TOWARDS AUTOMATED VESSEL DETECTION USING SPATIAL SCAN STATISTICS AND DISTRIBUTED ACOUSTIC SENSING

Jack Poole*, Philipp Anhaus, Angel Bueno Rodriguez,
Maurice Stephan, Enno Peters
German Aerospace Centre (DLR), Institute for the Protection of Maritime Infrastructures,
Bremerhaven, Fischkai 1, Germany
*Corresponding author: jack.poole@dlr.de

1 ABSTRACT

Situational awareness is a central task to managing potential risks to maritime infrastructure; however, conventional methods for capturing vessel movements are often susceptible to spoofing or sensitive to weather conditions. This issue motivates the methodology presented in this paper, which investigates distributed acoustic sensing (DAS) as a method to measure acoustics relating to vessels across wide regions, and proposes an anomaly detection method to process data. This paper aims to address two challenges for anomaly detection using DAS data: 1) increasing detection power via leveraging multiple DAS channels and 2) reducing the number of false alarms when monitoring large regions. For this purpose, a spatial scan statistic is implemented, as it presents a method for testing regions with various sizes, and often reduces false alarms related to testing multiple regions independently. Furthermore, it is proposed that model residuals can be used as features for the spatial scan statistic, since predictive models can be designed such that their residuals are insensitive to changes in the data relating to benign events. The proposed system is validated using an open dataset collected off the coast of Valencia. The results suggest that the method is able to consistently detect ships, with each ship crossing event in the test set being detected prior to crossing, and on average alarms are raised when ships are 2km from the cable. Furthermore, the spatial scan statistic is shown to both increase the detection range, while reducing false alarms when compared to testing each region independently.

2 INTRODUCTION

Maritime infrastructure such as pipelines, undersea cables and offshore wind turbines, have become central in the maritime domain. This increasing reliance on maritime infrastructure presents challenges in relation to managing maritime traffic, as well as the risk of sabotage or terrorism. Two prevalent methods for monitoring vessel movements include AIS messages[1], and satellite imaging[2]. However, neither of these methods can reliably capture vessel movements in all scenarios – AIS messages are not a legal requirement for all vessels and can be turned off, and satellite imaging is more challenging in poor visibility conditions. Thus, a more robust monitoring system requires additional sensing capabilities. This paper investigates distributed acoustic sensing (DAS)[3], as a cost-effective method for acoustic sensing across large regions, and presents a spatial anomaly detection method to detect vessels.

Distributed acoustic sensing is a technology that enables a fibre optic cable to be transformed into a dense acoustic sensing array[3]. Thus, acoustic signals from large regions can be measured, which in combination with automated methods for detection and characterisation, could allow for low-cost, comprehensive monitoring systems. Several approaches have been proposed for leveraging DAS data to predict vessel movements. These methods typically follow statistical/machine learning approaches[4;5;6;7;8;4] or focus purely on signal processing approaches[9;10;11;12]. It should also be noted that many alternative passive sonar systems have been used to successfully demonstrate automated vessel detection in acoustic data[13;14;15;16], using point sensors or arrays of point sensors. While the wider passive acoustics literature presents many tools that could be used to process DAS data, the spatial coverage of DAS presents novel challenges and opportunities.

The large coverage of DAS typically leads to many channels, which presents challenges relating to the optimal partitioning of the space to select subsets of channels, and developing modelling approaches to capture spatial dependencies. For example, one approach would be to jointly model the entire system of sensors; however, this would present a challenging modelling problem, and may lose sensitivity to local anomalies, as they become averaged out in a large feature pool. In addition, this approach would provide no information about the location of an anomaly. Another approach would be to partition the space into predefined regions, and to treat each partition as an independent prediction problem, which is the main approach taken in previous studies[5;6;7;8;4]. However, this approach has two main issues. First, a fixed partition must be defined; if the region is too small, there is a risk a feature set will not contain all sensors affected by a signal, reducing detection power, and if the region is much larger than the affected area, events may be missed because most channels will represent a normal signal. Second, if there are many partitions, false positive alarms will compound, adding to the manual effort required to investigate alarms.

These challenges motivates the investigation of spatial scan statistics in this paper, which present a statistical test for spatial anomalies of varying sizes[17;18]. By testing over different window sizes, the scan statistic circumvents the issues related to a fixed partition, increasing detection power in scenarios where anomalies may effect a changing number of sensor channels[18;19]. Furthermore, spatial scan statistics leverage the assumption that observations were drawn from a potentially large space, reducing the occurrence of false alarms caused by multiple testing[17;18]. Scan statistics present a promising method to leverage information across many DAS channels; as far as the authors are aware, they have not been previously investigated in the context of DAS. In addition, they have only been applied in a limited setting for object detection in the maritime domain[20], where a numerical dataset was used to investigate the scan statistic as a method for post-processing of signal-processing based binary detection labels using passive acoustic sensors.

