our current tank setup.

2. METHODS

The two calibration methods were tested and evaluated at the Ocean Sonics facility in Truro Heights, Nova Scotia.

Vibrating Column Test Fixture

A vibrating column calibration fixture was designed and built using readily available components. Due to the requirement to follow a similar procedure to the in-tank calibrations, the calibration fixture was designed to use a second reference hydrophone. A voice coil motor (VCM) was used to drive the bottom of the container. The VCM was able to be driven from a modified icTalk boardset which meant that interfacing with the calibration fixture was unchanged from the in-tank calibrations. The vibrating column tests were done with a steady state 5 second tone at each frequency step.

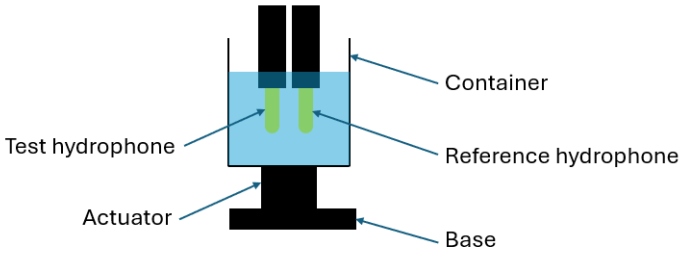


Figure 2: Vibrating column test fixture using a reference hydrophone and no positional feedback.

Initial tests were conducted to first confirm that the test fixture could produce acceptable sound levels at the frequency ranges of interest. Recordings of steady-state tones were made and a processing script was designed to calculate the recorded signal amplitudes. The calibration fixture was tested in the frequency range of 5 Hz to 200 Hz.

The hydrophone mount was designed to enable modification of the hydrophone position within the container to allow testing for any variation of signal intensity with the hydrophone position.

In-Tank Calibration Processing Modifications

To test the modifications of the in-tank calibration, the standard icTalk projector was replaced by the LF icTalk, which used a modified driver and low frequency transducer. This was necessary as the standard icTalk has a lower frequency limit of approximately 4 kHz. It was found that the LF icTalk did not produce clean tones; they seem to be contaminated with high frequency signals, possibly an issue with the amplifier. However, despite this, by filtering and comparing results with a standard icTalk calibration it was found that the sensitivity measurements were not greatly affected. For example, at 10 kHz, the LF icTalk-measured sensitivity was found to be -178.0 ± 0.1 dB re 1 V/uPA, over 4 different runs. The standard calibration gave a value of -178.0 dB V re 1 uPa. The filter used was a 4-pole Butterworth bandpass filter with lower cut-off at 75% of the tone

frequency and upper cut-off at 125% of the frequency. The filter was applied forward and then backward to preserve phase. Filtering heavily modified the waveform, and the effect of doing of this should be studied more in the future.

To evaluate the in-tank calibration modifications, we first looked at the effect of shortening the data window. For a valid test, we tested tone frequencies where the true sensitivity could be measured using the standard approach, i.e., there was more than one cycle of the tone available before the arrival of the first reflection. Tone frequencies of 5.1, 5.4, 5.7, and 6 kHz were selected for testing since these are near the bottom of the current in-tank calibration frequency range. For the test, the data window was progressively shortened, from 1 full cycle to an extreme 5% of a cycle. The idea was that if the sensitivity did not change much for the 6 kHz tone despite using a short data window, then the same should be true for lower frequency tones, validating the use of partial cycle data windows for low frequency calibrations.

Secondly, we looked at the effect of allowing the portion of the signal that contains reflections to be included in the data window. Again, the testing was conducted for tones with frequencies of 5.1, 5.4, 5.7, and 6 kHz. The data window length was fixed at 1 cycle and progressively shifted to the right by 31 μs to capture more and more of the reflection-containing part of the signal.

Finally, we tested the in-tank method modifications for frequencies between 200 Hz and 6 kHz, collecting data over multiple runs. For each tone frequency, the data window length varied and sometimes included reflections. Though the true sensitivity in this frequency range was not known, we wanted to see how much variability there was between runs. If there was a large difference in the sensitivity estimates between runs, that suggests that the method has poor repeatability and would be unsuitable for Ocean Sonics calibrations.

It should be noted that the reference hydrophone unit was calibrated by NPL down to 2 kHz, with measurements at points spaced by 500 Hz between 2 and 10 kHz. Below that, the sensitivity of the reference unit was estimated by interpolating between the pistonphone measurement at 250 Hz and the NPL value at 2 kHz. We recognize that this introduces some uncertainty in the sensitivity estimates of the UUT.

3. RESULTS

Vibrating Column Test Fixture Results

The vibrating column test fixture had some unexpected challenges, the main one being that the frequency range was limited more than was hoped by what appeared to be mechanical modes in the container in the 60-200 Hz range. This limited the useful calibration range to 1-55 Hz. The 60 Hz noise may have also been from mains interference but the higher frequency interference seemed likely to be a result of mechanical modes, which was roughly validated by a low-fidelity FEA model. This is visible in Figure 3 where the frequency responses become erratic in the upper portion of the frequency band. Even in the flat section there is more variation than would be desired on a run-to-run basis. However it should be noted that the average of a number of runs follows the expected curve

to within 0.5 dB so while the precision was an issue, the actual calibrated values were correct.

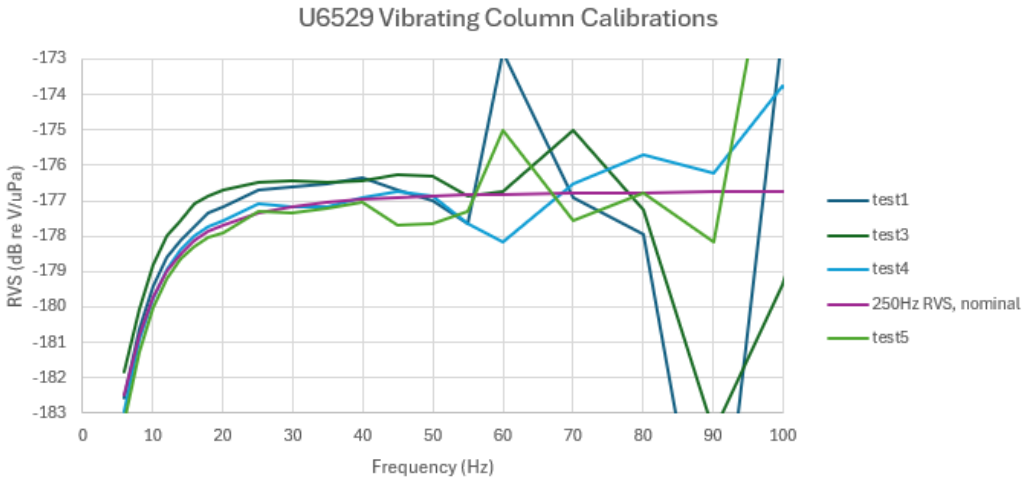


Figure 3: Calibration results in the vibrating column test fixture.

The investigation into the effect of the hydrophone positioning revealed inconsistency in the sound levels when the hydrophone position was modified laterally across with width of the container. This is shown in Figure 4 where the sound level as a function of hydrophone position is shown for a series of frequencies across two tests.

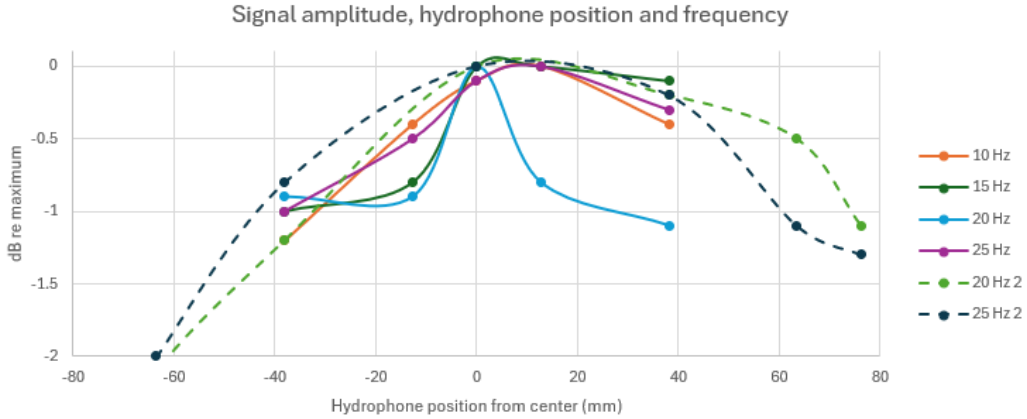


Figure 4: Measured sound level with lateral position inside vibrating column test fixture.

The results are consistent in that the highest sound levels were at the centre of the container but there was some further inconsistency between measurements that would not be acceptable in a calibration. The hydrophone sensor tips were placed at a spacing of 50mm so it appeared that amplitude errors of 0.5-1.0 dB could readily occur even if assuming good repeatability of the hydrophone positioning.

Overall, the vibrating column test fixture achieved some level of its objective of showing that very low frequency calibrations may be possible without significant changes to the existing calibration processes used with the in-tank calibrations. Though it was clear that further work would be needed on the specifics of the design, with a focus on the use of a more rigid container.

In-Tank Calibration Processing Modifications Results

Results for the in-tank calibration modification tests are shown in Figures 5 – 8.

Figure 5 shows an example of a 6 kHz tone with the data window used for processing highlighted by a red box. The original waveform (orange curve) is visibly very messy, as noted in the previous section, which is a known issue with the LF icTalk with short duration tones. The filtered waveform (blue curve) is very different from the original, however, agrees relatively well in the data window. The reference and UUT waveforms do agree (i.e. are scalar multiples of each other) up until the end of the data window, which is a critical assumption of the sensitivity calculation.

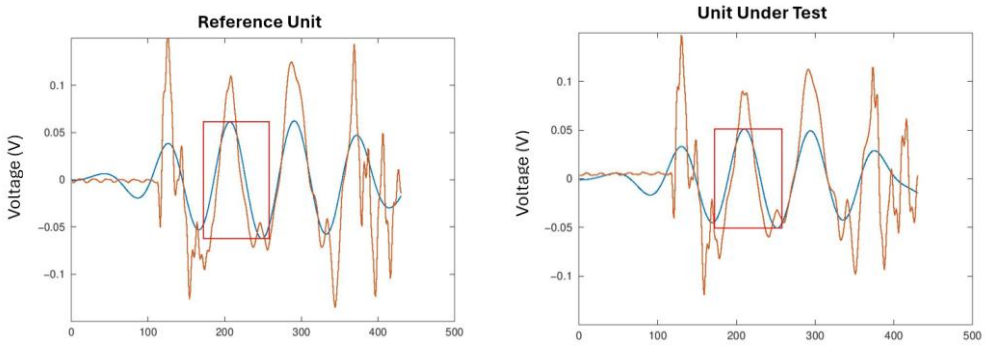


Figure 5. Example of data window (red) overlaid on filtered signal (blue) and original waveform (orange), Top: Reference unit, Bottom: Unit Under Test (UUT).

Reducing the data window length even down to an unreasonable 0.05 cycles did not have a large effect on the sensitivity calculation, as shown in Figure 6. The difference in sensitivity over all data window lengths tested was less than 0.2 dB re 1 V/uPa.

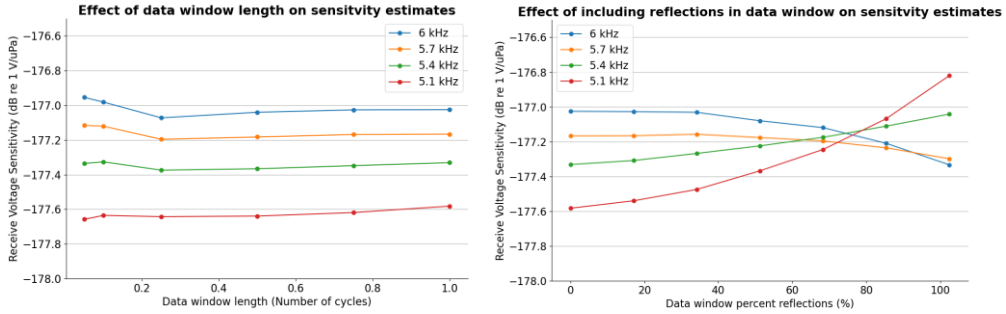


Figure 6. Left - Effect of using a data window of length less than one tone period on hydrophone sensitivity estimate. Right - Effect of including reflection-containing portion of the signal in the data window.

Allowing reflections in the data window had a greater effect on the sensitivity estimates than shortening the data window, which can be seen also in Figure 6. The changes in sensitivity were frequency dependent. However, what was consistent was that the sensitivity estimates diverged further from the original non-reflections estimate as more of the reflecting-containing portion of the signal was included. Interestingly, for all frequencies tested, the difference between the 0% reflections data window and the 100% reflections data window was less than 1 dB re 1 V/uPa. In particular, the difference between 0% and 40% reflections was under 0.2 dB re 1 V/uPa.

Figure 8 shows a subet of the low frequency calibrations performed using the in-tank calibration modifications. Below 3 kHz and above 500 Hz, the sensitivity estimates vary by up to +/- 1 dB re 1 V/uPa, however there were a variety of parameters being tested including tone intervals and processing parameters, so this variation does not represent the repeatability using a consistent testing protocol. Below 400 Hz, the results were erratic.

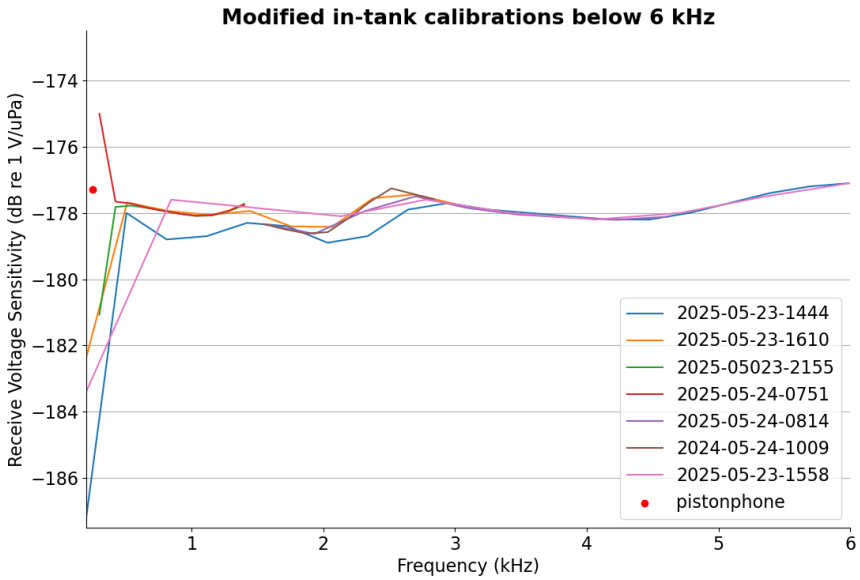


Figure 7. Multiple calibrations using modified in-tank method. The pistonphone measurement at 250 Hz is shown for reference.

Figure 8 provides a combined calibration curve for hydrophone S/N 2080 synthesized from a standard in-tank high frequency calibration, an in-tank low frequency calibration as described here, a GRAS 42AG calibrator measurement (250 Hz), and a vibrating column measurement. Overall, the agreement between the methods seems fairly good.

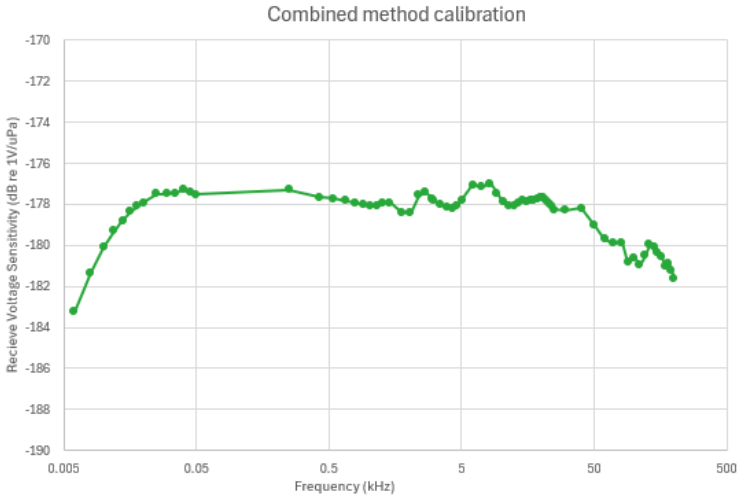


Figure 8: Combined calibration curve for hydrophone S/N 2080.

4. DISCUSSION

This paper outlines preliminary efforts to perform full-bandwidth calibrations choosing specifically techniques that can be performed using the icListen hydrophone and icTalk projector. The techniques used were selected to minimize the amount of effort to integrate multiple calibration methods into a single system. Two methods were used to extend the functional calibration minimum frequency from 4 kHz to 5 Hz. The synthesized calibration shown in Figure 8 suggests that the methods outlined here may be viable for this purpose however additional work will have to be done to further validate both methods and improve

upon the design of the vibrating column fixture used here. Ideally there would be overlap between each method, but it should be possible to extend the frequency range of the vibrating column to reach the 500 Hz or more. If further testing instead shows that this is not possible then there are other calibration methods that have been shown to work in the single Hz to 100's of Hz range that could be implemented. Regarding the sampling of data beyond the reflection-free period, there are trade-offs between the window length and calibration precision. Further work is needed to determine if it can be beneficial. Using a small percentages of a single wave period appears to have the potential to provide meaningful results when using averaged signals with a 90° offset. Even without extending into the reflection period, small cycle percentages could potentially reach 500 Hz.

REFERENCES

[1] **BSI 2007**. IEC 60556:2007 – Underwater acoustics – Hydrophones – Calibration in the frequency range 0,01 Hz to 1 MHz. International Electrotechnical Commission.

[2] **Metrology for low frequency sound and vibration.** https://www.ptb.de/empir2020/infra-auv/information-communication/publications/good-practice-guide/part-3-hydroacoustic-technology/developments-in-primary-calibration/ (Accessed: 25 May 2025)

[3] Drinnan, R. (2023). Low-Frequency Hydrophone Calibration Using Elastic Waveguides [Dalhousie University]. Dalhousie University Library: https://dalspace.library.dal.ca/items/6c0cc4be-7adf-4d11-9fa3-8b5eee638434

Structured Session
21. Underwater Communications and Networking

Organised by Charalampos Tsimenidis, Paul Mitchell and Konstantinos Pelekanakis

A Novel GRU-CNN Model for Self-Interference Cancellation in IBFD Underwater Communication Systems

Manar T. Alhamnday[1], Bilal A. Jebur[1], Charalampos Tsimenidis[2], and Shahid Mumtaz[2]

[1]Computer and Information Engineering Department, College of Electronics Engineering, Ninevah University, Mosul, Iraq
[2]Department of Engineering, School of Science and Technology, Nottingham Trent University, Nottingham, U.K.

Dr. Bilal A. Jebur, College of Electronics Engineering, Ninevah University, Mosul, Iraq, 41002, bilal.jebur@uoninevah.edu.iq

Abstract: *The in-band full-duplex (IBFD) scheme has attracted immense attention in the underwater communication society due to its ability to improve communication throughput. However, self-interference (SI) remains a major problem for the performance and efficiency of IBFD systems, as the transmitted signal from a device leaks into its own receiver, necessitating advanced cancellation techniques to enable effective communication. Recently, machine learning (ML), particularly deep learning (DL), has presented promising techniques for self-interference cancellation. In this work, we propose a deep learning-based hybrid model, termed GCCN, which synergistically integrates Gated Recurrent Units (GRU) with Convolutional Neural Networks (CNN) specially designed to enhance self-interference cancellation (SIC) in IBFD underwater communication systems. Our proposed model leverages the temporal processing skills of GRUs for sequential data, along with the spatial feature extraction strengths of CNNs for collecting complex patterns in data, to effectively mitigate SI. Experimental results demonstrate that the GCCN model outperforms traditional methods and standalone deep learning models in terms of interference cancellation performance and improves the bit error rate (BER), achieving significant improvements in signal quality and system throughput.*

Keywords: *in-band full-duplex, self-interference, GRU, CNN.*

1. INTRODUCTION

IBFD communication systems have received much attention recently due to the possibility of improving the spectral efficiency effectively through enabling sending and receiving simultaneously at the same frequency band [1][2]. However, a crucial barrier for the implementation of IBFD systems is self-interference. SI arises when the transmitted signal leaks into the receiver at the same location, and the leaked signal can swamp the receiver signal with noise, making it difficult to distinguish between the desired incoming signals and the SI signal generated by the transmission itself [3]. In order to eliminate or reduce SI, several interference cancellation techniques have been proposed [4]. Conventional SIC approaches achieved different degrees of success, but they frequently struggled with complicated and nonlinear interference. The advancements of recent years in the areas of ML and DL have generated better opportunities for addressing these difficulties.
In this paper, we propose a novel hybrid GRU-CNN model specifically designed for SI cancellation in IBFD underwater systems, combining the temporal processing capabilities of GRUs with the spatial feature extraction capabilities of CNNs. A proposed hybrid DL model shows promise to improve both the mean squared error (MSE) and efficiency of SI cancellation.

1.1. RELATED WORK

IBFD transmission is the main emphasis of modern wireless communication systems, as it improves spectrum efficiency. Different methods exist to tackle the problem of SI signals in IBFD systems. This segment surveys some of the existing literature on SIC methods while discussing traditional approaches and recent advances in ML.

Reference [5] demonstrates how SIC schemes enhance system performance in IBFD transmission systems. The two main categories of traditional SIC approaches are passive and active (analog and digital) SI cancellation. For instance, passive SIC employs a secondary transformer to reduce SI in the received signal, as reported in [6]. In contrast, active cancellation methods use the transmitter chain to produce duplicate signals that are then subtracted from the signals that are received. The authors in [7] introduce an adaptive SIC technique for underwater acoustic systems, utilizing the normalized least mean squares (NLMS) algorithm. In [8], the researchers suggested an SIC model based on separating the SI channel from the far channel by orthogonal pilots for improved performance in IBFD-UWA. A digitally assisted analog SIC (DAA-SIC) system for a complicated SI propagation channel was presented by researchers in [9], highlighting the significance of addressing hardware parameters to improve system performance. Conventional SIC technologies struggle in dynamic environments since they rely on static characteristics, which might result in lower communication efficiency [10].

Recent SIC methods have started to rely more heavily on ML techniques. Neural networks have been successfully applied to non-linear SIC in (FD) radios [11]. Advanced ML approaches including recurrent and complex value neural networks, has achieved remarkable success in eliminating SI signals [12]. Furthermore, researchers in [13] introduced a long short-term memory (LSTM) to cancel SI signals in co-frequency co-time FD systems.

Nowadays, studies combine several neural network architectures in order to take advantage of the special advantages of each to achieve the best performance. For example, researchers in [14] suggested a hybrid network that combines multi-head self-attention with a mix of 1D CNN and bidirectional LSTM to relieve inter-symbol interference in MIMO-FTN-OWC systems,

achieving a BER increase of up to 7 dB while lowering computational complexity by 31.15%. An integrated model that combines CNN, GRU, ResNet, and self-attention to reconstruct SI signals was proposed in [15]. The model increases its nonlinear SIC capacity by up to 28% relative to current neural network-based eliminators and polynomial eliminators.
Our proposed GCCN model capitalizes on these advancements by fusing temporal and spatial feature extraction to create solutions to the SI challenge in IBFD systems.

1.2. CONTRIBUTIONS

The paper's contributions are summarized as follows:

- Design a GCCN model that integrates the temporal abilities of GRUs with the feature learning capacity of CNNs to effectively SIC in IBFD underwater systems.
- Performance analysis of the GCCN model and evaluation of its effectiveness in removing self-interference.
- A comparison of the proposed GCCN model with standalone machine learning methods and traditional SIC techniques to prove the efficiency of the model.

1.3. ARRANGEMENT OF THE PAPER

The rest of the paper is arranged as follows: In Section 2, we present the system model, detailing the architecture of the IBFD system and the proposed model. Section 3 presents the simulation results of the proposed model. Finally, a conclusion and suggestions for potential avenues for future research are presented in Section 4.

2. SYSTEM MODEL

2.1. IBFD ARCHITECTURE

The IBFD system utilizes the orthogonal frequency division multiplexing (OFDM) technique. The block diagram architecture of an IBFD transceiver is illustrated in Fig. 1.

At the transmitter system, the information bits $d(m)$ are modulated to obtain X_t using quadrature amplitude modulation (QAM). After modulation, applying the Inverse Fast Fourier Transform (IFFT) to convert the signal to the time domain and adding a cyclic prefix (CP) to mitigate inter-symbol interference (ISI), can be expressed as

$$x_t(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_t(k) e^{j\frac{2\pi}{N}kn}, \tag{1}$$

and

$$x_{tcp}(n) = \begin{cases} x_t(N-L+n), & \text{for } n = 0, 1, \ldots, L-1 \\ x_t(n-L), & \text{for } n = L, L+1, \ldots, N+L-1 \end{cases} \tag{2}$$

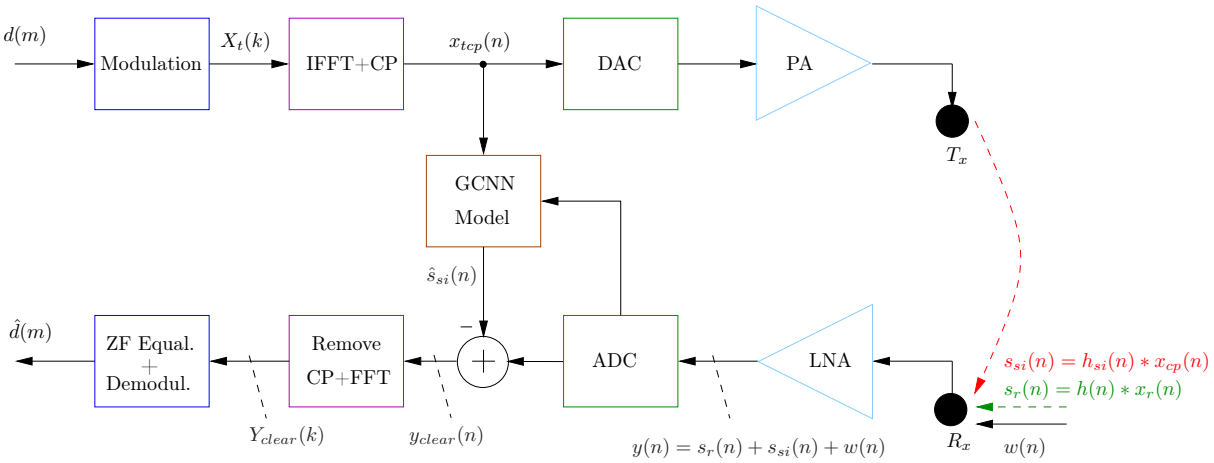


Figure 1: Architectural of IBFD system.

where L denotes the length of the CP. Next, the signal is converted to an analog format through a digital-to-analog converter (DAC) and amplified by a power amplifier (PA). Finally, the signal is transmitted by the transmitting antenna and transferred between nodes over a channel. The channel used is based on scientific measurements taken underwater, as described in [16].

Concurrently, at the receiver system, the system receives a signal that can be expressed as

$$y(n) = s_r(n) + s_{si}(n) + w(n), \tag{3}$$

where $w(n)$ denotes the additive white Gaussian noise (AWGN), and $s_r(n)$ denotes the signal received from the distant node given as

$$s_r(n) = \sum_{p=0}^{P-1} h(n-p)x_r(p), \tag{4}$$

and $s_{si}(n)$ denotes the SI signal given as

$$s_{si}(n) = \sum_{m=0}^{M-1} h_{si}(n-m)x_t(m). \tag{5}$$

After receiving the signal, it is amplified by a low-noise amplifier (LNA) to enhance the signal quality and is digitized using an analog-to-digital converter (ADC). Next, the predicted interference from the GCCN model $\hat{s}_{si}(n)$ is subtracted from the signal, which can be represented after subtracting as

$$y_{clear}(n) = s_r(n) + r_{si}(n) + w(n), \tag{6}$$

where r_{si} denotes the residual SI after cancellation, and can be written as

$$r_{si}(n) = s_{si}(n) - \hat{s}_{si}(n), \tag{7}$$

Subsequently, remove the CP from the clear signal and apply the Fast Fourier Transform (FFT) to convert the signal to the frequency domain, which can be expressed as

$$Y_{clear}(k) = S_r(k) + R_{si}(k) + W(k), \tag{8}$$

where $S_r(k)$ denotes the signal of interest, $R_{si}(k)$ denotes the residual SI, and $W(k)$ denotes the ambient noise in the frequency domain. Finally, the demodulation is applied to the signal using the same modulation scheme (QAM) to extract the received information.
The next subsection will cover a discussion on the GCCN model.

2.2. PROPOSED MODEL

In this part, we discuss the architecture of the suggested model, comprising the GRU and CNN architectures.

2.2.1. GRU

GRU networks are the most recent networks of RNN [17], which has a simpler architecture and lower computational demands than LSTM, where the input and the forget gates have been merged into one gate (update gate) [18]. Such simplification leads to a reduction in the parameters, meaning shorter training time durations are achieved while the temporal dependence is maintained in the long run.

The structure of the GRU is illustrated in Fig. 2. The GRU architecture has a two-gate [19]. The first gate is the reset gate $\mathbf{r}_t$, determining how much of the past it should forget, while the second gate is the update gate $\mathbf{z}_t$, determining how much of the previous information to retain and how much of the next information to incorporate. The GRU is computed by the following equations. Initially the reset gate is updated using

$$\mathbf{r}_t = \sigma(\mathbf{W}_r \cdot [\mathbf{h}_{t-1}, x_t] + \mathbf{b}_r). \tag{9}$$

Then, the update gate is computed as

$$\mathbf{z}_t = \sigma(\mathbf{W}_z \cdot [\mathbf{h}_{t-1}, x_t] + \mathbf{b}_z). \tag{10}$$

Finally, the hidden state candidate is selected as

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_{\tilde{h}} \cdot [\mathbf{r}_t \odot \mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_h), \tag{11}$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t, \tag{12}$$

where σ and tanh are the activation functions, $\odot$ is the Hadamard product, $\mathbf{h}_{t-1}$ is the prior hidden state, x_t is the input data at time t, $\mathbf{h}_t$ is the new hidden state, $\tilde{\mathbf{h}}_t$ is the candidate hidden state, $\mathbf{W}_{\tilde{h}}$ and $\mathbf{W}_z$ are the weight matrices, and $\mathbf{b}_z$ and $\mathbf{b}_h$i are the bias vectors. In this study, we employed a one-layer GRU network with 40 hidden neurons to learn temporal aspects of the data.

2.2.2. CNN

Convolutional neural networks enhance the ML capabilities of neural networks by incorporating convolutional layers into the architecture [21]. CNN is particularly suitable for processing network-like data, such as images [22] or sequential data [23], due to its ability to automatically

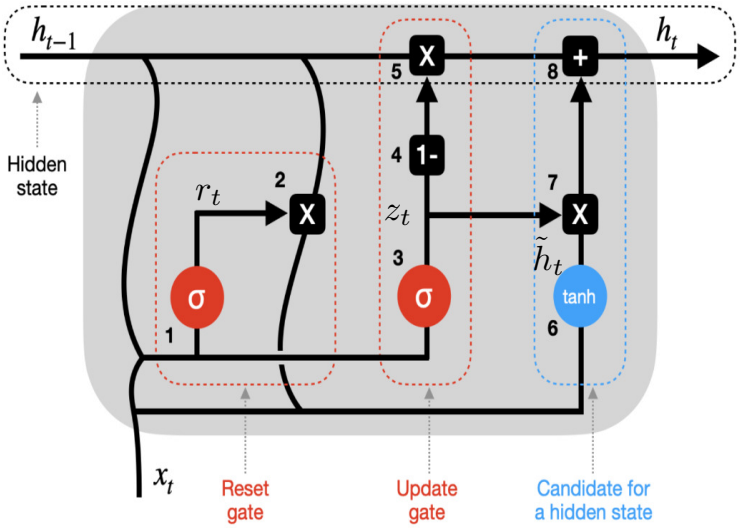


Figure 2: GRU structure [20].

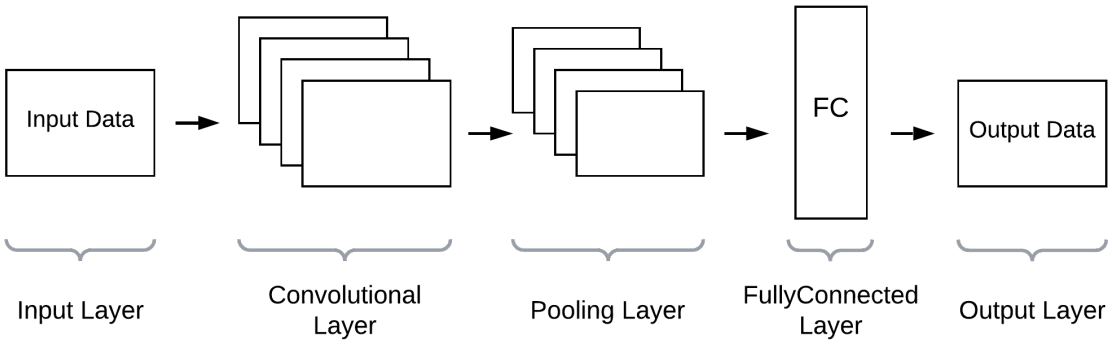


Figure 3: Simple CNN architecture.

learn the spatial hierarchy of features through convolutional layers. CNNs are composed of several layers, including input, convolution, pooling, and fully connected layers. Fig. 3 illustrates a minimal CNN architecture.

In this work, we employed 1D CNN, which contains a convolution layer, a batch normalization layer, and a ReLU activation, operates in conjunction with the GRU component. The CNN is used to extract spatial features from the GRU output.

2.2.3. GCCN MODEL

In this part, we present the architecture of the GCCN model, which integrates GRU and CNN algorithms.

The GCCN structure is initiated with an input layer that takes data. This data then passed

through a GRU layer, which deals with the sequential information. Following the GRU layer, a 1D CNN with a kernel size of (3, 128) is employed to extract spatial features from the sequential data. The result from the CNN is passed through fully connected layers, which are responsible for the predictions. Lastly, the results are obtained from the output layer. Fig. 4 shows the structure of the GCCN network.

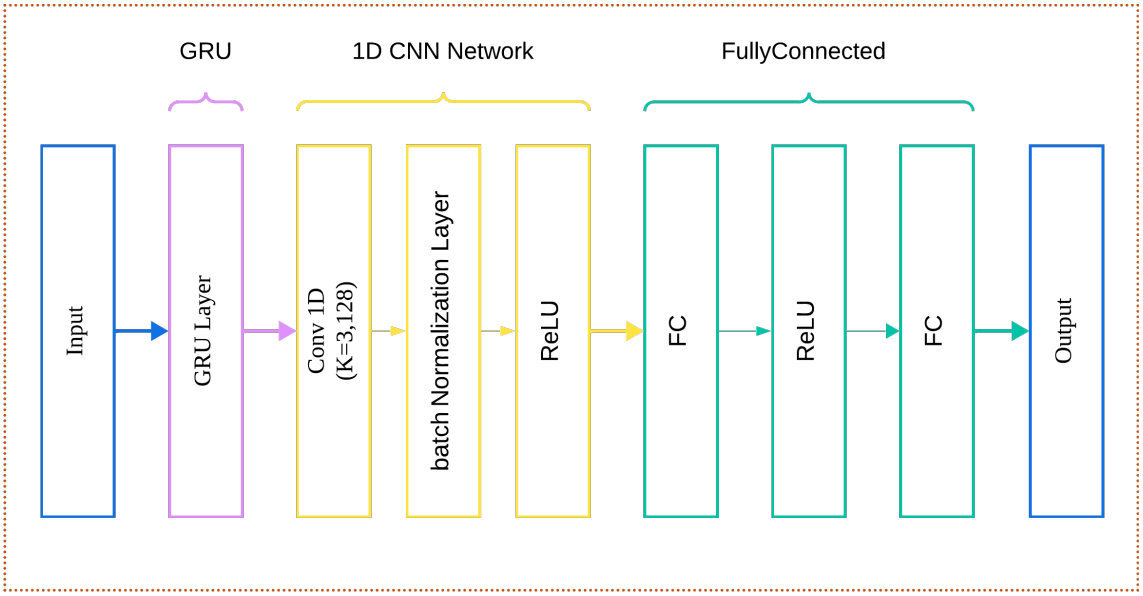


Figure 4: GCCN structure.

The GCCN model leverages the temporal processing capabilities of GRUs for sequential data, alongside the spatial feature extraction strengths of CNNs for collecting complex patterns in data to achieve reduced SI.

3. RESULTS

This section summarizes the results of MATLAB simulations of the suggested SI cancellation model. Table 1 displays the simulation parameters used in this work. The performance of the proposed model is evaluated using MSE, which measures the average squared difference between the estimated and actual signals, and cancellation ratio in decibels (C_{dB}), which measures the reduction in SI achieved by the model, calculated as

$$MSE = \frac{1}{N}\sum_{i=1}^{N} |r_{si}(i)|^2, \tag{13}$$

$$C_{dB} = 10\log_{10}\left(\frac{P_{SI}}{P_{RSI}}\right), \tag{14}$$

where N is the number of samples, P_{SI} is the power of the SI signal, and P_{RSI} is the power of the residual SI signal. The outcomes for the two metrics are summarized in Table 2.

Fig. 5 illustrates the performance of BER under various conditions, including theoretical performance, performance with interference, and performance after applying the SI cancellation using our model.

Parameter	Value
Optimizer	SGDM
Training SNR	30 dB
Loss Function	MSE
Input Features	2(1 real,1 imaginary)
GRU Hidden Size	40
Number of Epochs	500
Batch Size	200
InitialLearnRate	0.1
LearnRateDropFactor	0.1
LearnRateDropPeriod	150
Modulation Type	16-QAM
CP	$16, 32, 64, 128$

Table 1: Simulation parameters.

Metrics	MSE	C_{dB}
SNR=0	3.302e-04	44.865
SNR=5	1.801e-04	47.373
SNR=10	8.07e-05	50.94
SNR=15	4.672e-05	53.344
SNR=20	2.371e-05	56.245
SNR=25	1.083e-05	59.688
SNR=30	6.149e-06	62.135

Table 2: Performance of GCCN model.

Performance without interference is represented by the BER curve theory of the system. This curve is used as the base to compare with that of the performance of the proposed model. The orange curve illustrates poor BER performance when SI is present in the system. Furthermore, Fig. 5 displays the BER curves after cancelling SI using the GCCN model at various CP. The BER decreases significantly, putting it close to the theoretical performance. Fig. 6 illustrates the spectrum of the SI signal, alongside the residual signal spectrum after cancellations achieved with traditional methods (LMS and NLMS), as well as with individual ML methods (CNN and GRU). In addition, it presents the residual signal spectrum after cancellations using the proposed method. The ability of the GCCN model to model complex interference patterns and adapt to communication channels makes it the most efficient solution for SI cancellation.

4. CONCLUSION

In this paper, we proposed a GCCN model to eliminate SI caused by signal leakage to the receiver in IBFD underwater systems. Our approach leverages the strengths of GRUs and CNNs to further enhance performance in the SIC effectively. The GRU maintains temporal connectivity to analyze interference signals while the CNN performs spatial feature extraction from the signal. The results demonstrate the GCCN model outperforms individual machine learning approaches (GRU and CNN), and significantly outperforms traditional SIC methods (LMS and

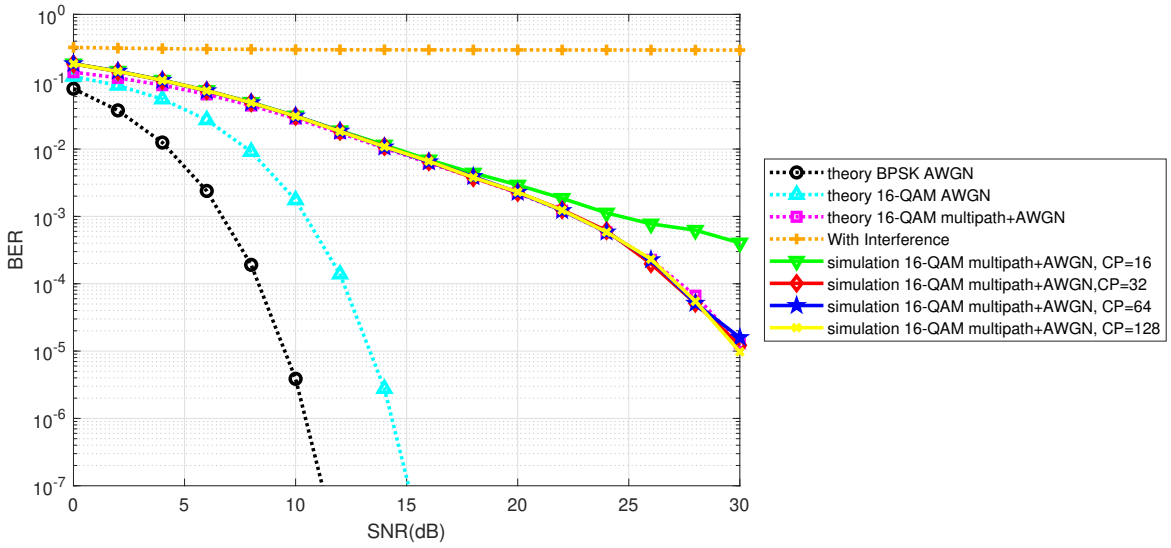


Figure 5: BER performance comparison of different conditions: theoretical performance, performance with interference, and performance after applying the proposed GCCN model for SI cancellation.

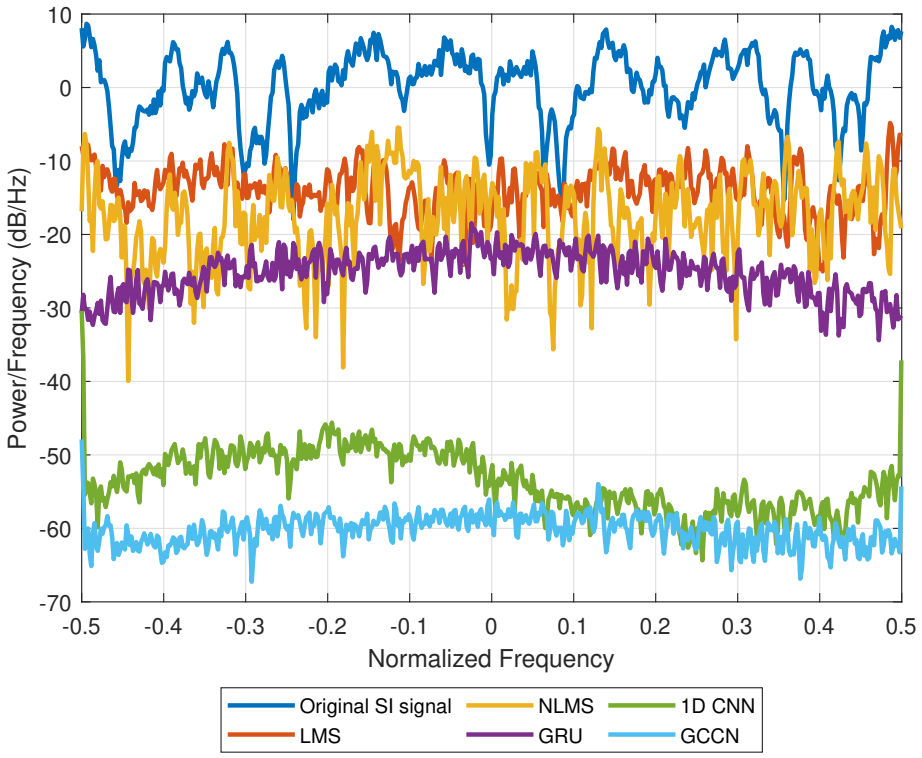


Figure 6: Signal spectrum after cancellation.

NLMS). Future work will focus on further optimizing the model architecture.

REFERENCES

[1] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, pp. 375–386, 2013.

[2] K. E. Kolodziej, B. T. Perry, and J. S. Herd, "In-band full-duplex technology: Techniques and systems survey," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 7, pp. 3025–3041, 2019.

[3] G. Liu, F. R. Yu, H. Ji, V. C. Leung, and X. Li, "In-band full-duplex relaying: A survey, research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 500–524, 2015.

[4] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE Journal on selected areas in communications*, vol. 32, no. 9, pp. 1637–1652, 2014.

[5] D. Kim, H. Lee, and D. Hong, "A survey of in-band full-duplex transmission: From the perspective of phy and mac layers," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2017–2046, 2015.

[6] Y. Wang, Y. Li, L. Shen, and Y. Zakharov, "Acoustic-domain self-interference cancellation for full-duplex underwater acoustic communication systems," in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1112–1116, IEEE, 2019.

[7] B. A. Jebur, C. T. Healy, C. C. Tsimenidis, J. Neasham, and J. Chambers, "In-band full-duplex interference for underwater acoustic communication systems," in *OCEANS 2019-Marseille*, pp. 1–6, IEEE, 2019.

[8] H. A. Naman and A. E. Abdelkareem, "Self-interference cancellation in underwater acoustic communications systems using orthogonal pilots in ibfd," *Acta Polytechnica*, vol. 63, no. 1, pp. 23–35, 2023.

[9] Y. Zhao, G. Qiao, Y. Lou, X. Liu, N. Zheng, and Y. Zhang, "Digitally assisted analog self-interference cancellation for in-band full-duplex underwater acoustic communication," in *2021 OES China Ocean Acoustics (COA)*, pp. 613–618, IEEE, 2021.

[10] S. Rahman Doha and A. Abdelhadi, "Deep learning in wireless communication receiver: A survey," *arXiv e-prints*, pp. arXiv–2501, 2025.

[11] A. Balatsoukas-Stimming, "Non-linear digital self-interference cancellation for in-band full-duplex radios using neural networks," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, IEEE, 2018.

[12] A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Advanced machine learning techniques for self-interference cancellation in full-duplex radios," in *2019 53rd Asilomar conference on signals, systems, and computers*, pp. 1149–1153, IEEE, 2019.

[13] Y. Wang, Y. Chen, C. Hu, Y. Li, S. Wang, J. Yu, F. Lu, Z. Fan, H. Du, and C. Ma, "Digital si cancellation based on lstm for co-frequency co-time full duplex communication systems," in *Journal of Physics: Conference Series*, vol. 2405, p. 012020, IOP Publishing, 2022.

[14] M. Cao, Q. Yang, G. Zhou, Y. Zhang, X. Zhang, and H. Wang, "A hybrid network integrating mhsa and 1d cnn–bi-lstm for interference mitigation in faster-than-nyquist mimo optical wireless communications," in *Photonics*, vol. 11, p. 982, MDPI, 2024.

[15] J. Liu and T. Ding, "Digital self-interference cancellation for full-duplex systems based on cnn and gru," *Electronics*, vol. 13, no. 15, p. 3041, 2024.

[16] C. T. Healy, B. A. Jebur, C. C. Tsimenidis, J. Neasham, and J. Chambers, "Experimental measurements and analysis of in-band full-duplex interference for underwater acoustic communication systems," in *OCEANS 2019 - Marseille*, pp. 1–5, 2019.

[17] A. K. Gizzini and M. Chafii, "Rnn based channel estimation in doubly selective environments," *IEEE Transactions on Machine Learning in Communications and Networking*, 2023.

[18] R. Cahuantzi, X. Chen, and S. Güttel, "A comparison of lstm and gru networks for learning symbolic sequences," in *Science and Information Conference*, pp. 771–785, Springer, 2023.

[19] C. Zeng, C. Ma, K. Wang, and Z. Cui, "Parking occupancy prediction method based on multi factors and stacked gru-lstm," *Ieee Access*, vol. 10, pp. 47361–47370, 2022.

[20] V. Patel, "Gru recurrent neural networks – a smart way to predict sequences in python," 2019. Accessed on 2025-06-06.

[21] Y. Zhao, D. Wang, L. Wang, and P. Liu, "A faster algorithm for reducing the computational complexity of convolutional neural networks," *Algorithms*, vol. 11, no. 10, p. 159, 2018.

[22] S. A. Rachmadinasya, F. H. Melvandino, H. Ramza, *et al.*, "Klasifikasi aktivitas olahraga berdasarkan citra foto dengan menggunakan metode convolutional neural network," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3s1, 2023.

[23] T. O. Omotehinwa, M. O. Lawrence, D. O. Oyewola, and E. G. Dada, "Bayesian optimization of one-dimensional convolutional neural networks (1d cnn) for early diagnosis of autistic spectrum disorder," *Journal of Computational Mathematics and Data Science*, vol. 13, p. 100105, 2024.

Robust channel comparison metrics for underwater physical layer authentication

Davide Eccher[1] and Paolo Casari[1]

[1]*DISI, University of Trento* and *CNIT*, Italy

Davide Eccher, DISI, Via Sommarive 9, 38123 Povo TN, Italy - davide.eccher-1@unitn.it.

Abstract: *Physical layer security enables authentication and privacy in communication systems by harvesting the randomness from wireless channel realizations. For this, it leverages unique channel features that legitimate parties (Alice and Bob) can reliably observe while the same features remain mostly secret to attackers (Eve). In underwater acoustic networks, the minimal correlation of acoustic channels in space provides a promising context for PLS. Our research focuses on implementing authentication via PLS using channel crafting. More specifically, we craft an artificial channel starting from a secret seed, and then precode the transmitted signal by convolving it with the crafted channel. The parameters of the channel impulse response (CIR), namely the delay and amplitude of each arrival, serve as the source of the seed for the crafting process. The receiver performs authentication by comparing the CIR extracted from the incoming signal with an expected CIR. Therefore, we need a mismatch metric that measures tDhe matching of two channels, indicating that such channels have been generated by the two ends of the same link. Based on this metric, a decision algorithm can perform authentication. In this paper, we present the implementation of four mismatch metrics for CIRs, which account for differences that originate from measurement errors, node movement, noise, etc. These disturbances make the delay and amplitude values of the arrivals fluctuate, even to the point of disappearing. Such differences may lead to increased false alarms and missed detections, and therefore a convenient CIR mismatch metric should be robust against them. Specifically, we propose to reward strong channel similarities more than we penalize large differences.*

Keywords: *Underwater acoustic networks; physical layer security; authentication; simulation.*

1. INTRODUCTION

Security is a matter of increasing concern in the underwater communication field, especially with the growing adoption of underwater acoustic networks in such critical applications as surveillance, drone operations, or defense. Implementing security features in an underwater network is particularly challenging because of the high variability, low throughput, and long propagation delays of the acoustic channel. This makes the overhead introduced by typical key-exchange algorithms inconvenient in underwater networks. Lighter approaches have been explored in the literature, one of which is to exploit the information conveyed by the communication channel to implement security features. Some of these techniques involve generating secret keys, for example, by using the route propagation times [1], or by exploiting the channel characteristics [2] - [4]. These can be also used to perform target detection as in [5], and other approaches exploit position information for authentication, such as in [6] or in [7].

We focus on physical layer authentication in underwater acoustic networks, where the features of the estimated multipath components of the acoustic channel provide a secret shared between two legitimate parties. Acoustic multipath propagation has two desirable properties in this respect: reciprocity and very limited spatial correlation. For instance, Fig. 1 shows that multipath patterns as observed by two legitimate parties Alice and Bob the two ends of a simulated channel. The patterns match very well; in contrast, the the arrivals of the channel extracted by a third node (Eve) at a different location have considerably different amplitude and delay.

By leveraging matching multipath patterns, we generate a new artificial channel that is convolved with the signal sent to Alice. Alice then compares the received channel with the one she had crafted to perform the authentication decision. The channel crafted by Bob will be much more similar to the reference channel generated by Alice than any intruding transmission crafted by Eve, who will inevitably make mistakes when estimating multipath propagation between Alice and Bob. However, accepting Bob's transmission and rejecting Eve's requires a metric that measures channel mismatch every time a node receives a transmission.

A *maximum time-reversal resonating strength* metric is proposed in [8] for authentication. This score is computed for a received signal based on a dataset of previous communications, and was not originally developed to reflect quasi-similarities between bidirectional channels. In this paper, we propose four different channel mismatch metrics, and comment on their performance in a simulated scenario.

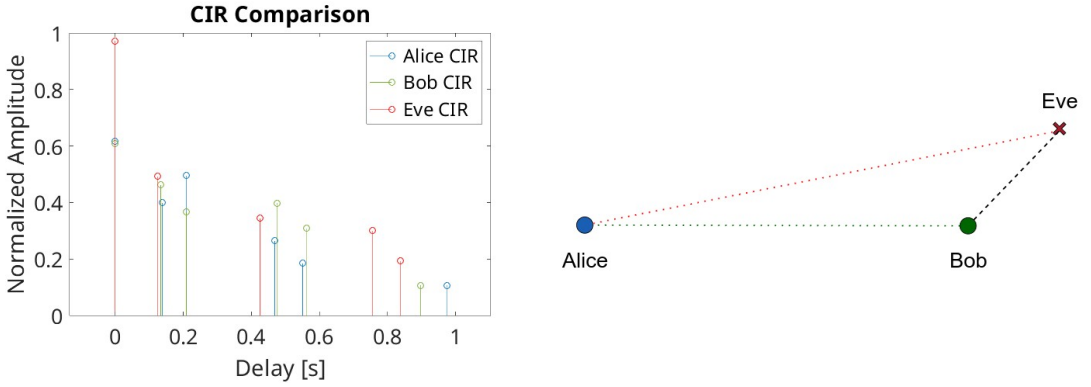


Figure 1: (Left) Comparison between CIRs of the AlicetoBob and BobtoAlice channels, against that of an attacker (Eve) arranged as in the topology on the right.

2. MISMATCH METRICS

The goal of a mismatch metric is to compare a received CIR against a reference one, and return a value that increases for increasing differences between the CIRs. Such differences may be due to the variability of underwater acoustic channels and can be observed from the different time of arrival and amplitude of their multipath components. Significant differences between CIRs may imply that an intruder is trying to impersonate a legitimate node. Hence, a mismatch metric should be able to detect these differences, while at the same time being robust against changes in the channels observed by Alice or Bob, which will also be subject to smaller but non-negligible fluctuations. Such fluctuation may include multipath components that even disappear temporarily due to, e.g., measurement errors, movement, noise, environmental changes, etc. An ideal metric should not penalize these changes too much over legitimate links in order to avoid false positives.

The metrics take as input both a reference CIR and a CIR to be compared against it. The CIRs are described in discrete terms, as arrays of delay-amplitude pairs, $\{(d_\mathrm{r}[i], a_\mathrm{r}[i])\}$, $i = 1, \ldots, N_\mathrm{r}$, and $\{(d_\mathrm{c}[j], a_\mathrm{c}[j])\}$, $j = 1, \ldots, N_\mathrm{c}$, where $d_\mathrm{r}[i]$ and $a_\mathrm{r}[i]$ represent the arrival delay and the complex amplitude of the ith multipath component of the reference CIR, $d_\mathrm{c}[j]$ and $a_\mathrm{c}[j]$ the same for the jth component of the CIR to be compared, whereas N_r and N_c are the number of components in the reference and compared CIR, respectively.

Before computing the mismatch metrics, we align the CIRs so that, for each channel, the first multipath arrival has delay $d[1] = 0$. Moreover, we join arrivals that are very close to each other. This kind of phenomenon is usually due to multiple reflections from extended surfaces or close-by scatterers. Small changes in the transmitting position cause small changes in the incidence angle that could create differences in the number of reflections and in the magnitude of constructive or destructive interference. Therefore, to mitigate this unpredictable behavior, we average the amplitudes of all components whose delays cluster within a range of 5 ms. After computing the mismatch metrics as described in the next subsections, we finally divide the obtained values by N_r. The above processing enables more comparable metric values across different scenarios or different time epochs. We now formalize the four metrics considered in this paper.

M_1: Full CIR comparison Here, we center a Gaussian kernel $g(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$ on each multipath component of the reference CIR, and use it to compute a mismatch between all pairs of multipath components on both CIRs, weighted with the reference amplitude. We exclude the first arrivals that, due to the initial alignment, match very well and may bias the metric. Formally, we compute

$$\widetilde{M_1} = \left(\sum_{i=2}^{N_\mathrm{r}} \sum_{j=2}^{N_\mathrm{c}} g\big(d_\mathrm{c}[j]; d_\mathrm{r}[i], \sigma_\mathrm{d}\big) \cdot g\big(a_\mathrm{c}[j]; a_\mathrm{r}[i], \sigma_\mathrm{a}\big) \cdot a_\mathrm{r}[i] \right)^{-1}, \tag{1}$$

where σ_a is the standard deviation of the amplitudes of the reference CIR's multipath components, and σ_d is computed such that a fraction Q of the area under the Gaussian kernel curve occurs within a delay interval proportional to the amplitude-weighed root mean-square delay difference

$$\hat{\tau} = \sqrt{\frac{\sum_{i=1}^{N_\mathrm{r}} (d_\mathrm{r}[i] - \bar{d}_\mathrm{r})^2 a_\mathrm{r}[i]}{\sum_{i=1}^{N_\mathrm{r}} a_\mathrm{r}[i]}}\,. \tag{2}$$

We divide this value by an additional parameter F to fine-tune the importance of delay differences over amplitude differences in the mismatch metric. Formally,

$$\sigma_{\mathrm{d}} = \frac{F\ \hat{\tau}}{\sqrt{2}\,\mathrm{erf}^{-1}(Q)}\ , \tag{3}$$

where $\mathrm{erf}(x) = \int_0^x \frac{2}{\sqrt{\pi}} \mathrm{e}^{-t^2}\mathrm{d}t$ is the Gaussian error function and $\bar{d}_{\mathrm{r}} = \frac{1}{N_{\mathrm{r}}}\sum_{i=1}^{N_{\mathrm{r}}} d_{\mathrm{r}}[i]$. This metric rewards multipath components with similar arrival delay more than components that are apart in time. Note that in (1) we take the reciprocal of the sum of the Gaussian kernel products, so that the mismatch decreases for CIRs with closely aligned arrivals.

For practical purposes, we impose a maximum mismatch value M_{max} to avoid that the metric approaches infinity for starkly different channels. Finally, we compute

$$M_1 = \widetilde{M_1}\ \cdot \max\left(1, N_{\mathrm{c}}/N_{\mathrm{r}}\right) \tag{4}$$

to balance channels with many more multipath components than the reference, which would otherwise be more likely to match reference multipath components and yield a low metric even if the channels are significantly different.

M_2: Optimal match This metric computes an optimal matching between each of the N_{r} components of the reference CIR and each of the components in the compared CIR, by minimizing the total delay difference between all matched components. The final value of the metric then depends on the delay and amplitude differences between matched pairs. Call $1 \leq C(i) \leq N_{\mathrm{c}}$ the index of the component of the compared CIR coupled to the ith component of the reference CIR, where the matching C can be any of the $N_{\mathrm{c}}!$ permutations of the indices from 1 to N_{c}. If one such permutation is $X(i)$, then

$$C = \arg\min_{X} \sum_{i=1}^{\min(N_{\mathrm{c}},N_{\mathrm{r}})} \left|d_{\mathrm{r}}[i] - d_{\mathrm{c}}[X(i)]\right| . \tag{5}$$

The optimal matching is computed via the Hungarian algorithm, which has complexity $O(N_{\mathrm{r}}^3)$. If $N_{\mathrm{c}} \neq N_{\mathrm{r}}$, only $\min(N_{\mathrm{c}}, N_{\mathrm{r}})$ matches are considered. However, if $N_{\mathrm{r}} > N_{\mathrm{c}}$, the remaining arrivals are matched with an artificial component having zero delay and amplitude: this prevents channels that are shorter than the reference from seeming accurate as a consequence of having a few matching components. After obtaining the optimal matching, we compute the metric as

$$M_2 = \sum_{i=1}^{N_{\mathrm{r}}} \left(g(0;0,\sigma_{\mathrm{d}}) - g(d_{\mathrm{c}}[C(i)];d_{\mathrm{r}}[i],\sigma_{\mathrm{d}})\right) \cdot \left(g(0;0,\sigma_{\mathrm{a}}) - g(a_{\mathrm{c}}[C(i)];a_{\mathrm{r}}[i],\sigma_{\mathrm{a}})\right) \cdot a_{\mathrm{r}}[i]\ . \tag{6}$$

Note that we use $g(0;0,\sigma) - g(x;\mu,\sigma)$, so that the higher the difference between the delay and amplitude of matched taps, the higher the metric value.

M_3: Interpolation difference This metric, denoted as M_3, oversamples the CIRs and then extends the duration of each arrival by an amount of time proportional to the normalized amplitude of the component (or until another arrival is reached). Such proportional amount is defined as the amplitude-weighed root mean-square delay difference $\hat{\tau}$ as defined in Eq. (2). Finally, the metric is computed by subtracting the two resulting CIRs and computing the average magnitude of the resulting difference. This approach is considered as a computationally inexpensive baseline.

M_4: Cumulative CIR difference Similar to the previous metric, this metric, denoted M_4, oversamples the CIRs and zero-pads the resulting CIR between subsequent multipath components. Then, it subtracts the interpolated CIRs and integrates the resulting differences. In this case, small errors in the delay values contribute slightly to the mean, whereas discrepancies in the amplitude values propagate to the cumulative sum and have a higher weight in the final mismatch value.

3. PERFORMANCE EVALUATION

We now set up a testing environment to evaluate the capability of the metrics to distinguish matching channels in an authentication context.

An attacker (Eve) tries to impersonate a legitimate node (Bob) at the receiver Alice, by sending a signal similar to the one that Bob would generate. We assume that Eve knows the propagation environment in detail, including the bathymetry, the sound speed profile, and Alice's location. Eve can also localize Bob with some error, and can simulate the acoustic channel between the estimated location of Bob and that of Alice. If the estimate of Bob's location is accurate, Eve will then obtain a channel similar to Bob's, which will likely yield a low mismatch metric. Conversely, we expect that in any sufficiently diverse environment, the discrepancies between Eve's estimated channel and Bob's real channel should increase as Eve's estimate of Bob's position is increasingly erroneous. Therefore, we evaluate the correlation between the computed mismatch metrics and the Eve's localization error for Bob's location. The objective of our evaluation is to identify a mismatch metric that yields low values on channels that should be quasi-reciprocal, while returning a higher value if channels are different, as a result of Eve localizing Bob at the wrong location. The latter aspect is crucial for the development of physical layer authentication schemes based on channel crafting (see also Section 1).

Simulation Environment To simulate the channels between two given locations, we use the Bellhop ray tracing software, with bathymetry and sound speed profile taken from the San Diego Bay Area. We analyze 10 scenarios with two different bathymetry shapes. For each scenario we place Alice and Bob at a random position, 500 to 1200 m apart from each other. We then generate about 1000 estimates of Bob's location in a circle of radius 700 meters around Bob. The depth of Alice, Bob, and all locations estimated by Eve have a random depth between 30 and 100 meters, as shown in Fig. 2. The parameters used by the mismatch metrics algorithms are set as $Q = 0.99$ and $F = 10$, see Eq. (3).

4. RESULTS

We proceed by presenting the results of our simulations. In Fig. 3, we show how the four proposed mismatch metrics change as a function of Eve's error on Bob's location estimate. The shaded area around the curves conveys the standard deviation of the metric as a function of distance. All metrics are normalized to the mismatch value for the true channel of Bob.

We observe that for distances close to 0, mismatch values approach 1, because Eve's estimated channel and Bob's actual channel almost coincide. As expected, the CIR measured in the Bob→Alice direction is the channel having the lowest possible mismatch with the Alice→Bob channel, compared to other CIRs measured at different locations around Bob's true one. The cases where the normalized mismatch is smaller than 1 mean that Eve generated a channel closer to the one generated by Alice than Bob's own channel. This is a rare occurrence, and is limited

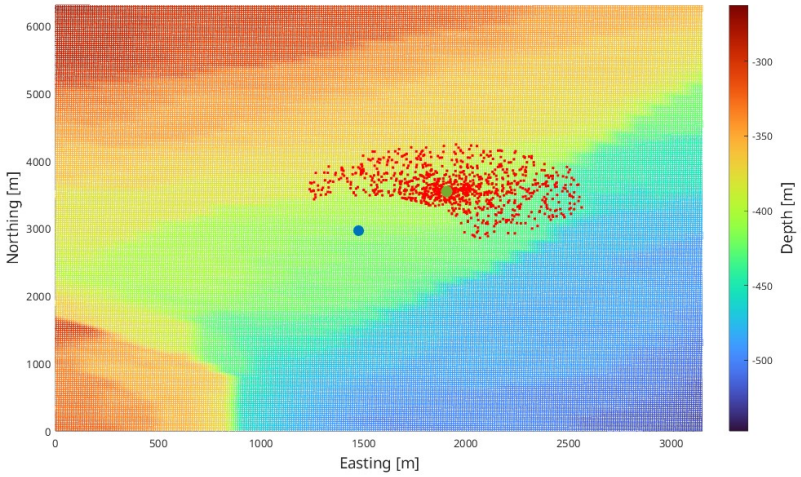


Figure 2: Simulation scenario and bathymetry. Alice: blue dot; Bob: green dot; Eve's estimates of Bob's location: red dots.

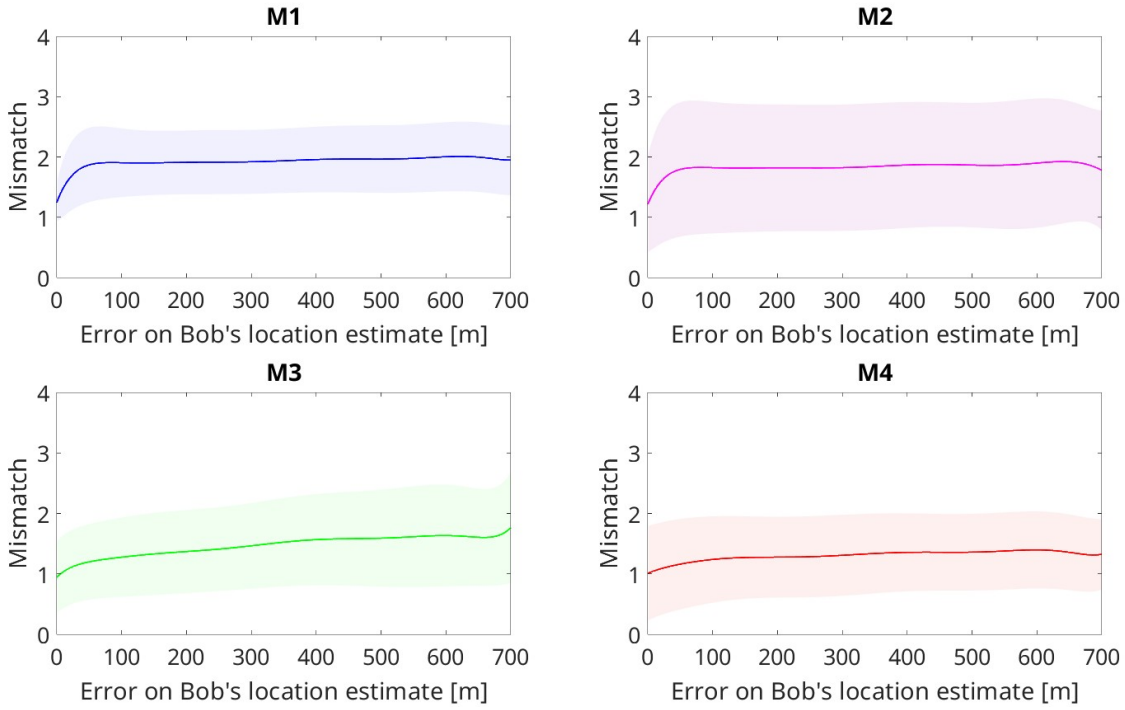


Figure 3: Channel mismatch as a function of distance between Bob and Eve's estimated position of Bob, normalized to the mismatch between Alice's and Bob's channels.

to those cases where Eve localizes Bob almost perfectly. From Fig. 3, we can also observe that spatial uncorrelation leads the mismatch metric to increase with increasing distance, until it tends to stabilize over an approximately constant value. This is because, beyond a certain distance, channels become uncorrelated. In our scenario, the uncorrelation distance is about 100 m from Bob.

Fig. 3 confirms that M_1 and M_2 yield similar results, and the best performance among the tested metrics. In particular, M_1 yields a smaller standard deviation in the results, as seen from the shaded areas around the solid curves. The baseline metrics M_3 and M_4 perform compara-

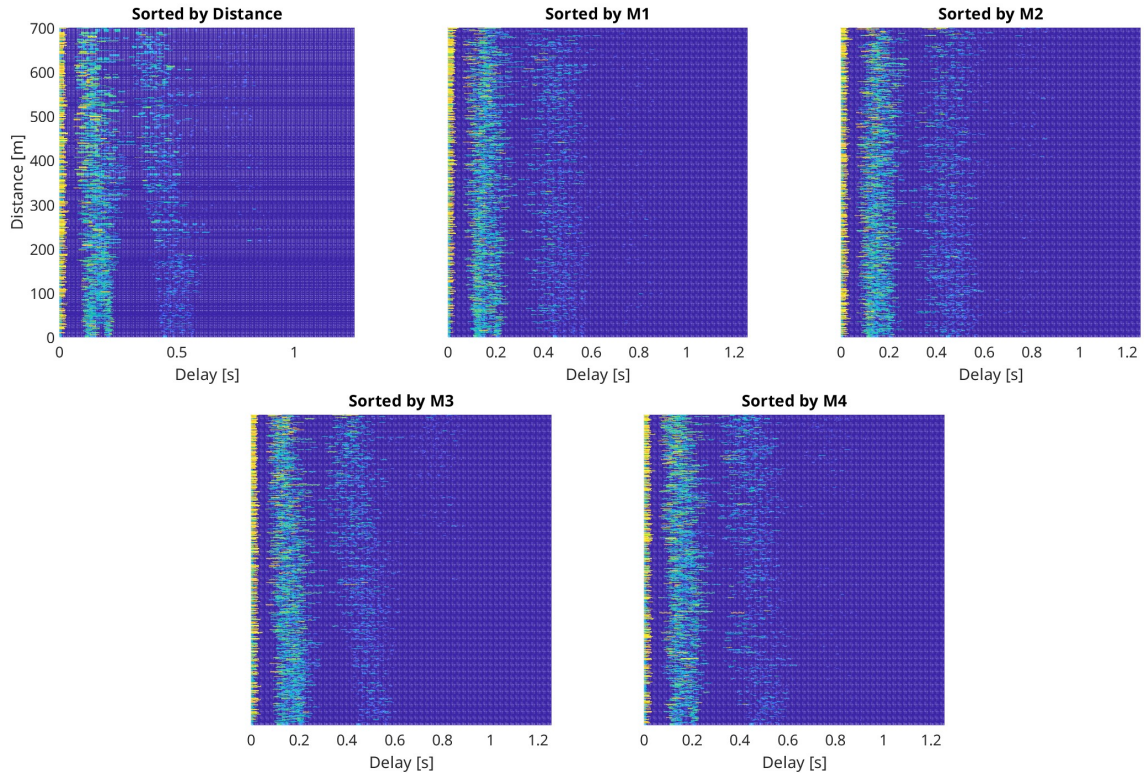


Figure 4: Channel impulse responses sorted by the actual distance between Eve's estimated position of Bob and Bob's true location (top-left), and by the four mismatch metrics (lowest mismatch at the bottom of each panel, highest mismatch at the top).

tively worse: M_4 barely differentiates between channels generated from locations close to Bob and far from it, whereas M_3 performs better only when Eve's error on Bob's position is large, hence channels are already starkly different.