To facilitate the application of scan statistics to real-time DAS data, this paper develops a framework to efficiently process many data channels and account for spurious fluctuations in the data. Specifically, a scan statistic is applied to a model residual instead of the raw signal. This approach follows ideas relating to modelling covariates[18;21;22;23], which is an important consideration for data where confounding influences will lead to significant variation; in DAS data there are a wide range of spurious effects that could influence the signal. In this paper, unsupervised modelling approaches (autoencoders[24]) are estimated using baseline data including no known vessels. This approach results in features with reduced sensitivity to patterns captured in the baseline dataset, while maintaining sensitivity to activity related to vessels. The resulting residual-based scan statistic is tested for the task of vessel detection using an open dataset collected near Valencia[25]. It is demonstrated to successfully identify all ten ship crossings in the investigated test set, and in most cases the first alarm was raised when the ship was still several kilometres from the cable.

The structure of this paper is as follows. Section 2 present the necessary background relating to scan statistics and introduces the idea of the use of model residuals as a feature. Section 3 then presents the case study methodology and results using an open dataset collected near Valencia. Finally, Section 4 presents conclusions and future work.

3 SPATIAL SCAN STATISTICS FOR ANOMALY DETECTION USING DISTRIBUTED ACOUSTIC SENSING

The spatial scan statistic was first proposed by Kuldorf to monitor disease outbreaks using counts of patients from hospitals[17]. Since its original implementation, there have been many variations of the

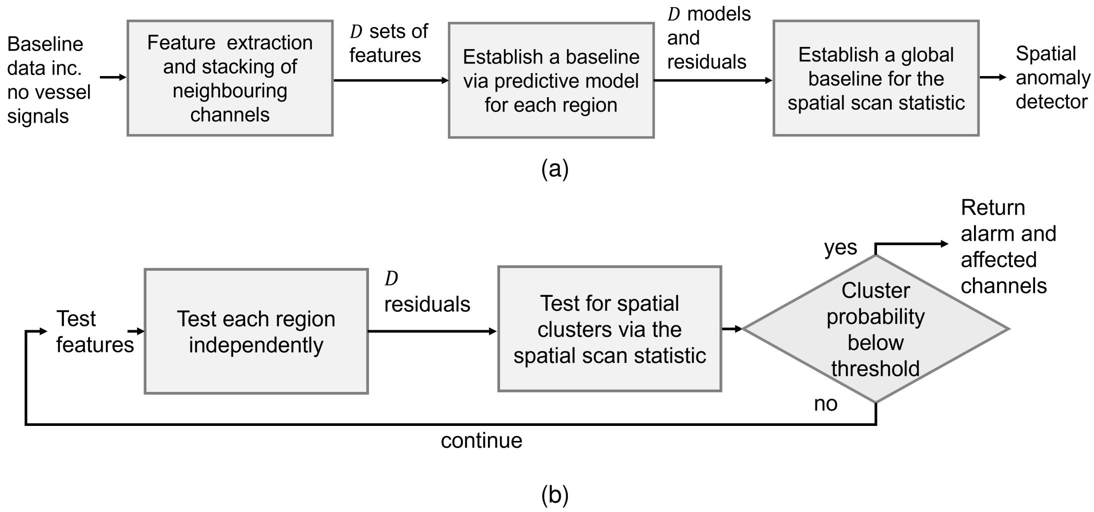


Figure 1: Outline of the proposed anomaly detection approach using DAS, where (a) shows the training process, while (b) presents the testing procedure.

scan statistic proposed[26], and it has found applications in disease/bio terrorism monitoring[18;19], crime surveillance[27;28], vehicle traffic analysis[29], target detection, structural damage localisation[30],[31;20], and environmental monitoring[32]. To apply scan statistics for real-time monitoring of DAS data, methods for accounting for benign variations and leveraging complex time series with high sampling rates are required. To this end, this section introduces a methodology for leveraging model residuals as a feature for the scan statistic. The general methodology is presented in Figure 1, with Figure 1(a) illustrating the general training methodology, and Figure 1(b) the testing process. First, the necessary background relating to the scan statistic is presented in the following section, and Section 2.2 discusses using model residuals as features.

3.1 Spatial scan statistics

The spatial scan statistic aims to monitor multiple spatially-related time series $X = \{\mathbf{x}_1, ..., \mathbf{x}_C\}$, to detect emerging spatial clusters of anomalous values[17]. Cluster detection is achieved by testing each set of indices S from a the set of index windows $S \in \mathcal{S}$. Subset testing of each region S is achieved by comparing a null hypothesis H_0, typically estimated using historical data, with an alternative hypothesis H_1. This paper follows an expectation-based approach[18], which calculates the alternative hypothesis H_1 via a maximum likelihood approach applied to a subset; the likelihood ratio is used as a test statistic and is calculated as follows,

$$F(S) = \frac{\max_{\theta_1(S)\in\Theta_1(S)} \prod_{i\in S} P(\mathbf{x}_i \mid H_1, \theta_1(S))}{\max_{\theta_0\in\Theta_0} \prod_{i\in S} P(\mathbf{x}_i \mid H_0, \theta_0)} \tag{1}$$

where $P(\mathbf{x}_i \mid H_0, \theta_0)$ represents the likelihood function, and θ_0 and $\theta_1(S)$ represent the estimated parameters for the null and alternative hypothesis respectively. The likelihood function can either be defined via empirical distributions, commonly Poisson, Bernoulli or Gaussian, or nonparametric approaches[26]. Each candidate region is tested and the cluster with the largest value is selected as follows,

$$S^* = \arg\max_{S\in\mathcal{S}} F(S) \tag{2}$$

where S^* represents the subset with the largest likelihood ratio from all subsets. After evaluating all regions in $\mathcal{S}$, anomalous regions are reported based on the probability associated with S^*; this paper implements Monte Carlo testing to obtain probabilities (p-values). By testing all candidate regions and selecting the most likely anomalous cluster, the spatial scan statistic adjusts for the multiple testing inherent to large spatial domains; thus, reducing false positive alarms [18].

Typically, the set of candidate locations $\mathcal{S}$, are chosen according to a *search window*, such as a circle around centroid locations, or by dividing the space into a grid. This paper aims to evaluate whether the spatial scan statistic can be used to improve detection by leveraging information across neighbouring DAS channels; in the investigated case the cable follows a straight line. As such, only one spatial dimension is considered; thus, the window used represented an interval of channels along the cable $\mathcal{S} = \{\{i, ..., j\} : 1 \leq i \leq j \leq C\}$. The spatial scan statistic could also be extended to consider measurements from more complex cable geometries; for example, by defining scan windows on graphs [26]. It can also be used to test over multiple time instances (the spatio-temporal scan statistic) [26]. The spatio-temporal scan statistic will be considering in future work.

Since this paper aims to monitor continuous model residuals, a Gaussian likelihood is implemented. One advantage of expectation-based scan statistics is that, when using an exponential likelihood function, computation time can by reduced significantly be pre-computing statistics [19]; this approach was used in the implementation in this paper.

For a more detailed explanation of the expectation-based spatial scan statistic, the interested reader may refer to [18].

3.2 Spatial scan statistics using model residuals

The aim of the spatial scan statistic is to monitor C time series simultaneously to identify spatially localised anomalies. The motivation of applying the scan statistic to DAS is to simultaneously monitor C time series relating a large number of channels. However, DAS time series will likely present a high-dimensional feature for hypothesis testing, representing multiple channels and potentially many frequencies, when frequency-based features are used. Furthermore, in many cases DAS measurements are affected by a range of confounding factors, including weather, cable degradation, and wildlife activity, which introduces non-stationary behaviour; thus, requiring more complex generative models for the null and alternative hypothesis to prevent false alarms [21;23].

One approach would be to extract a set of features relating to each channel (or a set of channels), which are only sensitive to changes in the signal caused by objects of interest. This paper proposes using the residual of a predictive model as a potential feature, which would allow complex phenomena to be implicitly modelled. This approach is similar to various approaches proposed for modelling covariates or functional data [18;21;22;23]; however, it allows for a more flexible specification of the mean function while also using computationally efficient scan statistics.

In this paper, multiple independent autoencoders were trained using DAS baseline data representing multiple regions along the cable, where data relating to ships within a certain distance threshold were removed using AIS data; multiple independent models were trained to account for variation in measurements across the cable. Autoencoders present a powerful approach for anomaly detection in high-dimensional data [24], and highlight deviations from the baseline via increased reconstruction error. By using reconstructive models, relationships implicit in the baseline data can be captured, i.e. given data from extreme weather conditions, a model should be able to accurately reconstruct the data, while phenomenon not included in the baseline dataset will be poorly represented, such as vessels. Future work will consider using different predictive models to improve robustness under changing conditions, including directly incorporating exogenous features, such as temperature and wind speed.

For more details on autoencoders the interested reader may refer to [24].

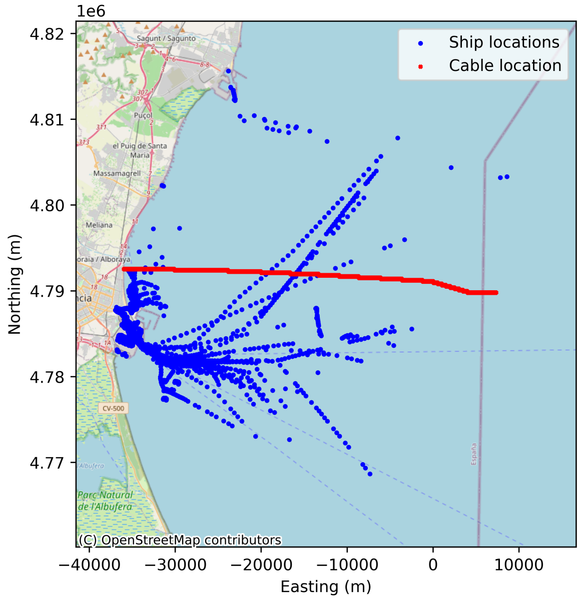


Figure 2: Example of AIS messages for 06/09/2020.