Fig. 4 illustrates Eve's and Bob's CIRs, where the arrival delay of each multipath component is shown on the x-axis and amplitudes are color-coded. The channels are sorted along the y-axis in order of increasing localization error for Bob (first panel) and in order of increasing mismatch with respect to Alice's channel in the remaining panels.

The first graph shows that our assumption that closer nodes result in better-matching channels is confirmed by the simulations. In fact, we can observe that the first channels in the sorting share a significant number of components with similar delays and amplitudes. We also observe that, beyond a certain localization error, Eve's channels become almost completely uncorrelated from Bob's and Alice's, as confirmed by the metric values in Fig. 3.

Fig. 4 also provides a qualitative understanding of the behavior and implicit priorities of the different mismatch metrics. The best-performing metrics give greater weight to channels with similar arrival delays, and specifically to the first arrivals, which usually have higher amplitudes. In the case of metric M_2, the higher standard deviation of its output results in a more irregular channel sorting, where multipath arrivals exhibit a comparatively more scattered pattern. As M_3 gives the same weight to all arrivals, the channels with the lowest values of the metric are those where arrivals are positioned along the same delay span of the reference channel. However, as shown previously in Fig. 3, this weighing strategy does not necessarily iDmply a better capability to differentiate legitimate transmissions from impersonation attacks.

5. CONCLUSIONS

We presented four different metrics to measure the superimposability of underwater acoustic channel impulse responses. These metrics convey the mismatch between the channels observed at the two ends of the same link. Hence, a high value of the metric can serve as an indication of an impersonation attack, whereby an attacker is trying to "guess" a channel impulse response by localizing one of the two communicating parties, and by reproducing the channel that should be observed at the estimated location. The ultimate purpose of our work is to provide metrics that are robust when employed to make authentication decisions.

We have showed that the two proposed metrics M_1 and M_2 achieve better results than the baseline metrics M_3 and M_4, because they correlate better with the distance between Bob and Eve's estimate of Bob's location, and reach a plateau only when such localization error is sufficiently large to lead to uncorrelated channel impulse responses. Our simulations using the well-known ray tracing software Bellhop show that arrival delays form a more reliable basis than arrival amplitudes when matching two channels, hence a mismatch metric should prioritize delay information. Moreover, higher-amplitude arrivals are more reliably measured at both ends of a link and should have a bigger weight in the matching metric, whereas less powerful arrivals are not a good indication of reciprocity.

Acknowledgments The authors warmly thank EvoLogics for the support and discussion on this research topic. This work was partially supported by the NATO Science for Peace and Security Programme under grant G7646 (SHIELD), and by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU – PNRR, Mission 4 Component 2, Investment 1.3 – PE RESTART Spoke 6 – Project EMBRACE (PE00000001, CUP E63C220020700069).

REFERENCES

[1] Diamant, Roee, et al. "Secret key generation from route propagation delays for underwater acoustic networks." *IEEE Trans. Inf. Forensics Security* **18**: 3318-3333 **2023**.

[2] Pan, Pan, et al. "A Secret key generation scheme exploiting spatio-temporal acoustic channel Characteristics for Underwater Sensor Networks." *IEEE Sensors J.* **2024**.

[3] Pelekanakis, Konstantinos, et al. "Physical layer security against an informed eavesdropper in underwater acoustic channels: Feature extraction and quantization." *UComms* **2021**.

[4] Sklivanitis, George, et al. "Physical layer security against an informed eavesdropper in underwater acoustic channels: Reconciliation and privacy amplification." *UComms* **2021**.

[5] Morozs, Nils, Paul D. Mitchell, and Yuriy Zakharov. "Target detection using underwater acoustic networking." *IEEE OCEANS* **2023**.

[6] Aman, Waqas, Saif Al-Kuwari, and Marwa Qaraqe. "Location-based physical layer authentication in underwater acoustic communication networks." *IEEE VTC* **2023**.

[7] Khalid, Muhammad, Ruiqin Zhao, and Nauman Ahmed. "Physical layer authentication in line-of-sight underwater acoustic sensor networks." *IEEE OCEANS* **2020**.

[8] Zhao, Ruiqin, et al. "Physical layer node authentication in underwater acoustic sensor networks using time-reversal." *IEEE Sensors J.* **22.4**: 3796-3809 **2022**.

A Lightweight AI-powered Framework for Multimodal Underwater Networks

Filippo Donegà[1], Filippo Bragato[2], Filippo Campagnaro[3], and Michele Zorzi[4]

[1-4]Department of Information Engineering, University of Padova

Filippo Donegà, Department of Information Engineering, University of Padova, Italy
E-mail: filippo.donega@unipd.it

Abstract: *Underwater acoustic communication enables numerous long-range applications, but the low bitrate and long propagation delay that characterize the underwater acoustic channel make it difficult to support demanding applications like real-time control of Remotely Operated Vehicles (ROVs). Alternative technologies such as optical, electromagnetic radio frequency, and magneto inductive communications could be preferable, providing significantly higher rates albeit at much shorter distances. Underwater multimodal networks, combining multiple communication technologies, currently represent the best strategy for optimizing the communication performance, given that no single underwater communication technique is universally superior. It is especially important to understand how to combine the transmission technologies optimally, and how to select which one to use depending on the channel conditions and the requirements imposed by the intended application. In particular, multimodal communication can be seen as an unknown Markov Decision Problem, and Reinforcement Learning (RL) is a practical way to deal with it. In this work we showcase how Artificial Intelligence can be used to optimize multimodal communication by presenting and evaluating through DESERT simulations a lightweight QoS-based RL framework that optimally selects the best transmission medium in multimodal underwater networks. Results highlight how this framework outperforms traditional approaches, achieving better performance in many communication parameters while still being suitable for practical deployments onboard lightweight single-board computers.*

Keywords: *Underwater wireless sensor networks, multimodal communication, reinforcement learning, AI, network simulation.*

1. INTRODUCTION

Acoustic communication represents the most mature and established underwater transmission technology, enabling robust connectivity over ranges of several kilometers. This makes it well-suited for many common Underwater Wireless Sensor Network (UWSN) applications, including coastal surveillance, underwater pipeline monitoring, and water quality assessment among others [1, 2]. However, the underwater acoustic channel is significantly affected by

environmental noise from vessels and wind, and further suffers from multipath propagation and Doppler effects, which particularly degrade performance in shallow, mobile, or noisy environments. Additionally, the slow speed of sound in water results in considerable propagation delays, and the data rates supported by acoustic modems are typically low [3].

These limitations make acoustic communication unsuitable for applications requiring higher throughput, such as real-time drone control or video streaming. In these instances, alternative communication methods—namely optical, Radio Frequency (RF), or Magneto Inductive (MI) technologies—may offer better performance.

Optical communication, in particular, supports significantly higher data rates, reaching several Gbps under ideal conditions, and is therefore more appropriate for data-intensive applications. However, while immune to multipath effects and acoustic noise, optical signals are highly susceptible to environmental factors such as water turbidity (which causes signal scattering and attenuation) and ambient light noise near the surface [4], and their effective range is typically limited to a few tens of meters.

Since no single technology performs best under all conditions, multimodal underwater networks (integrating multiple transmission technologies) have emerged as a promising solution, presenting the novel challenge of dynamically selecting the most appropriate communication mode in response to varying environmental conditions and application requirements.

Nevertheless, most existing multimodal solutions struggle with scalability and adaptability in dynamic underwater environments, often requiring the choice of arbitrary thresholds or parameters in order to work properly. Moreover, established solutions for handling multimodal communication in terrestrial networks are generally too expensive (computationally and/or energetically) and therefore difficult to apply to real-life underwater deployments, which are normally characterized by resource-constrained devices.

The use case guiding the design and evaluation of the proposed framework involves two mobile Autonomous Underwater Vehicles (AUVs) cooperating to manipulate a submerged object, each one equipped with both an acoustic and an optical modem. In such context, advanced computer vision and/or AI mechanisms would be used in close proximity, requiring the exchange of real-time position updates and state information between the drones, but acoustics would still be needed for exchanging reliable ranging data to allow target approach from a distance. Therefore each traffic is assumed to be related to a specific application, which in turn poses specific Quality of Service (QoS) requirements to be satisfied.

This paper presents a lightweight, adaptive, and scalable QoS-driven Reinforcement Learning (RL) framework that autonomously selects the optimal communication mode based on real-time channel conditions and specific application requirements. The proposed approach is designed to be efficient enough for low-power underwater nodes, with the additional advantage of supporting offline training for enhanced flexibility and deployment readiness in challenging environments.

2. REINFORCEMENT LEARNING AND MARKOV DECISION PROCESSES

Reinforcement Learning (RL) is a machine learning paradigm in which an agent learns to make decisions by interacting with an environment, receiving feedback in the form of rewards. More precisely, an RL agent aims at learning a policy that maximizes its cumulative expected reward, which is a problem that can be formalized with a Markov Decision Process (MDP) [5].

At each discrete time step t, the agent observes a state $S_t \in \mathcal{S}$, selects an action $A_t \in \mathcal{A}$, receives a reward $R_t = \mathcal{R}(S_t, A_t)$, and transitions to a new state S_{t+1} according to the

transition probability function $\mathcal{P}(S_{t+1}|S_t, A_t)$. Its behavior is governed by a policy π, defined as a probability distribution over the set of actions given a state:

$$\pi(a|s) = P(A_t = a \mid S_t = s). \tag{1}$$

The expected cumulative future reward G_t to be maximized is defined as the sum of the rewards obtained from the current time step to the end of the episode, and discounted by γ at each time step:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}. \tag{2}$$

The agent does not have access to the MDP's $\mathcal{P}$ and $\mathcal{R}$, and must learn it by interacting with the environment. The algorithm proposed in this work belongs to a specific class of RL algorithms, called *value-based* methods, in which the agent learns the value of each state-action pair $q_\pi(s, a)$, defined as the expected cumulative future reward when starting from state s and taking action a, and following policy π after the first step:

$$q_\pi(s, a) = \mathbb{E}_\pi[G_t|S_t = s, A_t = a]. \tag{3}$$

Afterwards, it can select the best action in each state by following a policy that is greedy with respect to the learned values, i.e.,

$$\pi^*(s) = \arg\max_a q_\pi(s, a). \tag{4}$$

Value-based methods typically alternate between two phases: the evaluation phase, where the agent learns the value of each state-action pair, and the improvement phase, where it updates its policy to be greedy with respect to the learned values. This guarantees that the agent will eventually converge to the optimal policy. Another important concept in RL is the distinction between on-policy and off-policy algorithms. In on-policy algorithms, the agent learns the value of the policy it is following, while in off-policy algorithms, the agent learns the value of a different policy at the expense of weaker convergence guarantees.

3. SYSTEM ARCHITECTURE

The system consists of two components at each node: the *agent* and the *wait system*. The agent is responsible for selecting the optimal communication technology for each transmission, while the wait system handles a finite-size queue of packets to be transmitted. When an application generates a packet, it enters the wait system's queue. This triggers the agent, which evaluates each packet and chooses the best transmission method. If a packet cannot be sent (e.g., due to poor channel conditions), it is re-queued and the process is paused to preserve packet order. Once all applications stop producing packets, the agent performs a final transmission attempt for any remaining packets, after which the system stops.

The state of the system is configurable and composed by a subset of 16 possible parameters, selectable through a mask, and normalized. These parameters include noise readings and power of the last received transmission for each technology, acoustic and optical channel-related metrics (i.e., temperature, attenuation coefficient, wind speed and intensity of shipping activity), spatial information like distance and depth, and size of Medium Access Control (MAC) and wait queues. The mask used for this work includes all of the aforementioned parameters.

For each packet stored in the wait queue, the RL agent can undertake three possible actions: forward it to the acoustic MAC module (i.e., send in acoustic), forward it to the optical MAC module (i.e., send in optical) or wait to send the packet (i.e., leave it in the queue and pause the process until next iteration).

Each agent receives a reward based on the QoS requirements of the packet being sent. The reward function is defined to guarantee that the agent learns to select the best transmission technology for each packet, while also considering the stability of the training.

The framework as presented in this paper involves two different types of traffic, as will be clarified in Section 4, and two dedicated agents: one aiming to maximize throughput (i.e., Maximum Throughput Agent (MTA)), and the other minimizing jitter (i.e., minimum Jitter Agent (mJA)).

Concerning the MTA, the reward function is based on the packet size, the reception of the packet, and the delay experienced by the packet. In particular, if the packet is successfully received, the reward is calculated as the packet size divided by the delay. If the agent chooses to wait and not send the packet, the reward is zero. Conversely, if the packet is sent but fails to reach its destination, a negative reward is assigned.

The mJA agent's reward function is based on the jitter experienced by the packet. In particular, if the packet is received, the reward is equal to a constant minus the difference between the current and previous packet's delay. If the packet is not sent, the reward is strongly negative to discourage the agent from waiting. If the packet is instead sent but not received, the reward is negative.

The learning agent is based on the Q-learning algorithm, which is a value-based RL algorithm. In particular, we implemented a continual non-episodic version with a linear approximator of the standard Q-learning algorithm. The agent is similar to the Differential semi-gradient Sarsa presented in [5], but it uses an off-policy approach to learn the value of the policy.

It is important to note that the introduction of the off-policy approach does not guarantee the convergence of the algorithm, since we are training the agent using a bootstrapped version of an off-policy algorithm with a linear approximator. The choice of this type of algorithms, that do not have a proof of convergence, is common in the RL ecosystem. Indeed, the algorithm is still able to learn the value of the policy and to select the best action in each state, like many other well-known algorithms that fall into this category, such as the DQN algorithm [6].

4. EXPERIMENTAL RESULTS

Each presented result is obtained by averaging over outcomes from 15 campaigns, where a campaign is defined as a set of simulations (or epochs) that are related to the same agent. Simulations are performed using DESERT Underwater [7], a publicly available underwater network simulator developed and maintained by the SIGNET group at the University of Padova. In this work, we define training and evaluation epochs differently: during a training epoch, the agent is trained using the RL algorithm described in Section 3, weights are updated at each step using the semi-gradient technique, and the agent utilizes an ϵ-greedy policy to explore the environment. Training epochs belonging to the same campaign share the same environment setup but with different realization (i.e., different RNG seed), hence the agent is trained in independent and identically distributed (iid) scenarios. Instead, during an evaluation epoch, the agent is always tested in the same environment, its weights are not updated, and ϵ is set to 0, meaning that the agent always selects the best action. Each campaign alternates 100 training epochs and 100 evaluation epochs, until reaching a total of 1000 training epochs. This way,

we can measure the performance of the agent during the training process, and evaluate its final performance.

The simulated scenario, approximating the use case described in Section 1, consists of a Remotely Operated Vehicle (ROV) approaching a static node at exponentially decreasing speed, starting from a distance of 2600 m. The communication is bidirectional and involves two different ongoing traffics, each one paired to a dedicated multimodal controller (and agent) that optimizes a different reward function. Table 1 reports traffic details and respective QoS requirements optimized by the relative RL agent.

#	Data type	Period	Packet size	Agent
1	Measurement data	60 s $\rightarrow$ 17 s Exponentially decreasing with distance	64 Bytes	MTA
2	Real-time control data	10 s	100 Bytes $\rightarrow$ 750 Bytes Exponentially increasing with distance	mJA

Table 1: Simulated traffic types and respective optimization goals.

To simulate acoustic transmission, DESERT's acoustic PHY layer (which models acoustic propagation according to [8]) is employed. Furthermore, acoustic interference is considered by utilizing DESERT's MEANPOWER model. Optical transmission is instead simulated using an extended version of DESERT's optical PHY layer, equipped with the capability of providing channel statistics to the multimodal controller via cross-layer messages. This layer models optical propagation according to [9], further adding background ambient light noise in the form of a per-depth look-up table. Optical interference is also taken into account by employing DESERT's MInterference/MIV plugin. Acoustic and optical noise factors are slowly and randomly varied throughout the simulation. More specifically, wind speed starts at 10 m/s and can get to a minimum of 0 m/s (calm) or a maximum of 20 m/s (gale). Instead, the attenuation coefficient starts at 0.30 and is capped between 0.15 (clear ocean) and 0.40 (coastal ocean). Acoustic PHY bitrate and central frequency are set to 2500 bps and 26 kHz respectively, whereas optical PHY bitrate and carrier are set to 1 MHz and 10 MHz.

The performance of the proposed system is evaluated by comparing it against two baselines, representative of much simpler methods for multimodal control:

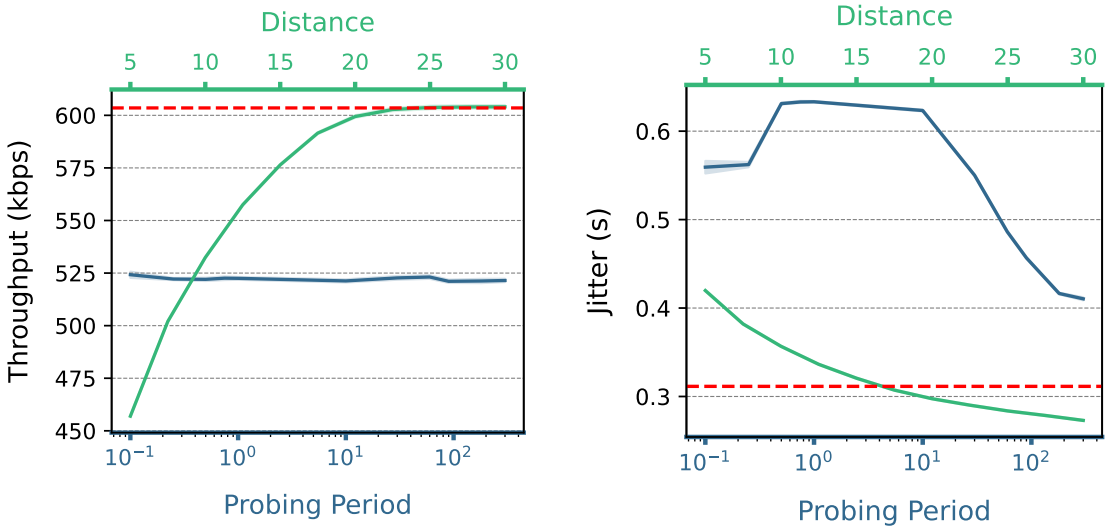
- *Range-based:* The choice of which technology to use for each packet transmission depends upon the distance between the two nodes. If the ROV is farther than a predefined distance d from the sink, acoustic communication is employed. Otherwise, every packet is sent in optical;
- *Optical probing:* Small probe packets are periodically sent in optical, and we call the period between probe transmissions τ. Once a probe packet is received, optical transmission is employed for a predetermined amount of time. In simulations, this time is set as equal to the inter-generation time of optical probes, allowing continuous optical transmission in case subsequent probes are correctly received.

Fig. 1 depicts the performance of the proposed framework in terms of throughput and jitter, compared with the two baselines. Results achieved by the RL agent are rendered as red dashed

lines, while the performance results for range-based and optical probing are represented by green and blue solid lines respectively.

In Fig. 1a, the displayed throughput is computed as the ratio between the total number of received packets and the overall time taken to receive them. The proposed framework (red dashed line) achieves a throughput over $600\,\text{kbps}$, which is more than $13\,\%$ higher than the throughput obtained by the probing method (blue solid line) and outperforms the range-based performance (green solid line) for $d < 20\,\text{m}$, getting instead similar results for $d > 20\,\text{m}$. This is due to the fact that the range-based method is not able to adapt to the changing conditions of the environment, while the proposed framework is able to adapt and to select the best transmission technology for each packet.

In Fig. 1b, jitter (also known as Packet Delay Variation (PDV)) is defined as the average difference between the delay of each packet and that of the subsequent one. The proposed framework achieves a jitter of $0.3\,\text{s}$, heavily outperforming the probing method, which tends to switch more often between the two technologies, and the range-based method when the distance is < 20 m.



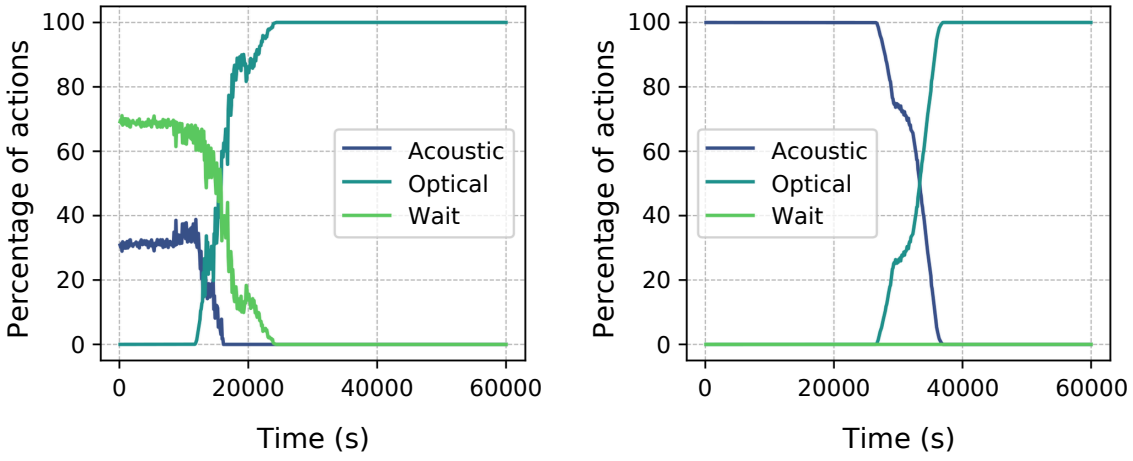
(a) Average throughput.

(b) Average jitter.

Figure 1: Comparison of the proposed framework (red dashed line) with the two baselines (green and blue solid lines for range-based and optical probing respectively) in terms of throughput and jitter.

In Fig. 2, the percentage of packets sent in each technology is displayed over time, showing how the RL agent changes the transmission technology throughout the course of a simulation. In particular, Fig. 2a shows that the MTA initially transmits a few packets acoustically while holding most of them in the wait queue. It begins sending packets only when the optical link is available. This behavior arises because using the acoustic link exclusively would congest the network, ultimately reducing throughput. Instead, in Fig. 2b, we can see that the mJA has a sudden transition from sending packets in the acoustic channel to sending them in the optical channel. This is due to the fact that the mJA agent is trained to minimize the jitter, and each switch between the two technologies introduces some jitter, hence the agent tries to minimize the number of switches between the two technologies.

To demonstrate that the presence of multiple agents does not affect the performance of the system, we also plot the actions taken by each agent over the training epochs. In Fig. 3, we

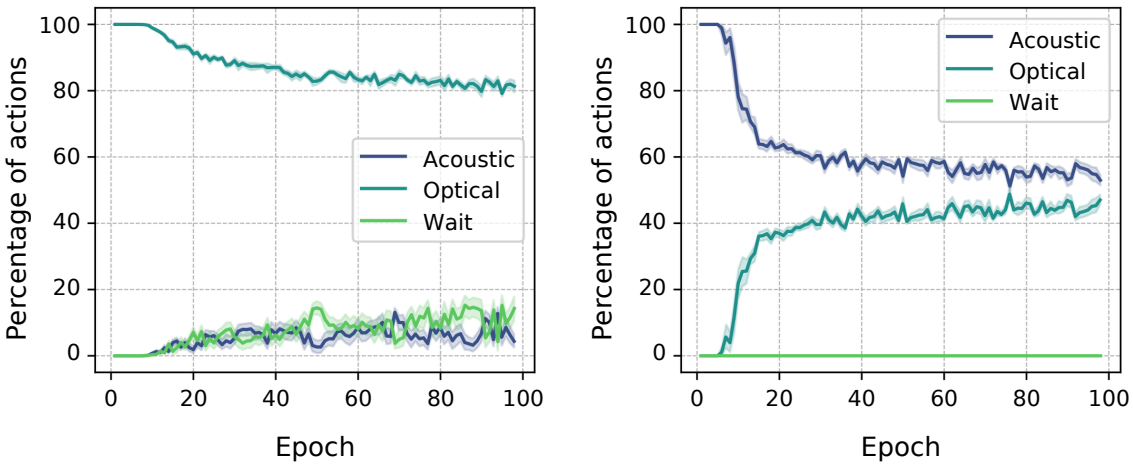


(a) MTA's actions. *(b) mJA's actions.*

Figure 2: Actions taken by the MTA and mJA agents over time.

display the actions taken by each agent during the training. The actions are the ones taken in the evaluation phase, and the results are averaged over 15 campaigns. Results show that the mJA agent is able to learn its policy faster than the MTA agent, but the evolution of the policy of the MTA does not affect the stability of the mJA agent.

This ensures that the training of the two agents is stable and that the presence of multiple agents does not affect the convergence of the system. Even if the two agents are trained in the same environment, they are able to learn their policy independently and to converge to the optimal policy.



(a) MTA's actions. *(b) mJA's actions.*

Figure 3: Actions taken by the MTA and mJA agents over training.

5. CONCLUSION

The proposed QoS-based AI framework for handling multimodal communication in underwater networks was tested in simulation and proved to perform better than the two baselines in

most cases. The main strengths of the proposed framework are its high customizability, its ease of extension to new applications with different QoS requirements, and its adaptability, since the RL agent can automatically adapt to new scenarios and use cases. Future work will involve agents supervision through multi-agent RL techniques, that enable the coordination between agents to maximize the total reward, the addition of new traffic types (e.g., video stream) and new QoS requirements to be optimized (e.g., Packet Reception Rate (PRR) optimization).

6. ACKNOWLEDGMENTS

This work was partially supported by the PRIN project BEASTIE (2022FZJ7MR), funded by the Italian Ministry of University and Research; and the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGeneration EU, partnership on "Telecommunications of the Future" (PE0000001-program "RESTART").

REFERENCES

[1] E. Felemban, F. K. Shaikh, U. M. Qureshi, A. A. Sheikh and S. B. Qaisar: "Underwater Sensor Network Applications: A Comprehensive Survey", *International Journal of Distributed Sensor Networks*, 11(11), 896832 (2015).

[2] F. Campagnaro, F. Steinmetz and B. C. Renner: "Survey on Low-Cost Underwater Sensor Networks: From Niche Applications to Everyday Use", *Journal of Marine Science and Engineering*, 11(1), 125 (2023).

[3] F. Campagnaro, R. Francescon, P. Casari, R. Diamant and M. Zorzi: "Multimodal Underwater Networks: Recent Advances and a Look Ahead" in *Proceedings of the 12th International Conference on Underwater Networks & Systems (WUWNet)*, 2017.

[4] A. Signori, F. Campagnaro and M. Zorzi: "Modeling the Performance of Optical Modems in the DESERT Underwater Network Simulator" in *IEEE Fourth Underwater Communications and Networking Conference (UComms)*, 2018.

[5] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, (MIT press Cambridge, 2018).

[6] H. Van Hasselt, A. Guez, and D. Silver: "Deep reinforcement learning with double q-learning" in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016.

[7] F. Campagnaro, R. Francescon, F. Guerra, F. Favaro, P. Casari, R. Diamant and M. Zorzi: "The DESERT Underwater Framework v2: Improved Capabilities and Extension Tools" in *IEEE Third Underwater Communications and Networking Conference (UComms)*, 2016.

[8] M. Stojanovic: "On the Relationship Between Capacity and Distance in an Underwater Acoustic Communication Channel" in *ACM SIGMOBILE Mobile Computing and Communications Review*, 11(4), 34-43, 2007.

[9] D. Anguita, D. Brizzolara, G. Parodi, and Q. Hu: "Optical Wireless Underwater Communication for AUV: Preliminary Simulation and Experimental Results" in *IEEE Oceans - Spain*, 2011.

Error Correction Code for FSCM-based Acoustic Underwater Communication

Julia Marie Wüpperling[1], Fabian Steinmetz[1], and Bernd-Christian Renner[1]

[1]Hamburg University of Technology, Institute for Autonomous Cyber-Physical Systems

Bernd-Christian Renner
Hamburg University of Technology
Institute for Autonomous Cyber-Physical Systems
Harburger Schloßstraße 28, 21079 Hamburg, Germany
christian.renner@tuhh.de

Abstract: *Efficient and reliable low-power acoustic underwater communication is essential for underwater applications, such as wireless sensor networks or swarms of miniature underwater robots. Moreover, underwater communication is already part of 6G. Our recent research focused on frequency shift chirp modulation (FSCM), which is already used for low-power over-water radio communication. However, the acoustic underwater channel differs in terms of propagation speed, Doppler shifts, attenuation, and multipath propagation. While we have already implemented the de-/modulation, in this paper we look into identifying an appropriate error correction code. First, we analyze the error distribution of our current FSCM implementation to understand the underlying causes of errors and their patterns. Second, we evaluate different error correction codes for FSCM through simulations in Watermark using real-world channel impulse responses. Our research demonstrates significant improvements in reliability for underwater communication, contributing to the underwater research community's pursuit of developing new solutions for low-power underwater applications.*

Keywords: *underwater communication, acoustic, FSCM, chirp, coding*

1. INTRODUCTION

Underwater Acoustic Communication (UWAC) systems are integral to various marine applications; e.g., the deployment of Underwater Wireless Sensor Networks (UWSNs) for environmental monitoring and scientific research. UWSNs may consist of a variety of sensor nodes, such as stationary seafloor units, surface buoys, Autonomous Underwater Vehicles (AUVs), and Autonomous Surface Vehicles (ASVs). These nodes collaboratively collect mission-specific data, ranging from seafloor pressure for tsunami detection to water quality parameters for monitoring marine ecosystem health [4]. To enable long-term autonomous operation, they may be equipped with low-power components and operate with limited computational resources [11].

Reliable communication is fundamental for successful data collection and remote control. However, utilizing underwater acoustic channels for signal transmission poses unique challenges, since phenomena such as absorption, multipath propagation, Doppler effects and environmental noise can significantly impair signal quality [6]. Especially in low-power systems, energy-efficient communication and avoidance of retransmissions is mandatory.

As we recently demonstrated in [12], the relatively new modulation technique Frequency Shift Chirp Modulation (FSCM), developed in the context of low-power radio communication, offers strong resilience against these impairments compared to traditional modulation schemes such as Frequency Shift Keying (FSK). To mitigate remaining transmission errors, channel coding is required for forward error correction [8]. Unitl now, channel coding for (underwater acoustic) FSCM has received little attention in existing research.

We address this gap by investigating the integration of FSCM with channel coding strategies suitable for low-power UWAC systems through the following contributions:

- **Analysis of Uncoded FSCM Transmissions:** In a simulation study in Section 3, we analyze how errors manifest in FSCM transmissions without channel coding. We study bit errors and packet reception success under various channel conditions.
- **Evaluation of Coded FSCM Transmissions:** In Section 4, we extend the previous study with coded FSCM. To assess the trade-off between reliability and transmission efficiency, improvements in packet reception are quantified alongside the impact on the data rate.
- **Recommendation of Practical Channel Codes:** Based on a final discussion in Section 4, we give recommendations for channel coding strategies in FSCM-based low-power UWAC systems.

2. RELATED WORK

In the last decade, LoRaWAN [5] was developed and implemented for over-water Low-Power Wide-Area Networks (LPWANs) such as Internet of Things (IoT) applications. LoRaWAN is composed of a physical (PHY) layer, the LoRa-PHY (frequently called LoRa), and a Medium Access Control (MAC) layer. The LoRa-PHY uses Chirp Spread Spectrum (CSS) modulation, which is usually called FSCM [18]. However, the LoRa-PHY is a proprietary standard, and implementation details are unspecified. Therefore, research has been conducted on packet synchronization and demodulation [16, 19], multipath propagation [3], error coding [1, 20], and real-world trials [5]. In general, FSCM showed remarkable resilience against multipath propagation and low Signal-to-Noise Ratios (SNRs) for over-water communication.

Motivated by the promising results from the over-water communication, researchers started to investigate FSCM for acoustic underwater communication [10, 7, 21]. However, the acoustic underwater channel differs from the over-water radio communication channel: The available bandwidth is smaller, the Doppler shifts are larger, and the acoustic channel is distorted by massive multipath propagation [15].

In previous publications, we showed resilience against multipath propagation [12] and developed methods to reduce the effect of Doppler shifts [13]. In real-world trials, we showed that our FSCM implementation for acoustic underwater communication achieves very low Bit Error Rates (BERs) [14]. However, we observed a difference between the static and mobile transmission channels. Our Doppler shift removal reduces the BERs, but the BER was higher compared to static scenarios. Until now, we have not applied any error correction codes to detect and correct transmission errors. Likewise, other researchers have also implemented FSCM without additional error coding [10, 7, 21]. Therefore, error correction codes for FSCM-based acoustic underwater communication are a prevailing research gap.

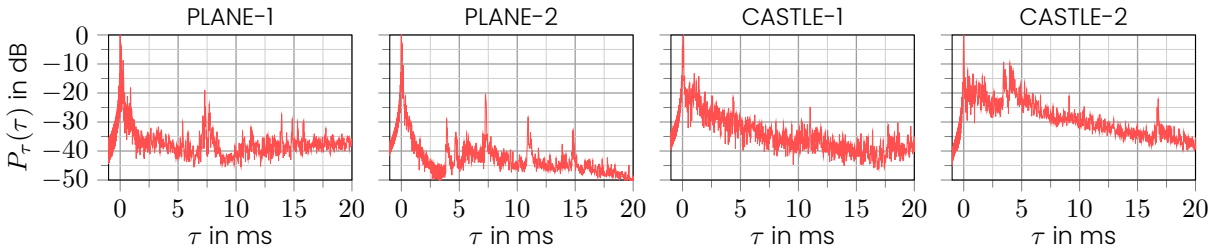


Figure 1: Power delay profiles of real-world acoustic underwater channels ($50\,\text{kHz} - 75\,\text{kHz}$*) from [12]. PLANE-1 and CASTLE-2 are used for the simulations.*

In over-water communication, on the contrary, FSCM has already been used with different codings. For example, LoRa uses a Hamming code [5]. Moreover, different codings have been investigated, for example, Hamming code [1], Reed-Solomon (RS) codes [2], and Low Density Parity Check (LDPC) codes [20]. We use this is a starting point for this paper.

3. ANALYSIS OF UNCODED FSCM TRANSMISSIONS

For selecting a favorable error coding and understanding the error characteristics of uncoded FSCM transmissions, we conducted a simulation study across diverse channel conditions.

3.1. SIMULATION SETUP AND METRICS

We ran simulations in MATLAB with our previously implemented differential FSCM modulation and demodulator [13]. To simulate the acoustic underwater channel, we use Watermark [17] with recorded channel impulse responses and apply configurable variations such as timing offsets, Doppler shifts, and noise levels. We compare demodulated output to the original data input and study the amount and distribution of bit errors. We transmit data packets with $128\,\text{B}$ of random data, modulated with spreading factors $\text{SF} \in \{6, 8, 10\}$. While a higher SF for the synchronization vs. data is favorable [14], we apply a common SF across the entire packet to limit the complexity. We use a subset of channel impulse responses from [12] shown in Fig. 1: PLANE-1 and CASTLE-2. They present challenging yet different conditions, where PLANE channels exhibit a dominant early-arriving path, followed by multiple closely spaced multipath components with gradually decreasing power. The CASTLE environments show denser and more prolonged multipath components, representing more complex conditions typically found in ports with multiple reflecting surfaces.

For each configuration, we evaluated 90 packets. The corresponding simulated acoustic signals are systematically modified to emulate realistic transmission impairments. Signal arrival times are varied within a range of $10\,\text{ms}$ to $100\,\text{ms}$ to simulate various propagation delays. Doppler shifts are introduced for relative velocities between $-3\,\text{m}\,\text{s}^{-1}$ and $3\,\text{m}\,\text{s}^{-1}$. We use SNR levels from $-6\,\text{dB}$ to $15\,\text{dB}$. Communication parameters—i.e., frequency band, sampling rate—match our ahoi modem [9]. Table 1 summarizes all simulation parameters.

Finally, the modified received signals are demodulated and compared to the original data input, which enables the direct assessment of the error characteristics of uncoded FSCM transmissions across the defined parameter space. We use Packet Success Rate (PSR)—the fraction of error-free packets—and BER—the fraction of erroneous bits—of all successfully synchronized packet transmissions per configuration for our evaluation. To assess the influence of packet length, we look into the first $8\,\text{B}$ of these packets separately.

Parameter	Symbol	Unit	Values
Packet size	n_{data}	B	128
Spreading Factor	SF	—	$6, 8, 10$
Bandwidth	BW	kHz	20
Carrier frequency	f_{c}	kHz	62.5
Sampling frequency	$f_{\text{s,tb}}$	kHz	200
Channel	—	—	PLANE-1, CASTLE-2
Number of evaluated packets	n_{pkt}	—	90
Time delay	δ_{t}	ms	$10 - 100$
Relative speed	v	$\text{m}\,\text{s}^{-1}$	$-3, -1.5, 0, 1.5, 3$
Signal-to-noise ratio	SNR	dB	$-6, -3, 0, 3, 6, 9, 12, 15$

Table 1: Simulation parameters for uncoded FSCM transmissions.

3.2. RESULTS

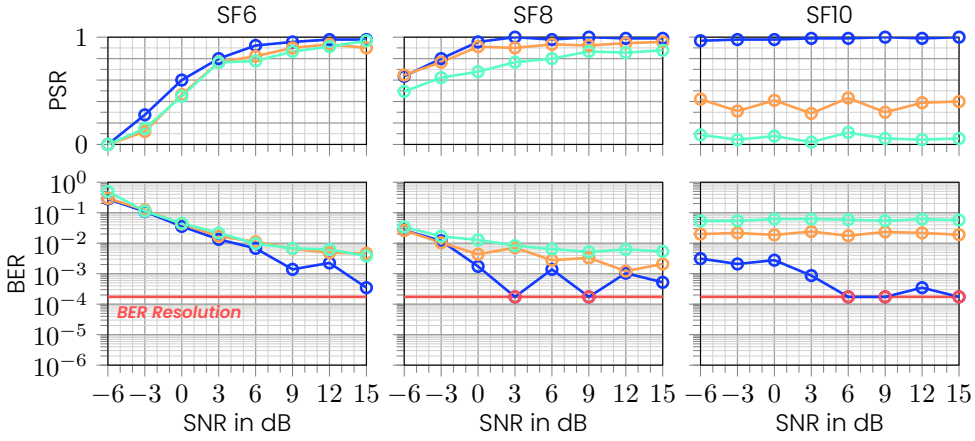
Figures 2 and 3 present the PSR and BER for $8\,\text{B}$ and $128\,\text{B}$ data packets in the PLANE-1 and CASTLE-2 channels, respectively. To indicate cases of error-free reception, the BER plots include a *BER resolution* line, indicating the smallest non-zero BER for the packet length. It is 1.74×10^{-4} for $8\,\text{B}$ packets and 1.09×10^{-5} for $128\,\text{B}$ packets. For configurations without any synchronized packets—thus, no packet was demodulated—the BER and PSR are set to 0.5 and 0, respectively. This only occured in few configurations for $\text{SF} = 6$ and a negative SNR.

Variations of PSR and BER values for adjacent SNR levels stem from the small number of transmitted packets. As a result, even minor differences in the number of correctly demodulated packets can noticeably shift PSR values, and few additional bit errors near the BER resolution threshold are observed as visible steps in the BER curves.

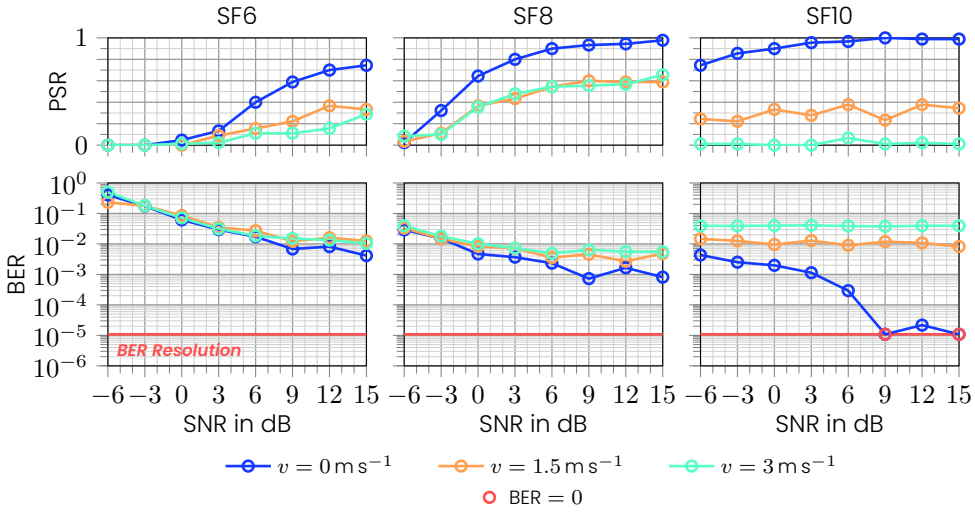
The PSR and BER trends in PLANE-1 exhibit similiar patterns across both packet sizes in Figs. 2(a) to 2(b). For all values of SF, the PSR increases with higher SNRs, while the BER decreases accordingly. This aligns with the expectation that lower noise levels improve demodulation. Only the combination of $\text{SF} = 10$ with non-zero relative speeds deviates from this trend, where both metrics remain nearly constant across all SNR levels. This suggests that at $\text{SF} = 10$ demodulation is primarily limited by Doppler effects rather than noise. In fact, as v increases, the PSR reduces notably, falling below $5\,\%$ for $v = 3\,\text{m}\,\text{s}^{-1}$, while the BER remains high. This observation supports an increased sensitivity for $\text{SF} = 10$ to motion-induced frequency distortions. In contrast, $\text{SF} = 6$ and $\text{SF} = 8$ show greater resilience to Doppler shifts, with less diverging PSR and BER curves across different relative velocities.

Conversely to the Doppler effect, variations in packet length have a stronger influence at $\text{SF} = 6$ and $\text{SF} = 8$. This is evident from the lower PSR for long packets ($128\,\text{B}$). One cause of this behavior lies in Doppler estimation accuracy: lower spreading factors result in shorter symbol durations, reducing Doppler estimation precision. While short packets may still be demodulated correctly under small inaccuracies in the estimated time scaling, the misalignment between the receiver's sampling window and the actual symbol positions progressively increases with packet length, leading to lower PSR and higher BER values. The increased symbol duration of $\text{SF} = 10$, in contrast, allows for more accurate Doppler estimation, resulting in similar PSR and BER values across both packet lengths.

The trends in CASTLE-2 generally follow those observed in PLANE-1, but show notable differences across spreading factors and mobility. For $\text{SF} = 6$, the PSR is lower, particularly for longer packets, where it remains below $40\,\%$. This degradation can be attributed to stronger multipath interference. According to the channel's Power Delay Profile (PDP) (see



(a) PSR and BER for 8 B uncoded data.



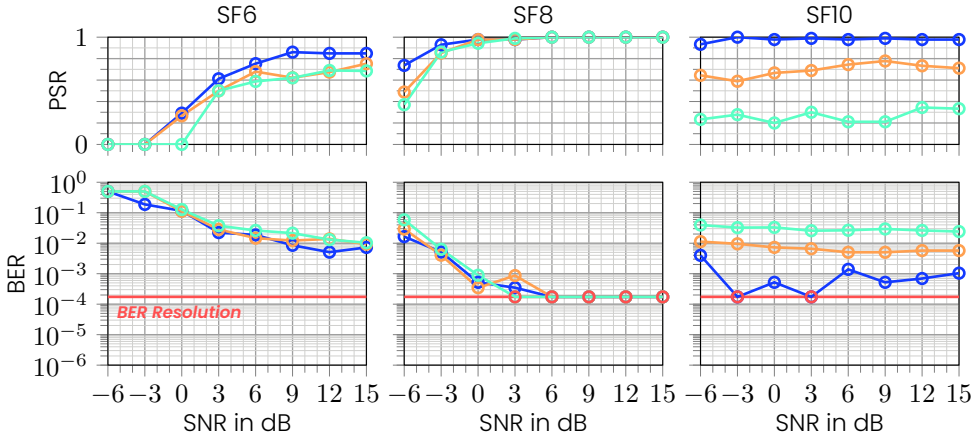
(b) PSR and BER for 128 B uncoded data.

Figure 2: PSR and BER of 8 B *and* 128 B *uncoded data in PLANE-1.*

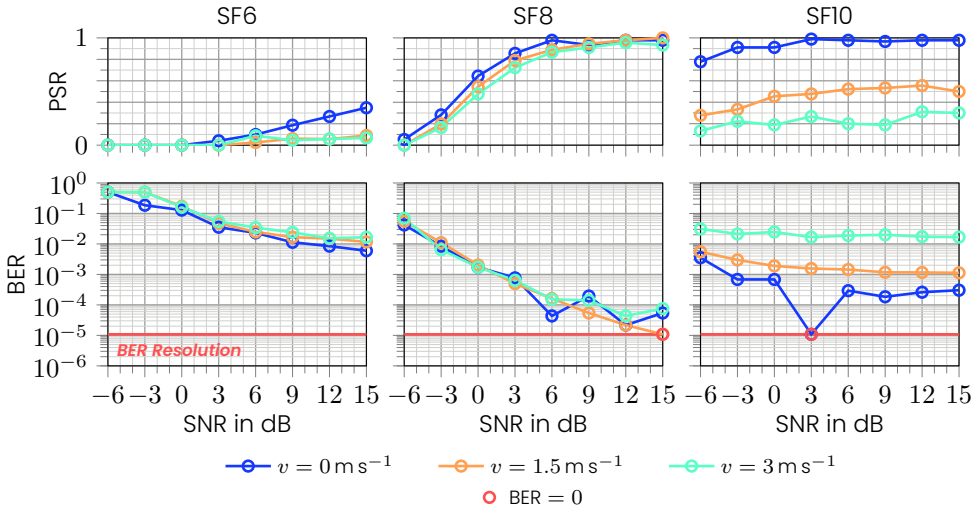
Fig. 1), CASTLE-2 exhibits strong reflections around 3 ms to 4 ms after the main path. With a symbol duration of only 3.2 ms at SF $= 6$, these reflections overlap with subsequent symbols and degrade demodulation accuracy in terms of inter-symbol interference. For SF $= 8$, however, CASTLE-2 produces less errors, especially at higher SNRs. Additionally, Doppler-induced packet loss is reduced, as the PSR and BER in mobile scenarios converge towards the stationary case. This behavior may be attributed to CASTLE-2's stronger and more sustained multipath components within the 12.8 ms symbol duration of SF $= 8$, which produces less inter-symbol interference. A similar improvement is observed for SF $= 10$, where the PSR in mobile scenarios is approximately 10 % higher compared to PLANE-1. Notably, the observed improvements in PSR and BER over PLANE-1 are more pronounced for shorter packets, while longer packets still experience substantial packet loss.

3.3. SUMMARY

The results indicate that uncoded FSCM transmissions exhibit limitations under challenging transmission conditions; e.g., significant Doppler shifts or symbol lengths in the order of the channel delay spread. SF $= 10$ enables reliable demodulation in stationary scenarios even at low SNR. In contrast, SF $= 6$ and SF $= 8$ are more robust to Doppler shifts, maintaining



(a) PSR and BER for 8 B uncoded data.



(b) PSR and BER for 128 B uncoded data.

Figure 3: PSR and BER of 8 B *and* 128 B *uncoded data in CASTLE-2.*

higher PSR values under motion. However, they require higher SNR levels to achieve reliable demodulation, with SF $= 6$ being particularly noise-sensitive.

Given these findings, channel coding is useful to improve communication reliability. Moreover, the results indicate that the BER has a high range across the parameters, and a considerable amount of packets can be transmitted without any bit error. We also found randomly distributed and burst bit errors in our study. Due to the page limitation, we omitted these results and the complex discussion.

4. EVALUATION OF CODED FSCM TRANSMISSIONS

To evaluate the effect of channel coding on FSCM-based low-power UWAC, we repeated the previous study with different coding techniques. We used available MATLAB functions. Based on the codes used in LoRa (cf. Section 2), the small packet sizes in acoustic underwater communication, and our results from Section 3, we chose $(7, 4)$, $(15, 11)$, and $(31, 26)$ Hamming[1] codes as well as the $(133, 171)$ convolutional code (CC) with code rates $R_\mathrm{c} \in \left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}\right\}$. All codes

[1]Note that we use the code rate of all codes in the plots, so the $(7, 4)$ Hamming code is named $4/7$ Ham(ming).

Parameter	Symbol	Unit	Values
Packet size	n_{data}	B	$8, 64$
Number of evaluated packets	n_{pkt}	—	2700
Relative speed	v	$\mathrm{m\,s^{-1}}$ (Doppler Shift)	$0, 0.75, 1.5, 3, 6$

Table 2: Simulation parameters for coded FSCM transmissions; parameters unchanged from Table 1 have been omitted.

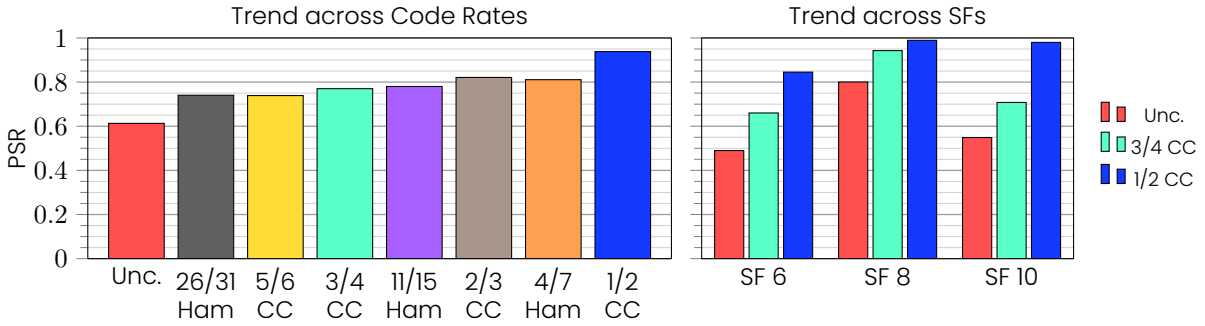


Figure 4: Mean PSR of FSCM transmissions vs. channel codes with block interleaving (left) and spreading factor (right). Results are limited to SNRs ≥ 0 dB.

are combined with block interleaving. We analyze the impact of the codes through PSRs—as an indicator for communication reliability—and data rate. Parameters that were changed from the previous study are summarized in Table 2.

4.1. PACKET SUCCESS RATE

Figure 4 (left plot) shows the mean PSRs (after decoding) for the different codes over all evaluated conditions, in decreasing order of code rate. The PSR increases as the code rate decreases, reflecting the stronger error correction due to added redundancy. The rate-1/2 convolutional code achieves the highest PSR of $93.8\,\%$, improving the uncoded case by approximately $30\,\%$ points. This gain exceeds that of all other evaluated codes by more than $10\,\%$ points. The remaining codes achieve PSRs between $73.9\,\%$ and $82.1\,\%$, with similar code rates yielding comparable results. Notably, the rate-2/3 convolutional code matches the $(7, 4)$ Hamming code in terms of PSR, despite its higher code rate.

When comparing the mean PSR vs. spreading factor, displayed in Fig. 4 (right plot) for uncoded transmissions and convolutional coding with $R_{\text{c}} \in \left\{\frac{3}{4}, \frac{1}{2}\right\}$, the impact of channel coding varies: the strongest gains occur at $\mathrm{SF} = 10$, where the rate-1/2 convolutional code raises the mean PSR to $98\,\%$, which is more than $20\,\%$ points higher than any other evaluated code. At $\mathrm{SF} = 8$, improvements are smaller, and even weaker codes achieve high PSRs, such as $94.3\,\%$ with the rate-3/4 convolutional code, reflecting the balanced resilience of $\mathrm{SF} = 8$ to noise and Doppler shifts observed in the uncoded analysis. At $\mathrm{SF} = 6$, coding still provides clear gains, but its increased susceptibility to multipath effects limits overall reliability, with mean PSRs remaining below $85\,\%$.

4.2. DATA RATE

Beyond error correction capability, the data rate is a key performance aspect of coded FSCM transmissions. The gross data rate of FSCM is the ratio of the number of bits SF and the du-

ration T_{sym} per symbol with $T_{\text{sym}} = 2^{\text{SF}}/\text{BW}$. If channel coding is used, the data rate R_{b} is proportional to the code rate R_{c} and the gross data rate. Relative to SF $= 6$, the data rates for SF $= 8$ and 10 are $33.3\,\%$ and $10.4\,\%$, respectively.

While both the spreading factor and code rate have a significant impact on the data rate of coded FSCM transmissions, the spreading factor is the primary determinant. Therefore, to achieve high data rates, high code rates and, in particular, low spreading factors should be selected. Simultaneously, higher data rates lead to decreased power consumption, as the number of samples per packet is directly proportional to the packet duration. This consideration is particularly relevant for low-power UWAC systems.

4.3. DISCUSSION OF CODING SELECTION FOR SYSTEM INTEGRATION

Stronger coding schemes, such as the rate-1/2 convolutional code, offer superior error correction but introduce substantial redundancy, thereby reducing the data rate. Consequently, optimal channel coding in FSCM-based low-power UWAC systems must balance error correction strength and data efficiency to ensure overall system suitability.

As shown in Sections 4.1 and 4.2, both error correction capability and data rate depend not only on the selected coding scheme, but also on the spreading factor. The spreading factor has the dominant effect on the data rate due to its exponential influence on symbol duration. In terms of communication reliability, however, higher spreading factors improve synchronization by enabling more precise Doppler estimation and achieve higher PSRs in multipath-dominated environments. As demonstrated in Fig. 4, SF $= 6$ provides the lowest reliability with mean PSRs below $85\,\%$, while SF $= 8$ and SF $= 10$ achieve nearly error-free transmission when supported with adequate coding. Between the two, SF $= 8$ offers higher data efficiency and achieves higher PSRs with weaker codings, making it the most suitable choice for FSCM-based low-power UWAC systems.

Beyond the spreading factor, coding effectiveness also varies with parameters such as data length, Doppler shift, and SNR. These variations highlight the need for adaptive coding strategies that respond to scenario-specific constraints. Due to the page limit, we concentrate on mobile scenarios and SF $= 8$. As shown in Fig. 5, convolutional and Hamming codes achieve similar PSRs for similar code rates. In general, the PSR is negatively affected by larger packet size and higher speed. For longer data packets, and particularly in the high-mobility scenario, the code rate has a stronger effect on the PSR.

A closer examination of the convolutional codes in Fig. 5 reveals that for short $(8\,\text{B})$ packets all rates achieve PSRs above $95\,\%$, even under high mobility. Thus, for short messages the rate-5/6 convolutional code and the $(31, 26)$ Hamming code appear to offer the most efficient choice. However, for very small packet sizes, the code and packet structure play a crucial role: For the proposed choices, the Hamming code operates on blocks of 26 data bits, whereas the convolutional coder adds one redundant bit at the beginning of each 5-bit input cycle. For a $7\,\text{B}$ packet, e.g., the coded data has a length of $93\,\text{bit}$ (3 blocks) and $68\,\text{bit}$ (11 cycles plus one input bit) for Hamming and convolutional coding, respectively. For SF $= 8$ in combination with block interleaving, $96\,\text{bit}$ and $72\,\text{bit}$ are transmitted. Therefore, convolutional codes are favorable due to the higher actual data rate at similar PSR.

For longer $(64\,\text{B})$ packets, however, mobility has a greater impact. In low-mobility conditions, codes with $R_{\text{c}} \leq \frac{3}{4}$ enable mean PSRs above $95\,\%$, but only the rate-1/2 convolutional code maintains this reliability in high-mobility scenarios. If packet retransmissions are tolerable, a lower code may be preferable, because the overall bandwidth usage is typically lower; i.e., the expected number of bits to be transmitted for successful packet transmission is lower.

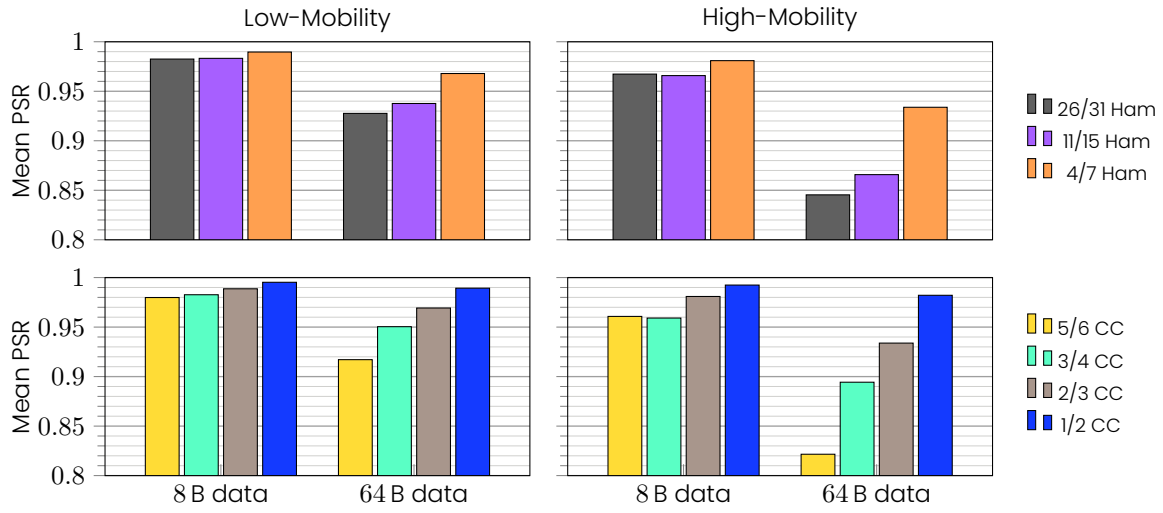


Figure 5: Mean PSRs for different channel codes with block interleaving for $\mathrm{SF} = 8$ *in low-mobility (*$v \in \{0, 0.75\}\,\mathrm{m\,s^{-1}}$*) and high-mobility (*$v \in \{1.5, 3, 6\}\,\mathrm{m\,s^{-1}}$*) scenarios.*

The exact spot of the break-even-point, though, depends on other parameters as, e.g., the length of the packet preamble.

In summary, robust and efficient low-power underwater communication using FSCM requires a joint selection of spreading factor and coding scheme. A spreading factor of $\mathrm{SF} = 8$ combined with the $(133, 171)$ convolutional code and a small set of puncturing patterns, enables a practical, flexible, and adaptive coding strategy. This approach maintains high PSRs across diverse underwater conditions while preserving high data rates.

5. CONCLUSION

We presented our results of a comprehensive simulation study of bit errors and packet success rates for uncoded and coded FSCM transmissions. We identified suitable codes and discussed the influence of several parameters. We found that a spreading factor $\mathrm{SF} = 8$, combined with the rate-adaptable $(133, 171)$ convolutional code and classical block interleaving is recommended for low-power UWAC systems. It achieves a balance between high reliability and data rate across diverse transmission conditions.

REFERENCES

[1] Q. Chen, L. Zhao, Y. Chu, and W. Guo. Research on Multipath Reception and Soft Decision Algorithms for Frequency Shift Chirp Modulation. In *Int. Symposium on Networks, Computers and Communications (ISNCC)*, Shenzhen, China, July 2022. IEEE.

[2] U. Coutaud and B. Tourancheau. Channel Coding for Better QoS in LoRa Networks. In *14th Int. Conf. on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Limassol, Cyprus, Oct. 2018. IEEE.

[3] C. Demeslay, P. Rostaing, and R. Gautier. Simple and Efficient LoRa Receiver Scheme for Multipath Channel. *IEEE Internet of Things Journal (IoT-J)*, 9(17), Sept. 2022.

[4] E. Felemban, F. K. Shaikh, U. M. Qureshi, A. A. Sheikh, and S. B. Qaisar. Underwater Sensor Network Applications: A Comprehensive Survey. *Int. Journal of Distributed Sensor Networks*, 11(11), Nov. 2015.

[5] J. C. Liando, A. Gamage, A. W. Tengourtius, and M. Li. Known and Unknown Facts of LoRa: Experiences from a Large-scale Measurement Study. *ACM Transactions on Sensor Networks*, 15(2), Feb. 2019.

[6] Y. Lou and N. Ahmed. *Underwater Communications and Networks*. Springer Int. Publishing, 2022.

[7] A. Petroni, G. Scarano, R. Cusani, and M. Biagit. Feasibility and Performance Analysis of Underwater Acoustic LoRa Modulation. In *IEEE Int. Conf. on Communications Workshops (ICC Workshops)*, Rome, Italy, May 2023.

[8] J. G. Proakis and M. Salehi. *Digital communications*. McGraw-Hill, Boston, Mass. [u.a.], 5th edition, 2008.

[9] B.-C. Renner, J. Heitmann, and F. Steinmetz. ahoi: Inexpensive, Low-power Communication and Localization for Underwater Sensor Networks and μAUVs. *ACM Transactions on Sensor Networks*, 16(2), Jan. 2020.

[10] M. Rezzouki and G. Ferre. Design and Implementation of Differential Chirp Spread Spectrum System for Underwater Acoustic Communication. In *OCEANS*, Hampton Roads, VA, USA, Oct. 2022. IEEE.

[11] A. Signori, F. Campagnaro, F. Steinmetz, B.-C. Renner, and M. Zorzi. Data Gathering from a Multimodal Dense Underwater Acoustic Sensor Network Deployed in Shallow Fresh Water Scenarios. *Journal of Sensor and Actuator Networks*, 8(4), Nov. 2019.

[12] F. Steinmetz and B.-C. Renner. Taking LoRa for a Dive: CSS for Low-Power Acoustic Underwater Communication. In *6th Underwater Communications and Networking Conf. (UComms)*, Lerici, Italy, Aug. 2022. IEEE.

[13] F. Steinmetz and B.-C. Renner. Doppler Shift and SFO Robust Synchronization for LoRa-Like Acoustic Underwater Communication. *IEEE Access*, 11, 2023.

[14] F. Steinmetz and C. Renner. Practical Evaluation of Differential Frequency Shift Chirp Modulation for Acoustic Underwater Communication. In *Proceedings of the 17th Int. Conf. on Underwater Networks & Systems*, WUWNet 2023, New York, NY, USA, Nov. 2023. ACM.

[15] M. Stojanovic and J. Preisig. Underwater acoustic communication channels: Propagation models and statistical characterization. *IEEE Communications Magazine*, 47(1), Jan. 2009.

[16] J. Tapparel, O. Afisiadis, P. Mayoraz, A. Balatsoukas-Stimming, and A. Burg. An Open-Source LoRa Physical Layer Prototype on GNU Radio. In *IEEE 21st Int. Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Atlanta, GA, USA, May 2020.

[17] P. A. van Walree, F.-X. Socheleau, R. Otnes, and T. Jenserud. The Watermark Benchmark for Underwater Acoustic Modulation Schemes. *IEEE Journal of Oceanic Engineering (JOE)*, 42(4):1007–1018, Oct. 2017.

[18] L. Vangelista. Frequency Shift Chirp Modulation: The LoRa Modulation. *IEEE Signal Processing Letters*, 24(12), Dec. 2017.

[19] M. Xhonneux, O. Afisiadis, D. Bol, and J. Louveaux. A Low-Complexity LoRa Synchronization Algorithm Robust to Sampling Time Offsets. *IEEE Internet of Things Journal (IoT-J)*, 9(5), Mar. 2022.

[20] K. Yang and W. Du. A Low-Density Parity-Check Coding Scheme for LoRa Networking. *ACM Transactions on Sensor Networks*, 20(4), July 2024.

[21] T. Ye, B. Qin, W. Cheng, Z. Peng, S. Song, and J. Cui. Recursive Distortion Recovery by Leveraging Interference in CSS-based UWAC Systems. In *ACM Int. Conf. on Underwater Networks & Systems (WUWNet)*, Šibenik, Croatia, Oct. 2024.

TARGET DETECTION IN AN UNDERWATER ACOUSTIC NETWORK: LAKE EXPERIMENT

Lu Shen*, Yuriy Zakharov*, Benjamin Henson*, Nils Morozs*, Benjamin Sherlock†, Paul D. Mitchell*

* School of Physics, Engineering and Technology, University of York, York YO10 5DD, UK

† School of Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, UK

Corresponding author: Lu Shen, lu.shen@york.ac.uk

Acknowledgements: This work was supported by U.K. EPSRC under Grant EP/V009591/1.

Abstract: *The aim of this work is to investigate the target detection performance of a distributed underwater acoustic (UWA) communication network with lake experimental data. In recent work, a target detection algorithm was proposed based on measurement of the time-varying channel impulse response (CIR) between a communication transmitter and receiver. The detector computes the normalized mean squared deviation (MSD) between a CIR of interest and a set of reference CIRs (estimated when the target is absent) and uses the minimum normalized MSD (nMSD) as the statistic for target detection. Before computing the nMSD, a filter is applied to the CIR to compensate for the change in delay and amplitude caused by environmental factors, such as node drift or moving surface waves. The performance of the detector was investigated in a small-scale lake experiment and showed clear indication of the target crossing the communication link. In this work, we improve the detector performance in fast-varying channels by updating the reference CIRs in time. The detection threshold is also time-varying and is computed using the current set of reference CIRs and a predefined false alarm rate. Instead of using the minimum nMSD, we exploit information from all nMSDs and make the detection decision by computing the probability of potential target crossing (percentage of nMSDs higher than the detection threshold). The performance of the new detector is investigated in an UWA network with nine communication links. Results show clear indication of target detection and a significant reduction of false alarms, compared to the previous work.*

Keywords: *Channel Estimation, Target Detection, Underwater Acoustic Network*

1. INTRODUCTION

Underwater target detection is a key part of an underwater monitoring and surveillance system. It is known that the forward scattering from a target generally has a higher intensity level compared to the back scattering [1]. Prior works perform target detection by building up an acoustic barrier with vertical source and receiver arrays covering the water column [1–3]. In this work,

we consider target detection within an underwater acoustic (UWA) network of acoustic sensor nodes communicating with each other. The key idea is to detect the forward scattering from the moving target by re-using the communication signals in the network. For every data packet, measurement of the channel impulse response (CIR) is available at the receiver. The detector can exploit this information and perform target detection based on changes in the CIRs between the data packets. Potential target detection information will be delivered to a master station on shore for joint processing and/or target trajectory estimation. Network protocol design for simultaneous UWA communication and target detection can be found in [4].

In [5], we proposed a detection algorithm based on the variation of the CIRs. The normalized mean squared deviation (nMSD) between a reference CIR (measured when the target is absent) and the CIR of interest (measured with the current data packet) is used as the statistic for target detection. Before computing the nMSD, a filter with a small delay spread is applied to one of the CIRs to compensate for the change in delay and amplitude caused by the slow motion of the nodes and environmental factors. To further reduce the impact of channel variation on the detection performance in lake experiments, a set of reference CIRs measured outside of the experimental interval are used for computing the nMSDs. These reference CIRs should cover the uncertainty area caused by the channel variation due to the changing environment. Only the minimum nMSD among these measurements is used as the statistic for target detection. Both the channel variation and a moving target can cause high nMSD. Using the minimum nMSD allows us to reduce the influence of the channel variation on the nMSD performance.

In [5], a fixed set of reference CIRs are chosen at the beginning and the end of the lake experiment. This works for experiments of short duration under relatively stable channel conditions. For long-term surveillence and monitoring, it is not practical to use a fixed set of reference CIRs as it requires a very large dataset to cover the uncertainty in channel variation due to the time-varying environment over the whole deployment period. In this work, we propose to update the reference CIRs in time so that the detection system can track the changes of the channel state with a small set of reference CIRs. In addition, we set detection threshold based on a pre-defined false alarm rate using the reference CIRs. In such a case, for every new received data packet, we will have a new set of reference CIRs and a time-varying detection threshold for target detection. Furthermore, instead of using a single (minimum) nMSD for target detection, we propose to exploit all nMSDs in the reference set to compute the probability of a target crossing event.

The performance of the new detector is investigated and compared with the detector proposed in [5] in a lake experiment with three in-house developed transmitter nodes and four hydrophones. Results indicate a good detection performance against the ground truth with the proposed detector and a significant reduction of false alarms compared to the previous work.

2. TARGET DETECTION BASED ON CIR VARIATION

2.1. COMPUTATION OF THE NORMALIZED MSD

The MSD can be computed using the frequency-domain representation of the channel:

$$\mathrm{MSD} = \sum_{k=0}^{K-1} |H_0(k) - H(k)|^2, \tag{1}$$

where H_0, H are the frequency responses corresponding to the reference CIR h_0 and the CIR of interest h, respectively, and they are computed using the fast Fourier Transform (FFT) of a size K.

Instead of computing the MSD using (1), we introduce an adaptive filter g with a small delay spread to compensate for the change in delay and amplitude caused by the slow motion of the nodes and the variations of the environment [5]. The frequency response of g can be expressed as a combination of complex exponentials with expansion coefficients $c(p)$ [6]:

$$G(k) = \sum_{p=0}^{2P} c(p) e^{j2\pi(-P+p)k/K}, \ k = 0, \ldots, K-1, \tag{2}$$

where the number of complex exponentials $2P+1$ corresponds to the delay spread of the filter g. The filter g is found by solving the following optimization problem:

$$\hat{G} = \underset{G}{\operatorname{argmin}} \sum_{k=0}^{K-1} |H_0(k)G(k) - H(k)|^2. \tag{3}$$

Detailed procedure of finding $\hat{G}$ is described in [5]. After obtaining $\hat{G}$, the deviation in the frequency response is computed as: $\Delta H(k) = H_0(k)\hat{G}(k) - H(k)$. Finally, the nMSD is computed as:

$$\mathrm{MSD}_{\mathrm{norm}} = \frac{\|\Delta H(k)\|_2^2}{\|H_0(k)\|_2 \|H(k)\|_2}, \tag{4}$$

where $\| \cdot \|_2$ denotes the Euclidean norm.

2.2. Computation of the Detection threshold

To compute the detection threshold, the nMSD between each pair of reference CIRs in the dataset needs to be computed. Assuming that the number of reference CIRs is N, we will have $N(N-1)$ measurements of the nMSD.

The detection threshold Γ is initialized to a very small value (starting with a high false alarm rate) and iteratively increased until the false alarm rate is equal to or lower than a pre-defined value a. In every iteration, we compare the nMSDs with Γ and count the number of times (N_{fa}) when the nMSD is higher than Γ. The false alarm rate P_{fa} is computed as: $P_{\mathrm{fa}} = N_{\mathrm{fa}}/(N(N-1))$. If the false alarm rate is higher than a, the threshold is exponentially increased as: $\Gamma = (1+\alpha)\Gamma$, where α represents the growth rate. The above procedure is repeated recursively until the pre-defined false alarm rate is met.

The algorithm for the threshold computation is summarized in Algorithm 1, where $\mathrm{MSD}_{\mathrm{norm}}(i,j)$ is the nMSD between the ith and the jth reference CIR.

2.3. Target detection

The performance of channel estimation is a key for target detection. To ensure high accuracy of the channel estimates, we only use the CIRs from perfectly demodulated data packets. These CIRs are further selected based on the signal to noise ratio (SNR) in the received data packet; if SNR is lower than a threshold, the CIR is discarded.

The reference CIRs are initialized with N CIRs measured assuming there are no targets in the vicinity of the communication link. The set of reference CIRs are updated as follows. For the nth data packet, we include a CIR measured at the $(n-l)$th data packet into the reference CIRs and remove the oldest reference CIR. As the target can be detected when it is in close proximity to the communication link, introducing a delay of l data packets allows us to ensure the reference CIR is not affected by the target. After updating the set of reference CIRs, a

Algorithm 1 Detection threshold computation

Initialization: $\Gamma = \Gamma_{\text{init}}$, $N_{\text{fa}} = 0$, $P_{\text{fa}} = 1$
while $P_{\text{fa}} > a$ **do**
    $\Gamma = (1 + \alpha)\Gamma$ and $N_{\text{fa}} = 0$
    Update MSD_{norm}
    for $i = 1, \ldots, N$ **do**
        for $j = 1, \ldots, N$ **do**
            if $\text{MSD}_{\text{norm}}(i, j) > \Gamma$ **then**
                $N_{\text{fa}} = N_{\text{fa}} + 1$
            end if
        end for
    end for
    $P_{\text{fa}} = N_{\text{fa}}/(N(N-1))$
end while

detection threshold is computed based on Algorithm 1. As only one of the reference CIRs is updated, only one row and one column of the matrix MSD_{norm} need to be re-computed for every data packet.

Then, we compute the nMSDs between the current CIR and the reference CIRs and compare them with the detection threshold Γ. The propability of the target crossing can be computed as: $P_{\text{d}} = N_{\text{d}}/N$, where N_{d} is the number of times when the nMSD is higher than Γ. The detector will only indicate a target being detected when $P_{\text{d}} > 0.5$. The algorithm for target detection is summarized in Algorithm 2, where h_0^i represents the ith reference CIR and h represents the CIR measured with the current data packet.

Algorithm 2 Target detection algorithm

while A data packet being received **do**
    Initialization: $N_{\text{d}} = 0$
    Replace h_0^0 in the reference CIRs with h_{n-l}
    Compute the detection threshold Γ using Algorithm 1
    for $i = 1, \ldots, N$ **do**
        Compute $\text{MSD}_{\text{norm}}(i)$ between h_0^i and h according to subsection 2.1
        if $\text{MSD}_{\text{norm}}(i) > \Gamma$ **then**
            $N_{\text{d}} = N_{\text{d}} + 1$
        end if
    end for
    $P_{\text{d}} = N_{\text{d}}/N$
    if $P_{\text{d}} > 0.5$ **then**
        Target detected!
    end if
end while

3. LAKE EXPERIMENT

3.1. EXPERIMENT TOPOLOGY

Fig. 1 shows the topology of the lake experiment and the target trajectory and measurements of the sound velocity. In total, three transmitter nodes (Tx1 to Tx3) and four hydrophones (H1 to H4) are deployed. The signal from H3 is not used due to the recorder artifacts, thus we have nine communication links in the experiment. Depth of the nodes and distance of the acoustic links

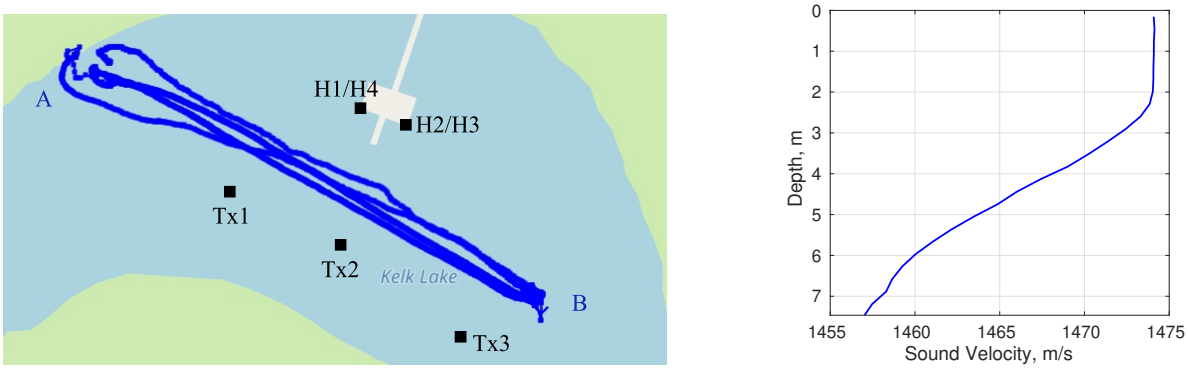


Figure 1: Topology of nodes and target trajectory (blue dots) in the lake experiment (left) and sound velocity measurements (right).

Table 1: Nodes depth and distance between nodes.

	H1/H4	H2	Tx1	Tx2	Tx3	Depth
Tx1	49 m	61 m		39 m	91 m	2.5 m
Tx2	46 m	46 m	39 m		53 m	2.5 m
Tx3	88 m	78 m	91 m	53 m		3.9 m
H1		17 m	49 m	46 m	88 m	1.8 m
H2	17 m		61 m	46 m	78 m	1.6 m
H4			61 m	46 m	78 m	5.8 m

are summarized in Table 1. The target used in this experiment is a 1 m $\times$ 0.9 m wooden board with gravels on both sides. The board is deployed at approximately 1.5 m depth. A buoy with a GPS recorder is attached to the board to record the target trajectory. During the experiment, the board crossed the network area four times between point A and point B. Initially, the board is towed by a canoe with a 7 m long rope when travelling from point A to point B. The board is then released from the canoe and pulled back to point A with a fishing line. This process was repeated twice within the experiment.

3.2. Data packet transmission

During the experiment, we have three unsynchronized transmitter nodes transmitting data packets at 32 kHz carrier frequency with 6 kHz frequency bandwidth. The duration of each data packet is 100 ms. The transmitter source level is 155 dB re 1μPa @1m. Each user transmits three data packets every ten seconds. As shown in Fig. 2, the gap between data packet transmission is 0.4 s, 1.2 s and 2.8 s for Tx1, Tx2 and Tx3, respectively. In each data packet, 200 bits of uncoded data are transmitted, where the last 16 bits are cyclic redundancy check (CRC) code.

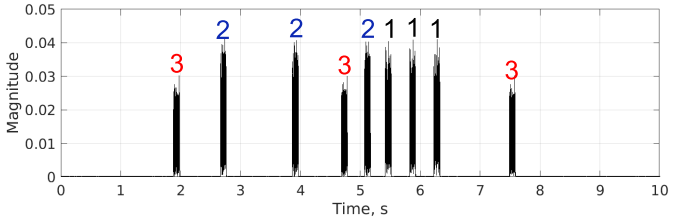


Figure 2: Ten seconds of baseband received data packets from Tx1, Tx2 and Tx3 (denoted as 1, 2 and 3, respectively).

To address the possible overlap in the received data packets from the unsynchronized transmitter

nodes, a multiuser UWA receiver with interference cancellation [7] is used to demodulate the data packets from three users and to provide channel estimates for target detection. After a data packet detection, the received signal segment is downsampled to twice the baseband sampling rate and de-multiplexed into two branches for channel estimation and equalization. The final CIR used for target detection is obtained by combining (multiplexing) the channel estimates from the two branches (with 1/(12 kHz) $\approx$ 0.08 ms delay resolution).

3.3. TARGET DETECTION PERFORMANCE

In this section, we investigate and compare the target detection performance of the new detector and the previous design in [5].

A fixed set of reference CIRs is used for target detection in [5]. For each acoustic link, 200 reference CIRs are measured when the target is far away from the link are used. Method for threshold computation was not proposed in [5]. Here we compute the threshold in a similar way as described in Algorithm 1. For every reference CIR, we compute the nMSDs with the rest of the reference CIRs, excluding two adjacent CIRs, which are supposed to be highly correlated with the current CIR, and use the minimum nMSD for target detection. The $(N-1)\times N$ nMSD matrix in Algorithm 1 is reduced to an $(N-3)\times 1$ column vector (removing the current reference CIR and two neighbouring ones). The acceptable false alarm rate is set as $a = 0.5\%$.

In Fig. 3, data packets with nMSD higher than the detection threshold are shown in blue stems. These potential crossings are validated with the ground truth of the target trajectory. If the target position is within 10 m distance from the crossing point indicated by the GPS, it is considered as a positive detection. If not, it is marked as false alarm by a black cross. All positive detections are denoted with a number indicating the associated crossing event (from 1 to 4).

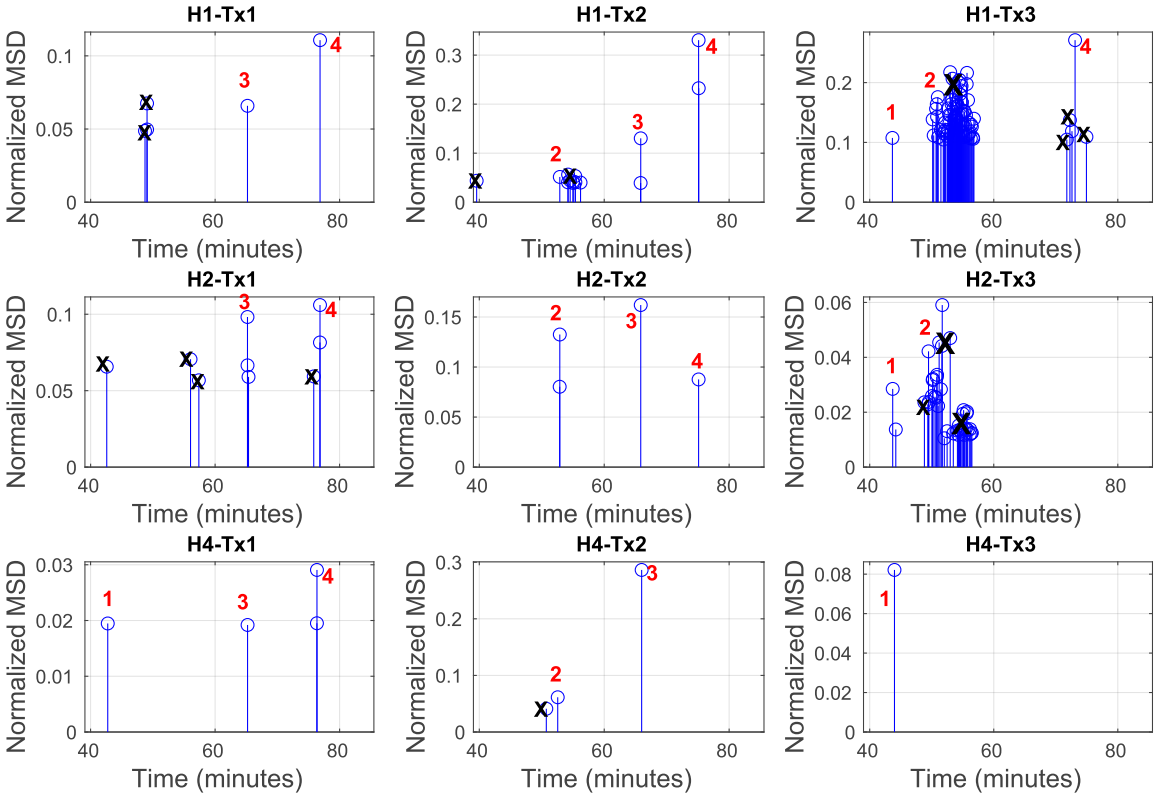
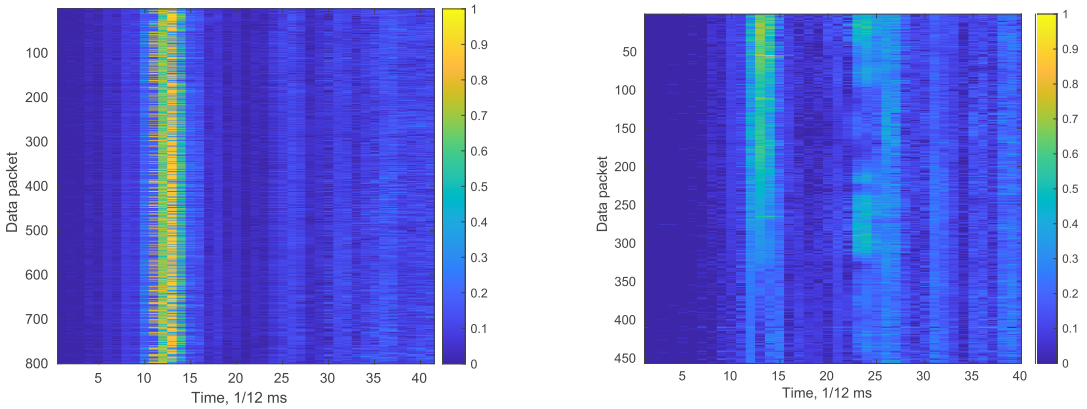


Figure 3: Normalised MSD of potential target crossings indicated by the detector in [5].

It can be seen from Fig. 3 that the detector can provide reasonable target detection performance for most of the acoustic links; at least four positive detections are reported in the network for



(a) CIRs between H1 and Tx1 .

(b) CIRs between H1 and Tx3.

Figure 4: Channel impulse responses of perfectly demodulated data packets between (a) H1 and Tx1 and (b) H1 and Tx3.

each target movement between A and B (see Fig. 1). However, there are many false alarms among these detection, especially for links H1-Tx3 and H2-Tx3. In Fig. 4, we show the CIRs obtained for the link between H1 and Tx1 and the link between H1 and Tx3, respectively. It can be seen that the multipath structure in Fig. 4(a) is relatively static, on the other hand, significant change in the CIRs can be observed in Fig. 4(b). This explains the sudden change in the nMSD shown in Fig. 3, which results in a large number of false alarms with a fixed detection threshold.

The sudden change in the channel state can be resolved by the new detector as both the reference CIRs and the associated detection threshold are changing in time. The number of reference CIRs used in the new detector is 50. For every data packet, the oldest reference CIR is replaced by a CIR measurement delayed by three data packets (corresponds to at least 10 second based on the transmission pattern) to ensure the reference CIR is measured when the target is not crossing the communication link.

Potential target crossing events in all acoustic links are shown with blue stems in Fig. 5, among which false alarms are marked by black cross symbols. Multiple positive detections are reported in the UWA network for each target movement between A and B. In addition, the number of false alarms is significantly reduced for the link between H1 and Tx3 and the link between H2 and Tx3. This demonstrates the effectiveness of the new detection algorithm.

4. Conclusions and Future Work

In this work, a new detector is proposed for target detection in an UWA communication network. The target detection is performed based on computing the nMSD between a CIR and a set of reference CIRs. To improve the detector performance in fast-varying channels, both the reference CIRs and the detection threshold are updated in time. The probability of a target crossing is then computed as the percentage of nMSDs higher than the detection threshold. If the probability is higher than 50%, the detector decides that there is a target crossing. The performance of the proposed detector is investigated and compared with our previous detector in a lake experiment with nine acoustic communication links. Results indicate a high detection performance has been achieved with less false alarms compared to the previous design.

The performance of the proposed detector will be further validated with sea experimental data. The next step of the research is to propose an algorithm for target trajectory estimation and prediction based on the detection information from all acoustic links in the UWA network.

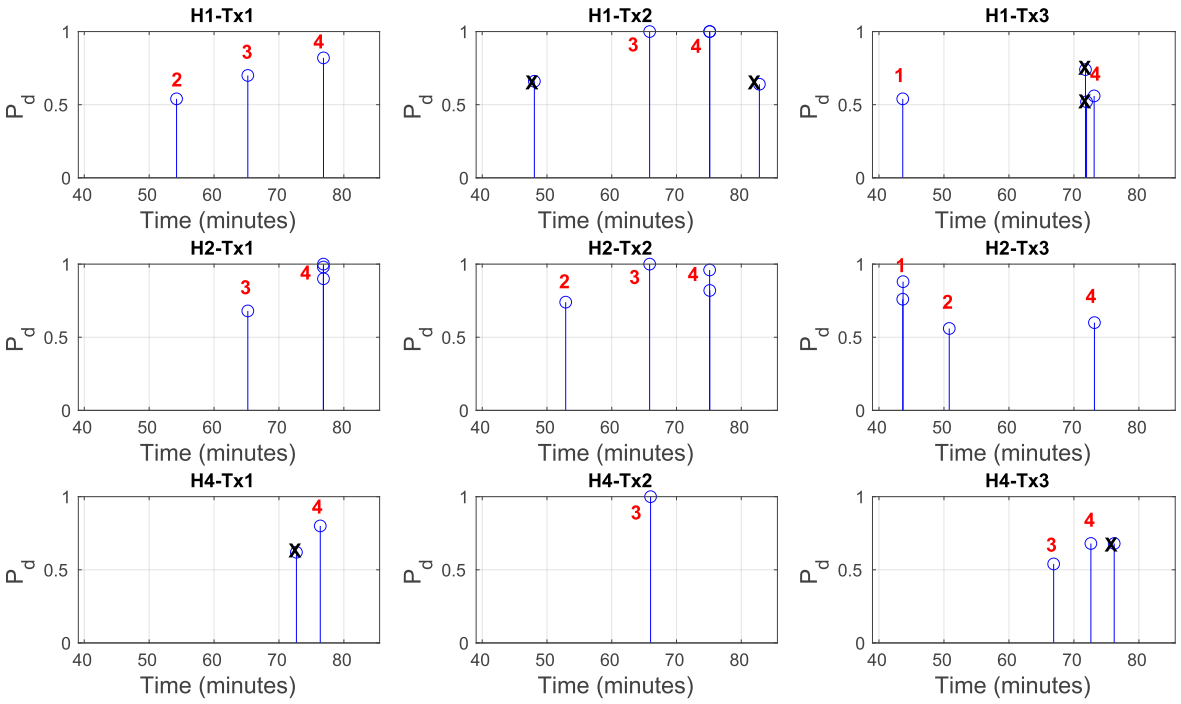


Figure 5: Probability of potential target crossing computed by the proposed detector.

REFERENCES

[1] Karim G Sabra, Stephane Conti, Philippe Roux, Tuncay Akal, William A Kuperman, J Mark Stevenson, Alessandra Tesei, and Piero Guerrini. Experimental demonstration of a high-frequency forward scattering acoustic barrier in a dynamic coastal environment. *The Journal of the Acoustical Society of America*, 127(6):3430–3439, 2010.

[2] Thomas Folegot, Giovanna Martinelli, Piero Guerrini, and J Mark Stevenson. An active acoustic tripwire for simultaneous detection and localization of multiple underwater intruders. *The Journal of the Acoustical Society of America*, 124(5):2852–2860, 2008.

[3] Bo Lei, Yixin Yang, Kunde Yang, and Yuanliang Ma. Detection of forward scattering from an intruder in a dynamic littoral environment. *The Journal of the Acoustical Society of America*, 141(3):1704–1710, 2017.

[4] Nils Morozs, Benoît Parrein, Lu Shen, Benjamin T Henson, and Paul D Mitchell. Network protocols for simultaneous underwater acoustic communication and target detection. In *Proceedings of the Underwater Acoustics Conference and Exhibition (UACE'23), Kalamata, Greece*, 2023.

[5] Lu Shen, Yuriy Zakharov, Benjamin Henson, Nils Morozs, Benoît Parrein, and Paul D Mitchell. Target detection using underwater acoustic communication links. *IEEE Journal of Oceanic Engineering*, 2024.

[6] Hua Yu, Aijun Song, Mohsen Badiey, Fangjiong Chen, and Fei Ji. Iterative estimation of doubly selective underwater acoustic channel using basis expansion models. *Ad Hoc Networks*, 34:52–61, 2015.

[7] Yuriy Zakharov, Lu Shen, Benjamin Henson, Nils Morozs, and Paul D Mitchell. Interference cancellation for UWA random access data packet transmission. *arXiv preprint arXiv:2401.10021, accepted by IEEE Journal of Oceanic Engineering*, 2024.

Structured Session
22. Underwater noise in the EU Marine Strategy Framework Directive: implementation, monitoring, assessment and measures

Organised by Peter Sigray, Aristides Prospathopoulos and Emmanuel Skarsoulis

Long-term measurements of ambient sounds in Cambridge Bay (Canada), 2015–2024 – Implications for extending the MSFD to Arctic waters

Philippe Blondel[1], Rhys Belcher[1] and Dylan Cooper[1]

[1]University of Bath; Claverton Down, Bath BA2 7AY, UK

Philippe Blondel, Department of Physics, University of Bath, Claverton Down, Bath BA2 7AY, UK: pyspb@bath.ac.uk.

Abstract: *Climate change in the Arctic enables increased access to human activities, affecting underwater soundscapes. It is therefore important to have complete guidelines to monitor impacts on natural environments. The EU Marine Strategy Framework Directive is the most complete, strongly inspiring emerging guidelines in other countries. Primary descriptor D11C2 addresses continuous low-frequency sounds and makes extensive use of third-octave "shipping bands" at 63 and 125 Hz. To address the lack of measurements, models often use ship tracks recorded by their Automatic Identification Systems (AIS). But not all ships in the Arctic use AIS, and winter ice also allows human activities other than shipping. We use sound measurements by Ocean Networks Canada in Cambridge Bay (Nunavut) between 2015 and 2024, focusing on the months of May (full ice cover, no shipping) and August (little to no ice, shipping activity). We show impacts beyond the "shipping bands". Baseline soundscapes vary with ice cover and AIS underestimates impactful activities of all types. Our results show that future guidelines will need adapting to the Arctic environments to fully measure the range of human impacts.*

Keywords: *Marine Strategy Framework Directive, human impacts, Arctic, shipping*

1. RATIONALE

The Arctic regions have long been shielded from human activities, because of their remoteness, the challenging weather conditions and the predominant ice cover. This is all changing now and changing fast. Arctic amplification means polar regions are warming three times faster than the global average [1]. Between 1979 and 2019, the Arctic sea ice declined by up to 43% in all regions, with sea ice becoming younger and thinner over this period [2, 3]. Economic developments and geostrategic considerations are leading to the development of offshore structures, in particular near-shore extraction facilities, expected to develop, in particular in Nunavut (Canada), Russia and Greenland and to be associated with marine shipping of the resources. The gradual openings of the Northern Sea Route and the Northwest Passage are now associated with the distinct possibility of trans-Arctic routes in the near future [4]. Analyses of current shipping patterns show that the Northwest Passage in particular can decrease voyages by 7,000 km [5], The distances travelled by ships along the Northwest Passage already tripled between 1990 and 2015 [6] and the southern route is often preferred because of its relative shelter. Cambridge Bay (Ekaluktutiak) in Nunavut is the largest stop for passenger and research vessels along this route (Fig. 1), even though the number of ships passing in the outer Dease Strait is still relatively low [7]. It is therefore a good location to assess these changes over the years.

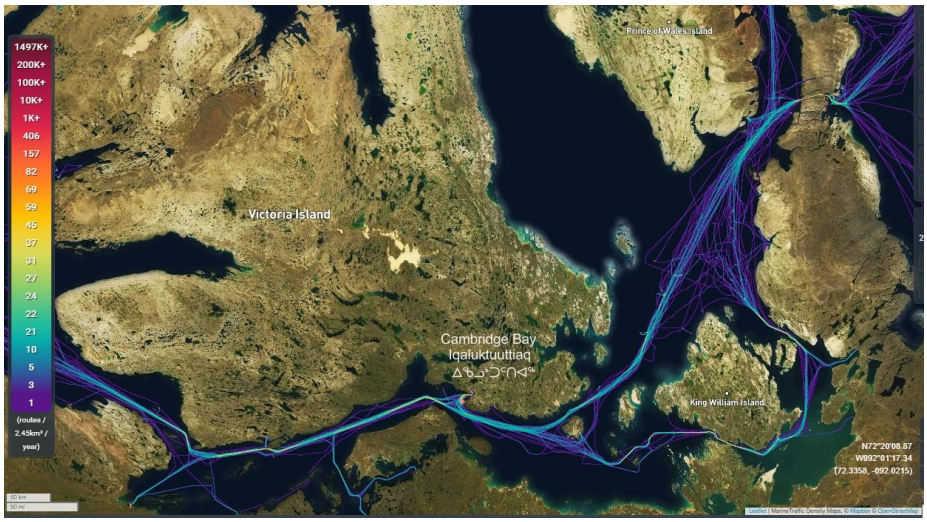


Fig.1: Cambridge Bay is on the southern route of the Northwest Passage. The most frequented shipping lanes show few vessels equipped with Automatic Identification Systems (2024 basemap from Marine Traffic, https://www.marinetraffic.com/).

The EU "Marine Strategy Framework Directive" (MSFD) is an important tool in the sustainable use of the oceans, providing a comprehensive series of metrics and contributing to addressing the UN Sustainable Development Goal SDG-14 "Life below water" (https://www.un.org/sustainabledevelopment/oceans/). It was augmented in 2022 by additional guidelines about setting threshold values for continuous underwater sounds [8], accounting for local conditions, acoustic baselines and marine life. The impact of human activities is measured along "shipping bands", third-octave frequency bands centred on 63 Hz and 125 Hz. This definition has been particularly useful to assess the impacts of marine traffic in different regions around Europe [9, 10] and in other oceans [11], including modelling the deep waters of the Arctic Ocean [12].

In the context of the shallow-water environment of Cambridge Bay, close to a harbour and town, ice-bound during winters, this article aims to address the following questions:

- What are the actual measurements of the "shipping bands"? How do they vary between seasons and between years?
- Are there other human activities that contribute to these bands? How do they compare with satellite monitoring of shipping activities?
- Beyond the MSFD "shipping bands", should other frequency bands be used to assess human impacts?

2. DATA AND METHOD

Ocean Networks Canada has been operating the Cambridge Bay Community Observatory for more than 10 years [13]. Soundscapes are measured continuously using an Ocean Sonics icListen HF hydrophone 8 m deep in shallow water (13 m). We use recordings from May and August, from 2015 to 2024. They are processed with PAMGuide [14], using the hydrophone sensitivity of -170 dB ref. 1 V/µPa and the frequency range of 10 Hz to 32 kHz. Time windows of 1 second, with a Hann filter and 50% overlap, were used to calculate broadband sound pressure levels (SPLs), power spectral densities (PSDs) and third-octave band levels (TOLs), aggregated every 5 minutes. There were occasional gaps in data coverage, due to hardware issues and the effects of COVID lockdowns on recovering data. Loud, continuous events are identified by their SPL exceeding the weekly average SPL by at least 10 dB for at least 1 minute (filtering out most ice-related processes, keeping most human-made sounds like shipping).

Regional ice information comes from weekly ice charts provided by the Canadian Ice Service, showing ice cover and ice types in Cambridge Bay and surrounding areas. These charts are based on analyses and combinations of satellite imagery, weather and oceanographic information and visual observations from ship and aircraft. Local ice thickness is provided by an ice profiler, operated by Ocean Networks Canada and located next to the hydrophone. Wind speeds and air temperatures were measured hourly at neighbouring weather station Cambridge Bay A, currently operated by NAV Canada. The average sound levels in each third-octave band were correlated with hourly wind speeds, in line with similar calculations by [15], as an indicator of Sea State. Hourly temperatures and their first derivatives were used to assess possible changes in ice dynamics.

Ship traffic information is coming from Automatic Identification Systems (AIS) collated by the Arctic Ship Traffic Database (ASTD) of the Arctic Council's working group on Protection of the Arctic Marine Environment (PAME). ASTD Level 3 data includes ship positions (every 6 minutes), ship type (e.g. cruise and passenger ships, general cargo, offshore supply and chemical tankers) and activity. The numbers of AIS entries for each month show the activity in Cambridge Bay and further out, even if the number of unique ships is small and some might be re-entering the area several times.

These strands of information have been used to focus on the contrasting months of May (maximum ice cover, no shipping) and August (smallest to no ice cover, higher shipping). An earlier study [16] had aurally identified different sounds (other than shipping) in May and August 2018, In later work [17], we had analysed the sound levels in the "shipping bands" of 63 Hz and 125 Hz for the months of May and August in 2015, 2016, 2018, 2019 and 2020. Here, we focus on both loud and continuous sounds, for a complete decade (2015 to 2024), with the added information of AIS to identify some ship types/activities but also to ascertain how much marine traffic is actually recorded with AIS (or not).

3. RESULTS

For the first four weeks of each month (to ensure statistical equivalence) and for all times where data was available, the variations of the third-octave levels for the 63-Hz "shipping band" (56.2–70.8 Hz) and its 125-Hz counterpart (112–141 Hz) are compared with those of the broadband (10 Hz – 32 kHz) Sound Pressure Levels between 2015 and 2024 (Fig. 2). The boxplots show the median values as central marks, and the 25th and 75th percentiles correspond to the edges of the boxes centred on the median values. Individual outliers are represented as crosses.

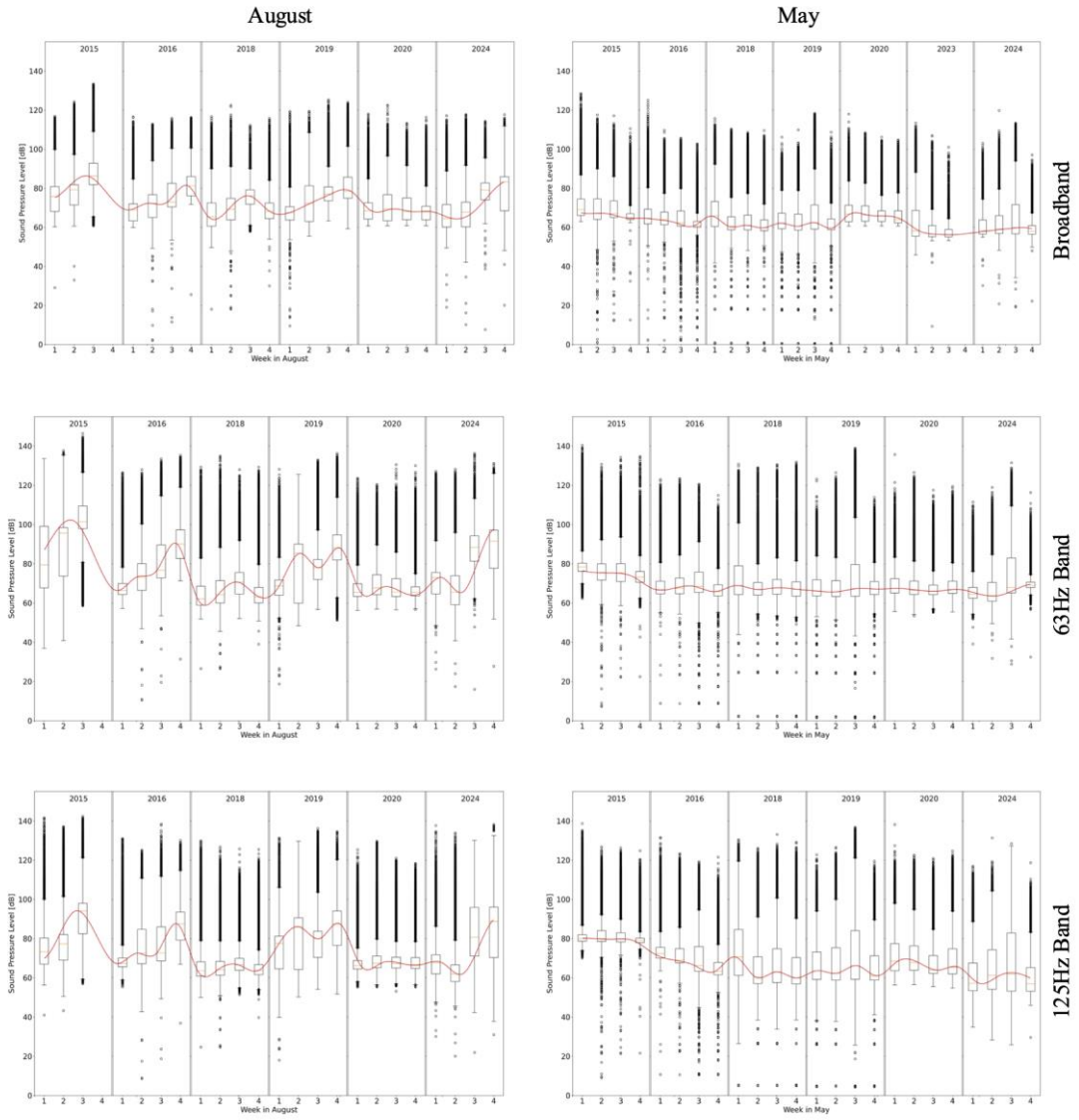


Fig.2: Boxplots showing the variation of Sound Pressure Levels, broadband (10 Hz – 32 kHz, top) and for the third-octave bands of 63 Hz (middle) and 125 (bottom) recommended by the MSFD, for the first 4 weeks of the months of August (left) and May (right), from 2015 to 2024. Months with not enough measurements are not represented. The red lines go through the mean values of each plot.

Broadband levels (Fig. 1, top) are generally louder by more than 10 dB in the summer months, with louder outliers present more than 10 times than the quieter ones. From week

to week, the average levels generally increase as the month progresses and navigation becomes easier. Values for August 2020 are affected by the Covid lockdown, decreasing to levels generally seen in the months of May. Broadband sound levels in the months of May are generally similar and much lower than in August, with much less variation from year to year and from week to week.

In August (Fig. 2, left), the two shipping bands (Fig.2, middle and bottom) follow each other and the broadband SPLs. They have respective Pearson's correlation coefficients r of 0.87 (63-Hz band) and 0.90 (125-Hz band), with p values of 1.1×10^{-8} and 2.3×10^{-10} respectively. These very strong correlations show the main contributions to the broadband levels are indeed coming from these two bands. However, comparisons with AIS records do not correlate with the numbers of ships each week (or the time they spend in the area) with the variations in either shipping band.

In May (Fig. 2, right), the two shipping bands also show levels varying in line with the broadband levels, generally at similar levels to August averages and decreasing with the years. The quieter outliers (broadband and in the shipping bands) also drastically decrease in numbers from 2020 onwards. Values in the 63-Hz band have a lower correlation with the broadband levels ($r = 0.68$ and $p = 1.5 \times 10^{-5}$). Conversely, values in the 125-Hz band have correlations with broadband levels similar to those from the months of August ($r = 0.88$ and $p = 1.9 \times 10^{-11}$). Ice charts (and local measurements) show thick ice cover and AIS records confirm there is no shipping at these times.

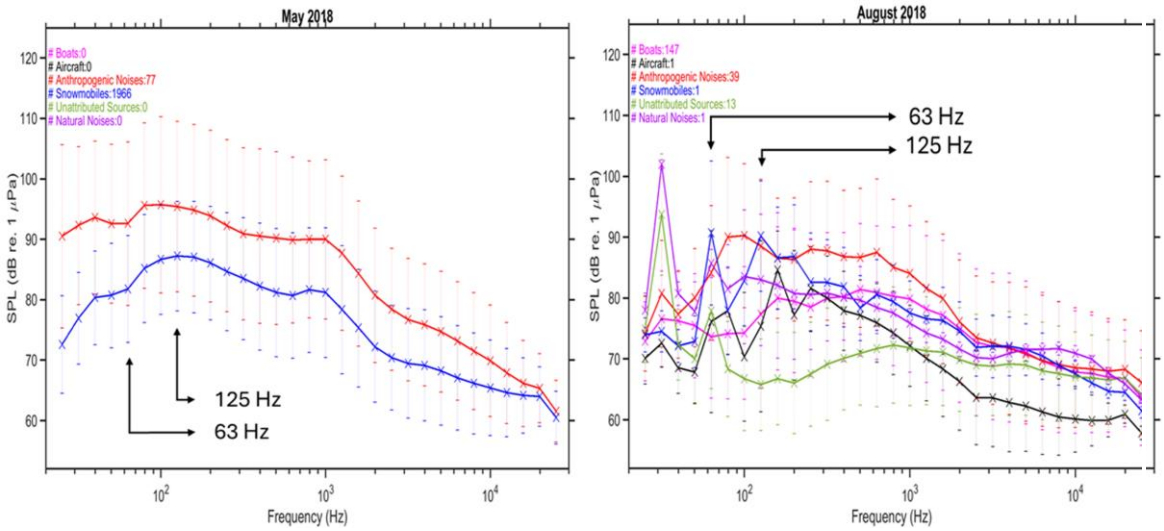


Fig.3: Aural classification of long, continuous sounds for the months of May and August 2018 (from [16]). They show that more ships are heard than are tracked with AIS and that other sound sources contribute to the MSFD bands.

Aural identification of recordings in 2018 showed there were many different sounds other than ships (Fig. 3; [16]), including aircraft (in summer) and snowmobiles (in winter). Aircraft sounds are similar to those presented in [18] and snowmobile sounds are similar to those presented in [19]. Other sounds were more difficult to identify with certainty and labelled "anthropogenic": they include machinery (onshore or on ships), sounds from repairs ("drilling" sounds) etc. Few ships are recorded using AIS, presumably because they are too small to legally need one, but hundreds of ships can still be heard in the data during the months of August, some for hours as they enter or leave Cambridge Bay [17] and the noise of machinery/lifting gear is also common. These ships create sounds generally at 125 Hz and below (including below 63 Hz) and extending into higher frequencies for shorter periods. Other sounds identified aurally as anthropogenic extend up to 700 Hz. Small vessels idling in the harbour are also a common occurrence. Because of

the proximity to the Cambridge Bay Airport (2 km from the hydrophone) and the Cambridge Bay Water Aerodrome (less than 2 km away), aircraft can be heard too. The sounds primarily come from the propellers, with harmonics between ca. 80 Hz and 1 kHz, and PSD levels up to 100 dB. Less aircraft can be heard than expected from flight schedules, presumably based on variations in approach paths and aircraft types. In the months of May, no ships were heard (obviously) but there were a large number of snowmobiles and possibly All-Terrain Vehicles. Cambridge Bay is also famed for its Omingmak Frolics, a week-long spring festival featuring snowmobile races (and usually held in May). Snowmobiles create loud tonals (below a few hundred Hz), with PSD levels up to 80 dB between 40 Hz and 1 kHz, extending into higher frequencies as they pass above the hydrophone (as evidenced by Lloyd's Mirror Effects).

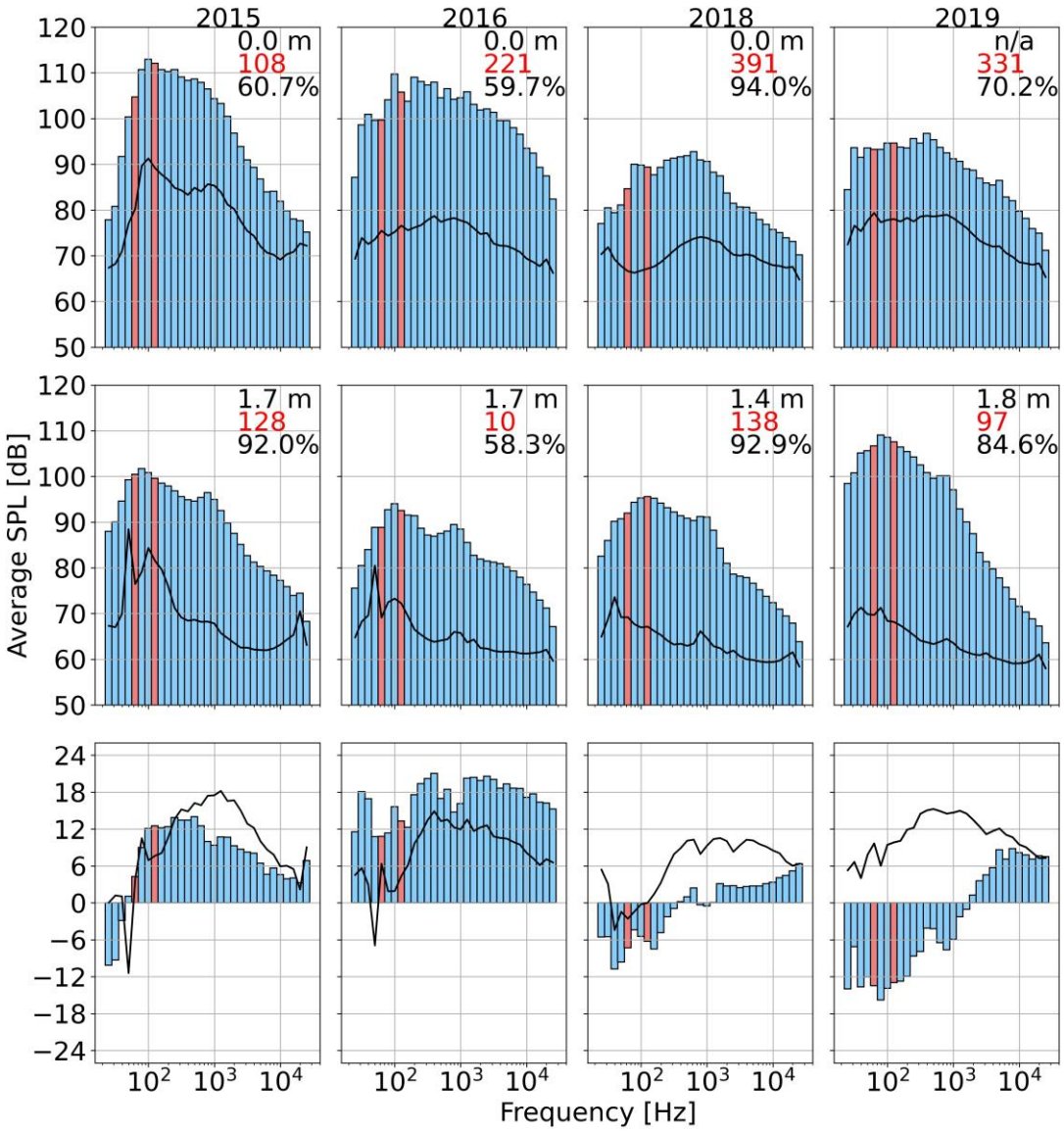


Fig.4: Frequency contributions of loud, continuous events for August (top) and May (middle) and their difference (bottom), with averages as black lines. MSFD bands are highlighted in red. Numbers show the ice draft, the number of continuous events identified and the record coverage for this month. For clarity, only some years are presented here.

Loud, continuous events extend well into the kHz range (Fig., 4) and the MSFD bands are not always the loudest. Sound levels in August are systematically louder in the 125-Hz

band than in the 63-Hz band. This matches measurements by [10] in a similar bay, attributing it to the increased presence of smaller vessels (with no AIS) and the effects of acoustic propagation. Individual recordings of ships show their sounds often extend into the kHz range, especially when audibly closer. Other studies showed similar results, with sounds from small vessels centred on 200 Hz [20], louder between 350 Hz and 2 kHz [21] and extending up to 10 kHz [22], depending on the types of ships and their exact environments. Conversely, in May, the MSFD bands are loud (even if there are no ships) but the frequency distributions of loud events tail off steeply above 1 kHz. Sound levels are also different between August (louder in 2015+2016) and May (louder in 2018+2019).

4. CONCLUSIONS

Actual measurements of the MSFD "shipping bands" show they are not always louder in summer months (when there is shipping) than in winter months (when there is no shipping because of the full ice cover). Comparisons between these seasons show there are other contributors to the MSFD bands (and other frequencies), including small vessels (with no AIS), overflying aircraft, snowmobiles and ATVs (in winter) and machinery noise (in both seasons). Satellite monitoring of shipping is still helpful, but in coastal environments like this one, they do not provide a full picture of the varied activities taking place at different seasons. Models of acoustic impacts based solely on AIS data are consequently also of limited use for this type of area. Focusing on loud, continuous events, it is also clear that they contribute to frequency bands well into the kHz region (in the open-water season) and up to 1 kHz (in the fully ice-covered season). These measurements can be extended to include different levels of ice cover (e.g. during melting or break-up), related to different types of human impacts (more or less shipping, more or less on-ice activities like snowmobiles). Extensions of the MSFD to Arctic waters should therefore use different baselines depending on the ice cover. They should also recognise that the 63-Hz and 125-Hz bands are not the most representative of human impacts, from shipping or other activities, and that frequency bands up to the kHz region should be monitored.

5. ACKNOWLEDGEMENTS

We are grateful to Ocean Networks Canada, NAV Canada and the Canadian Ice Service for their open-access provision of data. Access to the Arctic Ship Traffic Database was funded by the Department for Science, Innovation and Technology, as part of the United Kingdom – Arctic Council Working Groups – Research and Engagement Scheme 2024/25, working with the Arctic Council Working Groups, Norwegian Ministry of Foreign Affairs and the NERC Arctic Office.

REFERENCES

[1] **Zhou, W., L.R. Leung, J. Lu**. Steady threefold Arctic amplification of externally forced warming masked by natural variability. *Nat. Geosci.*, 17, 508–515, 2024,
[2] **Wadhams, P.**, *Ice in the Ocean*, CRC Press, 308 pp., 2000.
[3] **AMAP**, Arctic Climate Change Update 2021: Key Trends and Impacts. Summary for Policymakers**,** *AMAP* Tromsø, Norway. 16 pp., 2021.

[4] **Smith, L.C., S.R. Stephenson**, New Trans-Arctic shipping routes navigable by mid-century, *PNAS*, 110(13), E1191-E1195, 2013.
[5] **Herrmann, T.**, "Shipping Through the Northwest Passage: A Policy Brief", https://jsis.washington.edu/news/shipping-through-the-northwest-passage-a-policy-brief/, 2019 [Last accessed 20/05/2025]
[6] **Dawson, J. et al.,** Temporal and spatial patterns of ship traffic in the Canadian Arctic from 1990 to 2015, *Arctic*, 71(1), 15–26, 2018.
[7] **Halliday, W.D.**, Underwater noise from ship traffic near Cambridge Bay, Nunavut in 2017 and 2018, *Transport Canada. Wildlife Cons, Society, Canada*, 30 pp., 2021.
[8] **TG Noise**, Setting of EU Threshold Values for continuous underwater sound – MSFD Common Implementation Strategy, https://circabc.europa.eu/ui/welcome, 6 pp. 2022.
[9] **Merchant, N. D., Ph. Blondel, D.T. Dakin, J. Dorocicz**, Averaging underwater noise levels for environmental assessment of shipping. *JASA*, 132(4), EL343-EL349, 2012.
[10] **Garrett J.K., et al..**, Long-term underwater sound measurements in the shipping noise indicator bands 63 Hz and 125 Hz from the port of Falmouth Bay, UK, *Mar. Pollut. Bull.*, 110:438-448, 2016.
[11] **DFO**, Evaluation of a Proposed Approach for Offsetting Increases in Underwater Noise from Marine Shipping, Using Information on Southern Resident Killer Whales. *DFO Can. Sci. Advis. Sec*. Sci. Advis. Rep. 2025/001. 2025
[12] **Heaney K.D.et al.**, Modeled underwater sound levels in the Pan-Arctic due to increased shipping: Analysis from 2013 to 2019. *JASA*, 155 (1):707–721, 2024.
[13] **ONC Data Archive**, http://www.oceannetworks.ca, hydrophone data from 01 May 2015 to 31 August 2024 (May and August months only), Oceans Networks Canada, University of Victoria, Canada. Downloaded 2024.
[14] **Merchant, N.D., K.M. Fristrup, M.P. Johnson, P.L. Tyack, M.J. Witt, Ph. Blondel, S.E. Parks**, Measuring acoustic habitats, *Meth. Ecol. Evol.*, 6:25-265, 2015.
[15] **Bonnel, J., G.B. Kinda, D.P. Zitterbart,** Low-frequency ocean ambient noise on the Chukchi Shelf in the changing Arctic, *JASA*, 149 (6): 4061–4072, 2021
[16] **Blondel, Ph., F. Bichan, L. Lewry, H. Hallett, G. McCarthy,** Acoustic properties of Arctic sea ice, from year-long underwater measurements in Cambridge Bay, Canada, in *International Conference on Underwater Acoustics (ICUA)*, 2022.
[17] **Blondel, Ph., J. Davidge**, Shipping and ice-covered waters: how accurate are the MSFD "shipping bands"?, in *Underwater Acoustics Conference & Exhibition (UACE)*, M. Taroudakis (ed.), pp. 587-594, 2023
[18] **Buckingham, M.J., et al.**, Propeller noise from a light aircraft for low-frequency measurements of the speed of sound in a marine sediment, *J. Comp. Acoust.*, 10, pp. 445-464, 2002.
[19] **Cook, E., J. et al.,** Determining the speed-dependent source level of a snowmobile traveling on sea-ice, *JASA*, vol. 152, A72, 2022
[20] **Picciulin, M. et al.,** Are the 1/3-Octave Band 63- and 125-Hz Noise Levels Predictive of Vessel Activity? The Case in the Cres–Lošinj Archipelago (Northern Adriatic Sea, Croatia), in *The Effects of Noise on Aquatic Life*, A. N. Popper and A. Hawkins (eds), 2016.
[21] **Rako N., et al.**, Leisure boating noise as a trigger for the displacement of the bottlenose dolphins of the Cres-Lošinj archipelago (northern Adriatic Sea, Croatia), *Mar. Pollut. Bull.*, 68:77–84, 2013.
[22] **Erbe, C.**, Modeling Cumulative Sound Exposure Over Large Areas, Multiple Sources, and Long Durations, in *The Effects of Noise on Aquatic Life*, A.N. Popper and A. Hawkins (eds), 2012.

A BEM model for underwater acoustic radiation due to vibrating vehicle in the sea waveguide

John Prospathopoulos[1] Kostas Belibassakis[2],

[1] School of Mechanical Engineering, National Technical University of Athens (NTUA), Heroon Polytechniou 9, 15780 Athens, Greece
[2] School of Naval Architecture & Marine Engineering, National Technical University of Athens (NTUA), Heroon Polytechniou 9, 15780 Athens, Greece

Contact author: John Prospathopoulos, Tel.number: (+30)2107721048,
e-mail: jprosp@fluid.mech.ntua.gr

Abstract: *Underwater acoustic radiation behaviour of a flexible sea vehicle in finite water depth is investigated and numerical predictions are presented. To calculate the acoustic radiation field from vibrating ships and marine structures in the sea-acoustic waveguide representations are developed based on modal expansion of the solution of the Helmholtz equation, satisfying the free surface and the seabed boundary conditions. The eigenmodes and eigenfrequencies of the marine structure used in the above representation are calculated by finite element (FEM) solutions. Effects of the free surface and finite water depth are introduced by using a 3D Boundary Element Method (BEM) model, based on the ocean waveguide Green's function obtained by the multiple image-source method. The present model aims to the development and optimization of a computational model for vibro-acoustic calculation associated with the underwater noise radiated from ship and underwater vehicles (URN). Along with measurements, accurate URN modelling will enable the design of quieter vessels, thus contributing to the compliance with the threshold values for continuous noise developed under the European Commission's Marine Strategy Framework Directive (MSFD).*

Keywords: *Underwater noise radiation, structural vibration and noise, ocean acoustic waveguide, 3D-BEM modelling.*

1. INTRODUCTION

Reducing Underwater Radiated Noise (URN) generated by waterborne transport (shipping) is crucial to mitigate its impact on marine ecosystems. IMO [1] has issued guidelines aiming at the reduction of underwater noise from commercial shipping to address adverse impacts on marine life. Noise is generated by three ship components (machinery, propeller, hull) and the associated URN effects are mainly due to engine noise and vibration, propeller cavitation and fluid hull structure interaction and vibration [2].

Machinery noise excites vibrations of the hull structure resulting in radiation of low frequency noise into the sea environment. Underwater acoustic radiation induced by structural vibration and propagation in the ocean environment involves analyzing the interaction between the acoustic field and the structure. In this work, a hybrid scheme is implemented, integrating the Finite Element Method (FEM) [3] for the determination of structural modes of vibration, with the Boundary Element Method (BEM) for the scattering and propagation of vibrational noise into the sea acoustic waveguide [4].

First, the hydroelastic problem is formulated in the frequency domain for the underwater structure of a flexible vessel in the absence of external forces and damping. The eigenmodes and eigenfrequencies of the structure are calculated using FEM analysis. Then, the hydroacoustic problem concerning the acoustic radiation from vibrating marine structures in the sea acoustic waveguide is formulated using the appropriate Green's function, obtained by the multiple image-source method, along with the satisfaction of the free surface and the seabed boundary conditions. The 3D scattering problem is solved using a BEM model forced by boundary data calculated by the eigenmodes of the hydroelastic problem through the boundary condition on the vibrating hull surface. Results of the predicted acoustic field are presented for two vessels, a travelling ship and an Autonomous Underwater Vehicle (AUV).

2. PREDICTION HYDROELASTIC-HYDROACOUSTIC MODELS

2.1 The hydroelastic model

The equation of motion for the underwater structure is the basis of the hydroelastic model:

$$\mathbf{M}\mathbf{U}''(t)+\mathbf{C}\mathbf{U}'(t)+\mathbf{K}\mathbf{U}(t)=\mathbf{F}_a(t)+\mathbf{F}_b(t), \tag{1}$$

in which t is the time instant; $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ denote the mass matrix, damping matrix and stiffness matrix; $\mathbf{U}''(t)$, $\mathbf{U}'(t)$ and $\mathbf{U}(t)$ represent the acceleration, velocity and displacement vector; $\mathbf{F}_a(t)$, $\mathbf{F}_b(t)$ are the mechanical force vector and the hydrodynamic force vector respectively. Assuming time-harmonic vibrations, the following representation is used:

$$\mathbf{U}(t)=\mathrm{Re}\left\{U\exp(i\omega t)\right\},\quad \mathbf{F}_a(t)=\mathrm{Re}\left\{F_a\exp(i\omega t)\right\},\quad \mathbf{F}_b(t)=\mathrm{Re}\left\{F_b\exp(i\omega t)\right\}, \tag{2}$$

where ω is the angular frequency, $i^2 = -1$, $\mathrm{Re}\{\cdot\}$ denotes the real part of a quantity, and U, F_a and F_b represent the displacement vector, mechanical force vector and hydrodynamic force vector in the frequency domain. Therefore, Eq. (1)

$$-\omega^2 \mathbf{M}U + i\omega \mathbf{C}U + \mathbf{K}U = F_a + F_b \ . \tag{3}$$

In the absence of external forces and neglecting damping effects, Eq.(3) is simplified to the governing equation of the free vibration analysis of a linear elastic structure:

$$-\omega^2 \mathbf{M}U + \mathbf{K}U = 0 \ , \tag{4}$$

where U can be expressed by the modal functions as

$$U = \sum u_j c_j \ , \tag{5}$$

with u_j and c_j (j=1, 2, …) denoting the eigenmodes and the corresponding amplitudes respectively. By utilizing the modal superposition method, the corresponding eigenvalue problem can be formed as

$$\left(-\omega_j^2 \mathbf{M} + \mathbf{K}\right) u_j = 0 \ , \quad j = 1, 2, 3, ... \ , \tag{6}$$

where u_j is the j-th eigenmode and ω_j is the j-th eigenfrequency. It should be noted that the eigenmode corresponds to the mode shape of structural vibration, and the eigenfrequency corresponds to the natural frequency of structural vibration.

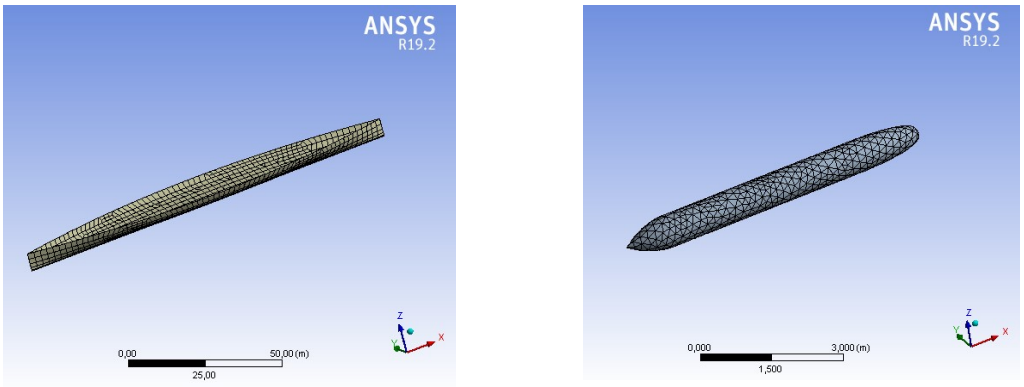


Fig. 1: Discretized vessels' surface using finite elements. Left: Surface ship. Right:AUV

The eigenvalue problem is solved using a FEM modal analysis implemented in the ANSYS Mechanical software [5]. Two vessels, considered as hollow flexible structures with vibrating surface are considered, a travelling ship and a submerged AUV. The ship hull is modelled using the parabolic Wigley hull geometry, with length and breadth chosen as $L = 130\,m$ and $B = 25\,m$, respectively, and the draft is set to $T = 6\,m$. The AUV geometry is modelled using the typical Myring geometry [6], with $L = 7\,m$, $a = 1.5\,m$, $b = 1\,m$, $d = 0.7\,m$ and $n = 3$. In both cases the material of the hull is considered structural steel with density $\rho = 7850\ kg/m^3$ and Young's modulus $E = 2 \cdot 10^{11}\ Pa$. The thickness of the ship surface is set to 1.6cm, whilst the thickness of the AUV surface is set to 5cm. For

the application of FEM, the ship surface is discretized with 900 elements, whilst the AUV surface is discretized with 3000 elements (Fig. 1).

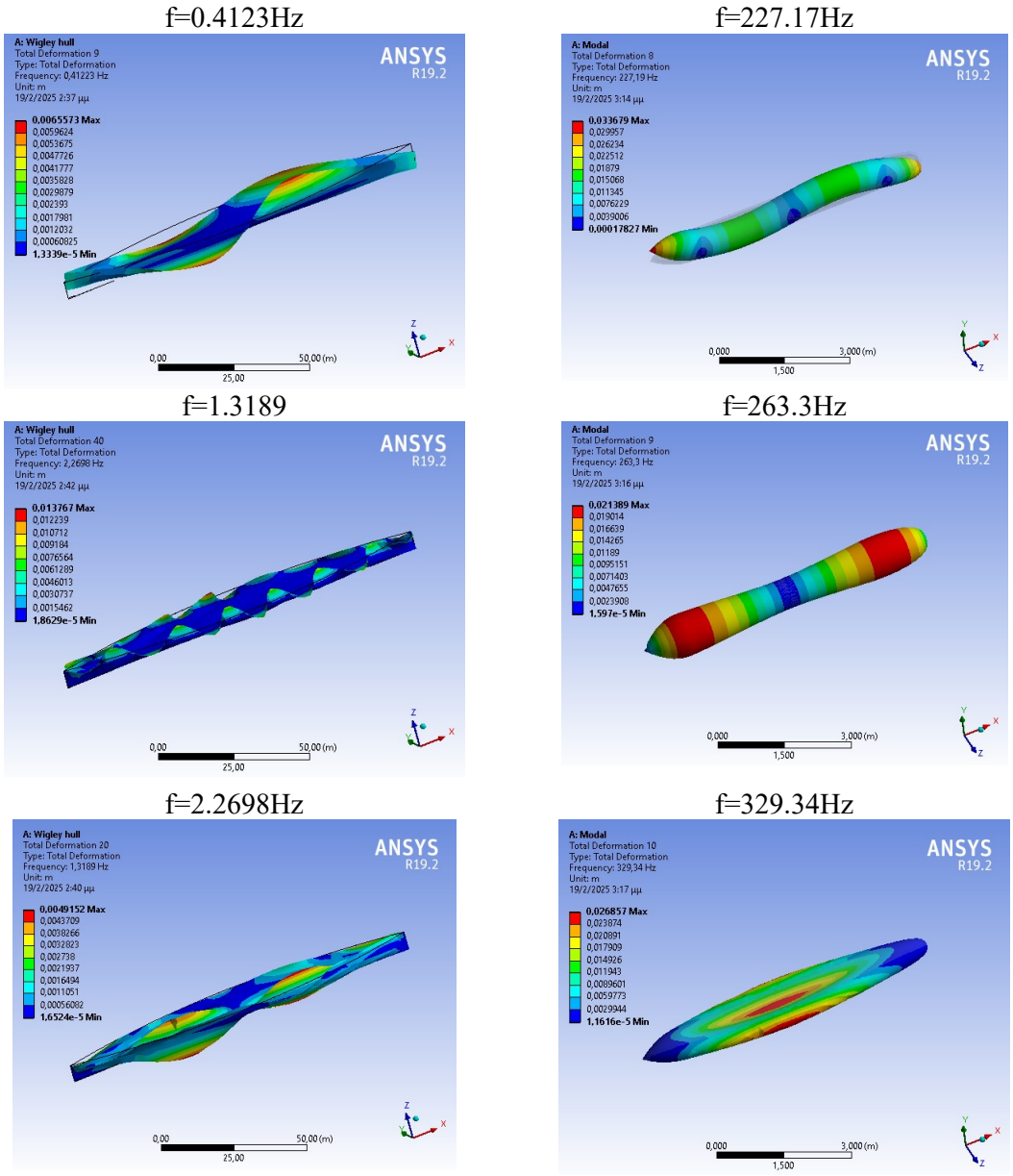


Fig. 2: Total deformations at different eigenfrequencies. Left: ship hull, Right: AUV

The deformations of the ship surface for selected eigenfrequencies are presented in Fig. 2 (left). They exhibit a wavy pattern with a characteristic length inversely proportional to the eigenfrequency, mainly in the direction of the long dimension of the ship. Note that the absolute value of deformations is not important since they correspond to the free response of the structure obtained by modal analysis. Due to the much smaller dimensions the eigen modes of the studied AUV are significantly larger. Deformations for selected eigenmodes are presented in Fig. 2 (right). Due to the closed surface and the smaller size, the deformation pattern is much different compared to the ship. It is interesting to observe that for the higher eigenfrequencies of 263 and 329 *Hz* deformations occur along the lateral directions *y* and *z*.

2.2 The hydroacoustic model in the frequency domain

In this part the hydroacoustic model is formulated to predict underwater acoustic radiation and propagation resulting from structural vibration in the ocean acoustic waveguide at specific eigenfrequencies. According to Eq.(5) the acoustic response for any excitation is obtained by superposition of the acoustic fields associated with the vibration modes, therefore the following approach provides the general treatment of the problem.

The acoustic waveguide consists solely of seawater, limited above by the free surface and below by the fully reflective, hard seabed surface, respectively. In the frequency domain the acoustic wave velocity potential φ induced by the structural vibrations will satisfy the Helmholtz equation. According to the modal superposition method, each mode φ_j of the velocity potential corresponding to the ω_j eigenfrequency of structural vibrations will also satisfy the Helmholtz equation, and the same applies for each mode of the acoustic pressure p_j. Therefore,

$$\Delta p_j + k_j^2 p_j = 0 \,, \tag{7}$$

where Δ denotes the Laplacian and k_j represents the wave number that corresponds to the ω_j eigenfrequency.

Given the displacements $u_j(\mathbf{x})$, as provided by the solution of the eigenvalue problem (6), the scattering problem is considered, excited by the structural vibrations on the wetted surface of a surface ship or AUV vessel, also accounting for the scattering effects of the free-surface and seabed boundary of the ocean acoustic waveguide. The mathematical description of the scattering of time-harmonic waves by an obstacle D leads to a Boundary Value Problem (BVP) for the Helmholtz equation. Considering the boundary of the obstacle ∂D to be acoustically hard the following simplified boundary condition is imposed on the wetted surface ∂D of the body:

$$\frac{\partial p_j(\mathbf{x})}{\partial n(\mathbf{x})} = i\omega_j u_j(\mathbf{x}) = g_j(\mathbf{x}), \quad j = 1,2,3,..... \quad , \quad \mathbf{x} \in \partial D \,, \tag{8}$$

indicating that the normal displacement at the interface between the structure and the sound field is continuous. In the above equation $\mathbf{n}(\mathbf{x})$ is the unit vector normal to ∂D with a direction towards the exterior of D.

In the case of the acoustic waveguide, $p_j(\mathbf{x})$ should also satisfy the homogeneous Dirichlet BC, $p_j = 0$ at the sea surface, $z = 0$, and the Neumann BC, $\partial p_j / \partial n = 0$ at the hard sea bottom, $z = -h$. The solution of the scattering problem is given by the integral

$$p_j(\mathbf{x}) = \int_{\partial D} G(\mathbf{x},\mathbf{y})\sigma_j(\mathbf{y})dS(\mathbf{y}), \qquad j = 1,2,3.... \quad \mathbf{x} \in R^3 \setminus \partial D \quad , \tag{9}$$

where $G(\mathbf{x},\mathbf{y})$ is the Green's function of the Helmholtz equation in the waveguide, given by either by a normal series expansion [7], or by a superposition of fields of monopole sources according to the multiple image model. The distribution of source intensities on

the body surface, $\sigma_j(\mathbf{y})$, is provided as the solution of the BVP, discretized and solved using a 3D BEM. More details can be found in [8].

The underwater sound pressure p in the ocean environment is approximated by the summation of the pressure contributions associated with each structural mode:

$$p = i\omega\rho_\alpha \sum p_j c_j \ , \tag{10}$$

where ρ_α is the water density, and the amplitudes c_j can be calculated to represent the structure response for any load. The sound pressure level SPL can be calculated using the modulus of the complex sound pressure $|p|$, $SPL = 20\log\left(|p|/p_{ref}\right)$, p_{ref} is the reference sound pressure.

3. NUMERICAL RESULTS AND DISCUSSION

The developed BEM is applied to the solution of the scattering problem and the prediction of the acoustic field in the case of noise emission from the vibrations of the two flexible vessels, for which a modal analysis has been performed as presented in section 2.1. The displacements calculated by the modal analysis are used in the boundary condition (8) on the wetted surface of the vessel. In both cases, an isovelocity environment $c = 1500\,m/s$ modelling an ocean acoustic waveguide of depth $h = 500\,m$ is considered. For the travelling ship the submergence depth is taken equal to 4.5m. The selected Froude number corresponds to a vibrational frequency of $4.1567Hz$. A number of $25 \times 10 = 250$ elements are proven to be sufficient for the discretization of the hull.

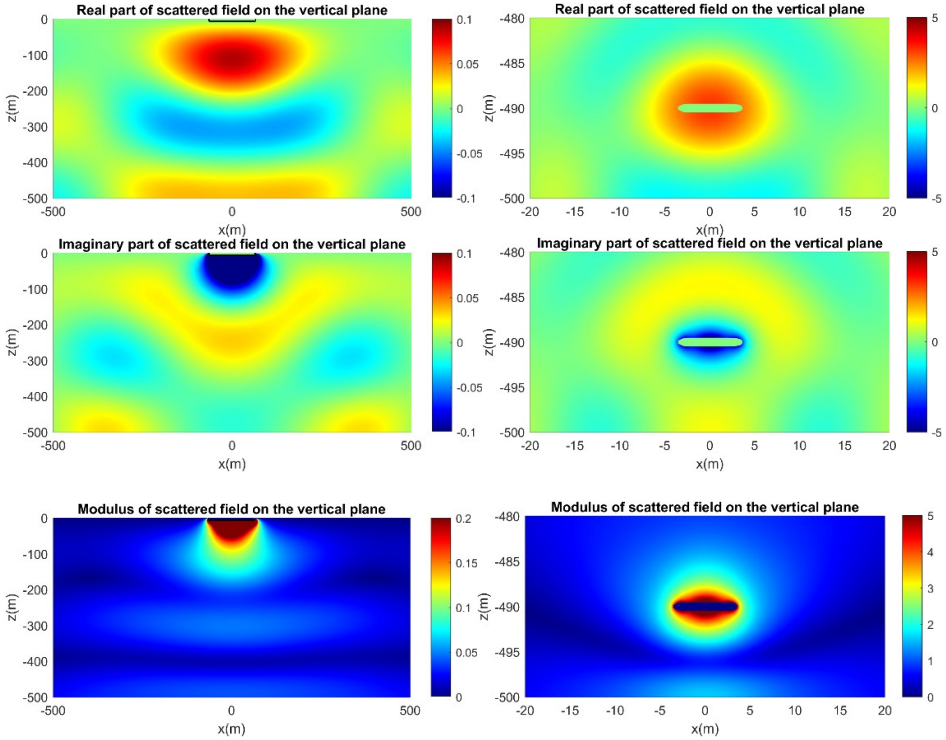


Fig. 3: Plots of acoustic field on a vertical plane passing through the major hull axis. Left: Surface ship at f=4.1567Hz. Right: AUV at f=91.963Hz.

The submergence depth of the AUV is taken as $490m$, $10m$ above the seabed. Compared to the ship, the size of the vessel is significantly smaller, and its submergence depth is large, therefore focus is made on a smaller region around the AUV hull, close to the seabed. A number of $25 \times 15 = 375$ elements are used for the discretization of the hull. The selected Froude number corresponds to a vibrational frequency of $91.963Hz$. In Fig. 3, the contours of the real and imaginary part, as well as the modulus of the acoustic field are presented on a vertical xz-plane passing through the major axis of the hull. The presence of both the seabed and the free surface significantly affects the directionality of the acoustic field. The pattern of the acoustic field in the z-direction is governed by the characteristic wavelength. In the case of AUV, due to the submergence of the whole vessel, noise is emitted to all directions. In the plot of the modulus the mirror effect caused by the presence of the seabed is pronounced, whilst, as expected, the effect of the free surface has weakened in the submergence depth of $490m$.

4. CONCLUSIONS

In the present work, the acoustic radiation from the vibration of flexible sea vehicles into the ocean acoustic waveguide is studied. The 3D scattering problem is treated by developing a 3D-BEM model in the frequency domain. Deformations of the wetted surface of the marine vehicles is introduced through the boundary condition indicating that the normal displacement at the interface between the structure and the sound field is continuous. The effects of the free surface and the seabed are considered through the corresponding boundary conditions of the 3D scattering problem. In the ship case, results show that the effects of the hull scattering, the seabed and the sea surface are important for the directionality of the generated noise. In the AUV case, due to the larger submergence depth, the effect of the free surface is weak, and the directionality is mainly affected by the body scattering and the presence of the seabed. This result may be further exploited to consider the response of the ship and AUV vibrations at certain excitations from the machinery noise which is one of the main components of noise generated from sea vehicles. The present model will contribute to the development of a computational model for URN modelling, aiming to the design of quieter vessels, in compliance with the threshold values for continuous noise developed under the European Commission's Marine Strategy Framework Directive (MSFD).