4 CASE STUDY: VALENCIA DAS DATASET

To test the potential of the residual-based scan statistic to detect vessels in a maritime environment, the dataset during the Valencia-Isalink experiment was used[25]. The following sections provide a brief description of the dataset and introduce the methodology for establishing a baseline and testing; results for the detection of vessels crossing and approaching the cable are then presented in Section 3.3.

4.1 Dataset description

The dataset was collected using a Febus Optics A1-R interrogator installed on a standard telecommunications cable connecting the Spanish peninsula and Mallorca[25]. Data was measured in the first 50km of the cable, over a period of seven days (01/09/2020-07/09/2020). Data were measured with a sampling frequency of 1000Hz, downsampled to 250Hz, using a gauge length of 30m, and a channel resolution of 16.8m.

To obtain labels, AIS data was leveraged across the same time period. An example day of vessel traffic (06/09/2020) is visualised in Figure 2. A geometry file is provided along with the DAS data, providing coordinates relating to underwater channels; this geometry is also visualised as a red line in Figure 2. It can be seen a large proportion of traffic in the region either begins or terminates its journeys in the port of Valencia, with some traffic also coming from Sagunto, and a moderate number of vessels took a route leading them to approach and/or cross the cable.

To generate labels from AIS signals at the same time scale as the DAS features, linear interpolation between two AIS messages was performed. Future work could consider more advanced methods for interpolation, such as autoregressive modelling[33]; however, linear interpolation should provide a suitable estimation for this application, assuming ships maintain a constant speed and route. To prevent incorrect labels from poor interpolation, vessel trajectories were filtered to remove cases where messages received either side of the cable were more than 15 minutes apart.

Two types of labels were used to validate the method: the minimum distance between a ship and a given DAS channel, and a binary label indicating a ship crossing the cable. To relate sensor channels with a their physical location, the geometry file provided with the dataset was used. Thus, distance to each channel could be computed using haversine distance, and ship crossing were found by looking at whether segments in the ship trajectory and cable geometry overlapped, as in[34].

The first 9,189m of the cable were on land[25].In addition, large numbers of channels in the first section of the cable underwater contained consistent high levels of noise. As most vessel crossings in this time period occurred further along the cable, the channels from the first 18.2km were not used. In addition, after initial analysis of the data, it was found that regions of high amplitude signal relating to ship crossings, as well as the region of high noise that can often be observed when a DAS cable transitions from land into the sea, were offset by a constant bias. Based on these observations, the geometry file was retrospectively corrected by shifting the the original channel indices by 58.

4.2 Feature extraction and training methodology

Data were split into a training, validation and testing set with a proportion of 60:10:30. To establish a baseline both for the autoencoders and the spatial scan statistic (the null hypothesis), AIS data was used to filter training data to only include instances where no ship was within a 5km radius of a given channel. This approach was used rather than directly using AIS data in training as the number of ship events are small, and data are imbalanced compared to the quantity of baseline data.

Features were obtained via Welches method, with a thirty-second time window, using a fifteen-second overlap – resulting in a system that evaluates and potentially raises an alarm every fifteen seconds. Frequency features in the range of 5 to 60Hz were selected to focus on a frequency range that is commonly related to ships[15;35]. Several channels were concatenated to jointly model multiple channels, allowing cross-channel relationships to be leveraged for data compression. In this paper, the FFT features from 10 channels are concatenated, resulting in a feature dimension of 363; this number was chosen to minimise correlation between the centre points of each set of channels.

There are a number of hyperparameters that must be selected both for the training of the autoencoders and to define the spatial scan statistic test. For the autoencoders, several trials were performed to select an architecture that produced low generalisation error to the validation set; the encoder and decoder both used two hidden layers with 100 and 75 nodes for the outer and inner layers respectively, and the latent space used 50 nodes. To establish a suitable length for time windows and the corresponding number of averages to use within Welches method cross-validation was performed using the validation set, which included ship crossings from the first five days of the dataset; windows including 150 samples were selected. For the spatial scan statistic, the search space must be defined. To limit computation, this paper limits the maximum locations to fifty. Standardisation was used as preprocessing for both the autoencoders, and residuals for the spatial scan statistics. As the residuals are skewed and prone to extreme values, a power transformation was applied to the data (the Yeo-Johnson transformation[24]), such that it better fits the Gaussian distribution assumption of the spatial scan statistic.

Method	True positive rate	Alarm rate	Average distance at first detection (m)	Average distance at last detection (m)
Independent detectors	1.00	0.15	1329	2355
Spatial scan	1.00	0.09	1879	2847

Table 1: The average performance of the independent anomaly detector results aggregated vs the residual-based spatial scan statistic, including proportion of ships detected from ten ships (the true positive rate for binary crossing indicator labels), the alarm rate - representing the proportion of 15-second time windows associated with an alarm - and the average distances of the vessels at first and last detection.