REFERENCES

[1] **IMO**, *Guidelines for the Reduction of Underwater Noise from Commercial Shipping to Address Adverse Impacts on Marine Life* London, 2014.

[2] **McKenna, M., Ross, D., Wiggins, S., Hildebrand**, **J.** Underwater radiated noise from modern commercial ships*, J Acoust Soc Am.*, 131(1), pp.92-103, 2012.

[3] **Cook, R.D., Malkus, D.S., Plesha M.E., Witt, R.J.**, *Concepts and Applications of Finite Element Analysis*, Wiley, New York, 2002

[4] **Belibassakis, K., Prospathopoulos, J.**, A 3d-BEM for underwater propeller noise propagation in the ocean environment including hull scattering effects, *Ocean Engineering*, 286 (1), 2023.

[5] **ANSYS** https://www.ansys.com/products/structures/ansys-mechanical

[6] **Joung, T-H, Sammut, K, He, F, Lee S-K**. Shape optimization of an autonomous underwater vehicle with a ducted propeller using computational fluid dynamics analysis, *Inter J Nav Archit Oc Engng*, 4(1), pp. 44-56, 2012.

[7] **Jensen, F.B., Kuperman, W.A., Porter, M.B., Schmidt, H.**, *Computational Ocean Acoustics*, 2nd edition, Springer New York, 2011.

[8] **Belibassakis, K., Prospathopoulos. J., Malefaki, I**. Scattering and Directionality Effects of Noise Generation from Flapping Thrusters Used for Propulsion of Small Ocean Vehicles MDPI, *J. Marine Science and Engineering* 10 (8), 1129, 2022.

Structured Session
23. Underwater noise modelling and measurements

Organised by David Barclay and Martin Siderius

Near-Field Testing and Evaluation Method for Stream Noise of Linear Array Based on Laser Doppler Velocimetry

Zeng Cao[1,2], Shiquan Wang[1,2], Haoyang Tong[1,2], and Yi Chen[1,2]

[1]Hangzhou Applied Acoustics Research Institute, Hangzhou 311400, China
[2] Zhejiang Key Laboratory of Acoustic Testing Technology

Zeng Cao, +86 18858391201, zeng_cao0520@163.com

Abstract: *Towed linear arrays have widespread applications in ocean monitoring and underwater detection, but their performance is often affected by stream noise. To accurately assess the stream noise characteristics of towed linear arrays in actual operational environments, this study proposes a near-field testing and evaluation method based on Laser Doppler Velocimetry (LDV). This method leverages the high-precision velocity measurement capabilities of LDV and integrates numerical analysis models to achieve precise testing and analysis of stream noise near the towed linear array. The results demonstrate that the LDV-based testing method can effectively capture the complex flow field characteristics around the towed linear array and accurately evaluate the stream noise levels. Compared to traditional methods, this approach enables testing within limited spaces, significantly reducing experimental costs and time. This research provides a novel technical means for evaluating stream noise in towed linear arrays, which is crucial for enhancing their performance in practical applications.*

Keywords: *Stream Noise, Linear Array, Laser Doppler Velocimetry*

1. INTRODUCTION

Linear arrays, particularly Towed Array Sonars (TAS), are essential tools in underwater acoustics for both defense and scientific research (e.g., marine mammal monitoring). Their effectiveness comes from deploying hydrophones over long distances, separating them from the towing platform's noise and allowing for directional sensing through beamforming. However, as the array moves through water, the flow over its surface generates significant self-noise, known as stream noise or flow noise.[1] This noise primarily arises from turbulent boundary layer pressure fluctuations and flow-induced vibrations.[1] Stream noise becomes a dominant limitation, especially at higher speeds, masking faint target signals and degrading the sonar's detection range and overall performance.[2] Accurately characterizing this noise, particularly in the near-field region close to the array surface where the noise originates, is crucial for understanding its mechanisms and developing mitigation strategies.[3]

The standard approach for measuring underwater sound, including stream noise, involves using hydrophones – sensors that convert pressure fluctuations into electrical signals.[4] In near-field testing, hydrophones are placed close to the array surface to capture the noise characteristics at the source.[5] However, this method suffers from a critical drawback: intrusiveness. The physical presence of the hydrophone and its mounting inevitably disturbs the local hydrodynamic flow field it aims to measure.[6] This disturbance alters the turbulent structures and pressure distribution.

Furthermore, the hydrophone measures a combination of true propagating sound and non-propagating hydrodynamic pressure fluctuations (pseudo-sound) associated with turbulent eddies. The intrusive sensor not only measures the flow's inherent pseudo-sound but can also generate additional pseudo-sound due to flow interaction with the sensor itself. This contamination makes it extremely difficult to isolate the true acoustic signal or accurately characterize the original hydrodynamic pressure field responsible for the noise.[7] Consequently, hydrophone-based near-field measurements can lead to significant inaccuracies, hindering the precise diagnosis of noise sources and the validation of noise reduction techniques. This limitation highlights a critical gap in current stream noise evaluation methods.

Laser Doppler Velocimetry (LDV) presents a powerful, non-intrusive optical technique for fluid velocity measurement.[5] It works by detecting the Doppler frequency shift of laser light scattered by small particles moving with the fluid.[8] By measuring this shift, LDV determines the fluid velocity components at a precise point without physical contact. The key advantage of LDV in this context is its non-intrusiveness. Unlike hydrophones, LDV probes the flow using light beams, eliminating the physical disturbance and sensor-induced pseudo-sound generation that compromise intrusive measurements.[8] LDV allows for the characterization of the undisturbed velocity field very close to the array surface. Since stream noise originates directly from hydrodynamic phenomena like turbulent velocity fluctuations, LDV provides direct access to the fundamental physics of noise generation. It can yield detailed data on turbulence intensity, velocity profiles, and flow structures near the wall. This accurate, non-intrusive velocity data is invaluable for understanding noise mechanisms, validating computational fluid dynamics (CFD) predictions, and assessing the effectiveness of noise control strategies, such as surface modifications. While LDV measures velocity, this information is directly linked to the pressure fluctuations responsible for noise and can be used to infer pressure characteristics without intrusive measurement errors.

This paper proposes and evaluates a near-field testing and evaluation methodology for linear array stream noise based on non-intrusive Laser Doppler Velocimetry (LDV). The primary objective is to overcome the accuracy limitations associated with conventional intrusive hydrophone techniques by leveraging LDV's ability to measure the undisturbed near-field flow dynamics. The paper will cover the theoretical background, describe the experimental setup using LDV with a linear array model in a flow facility, present the detailed velocity field measurements, analyze the results to understand the flow physics relevant to noise generation, and discuss the advantages of the LDV-based approach for stream noise assessment

2. THEORY FOR RADIATED NOISE CALCULATION

The predominant source of acoustic radiation from linear arrays is identified as trailing edge noise. This phenomenon is intrinsically linked to the turbulent boundary layer that develops along the linear arrays. Specifically, trailing edge noise is understood to arise from the scattering of acoustic waves generated by the turbulent flow as it convects past the hydrofoil's trailing edge.

Howe's empirical model provides a widely cited framework for predicting the characteristics of this radiated noise. The mean square sound pressure, $\langle p^2 \rangle$, is approximated by the following expression[9]:

$$< p^2 > \approx \rho_0^2 u_0^2 v^2 M_v \left(\frac{LI}{R^2} \right) \sin\alpha \sin^2\left(\frac{\theta}{2} \right) \cos^3\beta \quad (1)$$

where the $\langle p^2 \rangle$ is the mean square sound pressure, ρ0 is the fluid density, u0 is the uniform free stream velocity, v is the convection velocity, Mv is the turbulent convection Mach number, L is the, length of linear arrays, I is the displacement thickness of the turbulent boundary layer, R is the distance from the trailing edge to the observation point, β is the sweep angle of the trailing edge, α is the angle between the observation point and the hydrofoil trailing edge and θ is the direction angle.

2. NOISE MEASUREMENT TEST BASED ON LDV

The experiments were conducted in a recirculating water channel. This type of facility allows for controlled and continuous flow of water over an extended period, which is essential for detailed velocity measurements. The specific water channel to be used has overall dimensions of approximately 4.92 meters in length, 0.4 meters in width, and 2.34 meters in height. The test section, where the segment of the linear arrays was placed and measurements taken, has dimensions of 1.6 meters in length, 0.4 meters in width, and 0.4 meters in height. The water flow within the channel is generated by a pump system that allows for a range of flow velocities to be achieved in the test section. To ensure a uniform and stable flow entering the test section, several flow conditioning elements are incorporated upstream. The achievable range of flow velocities in this water channel spans from 0.5 m/s to 1.5 m/s in the test section.

The segment of the linear arrays is replaced by a cylinder. The submerged cylinder to be used in this experiment will be circular in cross-section. The cylinder has a diameter of 0.05 meters and a length of 0.15 meters, resulting in an aspect ratio (length/diameter) of 3. The cylinder was made of acetal (POM) 18, which provides a smooth surface finish suitable for noise measurement. The cylinder was mounted horizontal along the centerline

of the test section, fully submerged, and fixed in place using supports that minimize interference with the flow in the region of interest.

A two-component LDV system will be employed to measure the streamwise and vertical components of the fluid velocity in the wake of the cylinder. The transmitting and receiving optics will consist of an 83 mm diameter optical fiber probe with a front lens of 360 mm focal length. The system was operated in backscatter mode, where the probe serves as both the transmitter and receiver of the laser beams, allowing for measurements with optical access from only one side of the water channel.

The velocity measurements were taken on a grid of points in the symmetry plane downstream of the cylinder to characterize the wake region. The spatial resolution of the measurement grid is finer in the near wake region (closer to the cylinder) where velocity gradients are expected to be higher, and coarser further downstream. Specifically, in the region from 0.5D to 5D downstream of the cylinder and -1D to 1D in the transverse direction (where D is the cylinder diameter), the measurement points were spaced at intervals of 0.1D. Further downstream, the spacing will be increased to 0.25D. The test layout is shown in Fig. 1.

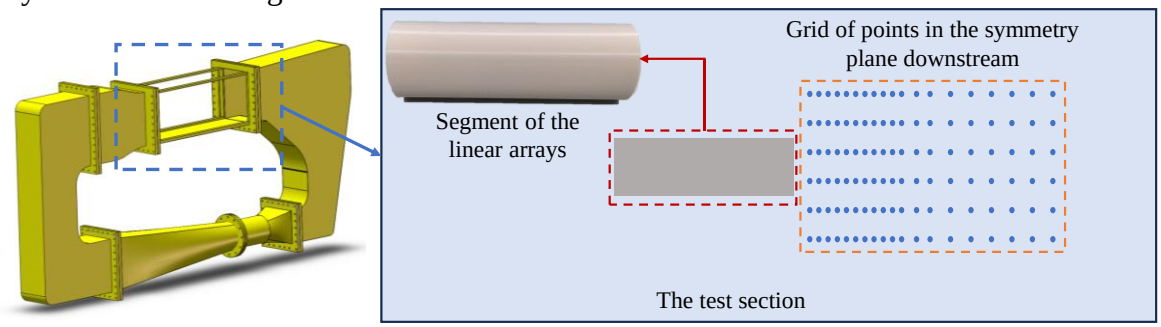


Fig.1: The test layout of the LDV flow noise test.

To facilitate comparison with the Laser Doppler Velocimetry (LDV) flow noise test results, the flow noise of the linear array segment was measured using an 8103 hydrophone array configuration with three equidistant measurement points, as depicted in the Fig. 2. The measurement points were uniformly spaced at intervals of 500 mm along the array segment.

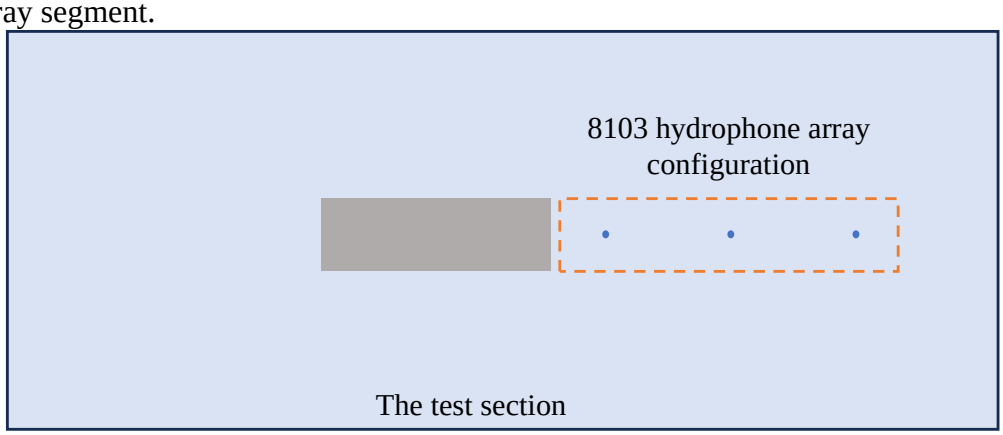


Fig.2: The layout of the hydrophone flow noise test.

3. RESULTS AND DISCUSSION

Fig. 3 presents a two-dimensional color contour map illustrating the velocity distribution within the designated test region under a nominal flow condition of 1 m/s.

According to the figure caption, this velocity field was derived using the finite difference method applied to data obtained from Laser Doppler Velocimetry (LDV) measurements. The map utilizes a color scale, shown on the right, to represent velocity magnitudes in meters per second (m/s). The scale ranges from approximately 0 m/s (indicated by dark blue) to a maximum of roughly 1.6 m/s (indicated by dark red). The velocity distribution exhibits a degree of symmetry about the horizontal centerline. The observed pattern, with a central low-velocity zone flanked by higher-velocity zones, is characteristic of flow phenomena because of wake formation behind an obstacle. The acceleration of the flow to speeds greater than the nominal 1 m/s in the outer regions suggests potential flow constriction. The regions of lowest velocity (dark blue) correspond to areas of flow separation or the near wake immediately downstream of an object.

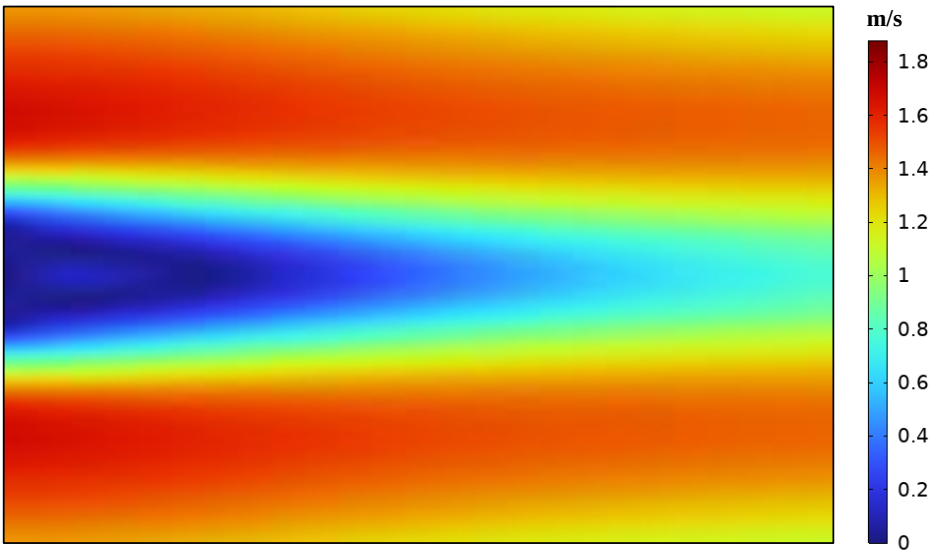


Fig.3: Velocity distribution maps of the test region at 1 m/s derived via the finite difference method based on LDV test results.

Furthermore, leveraging the detailed flow velocity distribution data obtained from the LDV tests, it becomes possible to estimate the radiated sound power. Specifically, by applying Equation (1), the sound power radiated in the direction of acoustic wave propagation at a distance of 1 meter can be calculated, which allows for the prediction of flow-induced noise based on the measured characteristics of the velocity field, as shown as in Fig. 4.

Fig.4 presents the comparison of the radiation noise level test results based on LDV and hydrophone. Both datasets demonstrate a clear positive correlation between flow velocity and radiation noise level; as the velocity increases, the noise level increases for both methods. This aligns with the fundamental principles of flow-induced noise generation, where higher flow speeds typically lead to more intense turbulent fluctuations and consequently higher noise levels. A significant observation is the consistent difference between the two datasets across the tested velocity range. The noise levels measured directly by the hydrophone are consistently higher than those estimated based on the LDV data. This discrepancy appears to widen slightly as the flow velocity increases.

While the LDV-based method likely focuses on estimating the noise generated directly by the flow phenomena around the object of interest (e.g., trailing edge noise, turbulence interaction), the hydrophone measures the total acoustic pressure at its location. This total pressure includes not only the direct flow-induced noise from the object but also any other noise sources present in the experimental facility. A plausible explanation for the elevated hydrophone readings is the potential contribution of facility background noise, particularly

noise originating from vibrations of the circulating water channel walls. Such structural vibrations, induced by the pump, the flow itself, or other machinery, can radiate sound into the water, which is then picked up by the hydrophone. This facility-related noise component would add to the actual flow noise generated by the test object, leading to higher overall measured levels compared to the LDV-based estimation, which may inherently exclude or underestimate these extraneous sources. Therefore, the higher noise levels recorded by the hydrophone might be attributed, at least in part, to the inclusion of wall vibration noise from the circulating water channel in the measurement results.

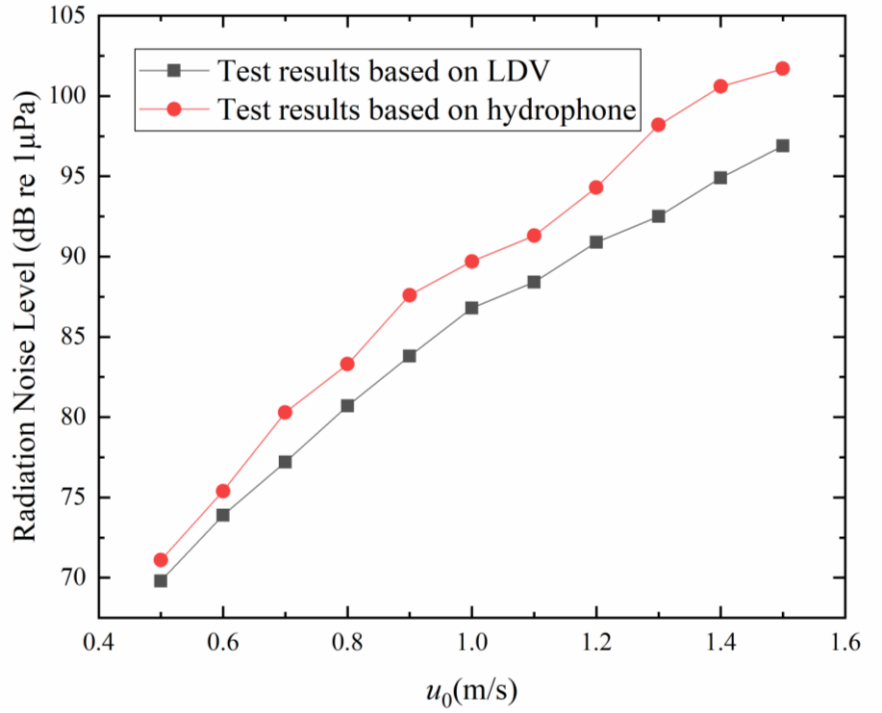


Fig.4 Comparison of the radiation noise level test results based on LDV and hydrophone.

4. CONCLUSION

This study proposed and validated a near-field testing and evaluation method for towed linear array stream noise based on Laser Doppler Velocimetry (LDV). Experimental results demonstrate that LDV effectively captures complex flow field characteristics near the array, enabling estimation of the associated stream noise. Compared to traditional hydrophones, the non-intrusive LDV method focuses more specifically on the array's self-generated noise. Consistently higher noise levels measured by hydrophones may be attributed to facility background noise, such as water channel wall vibrations. This LDV-based approach provides a valuable new pathway for accurate assessment and mechanistic study of linear array stream noise, contributing to performance optimization in practical applications.

5. ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 12304502, 12304529).

REFERENCES

[1] Kim, J. . (2022). Surface target tracking using towed array sonars with direct and bottom bounce underwater sound signals. IEEE Transactions on Signal and Information Processing over Networks, 8, 997-1007.
[2] Miller, H. A. , Smith, R. R. , & Nichols, C. S. . (1978). Solid filled vibration isolation module for a towed sonar array. US, US04090168A.
[3] William K. Blake, Mechanics of Flow-Induced Sound and Vibration, Academic Press, INC, pp. 595-620, 1986.
[4] YUAN Chi, SUN Chuan-dong, BAI Yong-lin, WANG Yi-shan, Precision and Errors Analysis of PIV Used for the Watertunnel Flow Field Displaying and Measuring system, Acta photonica sinica, 37(10), pp. 2005-2008, 2008.
[5] Humphrey, V. F. , Robinson, S. P. , Theobald, P. D. , Hayman, G. , & Cooling, M. P. . (2008). A comparison of hydrophone near-field scans and optical techniques for characterising high frequency sonar transducers. Journal of the Acoustical Society of America, 123(5), 3436.
[6] Sun, G. , Yang, D. , & Zhang, L. . (2002). Research on the method for measuring radiated noise by an underwater target in low frequency band based on the vector hydrophone. Shengxue Xuebao/Acta Acustica, 27(5), 429-434.
[7] Felli, M. , Grizzi, S. , & Falchi, M. . (2014). A novel approach for the isolation of the sound and pseudo-sound contributions from near-field pressure fluctuation measurements: analysis of the hydroacoustic and hydrodynamic perturbation in a propeller-rudder system. Experiments in Fluids, 55(1), 1-17.
[8] HUN Hai-bao, SONG Bao-wei, YUAN Chi, DU Xiao-xu, MAO Zhao-yong, Experimental Research on Particle Image Velocity Measure System in Fluid Field, Acta photonica sinica, 36(10), pp. 1928-1932, 2007.
[9] Howe, M. S., Noise Produced by a saw-toothed trailinged, Journal of the Acoustical Society Of America, 90, pp. 482-487, 1991.

A deep-sea ambient noise prediction method with few training samples based on VMD and LSTM

Guoli Song [1,2,3], Xinyi Guo [1,2,3], Qianchu Zhang[1,2], Jun Li[1,2], Qunyan Ren [1,2,3] and Li Ma

[1,2,3]

[1] Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China
[2] Key Laboratory of Underwater Acoustic Environment, Chinese Academy of Sciences, Beijing 100190, China
[3] University of Chinese Academy of Sciences, Beijing 100190, China

* Correspondence: Guoli Song, songguoli@mail.ioa.ac.cn, No. 21 North 4th Ring Road, Haidian District, 100190 Beijing, China

Abstract: *Deep-sea ambient noise levels are an important factor affecting underwater target detection, underwater acoustic communications, as well as ocean ambient monitoring and animal protection. To predict the deep-sea ambient noise spectral level, a prediction method is established combined with Variational Mode Decomposition (VMD) and Long Short-Term Memory Networks (LSTM). It can predict the ambient noise level under the condition of few training samples in the frequency band of 20 Hz - 10 kHz. The test data verify that the level of the ambient noise in the next 7 days can be predicted by using the current 3 days of deep-sea ambient noise data. The frequency and time trends of the forecast results are consistent with the measured noise trends. The average of the root mean square error (RMSE) is less than 1 dB in both the prediction band and the prediction time.*

Keywords: *Deep-sea ambient noise, noise prediction, Variational Mode Decomposition, Long Short-Term Memory Networks, few training samples*

1. INTRODUCTION

Forecasting the temporal variations of ocean ambient noise enables the formulation of targeted strategies to mitigate its adverse impacts on marine organisms and ecosystems [1], while also enhancing the performance of sonar systems [2] and improving the safety of maritime search and rescue operations and navigation systems.

Ocean ambient noise is influenced by multiple natural factors, including wind speed, waves, ocean currents, temperature and so on, leading to significant temporal variability [3]. The interplay of these environmental parameters introduces high uncertainty and complexity into the generation and propagation of ocean ambient noise, posing challenges for temporal forecasting, particularly under limited sample conditions.

Conventional approaches to ambient noise intensity prediction primarily rely on physical sound field modelling and statistical empirical models. By integrating dynamic sound field modelling [4] with noise source analysis, it becomes feasible to forecast ocean ambient noise fields. In recent years, the development and application of deep learning technology in underwater acoustics have provided new approaches for ocean ambient noise forecasting. Currently, neural networks have made achievements in underwater noise classification [5-6], ship radiated noise identification [7-9], and sound source localization [10-11]. Ocean ambient noise forecasting with machine learning integration has also yielded preliminary successes. Song implemented a Long Short-Term Memory (LSTM)-based approach to predict deep-sea ambient noise, achieving a prediction error of <1 dB for the 20 Hz - 5 kHz frequency band using 2,712 samples with a 9:1 training-test split ratio [12]. Yuan developed a linear time-series model for long-term ambient noise prediction. Based on 20 days of measured data, they achieved wide-band ocean ambient noise prediction [13]. It should be specially noted that the two methods above rely on long-term training samples to achieve robust performance.

This paper proposes an ocean ambient noise spectral level (NL) forecasting method for deep-sea environments under limited training samples by integrating Variational Mode Decomposition (VMD) and LSTM networks. The method enables prediction of ambient noise intensity across the 20 Hz to10 kHz and has been validated through sea trials.

2. VARIATIONAL MODE DECOMPOSITION OF AMBIENT NOISE

Variational Mode Decomposition is a fully non-recursive and adaptive variational mode decomposition method. It determines the number of decomposed modes based on the intrinsic characteristics of the signal. It also optimally matches the central frequency and finite bandwidth for each mode adaptively, thus effectively decomposing the Intrinsic Mode Functions (IMFs). Here, VMD is primarily employed to decompose the ambient noise spectral level at a specific central frequency into subcomponents.

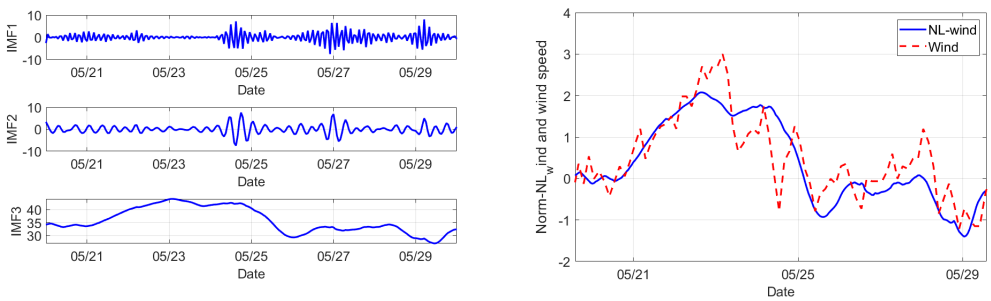


Fig.1: VMD decomposition results of NL and the trend of wind

As wind is one of the main factors of deep-sea ambient noise, the ambient noise spectral level time series is decomposed into two subcomponents - NL_wind and NL_other -via VMD, which means the wind-speed-related subcomponent and the composite subcomponent accounting for other environmental factors.

Fig. 1 displays the three subcomponents derived from VMD decomposition of the ambient noise spectral level at 10 kHz. Analytical results indicate that IMF3 exhibits long-term trend alignment with wind speed, therefore IMF3 is denoted as NL_wind, whereas the remaining components (IMF1, IMF2, and the residual) are summed to form NL_other.

3. VMD-LSTM-BASED AMBIENT NOISE PREDICTION METHOD

Building upon the decomposition of ambient noise subcomponents, the LSTM network is integrated to predict ambient noise intensity.

Let the two decomposed subcomponents of the noise spectral level at a specific central frequency f_c be denoted as $NL_wind_{f_c}=\left\{NL_wind_{f_ct}\right\},t=1,2,...,n$ and $NL_other_{f_c}=\left\{NL_other_{f_ct}\right\},t=1,2,...,n$. The two time-series subcomponents are used as inputs to the network and output the noise spectral level through the forecasting model. Specifically, the network utilizes two historical subcomponent sequences $NL_wind_{f_ct}$ and $NL_other_{f_ct}$ over a temporal window L=1 to predict $NL_{f_c(t+L)}$ .Then ,the predicted $NL_{f_c(t+L)}$ is decomposed via VMD into updated subcomponents $NL_wind_{f_c(t+L)}$ and $NL_other_{f_ct}$. These updated subcomponents are then fed back into the LSTM to predict $NL_{f_c(t+L+1)}$. This iterative process enables extended forecasting of spectral level across the frequency range of 20 Hz to 10 kHz, so that the variation curve of NL with frequency is obtained.

Fig. 2 illustrates the architecture of the deep-sea ambient noise intelligent learning and forecasting model, which comprises an input layer, LSTM layer, fully connected layer, and output layer. A dropout layer is incorporated between the LSTM layer and the fully connected layer to mitigate overfitting. The model is constructed with the following hyperparameters: 128 hidden neurons, dropout rate of 0.2, and initial learning rate of 0.0005.

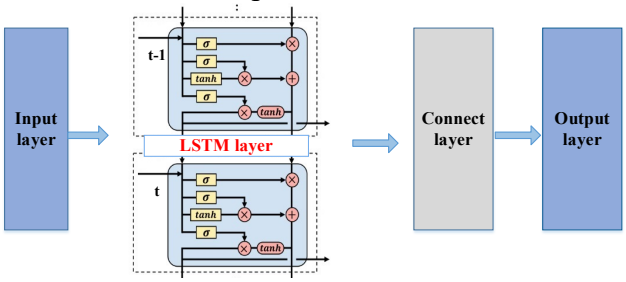


Fig.2: The prediction network of ambient noise intensity.

To assess the performance of the forecasting model, two mean root-mean-square errors (RMSEs) are defined in the frequency domain and time domain, respectively:

- Frequency-Averaged RMSE (denoted as $\overline{RMSE_F}$): The RMSE between predicted and ground-truth NL is averaged over N frequency bands across the entire forecasting duration.

$$\overline{RMSE_F}=1/N\sum_{i=1}^{N}\sqrt{\sum_{j=1}^{T}\left(NL_{pre}(i,j)-NL_{true}(i,j)\right)^2/T} \quad (1)$$

- Time-Averaged RMSE (denoted as $\overline{RMSE_T}$): The RMSE between predicted and ground-truth NL is averaged over T forecasting time steps across the entire frequency band.

$$\overline{RMSE_T} = 1/T \sum_{j=1}^{T} \sqrt{\sum_{i=1}^{N} \left(NL_{pre}(i,j) - NL_{true}(i,j) \right)^2 {}_i / N} \quad (2)$$

Where $NL_{pre}(i,j)$ is the predicted noise spectral level at the $i-th$ frequency band and $j-th$ time step, and $NL_{true}(i,j)$ is the measured noise spectral level at the $i-th$ frequency band and $j-th$ time step,

4. RESULTS AND ANALYSIS

4.1 Data Acquisition and Processing

The ambient noise data utilized in this study were collected in a deep-sea region using a subsurface hydrophone deployed at a depth of 2,672 meters. The system operated at a sampling rate of 64 kHz from 00:00 May 20 to 00:00 May 30, yielding 240 hours of valid data. Synchronous meteorological parameters (e.g., wind speed, temperature) were measured using meteorological sensors during the noise acquisition period.

For processing, the raw ambient noise data were segmented into 1-hour intervals. Spectral levels were computed for 1/3-octave bands within the 20 Hz -10 kHz frequency range, generating a 240-hour time series of noise spectral levels. Then it was decomposed into two subcomponents NL_wind and NL_other .

4.2 Prediction Results Analysis

The two subcomponents (NL_wind and NL_other) were partitioned into a training set (30%, 72 samples) and a test set (70%, 168 sample points). These subcomponents served as inputs to the LSTM network, generating forecasts of the ambient noise spectral level (denoted as the VMD-based method). For comparison, a baseline model using the ambient noise spectral level (NL) as input was also evaluated (denoted as the NL-based method).

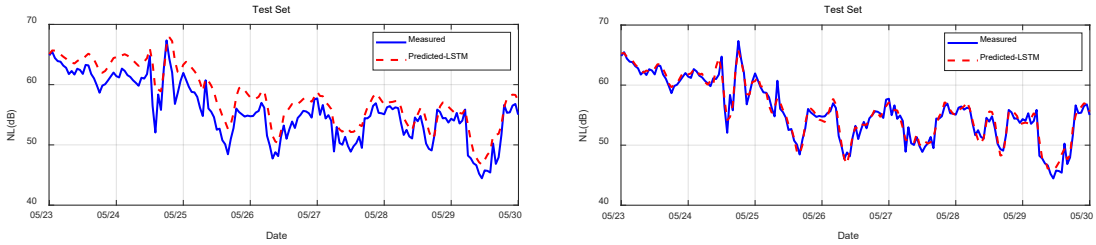


Fig.3: Comparison of Ambient Noise Spectral Level Predictions at Representative Frequencies (Left: NL-based method | Right: VMD-based method).

The Fig. 3 depicts the predicted noise intensity profiles obtained from the NL-based (left panels) and VMD-based (right panels) methods at 1 kHz. As can be seen from the figure, the VMD-based method significantly outperforms the NL-based approach.

Predictions were generated for all 1/3-octave bands within the 20 Hz -10 kHz range. The RMSEs between predicted and ground-truth values were analyzed, yielding frequency-dependent RMSE profiles shown in Fig. 4. Overall, the VMD-based method outperforms the NL-based approach. The frequency-averaged RMSE $\overline{RMSE_F}$ for the two methods are 2.23 dB (NL-based) and 0.96 dB (VMD-based), respectively.

Fig. 5 compares measured and predicted frequency-dependent noise spectral levels at hour 168 of the test set. The solid blue line represents the measured NL, and the dotted red line represents the predicted NL. The VMD-based predictions (fright figure) exhibit closer alignment with measured spectral trends across frequencies. At hour 168, the NL-based method has a RMSE of 1.94 dB, while the VMD-based method's RMSE is 1.07 dB.

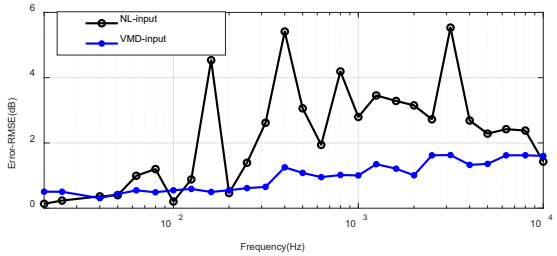


Fig.4: Frequency variation curve of RMSE for ambient noise forecasting.

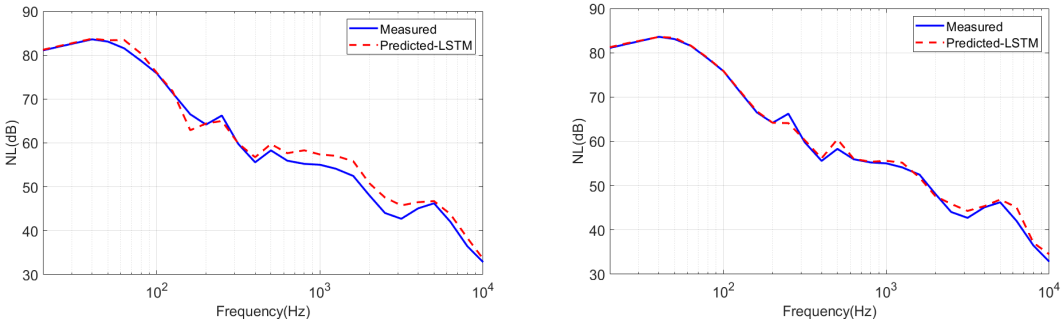


Fig.5: Frequency variation curves of forecasted ambient noise spectral level and the truth (Left: NL-based method | Right: VMD-based method).

Fig. 6 shows the RMSE variation over time across the entire frequency band. Overall, the VMD-based method outperforms the NL-based method. The time-averaged RMSE $\overline{RMSE_T}$ of the two methods are 2.50 dB and 0.93 dB, respectively.

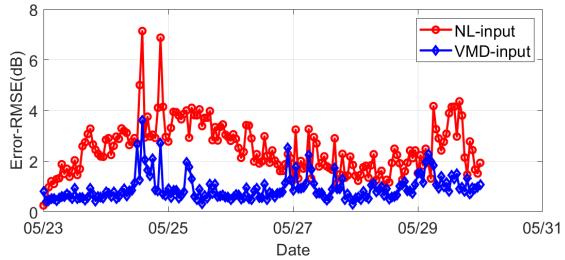


Fig.6: Time variation curve of RMSE for ambient noise forecasting.

In summary, the VMD-based method exhibits enhanced performance in terms of frequency-domain errors as well as temporal errors under the condition of small-samples compared to the NL-based baseline.

5. CONCLUSIONS

This study proposes a VMD-LSTM integrated method for forecasting deep-sea ambient noise spectral level, validated using 240 hours of field-measured marine data. The forecasting performance of the VMD-based and NL-based methods was comparatively analyzed across two dimensions: temporal variations at individual frequencies as well as frequency-dependent variations across the 20 Hz - 10 kHz band under limited training samples. Results show that the VMD-based method demonstrates superior performance with limited training data. Under a 3:7 training-test split ratio (i.e., using 3-day historical data to forecast the next 7 days), the model achieves:

- Frequency-averaged RMSE: 0.96 dB (across the entire forecasting duration).

- Time-averaged RMSE: 0.93 dB (across the 20 Hz -10 kHz band).

Both RMSE metrics are below 1 dB, underscoring the method's reliability for practical applications.

6. ACKNOWLEDGEMENTS

This research was funded by National Natural Science Foundation of China, grant number 12474445 and Basic and Frontier Exploration Project Independently Deployed by Institute of Acoustics, Chinese Academy of Sciences, grant number JCQY202408.

REFERENCES

[1] **L. S. Lemos, J. H. Haxel, A. Olsen**, Effects of vessel traffic and ocean noise on gray whale stress hormones, *Sci. Rep.*, volume (12), pp. 18580, 2022.
[2] **Z. Zhao**, Analysis and application of spatio-temporal correlation characteristics of measured ocean ambient noise, *Applied Acoustics*, volume (214), pp. 109653, 2023.
[3] **F. Li, D. Xu, J. Wang**, et al., Observations of wind-generated noise by the tropical cyclone, *J. Acoust. Soc. Am.*, 143(6), pp. 3312-3324, 2018.
[4] **L. Pang**, Methodology study on the fluctuations prediction of the acoustic filed in the sophisticated oceanographic environment, Beijing: University of Chinese Academy of Sciences,2019.
[5] **G. Song, X. Guo**, Underwater Noise Modeling and Its Application in Noise Classification with Small-Sized Samples, Electronics, volume (12), pp. 1-17, 2023.
[6] **G. Song, X. Guo**, **W. Wang**, Underwater Noise Classification based on Support Vector Machine, *IEEE/OES China Ocean Acoustics Conference COA 2021*, China, 14–17 July, 2021.
[7] **Z. Shan**, High Signal-to-Noise Ratio MEMS Noise Listener for Ship Noise Detection, Remote Sensing, volume (15), pp. 777, 2023.
[8] **S. Liu**, A fine-grained ship-radiated noise recognition system using deep hybrid neural networks with multi-scale features, Remote Sensing, volume (15), pp. 2068, 2023.
[9] **W. Fan, H. Yao, H. Wang**, Recognizing the State of Motion by Ship-Radiated Noise Using Time-Frequency Swin-Transformer, IEEE Journal of Oceanic Engineering, volume (15), pp. 2068, 2024.
[10] **C. Li, Z. Huang, J. Xu, X. Guo**, Multi-channel underwater target recognition using deep learning, ACTA ACOUSTICS, volume (45), pp. 506-514, 2020.
[11] **H. Niu, Z. Gong** Deep learning source localization using multi-frequency magnitude-only data, J. Acoust. Soc. Am., volume (146), pp. 211-222, 2019.
[12] **G. Song, X. Guo, L. Ma**, A Prediction Method for Deep Sea Ambient Noise Intensity Based on LSTM, IEEE/OES China Ocean Acoustics Conference COA 2024, 29-31 May, 2024.
[13] **B. Yuan, L. Lu**, Prediction of Full-Frequency Deep-Sea Noise Based on Sea Surface Wind Speed and Real-Time Noise Data, Remote Sensing, volume (17), pp.101, 2025.

Structured Session
25. Unmanned vehicles for underwater acoustic surveillance and monitoring

Organised by Alain Maguer

Autonomous Underwater Inspection of Inland Waterways with a Small Uncrewed Surface Vessel for Police and Emergency Response

Alfie Anthony Treloar[1], Zuhayr Rymansaib[1], Alireza Tamaddoni Nezhad[2], and Alan Hunter[1]

[1]Department of Mechanical Engineering, University of Bath, Bath BA2 7AY, United Kingdom
[2]School of Computer Science and Electronic Engineering, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom

Alfie Anthony Treloar, Department of Mechanical Engineering, University of Bath, Bath, BA2 7AY, aoat20@bath.ac.uk

Abstract: *The investigation of crime and emergency scenes in small inland bodies of water is typically a laborious process requiring a specialist team of divers. Also, visibility is often limited, and manual inspection of the area can thus expose divers to unseen hazards. While there is some use of sonar from crewed boats, this still requires skilled personnel to operate and interpret data.*
PRIME is an uncrewed surface vessel being developed to aid investigators by providing actionable information about the area of interest. It is intended to improve the efficiency and safety of divers by highlighting areas of interest and showing benign regions. This is achieved through a sensor payload including two sets of side-scan sonar sensors at different frequencies and a switchable frequency forward-looking sonar. The data from these sensors is combined to provide a detailed situation map, easily interpreted on shore by the dive team as they prepare to enter the water.
This paper describes the vessel and its sensor payload, as well as the software running onboard and the shoreside computer interface. Finally, we present the results of some recent field trials showing the real-time generation of a situation map.

Keywords: *uncrewed surface vessel, autonomy, side-scan sonar*

1. INTRODUCTION

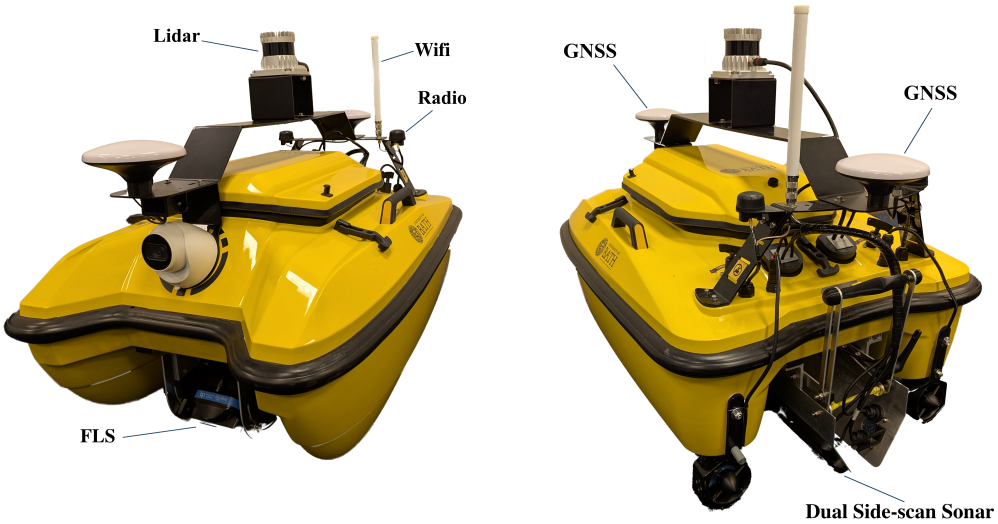
Searching for underwater evidence in response to emergencies is a challenging and labor intensive task. Specialized dive teams are manually required to systematically survey bodies of water with very low visibility and the possibility of concealed hazardous obstructions. Modern technology has been used to aid this process, such as towed side-scan sonar (SSS) and remotely operated vehicles (ROVs), but this still has the drawbacks of requiring skilled operators to use safely and effectively.

The Police Robot for Inspection and Mapping of underwater Evidence (PRIME) is an autonomous surface vessel (ASV) being developed at the University of Bath, UK[1]. The aim of the vessel is to assist emergency response dive teams by autonomously surveying a body of water prior to divers entering the water. The intended operating procedure is: deploy ASV to water body; autonomously scan and image the underwater scene with sonar; autonomously analyze gathered data and generate a situation map detailing potential areas of interest. This distilled information can then be used by the divers to prioritize regions within the search area and save time by avoiding featureless areas.

This paper presents the newest iteration of PRIME. We will present the new suite of sensors and some of the major changes to the software as well as some recent experimental results.

2. HARDWARE

The boat is dual hull with an underwater sensor payload carriage between the two hulls. This consists of two pairs of SSS sensors and a forward looking sonar (FLS) for gathering survey data. On the top are navigational sensors and a lidar for situational awareness to allow safe autonomous operation of the vessel.



(a) *(b)*

Figure 1: Annotated images illustrating the sensors on the boat.

2.1. SIDE-SCAN SONAR

The main data-gathering sensors on the vessel are Blueprint Subsea Starfish SSS sensors. These are mounted to a custom-built moving frame, shown in Figure 1b, that allows them to be deployed below the level of the hulls to maximize the scanned area. The vertical beamwidth of the sensors is 60°. The frequencies of the pairs of sensors are 450kHz and 1MHz respectively, which allows for different levels of detail in the data. The lower frequency sonar has better attenuation characteristics resulting in long range measurements and some penetration into the floor, whereas the higher frequency sonar results in images of higher resolution with limited imaging range.

2.2. FORWARD LOOKING SONAR

The FLS mounted on the front of the vessel is a Blueprint Subsea Oculus M750d, shown in Figure 1a. This compact multibeam imaging sonar has a switchable operating frequency of 750kHz or 1.2MHz, with the lower offering a wide 130° horizontal aperture, and the higher offering a narrower 80° aperture but better resolution. This makes it well suited for detailed inspection of areas of interest.

2.3. LIDAR

The lidar is an Ouster OS1 which has a 45° vertical field of view, up to 2048 horizontal resolution and an update rate of 10Hz. As well as lidar data, this also gathers ambient near-infrared light providing an image similar to a 360° monochromatic camera. This data is used to detect the shore or obstacles that would be in the path of the vessel, to aid in vessel navigation.

2.4. NAVIGATION

The onboard navigation system consists of a CubePilot Cube Orange+ autopilot controller which has three built-in inertial measurement units (IMUs) and is connected to dual differential global navigation satellite system (GNSS) antenna, which provide both position and heading. This is augmented by a static GNSS antenna at the base station that provides real-time kinematics (RTK) corrections to enable high accuracy positional estimates. This navigation system also enables the boat to be controlled over radio telemetry by a handheld controller to override autonomous operation whenever required.

2.5. COMPUTING

Inside the watertight compartment of the ASV, the main computer is an Intel NUC with an i7-10710U, 32GB RAM and 1TB of removable data storage. This interfaces with all the platform sensors and performs onboard processing of raw data. The computer can be accessed remotely via a base station connected locally over long-range WiFi.

3. SOFTWARE

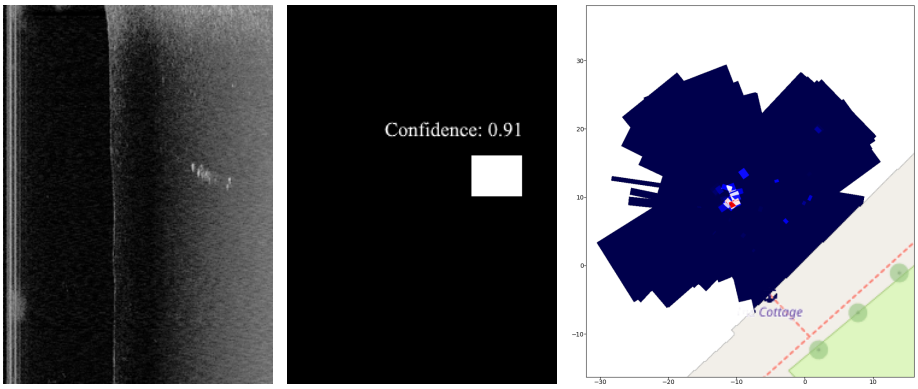
The software developed for control and data processing uses the ROS2 middleware as the framework to handle the passing of data between different processing nodes. This makes it straightforward to distribute processing across different computers, even across a WiFi network.

3.1. NAVIGATION AND GUIDANCE

Navigation is handled by the open source ArduPilot Rover firmware installed on the Cube-Orange+ controller. This is controlled either via Mission Planner software installed on a locally connected computer, or remotely via the MAVLink protocol using software such as QGround-Control or the mavros package within ROS2. These are used to access navigational information or push new waypoints to the autopilot.

3.2. DATA PROCESSING

The SSS data are used to develop a situation map that contains areas of interest. The boat follows a lawnmower pattern, such as that shown in Figure 3(b). The SSS data are accumulated as the vessel travels along a straight path. When it reaches a waypoint, a partition of data is returned, Figure 2a. This is projected onto the ground plane and contrast-stretched, before being processed with a Faster R-CNN[2] to estimate areas of interest in the partition, Figure 2b. The model for the Faster R-CNN was trained with 297 annotated images of data gathered from previous experiments of the same target[3]. Finally these are oriented appropriately using the GNSS data from the track, aggregated and then displayed on a map, resulting in Figure 2c.



(a) (b) (c)

Figure 2: (a) A contrast-stretched SSS data partition. (b) The corresponding detection from the Faster R-CNN. (c) The situation map accumulated from a complete mission involving multiple lawnmower patterns. The color corresponds to the cumulative confidence of each detection in that pixel.

3.3. CONTROL INTERFACE

To interact with the vessel, an interface was developed that can be run on a shoreside laptop connected to the boat over WiFi, Figure 3. This can be used to control the parameters of the sonar sensors, (a), and view the raw data feed from them to check they are operational. The map in the center of the interface has various layers that can show how much of the area has been covered and how many times, or the situation map which is updated in real-time as the boat carries out a survey. This can be interacted with by hovering over an area with the cursor which will bring up any detections in that area in a temporary window, as shown in Figure 3(c). This allows an operator to inspect the detections and the confidence value of the detection algorithm, which can then be used to inform the location of subsequent waypoints. Waypoints can be added by clicking on desired locations on the map, or by drawing an arbitrary shape on the map and filling it with a lawnmower pattern with a set spacing and angle, Figure 3(b).

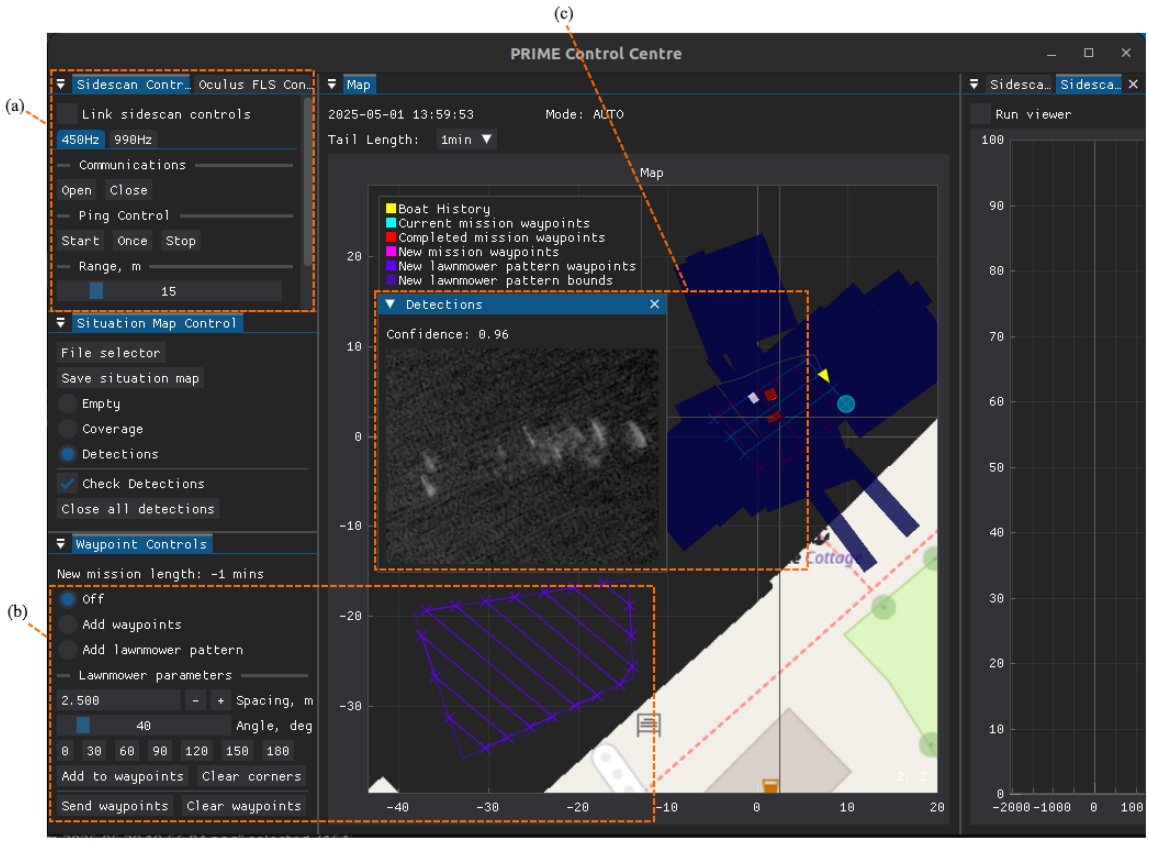


Figure 3: The control interface running in real-time during a field experiment. (a) Control of the sonar sensors. (b) Mission waypoint generation. (c) Detection inspector.

4. FIELD EXPERIMENTATION

The vessel and software has been tested and developed in various locations, one of which is Underfall Yard in Bristol, UK, shown in Figure 4a. The body shaped target, a clothed plastic mannequin shown in Figure 4b, was weighted and deployed to the bed of the river. This location was then surveyed using multiple lawnmower patterns set up for different view angles of the same region, running the software detailed in section 3.2. Figure 2a shows the data



(a) *(b)*

Figure 4: (a) PRIME during a field trial in Bristol. (b) The target mannequin prior to sinking.

from one leg of the lawnmower pattern in which the mannequin is present, and figure 2b shows the corresponding detection estimate from the detection algorithm. Figure 2c shows the resulting situation map from the accumulated data from the entire mission, showing some scattered false detections and numerous high confidence correct detections in the same spot, where the mannequin was deployed.

5. CONCLUSIONS & FUTURE WORK

This paper has detailed the progress made towards an autonomous survey vessel for aiding divers in an emergency response situation. The data processing chain has been demonstrated to produce a real-time situation map that can be used to make informed decisions about where to prioritize inspection.

Further developments include using the lidar data for obstacle avoidance and to use human-like machine learning to plan and adjust missions in response to changing environment.

6. ACKNOWLEDGEMENTS

This work was funded by EPSRC grant EP/X030156/1. We would also like to thank Bristol Harbour Master for enabling field experiments to be carried out at Underfall Yard, and DSTL for enabling field experiments at Stanborough Park.

REFERENCES

[1] Z. Rymansaib, B. Thomas, A. A. Treloar, B. Metcalfe, P. Wilson, and A. Hunter, “A prototype autonomous robot for underwater crime scene investigation and emergency response,” *Journal of Field Robotics*, vol. 40, no. 5, pp. 983–1002, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.22164

[2] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” 2016. [Online]. Available: https://arxiv.org/abs/1506.01497

[3] Z. Rymansaib, Y. Nga, A. Anthony Treloar, and A. Hunter, “Sidescan sonar images for training automated recognition of submerged body-like objects,” October 2024. [Online]. Available: https://researchdata.bath.ac.uk/1467/

Reinforcement Learning for Marine Path Planning

Edward Clark[1], Alfie Anthony Treloar[1], Alireza Tamaddoni Nezhad[2], and Alan Hunter[1]

[1]University of Bath, Bath BA2 7AY, United Kingdom
[2]University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom

Contact author: Edward Clark, eprc20@bath.ac.uk

Abstract: This paper investigates the use of reinforcement learning (RL) for marine path planning, focusing on active acoustic imaging with a side scan sonar. This is a high-level path planning problems in simulated marine environments, where the agent must learn to maximize coverage. The agent plans paths to maximize unique area coverage and target detection, considering sensor limitations such as nadir gaps and varying detection probabilities. We employ Proximal Policy Optimization (PPO) as the RL algorithm and systematically analyze the impact of key hyperparameters using Optuna for optimization. The study demonstrates that curriculum learning—progressively training agents from simpler to more complex environments—significantly improves learning efficiency and final performance. Our results highlight the importance of reward function design, state and action space formulation, and hyperparameter selection in achieving robust RL-based path planning. The findings provide practical guidance for deploying RL in real-world marine scenarios and underscore the value of curriculum learning and hyperparameter optimization in overcoming training challenges.

Keywords: *Machine Learning, Reinforcement learning, Path planning, Active acoustic imaging, Passive acoustic detection, Curriculum learning*

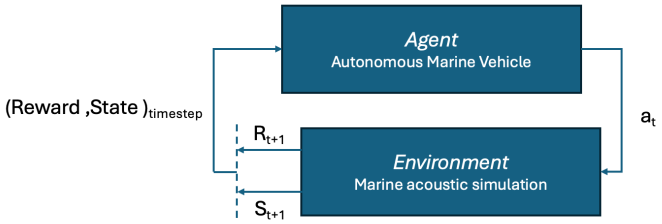


Figure 1: The RL cycle modified from Sutton and Barto [5]. The agent interacts with the environment, taking actions and receiving rewards and an updated state. The agent then updates its policy based on the rewards it receives.

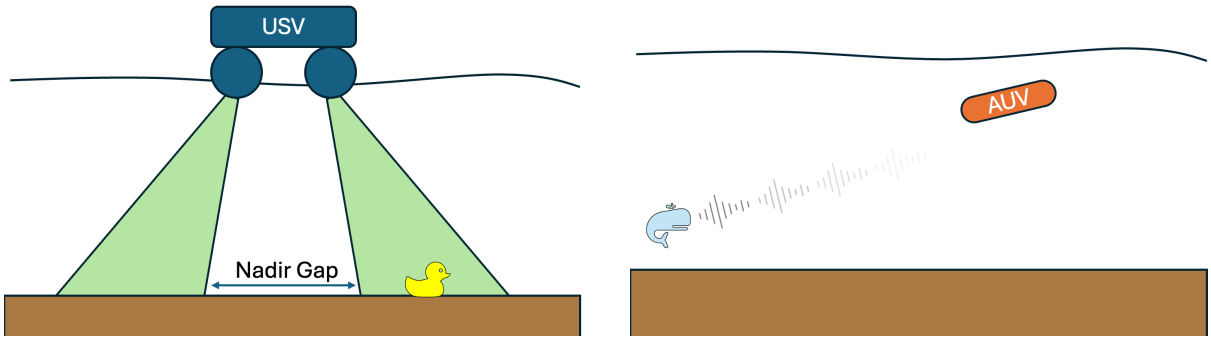
1. INTRODUCTION

Path planning is key to the success of many marine operations, including marine mammal monitoring, seabed investigation and many other tasks. Increasingly these tasks are being performed by autonomous vehicles, which require path planning algorithms to operate effectively. [2, 8, 3] The marine environment is complex and dynamic, with many factors that can affect the success of a path planning algorithm. Reinforcement learning (RL) has been shown to be effective in many path planning tasks, including robotics and autonomous vehicles [1, 3, 7]. It can learn from simulated environments and then transfer this knowledge to real-world scenarios. Marine path planning is expensive to run in the real world, and RL can help to reduce the cost of testing by allowing for simulation-based training. This paper explores the use of RL for marine path planning, with a focus on two case studies: active and passive sonar sensor path planning.

Reinforcement learning (RL) is a type of machine learning that allows an agent to learn from its environment by taking actions and receiving rewards. For a thorough introduction to RL, see Sutton and Barto.[5] An agent follows a policy, which is a mapping from states to actions, and receives a reward signal based on the actions it takes. In our case, the agent is a marine vehicle that is trying to plan a path through an environment to achieve a goal, such as detecting a target or classifying a signal. This is shown in Figure 1, which illustrates the RL cycle. Here a_t is selected by the agent based on its policy π, which maps the current state s_t to an action a_t. It follows that the agents policy is the set of 'instructions' it follows depending on what it observes, to achieve its goal such as turn to port or starboard, dive deeper or ascend. Previous work has looked at the application of RL to passive acoustic detection as seen in Figure 2b[2]. This paper extends the number of marine environments that RL has been successfully applied to for path planning, by investigating active acoustic imaging.

This paper aims to highlight the main considerations when using RL for marine path planning. This paper specifically focuses on PPO (Proximal Policy Optimization) as the RL algorithm, as it has been shown to be effective in many path planning tasks. We demonstrate how important hyperparameters such as batch size, learning rate and discount factor can affect the performance of the RL algorithm. We also explore how curriculum learning can be used to improve the performance of the RL algorithm by gradually increasing the difficulty of the task. Finally, we discuss the results and implications of our findings for future work in marine path planning using RL.

2. ACTIVE ACOUSTIC IMAGING



(a) Active acoustic Imaging *(b) Passive acoustic detection*

Figure 2: Examples of the active and passive acoustic path planning setups. (a) The agent is a USV with a side scan sonar (SSS) mounted on the bottom. The SAS has a fixed range depending on the water depth and a nadir gap where it cannot detect targets. The agent must plan a path to cover the area and detect targets. (b) The agent is an uncrewed underwater vehicle (UUV) with a hydrophone array. The UUV must plan a path to detect targets in the environment.

The active acoustic path planning is a task where an agent must plan to use a side scan sonar (SSS) to detect targets in a simulated environment. This task was developed to train a model (read policy) to deploy on a small (< 2m) unmanned surface vessel (USV) to detect small targets and seabed features.

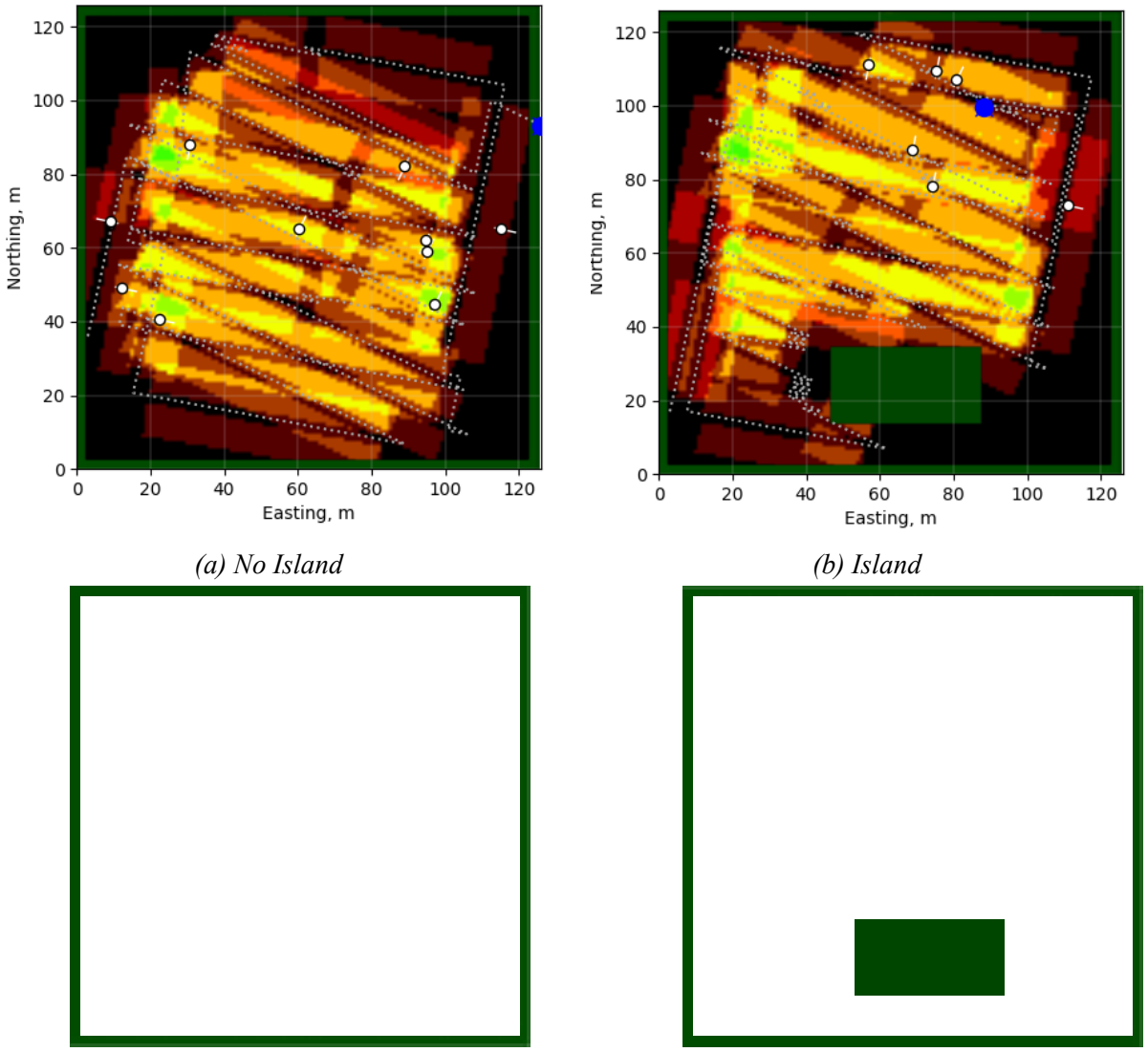
As shown in Figure 2a, the agent is a USV with a SSS mounted on the bottom. As the USV moves past a target it will detect the target with some probability, which is dependent on the target size, seabed type and relative position of the target to the USV. For example a long thin target side on be easily detected but end on it may not be detected at all. The agent must plan a 2D path to cover the area and detect targets, while covering the nadir gap where it cannot detect targets. A typical expert planning such a mission would also try to maximise the number of different angles the SSS sees the target from, to improve detection probability.

2.0.1. REWARD FUNCTION

To encourage the agent to make incremental gains in coverage, we employ an incremental coverage reward function. At each step, the reward is computed as the difference between the current coverage and the coverage at the previous step, i.e.,

$$r_n = \max(0, \text{coverage}_n - \text{coverage}_{n-1}) \tag{1}$$

where coverage_n is the area scanned by the agent at timestep n as a fraction of the total area. This formulation ensures that the agent is rewarded only for newly covered areas, promoting exploration and discouraging redundant actions. Negative differences (i.e., reductions in coverage) are clipped to zero, so the agent is not penalized for revisiting already covered regions but is incentivized to maximize unique coverage over time. This reward structure is particularly effective in environments where the objective is to maximize the area explored or surveyed.



(a) No Island *(b) Island*

(c) No Island *(d) Island*

Figure 3: Illustration of the two environments and comparison of the planned paths and corresponding coverage maps. The dark green regions indicate the bounds and obstacles in the environment. Coverage is shown by the shaded regions from a scale from red through yellow to green depending on the number of directions it has been scanned from. Contacts are shown as large white dots and the agents location is shown as a blue dot. The path of the agent is shown as a dotted white line. These paths are from the optimal RL learnt policy.

2.0.2. STATE & ACTION SPACE

The agent's state space is a 64x64x3 grid, where each cell represents a 2D area of the environment. The three channels represent the following:

- **Coverage Map:** A map indicating whether a cell has been covered by the agent.
- **Location Map:** A map indicating the location of the agent
- **Occupancy Map:** A map indicating the environment such as bounds and obstacles, where the agent cannot go.

Figure 3 shows an example of the three channels combined into a single image. In reality the maps are larger, but they are downsampled to 64x64 for computational efficiency. Using non square maps is also possible as these are padded before down or upsampling so the trained model can be used in different environments.

The action space is a discrete set of actions that the agent can take at each step. The agent can move in one of any 60 directions, giving it a 6 degree resolution.

3. METHODOLOGY

3.1. RL ALGORITHM

Proximal Policy Optimization (PPO) was used as the RL algorithm for training [4]. PPO is a policy gradient method that has been shown to be effective in many path planning tasks . There are a number of hyperparamters that should be tuned for PPO:

Table 1: Key Hyperparameters of Proximal Policy Optimization (PPO)

Hyperparameter	Description
N steps	Number of steps to run for each environment per update.
Batch Size	Number of samples used for each policy update.
Epochs	Number of updates at each learning phase
Discount Factor (γ)	Determines the importance of future rewards.
Learning Rate	Step size for updating the policy and value networks.
GAE Lambda (λ)	Controls bias-variance tradeoff in advantage estimation.
Clip Range	Limits policy update size to ensure stable learning.
Entropy Coefficient	Encourages exploration by adding entropy to the loss.
Value Function Coefficient	Scales the value loss in the total loss function.
Max Gradient Norm	Maximum value for gradient clipping to improve stability.

3.2. HYPERPARAMETER OPTIMISATION

Optuna is a hyperparameter optimisation framework that allows us to define a search space and an objective function to optimize. The search space is defined as a set of hyperparameters

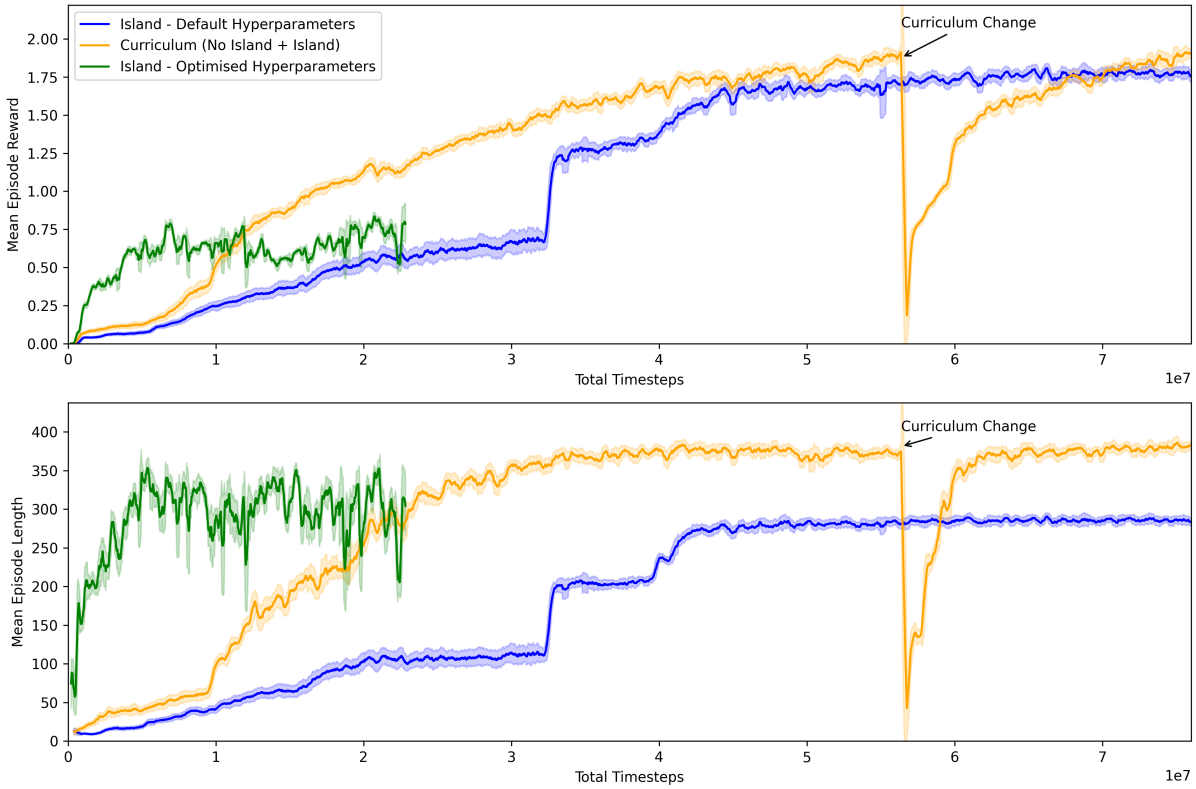


Figure 4: Learning curves for the active acoustic path planning task. For each curve the mean and standard deviation of the reward over a 200 episode rolling window is shown. The blue line shows learning with default hyperparameters in the Island environment. The orange line shows curriculum learning, where the agent is first trained in a simple environment with no islands and then transferred to the more complex environment with an island. The yellow line shows learning with optimised hyperparameters in the Island environment.

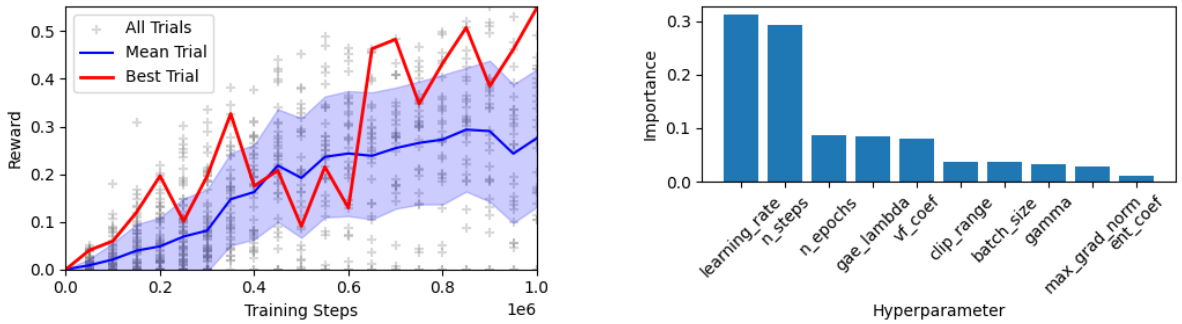
and their ranges, as shown in Table 1. The objective function is defined as the average reward over a number of episodes, which is used to evaluate the performance of the RL algorithm.

A TPE (Tree-structured Parzen Estimator) sampler was used to sample hyperparameters from the search space. This is a common sampler used in Optuna that is based on Bayesian optimization. 70 trials were computed which were allowed to run for 1 million steps. Trials were pruned using median pruning after 200,000 steps, which is a common practice to reduce the number of trials that are run to completion. This is where the reward is compared to the median reward of all trials at that step, and if it is below the median, the trial is pruned. The hyper parameter importance is calculated using the PED-ANOVA method [6], which is a method for calculating the importance of hyperparameters based on the variance of the reward.

4. RESULTS & DISCUSSION

The results of the training with and without curriculum learning are shown in Figure 5a. Three results are shown, the first is the performance of the RL algorithm with default hyperparameters in the Island environment. The second is the performance of the RL algorithm with curriculum learning, where the agent is first trained in a simple environment with no islands and

then transferred to the more complex environment with an island. The third is the performance of the RL algorithm with optimised hyperparameters in the Island environment. It is clear that curriculum learning improves the performance of the RL algorithm, as it is able to adapt to the new environment and improve its performance relative to simply training in the more complex environment.



(a) Hyper parameter optimisation *(b) PED-ANOVA Hyperparameter importance*

Figure 5: Hyper parameter tuning metrics(a) Results from 70 trials of hyper parameter optimisation. Raw data is shown in grey, the mean and standard deviation in blue. The best performing trial is shown in red. (b) Importance of hyperparameters across all trial runs. The importance is calculated using the PED-ANOVA method

The process for determining the best hyperparameters is shown in Figure 5. We see there are a number of trials that gain 0 reward at any stage of training. This implies that incorrect hyperparameter choice can lead to the RL algorithm not learning at all. The best performing trial is shown in red, and the mean and standard deviation of the all trials is shown in blue. Even though the best trial dips below the mean at some points, it is still able to achieve a higher reward than the mean at the end of training.

The hyperparameter importance is shown in Figure 5b. The most important hyperparameters are the the number of steps and the learning rate. The fact that the number of steps is the most important hyperparameter is not surprising, as it determines the diversity of the training data. The learning rate is also important, as it determines how quickly the agent learns from the environment. Too high a learning rate can lead to instability in the training process, while too low a learning rate can lead to slow convergence.

5. CONCLUSION

This paper has explored the use of reinforcement learning for marine path planning, with a focus on active acoustic imaging. We have shown how the environment and the task affect the RL approach, with differences in the state space, action space, reward function and hyperparameters. We have shown that curriculum learning can be used to improve the performance of the RL algorithm by gradually increasing the difficulty of the task. Hyperparameter optimisation is also important for the performance of the RL algorithm, with the number of steps and learning rate being the most important hyperparameters for the active acoustic task.

REFERENCES

[1] Ignacio Carlucho, Mariano De Paula, Sen Wang, Yvan Petillot, and Gerardo G. Acosta. Adaptive low-level control of autonomous underwater vehicles using deep reinforcement learning. *Robotics and Autonomous Systems*, 107:71–86, September 2018.

[2] Edward Clark, Alan Hunter, Olga Isupova, and Marcus Donnelly. Optimising sensor path planning with reinforcement learning and passive sonar modelling. In *7th Underwater Acoustics Conference and Exhibition, UACE 2023*, pages 411–418, 2023.

[3] Zhaolun Li and Xiaonan Luo. Autonomous underwater vehicles (AUVs) path planning based on Deep Reinforcement Learning. In *2021 11th International Conference on Intelligent Control and Information Processing (ICICIP)*, pages 125–129, December 2021.

[4] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms, August 2017.

[5] Richard S Sutton and Andrew G Barto. Reinforcement Learning: An Introduction. page 352.

[6] Shuhei Watanabe, Archit Bansal, and Frank Hutter. PED-ANOVA: Efficiently Quantifying Hyperparameter Importance in Arbitrary Subspaces, May 2023.

[7] Meng Xi, Jiachen Yang, Jiabao Wen, Hankai Liu, Yang Li, and Houbing Herbert Song. Comprehensive Ocean Information-Enabled AUV Path Planning Via Reinforcement Learning. *IEEE Internet of Things Journal*, 9(18):17440–17451, September 2022.

[8] Changxi You, Jianbo Lu, Dimitar Filev, and Panagiotis Tsiotras. Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning. *Robotics and Autonomous Systems*, 114:1–18, April 2019.

Acoustic Surveillance Using an Unmanned Blended Wing Glider

Scott Mahar[1], Gerald D'Spain[2]

[1]3550 General Atomics Court, San Diego, CA 92121
[2]Retired Scripps Institution of Oceanography, La Jolla, CA

Scott Mahar, 3550 General Atomics Court, San Diego, CA 92121, Scott.Mahar@GA.com

Abstract: *Unmanned underwater vehicles (UUV) are efficient task multipliers, able to perform underwater surveillance and monitoring, while freeing manned platforms to perform more complex work. Through optimizing the UUV as a blended wing glider with forward propulsion provided by an intermittently operated buoyancy engine, the acoustic and electromagnetic signatures are drastically reduced, while the endurance and range of the vehicle is significantly increased. Additionally, higher energy density systems such as fuel cell systems can further advance the surveillance capabilities of a UUV.*

General Atomics Electromagnetic Systems (GA-EMS) is developing and testing a novel fuel cell powered UUV: a large, blended wing glider, optimized for stealth with extended transit and zero surface expression. GA-EMS has hybridized an existing 20-foot blended wing glider with a fuel cell system and is presently testing this system at-sea. The glider's wingspan enables spatially separated towed arrays, as well as fixed horizontal and vertical arrays on the leading edges and in the tail. Data from at-sea testing will be presented, showing the ability of the blended-wing glider to resolve signal location and elevation angle of a controlled underwater source.

GA-EMS is also optimizing and fabricating a next generation 20-foot blended wing glider, due to be completed in 2025. At this wingspan, the glider can be easily handled, launched, and recovered for a modest cost, while still performing relevant surveillance missions; a

much larger glider is envisioned for a final product. This next generation glider will include a sizable payload volume with the ability to transport and deploy novel surveillance systems. Aspects of the design of this next-generation glider will be presented.

Keywords: *UUV, Glider, surveillance, hydrophone, LiFT^{TM}, fuel cell, hydrodynamic, range, persistence*

EXECUTIVE SUMMARY

Blended-wing gliders are the ideal unmanned underwater vehicles (UUV) for persistent acoustic surveillance due to the lack of a continuously running propulsor and low hydrodynamic noise. Optimization of the glider's shape and integration of novel energy-dense systems enhance the persistence of the glider. General Atomics Electromagnetic Systems (GA-EMS) has been developing power and energy systems for UUVs and working with Scripps Institution of Oceanography (SIO) Marine Physical Laboratory (MPL) at the University of California San Diego (UCSD) to optimize the design and acoustic sensing capabilities blended wing gliders. The team continues to modify and test the existing 20-foot ZRay glider, as well as design and fabricate a next generation glider (GRay). The team is working to scale up the vehicle to a large wingspan which will enable large acoustic sensor layouts in multiple dimensions, leading to novel acoustic sensing capabilities.

INTRODUCTION

Buoyancy-driven gliders propel themselves by changing the buoyancy of the vehicle, either by changing the mass (amount of water in ballast tanks), or by changing the volume of the vehicle using bladders. This change in buoyancy changes the vehicle's potential energy in the earth's gravitational field, which is converted to kinetic energy as the glider descends (or ascends) through the water column. At the top of the dive cycle, the vehicle pumps in water (or reduces its volume) to make it denser than the surrounding water and begins its descent. At the bottom of the dive cycle, the vehicle pumps water out of the ballast tanks (or increases its volume) to be less dense than the surrounding water and ascends.

In 2004, SIO partnered with Applied Physics Lab at the University of Washington under sponsorship of the Office of Naval Research to design, fabricate, and test blended wing gliders [1]. Throughout the development program, the XRay glider was tested from 2006 to 2008, and the ZRay glider was tested from 2010 to 2014. As shown in Fig. 1, the ZRay glider was able to demonstrate a lift-to-drag (L/D) ratio of 20:1 during at sea testing.

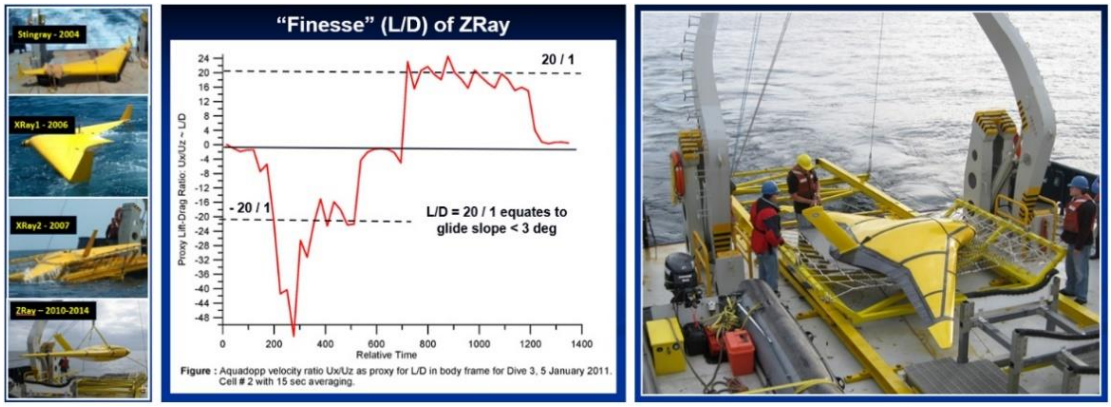


Fig. 1: The blended wing glider program led by SIO tested several gliders (left) and was able to verify the ZRay's L/D ratio of 20:1 (middle) during at-sea testing (right).

GA-EMS has been working with SIO to integrate an advanced power and energy system featuring a fuel cell and a GA-EMS pressure tolerant battery. This upgrade, as well as a modified hydrophone layout, was tested at sea in 2024 and 2025.

OPTIMIZATION OF UNDERSEA ACOUSTIC SURVEILLANCE PLATFORM

Silent operation and extended persistence of a UUV are two of the most important features of an undersea acoustic surveillance platform. Gliders are inherently quieter than prop-driven UUVs since the duty cycle of the pumps are 2-3% of the duty cycle of a propeller. Extended persistence is accomplished by optimizing the UUV's hydrodynamic efficiency and by implementing a power and energy system with a high energy density.

Maximizing the transit efficiency is accomplished hydrodynamically by minimizing the amount of lost kinetic energy. A hydrodynamically optimized winged glider leads to significantly less drag than a standard propeller-driven UUV [2]. Further hydrodynamic optimization of a glider is based on the mission goals. Gliders generally fall into one of two classes: 1) "profiling gliders" that have steep dive angles and are generally used to sample water column data; and 2) "cross country gliders" that are optimized to have the greatest horizontal transport efficiency. The "profiling gliders" such as the Spray [3], Seaglider [4], and Slocum [5] are based on a body of revolution with wings. A "cross country glider" with a blended wing utilizes the entire outer surface of the glider to generate lift. Since the lift-to-drag ratio is inversely proportional to the energy required to transit a certain horizontal distance, the vehicle's lift-to-drag ratio needs to be maximized to minimize the propulsion energy. GA-EMS developed a computational-fluid-dynamics-based software package to analyze glider performance as a function of 15 non-dimensionalized parameters. Utilizing a mix of proprietary and industry standard tools, a glider optimization tool was developed. This tool agrees qualitatively with previous analytical results in the literature [2], showing that blended wing gliders exhibit up to 40% better range than body-of-revolution gliders, whereas body-of-revolution gliders are better for faster speeds.

In addition to the optimization of the vehicle's form factor, persistence can be increased through maximizing the vehicle's energy storage. Power hungry sensors and systems, and the desire for longer range and more time on station, drive up the energy demand. The ultimate goal of realizing unlimited persistence through energy harvesting is under investigation [5,6]; however, it is yet to be widely utilized in underwater vehicles. The added complexity, size, and weight of the energy harvesting system needs to be accounted for, and the source of the energy harvesting will dictate where (geographically and in the water column) and how the vehicle can operate. The added complexity, strain on vehicle requirements, and operational limitations have generally led to battery powered vehicles.

GA-EMS specializes in developing safe, reliable power and energy systems for extreme environments. GA-EMS has developed a Lithium-ion Fault Tolerant (LiFT^{TM}) battery that prevents uncontrolled cascading failure, and a pressure tolerant battery (Fig. 2).

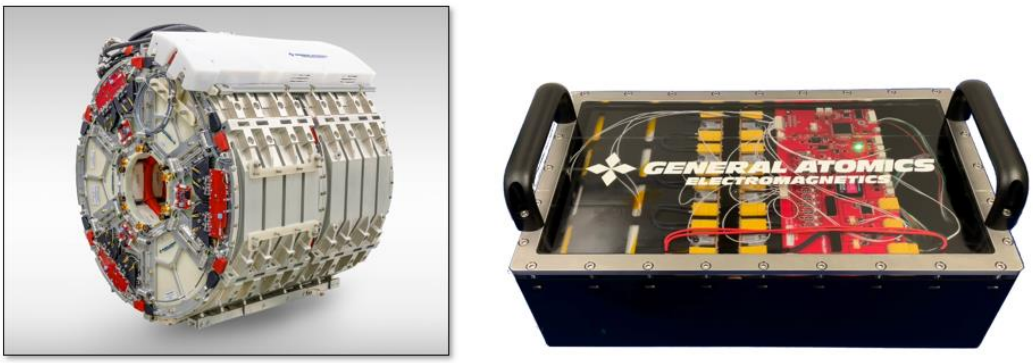


Fig. 2: The GA-EMS LiFT^{TM} battery (left), and pressure tolerant battery (right).

The LiFT^{TM} battery ensures the safety of personnel and equipment while keeping power available for high mission assurance. The assumption is that at some point, a lithium-ion cell is going to go into thermal runaway, and the LiFT^{TM} system will protect the remainder of the battery (and vehicle) from a catastrophic failure. The LiFT^{TM} system has been approved for use by the U.S. Navy for certain platforms.

GA-EMS has also developed a pressure tolerant battery capable of operation at full ocean depth. The absence of a pressure vessel enables an energy dense battery that can be conformal to a desired shape. This battery utilizes a dielectric fluid that aids with cell cooling, while also preventing fires and thermal runaway in the case of a cell failure. The fluid is also less dense that water, reducing the battery's overall weight in water.

Fuel cell systems have been under development for decades and have recently experienced exponential growth in the commercial sector driven by the transportation industry. A fuel cell system stores several times the energy of a battery of comparable size and weight. GA-EMS used a 300 W Baltic fuel cell and built the balance of plant to feed it reactants and manage the products. This custom fuel cell system, and the GA-EMS pressure tolerant battery were integrated into the ZRay and tested in February of 2025 (Fig. 3).

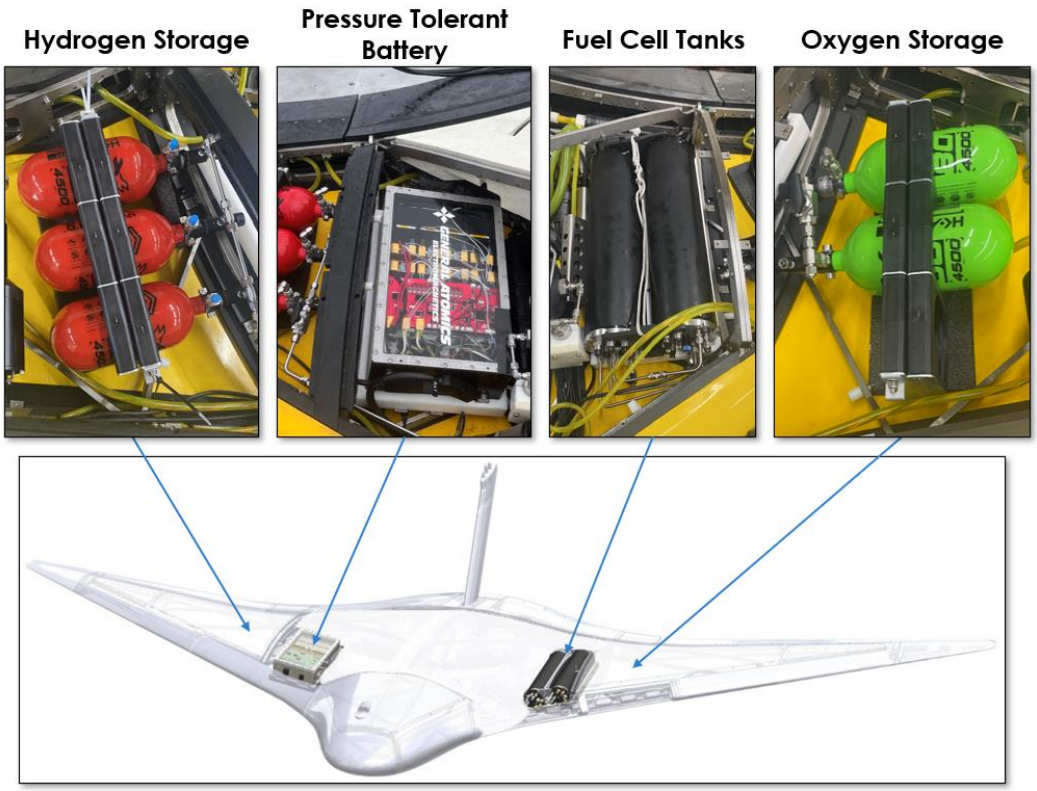


Fig. 3: ZRay hybridized with a GA-EMS pressure tolerant battery and fuel cell system.

Leveraging decades of work in fuel cell and battery development, this hybridized power and energy system significantly enhanced the energy density of the UUV. These advanced power and energy systems paired with the blended wing glider body enable an acoustic sensing platform with unprecidented persistence.

THE GLIDER AS AN ACOUSTICS PLATFORM

The ZRay glider was outfitted with passive hydrophone arrays inside the outer shell (see Fig. 4) for at-sea testing to demonstrate its acoustic surveillance capabilities.

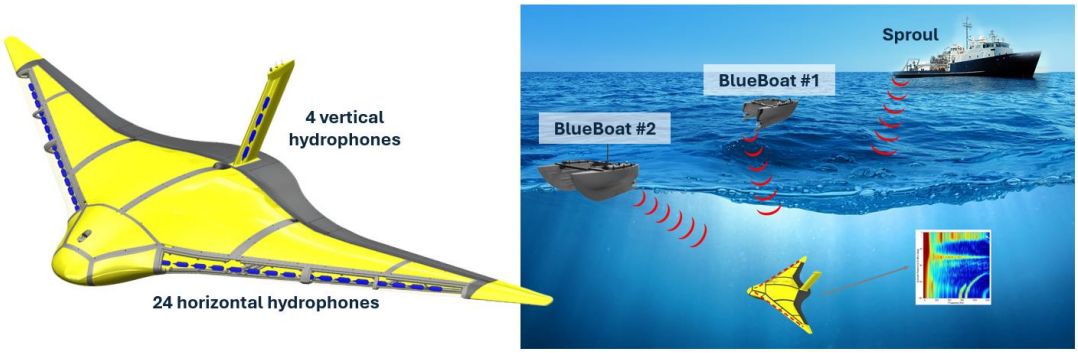


Fig. 4: The ZRay has been outfitted with multiple passive acoustic sensor systems.

The hydrophones are installed with 0.15 m inter-element spacing (half-wavelength spacing at 5 kHz). Each element is individually wrapped in open-cell foam to decrease the effects of non-acoustic pressure fluctuations (see below). The data acquisition system calibration curve in amplitude and phase has been calculated theoretically and parts of it have been confirmed by laboratory measurements. The effects of the analog to digital (A/D) converter settling time (i.e., the time delay in acquiring data samples across the array) are negligible up to 5 kHz, as determined by comparing results using time-aligned versus non-aligned time series input. Electronic self-noise of the system has been measured at sea and is more than 15 dB below ocean background noise above 7 Hz. A/D converter quantization noise is another 15 dB or more below electronic self-noise and so can be ignored.

The performance of passive acoustic arrays that move relative to the surrounding water can be adversely affected at low frequencies by noise from turbulent pressure fluctuations. This flow noise has much smaller spatial scales of coherence than low-frequency sound, so that coherence squared estimates between pairs of array elements can be used as a diagnostic of this contamination. As an example, Fig. 5 shows the spatial coherence squared levels for three time periods in the May, 2024 sea test, when the glider speed through the water was 0.14 m/sec (left), 0.4 m/sec (middle), and 0.5 m/sec (right). Note that coherence squared between any element and itself is unity across the band, explaining the horizontal solid red bar at reference element 9 in the three plots. In addition, element 1 was non-functional and element 28 was a non-acoustic electronically-equivalent dummy to measure electronic self-noise, resulting in the horizontal dark blue bars at the top and bottom of each plot.

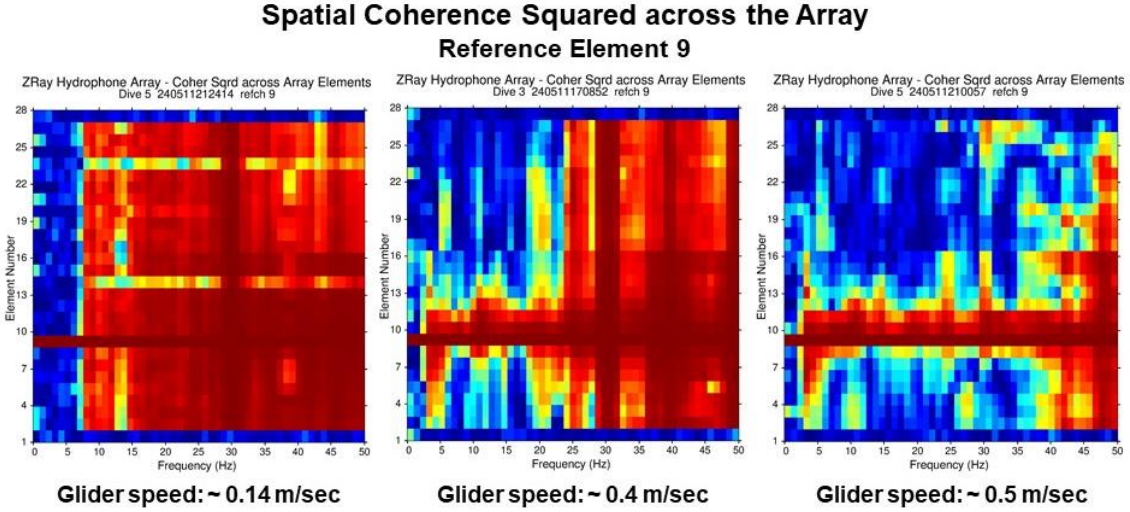


Fig. 5: Spatial coherence squared estimates between element 9 and all other hydrophones in the leading-edge array (elements 1-28 along the vertical axis) as a function of frequency (horizontal axis) from DC up to 50 Hz for three time periods in the May, 2024 sea test. The coherence squared values vary from 0 (dark blue) to 1 (dark red).

The level of coherence drops off rapidly with distance away from the reference (element 9) at the lowest frequencies (the blue areas above and below the dark horizontal red bar), and the frequency band of this rapid decay increases with increasing glider speed. The pressure spectral density curves over this low-coherence band are well fit by Eq. (8) of Strasberg, 1979, i.e.:

$$S_P(f) = 119 + 37 \bullet \log_{10}(U) - 27 \bullet \log_{10}(f) \quad (1)$$

which is derived assuming that pre-existing ocean turbulence of a given level (obtained from published ocean turbulence measurements) is advected past the hydrophones by fluid flow with relative speed *U*. In this equation, *U* has units of kts and the spectral density has the standard units of dB re 1 μPa²/Hz. Any turbulence generated by the glider itself is minimized by its wing shape and placement of the hydrophones in the leading edge.

Controlled acoustic sources (Lubell LL916) were deployed from autonomous surface vehicles ("BlueBoats") in all experiments to allow evaluation of array localization performance. These sources were suspended about 1 m below the sea surface and repeatedly transmitted a 60-sec waveform comprised of a) a set of 7 tones for 40 sec, followed by a 1-sec gap, and then b) an 18-sec-duration linear frequency modulated (LFM) upsweep, ending in a 1-sec gap. The latitude/longitude positions of these acoustic sources along with those of the ZRay glider and the R/V Sproul are plotted in Fig. 6 for two 1-min time periods in the Aug, 2024 (left plot) and Feb, 2025 (right plot) sea tests.

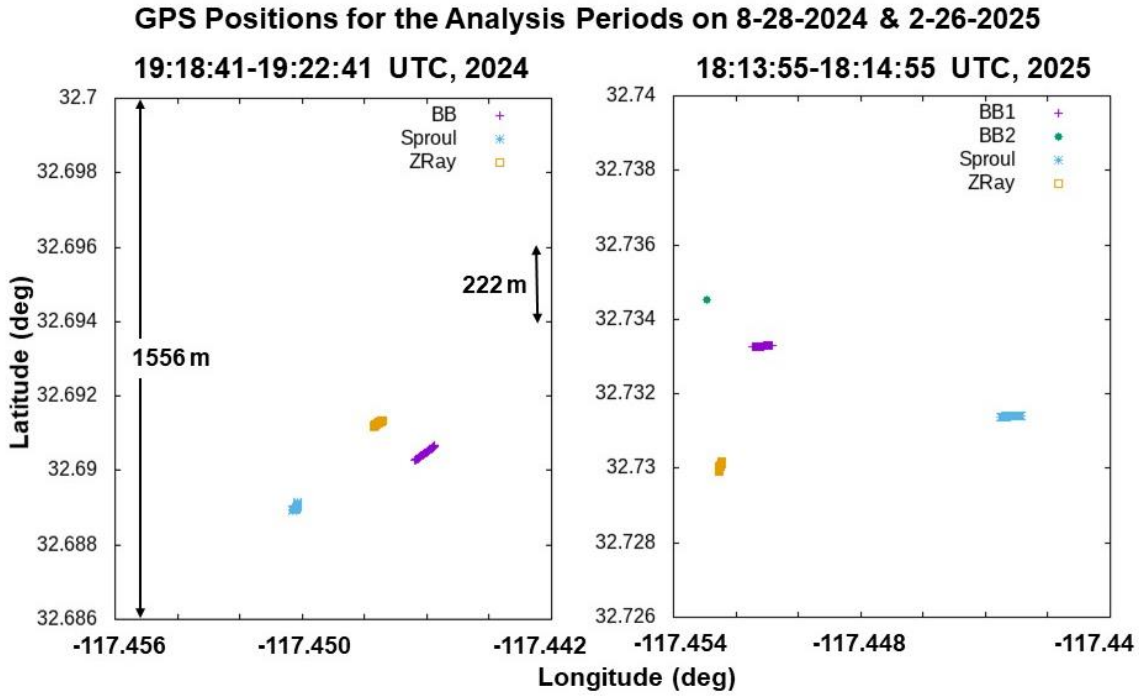


Fig. 6: The dead-reckoned positions in latitude and longitude of ZRay (gold), the BlueBoat vehicles with controlled acoustic sources (labelled "BB": purple, "BB1": purple, and "BB2": green), and the R/V Sproul GPS fixes (light blue).

Two BlueBoat ("BB") sources were deployed simultaneously during the 1-min time period in the February, 2025 sea test shown in the right plot in Fig. 6. These sources transmitted the same 40-sec comb of 7 tones, but starting at different times in the 60-sec cycle. In addition, their 18-sec LFM waveforms differed so that matched filtering of the

received time series allowed the 18-sec part of the 60-sec period that each source transmitted its LFM upsweep to be identified. Results of beamforming in azimuth at two tone frequencies over this 1-min period is shown in Fig. 7.

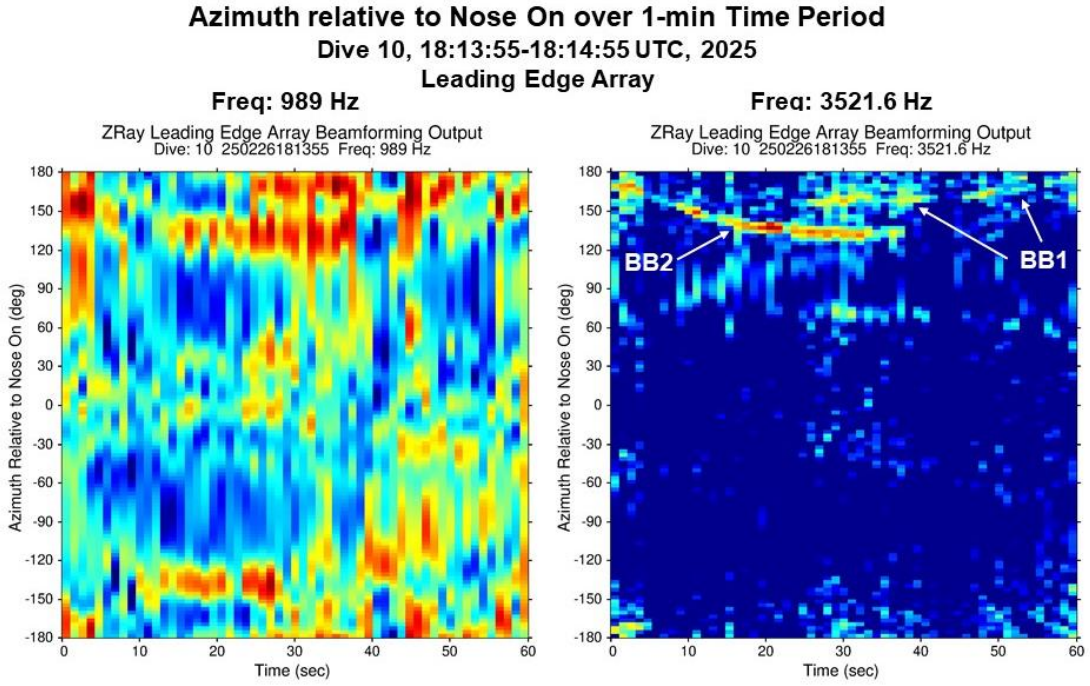


Fig. 7: White-noise-constrained adaptive beamforming output as a function of azimuth relative to glider heading (vertical axis) and time into a 1-min period (horizontal axis) at two tone frequencies – 989 Hz (left plot) and 3522 Hz (right plot). Positive azimuth angles represent angles to the starboard side and negative azimuths represent angles to port.

Two tracks on the starboard side of the glider (i.e., at the positive azimuth angles in the upper half of each plot) can be identified in the plots in Fig. 7. The azimuths of these tracks are consistent with the bearings of the BlueBoat acoustic sources from ZRay in Fig. 6 combined with the heading of ZRay to the south-southwest during this period. In fact, the sagging nature of the tracks in Fig. 7 is caused by the rotation of the glider heading from 180º true to 210º true and then back over this 1-min period. The matched-filter output indicates that BB1 transmitted its LFM upsweep over the initial third or so of the 60-sec period, so that it transmitted the 7-tone comb over the final two-thirds. In contrast, BB2 transmitted its LFM waveform over the final third and the 7-tone comb during the initial two-thirds. Therefore (as the labeling in white font in the right-hand plot of Fig. 7 indicates), the track at smaller azimuths over the initial portion of the 1-min period is that of the BB2 source, whereas the track near aft end-fire on the starboard side over the final two-thirds is the BB1 source. This separation in azimuth angle of the two sources again is consistent with the positions of the sources relative to the glider in the right-hand plot in Fig. 6.

In addition to the leading-edge hydrophone array, four hydrophone elements were placed inside the ZRay antenna mast to create vertical aperture, as illustrated in Fig. 4. An example of the beamforming output in elevation angle from this 4-element array is presented below. The data were collected over a 1-min period in the Aug, 2024 sea test. The left plot in Fig. 6 provides the locations of ZRay, the BlueBoat source, and the R/V Sproul over this 1-min period, showing the glider was less than 200 m from the source. The elevation angle of the arrival at 2623 Hz over this 1-min period is presented in Fig. 8.

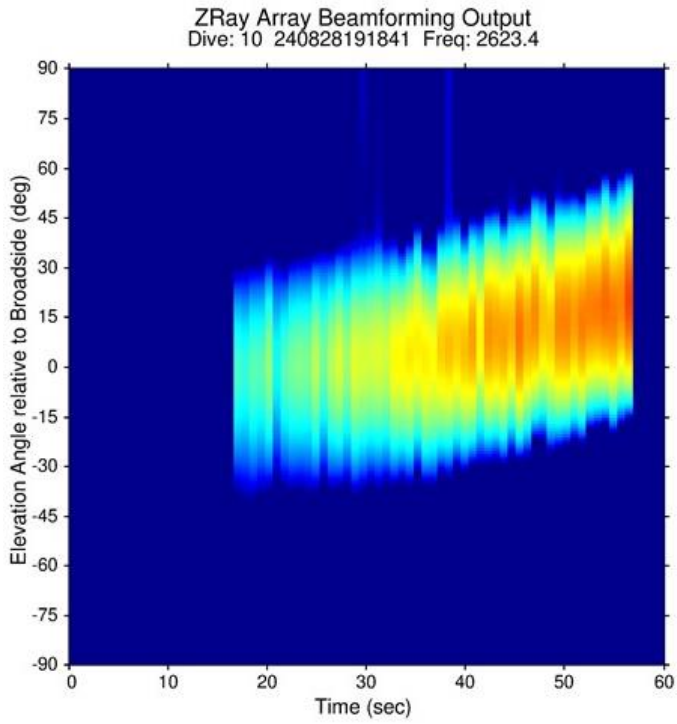


Fig. 8: White-noise-constrained adaptive beamforming output at a tone frequency of 2623 Hz as a function of elevation angle relative to the horizontal (vertical axis) and time into a 1-min period (horizontal axis). Positive elevation angles are upward looking and negative angles are downward looking.

The evolution of the arriving energy to higher elevation angles over time is partly due to the progressively increasing depth of the glider during this time, and partly due to a steadily increasing bias as a result of beamforming mismatch - the beamformer look direction in azimuth was fixed over the 1-min period, whereas the glider rotated in heading by nearly 30˚.

A detection/classification/localization/tracking (DCLT) algorithm string presently is being automated for real-time on-board array processing in future sea tests. It will allow rapid quantitative analyses of processing performance. In GA-EMS's next generation blended wing glider (GRay) outputs from this DCLT algorithm string will be communicated directly to the flight control computer to allow real-time adaptation of mission parameters based on the glider's measurements of the ocean sound field.

NEXT GENERATION GLIDER UNDER CONSTRUCTION

GA-EMS is presently designing and fabricating the next generation blended wing glider, GRay. This vehicle is comparable in wingspan to ZRay; however, advances in electronics, sensors, and batteries have enabled a significant payload bay in the GRay, as shown in Fig. 9. The body is slightly enlarged, while being mindful to minimize hydrodynamic degradations. The available cargo space is roughly 8 ft^3, cargo can be released through the top or bottom of the vehicle, and has 5, 12, 24, and 48 VDC available. This vehicle will be fabricated by the end of 2025, and at-sea testing will begin in 2026.

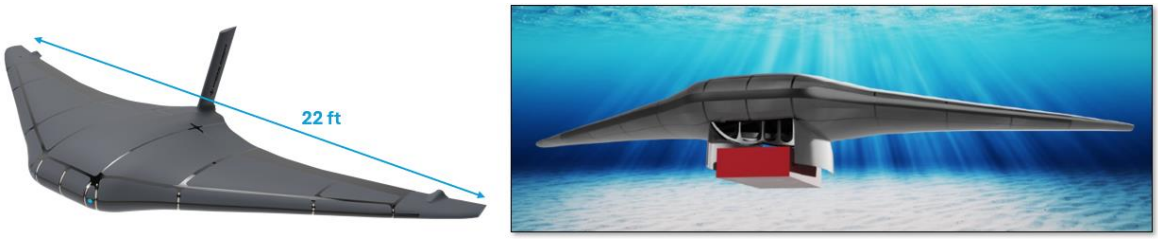


Fig. 9: The GA-EMS GRay glider being fabricated in 2025.

CONCLUSIONS

Persistence in underwater surveillance can be achieved through careful engineering design and implementation of appropriate subsystems. By optimizing the design of a blended wing glider, and integrating energy-dense energy storage systems, the blended wing glider can have extraordinary range and time on station. Due to how a buoyancy-driven glider operates, the vehicle is acoustically (and electromagnetically) virtually silent for the vast majority of the operational time, making it an ideal acoustic sensor platform. GA-EMS is working with SIO to hybridize the ZRay glider with a battery and fuel cell system, significantly increasing the persistence. GA-EMS is also fabricating a next generation GRay glider that will have a sizable payload bay for sensors and payloads. GA-EMS and SIO will be testing these vehicles at sea several times a year for the next few years, aiming to develop novel acoustic sensing capabilities. GA-EMS is looking for novel sensors that can be deployed and tested on the GRay to demonstrate future capabilities of a much larger glider.

ACKNOWLEDGEMENTS

The authors would like to thank GA-EMS for funding the recent glider work, as well as the technical team at SIO, particularly Dennis Rimington and Dave Price for their tireless work on the refurbishment and operation of the ZRay.

REFERENCES

[1] **G. L. D'Spain**, Final Report, https://apps.dtic.mil/sti/pdfs/ADA496168.pdf, *Office of Naval Research, 2009.*

[2] **Jenkins, S. A. and G. L. D'Spain**, *Springer Handbook of Ocean Engineering*, M. Dhanak and N. Xiros, Springer Cham, 301-321, 2016.

[3] **J. Sherman, R. Davis, W. Owens, J. Valdes**, The autonomous underwater glider Spray, *IEEE J. Oceanic Engin.,* 26 (4), 437-446, 2001.

[4] **C. Eriksen, T. Osse, R. Light, T. Wen, T. Lehman, P. Sabin, J. Ballard, A. Chiodi**, A long-range autonomous underwater vehicle for oceanographic research, *IEEE J. Oceanic Engin.,* 26 (4), 424-436, 2001.

[5] **D. Webb, P. Simonetti, C. Jones**, SLOCUM: An underwater glider propelled by environmental energy, *IEEE J. Oceanic Engin.*, 26 (4), 447-452, 2001.

[6] Y. Chao, Autonomous underwater vehicles and sensors powered by ocean thermal energy. In OCEANS 2016-Shanghai, pp. 1-4. IEEE, 2016.

[7] M. Strasberg, Nonacoustic noise interference in measurements of infrasonic ambient noise, *J. Acoust. Soc. Am.,* 66 (5), 1487-1493, 1979.

Concept demonstration of an ASW barrier using low-cost unmanned systems

Jeroen van de Sande[1], Peter Beerens[1], Koen Blom[1], Mathieu Colin[1], Martijn van Riet[1] and Otto Koetsier[2]

[1] Netherlands Organization for Applied Scientific Research (TNO); Oude Waalsdorperweg 63, 2597 AK, The Hague, The Netherlands; jeroen.vandesande@tno.nl
[2] Materiel and IT Command; PO Box 90125 3509 BB Utrecht, The Netherlands

Abstract: *The use of Maritime Unmanned Systems (MUS) in modern warfare has taken a flight over the past years. Advancement of technology in autonomous data processing and underwater communication and navigation is opening doors to more complex applications such as Anti-submarine Warfare (ASW). Traditional manned ASW capacity typically comes with high operational costs, limited availability and vulnerability in high-risk areas. By leveraging new technologies, augmentation of existing ASW capabilities with low-cost, unmanned systems can serve as a force multiplier. A particular use case in this context is the development of an autonomous NATO ASW barrier designed for near-future, small-scale conflicts. For this purpose, a network of static and dynamic nodes function as a 'trip-wire' that is triggered when a submarine crosses it. One or more active dipping sonars combined with passive nodes perform multi-static Doppler-based detection of (slow) intruders. At REPMUS 2024 a concept implementation of such an ASW barrier has been successfully demonstrated in collaboration with other NATO partners. Experimental passive bottom nodes were deployed on the seabed together with a surface vehicle equipped with dipping sonar. All assets performed fully autonomous on-board target detection, tracking and acoustic communication of results using the Smart Adaptive Long and Short-range underwater Acoustic Network (SALSA) protocol stack. Interoperability between assets of the different NATO countries was achieved through the use of the Collaborative Autonomous Tasking Layer (CATL) protocol, allowing target prosecution by other partners. The experiments demonstrated the potential of the use of unmanned systems for augmenting traditional ASW capacity. This paper will elaborate on the implementation of this system-of-systems and show achieved results.*

Keywords: *ASW, ASW barrier, unmanned systems, dipping sonar, multi-static sonar, autonomy, underwater communication*

1. INTRODUCTION

The geopolitical developments of recent years have put Anti-Submarine Warfare (ASW) high on the agenda as a key capability in naval warfare. Modern ASW faces new challenges, however, such as widespread availabilities of technologies that enhance submarine stealth but also challenges in recruitment of personnel required to expand forces [1]. And although technological advancements have improved submarine detection capabilities, the high costs of state-of-the-art frigates limit their availability. As a positive development, technology in the field of marine robotics is becoming more mature. Hurdles such as underwater navigation, underwater communication and autonomy are gradually being tackled, as demonstrated by the rapid introduction of unmanned systems in fields like Mine Counter Measures (MCM). In this paper, another step in the application of Maritime Unmanned Systems (MUS) is taken by investigating their potential for augmenting traditional ASW assets.

In addition to the already existing challenges in MCM, in ASW one is dealing with a mobile, stealthy threat operated by skilled personnel. At the same time, the geographical scales of the areas to be monitored are substantially larger than in MCM. This will require sensors with substantial detection ranges to make their use worth the efforts. Still, successful augmentation of traditional assets by MUS will reward itself by enhancing detection capabilities through parallel deployment of low-cost systems. Additionally, by deploying MUS as forward detection units, frigates can operate outside the range of enemy submarines, making ASW operations safer and more efficient.

In this context The Netherlands is conducting a study to evaluate the performance of a system-of-systems, referred to as a NATO ASW barrier, consisting of static and dynamic nodes that prevent hostile submarines from crossing designated lines. This virtual barrier is intended for monitoring chokepoints, establishing harbour blockades and protecting mobile sea bases against mobile adversaries. Building up the barrier from small, low-cost assets should make the solution scalable and labour-extensive. At the REPMUS ’24 exercise in Portugal a successful demonstration of an implementation of this concept was given. In this paper we will elaborate on the details of the setup and provide experimental results. Particular attention will be given to the fully autonomous multi-static sonar processing chain; dedicated underwater communication protocols and interoperability with the collaborating NATO partners.

2. CONCEPT

Fig. 1 depicts an overview of the full engagement chain of our envisioned implementation of an ASW barrier for an example scenario in which an enemy submarine intends crossing a chokepoint or exiting a harbour. Initially, detection of potential intruders is performed using a combination of active and passive assets. After detection, handover of contacts is done to allow confirmation by other sensors such as sonobuoys or Unmanned Aerial Vehicles (UAVs) carrying a Magnetic Anomaly Detector (MAD). In case of positive confirmation, engagement can be done to refrain the hostile submarine from continuing its mission. In this paper, the emphasis will be on the detection phase up to handover for confirmation.

Certain design decisions were made in earlier studies [2][3] to make this proposed solution scalable and remain low-cost. The barrier uses active (multi-static) sonar for target detection. The reduction in passive signatures of modern submarines would put too high a limitation on the detection range when simply relying on passive sonar. Doppler-sensitive pulses transmitted by the active source(s), are used in combination with passive, man-portable, stationary nodes, which allows detection of slow intruders and strongly alleviates reverberation limitations that would otherwise be experienced as a consequence of using small sensor nodes.

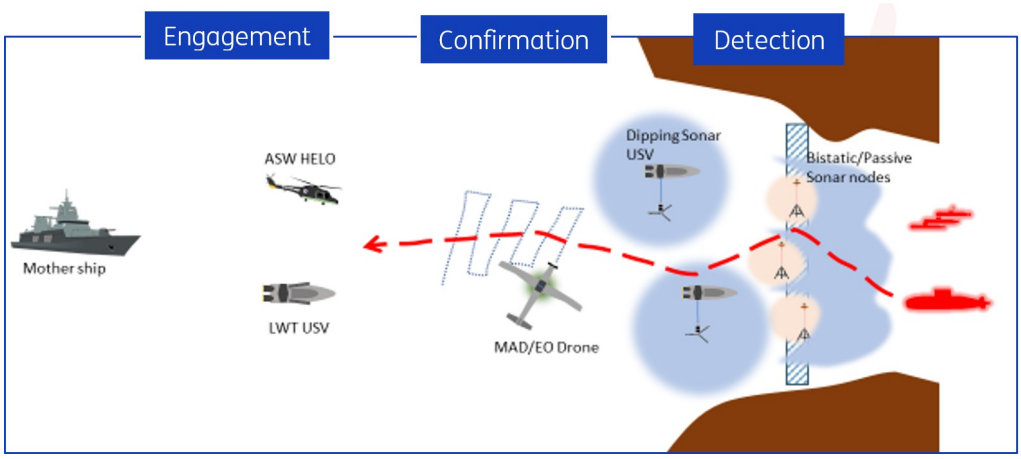


Fig. 1. Schematic overview of the envisaged engagement chain for an ASW barrier monitoring the passage of submarines through a chokepoint or near a harbour.

The passive nodes are typically being distributed in a line configuration on the seabed with several kilometres inter-node distance, together forming a virtual barrier. The active source(s) are represented by one or more dipping sonars operated from Unmanned Surface Vessels. The dipping sonars are stationary when active but are dynamic in nature to allow relocation when deemed necessary.

Since the passive nodes are not wired, they rely on acoustic underwater communications (UComms) to transfer their data. The typical low data throughput of UComms requires significant data reduction by the nodes. The nodes are equipped with processing hardware on which a high level of autonomy is required, covering the construction of sonar images from raw hydrophone data up to and including target detection, tracking and communication.

The development of an ASW barrier is a NATO initiative. As a result, emphasis is put on making all assets interoperable between nations, at all stages of the engagement chain. This means a high level of standardization is used and/or developed. Operations and communications follow existing or upcoming STANAGs as much as possible.

3. CONCEPT DEMONSTRATION

3.1. Test setup

In 2024 the Robotic Experimentation and Prototyping for Maritime Unmanned Systems (REPMUS) exercise, led by the Portuguese navy, was joined by The Netherlands. An experimental barrier was deployed and tested with partner nations Sweden, Germany, the United Kingdom, Norway and Finland. The setup (Fig. 2) consisted of an experimental

dipping sonar (Fig. 3), launched from a RHIB, and passive NILUS bottom nodes [4] of the different nations. The civilian ship GeoSea served as Command, Control and Communications (C3) station. The SEMA AUV from RTSYS carrying an echo repeater was used as ASW test target.

Autonomous processing on the NILUS nodes delivered contact tracks that were encoded and transmitted acoustically to the RHIB. Results were forwarded to the C3 station where they were fused to create a georeferenced picture of contact tracks in the vicinity. These tracks were communicated to the Maritime Operations Centre (MOC) on shore, using the Collaborative Autonomy Tasking Layer (CATL) [5] over the C3MRE network of CMRE. CATL is an interoperability standard under development, that enables collaborative autonomy between unmanned systems of NATO countries. After reception at the MOC, the target confirmation process could be commenced.

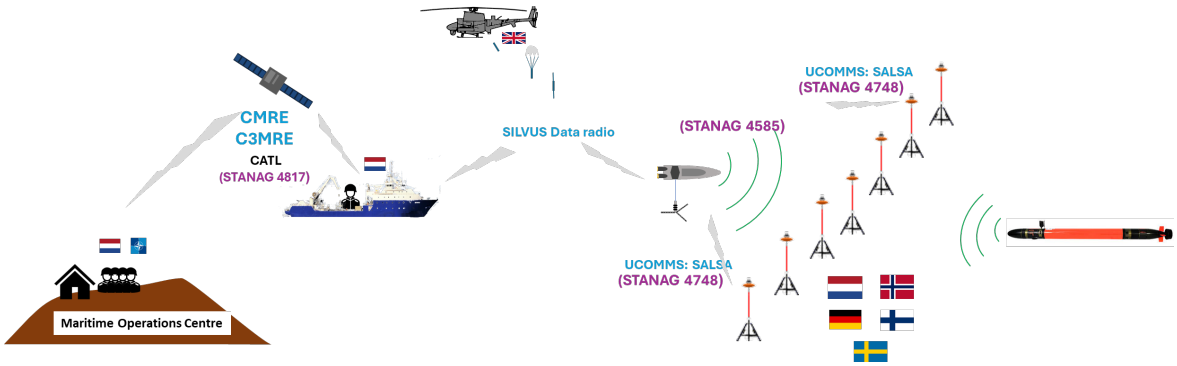


Fig. 2: Overview of deployed ASW barrier testbed at REPMUS '24.



Fig. 3: Left: TNO experimental dipping sonar, right: NILUS node. From [2].

3.2. Detection and tracking

An experimental dipping sonar transmitted Generalized Sinusoidal Frequency Modulated (GSFM) pulses [6] to allow robust, Doppler-sensitive detection of moving targets. Autonomous processing on both the dipping sonar and the bottom nodes converted raw hydrophone data into sonar images after which contact detection was performed (Fig. 4). Contact tracking on the nodes provided reduction of false alarms and target motion analysis. Tracking was done in range-doppler-SNR domain since no bearing information

was available due to the use of only a single receiver on the nodes. Note that the detection ranges are relatively low, as a result of using low source levels on the dipping sonar. Operational detection distances will substantially increase when using source levels that are more realistic for ASW operations and when equipping the bottom nodes with an array of receivers.

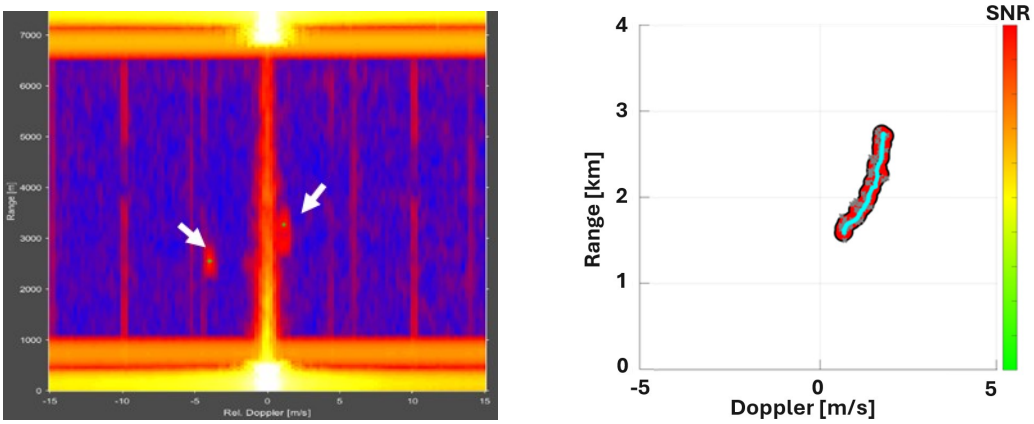


Fig. 4: Left: Range-doppler display of matched-filtered data of a single ping of the dipping sonar, including two detections [2]. Right: an example of a contact track constructed over the course of several pings.

3.3. Underwater acoustic communication

All bottom nodes were running the Smart Adaptive Long and Short-range underwater Acoustic Network (SALSA) protocol stack [7], together forming a communication network for transmitting and relaying tracking results to the gateway at the RHIB. Frequency Repetition Spread Spectrum (FRSS) [8] was used as physical layer protocol, allowing transmission of 256 bits per acoustic message of which 160 bits were available for contact track information. A dedicated encoding protocol allowed the efficient transmission of three contact tracks per message, leaving sufficient detail for proper fusion of node results at the surface. Fig. 5 shows an example of contact tracks on one of the nodes (left) and their decoded versions at the C3 station (right).

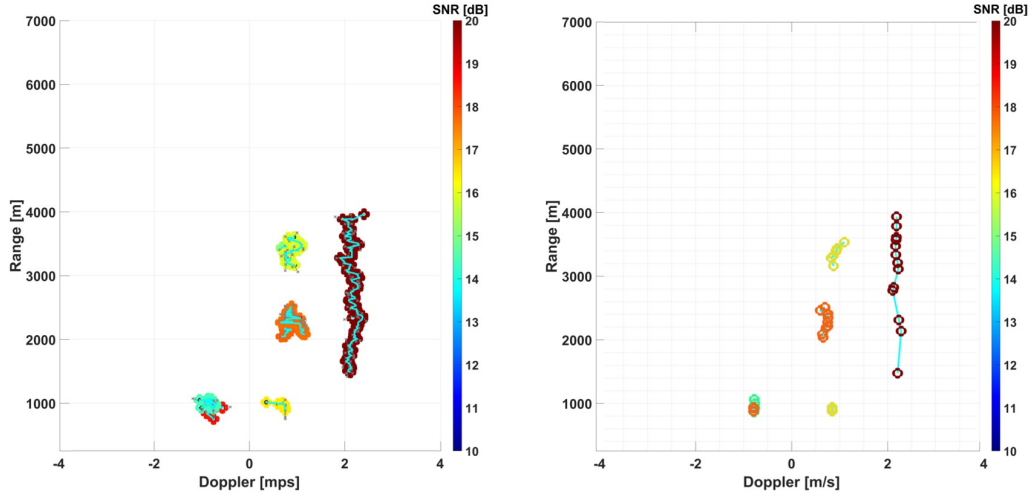


Fig. 5: Track communication protocol. Left: All range-doppler-SNR tracks at one of the bottom nodes built up over the course of 2 hours during one of the experiments. Right: The same tracks after encoding, acoustic communication and decoding at the C&C station, transmitted in three FRSS messages.

3.4. Track fusion

At the C3 station the tracks of all nodes were collected and fused to create a georeferenced picture of contacts in the vicinity. Using source and receiver positions, the contact tracks were converted to particles describing bistatic ellipses. Intersection points on these ellipses provided a georeferenced contact location. Aggregation of these points resulted in georeferenced contact tracks (Fig. 6). Since tracking on the nodes was performed in the range-doppler(-SNR) domain, three nodes with a target track are required to deliver a unique solution. The intended addition of bearing information will in principle require only a single node to provide a georeferenced contact position.

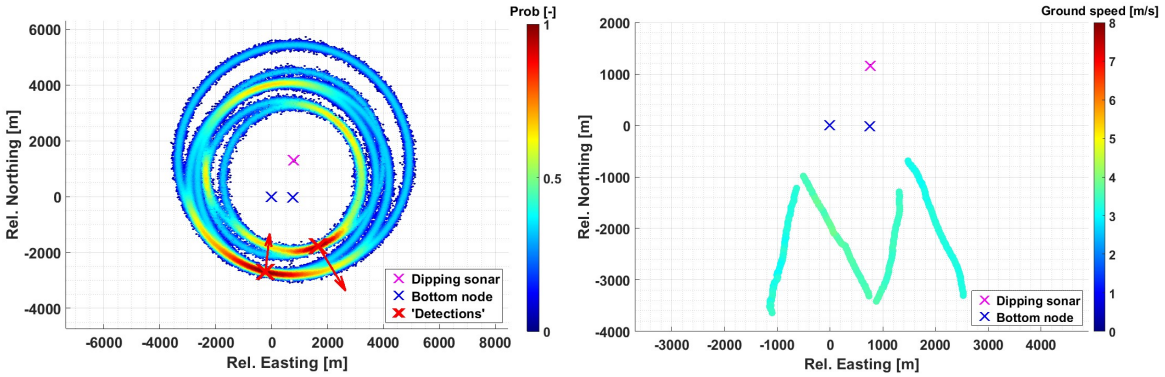


Fig. 6: Fusion process at the C3 station. Left: snapshot of particles describing bi-static ellipses from contact tracks of three assets at a single time instant. Intersection points (red crosses) lead to georeferenced contact positions. Right: aggregated contact positions resulting in georeferenced contact tracks.

SUMMARY AND DISCUSSION

In this paper, the results of a successful demonstration of a NATO ASW barrier were provided. The use of small, unmanned, low-cost systems and the emphasis on interoperability with other nations makes this solution scalable to larger scenarios. The developed autonomous processing capabilities on the nodes proves to deliver substantial data reduction and allows operating in communication-degraded circumstances. The high level of autonomy on all layers of operation complies with the demand for little to no operator involvement.

The current demonstration used bottom nodes as passive receive elements. To increase the range of opportunities in which the developed building blocks can be exploited, future work will focus on the augmentation of this system-of-systems with sonobuoys. At the same time, developments will focus on further reducing the amount of communication by the nodes and increasing area coverage by supporting the use of multiple dipping sonars in parallel.

4. ACKNOWLEDGEMENTS

This work has been financed by the NL MoD and supported by the Material and IT Command and the Royal Netherlands Navy, as part of a research programme on

increasing the autonomy of Maritime Unmanned Systems. In addition special thanks go to the GeoSea crew and the participating NATO countries for their collaboration as well as the Portuguese Navy for hosting the REPMUS ’24 exercise.

REFERENCES

[1] **Barno D., Bensahel N.**, Addressing The U.S. Military Recruiting Crisis, *warontherocks.com*.

[2] **Colin, M.E.G.D., Prior, M.K, Beerens, S.P., Van Heteren, M., van Ijsselmuide, S.P., Van de Sande, J.J.M., Koetsier, J.O, Brouns, E., Van Spellen, M., Van den Bos, R.**, Modelling and Experimentation for an unmanned ASW barrier, *Proceedings of UACE 2023*

[3] **Koetsier, J.O., Beerens, S.P., Colin, M.E.G.D.,** Doppler detection for ASW barrier with maritime unmanned systems, *UDT Rostock,* Rostock, Germany, 2023

[4] **Otnes, R., van der Spek, E., Colin, M.E.G.D., Dol, H.S., Solberg, C.E.,** Easily deployable underwater sensor nodes–the NILUS MK2 demonstrator system, *UDT Europe,* Oslo, Norway, 2016

[5] **Van Riet, M.W.G., et al,** Experimental Demonstrations of CATL, the Collaborative Autonomy Tasking Layer, *OCEANS Limerick*, Limerick, Ireland, 2023

[6] **Doisy, Y., Deruaz, L., van Ijsselmuide, S.P., Beerens, S.P., Been, R.,** Reverberation suppression using wideband Doppler-sensitive pulses, *IEEE journal of Oceanic Engineering*, 33(4), 419-433, 2008

[7] **Dol, H.S.**, EDA-SALSA: Towards smart adaptive underwater acoustic networking, *OCEANS Marseille,* Marseille, France, 2019

[8] **Van Walree, P., Buen, H., Otnes, R.**, A Performance Comparison between DSSS, M-FSK, and Frequency-division Multiplexing in Underwater Acoustic Channels, *Ucomms14 Conference*, Sestri Levante, Italy, 2014

Regular Session
27. Radiated noise from Ships and Surface Platforms

Soundscape modelling in a marine area designated for floating offshore wind turbine (FOWT) development in northern Portugal

Marcos Ricardo Souza[1], Tiago Gomes[1], Guilherme Vaz[1,2]

[1]blueOASIS, Portugal
[2]University of Southampton, United Kingdom

Marcos Ricardo Souza. Rua do Norte, 16, 2655-321, Ericeira, Portugal.
msouza@blueoasis.pt

Abstract: *Environmental impact assessment and planning are crucial in offshore energy production, where underwater acoustics play a key role. As wind turbines generate energy, vibrations from moving components in the generator are also produced. These vibrations travel through the structure and are finally radiated to the underwater environment. Building on the FLOATFARM project, this study develops sound maps incorporating various anthropogenic sources for Viana do Castelo, Portugal, where floating offshore wind farms are planned. The Source Level of an individual floating wind turbine (FOWT) is estimated using Finite Element Analysis based on structural vibrations. Acoustic-structure interaction is considered, and radiated power is computed to calculate the acoustic Source Level. Underwater noise propagation is then modelled using blueOASIS's RAINDROP, a tool that streamlines noise map creation. The model integrates site-specific bathymetry and AIS data to account for vessel contributions to the soundscape, alongside the windfarms and multiple FOWTs. Results include SPL levels around the wind farms at assumed different power generation conditions. Further advancements are needed in powertrain modelling and vibration and acoustic data validation. Nevertheless, this study proposes a framework for characterizing FOWT structure-borne noise and its underwater propagation, contributing to a more comprehensive understanding of offshore wind farm acoustics, even during project design stages.*

Keywords: *Underwater Acoustics, Floating Offshore Windfarm, Renewables, RAINDROP*

1. INTRODUCTION

In 1990, Sweden installed the world's first offshore wind turbine [1], marking the start of a rapidly advancing field in renewable energy. The 1990s also marked the starting point of floating offshore wind turbine research [2]. While most of the operational offshore wind turbines are bottom-fixed structure[1], it is expected that the 2023 capacity of 121 MW [2] will have increased to 264 GW by 2050 [3].

Floating offshore wind farm development presents many advantages over bottom-fixed (or onshore) windfarms, such as extended water depth range [4], [5], insensitivity to seabed properties [5], invisibility [6](avoiding the "not in my backyard" problem). However, they still affect their environment and require detailed planning and environmental impact assessments. The work presented in this paper aims at helping facilitate these assessments by providing a framework for numerical simulations that can quantify the underwater acoustic impact due to floating offshore windfarms also combining it with shipping noise. The methodology is applied to a windfarm (800 MW) planned for Viana do Castelo, in the north of Portugal, shown in Fig. 1.



Fig. 1: Portugal and the planned Viana do Castelo wind farm area.

The paper is organised as follows: section 2 covers the source characterisation and underwater sound propagation; section 3 presents the results; section 4 provides the final remarks.

2. NUMERICAL MODELLING

For the underwater soundscape analysis, it is first required to establish the source characteristics. This is done via a finite element model of a single floating offshore wind turbine, where its Source Level (SL) is estimated from the calculated radiated sound power W. After the source is characterised, noise propagation maps are calculated considering multiple wind turbines and shipping traffic. The methodology is shown schematically in Fig. 2.

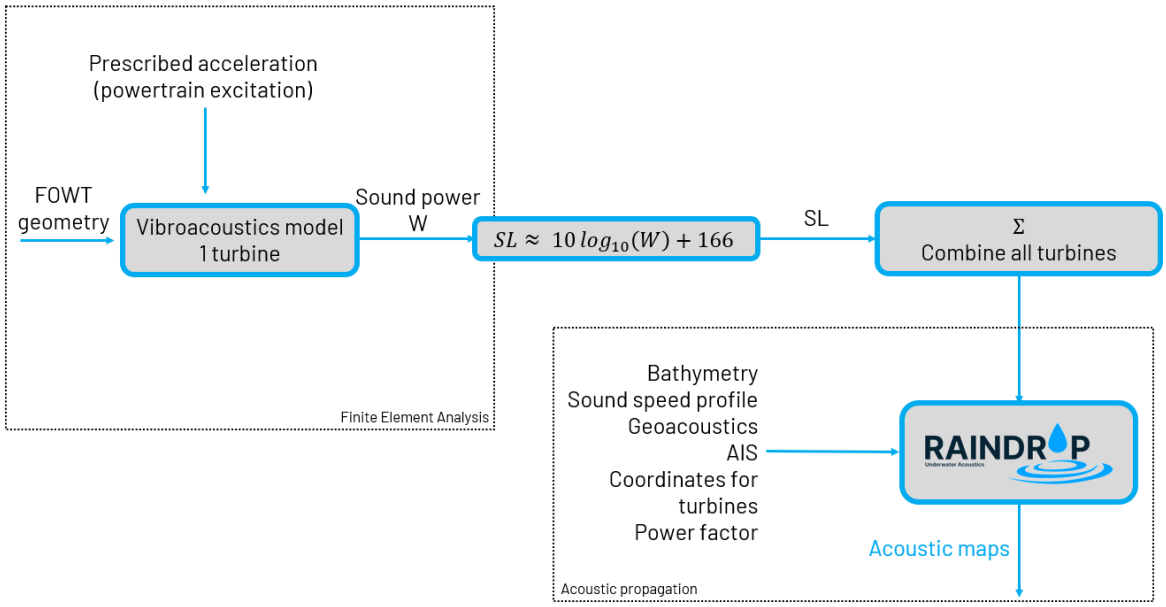


Fig. 2: Methodology overview.

2.1. Finite element analysis: source characterisation

The model was built in COMSOL 6.3 using an axisymmetric shell representing a spar-type substructure, such as the Hywind wind turbine [7]. Structural steel properties were used, with a damping ratio of 0.001. The thickness of the shell varies from 19 mm to 27mm. The diameter at the top of the tower is 6.5m and 7.5 where it connects to the spar buoy. For the floater, the diameter varies from 14.4 m at its lowest point to 7.5 m. The *Pressure Acoustics, Frequency Domain* physics is combined with the structural *Shell* physics via the *Acoustic-Structure Interaction* multiphysics, following Eqs. (1)-(2) [8]:

$$-\mathbf{n}\left(-\frac{1}{\rho_c}\nabla p_t\right) = -\mathbf{n}\mathbf{u_{tt}} \tag{1}$$

$$\boldsymbol{F}_A = p_t\boldsymbol{n} \tag{2}$$

where $\mathbf{n}$ is the surface normal, ρ_c is the fluid density (water in this case), p_t is the total acoustic pressure, $\mathbf{u_{tt}}$ is the out-of-plane surface acceleration of the structure and $\boldsymbol{F}_A$ is the force per unit area experienced by the structure.

A radial acceleration mimicking the excitation from the powertrain is prescribed at the top of the tower. The water domain size changes with the frequency being analysed in order to accommodate at least 3 wavelengths, aiming at having free field propagation. The air-water interface is modelled as a soft boundary, meaning that the pressure vanishes at the interface; and perfectly matched boundary is set at the other water boundaries. This avoids reflections at the boundaries, creating anechoic boundaries. The meshed model is show in Fig. 3(a).

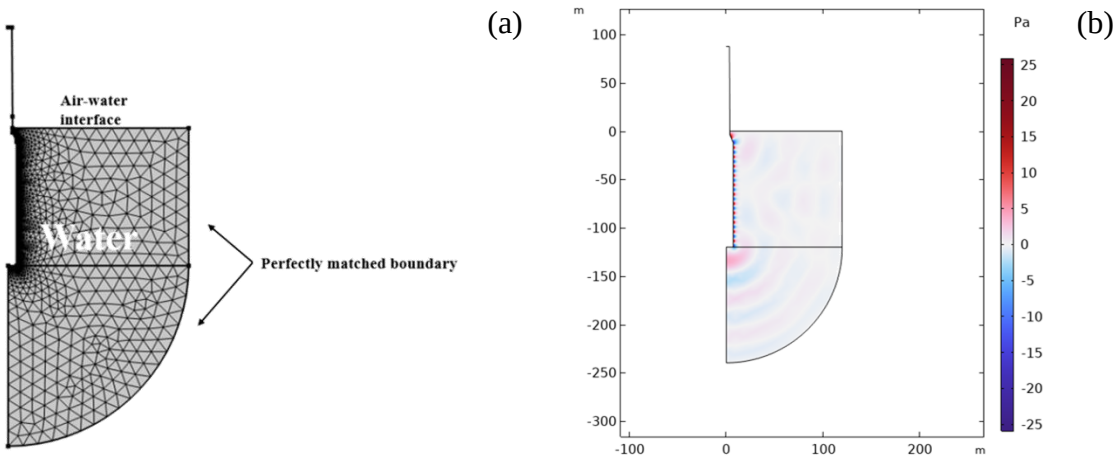


Fig. 3: Meshed model in (a) and acoustic pressure at 40Hz in (b)

Vibroacoustic data for offshore wind turbine is scarce in the literature, especially for floating ones. The excitation at the top of the tower was based on the report of Lindell [9], that has the radial acceleration of the structure at points close to the water surface. A unit-amplitude broadband acceleration was first applied at the tower top in the finite element model. The resulting acceleration was computed at a point 10 meters above sea level. This allowed the derivation of a transfer function between the top prescribed acceleration and the response down the structure. Using this transfer function in conjunction with the measured acceleration spectrum from Lindell, an equivalent prescribed spectrum was inferred that would produce comparable radial acceleration levels in the numerical model. An example of the resulting acoustic pressure is shown in Fig. 3(b).