To test the effectiveness of post-processing residuals via the spatial scan statistic, a comparison is presented using autoencoders as independent anomaly detectors, indicating alarms by thresholding the reconstruction error. This method provides independent alarms relating to sensor data in regions of the cable, representing a similar approach to previous literature. To provide a single alarm for a given time instance, alarms are aggregated by setting the label as true if any label is positive. The probability threshold for raising an alarm was defined as the 99.9th percentile for both methods.

4.3 Results

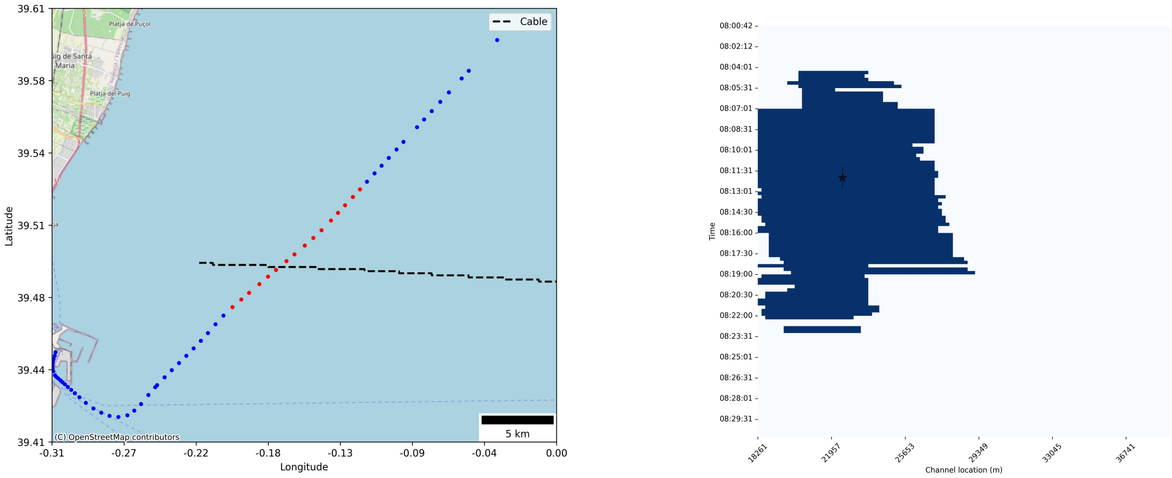
Table 1 summarises the performance of the residual-based scan statistic for raising alarms related ship crossing events specifically, where "independent detectors" refers to the independent anomaly detector approach. In the test period, ten ships crossed the cable, of which all raised alarms associated with the estimated crossing time, as shown by the "true positive rate" in Table 1.

On average the spatial scan statistic first raised alarms when ships were 1879m from the cable, and continued indicating the presence of an anomaly until vessels had moved 2847m from the cable. This result demonstrates the potential for detecting vessels via DAS. The cause of the asymmetry between the average first and last detection distance requires further investigation; it may be caused by changing conditions along the cable or an imbalance in vessel behaviours in the test set. However, since the test dataset only includes ten crossings, strong conclusions cannot be drawn.

Using, the spatial scan statistic provided an improvement in the detection range from the cable of around 500m, demonstrating the benefit of combining information from detectors across multiple channels. In addition, the spatial scan statistic also led to a lower overall proportion of time instances where an alarm was raised, shown by the alarm rate in Table 1. This result is likely because it is more robust to spurious spikes in single or small sets of channels. Although, it would clearly still be beneficial to further reduce this alarm rate to prevent wasted personnel time in investigating false alarms.

An example of a single ship trajectory during the test period is shown in Figure 3(a), where each point represents the location of an AIS message, and blue and red points indicate no alarm, or alarm raised, respectively. Meanwhile, Figure 3(b) provides a shows the time and location of anomalous windows with a dark blue line, and the ship crossing time and location is indicated by the star. It can be seen in this case an alarm is raised when a ship was at a distance of 3074m from the cable, and alarms were continually raised while it was near the cable. This result suggests incorporating information from multiple time windows may help reduce false positives. For example, a basic approach would be to only raise an alarm in the case N consecutive alarms were raise. A more advanced approach would be to leverage a scan statistic with flexible scan windows to leverage anomalous signals across multiple channels and time instances.

It can be seen in this case, while the ship crossing is within the centre of the window, a wide range of channels are flagged as anomalous in Figure 3(b), suggesting a potentially limited utility of the



(a) (b)

Figure 3: Example of a single case of vessel detection. The path reported via AIS is indicated by the red and blue dots in (a), indicating if an alarm was raised, or not raised, at the time of the AIS message, respectively. The section of the DAS cable used for anomaly detection is indicated by the black line. Panel (b) shows the time and location of the binary prediction labels and the associated spatial window, shown by the dark blue lines, and the estimated crossing time and location is shown by the star marker.

Distance	Independent detectors				Scan statistic			
threshold (m)	TP	FP	FN	TN	TP	FP	FN	TN
500	113	1212	14	7785	108	710	19	8287
1000	196	1129	59	7740	188	630	67	8239
5000	501	824	1161	6638	471	347	1191	7115

Table 2: Number of true positives (TP), false positives (FP), false negatives (FN), and true positives (TN), when considering all time instances as positive label where ship distances are below the threshold.

method for source localisation, even though it could still present a useful tool for quantifying the extent acoustic waves can be measured across a cable. This result may suggest source localisation could be challenging with unsupervised methods, as the signal extends across a large region. Furthermore, the movement of the ship can be seen from the location of clusters increasing along the cable between 8:14 and 8:19, which corresponds with the trajectory shown in Figure 3(a). However, at the end of the detection period, the detection window constricts, and no longer corresponds to the location of the ship. There are several reasons that may have caused this result: the cable depth is increasing along the cable, so perhaps only the channels at a shallower depths measured a signal above the noise floor, or the wave arrival direction may have led to larger axial strain in these locations[36].

The presented method it not necessarily only sensitive to vessel crossings, but any deviation from the baseline – including ships approaching the cable. Table 2 presents the counts of the true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN), where a binary label is considered true, if a ship is within the distance threshold. It can be seen that alarms are raised for the large majority thirty-second windows when a ship is within 500m of a channel, while detection rate drops to around 30% when the distance threshold is 5000m. Furthermore, using the scan statistic led to a slight reduction in true positive predictions; however, it has also significantly reduced the number of false positive alarms at each distance threshold.

Proportion of alarms within the window	Average window width (m)	Average distance from window centre to ship (m)
1.00	5580	601

Table 3: Summary of the correspondence between anomalous regions and ship position during crossing events.

Another aspect of the scan statistic is it returns a set of indices for channels with high deviations in the signal. Assuming an anomalous signal originates from a vessel, these indices should correspond to channels where the acoustic signal has not yet attenuated close to or below the level of ambient noise. While this output does not directly correspond to the true location of a source, it may be reasonable to assume the source is within the window, which was the case for all crossing trajectories in the test set, as shown in Table 3.

It can also be seen with the current implementation of the spatial scan statistic, the average window size at the time of crossing was large. Considering the detection range indicated in Table 1, it is feasible this represents the region where channel measurements are significantly affected. This results further suggests the method could provide a useful scientific tool for assessing the sensitivity of DAS measurements. Furthermore, it could be informative for downstream analysis to predict source location; for example, by selecting channels to be used in more computationally intensive methods, or providing additional information for predictive models. It is important to note, spatial scan statistics can introduce a bias towards larger windows[37]; this phenomenon should be further investigated and methods for penalising the spatial scan statistic could be implemented. It is also worth noting that in many cases a more precise indication of the crossing location might be indicated by signal intensity, or autoencoder residual values.

A naive approach to obtain a more precise location would be to assume a source is at the centre of a window; the average distance from the centre is also shown in Table 3. It can be seen that while this assumption does result in a general indication of vessel location, it still presents a large possible region. Furthermore, how close the vessel is to the centre of the window will be dependent on several factors, such as the approach angle and environment. It should also be again emphasised that these results rely an ad-hoc recalibration of the cable location (see Section 3.2); the localisation capabilities require validation on an additional dataset with accurate calibration.

5 DISCUSSION AND CONCLUSIONS

This paper has presented an initial anomaly detection approach for DAS data using model residuals and a spatial scan statistic. This approach was shown to consistently detect vessel crossings with an average detection range of approximately 2-3km. In addition, aggregating information across channels via the spatial scan statistic provided a reduction in false positive predictions, as well as an increased detection range of approximately 500m in comparison to independent anomaly detectors. Furthermore, the spatial anomalies indicate channels with abnormal signals, providing an initial indication of source location and a method to investigate to what extent signals can be measured across a DAS cable.

While this paper provides a positive indication that vessel detection could be reliably achieved with DAS and cheap-to-train anomaly detection systems, there are several limitations of the presented study. First, the methods presented should be further validated using larger DAS datasets, which include measurements of multiple ship crossings and well-calibrated cable locations. Part of this analysis should include the effects of various vessel characteristics, weather events, cable types, and changing water depths and bathymetries. Furthermore, one of the core advantages of anomaly detection is that any object not included in the baseline may be detectable; as such, the methodology should be tested for detecting a variety of objects.