COMSOL also provides the radiated sound power. This is converted to Source Level with the following assumptions: the equivalent source is a monopole at a depth greater than 1 m and at more than 1 m from the seafloor. Meaning that SL will be estimated as the sound pressure level at 1 m from the monopole source based on the area of a 1 m radius sphere and the calculated sound power. Therefore, the following can be written [10],

$$W = IA = \frac{p^2}{\rho c} A \tag{3}$$

$$p^2 = \frac{W \rho c}{4\pi} \tag{4}$$

where W is the radiated sound power, I is the sound intensity, A is the area of the unit sphere 4π, p is the acoustic pressure, ρ is the water density and c is the water speed of sound. Moreover,

$$SPL = 10 \log_{10}(p^2/p_{ref}^2),\ p_{ref} = 10^{-6}\ Pa. \tag{5}$$

Therefore, at 1 m,

$$SL = 10 \log_{10}(p^2) - 10 \log_{10}\left(p_{ref}^2\right) \tag{6}$$

$$SL = 10\,log_{10}(W) + 10\,log_{10}(\rho c) - 10\,log_{10}(4\pi) - 20\,log_{10}(p_{ref}) \tag{7}$$

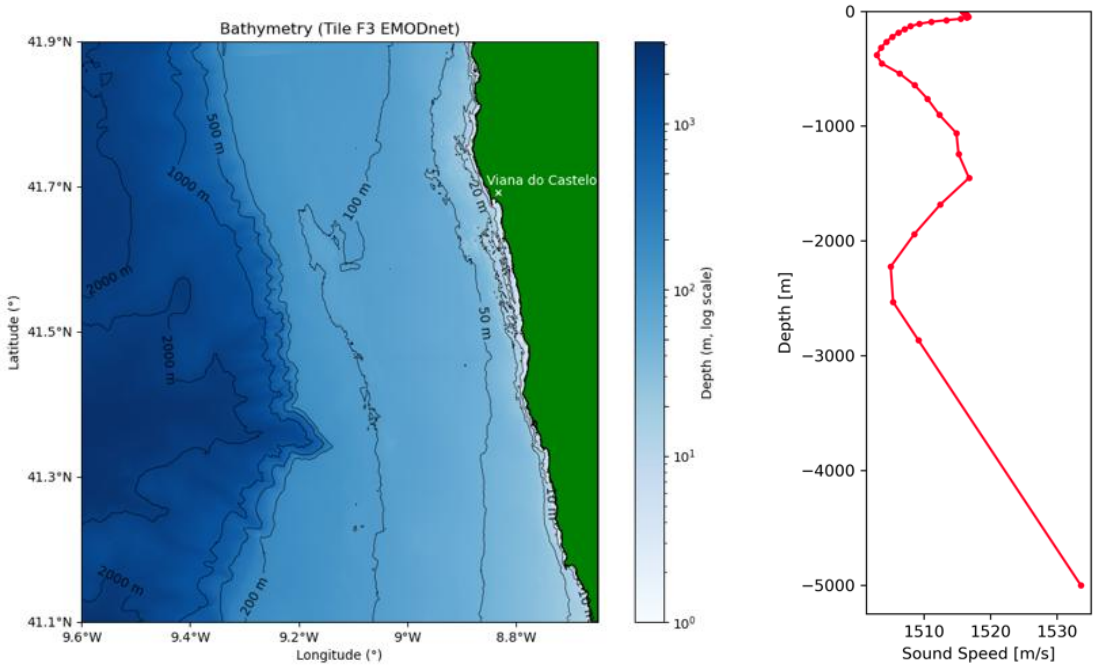
$$SL \approx 10\,log_{10}(W) + 61.8 - 15.9 + 120 \;\approx\; 10\,log_{10}(W) + 166 \tag{8}$$

2.2. Sound propagation

RAINDROP[1] is a holistic framework for computational ocean acoustics, developed by blueOASIS, that targets Digital Twins of the Ocean applications. It seamlessly integrates Python code with extensively validated acoustic propagation solvers. In this paper, KRAKEN [11] was used as the propagation solver. It can efficiently process complex marine data, including bathymetry, metocean information, and ship AIS data, along with custom acoustic sources, such as wind farms.

The bathymetry data was obtained from EMODnet [12], while the water properties to calculate the sound speed profile come from the Atlantic-Iberian Biscay Irish-Ocean dataset [13], following the UNESCO formula [14], both shown in Fig. 4. For convenience, the SSP was obtained for the point at the greatest depth and averaged for the month of December 2024. Also, the seafloor was considered to have uniform geoacoustic properties summarised in Table 1 [15]. However, RAINDROP accepts different properties in a similar fashion to bathymetry. Based on the AIS data, the SL of any vessels presenting in the area of interest are automatically calculated by the toolbox based on the JOMOPANS-ECHO model [16].

For the anticipated capacity of 800 MW and considering that each Hywind turbine has a capacity of 6MW, the Viana do Castelo wind farm would need 134 turbines.



(a) (b)

Fig. 4: RAINDROP inputs – in (a), bathymetry and in (b) sound speed profile

[1] https://blueoasis.pt/raindrop/

Property	Value
Density [g/cm^3]	2.1
P-wave speed [m/s]	1885
S-wave speed [m/s]	0
P-wave attenuation [dB]	34
S-wave attenuation [dB]	25

Table 1: Geoacoustic properties of the seabed

3. RESULTS

The EU Marine Strategy Framework Directive recommends assessing the 63 Hz and the 125 Hz third-octave frequency bands for environmental impact of shipping noise. The results presented in this paper are for the 63 Hz band.

Firstly, to help define the baseline of the soundscape, RAINDROP was run without the windfarm (marked by the light grey polygon), to assess the impact of shipping traffic. For a randomly selected instant in time, the map is produced, with three fishing vessels and a bulker from the AIS available. The map is shown in Fig. 6. The "louder" bulker produces SPL close to 150 dB close to it, but with most of the noise being between 100-120 dB, concentrated in an area with a radius of around 15 km.

The transfer function $H = a_{down}/a_{top}$ between the prescribed acceleration at top and the acceleration 10 m from the water surface is shown in Fig. 5. These are accelerations in the radial direction of the structure. As discussed previously, this was constructed from a unit acceleration first, which was then used to adjust the prescribed acceleration at the top in order to obtain similar levels of acceleration reported in the literature closer to the water surface.

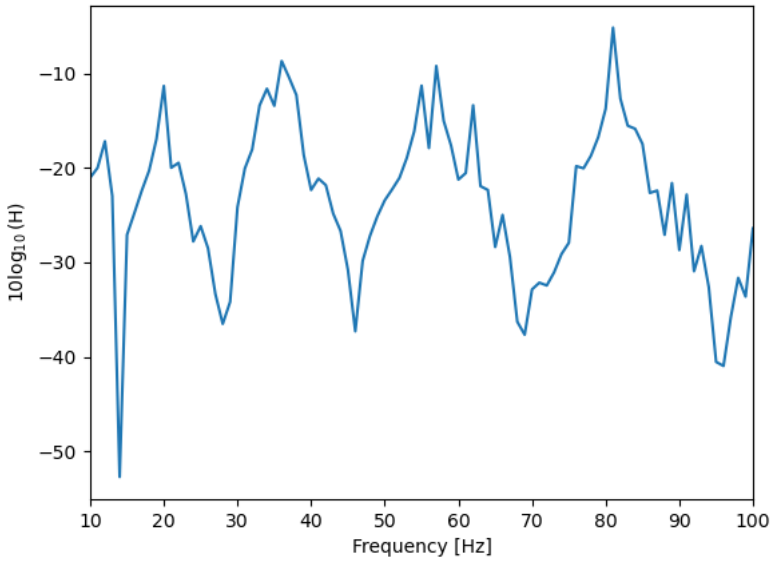


Fig. 5: Acceleration transfer function.

The structural model for the tower, at the 63 Hz, calculates an acceleration of around 0.02 m/s^2. The calculated sound power is 0.0479 W, leading to a SL estimation of 152.6 dB re 1 µPa @ 1 m. Literature reports measured median SL between 149-155 dB re 1 µPa @ 1 m [17] for the Hywind project at 63 Hz depending on the wind speed. However, it is

important to note that the results in literature include additional noise sources such as the moorings. These are expected to contribute more at higher frequencies, but might have some influence at this particular band, as well. Moreover, the highest SL measured occurs at 80 Hz. Regardless, the proposed methodology produces comparable SL and with adjusted input excitation, better agreement is expected. The map with the windfarm and shipping traffic is shown in Fig. 7.

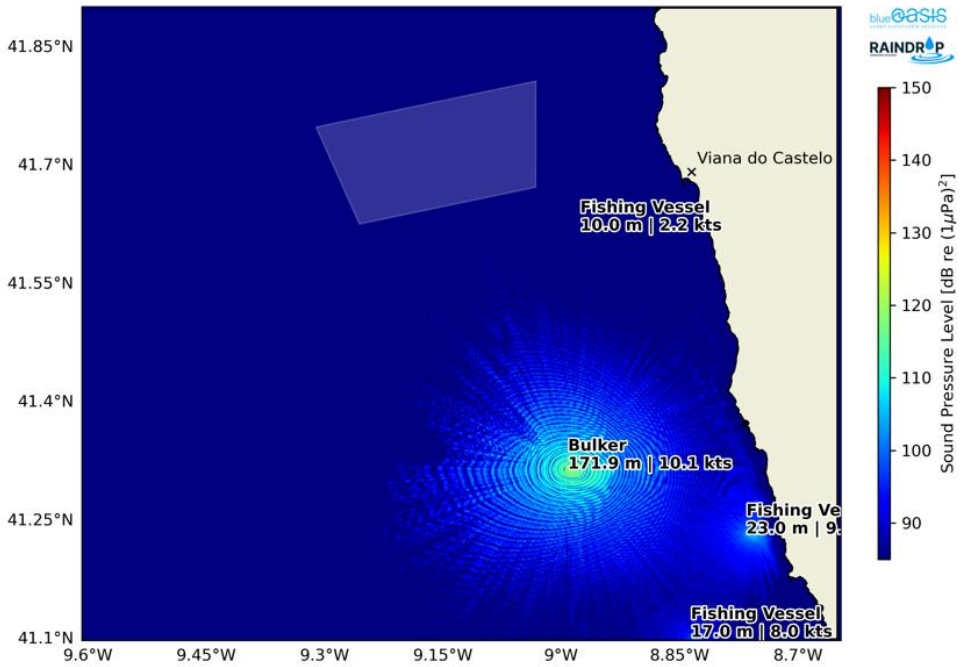


Fig. 6: Acoustic map at 63 Hz with vessels in the Viana do Castelo area.

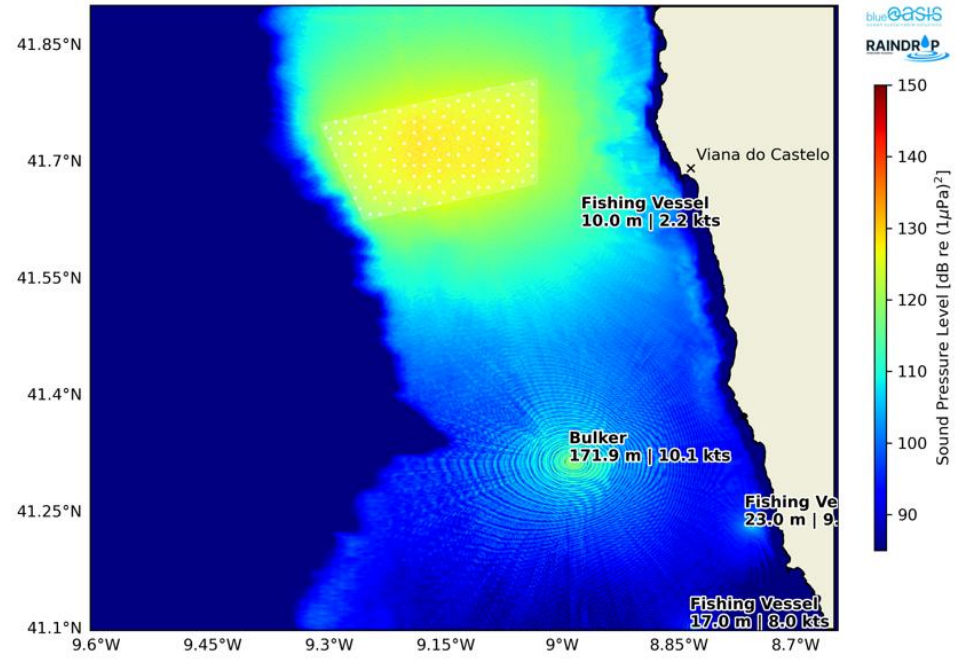


Fig. 7: Acoustic map at 63Hz considering the planed 800 MW windfarm and shipping.

The results consider that all of 134 turbines are a monopole with a SL of 152.6 dB re 1 µPa @ 1 m. This leads to similar SPL from the bulker, also extending more than 15 km

from the windfarm, but in a more uniform area. It stands to reason that the SL will be proportional to the operating condition of the turbine and therefore it is unlikely that all of them are under the same condition. This means that in the real operation, some fluctuation will be present. Moreover, the prescribed acceleration at the top of the tower is not validated, which might lead to deviations in the calculated of the SL. The windfarm operating at different conditions is also shown in Fig. 8. Note the power factor turbine $p_{fac.turb}$ percentage. The calculated radiated sound power W in these cases is directly multiplied by this value. Therefore, if it says 9%, the estimated W is multiplied by 0.09 leading to a SL of 142.1 dB re 1 µPa @ 1 m.

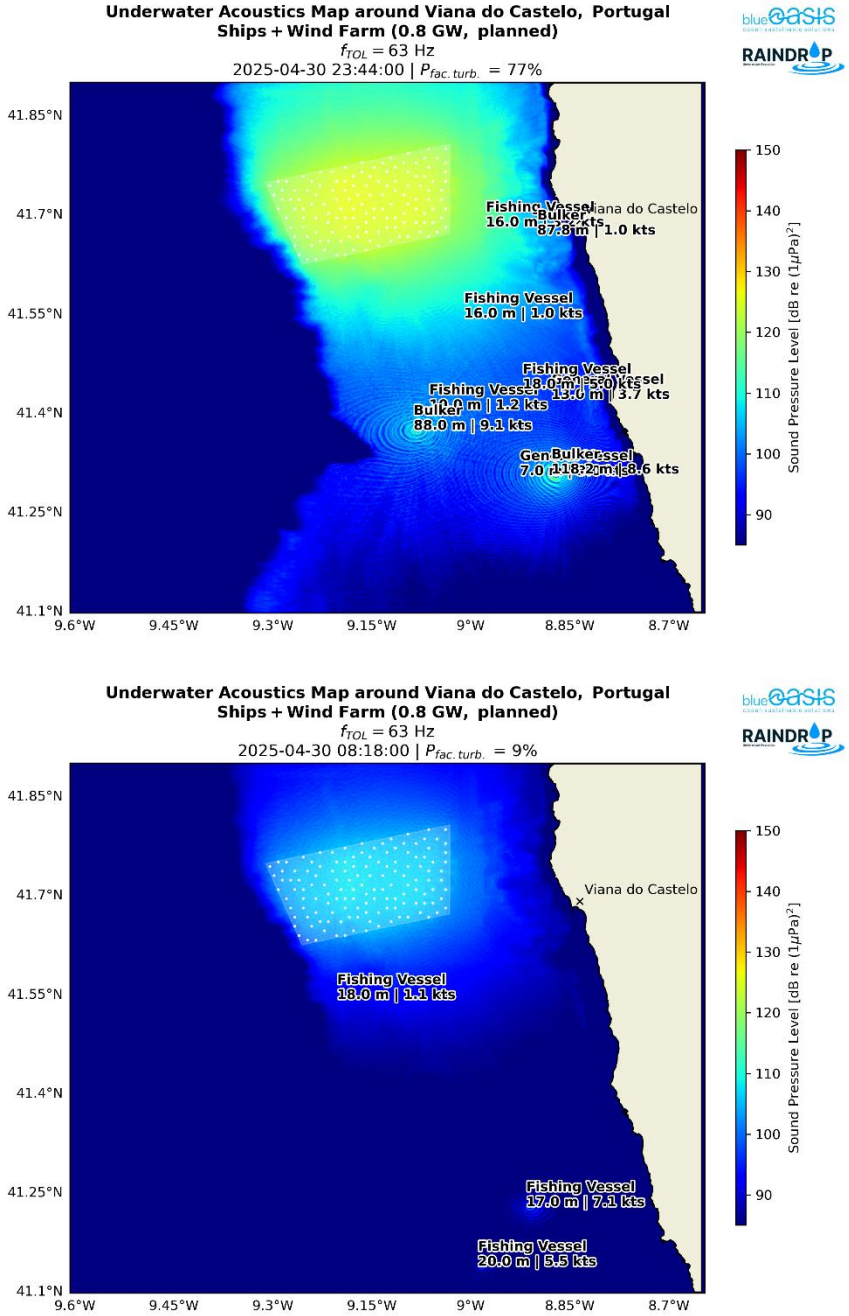


Fig. 8: Windfarm at different operating conditions and different vessels in the area

4. FINAL REMARKS

This paper presents a methodology for estimating the underwater acoustic impact of floating offshore wind turbines and wind farms, starting from finite element modeling and extending to acoustic propagation analysis. This approach helps to enable effective planning and environmental impact assessments, as well as guidance on noise mitigation strategies.

A single turbine is first modeled to estimate its Source Level based on radiated sound power derived from the finite element simulation. Once characterized, this source can be replicated across the planned wind farm layout. By integrating this with established underwater acoustic solvers, the overall acoustic footprint of the wind farm is calculated. Results show that the SPL footprint of the wind farm can be comparable to shipping traffic and can extend to a larger area. However, the actual impact for a real operating wind turbine requires further validation, due to the uncertainty in the excitation from the generator. Regardless of challenges in defining model inputs, the estimated SL is within the range of available experimental data. Therefore, the methodology holds just requiring fine tuning of the input date. The results shown here also consider that every wind turbine is operating with the same condition and therefore has the same SL. This can be changed in RAINDROP with either the use of a power factor that changes the SL depending on the operating conditions of each specific turbine or the farm as a whole.

RAINDROP can also provide results in terms of other quantities such as Level of Onset of Biological Adverse Effects, LOBE. Future work will focus on simulating the wind-loaded generator powertrain for more accurate excitation modeling and employing reduced-order modeling techniques to accommodate more complex turbine geometries beyond the axisymmetric model used in this study and different vibroacoustics modelling, such as Statistical Energy Analysis.

ACKNOWLEDGEMENTS

This research is part of the FLOATFARM project, funded by the European Union under the Horizon Europe programme Grant Agreement No. 101136091. The authors also acknowledge the University of Southampton for providing access to essential software resources and Marine Traffic for their AIS trial.

REFERENCES

[1] M. Barooni, T. Ashuri, D. Velioglu Sogut, S. Wood, and S. Ghaderpour Taleghani, “Floating Offshore Wind Turbines: Current Status and Future Prospects,” *Energies*, vol. 16, no. 1, Art. no. 1, Jan. 2023, doi: 10.3390/en16010002.

[2] “Evolution of floating offshore wind platforms: A review of at-sea devices,” *Renewable and Sustainable Energy Reviews*, vol. 183, p. 113416, Sep. 2023, doi: 10.1016/j.rser.2023.113416.

[3] “DNV: Floating offshore wind: The next five years,” DNV, 2022.

[4] “Economic feasibility of floating offshore wind farms,” *Energy*, vol. 112, pp. 868–882, Oct. 2016, doi: 10.1016/j.energy.2016.06.135.

[5] "Technical and economic aspects of a floating offshore wind farm," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 74–76, pp. 399–410, Apr. 1998, doi: 10.1016/S0167-6105(98)00036-1.
[6] P. Sclavounos, "Floating Offshore Wind Turbines," *Marine Technology Society Journal*, vol. 42, no. 2, pp. 39–43, Jun. 2008, doi: 10.4031/002533208786829151.
[7] K. E. Steen, "Hywind Scotland – status and plans – Statoil," presented at the EERA DeepWind' 2016, Trondheim, Norway, 2016.
[8] "COMSOL Acoustics Module User's Guide." COMSOL. Accessed: May 26, 2025. [Online]. Available: https://doc.comsol.com/5.4/doc/com.comsol.help.aco/AcousticsModuleUsersGuide.pdf
[9] H. Lindell, *Utgrunden off-shore wind farm - Measurements of underwater noise*. 2003.
[10] F. Fahy, "Sound Energy and Intensity," in *Foundations of Engineering Acoustics*, Academic Press, 2001, pp. 74–95. doi: 10.1016/B978-012247665-5/50006-0.
[11] M. B. Porter, "The KRAKEN normal mode program," *SACLANT Undersea Research Centre Memorandum (SM-245)*, 1991, Accessed: May 26, 2025. [Online]. Available: https://ui.adsabs.harvard.edu/abs/1992knmp.book.....P/abstract
[12] "EMODnet Map Viewer." Accessed: May 27, 2025. [Online]. Available: https://emodnet.ec.europa.eu/geoviewer/
[13] Copernicus, "Atlantic-Iberian Biscay Irish- Ocean Physics Analysis and Forecast." Accessed: May 27, 2025. [Online]. Available: https://data.marine.copernicus.eu/product/IBI_ANALYSISFORECAST_PHY_005_001/description
[14] C.-T. Chen and F. J. Millero, "Speed of sound in seawater at high pressures," *The Journal of the Acoustical Society of America*, vol. 62, no. 5, pp. 1129–1135, Nov. 1977, doi: 10.1121/1.381646.
[15] P. Pregitzer, "Underwater Acoustic Impact of Marine Renewable Energy Devices: Modelling Approaches," Master of Science, Instituto Superior Técnico, Lisboa, Portugal, 2021.
[16] A. MacGillivray and C. De Jong, "A Reference Spectrum Model for Estimating Source Levels of Marine Shipping Based on Automated Identification System Data," *JMSE*, vol. 9, no. 4, p. 369, Mar. 2021, doi: 10.3390/jmse9040369.
[17] F. Pace *et al.*, "Underwater Sound Emissions from the Moorings of Floating Wind Turbines: HYWIND Scotland Case Study," in *The Effects of Noise on Aquatic Life: Principles and Practical Considerations*, A. N. Popper, J. A. Sisneros, A. D. Hawkins, and F. Thomsen, Eds., Cham: Springer International Publishing, 2024, pp. 179–201. doi: 10.1007/978-3-031-50256-9_121.

The Impact of Electric Propulsion Systems on Underwater Radiated Noise

Andrea Grech La Rosa[1], Peng Wu[1], and Tom Smith[1]

[1]Department of Mechanical Engineering, University College London, London, WC1E 6BT

Andrea Grech La Rosa, Department of Mechanical Engineering, University College London, Gower Street, London, WC1E 6BT, a.grechlarosa@ucl.ac.uk

Abstract: *Electric propulsion systems are becoming more widespread on marine vessels. There is considerable interest in understanding the acoustical impact of such systems, particularly if it can be shown that they can lead to meaningful reductions in URN. It is sometimes assumed that electric propulsion systems will be quieter than their combustion counterparts. However, there is only limited data on their overall impact on URN available at present and this paints a complex picture. On smaller vessels, research has shown that engine noise due to combustion processes can dominate over a larger speed range. Combustion outboard engines also typically expel their exhaust gases through the propeller hub, which is a significant source of low-frequency broadband noise. However, depending on the power electronics and propulsion architecture, electric motors can produce high amplitude high-frequency tonal noise which can be very detrimental to many marine mammals. In this work, the impact of marine electrification on underwater radiated noise is discussed for small craft. Trials data are presented for two vessels: one powered by a conventional outboard engine and the other by an electric motor. The electric vessel is found to be significantly quieter, with a broadband reduction of 29 dB re 1μPa m. Compared to the conventional outboard, this is due to a lack of combustion noise and no exhaust gases. Crucially, the motor does not radiate high frequency tonal noise into the water. This work demonstrates that electric propulsion systems can be significantly quieter but that care must be taken to ensure that high-frequency noise from the motor does not radiate into the water.*

Keywords: *electric propulsion, underwater radiated noise, small boats, field trials*

1. INTRODUCTION

Small outboard-powered vessels can be the dominant noise sources in shallow coastal waters [1]. They are extremely common across the globe and operate in a wide variety of environments. Some studies have shown that engine noise can dominate over a wide speed range [2]. It was found that the noise at the cylinder and engine firing rates dominated the narrowband spectrum for a RIB powered by an internal combustion (IC) outboard engine at 10 knots and 26 knots. Small boat noise has been highlighted as an issue in coral reef environments [3]. It is noted that a few studies show that some marine life can become acclimatised to the elevated noise levels from small boats [4], but the net impact on many marine ecosystems remains negative.

The electrification of marine vessels is an evolving field driven by factors such as environmental concerns, advancements in battery technology, and regulatory pressure to reduce emissions [5] that may reduce both underwater and airborne noise.

There is an assumption that electric vessels produce very little underwater noise, but there are only a few studies in the literature that quantify the noise levels and make comparisons with conventionally powered vessels [6]. The current data do not allow for any clear conclusions to be drawn as to the impact of marine electrification on URN.

1.1. Electric Outboards

Outboard motors highlight the growing trend toward electrification within the recreational and commercial boating market [7]. Brushless DC motors are adopted by small boats in the electric outboard segment [8] while larger vessels often utilise AC motors to handle higher power loads [9].

An important consideration when using electric motors for propulsion is the potential to introduce high-frequency tonal noise into the underwater environment. These tones arise from the pulse-width modulation (PWM) used to control the motor. This issue was highlighted in a recent study that compared electric boat noise to that from a conventional outboard-powered vessel [6].

PWM noise can originate from the rapid switching of voltages used to control motor speed and torque in electric propulsion systems. In the context of motor control, PWM modulates the duration of “on” and “off” cycles in the power supply to the motor, allowing precise control over motor performance while minimising energy loss. This switching causes high-frequency electromagnetic interference (EMI) and acoustic noise, spreading through the motor structure and surrounding areas, including water [10]. This phenomenon is not limited to a specific type of motor but is prevalent in both AC and DC systems using PWM-based motor drives. The magnitude of noise depends on motor design, switching frequency, and the type of PWM strategy employed.

The transmission of PWM noise into water is influenced by factors such as the physical location of the motor and power electronics within the vessel, the material and design of the mountings, and the proximity of components to the hull. Effective mitigation measures to reduce this noise include using optimised PWM switching frequencies that minimise harmonics in sensitive frequency ranges, and incorporating EMI filters in the drive circuits [11]. Proper shielding of cables and enclosures also prevents electromagnetic noise leakage, ensuring minimal impact on the marine environment and compliance with environmental standards.

2. TRIALS

This work presents experimental results from a trial conducted on two small vessels: one fitted with a conventional internal combustion (IC) outboard engine and one fitted with an electric outboard. The aim is to provide comparative data to quantify the difference between the noise levels produced by these vessels. The study also aims to detail the makeup of the signature and how this is changes between the two propulsion types.

The vessels used for the trial are a 5.8 m rigid inflatable boat (RIB) and a 5.8 m "Piscator". The two vessels are shown in Figure 1. Both vessels are of the same length have planing hulls, and are powered by a single outboard engine with a 3-bladed propeller. The RIB is powered by a 52 kW 4-stroke, 4-cylinder IC outboard engine, and is referred to as the internal combustion engine boat (ICEB) in this work. The Piscator is powered by a 40 kW electric outboard from RAD Propulsion. This will subsequently be referred to as the electric boat (EB). Parameters for both vessels are shown in Table 1. Whilst the vessels are not identical, they are more similar in size and propulsion configuration than has been used in other comparative studies of electric boat noise [6], [12].

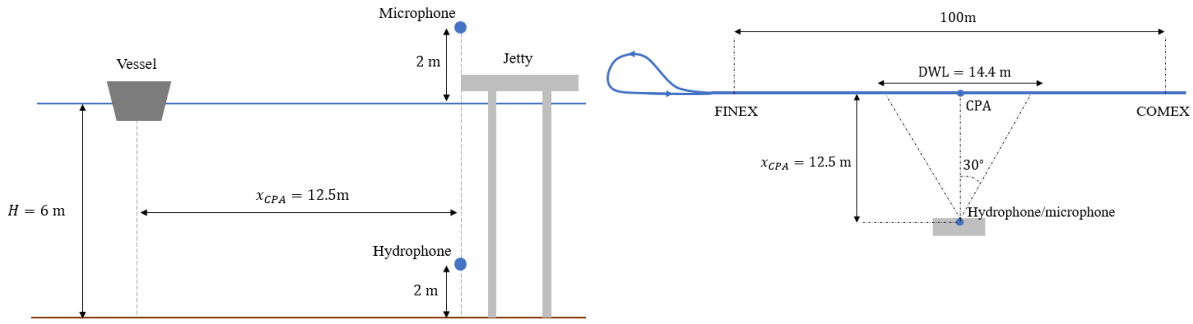


Fig.1: Images of the two vessels used for the trial

Table 1: Parameters of the two trial vessels

Vessel	ICEB	*EB*
Installed Power	52 kW	40 kW
Gear Ratio (g)	2.33	1.5
Propeller blades (z)	3	3
Propeller diameter (D)	0.34 m	0.33m
Hub-diameter ratio (h/D)	0.3	0.23

The trial was conducted in a river estuary which is very sheltered and representative of the typical operating area for these vessels. The water conditions were commensurate with sea state 1 and the wind speed was below 10 knots throughout the trial. The trial was conducted close to slack tide, minimising discrepancies due to speed over ground being substantially different from the speed through water. The flow speed was estimated to be less than 0.5 knots, and double runs were conducted with and against the flow as illustrated in Figure 2. Four double runs were conducted for both vessels at 6 knots resulting in eight sets of underwater noise measurements for each. The background noise was monitored throughout and no other vessels were present in the area during the trials.



(a) (b)

Fig.2: Illustrations of the trial setup (not to scale). (a) Illustration of equipment positioning and depths. (b) Illustration of track.

Underwater noise measurements were taken using a Neptune D70 hydrophone sampled at 100 kHz. The hydrophone was also located at a horizontal distance of 12.5 m from the closest point of approach and 2 m above the river bed. This is closer to the CPA than in some other studies of small vessel URN [2], [13] and closer than recommended by ISO17208-2 [14]. This was done in anticipation of the low frequency noise being very low for the electric vessel and so seeks to improve the signal-to-noise ratio.

Locating the hydrophone closer to the CPA reduces the data window length available without increasing the maximum angle between the source and receiver. Having too great an angle could introduce uncertainty due to unknown directivity patterns, making the data more difficult to interpret and compare with other sources. To mitigate this, four double runs were carried out for each vessel.

3. RESULTS AND DISCUSSION

Results are presented as either received levels (L_p) or 1 m equivalent source levels (L_S). Power spectral density (PSD) estimates are obtained via Welch's method. For each run, the data within the data window length are windowed to obtain 1 s segments with a 50% overlap. A Hanning window is applied to each segment. The PSD estimates for each window are then averaged to obtain the PSD for that particular run.

The 1 m equivalent source levels are then computed using the Seabed Critical Angle (SCA) [15]. This approach accounts for the Lloyd's Mirror effect at the free surface and reflection and absorption at the seabed.

The decidecade band 1 m equivalent source levels are shown in Figure 3 for both vessels. The overall source levels, computed by integrating over the decidecade bands, are 174 dB re 1µPa m for the ICEB and 145 dB re 1µPa m for the EB. The source level for the ICEB considered in this study is higher than has been reported in some other studies of IC powered RIBs. For example, a 7 m RIB travelling at 6 knots has a reported source level of 166 dB in and a 5 m RIB had a source level of 167 dB re 1µPa m [2]. However, a 5 m motorboat with a 15 hp engine had a reported broadband source level of 173 dB when travelling at 5.6 knots [13], which is similar to the result seen here.

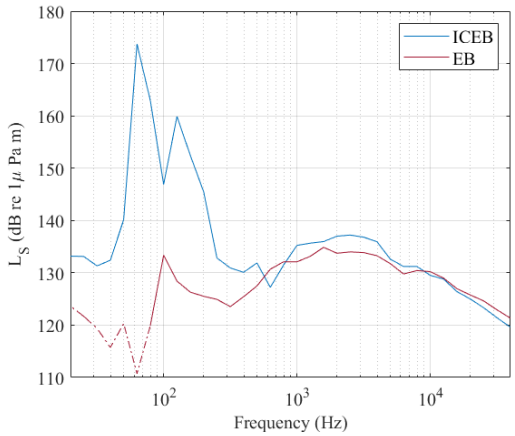


Fig.3: Decidecade band underwater source levels for both vessels. Dotted lines are used to indicate where the received level was within 3 dB of the background noise and denotes the maximum level that the source level can be at the given decidecade band.

The electric boat is significantly quieter than all comparable data for IC engines at this speed, with the broadband source level being 29 dB lower than for the ICEB. The large difference is entirely due to a large drop in low frequency noise for the electric boat. Despite the hydrophone being located only 12.5 m from the CPA, the EB noise levels at frequencies below 80 Hz are not sufficiently above the background noise levels to reliably measure them. The EB source levels are below those for the ICEB for all frequency bands below 500 Hz, where the engine noise typically is greatest [2], with the levels being comparable at higher frequencies. The acoustic signature for the EB is dominated by the 1 kHz $\leq$ f $\leq$ 5 kHz bands which is very unusual for a vessel travelling at this speed. The noise at these frequencies made up of flow noise, spray, and cavitation collapse noise if cavitation is present. The levels reported at the mid-high frequencies are comparable with other studies [2], [13].

To better understand the reasons for the differences in the results and to provide insights into the makeup of the acoustic signatures, the narrowband spectra are shown in Figure 4 for both vessels, presented as received levels.

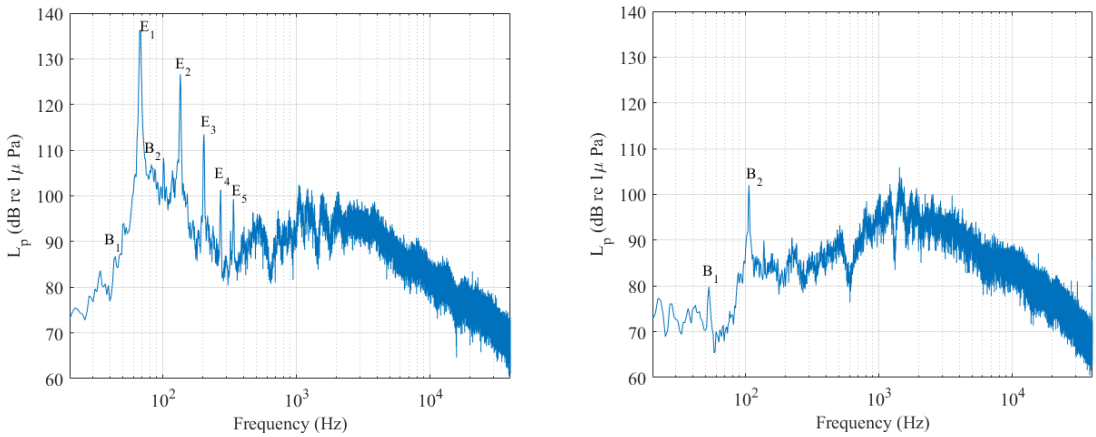


Fig.4: Narrowband received levels for (a) ICEB and (b) EB. "B" denotes the propeller blade rate and harmonics and "E" denotes the engine speed and harmonics.

The narrowband spectrum for the ICEB is dominated by tonal noise at the engine firing rate with the first 5 harmonics are clearly visible in the spectrum. These are super-imposed onto a low frequency broadband hump, which is from the noise produced by the exhaust gases which are expelled through the hub [16]. Tonal noise at the blade rates is just visible but is largely masked by the exhaust noise. For the electric boat, the blade rate is clearly visible in the spectrum but no other significant tonal components are present. The source levels at the blade rate and harmonics are comparable for both vessels, although slightly higher for the ICEB.

Interestingly, the high-frequency tonal noise reported by Gaggero et al. [6], which results from the motor control, is not present here. A number of design approaches can be taken to prevent the high frequency noise transmitting into the environment, as discussed in the introduction. Details of the outboard design and layout are not available, but it is known that the motor sits slightly above the waterline for the RAD motor, whereas for the motor tested by Gaggero et al. [6] the motor is in the hub. What these two sets of results show is that electric motors can produce high frequency tonal noise, but this is not a foregone conclusion, and it can be mitigated.

4. CONCLUSIONS

This paper has discussed the acoustics of electric outboards on small boats and presented a comparison of an internal combustion outboard engine and an electric propulsion system. Underwater noise levels have been presented for the two vessels travelling at 6 knots.

The underwater noise from the EB is substantially lower than for the ICEB. The low frequency noise reduction comes about due to the removal of the engine tones associated with combustion processes as well as the lack of exhaust gases. The large reduction in the URN from the EB compared to the ICEB demonstrates the potential for electric vessels to significantly reduce the acoustical impact of small craft on the marine environment, at least at low speeds.

One important conclusion of this study is it shows that electric propulsion need not lead to high frequency underwater noise pollution. Future work should explore methods to prevent high-frequency noise from radiating into the water for various electric propulsion units, especially those with motors located in the hub.

Future trials should consider a broader range of operating conditions to understand how the reductions in engine noise change at higher speeds. The potential benefits of having a reduced hub-diameter ratio could also be quantified.

5. ACKNOWLEDGEMENTS

The authors would like to thank Pete Byford and Lawrence Herbert of RAD Propulsion for providing access to the vessels used in these trials.

REFERENCES

[1] L. Wilson, M. K. Pine, and C. A. Radford, “Small recreational boats: A ubiquitous source of sound pollution in shallow coastal habitats,” *Mar Pollut Bull*, vol. 174, p. 113295, 2022.

[2] T. A. Smith, A. G. La Rosa, and B. Wood, “Underwater radiated noise from small craft in shallow water: Effects of speed and running attitude,” *Ocean Engineering*, vol. 306, p. 118040, 2024.
[3] D. Lecchini *et al.*, “Boat noise prevents soundscape-based habitat selection by coral planulae,” *Sci Rep*, vol. 8, no. 1, p. 9283, 2018.
[4] H. R. Harding, T. A. C. Gordon, R. E. Hsuan, A. C. E. Mackaness, A. N. Radford, and S. D. Simpson, “Fish in habitats with higher motorboat disturbance show reduced sensitivity to motorboat noise,” *Biol Lett*, vol. 14, no. 10, p. 20180441, 2018.
[5] E. A. Sciberras, B. Zahawi, and D. J. Atkinson, “Reducing shipboard emissions–Assessment of the role of electrical technologies,” *Transp Res D Transp Environ*, vol. 51, pp. 227–239, 2017.
[6] T. Gaggero *et al.*, “Electric boat underwater radiated noise and its potential impact on species of conservation interest,” *Mar Pollut Bull*, vol. 199, p. 115937, 2024.
[7] M. Porru, M. Pisano, A. Serpi, and F. Pilo, “Electrification of Leisure Boats: a commercial State-of-the-Art,” in *2020 IEEE Vehicle power and propulsion conference (VPPC)*, 2020, pp. 1–6.
[8] S. Calligaro, D. Frezza, R. Petrella, M. Bortolozzi, M. Mezzarobba, and A. Tessarolo, “A fully-integrated fault-tolerant multi-phase electric drive for outboard sailing boat propulsion,” in *2019 21st European Conference on Power Electronics and Applications (EPE’19 ECCE Europe)*, 2019, pp. 1–10.
[9] M. Chai, D. R. Bonthapalle, L. Sobrayen, S. K. Panda, D. Wu, and X. Chen, “Alternating current and direct current-based electrical systems for marine vessels with electric propulsion drives,” *Appl Energy*, vol. 231, pp. 747–756, 2018.
[10] K. Borisov, T. E. Calvert, J. A. Kleppe, E. Martin, and A. M. Trzynadlowski, “Experimental investigation of a naval propulsion drive model with the PWM-based attenuation of the acoustic and electromagnetic noise,” *IEEE Transactions on Industrial Electronics*, vol. 53, no. 2, pp. 450–457, 2006.
[11] A. Kalair, N. Abas, A. R. Kalair, Z. Saleem, and N. Khan, “Review of harmonic analysis, modeling and mitigation techniques,” *Renewable and Sustainable Energy Reviews*, vol. 78, pp. 1152–1187, 2017.
[12] Y. Alfio, “Field Study on Underwater Noise Emitted by Small Tourist Boats. Comparison Between the Use of Electric and Combustion Motors,” *Archives of Acoustics*, vol. 48, no. 3, pp. 347–358, 2023.
[13] M. Picciulin *et al.*, “Characterization of the underwater noise produced by recreational and small fishing boats (< 14 m) in the shallow-water of the Cres-Lošinj Natura 2000 SCI,” *Mar Pollut Bull*, vol. 183, p. 114050, 2022.
[14] ISO17208-2, *Underwater acoustics — Quantities and procedures for description and measurement of underwater sound from ships. Part 2: Determination of source levels from deep water measurements*. International Standards Organisation, 2019.
[15] A. O. MacGillivray, S. B. Martin, M. A. Ainslie, J. N. Dolman, Z. Li, and G. A. Warner, “Measuring vessel underwater radiated noise in shallow water,” *J Acoust Soc Am*, vol. 153, no. 3, pp. 1506–1524, 2023.
[16] T. A. Smith, A. G. La Rosa, G. Piggott, J. A. N. Gaivota, and S. S. McMorran, “Propeller cavitation on small craft: Underwater noise measurements and visualisation from full-scale trials,” *Ocean Engineering*, vol. 317, p. 120024, 2025.

Cavitation on small craft propellers and its contribution to underwater radiated noise

Tom A. Smith[1], Andrea Grech La Rosa[1], Solomon S. McMorran[2], Giles Piggott[3], and João A. N. Gaivota [4]

[1]Department of Mechanical Engineering, University College London, London, WC1E 6BT
[2]Royal Canadian Navy
[3]Royal Navy
[4]Portuguese Navy

Tom Smith, Department of Mechanical Engineering, University College London, Gower Street, London, WC1E 6BT, tom.smith.17@ucl.ac.uk

Abstract: *Cavitation is a major contributor to underwater radiated noise (URN) from commercial ships but less is known about the dynamics, inception speed, and contribution of cavitation to URN for small vessels. Small vessels often use propellers located very close to the free surface and which turn at high speeds. This means that the cavitation dynamics are likely to be different from those on larger ships. Additionally, many small vessels are powered by internal combustion outboard engines, which typically expel exhaust gases through the hub, leading to a bubbly hub vortex. In this work, acoustic data from small boat trials are presented alongside underwater camera footage to show how cavitation develops on small high speed propellers and how it contributes to the acoustic signature. The results show that the cavitation inception speeds are typically 5-6 knots, and this is accompanied by a sharp rise in noise across a broad frequency range. Visual footage shows that tip vortex cavitation (TVC) dominates at low speeds. Results are also presented that provide insights into how the material state of the propeller can influence the cavitation dynamics. Whilst rough and fouled propellers might be assumed to cavitate earlier and hence be louder than their smooth counterparts, it is shown here that this is not always the case. When tip vortex cavitation dominates, small amounts of roughness can actually reduce URN levels by disrupting the formation of the tip vortices.*

Keywords: *propeller cavitation, underwater radiated noise, experimental trials*

1. INTRODUCTION

Propeller cavitation is a significant contributor to radiated noise from ships and can also lead to efficiency losses, vibration, and erosion [1]. Over past decades there have been a multitude of studies on ship propeller cavitation, and this work has led to significant design improvements. Despite this, cavitation remains a problem on vessels of all sizes, and is extremely difficult to avoid at higher speeds [2]. For larger vessels, it typically dominates the acoustic signature at higher speeds, producing noise across a wide frequency range.

Recent research has highlighted the acoustical importance of small boats, particularly in shallow and coastal waters where they can dominate the soundscape [3, 4]. This has prompted increased interest in quantifying levels of URN from small craft and understanding the underlying sources [5, 6, 7]. These studies show that the relationship between speed and noise levels is less clear than it is for larger vessels, and this may partly be due to cavitation not always being worse at higher speeds [8]. Few studies have discussed cavitation on small craft in any detail, and visual footage from field trials is rare. As a result, cavitation inception speeds are not readily available, and the contribution of cavitation noise to the overall noise levels is not known for small vessels. In many cases, propellers are "off-the-shelf" and many have not undergone testing to determine their cavitation performance. This, together with the higher rotational speeds and their proximity to the free surface make it likely that the cavitation performance will be different for small high-speed propellers than it is for a typical large ship propeller.

This paper presents recent experimental research carried out into the acoustical impact of propeller cavitation on small boats powered by internal combustion engine (ICE) outboard motors. Underwater noise data are presented for two vessels: a 5 m and a 6 m rigid inflatable boat (RIB) alongside camera footage for the 5 m RIB. The results consider three areas: the role of exhaust gases, the cavitation inception speed and the change in signature due to the onset of cavitation, and the role of the propeller surface quality.

2. METHODS

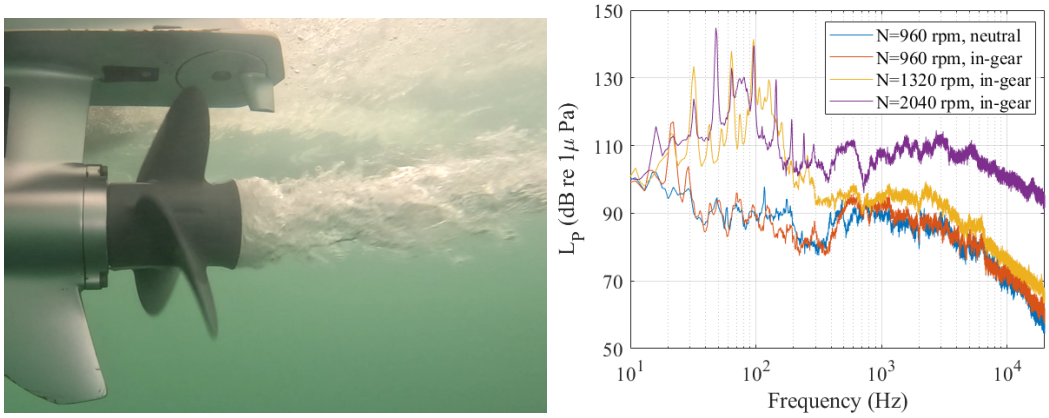
The results presented in this work pertain to trials carried out in July 2024 off the south coast of England. The weather conditions were calm and commensurate with sea state 2. Two RIBs were used and their principal particulars are shown in Table 1.

A detailed description of the trial setup is given in [9] and so only a brief description is given here. Acoustic measurements were made with two RS Aqua/Turbulent Research acoustic recorders sampling at 384 kHz. These were located at a horizontal distance of 20 m from the closest point of approach and 3 and 5 m above the seabed, which consisted of sand and mud. As discussed in [9], this is closer than is recommended by ISO 17208 and was motivated by the desire to improve the signal-to-noise ratio, particularly as the focus of this study is to try to understand the contributions of particular types of cavitation to the acoustic signature. Camera footage was obtained using a GoPro Hero 11 recording at 240 frames per second. This was mounted in a custom-made faired frame. Trials were carried out using double runs, with three double runs per speed. Additional trials were carried out where the vessel speed was increased incrementally to determine the cavitation inception speed using the camera.

The data are analysed using the approach set out in [9] and uncertainty analysis is presented in [9, 10]. Source levels are computed from the received levels using the Seabed Critical Angle method [11].

	5 m RIB	6 m RIB
Length overall (m)	5.0	5.9
Beam (m)	2.2	2.5
Max engine power (kW)	37	104
Number of cylinders	3	4
Gearbox ratio, g	2.1	2.6
Number of blades, z	3	3
Propeller diameter, D (m)	0.28	0.36
Pitch-diameter ratio, P/D	1.17	not known

Table 1: Vessel particulars for the two boats



(a) *(b)*

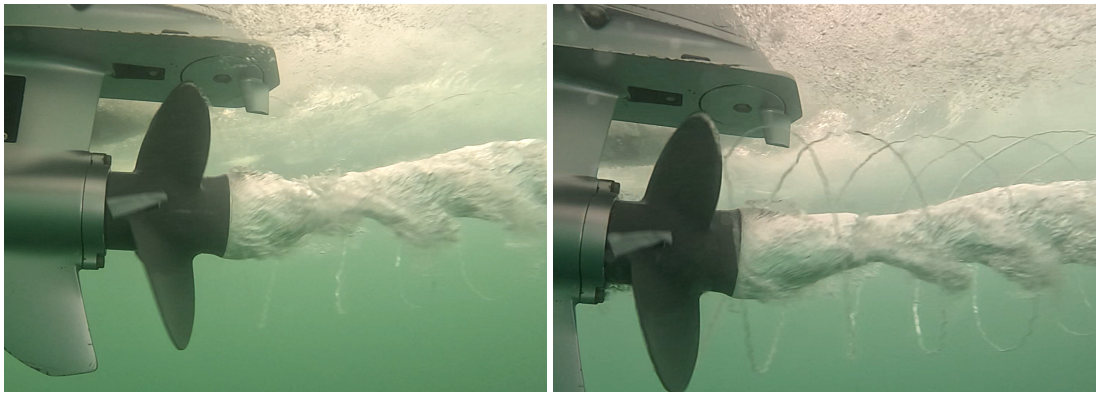
Figure 1: (a) hub gases being expelled through the hub of the 5 m RIB at 4 knots. (b) narrow-band spectra for the vessel moored up with the engine running at different speeds.

3. RESULTS

3.1. EXHAUST GASES

ICE outboard motors typically have two exhausts: one above the waterline and one in the propeller hub. When idling or at very low speeds, gases are expelled through a port above the waterline. As the speed increases (usually at around 3-4 knots), the increased pressure in the engine and the reduced pressure in the hub vortex results in the gases being expelled through the hub. This is shown in Figure 1(a) which is for the 5 m RIB travelling at 4 knots.

This has two important consequences for underwater noise. Firstly, it results in the hub being relatively large as a proportion of the overall propeller diameter. Thus, for a given diameter, there is less area providing thrust which increases the load per unit area and hence the propensity for the propeller to cavitate. Secondly, the expulsion of gas into the low pressure hub vortex core creates noise. To determine how this contributes to the acoustic signature, trials have been carried out with the vessel tied alongside and doing a bollard pull. This has been done to eliminate noise from spray and other hydrodynamic sources. The engine was run at a series of different speeds and a hydrophone located 2 m away from the propeller recorded the sound. Narrowband spectra are shown in Figure 1(b) for the vessel when idling and at two higher engine



(a) *(b)*

Figure 2: Tip vortex cavitation on the 5 m RIB at (a) 5 knots and (b) 6 knots.

speeds. When idling and not in gear, the noise levels are only very slightly above background noise. When idling but in gear, tonal components at the engine firing rate and propeller rate are observed but no other significant changes are observed. At 1320 RPM, the exhaust gases are expelled through the hub but no cavitation is observed. In addition to an increase in tonal components at the engine and propeller rates, there is a pronounced low-frequency broadband hump from 10 to 300 Hz. At 2040 RPM, the low frequency noise levels are similar, but there is a significant increase in broadband levels at mid and high frequencies. The camera footage shows that cavitation is present at this speed and so this noise is likely dominated by cavitation collapse. From this, we can deduce that the exhaust gases result in a low frequency broadband hump in the acoustic spectrum.

3.2. CAVITATION INCEPTION AND TIP VORTEX CAVITATION

No consistent cavitation was observed on either vessel below 5 knots. However, occasional bubble collapse signals were measured at 4 knots. Analysis of the camera footage showed that this is due to pre-existing bubbles in the water being drawn close to the propeller tip where they expand and then collapse. Further details can be found in [9].

TVC emerges on the propeller on the 5 m RIB consistently at 5 knots. At 6 knots, fully developed TVC is observed, with cavitating tip vortices propagating several propeller diameters downstream before they break up and collapse. Images of the cavitation pattern at 5 and 6 knots are shown in Figure 2. The onset and development of cavitation on the 6 m RIB follows a similar pattern, but this occurs at a slightly higher speed, with TVC first appearing at 6 knots but not fully developing until around 8 knots. This is partly because the 6 m RIB uses a larger propeller that operates at a lower speed. At 5 knots, the cavitation number calculated using the propeller rotation rate is $\sigma_n = 0.76$ for the 5 m RIB and $\sigma_n = 1.32$ for the 6 m RIB. By 6 knots, this has reduced to $\sigma_n = 0.90$ for the 6 m RIB, and so cavitation first occurs at a similar cavitation number for the two vessels. Narrowband spectra for the two vessels are shown in Figure 3 at 6 and 10 knots. For the 5 m RIB at 6 knots, the blade rate is very pronounced and there is a broadband hump centred around 400 Hz. This is characteristic of TVC noise [12]. This is far less prominent for the 6 m RIB due to the later emergence of cavitation. The increase in high frequency noise on the 6 m RIB going from 6 to 10 knots coincides with the increasing levels of cavitation. At 10 knots, a broadband hump is visible for both vessels but is far less prominent

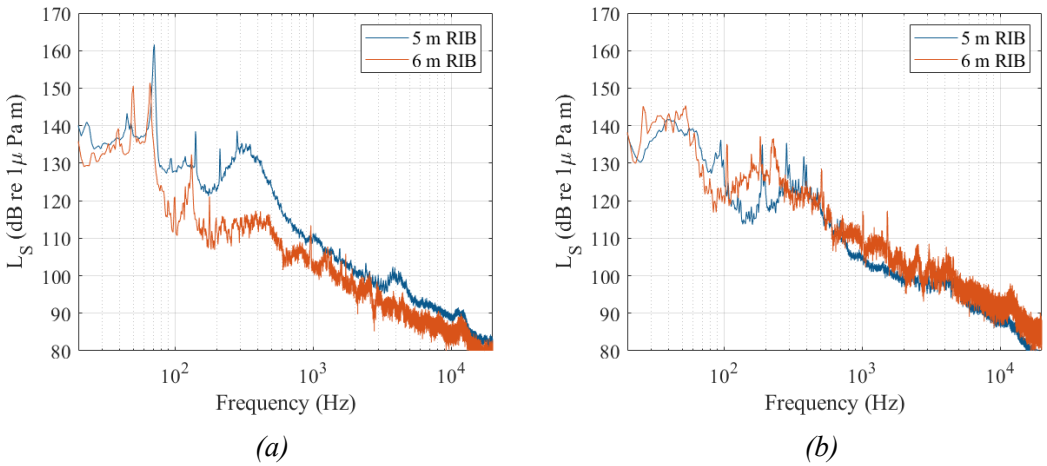


Figure 3: Narrowband source levels for the two vessels at (a) 6 knots and (b) 10 knots

and the broadband noise levels are actually lower for the 5 m RIB at 10 knots than at 6 knots. It was not possible to obtain good camera footage at 10 knots due to the large number of bubbles being entrained in the flow. It may therefore be that these bubbles are attenuating some of the noise and this may partly explain the lower noise levels. It may also be that higher turbulence levels disrupt the formation of the tip vortices. Changes in propeller loading also occur over this speed range, as noted in [9]. Further research is needed to better understand the mechanisms here, as the results show that speeding up can be beneficial from a noise perspective, which is generally not the case for larger vessels.

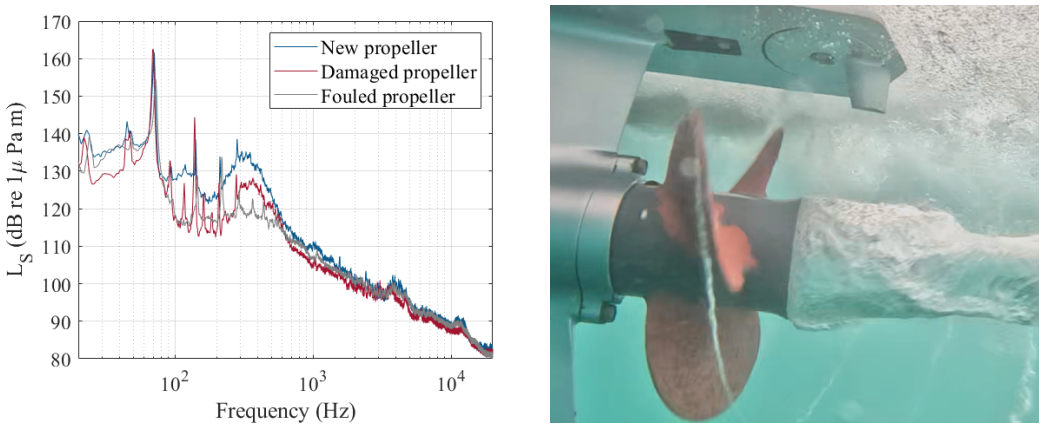
3.3. THE ROLE OF DAMAGE AND ROUGHNESS

It might reasonably be assumed that a rough or damaged propeller would exhibit cavitation at a lower speed than a clean propeller and that the levels of cavitation at a given speed would be worse. A rough propeller will be less efficient and must therefore run faster to achieve the same vessel speed. Roughness elements can also act as nucleation sites by stabilising gas volumes at the propeller surface. The pressure fluctuations due to turbulence created by roughness may also lead to the local pressure dropping below the threshold pressure earlier.

However when TVC is present, studies have shown that roughness at the tip can effectively reduce TVC by modifying the roll-up process and leading to a weaker vortex strength [13, 14]. To provide further insights into how roughness and damage/wear affects propeller noise levels, trials have been conducted at 6 and 10 knots using a clean, slightly damaged, and an artificially roughened propeller. This was roughened by applying 1-2 mm elements randomly across the blade surfaces. The details of the material state of these propellers are given in [10] and photographs are shown in Figure 4. The narrowband source levels are shown in Figure 5 and this confirms the findings of the aforementioned studies. Even a very moderate level of surface wear or roughness can be effective at alleviating TVC. The propeller still cavitates at 6 knots, as shown in Figure 5(b), but it is far more fragmented and does not extend as far downstream. As a result, the noise due to collapse (typically high frequency) is the same but the lower frequency noise associated with the oscillating cavity volumes is significantly reduced. Further research is needed to understand the minimum levels of roughness required to achieve this and to determine whether this effect is observed over a wider range of operating conditions.



Figure 4: The clean (left), damaged (centre), and artificially roughened (right) propellers.



(a) *(b)*

Figure 5: (a) Narrowband spectra for the 5 m RIB at 6 knots with the three propellers. (b) Cavitation on the roughened propeller at 6 knots. Adapted from [10].

4. CONCLUSIONS

This work has presented an overview of recent research carried out into propeller cavitation on small craft. It has been shown that cavitation inception speeds can be as low as 5 knots in calm water, suggesting that cavitation is likely to be present on these vessels at most operating speeds. Tip vortex cavitation has been shown to appear first, accompanied by a significant rise in noise at both low and high frequencies. However, this noise may not always persist and has been shown to sometimes decrease as the vessel speed increases. Furthermore, trials have shown that a rough or damaged propeller can actually produce less noise. Based on previous studies [14], this is because the roughness (or roughness elements associated with tip damage) alter the vortex roll-up process and reduces its strength.

Further research is planned to investigate cavitation levels at higher speeds and to further investigate ways of mitigating tip vortex cavitation without reducing the propeller efficiency.

REFERENCES

[1] T. J. van Terwisga, M. van Rijsbergen, E. van Wijngaarden, J. Bosschers, S. Schenke, and T. Melissaris, “Cavitation nuisance in ship propulsion: A review of developments,” *Cavitation and Bubble Dynamics*, pp. 73–109, 2021.

[2] R. Leaper and M. Renilson, “A review of practical methods for reducing underwater noise

pollution from large commercial vessels," *International Journal of Maritime Engineering*, vol. 154, no. A2, 2012.

[3] L. Wilson, M. K. Pine, and C. A. Radford, "Small recreational boats: A ubiquitous source of sound pollution in shallow coastal habitats," *Marine Pollution Bulletin*, vol. 174, p. 113295, 2022.

[4] L. Hermannsen, L. Mikkelsen, J. Tougaard, K. Beedholm, M. Johnson, and P. T. Madsen, "Recreational vessels without automatic identification system (ais) dominate anthropogenic noise contributions to a shallow water soundscape," *Scientific Reports*, vol. 9, no. 1, p. 15477, 2019.

[5] M. Picciulin, E. Armelloni, R. Falkner, N. Rako-Gospić, M. Radulović, G. Pleslić, S. Muslim, H. Mihanović, and T. Gaggero, "Characterization of the underwater noise produced by recreational and small fishing boats (< 14 m) in the shallow-water of the cres-lošinj natura 2000 sci," *Marine Pollution Bulletin*, vol. 183, p. 114050, 2022.

[6] T. A. Smith, A. G. La Rosa, and B. Wood, "Underwater radiated noise from small craft in shallow water: Effects of speed and running attitude," *Ocean Engineering*, vol. 306, p. 118040, 2024.

[7] M. Shipton, J. Obradović, F. Ferreira, N. Mišković, T. Bulat, N. Cukrov, and R. Diamant, "A database of underwater radiated noise from small vessels in the coastal area," *Scientific Data*, vol. 12, no. 1, p. 289, 2025.

[8] M. Svedendahl, E. Lalander, P. Sigray, M. Östberg, M. H. Andersson, M. Svedendahl, E. Lalander, M. Östberg, and H. Mathias, "Underwater acoustic signatures from recreational boats," *Tech. Rep*, 2021.

[9] T. A. Smith, A. G. La Rosa, G. Piggott, J. A. N. Gaivota, and S. S. McMorran, "Propeller cavitation on small craft: Underwater noise measurements and visualisation from full-scale trials," *Ocean Engineering*, vol. 317, p. 120024, 2025.

[10] T. A. Smith, A. Grech La Rosa, G. Piggott, J. A. N. Gaivota, and S. S. McMorran, "An experimental study of underwater radiated noise from a small vessel with damaged and fouled propellers," *JASA Express Letters*, vol. 4, no. 12, 2024.

[11] A. O. MacGillivray, S. B. Martin, M. A. Ainslie, J. N. Dolman, Z. Li, and G. A. Warner, "Measuring vessel underwater radiated noise in shallow water," *The Journal of the Acoustical Society of America*, vol. 153, no. 3, pp. 1506–1524, 2023.

[12] J. Bosschers, "Mechanisms of low-frequency broadband noise by cavitating tip-vortices on marine propellers," in *Sixth International Symposium on Marine Propulsors SMP*, vol. 19, 2019.

[13] S. Sezen, D. Uzun, O. Turan, and M. Atlar, "Influence of roughness on propeller performance with a view to mitigating tip vortex cavitation," *Ocean Engineering*, vol. 239, p. 109703, 2021.

[14] A. Asnaghi, U. Svennberg, R. Gustafsson, and R. E. Bensow, "Investigations of tip vortex mitigation by using roughness," *Physics of Fluids*, vol. 32, no. 6, 2020.

Regular Session
28. Signal and Image Processing

Hierarchical Multi-Target Tracking in Underwater Environments

Joshua J. Wakefield[1], Stewart G. Haslinger[2], and Jason F. Ralph[1]

[1]Department of Electrical Engineering & Electronics, University of Liverpool
[2]Department of Mathematical Sciences, University of Liverpool

Joshua J. Wakefield, University of Liverpool, Liverpool, L69 7ZF, sgjwakef@liverpool.ac.uk

Abstract: *This paper presents a comparative analysis of tracking filters for passive sonar applications, focusing on scalability under varying signal-to-noise ratio (SNR) conditions. Unlike active sonar, passive systems rely solely on received signals, making tracking particularly challenging due to high noise levels, multipath propagation, and limited measurement information. Existing approaches often apply a single filtering method tailored to specific scenarios; however, no single filter performs optimally across all conditions. While more flexible methods may offer robustness, they typically incur greater computational costs, limiting their practicality for real-time use. Simulation results show that linear Gaussian filters, such as the Kalman filter, are highly effective for high-SNR sources where measurements are reliable and ambiguity is minimal. In these cases, more complex methods offer minimal benefit and can be computationally inefficient. Conversely, sequential Monte Carlo methods, such as particle filters, provide superior performance in low-SNR environments, where non-linearities and sparse detections increase tracking ambiguity at the cost of higher computational load. These findings support a scalable, adaptive filtering strategy in which different filters are used for different sources based on their observed SNR. This approach balances accuracy and efficiency by aligning algorithmic complexity with the difficulty of the tracking task.*

Keywords: *Target tracking, passive sonar, particle filter, kalman filter, state estimation*

1. INTRODUCTION

Passive sonar is a widely used tool for underwater surveillance to detect, track, and monitor the movement of sea vessels, both on the surface and underwater, as well as to observe marine animals and environmental conditions.

Traditional tracking systems often rely on a single filtering method to estimate source trajectories from noisy measurements. Kalman filter variants are commonly employed in underwater tracking, where motion models are approximately linear and noise is Gaussian. However, their performance degrades in the presence of clutter, ambiguous data associations, non-linear measurements, non-Gaussian noise, or multi-modal posteriors [8]. In such conditions, particle filters are often preferred due to their ability to better approximate complex dynamics and distributions [16]. However, this robustness comes at the cost of increased computational complexity, which limits real-time applicability, particularly in systems with power or hardware constraints.

This work supports a hierarchical tracking strategy in which different filters are selected to operate in parallel, with selection informed by each source's observed signal-to-noise ratio (SNR). Rather than proposing a new filter, we analyse the performance of Kalman and particle filters across a range of SNR conditions in two scenarios: *crossing contacts*, where source trajectories intersect, and *close contacts*, where source trajectories remain in close angular proximity for extended periods. Simulation results show that no single method performs optimally across all cases: Kalman filters are well-suited to high-SNR sources with reliable measurements, while particle filters maintain performance in low-SNR settings. By matching filter complexity to tracking difficulty, this approach offers a scalable and resource-efficient solution for multi-target tracking.

2. SIGNAL PROPAGATION AND PROCESSING

We model the signal emitted by a source as a complex sinusoid comprising multiple tonal components, which reflect the mechanical gearing of onboard machinery and form a distinctive narrowband acoustic signature [5]. The source also generates broadband noise, modelled as pink noise (also known as $\frac{1}{f}$ noise), which captures irregular and diverse sounds generated onboard, such as those from human activity, mechanical operations, and other vibrations [3]. Ambient ocean noise, arising from both natural and anthropogenic sources, is modelled as additive white Gaussian noise, consistent with conventional ocean noise characterisations [15].

Acoustic propagation is simulated using Bellhop, a ray-tracing model [12]. For simplicity, the environment is modelled with a flat seabed and a constant ocean depth of 5 km. The sound speed follows the Munk profile, with depth normalised around 1300 m and a surface sound speed of 1500 m/s, consistent with standard deep-ocean conditions [11]. Bellhop provides both transmission loss, which quantifies the reduction in acoustic intensity with range, and ray arrival times. Transmission loss is used to scale the received signal amplitude, while arrival times determine propagation delays and corresponding phase shifts at the hydrophones, which are configured in a static horizontal linear array placed at a depth of 1 km.

The complete signal received at the i-th hydrophone in the array can be expressed as

$$s_i(t) = \left[\sum_{n=1}^{N_t} a_n \exp\left(j2\pi f_n(t + \tau_p + \tau_i) + j\phi_n\right) \right] + \eta_p(t, \tau_p, \tau_i) + \eta_w(t), \tag{1}$$

where N_t is the number of tonal components; a_n, f_n, and ϕ_n are the amplitude, frequency, and

phase of the n-th tonal, respectively; τ_p is the propagation delay from the source to a reference sensor; and τ_i is the relative time delay at the i-th hydrophone due to the array geometry. The terms $\eta_p(\cdot)$ and $\eta_w(\cdot)$ represent the pink and white noise processes, respectively.

To estimate the direction of arrival (DoA) of acoustic signals, we apply a frequency-domain delay-and-sum (DAS) beamformer to the hydrophone data [13]. This method phase-aligns the frequency components of the signals across the array to enhance energy from specific directions while suppressing noise and interference.

Detection is performed in two stages. First, the SNR is computed across discrete bearing angles, obtained by partitioning the angular field of view. A fixed threshold is applied to identify candidate detections. Due to the finite resolution of the beamformer and the presence of sidelobes, strong signals may exceed the threshold in several adjacent bearing angles. To consolidate these into single detections, a peak-finding algorithm identifies local SNR maxima at each time step. The resulting bearing measurements are then associated with tracks over time, enabling the estimation of each source's trajectory and dynamics.

3. TARGET TRACKING

Bayesian target tracking provides a probabilistic framework for recursively estimating the state of a dynamic system given noisy and possibly incomplete measurements. The belief about the system's state is represented by a probability density function (PDF), which is updated as new observations become available through two key steps: prediction and update.

In the prediction step, the prior PDF is propagated forward using the system's motion model, described by the Chapman-Kolmogorov equation [14],

$$p(\mathrm{x}_k \mid \mathrm{z}_{1:k-1}) = \int p(\mathrm{x}_k \mid \mathrm{x}_{k-1}) p(\mathrm{x}_{k-1} \mid \mathrm{z}_{1:k-1}) \, d\mathrm{x}_{k-1}, \tag{2}$$

where x_k and z_k denote the system state and measurement at time step k, respectively. The transition model $p(\mathrm{x}_k \mid \mathrm{x}_{k-1})$ reflects the system dynamics and process noise.

In the update step, the prior is refined using Bayes' rule,

$$p(\mathrm{x}_k \mid \mathrm{z}_{1:k}) = \frac{p(\mathrm{z}_k \mid \mathrm{x}_k) p(\mathrm{x}_k \mid \mathrm{z}_{1:k-1})}{p(\mathrm{z}_k \mid \mathrm{z}_{1:k-1})}, \tag{3}$$

where $p(\mathrm{z}_k \mid \mathrm{x}_k)$ is the likelihood of the measurement given the predicted state, and the denominator normalises the posterior distribution.

In this work, Bayesian target tracking is applied to a passive sonar system. The problem is treated as two-dimensional since all vessels operate on or near the surface relative to the range between the sources and the sensor array, and tracking is performed in the angular domain. The true state of target j at time step k is represented by

$$\mathrm{x}_k^{(j)} = \begin{bmatrix} \theta_k^{(j)} & \dot{\theta}_k^{(j)} \end{bmatrix}^T, \tag{4}$$

where $\theta_k^{(j)}$ denotes the bearing and $\dot{\theta}_k^{(j)}$ the bearing rate of the j-th target. The estimated state $\tilde{\mathrm{x}}_k^{(j)}$ is a filtered approximation of $\mathrm{x}_k^{(j)}$, computed recursively as part of the tracking algorithm.