While several studies have demonstrated successful detection on regions (or patches) of DAS data, this paper highlights the challenge of high false positive rates caused by multiple testing when monitoring large regions via DAS. Spatial scan statistics are demonstrated to be a potential solution to reduce false alarms; however, the reduction of benign alarms while maintaining predictive power should be a major focus of future work. The first way alarms could be reduced in the presented framework would be to refine the predictive modelling stage; for example, exogenous variables (temperature, wind speed etc.) could be incorporated; larger, more varied baseline datasets could be used; and methods for modelling temporal trends could be incorporated. In the post-processing of model residuals, there are several potential improvements to reduce false positives. One approach could be to improve scan window selection to perform tests on moving objects over multiple time instance; two options would be to use kernel-spatial scan methods[28], or subset scans[19;37]. Another potential direction of research would be to investigate methods for establishing more representative null hypotheses.

6 ACKNOWLEDGEMENTS

The scientific work presented here was partly funded by DLR and by the German Federal Ministry for Economic Affairs and Energy in the context of the AFKUI project. The responsibility for the content of this publication lies with the author. The authors would like to acknowledge and thank the authors of The Valencia-Islalink Distributed Acoustic Sensing Experiment[25].

REFERENCES

[1] Thomas Stach, Yann Kinkel, Manfred Constapel, and Hans-Christoph Burmeister. Maritime anomaly detection for vessel traffic services: A survey. *Journal of Marine Science and Engineering*, 11(6):1174, 2023.

[2] Giovanni Soldi, Domenico Gaglione, Nicola Forti, Alessio Di Simone, Filippo Cristian Daffinà, Gianfausto Bottini, Dino Quattrociocchi, Leonardo M Millefiori, Paolo Braca, Sandro Carniel, et al. Space-based global maritime surveillance. part i: Satellite technologies. *IEEE Aerospace and Electronic Systems Magazine*, 36(9):8–28, 2021.

[3] Arthur H Hartog. *An introduction to distributed optical fibre sensors*. CRC press, 2017.

[4] Erick Eduardo Ramirez-Torres, Javier Macias-Guarasa, Daniel Pizarro-Perez, Javier Tejedor, Sira Elena Palazuelos-Cagigas, Pedro J Vidal-Moreno, Sonia Martin-Lopez, Miguel Gonzalez-Herraez, and Roel Vanthillo. Vessel detection and localization using distributed acoustic sensing in submarine optical fiber cables. *arXiv preprint arXiv:2509.11614*, 2025.

[5] Jiajia Huang, Wei Liu, Vuong Nhu Khue, Hailiang Zhang, Hui Dong, Yixin Wang, Meng Fai Yue, Kum Yeow Lum, Gary Jia Rui Chew, Bien Aik Tan, et al. Ai-driven automatic ship detection using intensity-based das on submarine cables. In *2025 IEEE Photonics Conference (IPC)*, pages 1–2. IEEE, 2025.

[6] Wenjin Huang, Shaoyi Chen, Yichang Wu, Ruihua Li, Tianrui Li, Yihua Huang, Xiaochun Cao, and Zhaohui Li. Daship: A large-scale annotated dataset for ship detection using distributed acoustic sensing technique. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:4093–4107, 2025.

[7] Chun Shan, Yankai Lu, and Yong Zeng. A distributed acoustic sensing technology method for ship feature extraction based on multitask learning. In *Second Distributed Optical Fiber Sensing*

Technology and Applications Conference (DOFS 2025), volume 14112, pages 404–417. SPIE, 2026.

[8] Hasse B Pedersen, Peder Heiselberg, Henning Heiselberg, Arnhold Simonsen, and Kristian Aalling Sørensen. A feasibility study of automated detection and classification of signals in distributed acoustic sensing. *Sensors*, 25(17):5445, 2025.

[9] Diane Rivet, Benoit de Cacqueray, Anthony Sladen, Aurélien Roques, and Gaëtan Calbris. Preliminary assessment of ship detection and trajectory evaluation using distributed acoustic sensing on an optical fiber telecom cable. *The Journal of the Acoustical Society of America*, 149(4):2615–2627, 2021.

[10] Lukas Thiem, Susann Wienecke, Kittinat Taweesintananon, Melvin Vaupel, and Martin Landrø. Ship noise characterization for marine traffic monitoring using distributed acoustic sensing. In *2023 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea)*, pages 334–339. IEEE, 2023.

[11] Bob Paap, Vincent Vandeweijer, Jan-Diederik van Wees, and Dirk Kraaijpoel. Leveraging distributed acoustic sensing for monitoring vessels using submarine fiber-optic cables. *Applied Ocean Research*, 154:104422, 2025.

[12] Jie Shao, Yibo Wang, Yixin Zhang, Xuping Zhang, and Chi Zhang. Tracking moving ships using distributed acoustic sensing data. *IEEE Geoscience and Remote Sensing Letters*, 22:1–5, 2025.

[13] Andrzej Zak. Ships classification basing on acoustic signatures. *WSEAS Transactions on Signal Processing*, 4(4):137–149, 2008.

[14] Michael Bruno, Kil Woo Chung, Hady Salloum, Alexander Sedunov, Nikolay Sedunov, Alexander Sutin, Hans Graber, and Paul Mallas. Concurrent use of satellite imaging and passive acoustics for maritime domain awareness. In *2010 international waterside security conference*, pages 1–8. IEEE, 2010.