The target is assumed to follow a nearly constant angular velocity model, where process noise allows small deviations due to unmodelled dynamics such as acceleration,

$$\mathrm{x}_k = \mathbf{f}(\mathrm{x}_{k-1}, \mathrm{v}_{k-1}) = F_k \mathrm{x}_{k-1} + \mathrm{v}_{k-1}, \quad \mathrm{v}_{k-1} \sim \mathcal{N}(0, Q_{k-1}), \tag{5}$$

where F_k is the state transition matrix representing deterministic motion and v_{k-1} is zero-mean Gaussian noise with covariance Q_{k-1}.

In this setup, the measurement provides a noisy observation of the true bearing. Although beamforming is inherently non-linear, it is approximated here by a linear measurement model,

$$\mathrm{z}_k = \mathbf{h}(\mathrm{x}_k, \mathrm{n}_k) = H_k \mathrm{x}_k + \mathrm{n}_k, \quad \mathrm{n}_k \sim \mathcal{N}(0, R_k), \tag{6}$$

where $H_k = \begin{bmatrix} 1 & 0 \end{bmatrix}$ extracts the bearing component from the state, and n_k is zero-mean Gaussian noise with variance R_k, representing sensor and environmental uncertainties. All angular operations are performed with appropriate wrapping to the interval $(-\pi, \pi]$.

The Kalman filter estimates the target's state $\tilde{\mathrm{x}}_k^{(j)}$ and its associated covariance $\tilde{P}_k^{(j)}$ through recursive prediction and update steps [14]. In the prediction step, the motion model $\mathbf{f}(\cdot)$ is applied without process noise to propagate the state estimate, while the effect of process noise is incorporated in the evolution of the covariance:

$$\tilde{\mathrm{x}}_k^{(j)} = \mathbf{f}(\tilde{\mathrm{x}}_{k-1}^{(j)}, 0), \quad \tilde{P}_k^{(j)} = F_k \tilde{P}_{k-1}^{(j)} F_k^T + Q_{k-1}. \tag{7}$$

Upon receiving a new measurement z_k, the filter updates the state and covariance as

$$\tilde{\mathrm{x}}_k^{(j)} = \tilde{\mathrm{x}}_k^{(j)} + K_k(\mathrm{z}_k - \mathbf{h}(\tilde{\mathrm{x}}_k^{(j)}, 0)), \quad \tilde{P}_k^{(j)} = (I - K_k H_k)\tilde{P}_k^{(j)}, \tag{8}$$

where the innovation $\mathrm{z}_k - \mathbf{h}(\tilde{\mathrm{x}}_k^{(j)}, 0)$ quantifies the difference between the actual and predicted measurements, and I is the identity matrix. The Kalman gain,

$$K_k = \tilde{P}_k^{(j)} H_k^T (H_k \tilde{P}_k^{(j)} H_k^T + R_k)^{-1}, \tag{9}$$

determines the relative weight of the new measurement compared to the predicted state.

The particle filter approximates the posterior distribution for each target j using a weighted set of samples, or particles, $\{\tilde{\mathrm{x}}_k^{(i,j)}, w_k^{(i,j)}\}_{i=1}^{N_s}$ [2].

Each particle is propagated independently according to the motion model,

$$\tilde{\mathrm{x}}_k^{(i,j)} = \mathbf{f}(\tilde{\mathrm{x}}_{k-1}^{(i,j)}, \mathrm{v}_{k-1}^{(i,j)}), \quad \text{for } i = 1, \ldots, N_s, \tag{10}$$

where N_s is the number of particles. The particle filter represents uncertainty non-parametrically through the distribution of particles in the target's state space, explicitly incorporating process noise during propagation.

Particle weights are updated based on the likelihood of the new measurement given the predicted state,

$$w_k^{(i,j)} = w_{k-1}^{(i,j)} \cdot p(\mathrm{z}_k \mid \mathbf{h}(\tilde{\mathrm{x}}_k^{(i,j)}, 0)), \quad \text{for } i = 1, \ldots, N_s, \tag{11}$$

where $p(\cdot)$ denotes the likelihood function. Particles that better match the measurement receive higher weights. The weights are then normalised to satisfy $\sum_{i=1}^{N_s} w_k^{(i,j)} = 1$.

Over time, many particles may receive negligible weights, a problem known as *particle degeneracy* [6]. To mitigate this, resampling is performed: low-weight particles are discarded, while those with high weights are replicated proportionally to their weights. This step reallocates computational effort toward more probable regions of the state space and helps maintain a representative and diverse particle set.

The ocean environment is inherently noisy and often produces spurious detections, referred to as *clutter*, which introduces additional uncertainty and complexity. To address this, detections are associated with tracks probabilistically using a well-established data association method,

Joint Probabilistic Data Association (JPDA), which jointly considers all valid associations and assigns likelihood-based weights [4].

A valid association assigns each detection to at most one track and requires that the detection lies within a specified confidence region of the predicted measurement, thereby defining a hypothesis. Missed detections, cases where a track is not associated with any measurement, are also accounted for. Such missed detection hypotheses are assigned likelihoods corresponding to the probability of missed detection and are always considered valid.

Let $\mathcal{H}_k$ denote the set of all joint association hypotheses at time step k, and let $\mathcal{H}_k^{(j)}$ be the set of single-target hypotheses induced for target j. For each $h \in \mathcal{H}_k^{(j)}$, the marginal association likelihood is

$$p(h) = \begin{cases} 1 - (p_d^{(j)} \cdot p_g), & \text{if } h \text{ corresponds to a missed detection,} \\ \dfrac{p(\mathrm{z}_k \mid \mathbf{h}(\tilde{\mathrm{x}}_k^{(j)}, 0)) \cdot p_d^{(j)}}{\kappa}, & \text{otherwise,} \end{cases} \tag{12}$$

where $p_d^{(j)}$ is the detection probability for target j, p_g the gating probability, and κ the spatial clutter density [4]. The missed detection case captures both failure to detect and detections falling outside the gate. The valid association case combines the measurement likelihood and detection probability, normalised by clutter density.

The detection probability p_d depends on factors such as sensor characteristics and target observability. For simplicity, we model it as a function of the target's SNR, using a relationship from the receiver operating characteristic curve for Rayleigh-distributed measurements [7]:

$$p_d = p_{\text{fa}}^{\frac{1}{1+\gamma(\mathrm{x}_k)}}, \quad \text{where} \quad \gamma(\mathrm{x}_k) = 10^{\text{SNR}(\mathrm{x}_k)/10}, \tag{13}$$

with p_{fa} denoting the false alarm probability, and $\text{SNR}(\mathrm{x}_k)$ the target's SNR in decibels at reception [1]. This formulation captures the intuition that stronger signals are more likely to be detected than weaker ones.

The false alarm probability p_{fa} depends on the SNR detection threshold [1]. Assuming a Rayleigh distribution, typical for zero-mean Gaussian noise amplitude detection, it is given by

$$p_{\text{fa}} = \exp\left(-\tfrac{1}{2}\gamma_{\text{thresh}}\right), \quad \text{where} \quad \gamma_{\text{thresh}} = 10^{\text{SNR}_{\text{thresh}}/10}. \tag{14}$$

By considering all hypotheses, weighted by their likelihoods, the filter computes an *expected* state estimate that captures both measurement uncertainty and ambiguity in data association.

For the Kalman filter, each hypothesis contributes a Gaussian component to a mixture model. The final posterior is approximated by reducing the Gaussian mixture to a single Gaussian,

$$\tilde{\mathrm{x}}_k^{(j)} = \sum_{h \in \mathcal{H}_k^{(j)}} p(h) \cdot \tilde{\mathrm{x}}_k^{(h,j)}, \quad \tilde{P}_k^{(j)} = \sum_{h \in \mathcal{H}_k^{(j)}} p(h) \cdot \left(\tilde{P}_k^{(h,j)} + (\tilde{\mathrm{x}}_k^{(h,j)} - \tilde{\mathrm{x}}_k^{(j)})(\tilde{\mathrm{x}}_k^{(h,j)} - \tilde{\mathrm{x}}_k^{(j)})^T \right), \tag{15}$$

where $\tilde{\mathrm{x}}_k^{(h,j)}$ and $\tilde{P}_k^{(h,j)}$ denote state and covariance estimates for target j under hypothesis h.

In the case of the particle filter, the expected estimate is computed by updating each particle's weight through a weighted sum over hypotheses,

$$w_k^{(i,j)} = \sum_{h \in \mathcal{H}_k^{(j)}} p(h) \cdot w_k^{(i,h,j)}, \quad \text{for } i = 1, \ldots, N_s. \tag{16}$$

where $w_k^{(i,h,j)}$ is the weight of particle i under hypothesis h for target j. Resampling is performed only after computing the combined weights, not for each hypothesis. This produces the final particle set $\{\tilde{\mathrm{x}}_k^{(i,j)}, w_k^{(i,j)}\}_{i=1}^{N_s}$ and implements the *expected likelihood particle filter* (ELPF) [9].

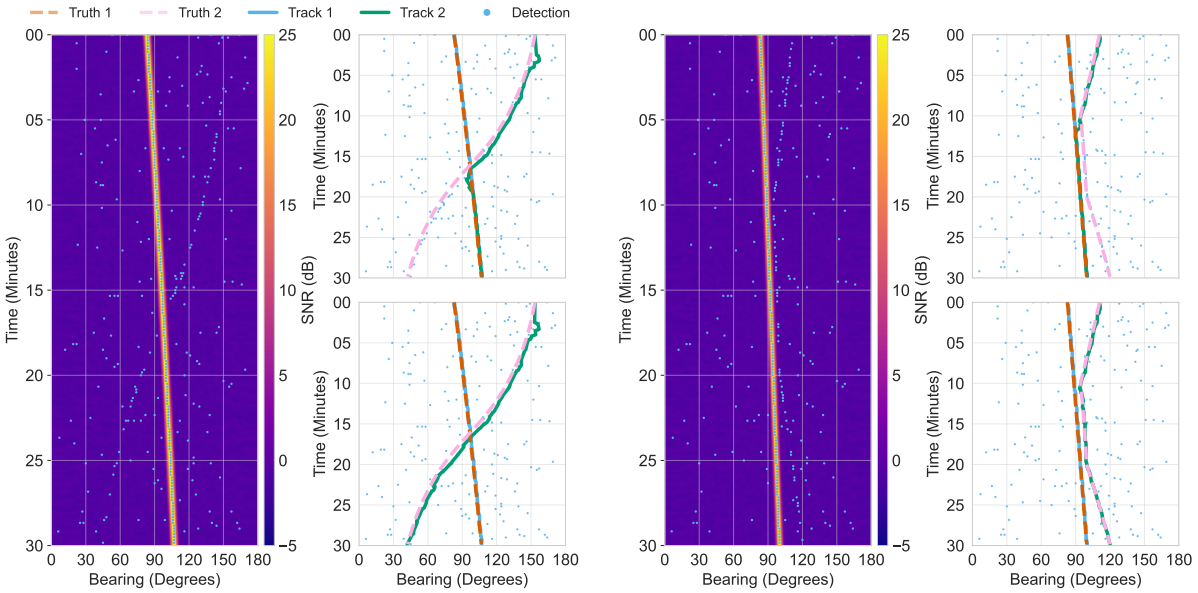
Although both filters use the same Bayesian principles, their computational demands differ. The Kalman filter efficiently updates the state and covariance once per track each time step. The particle filter propagates and updates many particles, then resamples to prevent degeneracy. These steps increase computational load, especially as particle count grows, limiting scalability.

4. RESULTS

Tracking performance is evaluated in two scenarios, both with sources located North of the sensor array's broadside. In the crossing contacts case, Source 1, modelled after a loud modern cargo ship [10], is 40 km away, while Source 2 is quieter and 10 km away; both travel on intersecting paths at 10 m/s. In the close contacts scenario, both sources are approximately 60 km away and remain in close angular proximity for extended periods, challenging the filter's ability to discriminate overlapping measurements. Performance is assessed over 25 SNR values from –10 dB to 2 dB in 0.5 dB steps, with 100 Monte Carlo trials per SNR. SNR is controlled by varying ambient noise between 68 dB and 80 dB, with a fixed detection threshold of 0 dB. Signal parameters are summarised in Table 1, and example trials are shown in Figure 1.

Parameter	Units	Source 1	Source 2
Source Level	dB re 1 µPa at 1 m	[168, 155, 158]	[110, 118]
Frequency	Hz	[97, 72, 580]	[650, 370]
Phase	Degrees	[–112, –151, 86]	[15, 40]

Table 1: Signal tonal parameters for the two sources used in the example tracking scenarios.



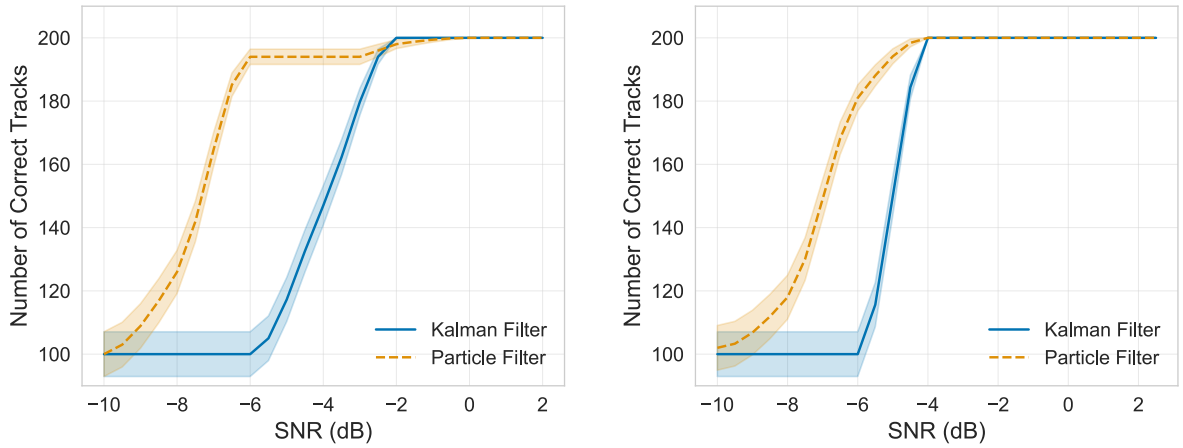
(a) Crossing contacts. *(b) Close contacts.*

Figure 1: Two scenarios showing beamformer output (left), Kalman filter tracks (upper right), and particle filter tracks (lower right). Source/Truth 2 has a SNR of –6 dB.

As shown in Figure 2, the Kalman filter's performance declines sharply below –2 dB in the crossing contacts scenario and below –4 dB in the close contacts case, failing to track the weaker

source at –6 dB and below. The particle filter remains effective down to –10 dB, showing greater robustness at low SNR, while the Kalman filter performs optimally at higher SNRs.

A closer analysis reveals that under low-SNR conditions, the Kalman filter often produces two tracks that follow the same strong source, neglecting the weaker contact entirely. This occurs because the Kalman filter, being a unimodal estimator, tends to converge on the most consistent measurement stream. Degradation of the particle filter's performance is typically due to ambiguity in measurement-to-track association, leading to tracks that intermittently swap between sources. This reflects the particle filter's sensitivity to multi-modal uncertainty.



(a) Crossing contacts. *(b) Close contacts.*

Figure 2: Tracking accuracy across SNR values, showing the number of correctly tracked sources. A track is correct if it remains within 5 degrees of the ground truth bearing. Shaded regions show ± 1 *standard deviation over 100 trials.*

In practice, a SNR threshold may serve as a decision boundary for selecting between Kalman and particle filters in real-time systems. This concept naturally extends to a scalable, hierarchical tracking framework, where multiple filters are available and selected adaptively based on the operating environment. By leveraging the strengths of different algorithms, such a system can maintain robust tracking while optimising computational efficiency across varying conditions.

5. CONCLUSIONS

A comparative analysis of Bayesian tracking filters for passive sonar in bearings-only scenarios has been presented, focusing on how signal-to-noise ratio (SNR) affects filter performance. This work promotes a scalable, hierarchical approach to multi-target tracking. Results from simulations demonstrate that while Kalman filters offer efficient tracking under high-SNR conditions, particle filters are more robust in low-SNR environments where non-linearities and ambiguities are more pronounced. These findings suggest that no single filter is universally optimal across all operating conditions. Instead, identifying a practical decision boundary, such as an SNR threshold, for selecting between filters enables hierarchical multi-target filtering strategies that balance tracking accuracy and computational efficiency by matching algorithm complexity to the difficulty of the tracking task, supporting practical real-time sonar applications.

6. ACKNOWLEDGEMENTS

This work was supported by the EPSRC Centre for Doctoral Training in Distributed Algorithms through a Research Studentship under Grant EP/S023445/1.

REFERENCES

[1] M. A. Ainslie. *Principles of Sonar Performance Modelling*. Springer, 2010.

[2] M Sanjeev Arulampalam et al. "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking". In: *IEEE Transactions on signal processing* 50.2 (2002), pp. 174–188.

[3] Paul T Arveson and David J Vendittis. "Radiated noise characteristics of a modern cargo ship". In: *The Journal of the Acoustical Society of America* 107 (2000), pp. 118–129.

[4] Yaakov Bar-Shalom, Fred Daum, and Jim Huang. "The probabilistic data association filter". In: *IEEE Control Systems Magazine* 29.6 (2009), pp. 82–100.

[5] Richard M Heitmeyer, Stephen C Wales, and Lisa A Pflug. "Shipping noise predictions: capabilities and limitations". In: *Marine Technology Society Journal* 37.4 (2003), pp. 54–65.

[6] Genshiro Kitagawa. "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models". In: *Journal of Computational and Graphical Statistics* 5 (1996), pp. 1–25.

[7] A Vincent van Leijen, Willem L van Norden, and Fok Bolderheij. "Unification of radar and sonar coverage modeling". In: *2009 12th International Conference on Information Fusion*. IEEE. 2009, pp. 1673–1678.

[8] Junhai Luo, Ying Han, and Liying Fan. "Underwater acoustic target tracking: A review". In: *Sensors* 18.1 (2018), p. 112.

[9] Alan Marrs, Simon Maskell, and Yaakov Bar-Shalom. "Expected likelihood for tracking in clutter with particle filters". In: *Signal and Data Processing of Small Targets 2002*. Vol. 4728. SPIE. 2002, pp. 230–239.

[10] Megan F McKenna et al. "Underwater radiated noise from modern commercial ships". In: *The Journal of the Acoustical Society of America* 131 (2012), pp. 92–103.

[11] Walter H Munk. "Sound channel in an exponentially stratified ocean, with application to SOFAR". In: *The Journal of the Acoustical Society of America* 55 (1974), pp. 220–226.

[12] Michael B Porter. "The bellhop manual and user's guide: Preliminary draft". In: *Heat, Light, and Sound Research, Inc., CA, USA, Tech. Rep* 260 (2011).

[13] Barry D Van Veen and Kevin M Buckley. "Beamforming: A versatile approach to spatial filtering". In: *IEEE assp magazine* 5.2 (1988), pp. 4–24.

[14] Greg Welch, Gary Bishop, et al. "An introduction to the Kalman filter". In: (1995).

[15] Gordon M Wenz. "Acoustic ambient noise in the ocean: spectra and sources". In: *The journal of the acoustical society of America* 34.12 (1962), pp. 1936–1956.

[16] Wei Yi et al. "Particle filtering based track-before-detect method for passive array sonar systems". In: *Signal Processing* 165 (2019), pp. 303–314.

Dynamic region method for improving adaptive quantization during sonar data transmission

Hyunjae Song[1], Minho Kim[1], Youngsub Lim[1], and Youngsin Kim[1]

[1]Underwater Surveillance Technology Research and Development, LIG Nex1 Co., Ltd.

Contact author: Hyunjae Song
Address: (13488) 333, Pangyo-ro, Bundang-gu, Seongnam-si, Gyeonggi-do, South Korea
E-mail: hyunjae.song@lignex1.co.kr

Abstract: *Sonar system required massive amount of data on long period of time by receiving data repeatedly. There are many previous study for reduce data rate, sampling reduction, vector quantization. But these studies require to change exist systems. In this paper, I search adaptive quantization bit to fit sonar system and suggest dynamic region method that divide region along with characteristics of signal. Short-time Fourier transform (STFT) analysis and percentage error were conducted. there is less noise occurred from 5-bit re-quantization result. Dynamic interval region method (DIRM) has lower percentage error in every re-quantization bit, maximum percentage error difference is 22.61 % at 4-bit re-quantization result. We demonstrated that our proposed method is capable of effectively applying adaptive quantization method at sonar system and reduce data rate for improving quantization error. It is necessary that address to compare final image conducted rest of sonar signal processing providing to operator.*

Keywords: *Sonar, Adaptive quantization, Data rate, STFT, Dynamic interval region method*

INTRODUCTION

Sonar system conducts sampling and quantization ultrasound signal obtained from acoustic sensor. It transmits data to operator via cable or wireless communication. It requires too much data because it repeats transmitting and receiving. [1] Recently, high-end sonar systems, such as Sonobuoy or Multistatic, require more data because they use several, high frequency sensors simultaneous. But it is difficult to implement high data transfer rate system. Although research has been conducted on compression techniques applied to sonar images, such as a compression technique using the discrete wavelet transform technique [1], a technique using non-uniform quantization and noise-shaping feedback coding [2], and the SPIRT algorithm proposed to complement the weakness of being vulnerable to changes in the input signal [3, 4], a lossless compression technique using the data reordering technique to remove redundancy in data, and a Deflate/inflate algorithm [5, 6]. These are studies on compression techniques applied to sonar images, such as reduce the sampling rate methods to reduce the data rate [7, 8, 9], vector quantization [10, 11], subband partitioning based on Quadrature Mirror Filter (QMF) and methods applying linear prediction coding and entropy coding techniques [12] These methods have been proposed, but changes to the existing system are required.

Recently, adaptive quantization research has been conducted in the field of ultrasound imaging to reduce the amount of data transmitted while using the existing system. [13] This method divides data already quantized into M-bits into regions with equal intervals and obtains the local maxima of each region. It is a technology that re-quantizes the data into N-bits smaller than the initial quantization M-bits, transmits the data, and then multiplies the local maxima to restore the signal. The adaptive quantization method reduces the amount of transmitted data by re-quantizing using only 3 bits, which is less than the 14-bit ADC used in the existing ultrasound imaging system, but signal loss occurs. Since medical ultrasound provides the results as images, the results of evaluating the ultrasound image loss and CNR show that the loss is small and can be applied to ultrasound imaging systems.

In this study, we propose a method to minimize signal loss by reducing the amount of transmitted data and dynamically dividing the regions according to the characteristics of the signal to perform adaptive quantization as the first attempt to apply adaptive quantization to the existing sonar system. To verify this study, we compared the signals restored by the existing equal interval region method (EIRM) that divides the regions into equal intervals using data collected from the currently operating sonar system and the proposed dynamic interval method. As a result, the DIRM proposed in this study can reduce the amount of data transmitted by 6-bits re-quantization to a 16-bits existing sonar system and has a maximum 22.61% lower error rate than the EIRM, which is expected to allow the simultaneous operation of higher frequency acoustic sensors.

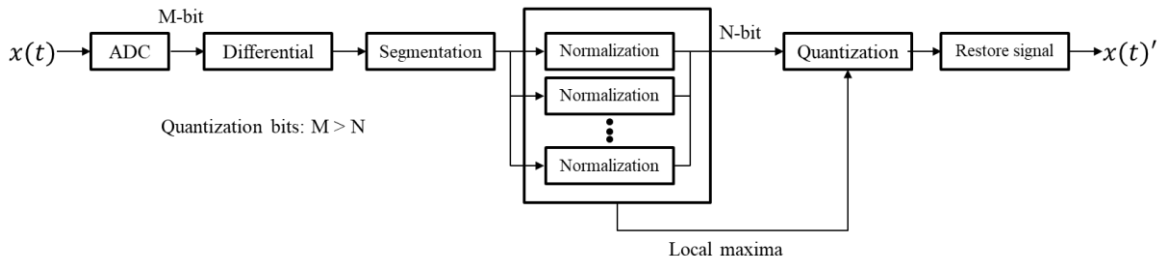


Fig. 1: Proposed Adaptive quantization block diagram with DIRM

METHODS

A. Data

The acoustic signals used in the study were transmitted for 0.2 seconds at 2 second intervals from a circular array sensor and received by an acoustic sensor with a center frequency of 80 kHz and a bandwidth of 50 kHz. Since the transmitted signal is included in the received acoustic signal, it was removed before use.

B. Dynamic interval region method

Since the adaptive quantization technique performs normalization by region, it can minimize signal loss when the dynamic range of each region is set small. The existing EIRM does not consider the characteristics of the signal, so there are large dynamic range regions, but the proposed DIRM differentiates the signal and divides the region when it exceeds the threshold, so it minimizes the dynamic range of each region.

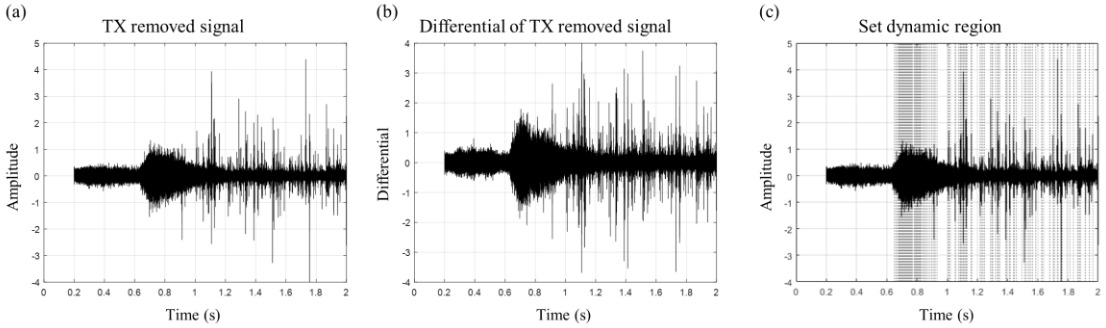


Fig. 2:TX removed signal (a), Differentiation of the signal (b) and the region divided into DIRM (c). The boundaries of the region set according to the change threshold are indicated by dots

The threshold was arbitrarily set to an absolute value of 1, and the boundaries of adjacent regions were set to the same region. The number of regions divided was 196. Normalization was performed for each region divided, and local maxima were calculated. Local maxima are used to restore data re-quantized to low bits.

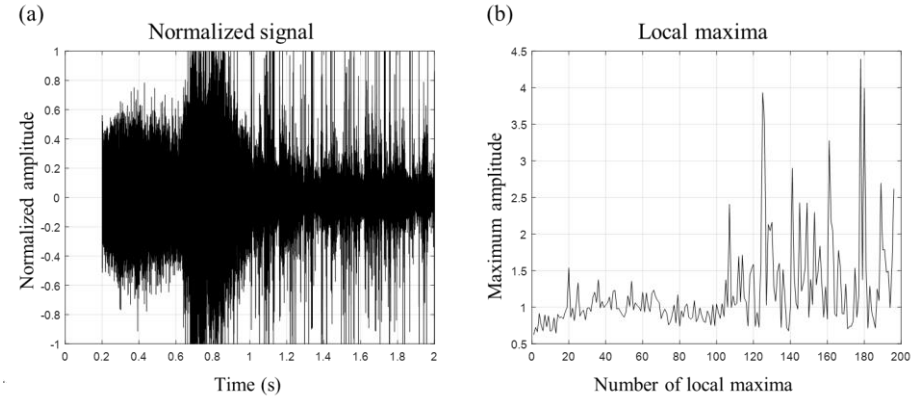


Fig. 3:(a) Normalization was performed by dividing the area by the change threshold (Fig. 2b), and (b) the local maxima, which is the maximum value for each area, was obtained.

C. Short Time Fourier Transform (STFT)

STFT is a method of analysing received acoustic signals according to time and frequency axes, used for target identification and detection in sonar systems. Since acoustic signals are randomly received and statistical characteristics of long-term signals change over time, Fast Fourier Transform (FFT) of the entire signal cannot identify the characteristics of the changing signal. If it is assumed that the characteristics of the signal change relatively slowly and are stationary in a short period of time, it can be analysed using STFT, which divides it into short segments and then performs FFT on each short segment

D. Percentage error rate

The percent error rate is a method of measuring how close two values are. It is essential to evaluate how much of the original data is restored after the signal is re-quantized to a lower bit to reduce the amount of data to be transmitted. The percent error rate was calculated using the STFT Magnitude of the restored signal and the original signal and used for quantitative evaluation. Here, x(t) is the original signal and x' (t) is the quantized restored signal. The threshold was arbitrarily set to an absolute value of 1, and the boundaries of adjacent regions were set to the same region. The number of regions divided was 196. Normalization was performed for each region divided, and local maxima were calculated. Local maxima are used to restore data re-quantized to low bits.

$$\text{Percentage error} = \frac{\sum_{t=1}^{N} |x(t) - x'(t)|}{\sum_{t=1}^{N} |x(t)|} * 100 \quad (1)$$

RESULTS

A, Qualitative evaluation

For qualitative evaluation, we performed STFT for each re-quantization bit. The normalized data by region was re-quantized to 1–15 bits, which is smaller than the 16 bits of the existing sonar system's quantization bits, and only the results for re-quantization 6-bit where the difference from the original is visible to the naked eye, are presented. When comparing the STFT results, the background noise gradually disappeared as the bit increased, and from the 6-bit re-quantization result, the noise components of the restored signal began to differ visually (red arrow) when DIRM (Figure 4b) was performed rather than EIRM (Figure 4a).

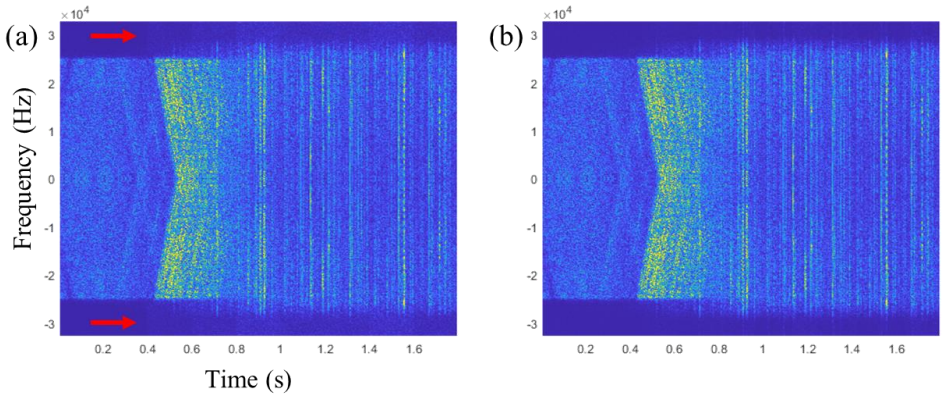


Fig. 4:STFT result of EIRM (a) and DIRM (b)

B. Quantitative evaluation

The error rates of the signals restored after re-quantization with the two methods were measured using Equation 1. When EIRM was performed, the error rates were 31.73%, 16.47%, and 8.20% at 5, 6, and 7 re-quantization bits, respectively, while the DIRM results showed low error rates at all re-quantization bits, at 13.99%, 6.89%, and 3.43%. The largest error rate difference occurred at 4 re-quantization bits, and the DIRM result showed more than twice the error rate up to 5 re-quantization bits compared to EIRM, showing greater similarity to the original signal.

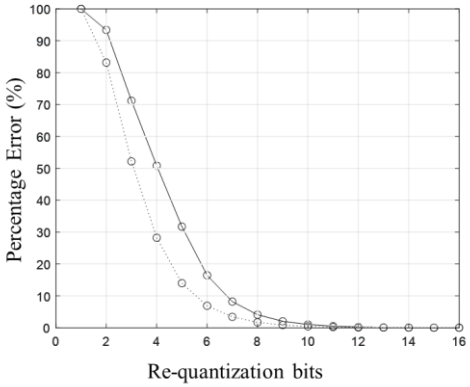


Fig. 5:Error rate according to re-quantization bit when DIRM (dot) and EIRM (line) are used

Re-quantization bit	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Error rate of Equal interval region (%)	100	93.39	71.21	50.85	31.73	16.47	8.20	4.10	2.03	1.02	0.51	0.25	0.13	0.06	0.03	0.02
Error rate of Dynamic interval region (%)	100	83.19	52.17	28.24	13.99	6.89	3.43	1.71	0.85	0.42	0.21	0.11	0.05	0.03	0.01	0.01
Difference of error rate (%)	0	10.21	19.04	22.61	17.74	9.57	4.77	2.39	1.18	0.60	0.30	0.15	0.07	0.04	0.02	0.01

Table 1:Error rate of EIRM and DIRM according to re-quantization bit

DISCUSSIONS & CONCLUSION

In this paper, we first applied adaptive quantization technique to the sonar system under study to reduce the transmission amount with minimizing data loss. We proposed a DIRM to reduce signal loss compared to the EIRM used in adaptive quantization, and proved it quantitatively and qualitatively. From the result of 5-bit re-quantization by DIRM, it was confirmed that less noise occurred in the restored signal even with the naked eye. In addition, when the error rate of the restored signal for each re-quantization bit was calculated, the DIRM had a lower error rate than the EIRM for all re-quantization bits, and the maximum difference was 22.61% at 4 bits.

There are still many aspects that need to be addressed to ensure the practicality of this method. The scope of this study does not yet include sonar images provided to the operator, and the change threshold was arbitrarily set to half of the maximum change threshold. Instead, the error rate was evaluated by focusing on the reduction of signal loss in the existing adaptive quantization. The author's future research will study whether there is a difference from the original signal when performing signal processing up to detection and identification of the restored signal after re-quantization.

ACKNOWLEDGEMENTS

This work was supported by Korea Research Institute for defense Technology planning and advancement(KRIT) grant funded by the Korea government (DAPA-Defense Acquisition Program Administration), (No. 21-107-B00-008(KRIT-CT-23-009), The technical research for Underwater surveillance in open-sea and the analytic technique of acoustical environments in deep water, 2023)

REFERENCES

[1] **Cunha, R. A., M. T. Figueiredo, and C. J. Silvestre**, "Simultaneous compression and denoising of side scan sonar images using the discrete wavelet transform", OCEANS 2000 MTS/IEEE Conference and Exhibition. Conference Proceedings (Cat. No. 00CH37158). Vol. 1. IEEE, 2000.
[2] **Wong, Lok S., Gregory E. Allen, and Brian L. Evans**, "Sonar data compression using non-uniform quantization and noise shaping". 2014 48th Asilomar Conference on Signals, Systems and Computers. IEEE, 2014.
[3] **Bin, Li, and Meng Qinggang**, "An improved SPIHT wavelet transform in the underwater acoustic image compression", Proceedings of 2013 2nd International Conference on Measurement, Information and Control. Vol. 2. IEEE, 2013.
[4] **Bin, Li, and Yang Xiaolei**, "An adaptive lifting wavelet transform in the underwater acoustic image compression." Proceedings of 2012 International Conference on Measurement, Information and Control. Vol. 1. IEEE, 2012.
[5] **Chybicki, Andrzej, Marek Moszynski, and Pawel Pocwiardowski**, "Applications of compression techniques for reducing the size of multibeam sonar records." 2008 1st International Conference on Information Technology. IEEE, 2008.
[6] **Antaneus Feldspar**, “An Explanation of the Deflate Algorithm”, http://zlib.net/
[7] **R. G. Vaughan, N. L. Scott, and D. R. White**, “The theory of bandpass sampling.” IEEE Trans. Signal Process. 39, pp.1973-1984, 1991
[8] **M. Mishali and Y. C. Eldar**, “Sub-nyquist sampling: Bridging theory and practice.” IEEE Signal Process. Mag. 28, pp.98-124, 2011
[9] **J. Kang, H. Yoon, C. Yoon, and S. Y. Emelianov**, “High-frequency ultrasound imaging with sub-nyquist sampling.” IEEE Trans. Ultrason. Ferroelectr. Freq. Control, 69, 2001-2009, 2022
[10] **P. C. Huang and P. C. Li**, “Ultrasound imaging using vector quantization of rf channel data.” IEEE IUS, pp.1303-1306, 2012
[11] **S. M. Hosseini and A. R. Naghsh-Nilchi**, “Medical ultrasound image compression using contextual vector quantization.” Comput. Biol. Med. 42, pp.743-750, 2012
[12] **Kim, Yong Guk, et al**. "An efficient compression method of underwater acoustic sensor signals for underwater surveillance." Sensors 22.9,2022
[13] **D. Jang and H. Yoon**, “Adaptive quantization for effective data-rate reduction in ultrafast ultrasound imaging.” The Journal of the Acoustical Society of Korea, 42(5), 2023

Regular Session
29. Underwater Communication Navigation and Positioning

Uplink Multi-User Underwater Acoustic Orthogonal Frequency Division Multiplexing Communication Based on Vertical Vector Sensor Array

Wei Men[1], Bowen Dong[2], Luxing Zhang[1], Yanan Tian[3,4], and Yong Ren[1]

[1]Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China
[2]College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China
[3]Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China
[4]Key Laboratory of Science and Technology on Advanced Underwater Acoustic Signal Processing,Chinese Academy of Sciences, Beijing, 100190, China

Wei Men, menwei@tsinghua.edu.cn.

Abstract: *Uplink multi-user communication (UMC) is one of the key technologies for establishing underwater acoustic (UWA) sensor networks. However, the long delay and narrow bandwidth of UWA channels pose significant challenges for UMC. In this paper, space-division multiple access is adopted to study a novel UWA UMC method based on a vertical vector sensor array (VSA). Compared with traditional pressure sensor arrays, VSAs offer additional gains and achieve satisfactory UMC performance with smaller array apertures. This avoids the beam pointing errors and signal mismatches caused by the decrease in spatial channel coherence of large-aperture arrays. In particular, a maximum likelihood-based joint angle-delay channel estimation method is researched to acquire the channel state information for each user. To enhance the desired user signal and restrain interference from other users, the vector-wideband-null-steering beamforming method is thoroughly investigated. The performance of the UMC system was analyzed through the South China Sea acoustic channel. Simulation results demonstrate that under an signal-to-noise ratio (SNR) of 20 dB, the proposed method reduces the mean square error (MSE) of data detection by 14 dB compared to traditional pressure sensor arrays. Furthermore, a* 10^{-4} *bit error rate level is achieved in a three-user UWA communication using a 24-element vector sensor array, even at an SNR as low as -5 dB.*

Keywords: *Uplink multi-user communication, vector sensor array, channel estimation, beamforming.*

1. INTRODUCTION

With the continuous advancement of marine technology, underwater operations are transitioning from a single system architecture to a multi-platform network cluster[1]. In this context, uplink multi-user communication (UMC) technologies are critical for enhancing information exchange efficiency among submarines and underwater vehicles. However, mutual interference between users during data detection must be mitigated. Existing multiple access methods commonly employed for this purpose include time division multiple access (TDMA), frequency division multiple access (FDMA), and space division multiple access (SDMA) [2, 3, 4]. Among these techniques, SDMA enables simultaneous full-bandwidth communication during identical time intervals. This method demonstrates superior spectral efficiency, offering distinct advantages over alternative multiple access schemes.

Recently, a vertical sensor array-based multi-user underwater acoustic communication method was proposed in [5]. However, larger array apertures lead to degradation of spatial channel coherence. This results in beam pointing errors and phase mismatches, ultimately causing a decline in spatial gain as the number of elements increases. Unlike conventional pressure hydrophone arrays with identical apertures, vector sensors simultaneously measure acoustic pressure and particle vibration velocity within the sound field. This capability enables vector arrays to generate threefold signal outputs in two-dimensional space, thereby achieving higher array gain. Therefore, our research developed an underwater acoustic (UWA) UMC method based on vertical vector sensor arrays (VSAs). The communication performance was analyzed through the South China Sea acoustic channel. Simulation results verify the feasibility and effectiveness of the proposed scheme.

2. SYSTEM MODEL

A multi-user CP-orthogonal frequency-division multiplexing (OFDM) system is considered, which consists of M equally spaced two-dimensional vector sensor receiving elements and U users. The inter-element spacing is denoted by d. Each vector sensor is equipped with one acoustic pressure channel and two orthogonal particle velocity channels. The y-axis of the particle velocity channels is aligned with the axial direction of the array. The positive x-axis is defined as the rightward orthogonal direction, obtained by rotating the positive y-axis clockwise by 90°. Each user is equipped with a single-element transmitter. One OFDM symbol period is denoted as T, and the subcarrier spacing is given by $\Delta f = 1/T$. The set of subcarrier indices is defined as $S_A = \{-K/2, -K/2+1, \ldots, K/2-1\}$, which includes both data and pilot subcarriers. The center frequency of the subcarriers is f_c, and the frequency of the k-th subcarrier is expressed as $f_k = f_c + k/T, k \in S_A$. The original information bits are encoded using a convolutional code and modulated via 4-QAM. Each user adopts comb-type pilots with an equidistant spacing of Δk_{pilot}. The number of pilot subcarriers is $K_{\text{pilot}} = K/\Delta k_{\text{pilot}}$. The pilot subcarrier indices are allocated as $S_p = \{-K/2, -K/2 + \Delta k_{\text{pilot}}, -K/2 + 2\Delta k_{\text{pilot}}, \ldots\}$.

The signal processing at the receiver is divided into two stages: passband signal preprocessing and baseband signal processing.In the passband signal preprocessing stage, doppler factor estimation is performed for the received passband signal of each array element, followed by correction for time-domain expansion or compression induced by the Doppler effect.

The doppler-corrected signal is down-converted and sampled at the baseband sampling rate. After removing the cyclic prefix and performing DFT demodulation, the frequency-domain

received symbol on the k-th subcarrier for user u is expressed as:

$$\mathbf{z}[k,u] = \sum_{p=0}^{N_u-1} \left(H_p[k,u]s[k,u]\right) \otimes \mathbf{u}_{p,u} + \mathbf{w}[k,u], \quad k \in S_A, \quad u = 1,2,\ldots,U, \tag{1}$$

where $\mathbf{w}[k,u]$ represents the frequency-domain discrete Gaussian white noise with zero mean, and $s[k,u]$ denotes the frequency-domain transmitted symbol on the k-th subcarrier for user u. The vector sensor response matrix is defined as $\mathbf{u}_{p,u} = [1,\ \cos\theta_{p,u},\ \sin\theta_{p,u}]$ and the discrete baseband channel transfer function for the p-th path of user u on the k-th subcarrier is given by:

$$H_p[k,u] = A_{p,u}e^{-j2\pi k\Delta f\tau_{p,u}}\mathbf{a}\left(2\pi f_k\frac{d}{c}\sin\theta_{p,u}\right), \tag{2}$$

where N_u denotes the number of propagation paths for user u, $A_{p,u}$ represents the amplitude of the p-th path for user u, $\tau_{p,u}$ is the time delay of the p-th path for user u, $\theta_{p,u}$ denotes the elevation angle of the p-th path for user u, and c denotes the sound speed in water. The array steering vector is expressed as $\mathbf{a}(\phi) = \left[1, e^{-j\phi}, \ldots, e^{-j(M-1)\phi}\right]$.

The baseband signal processing comprises four stages: channel estimation, beamforming, channel equalization based on spatial diversity combining and demodulation. Detailed descriptions and analyses will be provided in Chapter 3.

3. BASEBAND SIGNAL PROCESSING FOR VERTICAL VSA

3.1. JOINT ANGLE-DELAY CHANNEL ESTIMATION

The pilot symbols are utilized for maximum likelihood-based joint angle-delay channel estimation of the vector array. This process estimates the elevation angle $\hat{\theta}_{p,u}$ and time delay $\hat{\tau}_{p,u}$ for each path of every user. The angle and delay estimation problem is formulated as:

$$\mathrm{L}_{p,u}[\theta,\tau] = \left(\mathbf{D}[\tau]\mathbf{s}_u\right)^{\mathrm{H}}\mathrm{diag}\left(\mathbf{T}^{\mathrm{H}}[\theta]\mathbf{z}_u\right), \tag{3}$$

where $\mathbf{s}_u = \mathrm{diag}\left(s[k,u]\right),\ k \in S_p$, represents the diagonal matrix composed of transmitted pilot symbols for user u. The phase shift diagonal estimation matrix, caused by signal propagation delay, is defined as $\mathbf{D}\left[\tau\right] = \mathrm{diag}\left[e^{-j2\pi k\Delta f\tau}\right], k \in S_p$ and the VSA response estimation matrix, induced by the elevation angle is expressed as $\mathbf{T}(\theta) = \mathbf{a}\left(2\pi f_k\frac{d}{c}\sin\theta\right)\otimes\mathbf{v}(\theta), k \in S_p$. The three-channel combining matrix for the vector sensor, designed under the maximum SNR criterion, is constructed as $\mathbf{v}(\hat{\theta}) = \left[1,\ 2\cos\hat{\theta},\ 2\sin\hat{\theta}\right]$. The angle and time delay estimates of the path p are obtained by searching over the two-dimensional grid:

$$(\hat{\theta}_{p,u}, \hat{\tau}_{p,u}) = \arg\max_{\theta,\tau}\left|\mathbf{L}_{p,u}\right|^2. \tag{4}$$

To estimate subsequent paths, the current strongest path must be subtracted from the matrix $\mathbf{L}_{p,u}$. The elements in the feature function matrix $\mathbf{G}$ is constructed as:

$$\mathrm{G}[\tau-\hat{\tau}_{p,u}, \theta-\hat{\theta}_{p,u}] = \mathrm{Sum}\left(\mathbf{T}^{\mathrm{H}}[\theta]\mathbf{D}^{\mathrm{H}}[\tau]\right), \tag{5}$$

where $\mathrm{Sum}\left(\mathbf{T}^{\mathrm{H}}[\theta]\mathbf{D}^{\mathrm{H}}[\tau]\right)$ denotes the summation of all elements in the matrix $\mathbf{T}^{\mathrm{H}}[\theta]\mathbf{D}^{\mathrm{H}}[\tau]$. The maximum likelihood function matrix for the next path $p+1$ is then updated by subtracting the feature function matrix from the current maximum likelihood function matrix:

$$\mathbf{L}_{p+1,u} = \mathbf{L}_{p,u} - \hat{A}_{p,u}\mathbf{G}, \tag{6}$$

where the amplitude estimate $\hat{A}_{p,u}$ is computed as:

$$\hat{A}_{p,u} = \frac{\mathrm{L}_{p,u}[\hat{\theta}_{p,u}, \hat{\tau}_{p,u}]}{\max(\mathbf{G})}. \tag{7}$$

The iterative process is terminated when the predefined maximum number of paths P is reached. For each user, the elevation angle estimates $\hat{\theta}_{p,u}$ and time delay estimates $\hat{\tau}_{p,u}$ are obtained sequentially for all paths.

3.2. BEAMFORMING

The estimated elevation angles $\hat{\theta}_{p,u}$ are utilized to separate data symbols from different paths using the vector-wideband-null-steering beamforming method. This technique achieves spatial diversity gain in specified directions. Taking the k-th subcarrier of user u and its p-th path as an example, the beamforming weight vector is constructed as:

$$\mathbf{w}_k(\hat{\theta}_{p,u}) = \mathbf{a}\left(2\pi f_k \frac{d}{c} \sin\hat{\theta}_{p,u}\right). \tag{8}$$

The null-steering projection matrix for interfering paths is defined as:

$$\mathbf{P}_k = \mathbf{I} - \mathbf{W}_k \left(\mathbf{W}_k^{\mathrm{H}} \mathbf{W}_k\right)^{-1} \mathbf{W}_k^{\mathrm{H}}, \tag{9}$$

where each column of $\mathbf{W}_k$ corresponds to the beamforming weight vector constructed in the direction of an interfering path. The desired beamforming weight vector is projected onto the orthogonal subspace excluding interference directions using $\mathbf{P}_k$, obtaining the null-steered weight vector for the vector array:

$$\tilde{\mathbf{w}}_k(\hat{\theta}_{p,u}) = \mathrm{c}_k[\mathbf{P}_k \mathbf{w}_k(\hat{\theta}_{p,u})] \otimes \mathbf{v}(\hat{\theta}_{p,u}), \tag{10}$$

where c_k is a normalization constant ensuring unit Euclidean norm $\left\|\tilde{\mathbf{w}}_k(\hat{\theta}_{p,u})\right\|_2 = 1$. In scenarios where interfering paths are angularly close to the desired path, null-steering processing may inadvertently degrade the beamforming performance of the desired path. To address this limitation, serial interference cancellation is employed as an alternative to null-steering for mitigating the impact of interfering paths[6]. Further, the beamformed output symbol for the k-th subcarrier of user u in the direction of the desired path p is computed as:

$$\hat{b}_{k,p,u} = \tilde{\mathbf{w}}_k^{\mathrm{H}}(\hat{\theta}_{p,u}) \mathbf{z}_{k,u}. \tag{11}$$

By repeating this process for all data subcarriers, the vector-wideband-null-steering beamformed output symbols $\hat{\mathbf{b}}_{p,u}$ for user u in the direction of path p are obtained.

4. SIMULATION RESULTS

The UMC performance of the proposed method is evaluated based on the South China Sea acoustic channel simulation. The evaluation metrics include the root mean square error (RMSE) of angle and delay estimation ,the mean squared error (MSE) of data detection and the BER. The RMSE for angle estimation is defined as $\mathrm{RMSE} = \sqrt{\sum_{p=0}^{P-1}\left(\hat{\theta}_{p,u} - \theta_{p,u}\right)^2 / P}$ and the MSE

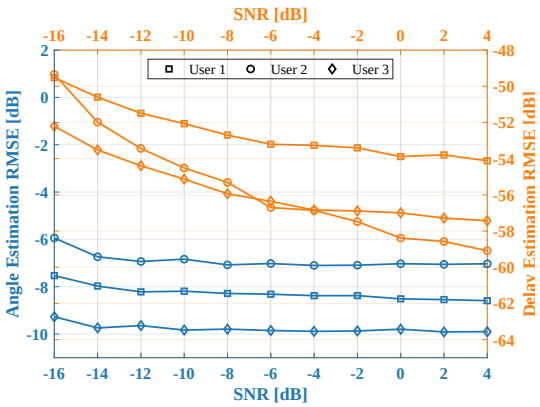
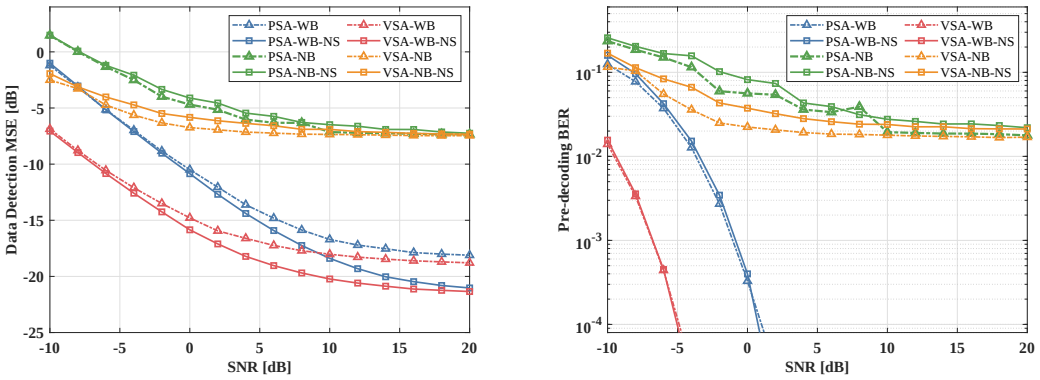


Figure 1: RMSE performance of angle-delay estimation for VSA at various SNRs.



(a) *(b)*

Figure 2: Communication performance of PSA and VSA under different baseband signal processing methods at various SNRs. (a) Data Detection MSE, (b) Pre-decoding BER.

of data detection is defined as MSE $= \left\|\hat{\mathbf{b}}_{k,u} - \mathbf{b}_{k,u}\right\|_2^2 / K$. The simulations use a CP-OFDM system. The transmitter is a single-element source, and the receiver is a 24-element vertical sensor array with an inter-element spacing of 0.15 m. The acoustic channel is modeled using the Bellhop ray-tracing tool for a shallow water environment in the South China Sea, with a sea depth of 170 m and no relative motion between the transmitter and receiver.

First, in the three-user communication scenario, the RMSE performance of angle-delay estimation for the VSA is evaluated under different SNR conditions. As shown in Fig. 1, the angle estimation RMSE for all three users remain nearly identical across the simulated SNR range. The time-delay estimation RMSE decrease progressively with increasing SNR.

The data detection MSE and pre-decoding BER of User 1 are evaluated under different SNR conditions in a three-user communication scenario. Eight distinct schemes are compared, incorporating variations across three dimensions: sensor type, which includes pressure sensor array (PSA) and VSA; beamforming method, either wideband (WB) or narrowband (NB); and null-steering (NS) processing, which is either enabled or disabled. As shown in Fig. 2, the four WB beamforming groups significantly outperform the four NB beamforming groups across all SNRs. This is attributed to the broadband nature of the CP-OFDM system. The superior performance of the VSA groups over the PSA groups with identical signal processing configurations is driven by the joint three-channel processing gain of the vector sensor. However, at low SNR ranges, the advantage of the VSA is less pronounced due to dominant noise effects.

The performance gap widens as the SNR increases. Among the four WB beamforming groups, the two null-steering (NS)-enabled schemes achieve lower MSE and BER compared to their NS-disabled counterparts. In contrast, for the four NB beamforming groups, the NS-enabled schemes exhibit degraded performance relative to the NS-disabled schemes. This degradation is caused by increased angle estimation errors in NB beamforming. The PSA-NB schemes demonstrate the poorest and most unstable performance, particularly at low SNRs. The optimal scheme is the proposed VSA-WB-NS method. At an SNR of 0 dB, the MSE of the VSA is reduced by approximately 5 dB compared to the PSA under WB beamforming. At 20 dB SNR, the MSE reduction reaches 14 dB relative to the PSA-NB scheme. Furthermore, the pre-decoding BER drops below 10^{-4} when the SNR exceeds -5 dB.

5. CONCLUSION

This paper focuses on UWA UMC based on vertical vector sensor arrays. A joint angle-delay channel estimation method and a vector-wideband-null-steering beamforming method are proposed. Simulation test results demonstrate that our scheme achieves a BER on the order of 10^{-4} in a three-user UWA communication scenario, even at a SNR as low as -5 dB. Future research will be focused on the investigation of vertical VSA-based multi-user communication under time-varying channel conditions.

6. ACKNOWLEDGEMENTS

This work was partly supported by the National Natural Science Foundation of China under Grant 62401317 and 62401559, partly supported by the China Postdoctoral Science Foundation under Grant 2024M761710 and GZB20230323.

REFERENCES

[1] W. Men, J. Wang, Z. Li, W. Mi, J. Yin and Y. Ren, “AUV-aided deep-sea Internet of Things: A machine learning perspective,” IEEE Network, early access, Dec. 2024. doi: 10.1109/MNET.2024.3518780.

[2] S. Cang, J. Swärd, X. Sheng, and A. Jakobsson, “Toeplitz-based blind deconvolution of underwater acoustic channels using wideband integrated dictionaries,” Signal Processing, vol. 179, pp: 107812, Feb. 2021.

[3] Y. Zhang, Y. Huang, L. Wan H. Zhou, S. Zhou, and X. Shen,“Adaptive OFDMA for downlink underwater acoustic communications” in 2014 Oceans - St. John’s, St. John’s, NL, Canada, 2014, , pp. 1-5.

[4] J. Yin, P. Du, G. Yang, and H. Zhou, “Space-division multiple access for CDMA multiuser underwater nacoustic commuications” Journal of Systems Engineering and Electronics, vol. 26, no. 6, pp. 1184-1190, Dec. 2015.

[5] D. A. Cuji, Z. Li and M. Stojanovic, “Joint beamforming and tracking for multi-user acoustic communications” in OCEANS 2023-Limerick, Limerick, Ireland, 2023, pp. 1-4.

[6] N. I. Miridakis and D. D. Vergados, “A survey on the successive interference cancellation performance for single-antenna and multiple-antenna OFDM systems,” IEEE Communications Surveys & Tutorials, vol. 15, no. 1, pp. 312-335, Apr. 2013.

Real-Time Underwater Localization System Based on Position-Keeping USVs for Multiple AUVs

Youngsun Moon[1], Minho Kim[2], Gwangho Moon[3], Seokjong Yong[4] and Myungjun Kim[5]

[1]Youngsun Moon, LIG Nex1, Republic of Korea
[2]Minho Kim, LIG Nex1, Republic of Korea
[3]Gwanho Moon, LIG Nex1, Republic of Korea
[4]Seokjong Yong, LIG Nex1, Republic of Korea
[5]Myungjun Kim, LIG Nex1, Republic of Korea

Youngsun Moon, 13488 333, Pangyo-ro, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea, Email: youngsun.moon@lignex1.com

Abstract: *With the advancement of Autonomous Underwater Vehicles (AUVs), the demand for evaluating autonomy and navigation performance has increased significantly. However, real-time localization in underwater environments remains a major challenge due to the infeasibility of GPS and conventional wireless communication. In many countries, localization is typically performed using fixed seabed infrastructure or drifting buoys. However, in Korean coastal waters, securing dedicated test ranges is difficult due to dense maritime traffic and intensive fishing activity, limiting the feasibility of such infrastructure-based approaches. To overcome these limitations, we propose a rapidly deployable localization system utilizing Unmanned Surface Vehicles (USVs) equipped with position-keeping capabilities. The system consists of acoustic transmitters mounted on AUVs, acoustic receivers onboard USVs, and an operation control unit located on a mothership. It enables simultaneous tracking of multiple AUVs using up to nine USVs and six auxiliary drifting buoys, depending on the operational scenario. Acoustic signals transmitted by each AUV are detected by USV-mounted hydrophones at ranges up to 5 km. The received signals and operational data are transmitted to the mothership for integrated control and monitoring. Real-time positioning is performed using Time of Arrival (ToA) and Time Difference of Arrival (TDoA) techniques. Field experiments conducted at a depth of 1,000 m demonstrated an average localization accuracy of approximately 15 m at*

distances of 1, 3, and 5 km, confirming the system's effectiveness and operational reliability in dynamic marine environments.

Keywords: *underwater localization, Real-Time tracking, underwater communication, AUV, USV, Time of Arrival*

INTRODUCTION

Autonomous Underwater Vehicles (AUVs) have increasingly emerged as critical tools in various marine applications, including ocean exploration, environmental monitoring, and underwater defense operations [1], [2]. Accurate real-time localization of AUVs is fundamental to successful mission execution, particularly given the complexities of underwater environments. Due to the impracticality of Global Positioning Systems (GPS) underwater, acoustic-based localization methods are predominantly employed [3], [4].

Conventional acoustic localization systems commonly utilize fixed seabed infrastructure or drifting buoys equipped with acoustic sensors [5], [6]. Although effective, these approaches encounter significant challenges in regions characterized by intense maritime traffic or extensive fishing activities, such as Korean waters. Establishing permanent or semi-permanent localization infrastructure in such dynamic marine areas is often infeasible due to operational constraints, environmental regulations, and the high risk of damage or loss.

Addressing these challenges, recent research efforts have explored alternative, flexible approaches that utilize mobile platforms for underwater localization. Particularly, Unmanned Surface Vehicles (USVs) equipped with precise station-keeping capabilities have demonstrated potential in providing dynamically adaptable localization solutions [10].

In this paper, we propose a novel real-time underwater localization system leveraging multiple position-keeping USVs, designed explicitly for dynamic and complex marine environments. The proposed system integrates acoustic transmitters mounted on AUVs, USV-based acoustic receivers, and a centralized operation control unit located on board a mothership. This configuration allows simultaneous real-time tracking of multiple submerged vehicles utilizing acoustic-based Time Difference of Arrival (TDoA) and Time of Arrival (ToA) localization techniques combined with advanced tracking algorithms for enhanced accuracy [2], [7].

To validate the effectiveness of our proposed system, we conducted extensive simulations and field trials at depths of approximately 1,000 m. The experimental results demonstrate reliable tracking capabilities, achieving positional accuracy within 15 m for targets at distances up to 5 km. These outcomes validate the proposed USV-based localization framework as a viable and robust alternative, capable of addressing the limitations posed by traditional methods in highly dynamic marine operational environments.

UNDERWATER LOCALIZATION SYSTEM DESIGN

This section provides an overview of the system architecture and outlines key design considerations. Specific implementation details have been omitted or anonymized due to

technical confidentiality and security restrictions. The proposed real-time underwater tracking system consists of four main components: an acoustic transmitter, an acoustic receiver, an operation control unit, and a ground control station. A conceptual diagram of the system operation is illustrated in Fig. 1.

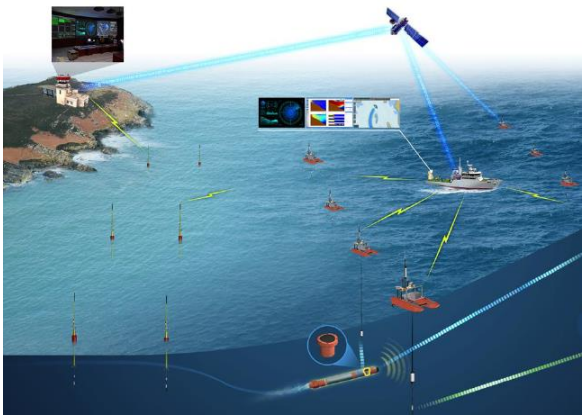


Fig. 1:Real-Time underwater localization system concept diagram

2.1 Acoustic Transmitter

The acoustic transmitter is mounted on the AUV and emits signals detectable by the acoustic receivers. All system components are synchronized to a common GPS time reference to ensure temporal alignment. Transmitted signals are detected by receivers onboard USVs using matched filtering techniques, and the demodulated results are forwarded to the operation control unit on the mothership. The control unit integrates information from multiple receivers to estimate the position of the AUV and transmits the estimated location to the ground station via satellite communication.

The acoustic transmitter is installed on the AUV and is responsible for emitting signals that can be tracked by the receivers. The transmitted signal consists of three components: a tracking signal, an identification (ID) signal, and a depth information signal, as shown in Fig. 2.

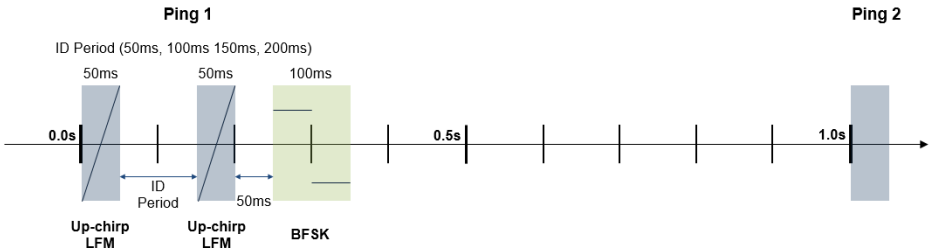


Fig. 2:Transmitted signal concept

The tracking signal is implemented as a linear frequency modulated (LFM) waveform of OOO ms duration, optimized for tracking fast-moving AUVs. Although Doppler effects may cause detection errors, the waveform is tailored to meet the requirements of high-speed underwater platforms. Before AUV deployment, GPS modules are used to synchronize the signal with GPS Pulse Per Second (PPS) timing, enabling Time of Arrival (ToA)-based ranging upon signal reception at the USVs.

The ID signal is introduced to resolve ambiguities arising from the geometry and spacing of the receivers, which may cause time-delay overlaps during detection. In practical scenarios, identifying which combination of detected signals should be used for position

estimation becomes computationally expensive, especially when signals are intermittently detected. To address this, four ID-coded versions of the tracking signal are generated by applying predefined delays to the same waveform. This approach enables unambiguous signal identification at distances of up to 6 km, while also reducing the computational burden of signal matching.

Finally, to compensate for the degraded accuracy in depth (Z-axis) estimation due to the planar geometry of the receiver array, an additional depth signal is incorporated. This signal is implemented as a binary frequency shift keying (BFSK) waveform occupying two sub-bands within the tracking signal spectrum. The total duration of the signal is 000 ms, divided into N bits for BFSK encoding.

2.2 Acoustic Receiver

The acoustic receivers are mounted on twin-hull unmanned surface vehicles (USVs), as illustrated in Fig. 3. This configuration facilitates the integration of a winch system, which is advantageous for deploying and retrieving sensors. Additionally, the physical separation between the signal processing and communication modules and the propulsion unit minimizes the impact of potential water ingress caused by hull damage and provides additional buoyancy.

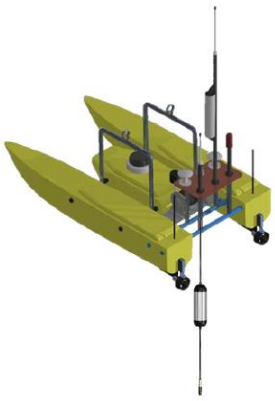


Fig. 3: Position keeping unmanned surface vehicle with acoustic receiver

Each USV is equipped with a navigation controller, which receives commands from the operation control unit and executes them accordingly. The USV travels to the designated position at a limited speed and maintains its location upon arrival. Since the thruster operates intermittently during station-keeping, the detection of acoustic signals is not significantly affected.

A 16-bit analog-to-digital converter (ADC) is used for data acquisition, and a B&K 8106 hydrophone is installed to ensure high-sensitivity acoustic reception. To reduce the influence of ambient noise and enable signal detection over a wide area, analog filters targeting specific frequency bands are applied. The overall signal processing flow is shown in Fig. 4.

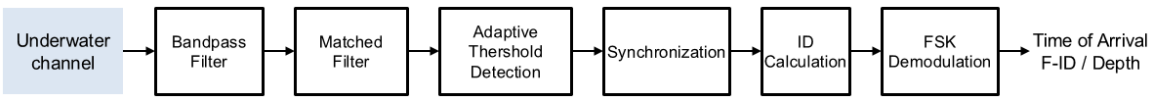


Fig. 4: Signal processing flow chart

For each received signal, a band-pass filter is applied based on the frequency band assigned to each target, followed by detection using a matched filter. When applied to LFM signals, the matched filter yields a signal-to-noise ratio (SNR) gain and enhances

detection performance even under Doppler conditions. The ID and depth information are then demodulated from the detected signals and transmitted to the operation control unit on the mothership.

2.3 Operation Control

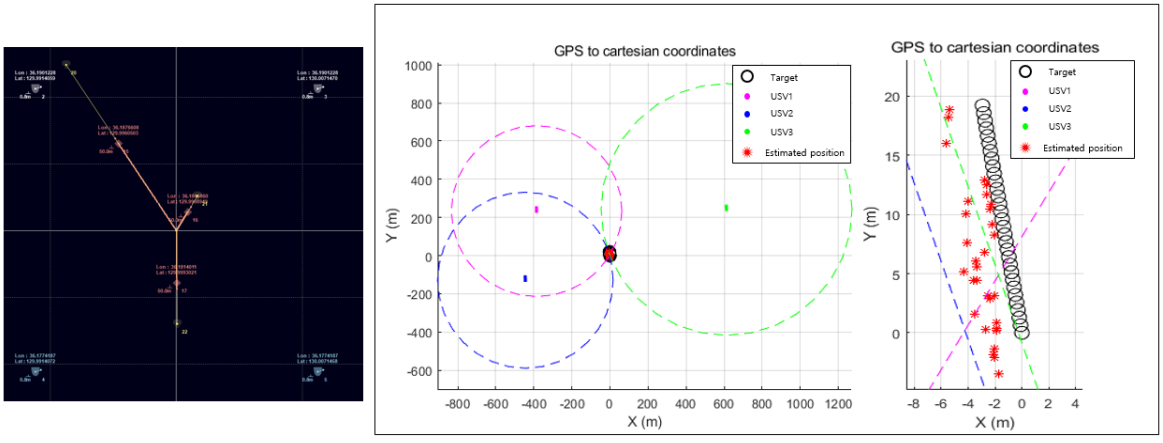
The operation control unit is installed on the mothership and communicates with all station-keeping USVs to receive their data. It has the capability to simultaneously control all USVs, including navigation, signal processing, and mechanical actuation.

When an acoustic signal is detected by a USV, the corresponding detection result is transmitted to the operation control unit. The unit integrates the received information from multiple USVs and performs real-time localization of the AUV. The localization algorithm primarily relies on the Time of Arrival (ToA) method and applies correction using dynamic filters such as the Kalman filter.

3. RESULTS AND CONCLUSION

To validate the system's multi-target tracking functionality, analog signals generated by a signal generator were injected into the receiver under near-realistic conditions. This pre-test, performed in the absence of multiple AUVs due to infrastructure limitations, confirmed that signals transmitted over three distinct frequency bands could be independently detected and tracked without interference shown in Fig. 5-(a)

Subsequent sea trials were conducted in deep-sea conditions in the East Sea of Korea, at depths exceeding 1,000 meters. Three USVs were deployed in a triangular formation, and the system was evaluated at target distances of 1 km, 3 km, and 5 km. Fig. 5-(b) shows that even without applying tracking filters, the system achieved an average localization error of approximately 4 m when the target was centered within the USV array. As the distance increased, the average error grew to 11 m at 3 km and 30 m at 5 km, demonstrating effective long-range tracking performance.



(a) (b)

Fig. 5:Localization results a) multi-target simultaneous tracking result using a signal generator b)sea trial result of the target positioned at the center of the USV array

The signal of the ID and BFSK depth signal decoding results, evaluated at 3 km and 5 km, are shown in Fig. 6. ID decoding achieved perfect accuracy in both scenarios, verifying the reliability of the coded ID system under real oceanic conditions. In contrast, BFSK depth decoding exhibited occasional bit errors at the final position, resulting in

depth estimation errors of up to 1 m. This degradation is attributed to synchronization jitter caused by doppler-induced distortions in the reference tracking signal.

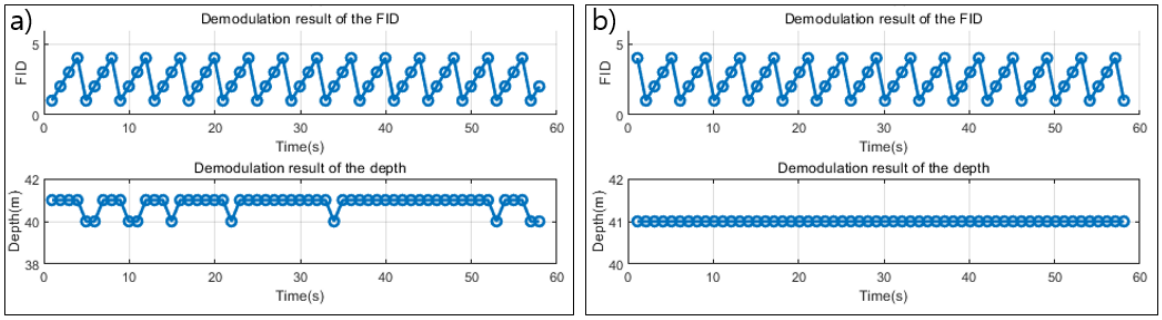


Fig. 6: Signal ID and BFSK demodulation results a) at 3km, b) 5km

Overall, the results confirm that the proposed system enables robust and simultaneous tracking of multiple AUVs, and that its ID and BFSK decoding modules are viable for practical deployment in long-range marine operations. Future work will focus on Doppler-resilient signal design and adaptive filtering to further improve performance under high-speed doppler conditions.