[15] Kil Woo Chung, Alexander Sutin, Alexander Sedunov, and Michael Bruno. Demon acoustic ship signature measurements in an urban harbor. *Advances in Acoustics and Vibration*, 2011(1):952798, 2011.

[16] Chenyang Zhu, Heriberto Garcia, Anna Kaplan, Matthew Schinault, Nils Olav Handegard, Olav Rune Godø, Wei Huang, and Purnima Ratilal. Detection, localization and classification of multiple mechanized ocean vessels over continental-shelf scale regions with passive ocean acoustic waveguide remote sensing. *Remote Sensing*, 10(11):1699, 2018.

[17] Martin Kulldorff. A spatial scan statistic. *Communications in Statistics-Theory and methods*, 26(6):1481–1496, 1997.

[18] Daniel B Neill. Expectation-based scan statistics for monitoring spatial time series data. *International Journal of Forecasting*, 25(3):498–517, 2009.

[19] Daniel B Neill. Fast subset scan for spatial pattern detection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 74(2):337–360, 2012.

[20] Xiufeng Song, Peter Willett, Joseph Glaz, and Shengli Zhou. Active detection with a barrier sensor network using a scan statistic. *IEEE Journal of Oceanic Engineering*, 37(1):66–74, 2011.

[21] Tonglin Zhang and Ge Lin. Spatial scan statistics in loglinear models. *Computational Statistics & Data Analysis*, 53(8):2851–2858, 2009.

[22] Ying Liu, Yawen Liu, and Tonglin Zhang. Wald-based spatial scan statistics for cluster detection. *Computational Statistics & Data Analysis*, 127:298–310, 2018.

[23] Takayuki Kawashima, Daisuke Yoneoka, Yuta Tanoue, Akifumi Eguchi, and Shuhei Nomura. New insight of spatial scan statistics via regression model. *arXiv preprint arXiv:2502.06347*, 2025.

[24] Kevin P Murphy. *Probabilistic machine learning: an introduction*. MIT press, 2022.

[25] Zack Spica, Beatriz Gaite, and Sandra Ruiz Barajas. The valencia-islalink distributed acoustic sensing experiment. *International Federation of Digital Seismograph Networks*, 2020.

[26] Ali Abolhassani and Marcos O Prates. An up-to-date review of scan statistics. *Statistic Surveys*, 15:111–153, 2021.

[27] Wilpen Gorr and Richard Harries. Introduction to crime forecasting. *International journal of forecasting*, 19(4):551–555, 2003.

[28] Mingxuan Han, Michael Matheny, and Jeff M Phillips. The kernel spatial scan statistic. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 349–358, 2019.

[29] Qiliang Liu, Sancheng Zhu, Min Deng, Wenkai Liu, and Zhihui Wu. A spatial scan statistic to detect spatial communities of vehicle movements on urban road networks. *Geographical Analysis*, 54(1):124–148, 2022.

[30] James Hensman, Keith Worden, Mark Eaton, Rhys Pullin, Karen Holford, and Sam Evans. Spatial scanning for anomaly detection in acoustic emission testing of an aerospace structure. *Mechanical Systems and Signal Processing*, 25(7):2462–2474, 2011.

[31] Marco Guerriero, Peter Willett, and Joseph Glaz. Distributed target detection in sensor networks using scan statistics. *IEEE Transactions on Signal Processing*, 57(7):2629–2639, 2009.

[32] Anastassia Ailamaki, Christos Faloutos, Paul S Fischbeck, Mitchell J Small, and Jeanne VanBriesen. An environmental sensor network to determine drinking water quality and security. *ACM Sigmod Record*, 32(4):47–52, 2003.

[33] Yanxiao Jiang, Ying Chang, Benkui Zhang, Yunping Ge, and Xinyu Zhao. A review of vessel trajectory prediction research: Methodological evolution, current challenges, and future directions. *International Journal of Naval Architecture and Ocean Engineering*, page 100754, 2026.

[34] Angel Bueno Rodriguez, Felix Sattler, Tino Flenker, Enno Peters, Sarah Barnes, and Maurice Stephan. Dasboot: an annotation toolkit for das-based maritime surveillance. In *Artificial Intelligence for Security and Defence Applications II*, volume 13206, pages 78–85. SPIE, 2024.

[35] Clausius DG Reis, Linilson Rodrigues Padovese, and Maria CF De Oliveira. Automatic detection of vessel signatures in audio recordings with spectral amplitude variation signature. *Methods in Ecology and Evolution*, 10(9):1501–1516, 2019.

[36] Brian LN Kennett. The seismic wavefield as seen by distributed acoustic sensing arrays: local, regional and teleseismic sources. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 478(2258), 2022.

[37] Skyler Speakman, Sriram Somanchi, Edward McFowland III, and Daniel B Neill. Penalized fast subset scanning. *Journal of Computational and Graphical Statistics*, 25(2):382–404, 2016.