[1] **P. J. B. Sánchez, M. Papaelias, and F. P. G. Márquez**, "Autonomous Underwater Vehicles: Instrumentation and Measurements," IEEE Instrumentation & Measurement Magazine, vol. 23, no. 2, pp. 105–114, 2020.

[2] **P. Wu, S. Su, Z. Zuo, X. Guo, B. Sun, and X. Wen,** "Time Difference of Arrival (TDoA) Localization Combining Weighted Least Squares and Firefly Algorithm," Sensors, vol. 19, no. 11, p. 2554, 2019.

[3] **B. Xu, A. A. Razzaqi, and X. Wang,** "Optimal Sensor Formation for 3D Cooperative Localization of AUVs Using Time Difference of Arrival (TDOA) Method," Sensors, vol. 18, no. 12, p. 4442, 2018.

[4] A. K. M. M. Islam and A. E. Kamal, "A Review of Underwater Localization Techniques, Algorithms, and Challenges," Sensors, vol. 20, no. 1, p. 14, 2020.

[5] **P. Casari and M. Zorzi,** "Protocol Design Issues in Underwater Acoustic Networks," Computer Communications, vol. 34, no. 17, pp. 2013–2025, 2011.

[6] **J. Heidemann, Y. Li, A. Syed, J. Wills, and W. Ye,** "Underwater Sensor Networking: Research Challenges and Potential Applications," in Proceedings of the IEEE Wireless Communications and Networking Conference, Las Vegas, NV, USA, 2006, pp. 228–235.

[7] **H. Yan, Z. J. Shi, and J.-H. Cui,** "DBR: Depth-Based Routing for Underwater Sensor Networks," in Proceedings of the 7th International IFIP-TC6 Networking Conference, Singapore, 2008, pp. 72–86.

[8] **N. Farr, A. Bowen, J. Ware, C. Pontbriand, and M. Tivey,** "An Integrated, Underwater Optical / Acoustical Communications System," in Proceedings of the IEEE OCEANS Conference, Bremen, Germany, 2009, pp. 1–6.

[9] M. Stojanovic and J. Preisig, "Underwater Acoustic Communication Channels: Propagation Models and Statistical Characterization," IEEE Communications Magazine, vol. 47, no. 1, pp. 84–89, 2009.

[10] S. M. Yazdani, M. Nazemizadeh, and A. Barati, "Dynamic Positioning Systems for Marine Vehicles: State-of-the-Art and Future Trends," IEEE Transactions on Control Systems Technology, vol. 26, no. 6, pp. 2021–2034, 2018.

Regular Session
30. Underwater Sensors Technology

Determination of Matched Filter Broadband Beampattern for Arbitrary Antenna Geometries

Christoph Zimmer[1], Johannes Groen[2], Dieter Kraus[3], Nils Theuerkauf[4], Anton Kummert[5]

[1,2,4]ATLAS ELEKTRONIK GmbH, Sebaldsbrücker Heerstr. 235, 28309 Bremen
[3]City University of Applied Sciences Bremen, Neustadtswall 30, 28199 Bremen
[5]Bergische Universität Wuppertal, Rainer-Gruenter-Str. 21, 42119 Wuppertal

Contact: Christoph Zimmer, Christoph.Zimmer@atlas-elektronik.com

Abstract: *The paper addresses the determination of a broadband beamforming function for arbitrary antenna geometries and pulse types based on the application of a matched filter. The proposed calculation method is based on the antenna Point Spread Function (PSF). It is shown that peak sidelobe levels do not occur at the target distance but rather for specific time delays. As the target distance is usually unknown and a multitude of scatterers is present the matched filter broadband beampattern is defined by the maximum level of the PSF for all time delays. This definition avoids a general overestimation of sidelobe suppression capabilities of the given system which is particularly prominent in sparse- or random arrays where high sidelobes for individual frequencies are compensated using a large bandwidth. The proposed calculation method determines the broadband beampattern for arbitrary given antenna geometries also incorporating the applied pulse type and envelope. It is detailed, that the aforementioned antenna PSF can be efficiently calculated in the frequency domain using narrowband beampatterns and shows how this calculation can be simplified under common assumptions. The algorithm is additionally refined since relevant levels in the PSF can only appear for delays shorter or equal to the travel time over the antenna aperture for the respective angle. For example, given universal linear antennas only levels inside a cone of the PSF have to be calculated. For common equidistant linear antennas the highest levels appear for delays with minimal pulse overlap and hence on the cone edges. Hence, by focusing on the relevant sidelobe contributions identified, the overall computational complexity can be significantly reduced.*

Keywords: *Beampattern, Matched Filter, Array Design*

1. INTRODUCTION

The aim of beamforming is to direct the transmitted power of an antenna towards a desired direction or area. A typical model for antenna application applies discrete sources to calculate the beampattern, i.e. the angular sensitivity of an antenna [1] for discrete frequencies. The beampattern can be refined by the application of complex shading coefficients to affect side lobe levels and beamwidth [2].

For active sonar system applications like minehunting and environmental imaging often broadband signals like linear frequency modulated signals (LFM) or non-linear frequency modulated signals, e.g. the hyperbolic frequency modulated signal (HFM), are applied. The broadband beampattern shall incorporate the matched filter (MF) which is predominantly used in active sonar applications and was previously attempted with mediocre estimation quality and high computational effort [3].

2. NARROWBAND BEAMPATTERN CALCULATION

The sound pressure resulting from a given N element antenna array at an observation point can be modelled by the superposition of the individual element pressures. Assuming large distances R to the observation point compared to the extent of the aperture, i.e. far field approximation, the pressure is given by

$$p(t,\phi,\theta|f) = \sum_{n=1}^{N} p_n(t,\phi,\theta|f) \cong \frac{e^{j(2\pi f(t-R/c))}}{R} bp_N(\phi,\theta|f), \tag{1}$$

where t is the time, ϕ and θ are Azimuth and Elevation, f is the frequency and c is the sound speed in water. The equations component only depending on the array geometry, orientation and frequency is denoted as the complex beampattern

$$bp_N(\phi,\theta|f) = \sum_{n=1}^{N} Q_n(f) c_n(\phi,\theta|f) e^{j2\pi f(r_n/c)\cos(\angle(\boldsymbol{O}(\phi,\theta),\boldsymbol{r}_n))}, \tag{2}$$

where $Q_n(f)$ is a complex shading coefficient, $c_n(\varphi,\theta|f)$ is the element characteristic, $\boldsymbol{r}_n$ is the vector pointing towards the n-th element and $\boldsymbol{O}(\varphi,\theta)$ is the unit vector pointing towards the observation point represented in Cartesian coordinates. The logarithm of the squared magnitude of the complex narrowband beampattern given by

$$BP_N(\phi,\theta|f) = 10\log_{10}(|bp_N(\phi,\theta|f)|^2) \tag{3}$$

is denoted as narrowband beampattern.

3. MATCHED FILTER BEAMPATTERN CALCULATION

The time domain matched filter output is calculated based on a transmitted signal $s_T(t)$ with pulse length T_P and the signal $s_o(t)$ observed at an observation point given by the superposition of the signals of the individual elements. Matched filtering of $s_o(t)$ provides the cross-correlation function

$$r_{s_o,s_T}(\tau|\phi,\theta) = \frac{1}{T_P}\int_{\tau}^{\tau+T_p} s_o(t), s_T^*(t-\tau)\, dt. \tag{4}$$

As the target range and location is unknown the corresponding time domain matched filter beampattern is given by the maximum of the cross-correlation function, i.e.

$$BP_{MF}(\phi,\theta) = 10\log_{10}\left(\max_{\tau}\left|r_{s_o,s_T}(\tau|\phi,\theta)\right|^2\right). \tag{5}$$

Since many properties in antenna design are frequency dependent as e.g. the element characteristic it is advantageous to calculate the broadband beampattern in the frequency domain. In [4] it is shown that for a band limited transmitted signal the signal at the observation point can be modelled by the superposition of all frequency components given by

$$\begin{aligned} s_o(t) = p(t,\phi,\theta) &= \int_{-\infty}^{\infty}\sum_{n=1}^{N} p(t,\phi,\theta|f)\, df \\ &= \frac{1}{R}\int_{-\infty}^{\infty}\sum_{n=1}^{N} S_T(f)e^{j(2\pi f(t-R/c))}Q_n(f)c_n(\phi,\theta|f)e^{j2\pi f\left(\frac{r_n}{c}\right)\cos\left(\angle(\boldsymbol{O}(\phi,\theta),\boldsymbol{r}_n)\right)}\, df. \end{aligned} \tag{6}$$

Comparing Eq. (6) with the well-known relation

$$s_o(t) = \int_{-\infty}^{\infty} S_o(f)e^{j(2\pi ft)}df \tag{7}$$

is leading to the relationship

$$S_o(f) = \frac{1}{R}\sum_{n=1}^{N} S_T(f)e^{-j2\pi(R/c)}Q_n(f)c_n(\phi,\theta|f)e^{j2\pi f\left(\frac{r_n}{c}\right)\cos\left(\angle(\boldsymbol{O}(\phi,\theta),\boldsymbol{r}_n)\right)}. \tag{8}$$

Following Eq. (4) the output signal of the matched filter in the frequency domain can be represented by

$$R_{s_o s_T}(f) = \int_{-\infty}^{\infty} r_{s_o,s_T}(\tau|\phi,\theta) e^{-j2\pi f\tau} d\tau = \frac{1}{T_p} S_o(f) S_T^*(f) = \frac{|S_T(f)|^2}{T_p R} e^{-j2\pi f(R/c)} bp_N(\phi,\theta|f). \tag{9}$$

Hence the cross-correlation function can be expressed by

$$r_{s_o s_T}(\tau|\phi,\theta) = \int_{-\infty}^{\infty} R_{s_o s_T}(f) e^{j2\pi f\tau} df = \frac{1}{T_p R} \int_{-\infty}^{\infty} |S_T(f)|^2 e^{-j\left(2\pi f(R/c)\right)} bp_N(\phi,\theta|f) e^{j2\pi f\tau} df. \tag{10}$$

By selecting $\tau = \tau_0 = R/c$ and assuming $|S_T(f)|^2$ to be constant over $B = [f_1, f_2]$ Eq. (10) yields the beampattern at target range which is a typically used approximation for broadband beampatterns given by

$$BP_{B,\tau_0}(\phi,\theta) = 10\log_{10}\left(\left|r_{s_o s_T}(\tau = \tau_0|\phi,\theta)\right|^2\right) \approx 10\log_{10}\left(\left|\frac{1}{b}\int_B bp_N(\phi,\theta|f) df\right|^2\right). \tag{11}$$

The matched filter beampattern at the target range may still differ from the time domain matched filter result in Eq. (5) since it is not necessarily maximum for $\tau = \tau_0$. Defining the time delay centered around the observation point with $\tilde{\tau} = \tau - \tau_0$ Eq. (10) becomes

$$\tilde{r}_{s_o s_T}(\tilde{\tau}|\phi,\theta) = r_{s_o s_T}(\tilde{\tau} + \tau_0|\varphi,\theta) = \frac{1}{T_p R} \int_{-\infty}^{\infty} |S_T(f)|^2 bp_N(\phi,\theta|f) e^{j2\pi f\tilde{\tau}} df. \tag{12}$$

Following the definition of the time domain matched filter beampattern in Eq. (5)

$$BP_{MF}(\phi,\theta) = 10\log_{10}\left(\max_{\tilde{\tau}}\left|\tilde{r}_{s_o,s_T}(\tilde{\tau}|\phi,\theta)\right|^2\right), \tag{13}$$

where the integration in Eq. (12) is usually approximated by the Riemann sum over a set of frequencies $f_i \in B, i = 1,\ldots,N_f$ which leads to the approximation

$$BP_B(\phi,\theta) \approx 10\log_{10}\left(\max_{\tilde{\tau}}\left|\sum_{i=1}^{N_f} |S_T(f_i)|^2 bp_N(\phi,\theta|f_i) e^{j2\pi f\tilde{\tau}} \Big/ \sum_{i=1}^{N_f} |S_T(f_i)|^2\right|^2\right). \tag{14}$$

4. ANTENNA PATTERN EXAMPLES

The differences between the aforementioned beampattern calculations are illustrated with a linear antenna consisting of $N = 21$ equispaced elements which are spaced by $d = 2\lambda = 2c/f_c$ with respect to the band center frequency $f_c = (f_1 + f_2)/2$.

As signal a $T_p = 10$ ms length LFM with a center frequency $f_c = 50$ kHz and $b = 25$ kHz bandwidth is used which approximately fulfills the assumption of a bandlimited signal with constant level in the frequency domain over the selected bandwidth. Example results are obtained without amplitude or phase shading applied. The corresponding beampatterns are given in Fig. 1.

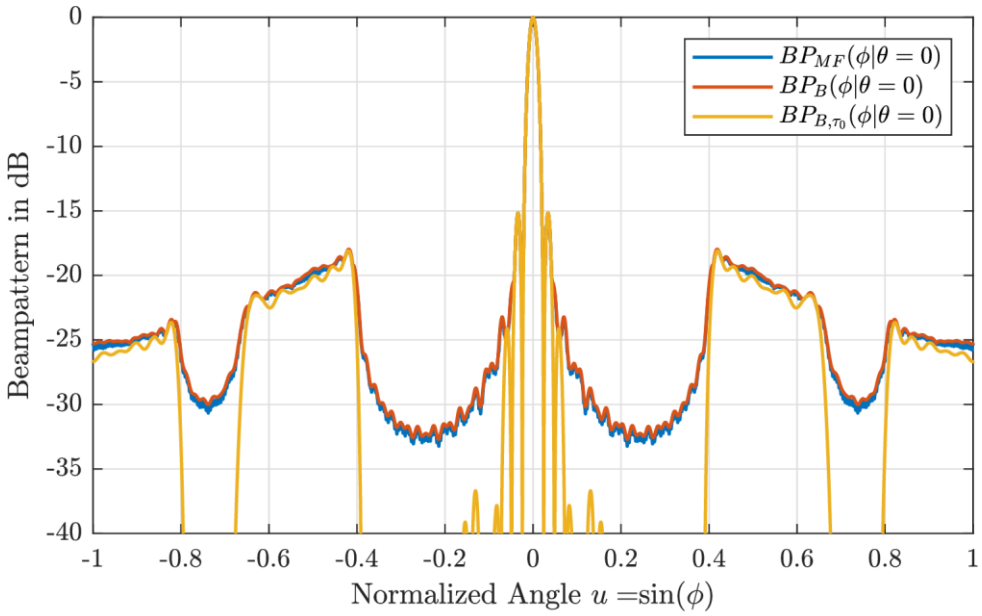


Fig.1: Comparison of the beampattern calculation methods for a uniform linear array

The frequency domain broadband beampattern approximates the time domain matched filter beampattern up to minor errors induced by sampling and band limitation. The matched filter beampattern combines beampattern calculation with the point spread function (PSF) of the applied antenna.

The matched filter beampattern at target range matches well for the main lobe, the first side lobe as well as for the grating lobes, where some underestimation is visible. Nevertheless, all other side lobes are significantly underestimated. This shows, that only calculating the pattern at the target range overestimates the antenna performance. The origin of the deviation can be found by analysing the matched filter output $10\log_{10}\left(\left|\tilde{r}_{s_o,s_T}(\tilde{\tau}|\phi,\theta)\right|^2\right)$ which is given in Fig. 2.

For uniform linear arrays the highest side lobe levels are located on two linear graphs in an "X"-shape. The slope of the graphs is determined by the aperture of the antenna and hence the width increases linearly with $u = \sin(\phi)$. The grating lobes which appear due to the spatial undersampling provide an approximately constant level for all $|\tilde{\tau}| < uL_a/(2c)$ where L_a is the aperture length. Aside from the grating lobes the side lobes take their maximum for delays $\tilde{\tau} \simeq uL_a/(2c)$ and hence, with minimal signal overlap. This

knowledge allows an efficient calculation of the relevant side or grating lobe levels for equispaced arrays.

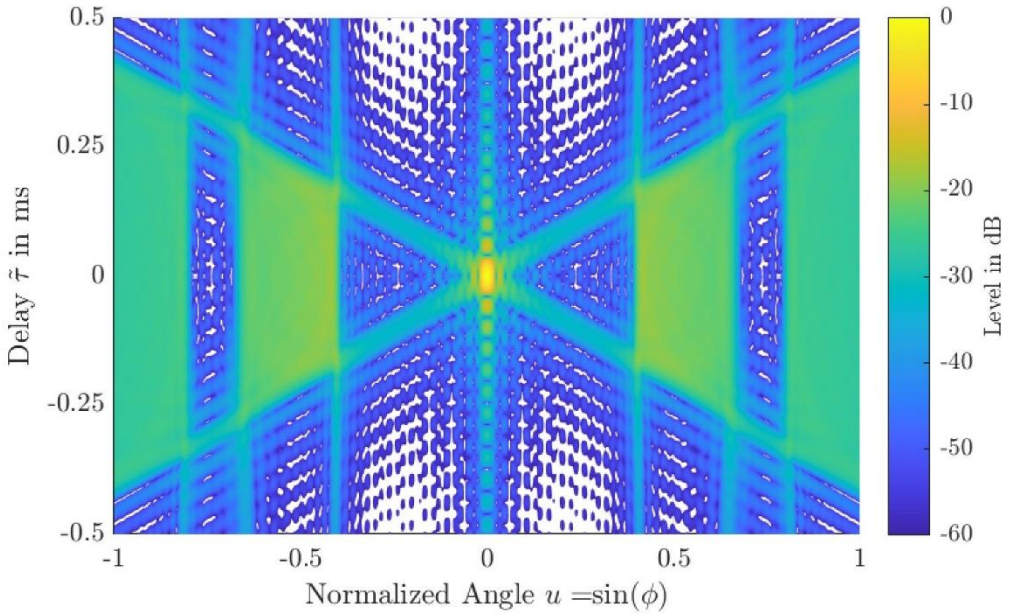


Fig.2: Magnitude of the matched filter output for a uniform linear array

5. CONCLUSION

In this paper a method for broadband beampattern calculation incorporating the matched filter is presented. It allows the calculation in the frequency domain which yields accurate results, is computationally efficient and enables for approximations with minor errors. The presented method is derived using the analogy to the Fourier transform and offers the calculation of the matched filter beampattern from precalculated narrowband beampattern.

The introduced calculation method is compared to a typically applied broadband pattern approximation which shows significant differences for the side lobe level calculation and hence avoids an overestimation of the antenna performance.

While the performance impact for the exemplary linear array is debatable the impact is significantly higher on random array configurations or for optimized shading coefficients which is further highlighted in [4].

REFERENCES

[1] **Urban, H. G.**, *Handbook of Underwater Acoustic Engineering*, Verlag STN ATLAS Elektronik GmbH, 2002.
[2] **Blondel, P.**, *Handbook of Side Scan Sonar*, Springer, 2009.
[3] **Zimmer, C. et. al.**, Transmission Pattern Optimization for Broadband Active Sonar Antennas, In *UACE 2017 Proceedings*, 2017.
[4] **Zimmer, C.**, *Optimization of Underwater Acoustic Transducers and Antennas with Global Optimization Techniques*, Ph.D-Thesis, Shaker Verlag, 2025.

QUANTUM SPIN STATES IN WATER: ORTHO-PARA TRANSITION THEORY AND ITS IMPACT ON ENVIRONMENTAL AND INDUSTRIAL TECHNOLOGIES

Let him who seeks continue seeking until he finds. When he finds, he will become troubled. When he becomes troubled, he will be astonished, and he will rule over the All."

—Gospel of Thomas (2:1–4)

S.M. Pershin[1], I. Bjørnø[2], M.Ya. Grishin[1]

[1] Prokhorov General Physics Institute of the Russian Academy of Sciences, Moscow, Russia

[2] RKMD, Denmark

Abstract: *Recent discoveries reveal that light-induced water evaporation—the "photomolecular effect"—is governed by quantum spin isomerism in H_2O. This study unifies ortho-para transitions, Djanibekov-like rotational instabilities, and Vernadsky's noosphere concept to explain how spin-state dynamics influence atmospheric, oceanic, and industrial systems. Polarized light (532 nm) drives para→ortho conversion via resonance with rotational energy levels, while π-turn flips of ortho-H_2O enhance interfacial processes. These phenomena challenge classical fluid dynamics, offering pathways for climate regulation, sustainable desalination, and pollution control. By framing water's quantum behavior within the noosphere—a biosphere shaped by human intellect—we propose novel solutions for global environmental challenges.*

INTRODUCTION

Water, often regarded as a simple molecule, exhibits complex quantum behaviors due to nuclear spin isomerism. The two spin isomers of water—ortho-H_2O (triplet state with parallel nuclear spins) and para-H_2O (singlet state with antiparallel nuclear spins)—have distinct magnetic properties and rotational dynamics. These isomers

influence hydrogen bonding, molecular interactions, and responses to electromagnetic fields.

The discovery of the photomolecular effect, where water undergoes evaporation upon exposure to light without significant thermal input, challenges traditional thermodynamic understanding. This phenomenon is attributed to light-induced transitions between the ortho and para spin states of water molecules, leading to changes in hydrogen bonding and molecular motion.

Furthermore, the Djanibekov effect, observed in rotating bodies exhibiting sudden flips due to asymmetrical mass distribution, finds a parallel in the rotational behavior of ortho-H_2O molecules. These π-turn flips at the molecular level have implications for fluid dynamics and interfacial processes.

Integrating these quantum phenomena with Vernadsky's concept of the noosphere—a stage of Earth's development where human cognition becomes a significant geological force—offers a framework to explore how manipulating water's quantum states can address environmental and industrial challenges.

QUANTUM SPIN DYNAMICS AND ENVIRONMENTAL INTERFACES

The ortho and para isomers of water differ not only in their nuclear spin configurations but also in their physical and chemical behaviors. Ortho-H_2O, being magnetic and rotationally active, tends to reside in icelike cavities such as aquaporins and nanoparticle hydration shells. In contrast, para-H_2O is non-magnetic and stabilizes structured hydrogen-bonded networks.

In liquid water, the ortho/para ratio is approximately 1:1, whereas in vapor, it shifts to about 3:1, indicating a non-equilibrium state that drives metastability. Exposure to polarized light at 532 nm can resonantly excite para-H_2O molecules, inducing transitions to the ortho state. This process releases energy sufficient to disrupt hydrogen bonds, facilitating the photomolecular effect where water evaporates without significant heat input .

Triplet oxygen (O_2), with its unpaired electrons, acts as a catalyst in spin conversion processes, linking atmospheric chemistry to quantum hydrodynamics. The presence of O_2 enhances the efficiency of para-to-ortho transitions, further influencing evaporation rates and interfacial dynamics.

The Djanibekov effect, characterized by sudden flips in rotating bodies due to asymmetrical mass distribution, is mirrored in the behavior of ortho-H_2O molecules. Their asymmetric rotation can lead to π-turn flips, impacting various systems:

Atmospheric Impact: These molecular flips enhance photon absorption at air-water interfaces, influencing cloud condensation nuclei formation and, consequently, weather patterns.

Biological Systems: In erythrocytes, the deformability facilitated by aquaporin-mediated nanofluidics is affected by the rotational dynamics of ortho-H_2O, optimizing oxygen transport.

Industrial Applications: Spin-selective membranes that filter para-H_2O demonstrate efficiency gains in desalination processes by resisting salt ions, attributed to the distinct behaviors of the water isomers.

NOOSPHERIC APPLICATIONS: BRIDGING QUANTUM AND GLOBAL SYSTEMS

The manipulation of water's quantum spin states offers innovative approaches to address environmental challenges:

Climate Regulation via Spin-State Engineering: Ortho-rich vapor increases light scattering, enhancing cloud albedo and potentially reflecting more sunlight to cool the Earth. Conversely, para-H_2O dominates infrared absorption; adjusting atmospheric ratios could modulate greenhouse effects. Such interventions align with Vernadsky's noosphere concept, where human intellect actively shapes Earth's systems.

Sustainable Desalination and Pollution Remediation: Solar-driven membranes utilizing green light at 532 nm can induce para-to-ortho transitions, reducing water viscosity and enhancing evaporation rates in desalination. Additionally, the free rotation of ortho-H_2O disrupts pollutant adhesion, improving the efficiency of nanofluidic filtration systems.

Quantum Hydrology in the Anthropocene: Spin states influence groundwater dynamics; para-dominated layers can trap heavy metals, affecting contaminant transport. Moreover, the rotational behavior of ortho-H_2O inhibits ice formation, suggesting applications in geoengineering to preserve polar ice caps.

PARADOXES AND FUTURE DIRECTIONS

Despite advancements in understanding water's quantum spin dynamics, several paradoxes challenge existing theories and applications:

1. Quantum Viscosity Paradox

Observations indicate that water with a high concentration of ortho-H_2O exhibits unusually low viscosity in nanoscale channels, such as aquaporins or synthetic membranes. This behavior resembles superfluidity, where friction approaches zero. The paradox lies in how macroscopic fluid properties, traditionally governed by classical hydrodynamics, can be significantly altered by quantum nuclear spin states. This raises questions about the existence of a quasi-superfluid phase of water at room temperature, limited to nanoscale environments, and its implications for biological and industrial fluid transport.

2. Hydration Photonics and Nonthermal Crystallization

In biophysics, ortho-H_2O's high rotational freedom and asymmetric magnetic field oscillations may generate ultrafast electromagnetic pulses, termed "hydration photons." These pulses could trigger protein crystallization without freezing or supersaturation, challenging traditional thermodynamic models. If validated, this mechanism could revolutionize nonthermal preservation of biomolecules, enabling cryo-free storage and precise control of biomineralization, while prompting further investigation into quantum coherence in biological systems.

3. Noospheric Ethics of Quantum Climate Engineering

The potential to manipulate global spin-state ratios through technologies like lasers, nanoparticles, or aerosol catalysts introduces ethical considerations. While such interventions could aid in climate regulation, the unintended consequences on weather systems and biological processes are not fully understood. This necessitates the

development of a "noospheric code"—an ethical framework guiding the responsible application of quantum environmental technologies, ensuring that human interventions align with planetary stewardship principles.

CONCLUSION

Water's quantum spin isomers significantly influence environmental and technological systems, from climate dynamics to industrial processes. By harnessing these quantum properties through the lens of Vernadsky's noosphere, humanity can develop sustainable solutions to pressing global challenges. However, the paradoxes identified underscore the need for cautious advancement, ensuring that the manipulation of quantum states serves to heal rather than inadvertently harm our planet.

REFERENCES

1. **Vernadsky, V. I.** (1945). The Biosphere and Noosphere. Moscow: Nauka. A foundational work introducing the concept of the noosphere as a geologic force of human cognition.
2. **Pershin, S. M., Bjørnø, I., & Grishin, M. Y.** (2024). Water Photomolecular Evaporation Due to Light-Mediated Ortho-Para Spin Transitions. Open Journal of Applied Sciences, 14(8), 1051–1065. https://doi.org/10.4236/ojapps.2024.148147
3. **Pershin, S. M., Bjørnø, I., & Grishin, M. Y.** (2025). Ortho/Para H_2O in Icelike Nanocavities: Quantum Hydrodynamics and Environmental Interfaces. Journal of Experimental and Theoretical Physics (JETP), 130(3), 401–416. https://doi.org/10.7868/S004445102503005X
4. **Pershin, S. M., & Bjørnø, I.** (2024). Light-Mediated Spin Transitions in Water and Their Applications in Nanofluidics. Applied Sciences, 14(11), 2209. https://doi.org/10.3390/app14112209
5. **Felts, A. C., Slimani, A., Cain, J. M., Andrus, M. J., Ahir, A. R., Abboud, K. A., Meisel, M. W., Boukheddaden, K., & Talham, D. R.** (2018). Control of the Speed of a Light-Induced Spin Transition through Mesoscale Core–Shell Architecture. Journal of the American Chemical Society, 140(17), 5814–5824. https://doi.org/10.1021/jacs.8b02148

Poster Session

Distance estimation of near-surface moving targets by deep sonar

Xingyuan Pei[1,2], Xiaochuan Ma[1,2], Chao Feng[1], Yu Liu[1], and Xiaomei Wang[3]

[1]Institute of Acoustics, Chinese Academy of Sciences, Beijing, P. R. China
[2]University of Chinese Academy of Sciences, Beijing, P. R. China
[3]National Satellite Ocean Application Service, Beijing, P. R. China

Xiaochuan Ma, No.2l North Fourth Ring West Road, Haidian District, 100190 Beijing, China, e-mail: maxc@mail.ioa.ac.cn.

Abstract: *Using deep sonar for near-surface target positioning, an inaccurate sound speed profile can result in significant deviation in horizontal distance estimation. Based on a few model assumptions, we propose a high-precision horizontal distance estimation polynomial that eliminates the need for knowledge of the sound speed profile. The input parameters to the polynomial are the echo time delay and incident angle. The polynomial coefficients are only related to the target depth and the sound speed profile. If the depth of the moving target remains constant, the polynomial coefficients can be obtained by estimating the time delays and incident angles of multiple echoes from the moving target. The target's horizontal distance can be obtained by inputting the coefficients, echo time delay and incident angle into the polynomial. Under the MUNK standard sound speed profile, simulation results verify that the proposed method's horizontal distance estimation deviation does not exceed 20 meters. This method is suitable for the design of deep-sea navigation and detection systems.*

Keywords: *Deep sea, Large receiving depth, Target distance estimation.*

1. INTRODUCTION

In the deep sea, deploying sonar near the seabed to detect near-surface targets presents advantages such as stable acoustic channels, minimal propagation loss, and the absence of acoustic shadow zones for near-surface targets within medium detection ranges, as the direct sound paths of target echoes do not contact the sea surface or seabed. The detection framework has great prospects for application. In the deep sea, the underwater sound speed varies vertically within a certain range, and the curve that describes the variation of sound speed with depth in the deep sea is referred to as the sound speed profile[1]. According to ray theory, due to the variation of sound speed with depth in the deep sea, sound rays are curved. Consequently, underwater target positioning methods based on the assumption of straight sound rays will result in significant deviations. Traditional passive[2, 3, 4, 5] and active[6, 7] positioning methods generally require precise sound speed profile parameters, which is unfavorable for acoustic positioning in unfamiliar maritime areas. In the deep sea, this paper proposes a target distance estimation method that does not require sound speed profile parameters for the scenario of deploying the monostatic sonar near the seabed to actively detect underwater targets.

2. DEEP-SEA UNDERWATER SOUND PROPAGATION MODEL

2.1. DEEP SEA SOUND SPEED PROFILE

In the deep sea, the underwater sound speed at the same depth is generally the same over a larger area. Except for the sound speed near the sea surface, which is greatly influenced by the ocean environment, the sound speed initially decreases then increases with increasing depth. A typical model for deep-sea sound speed profiles is the Munk sound speed standard model[8]. The Munk sound speed standard model described as

$$c(z) = c_{\text{min}}[1 + \varepsilon(\eta + e^{-\eta} - 1)]. \tag{1}$$

where $\eta = \frac{2(z-z_0)}{B}$, z is depth. Typical data given by Munk is $B = 1000$ m, $z_0 = 1000$ m, $c_{\text{min}} = 1500$ m/s, $\varepsilon = 0.57 \times 10^{-2}$.

2.2. DEEP-SEA NEAR-SEAFLOOR SONAR BOTTOM-UP DETECTION MODEL

In the deep sea, when the frequency of sound waves is greater than 500 Hz, it is suitable to describe underwater sound propagation with the ray model[9]. the schematic diagram of a monostatic sonar detecting underwater targets near the seabed is shown in Fig. 1, where z_N and z_T are monostatic sonar depth and target depth, respectively. R_T is the horizontal distance to the target from the sonar. Because the sound speed profile $c(z)$ only varies with depth, the direct sound ray to the target is isotropic in the horizontal direction, and the sonar, target, and sound ray lie within the same vertical plane.

According to [7] and [10], the echo parameters of the target and the parameters of the sound

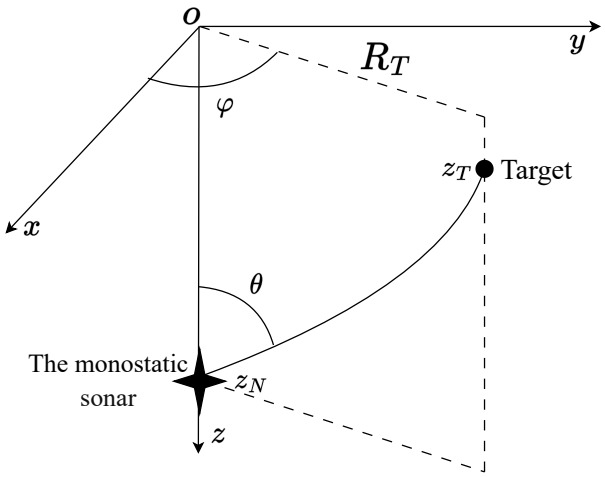


Figure 1: Deep-sea near-seafloor sonar bottom-up detection model.

speed profile satisfy the following three equations:

$$\frac{\tau_r}{2} = \frac{1}{c(z_N)} \int_{z_T}^{z_N} \frac{\left[c(z_N)/c(z)\right]^2}{\sqrt{\left[c(z_N)/c(z)\right]^2 - \sin^2(\theta_r)}} \mathrm{d}z, \tag{2}$$

$$\frac{\tau_r}{2} = \frac{1}{c(z_N)} \left[R_T \sin(\theta_r) + \int_{z_T}^{z_N} \sqrt{\left[c(z_N)/c(z)\right]^2 - \sin^2\left[\theta_r\right]} \mathrm{d}z \right], \tag{3}$$

$$R_T = \int_{z_T}^{z_N} \frac{\sin(\theta_r)}{\sqrt{\left[c(z_N)/c(z)\right]^2 - \sin^2(\theta_r)}} \mathrm{d}z, \tag{4}$$

where τ_r is the two-way delay of the target echo, θ_r is the target's incident angle, where the incident angle refers to the angle between the direction of the echo and the perpendicular line.

3. DISTANCE ESTIMATION METHOD FOR A MOVING TARGET WITH CONSTANT DEPTH

According to the Taylor expansion, Eq. 2 can be denoted as

$$\tau_r = \sum_{n=0}^{\infty} \frac{1}{\cos^{2n+1}(\theta_r)} a_n(z_T), \tag{5}$$

where

$$a_0(z_T) = \int_{z_T}^{z_N} \frac{2c(z_N)\mathrm{d}z}{\left[c\left(z\right)\right]^2}, a_n(z_T) = \frac{(-1)^n(2n-1)!!}{(2n)!!} \int_{z_T}^{z_N} \frac{2c(z_N)x^n\left(z\right)\mathrm{d}z}{\left[c\left(z\right)\right]^2}, n = 1, 2, \cdots \infty. \tag{6}$$

Eq. 3 can be denoted as

$$R_T \sin\left(\theta_r\right) = \frac{c(z_N)}{2}\tau_r - \sum_{n=0}^{\infty} \frac{1}{\cos^{2n-1}(\theta_r)} b_n(z_T), \tag{7}$$

where

$$b_0(z_T) = \int_{z_T}^{z_N} \mathrm{dz}, b_n(z_T) = \frac{(-1)^{n+1}(2n-3)!!}{(2n)!!} \int_{z_T}^{z_N} x^n(z)\mathrm{d}z, n = 1, 2, \cdots, \infty. \tag{8}$$

Eq. 4 can be denoted as

$$R_T = \tan(\theta_r) \left(\sum_{n=0}^{\infty} \frac{1}{\cos^{2n}(\theta_r)} f_n\left(z_T\right) \right), \tag{9}$$

where

$$f_0\left(z_T\right) = \int_{z_T}^{z_N} \mathrm{dz}, f_n\left(z_T\right) = \frac{(-1)^n(2n-1)!!}{(2n)!!} \int_{z_T}^{z_N} x^n\left(z\right) \mathrm{d}z, n = 1, 2, \cdots, \infty. \tag{10}$$

The convergence condition of the Taylor expansion is

$$\left| \frac{x(z)}{\left[\cos\left(\theta_r\right)\right]^2} \right| < 1, z \in [0, z_N], \tag{11}$$

where $x(z) = \left[c(z_N)/c\left(z\right)\right]^2 - 1$. Since the underwater sound speed generally varies from 1450 m/s to 1540 m/s, $|x(z)| << 1$, $\left[c(z_N)/c(z)\right]^2 \approx 1$, there is the following equation

$$b_n(z_T) \approx \frac{c(z_N)a_n(z_T)}{2\left(1-2n\right)}, f_n\left(z_T\right) \approx \frac{c(z_N)}{2} a_n\left(z_T\right), n = 1, 2, \cdots, \infty. \tag{12}$$

According to Eq. 6, when n is larger, $a_n(z_T)$ approaches zero. Combining Eq. 7, Eq. 9, and Eq. 12, the following equation exists

$$R_T \approx \frac{\frac{c(z_N)}{2} \left[\frac{\tau_r}{\cos(\theta_r)} + \sum_{n=1}^{M-1} \frac{a_n(z_T)}{\cos^{2n}(\theta_r)} \frac{2n}{2n-1} \right]}{\left[\tan\left(\theta_r\right) + \cot(\theta_r)\right]}. \tag{13}$$

According to Eq. 13, $a_n\left(z_T\right)$ is required to estimate R_T. The estimation method for $a_n\left(z_T\right)$ is as follows:

According to Eq. 5, if incident angles satisfy Eq. 11, multiple echo parameters $(\theta_{r_i}, \tau_{r_i}), i = 0, 1, ..Q$ for targets at the same depth, but different horizontal position satisfy

$$\tau = \mathbf{A}\mathbf{a}(z_T) + \Delta\mathbf{E}, \tag{14}$$

where $\tau = \left[\begin{array}{cccc} \tau_{r_0} & \tau_{r_1} & \cdots & \tau_{r_Q} \end{array} \right]^{\mathrm{T}}$, $\mathbf{a}(z_T) = \left[\begin{array}{cccc} a_0(z_T) & a_1(z_T) & \cdots & a_{M-1}(z_T) \end{array} \right]^{\mathrm{T}}$, $\Delta\mathbf{E} = \left[\begin{array}{cccc} \Delta e_0 & \Delta e_1 & \cdots & \Delta e_Q \end{array} \right]^{\mathrm{T}}$. and $\mathbf{A} = \begin{bmatrix} \cos^{-1}(\theta_{r0}) & \cos^{-(2n+1)}(\theta_{r0}) & \cdots & \cos^{-(2M-1)}(\theta_{r0}) \\ \cos^{-1}(\theta_{r1}) & \cos^{-(2n+1)}(\theta_{r1}) & \cdots & \cos^{-(2M-1)}(\theta_{r1}) \\ \vdots & \vdots & \ddots & \vdots \\ \cos^{-1}(\theta_{rQ}) & \cos^{-(2n+1)}(\theta_{rQ}) & \cdots & \cos^{-(2M-1)}(\theta_{rQ}) \end{bmatrix}$.

According to the least squares method, $\mathbf{a}(z_T)$ is denoted as

$$\mathbf{a}(z_T) = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\tau. \tag{15}$$

We refer to the aforementioned method as the Moving Target Distance Estimation Method(MTDE). If the depth of the moving target remains unchanged, the steps for estimating the target's distance are as follows:

1. Track moving targets and estimate the target's multiple echo parameters $(\theta_{r_i}, \tau_{r_i}), i = 0, 1, ..Q$ at different locations.
2. Construct **A** and τ according to Eq. 14.
3. Calculate $\mathbf{a}(z_T)$ according to Eq. 15.
4. Calculate the target's horizontal distance R_T by substituting the target echo parameters (θ_r, τ_r) at the current position and $\mathbf{a}(z_T)$ into Eq. 13.

4. SIMULATION ANALYSIS

The true value of the target distance is R_T, while the estimated value is $\hat{R}_T$. We define the target's distance estimation deviation is $\Delta R = \left| R_T - \hat{R}_T \right|$. Using the Munk standard sound speed profile as a background, the depth of the monostatic sonar is set at 4000 meters underwater. Simulate a moving target with a depth of 100 m approaching sonar from a horizontal distance of 14 km, and calculate the corresponding target echo parameters. What's more, only the sound speed $c(z_N)$ at the depth at which the sonar is located is known when estimating the target distance. The estimated target horizontal distance calculated using a geometric method based on the assumption of a straight-line sound path is

$$\hat{R}_{T_i} = \frac{\tau_{ri} \sin(\theta_{ri})}{2} c(z_N). \tag{16}$$

Select M equals 5, the comparison results among the MTDE method, the geometrical method and the real horizontal distance of the target are shown in Fig. 2.

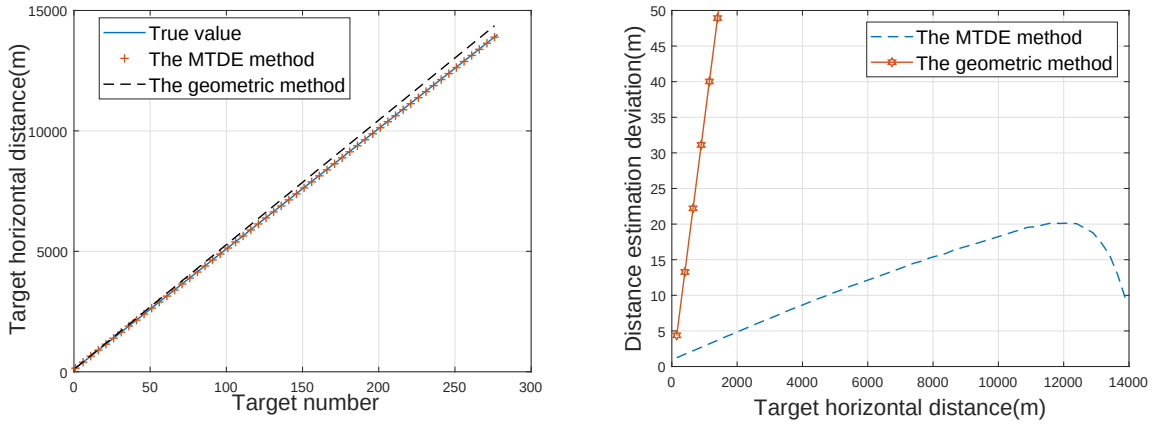


Figure 2: The comparison between the distance estimation and the true value is shown in (a), and the deviation of distance estimation is shown in (b).

With the Munk standard sound speed profile as the background, the pictures in Fig. 2 show that the estimation results of the MTDE method are more accurate, and the distance estimation deviation is less than 20 m.

5. CONCLUSION

In the deep sea, against the background of placing a monostatic sonar system near the seabed for underwater target detection, this paper derives a distance estimation polynomial with the target's incident angle and time delay as independent variables. If the moving target's depth remains constant, this paper introduces a solution for estimating the polynomial coefficients using echo parameters, and further proposes a target distance estimation method that does not require sound speed profile parameters. Under the MUNK standard sound speed profile, simulation results verify that the horizontal distance estimation deviation of the proposed method does not exceed 20 meters.

6. ACKNOWLEDGEMENTS

This work is supported by the National Key R&D Program of China under Grant 2022YFC3101900.

REFERENCES

[1] Urick, RobertJ. *Principles of underwater sound /-3rd ed*. (1983).

[2] Duan, Rui, Yang, Kun-De, Ma, Yuan-Liang, and Lei, Bo. "A reliable acoustic path: Physical properties and a source localization method", *Chinese Physics B* **21**(12), 124301 **(2012)**.

[3] Li, Hui, Yang, Kunde, Duan, Rui, and Yang, Qiulong. "Multipath-based passive source range localization with a single hydrophone in deep ocean" in *OCEANS 2016-Shanghai*, IEEE, 1–5 **(2016)**.

[4] Li, Hui, Xu, Zhezhen, Yang, Kunde, and Duan, Rui. "Use of multipath time-delay ratio for source depth estimation with a vertical line array in deep water", *The Journal of the Acoustical Society of America* **149**(1), 524–539 **(2021)**.

[5] Yang, Kun-De, Yang, Qiu-Long, Guo, Xiao-Le, and Cao, Ran. "A simple method for source depth estimation with multi-path time delay in deep ocean", *Chinese Physics Letters* **33**(12), 124302 **(2016)**.

[6] Chen, Chunhui, Ruang, Haining, Chi, Cheng, Li, Yu, and Jin, Shenglong. "Estimating Target Depth for Deep-Sea Active Sonars Through Channel-Impulse-Response Matching" in *OCEANS 2024-Singapore*, IEEE, 1–5 **(2024)**.

[7] Zhang, Tongwei, Yan, Lei, Han, Guangjie, and Peng, Yan. "Fast and accurate underwater acoustic horizontal ranging algorithm for an arbitrary sound-speed profile in the deep sea", *IEEE Internet of Things Journal* **9**(1), 755–769 **(2021)**.

[8] Munk, Walter H. "Sound channel in an exponentially stratified ocean, with application to SOFAR", *The Journal of the Acoustical Society of America* **55**(2), 220–226 **(1974)**.

[9] Etter, Paul C. *Underwater acoustic modeling and simulation*. CRC Press **(2018)**.

[10] Liu Bosheng. *Principles of underwater acoustics, 3rd ed*. Beijng: Science Press **(2019)**.

Clutter Reduction in Active Sonar Using a Clustering Method

MyoungJun Cheong[1], Juho Kim[1], and Hyeon-Deok Cho[1]

[1]Agency for Defense Development (ADD), Republic of Korea

MyoungJun Cheong, ADD, 51678 Jinhae P.O.Box 18, Changwon, Republic of Korea, mjjeong@add.re.kr

Abstract: *In shallow-water environments, active sonar pulses are repeatedly reflected off the seabed and sea surface as they travel to underwater targets. This generates heavy clutter and multiple measurements from targets, making it challenging to design algorithms that effectively remove clutter while preserving target measurements for tracking. Clutter reduction methods, such as k-means clustering method and hough transform-based track initialization method, have been proposed to mitigate heavy clutter. However, these approaches rely on the assumption that targets follow constant-course maneuvers to distinguish them from clutter, which limits their effectiveness in scenarios involving random target maneuvers. In this paper, we propose a novel clutter reduction method to address heavy clutter in shallow-water multipath detection environments. It applies HDBSCAN clustering method to segment measurements into multiple clusters, followed by a cluster classifier to differentiate target clusters from clutter clusters. The cluster classifier considers the spatial and temporal consistency of accumulated measurements to distinguish between target clusters and clutter clusters. The performance of the proposed clutter reduction method is evaluated using sea trial data collected near Jeju Island, South Korea, with an Active Towed Array Sonar system deployed from ADD Research Vessel CheongHae.*

Keywords: *Active sonar, Clutter rejection, Clustering, Classification, Multipath, HDBSCAN*

1 INTRODUCTION

Active sonar has become the preferred method for detecting submarines in modern ASW. The traditional method of detecting active sonar targets is the thresholded matched filter detector. The conventional thresholded matched filter detector declares a target present when the normalized matched filter output exceeds a predefined threshold [1]. However, this approach has limitations, particularly in complex underwater environments, as it results in a higher incidence of false alarms, generally called clutter. When real targets and clutter are detected simultaneously, operators should manually distinguish between them. Manually classifying sonar signals by operators can cause delays in target declaration and increase a system's workload.

In previous studies, researchers attempted to remove clutter by tuning thresholds based on the signal characteristics of clutter [2]. However, this approach has limitations, as it is not robust account for the highly variable nature of clutter, which differs significantly depending on the underwater environments. In particular, the Yellow Sea and South Sea of Korea are shallow-water environments where multipath detection is prevalent. In such shallow-water environments, heavy clutter is generated by underwater objects and seabed reflections, further complicating the distinction between targets and clutter. In shallow-water environments, active sonar pulses are repeatedly reflected off the seabed and sea surface as they travel to targets. It generates heavy clutter and multiple measurements from targets, making it challenging to design algorithms that effectively remove clutter while preserving target measurements for tracking.

Recently, clustering-based clutter removal methods, such as k-means clustering method and hough transform-based track initialization method, have been proposed for heavy clutter environments [3],[4]. However, previous methods often require predefined parameters and operate under the assumption that target movement follows a constant course and speed. This assumption limits performance in multipath detection environments and during random target maneuvers.

In this paper, a novel clutter removal algorithm is proposed. The proposed method is independent of target movement and remains effective in multipath detection environments, where a target produces multiple measurements amidst high-density clutter. The proposed clutter removal method distinguishes targets from clutter by analysing the spatial and temporal consistency of accumulated measurements. It employs clustering techniques to group measurements into clusters and applies a cluster classifier, based on their temporal and spatial distribution, to differentiate target clusters from clutter clusters. The performance of the proposed clutter removal technique has been validated using sea experimental data, demonstrating its effectiveness.

2 PROPOSED METHOD

2.1 Sea Trial Data

Sea experiment was conducted at the eastern sea of Jeju Island in september 2022, as depicted in Fig.1. The sea trial data was obtained using the Active Towed Array Sonar installed on the Research Vessel (RV) CheongHae of Agency for Defense Development (ADD). The towed active sonar system consists of a variable-depth active sensor for

transmitting signals and a towed linear array of sensors for receiving signals. The active sensor used a LFM pulse with a bandwidth of *400* Hz and a pulse length of *1* second. The transmission frequency was similar to that used in Murphy's research [5]. The receiving sensor is an array of triplet sensors designed to receive signals within the transmission frequency range. The sea experiment was conducted in a shallow environment at a depth of approximately *150* meters, with active and passive sensors' depth of *50* meters. The submarine involved in the sea experiment was a medium-sized submarine. In the sea experimental environment, the sound propagation path involved repeated reflections from the sea bottom and the surface, resulting in a multipath detection environment, as shown in Fig.1. After obtaining the beamformed sensor signals from the sea experiment, matched filtering and two-dimensional normalization using CFAR are performed. Measurements exceeding a specific SNR threshold are extracted, and the accumulated results over a total of *120* pings are shown in Fig.1. The target moved in a clockwise circular motion at approximately *9* km and a speed of about *4* knots, while the RV CheongHae moved westward at a constant speed of *5* knots.

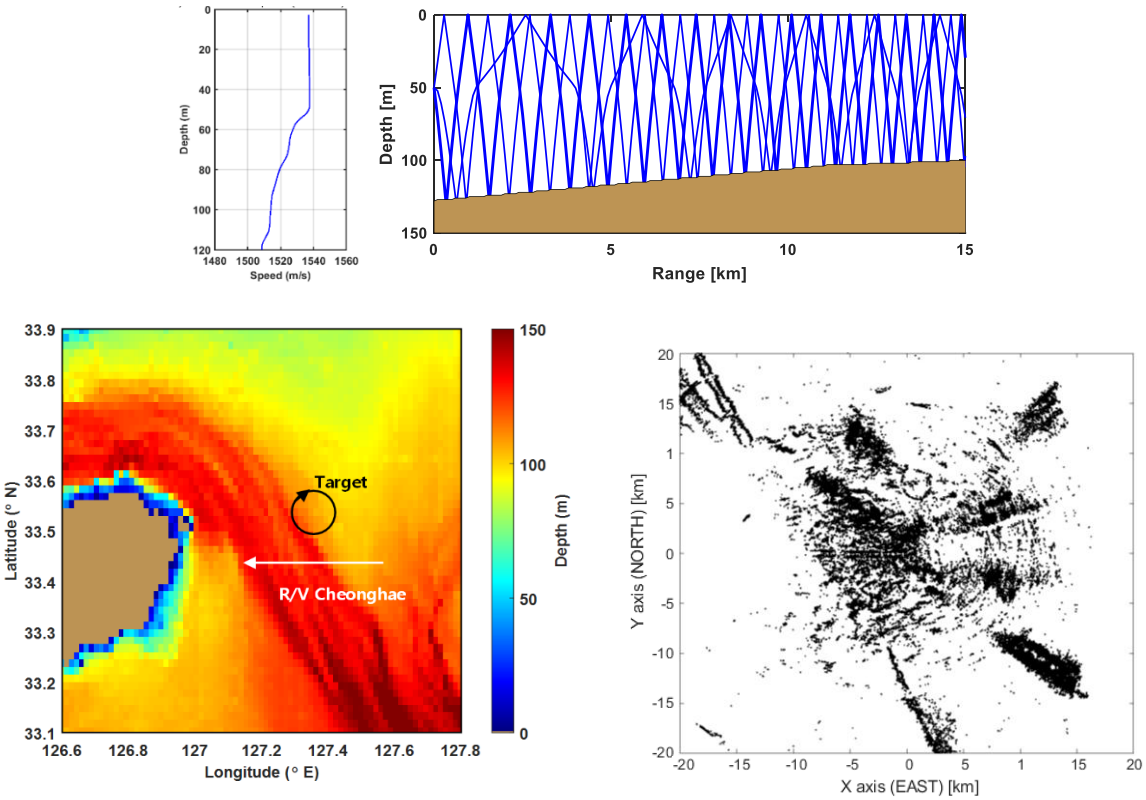


Fig.1: The top graph shows sound velocity profile and the ray trace. The bottom left panel shows rendering of the bathymetry data and navigation path. The bottom right panel shows accumulated measurements (120 pings) which exceed a predefined SNR threshold.

2.2 Clustering Method

HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) is a type of density-based clustering technique that, unlike DBSCAN, effectively groups data with varying characteristics by allowing different densities for each cluster [6]. Moreover, it offers the advantage of lower tuning costs due to fewer parameters compared

to DBSCAN. It is not necessary for the HDBSCAN method to define a minimum distance ε to determine core points; instead, only the minimum number of neighbors *minPts* needs to be defined. The HDBSCAN technique uses mutual reachability distance between all points to form a hierarchical structure rather than the DBSCAN method checks whether the Euclidean distance between data points is less than the minimum distance ε. The mutual reachability distance is defined as shown in equation (1).

$$d_{mreach,k}(a,b) = \max\{core_k(a), core_k(b), d(a,b)\} \tag{1}$$

where, $core_k(a)$ is the core distance of a point, indicating the radius of a cluster that contains at least k points surrounding the core point. $d(a,b)$ is the distance between the core point and the center of the cluster. The hyperparameter k is typically set to the same value as *minPts*. A hierarchical structure of data points can be established by gradually reducing the mutual reachability distance from a value large enough to group all points into a single cluster to a smaller value where each point is classified as its own cluster. The hierarchical structure can be visualized as branches extending from a single cluster to multiple clusters when represented using the inverse of the mutual reachability distance, arranged in ascending order from smaller to larger values. The HDBSCAN technique identifies long-length branches as valid clusters while disregarding clusters with short-length branches within the tree-like hierarchical structure.

2.3 Classification of Each Cluster

After clustering, the proposed method extracts features from each cluster. Using these features, a classification algorithm determines whether each cluster represents a target or clutter. The proposed classification method utilizes a combination of three features to differentiate between target clusters and clutter clusters. The first feature is temporal and spatial distribution characteristics. The second feature is the number of measurements within a cluster. The third feature is ping consistency of measurements within a cluster.

The first feature indicates whether the spatial and temporal distribution of accumulated measurements can be approximated by a polynomial function. The polynomial f approximates time intervals as a function of space i.e. cartesian coordinates of measurement x_n and y_n. Since M accumulated measurements over N pings have a time interval t_n, it is calculated by multiplying the ping number by the sonar pulse repetition interval, the residual r can be obtained using the difference from the polynomial approximation.

$$r = \sum_{n=1}^{M} (t_n - f(x_n, y_n))^2, \tag{2}$$

In this paper we choose the polynomial as $f(x_n, y_n) = ax_n + by_n + c$. But it can be expended in the quadratic form as $f(x_n, y_n) = a{x_n}^2 + b{y_n}^2 + cx_n y_n + dx_n + ey_n + f$. Then, target clusters and clutter clusters can be distinguished by comparing their residuals. The coefficients of the polynomial can be easily obtained by the least square method.

$$\boldsymbol{X} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}\text{ , where }\boldsymbol{A} = \begin{bmatrix} x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots \\ x_M & y_M & 1 \end{bmatrix},\ \boldsymbol{b} = [t_1\ \ t_2\ \ \cdots\ \ t_M]^T \tag{3}$$

The second feature, the number of measurements within a cluster, is influenced by the probability of consistently detecting a submarine over N pings, leading to the accumulation of at least Np_c measurements, where p_c represents the average continuous detection probability for the target over pings. The maximum number of measurements within a target cluster could be as high as $m_t N$, where m_t represents the number of multiple measurements from a target obtained from a single ping. These multiple measurements are caused by multipath detection environments. Clutter clusters caused by high ambient noise often contain an excessive number of measurements, exceeding $m_t N$.

The third feature, the ping consistency between measurements within a cluster, indicates that measurements originating from a target are likely to follow a consistent sequence of ping numbers. In contrast, measurements within clutter clusters may lack a consistent sequence of ping numbers due to variable SNR and a spatially uniform distribution. Although ping consistency can be incorporated into distance calculation for clustering, situations may arise where too many measurements are detected from a single ping or where sequence of ping numbers in the measurements in a cluster is inconsistent.

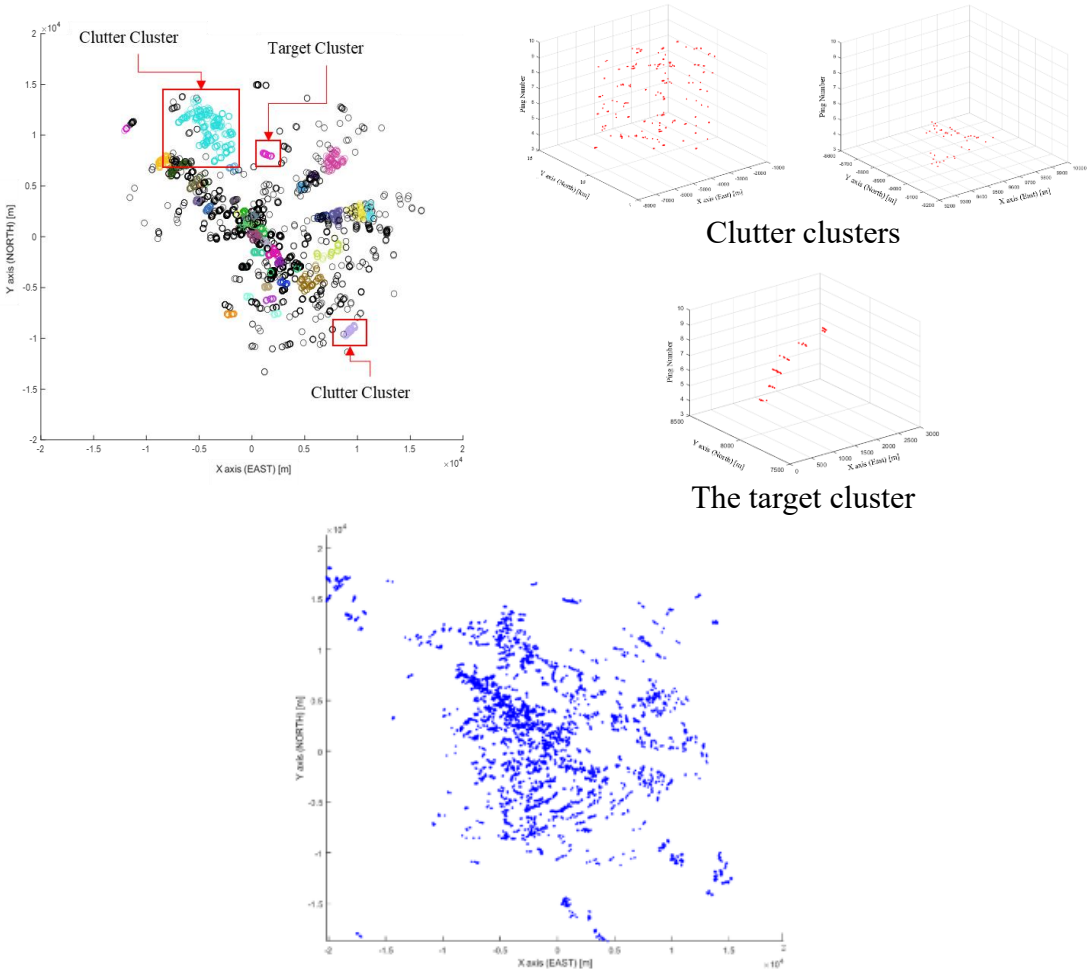


Fig.2: The top left panel shows the clustering result of accumulated measurements (ping #4~9), different color for each cluster. The top right panel shows temporal and spatial distribution of clutter clusters and the target cluster. Red dots are accumulated measurements within a cluster. The bottom panel shows clutter reduction result.

3 RESULTS

The clutter removal performance of the proposed algorithm was evaluated in two stages. First, the performance was analysed using the HDBSCAN clustering method alone. Then, we investigated the cluster classification results after the clustering. By removing non-clustered measurements, it was observed that 36% of clutter was reduced using the clustering method alone. The parameter *minPts* for the HDBSCAN was set to *3*, which is the half of the number of accumulated pings. In the second stage, we classified every cluster, and 85% clutter measurements were reduced. The parameters used to distinguish between target clusters and clutter clusters were as follows: the polynomial approximation residual threshold for temporal and spatial distribution characteristics was *3*. The accumulated number of pings N was *6*, and the number of measurements in an extended target form m_t was set to *10* based on the analysis of the sea experimental data. The average continuous detection probability p_c between pings was set to *0.5*. The ping consistency was set to *2*, meaning the proposed method allows for one missing ping. The results of clustering and cluster classification process are presented in Fig.2.

4 CONCLUSION

A novel method employing density-based clustering combined with cluster-specific classification has been proposed for effective clutter removal in heavy-clutter active sonar detection environments. The proposed approach simplifies clustering application in real systems by requiring only the definition of the minimum number of neighbors. The cluster classification algorithm is generalized and adaptable to various target maneuvers and underwater acoustic conditions. However, optimization of certain algorithm parameters is necessary, considering the specific characteristics of active sonar detection. Furthermore, a comprehensive performance evaluation using diverse sea trial datasets is required, as the current analysis is based on only a single set of sea trial data.

REFERENCES

[1] **R. J. Urick,** *Principles of Underwater Sound*, McGraw-Hill, 3rd edition, 1983.
[2] **J. M. Fialkowski and R. C. Gauss.** Methods for identifying and controlling sonar clutter. *IEEE J. Ocean. Eng*., vol.35, pp.330–354, 2010.
[3] **E. Hanusa, D. Krout, and M. R. Gupta.** Clutter rejection by clustering likelihood-based similarities. In *IEEE Int. Conf. on Information Fusion*, pp.1–6, 2011.
[4] **K. M. Alexiev**. Multiple target tracking using Hough transform PMHT algorithm. In *Proceedings First International IEEE Symposium Intelligent Systems*, pp.227–232, Varna, Bulgaria, 2002.
[5] **S. M. Murphy and P. C. Hines.** Examining the robustness of automated aural classification of active sonar echoes. *J. Acoust. Soc. Am*., vol.135, pp.626–636, 2014.
[6] **R. J. Campello, D. Moulavi, and J. Sander.** Density-based clustering based on hierarchical density estimates. In *Advances in Knowledge Discovery and Data Mining, 17th Pacific-Asia Conference PAKDD, Proceedings Part II*, pp.160–172, 2013.

AUTHORS' INDEX

A

Abbot Phil: 275
Abbot Ted: 275
Abraham Douglas: 325
Abrahamsson Per: 11
Aksoy Serkan: 439
Alhamnday Manar: 459
Antipina Marina: 159
Antoine Samuel: 37
Apte Sudhanshu: 359, 367
Audone Gianluca: 45

B

Barnes Andrew: 83, 135
Bazin Sara: 53
Beerens Peter: 565
Belcher Rhys: 507
Beleza Vaz Guilherme: 143, 175, 575
Belibassakis Kostas: 515
Bennaceur Iannis: 185
Berg Henrik: 231
Berntsen Bjarte: 231
Bezotvetnykh Vladimir: 191
Biber Alper: 439
Said Biskin Osman: 439
Blachet Antoine: 69
Blom Koen: 565
Blondel Philippe: 3, 45, 507
Bodjona Devi: 191
Bolt Laird: 153
Bonnett Blair: 359, 367
Boudraa Abdel: 99
Bragato Filippo: 479
Brander Tommi: 231
Browning John: 275
Budd Obe Chris: 45
Burgschweiger Ralf: 349
Butler William: 83, 135

C

Campagnaro Filippo: 479
Canny Khairina Anindya: 259
Cao Zeng: 525
Carrillo La Rosa Luis Lorenzo: 213
Carter Oscar: 91
Casari Paolo: 471
Chalindar Bruno: 295

Chen Yi: 415, 423, 525
Cheong MyoungJun: 647
Chilvers Elizabeth: 343
Cho Hyeon-Deok: 647
Chu Dezhang: 219
Clark Edward: 239, 547
Cleverly Jonathan: 3
Colin Mathieu: 565
Cooper Dylan: 507
Corakci Ata Can: 439
Correnty Siobhán: 105
Cristol Xavier: 69, 185, 295
Cruz Erica: 143

D

D'Spain Gerald: 555
Donegà Filippo: 479
Dong Bowen: 619
Dong Hefeng: 61
Donnelly Marcus: 547
Dzieciuch Matthew: 3

E

Eccher Davide: 471
Ehrlich Jan: 349
Espinosa Victor: 207, 213
Espinosa Roselló Víctor: 225

F

Fattaccioli Dominique: 69
Fauvel Hugo: 53
Feng Chao: 641
Fickenscher Thomas: 359, 367
Fisichella Nicolas: 295
Fortune Sarah: 159
Friese Jake: 275
Fuchs Louise: 11

G

Gaivota Joao: 593
Giorli Giacomo: 317
Glowacki Oskar: 29
Gomes Tiago: 175, 575
Gorska Natalia: 29, 219
Gourret Yohann: 231
Grech La Rosa Andrea: 585, 593
Groen Johannes: 633
Guo Jing: 111, 395
Guo Xinyi: 533

H
Haslinger Stewart: 167, 603
Hedlund William: 11
Henman Finn: 167
Henson Benjamin: 497
Hjelmervik KarlThomas: 231, 309
Hostombe Kevin: 335
Hunter Alan: 83, 135, 239, 541, 547

I
Illich Eric: 275
Impraimakis Marios: 83, 135
Ivakin Anatoliy: 285
Ivansson Sven: 197
Ivkovic Nathan: 185

J
Jebur Bilal: 459
Jia Guanghui: 415

K
Kaushikk V.N.: 359
Kean Caitlin: 429
Kim Juho: 647
Kim Minho: 611, 625
Kim Myungjun: 625
Kim Youngshin: 611
Kinda Gnouregma Bazile: 99
Kirkham Jay: 159
Kmelnitsky Vitaly: 275
Koetsier Otto: 565
Korenaga Jun: 53
Kovzel Dmitrii: 191
Kraus Dieter: 633
Kummert Anton: 633

L
Labat Valérie: 99
Ladino Velasquez Anderson: 225
Lavayssière Aude: 53
Lehmann Benjamin: 375
Lim Youngsub: 611
Liu Yu: 641
Ljungberg Rydin Ylva: 105
Lloyd Wil: 135
Logan Aaron: 275
Lundberg Nordenvad Magnus:
Lunkov Andrey: 191

M
Ma Li: 111, 395, 533
Ma XiaoChuan: 641
MacGillivray Michael: 153
Madugeta Vidanamesthrige Pavani Vithana: 29
Mahar Scott: 555
Mahiout Thomas: 69
Manzie Chris: 239
Martin S. Bruce: 153
McMorran Solomon: 593
Men Wei: 619
Menetrier Baptiste: 99
Mitchell Paul: 497
Moon Gwangho: 625
Moon Youngsun: 625
Morell-Monzó Sergio: 207, 213
Morozs Nils: 497
Moszyński Marek: 225
Mumtaz Shahid: 459
Murzyniec Michaela: 301

N
Nunes Matthew: 45

O
Oppeneer Victor: 317
Östberg Martin: 105

P
Papili Sonia: 77
Pei Xingyuan: 641
Pena Hector: 231
Peng Yaxi: 259
Peña Hector: 207
Pershin S.M.: 639A
Piggott Giles: 593
Pinson Samuel: 99
Premus Vincent E.: 275
Prospathopoulos John: 515
Puig-Pons Vicente: 207, 213
Pérez-Arjona Isabel: 207, 213, 225

R
Rabeson Herald: 295
Ralph Jason: 603
Ram Matthias: 249
Raumer Pierre-Yves: 53
Real Gaultier: 69
Ren Qunyan: 533

Renner Bernd Christian: 487
Ricardo Souza Marcos: 575
Risser Nolwenn: 317
Rondeau Adrien: 185
Royer Jean-Yves: 53
Russo Aniello: 317
Rygh JonasHalse: 309
Rymansaib Zuhayr: 541

S

Sachau Delf: 335, 349
Safran Romain: 53
Sagen Hanne: 3
Saydam Talha: 439
Schindall Jeffrey: 21
Schmaljohann Holger: 367
Schmidt Beata: 219
Schoof Eric: 239
Schröder Sven: 301
Schäfer Ingo: 349
Seto Mae: 153, 159
Shen Lu: 497
Sherlock Benjamin: 497
Sidorov Danila: 191
Simurda Matej: 267
Skog Isaac: 105
Slater William: 429
Smith Dale: 91
Smith Harrison: 83, 135
Smith Tom: 585, 593
Song Guoli: 533
Song Hyunjae: 611
Steiniger Yannik: 301, 375, 385
Steinmetz Fabian: 487
Storheim Espen: 3

T

Tanrin Jonathan: 53
Taroudakis Michael: 61
Theckes Benoit: 185
Theuerkauf Nils: 633
Tian Yanan: 619
Treloar Alfie Anthony: 541, 547
Tsetas Athanasios: 259
Tsigkinopoulou Areti Ioanna: 267
Tsimenidis Charalampos: 459
Tsouvalas Apostolos: 259
Turgut Altan: 21

V
van de Sande Jeroen: 565
van Riet Martijn: 565
van Toor Astrid: 143
van Vossen Robbert: 317
von Pein Jonas: 249
Vergoz Julien: 53
Vijay Ingale Vaibhav: 53

W
Wakefield Joshua: 167, 603
Wang Xiaomei: 641
Wood Nicolas: 447
Wu Boyue: 423
Wu Peng: 585
Wu Xiaojian: 525
Wüpperling Julia: 487

Y
Yang Liuqing: 415, 423
Yong Seokjeong: 625
Yuan Bo: 533

Z
Zakharov Yuriy: 497
Zeng Juan: 111, 395
Zhang Lu: 619
Zhang Qianchu: 533
Zimmer Christoph: 633
Zorzi Michele: 479
Żytko Aleksander: 219



FORTH

CCBS Greece
Conference tailor made