IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-2010-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IGARSS52108.2023.10281607

DDM-FORMER: GLOBAL OCEAN WIND SPEED RETRIEVAL WITH TRANSFORMER NETWORKS

Daixin Zhao [1,2], *Konrad Heidler* [1], *Milad Asgarimehr* [3,4], *Caroline Arnold* [5], *Tianqi Xiao* [3,4], *Jens Wickert* [3,4], *Xiao Xiang Zhu* [1], *Lichao Mou* [2,1]

[1] Technische Universität München (TUM), Germany
[2] German Aerospace Center (DLR), Germany
[3] German Research Centre for Geosciences (GFZ), Germany
[4] Technische Universität Berlin (TU Berlin), Germany
[5] German Climate Computing Center (DKRZ), Germany

ABSTRACT

As a novel remote sensing technique, GNSS reflectometry (GNSS-R) opens a new era of retrieving Earth surface parameters. Several studies employ the combination of deep learning and GNSS-R observable delay-Doppler maps (DDMs) to generate ocean wind speed estimation. Unlike these methods that often use convolutional neural networks (CNNs) with inductive bias, we proposed a Transformer-based model, named DDM-Former, to exploit fine-grained delay-Doppler correlation independently. Our model is evaluated on the Cyclone GNSS (CYGNSS) version 3.0 dataset and shown to outperform the other retrieval methods.

***Index Terms*—** Cyclone GNSS, deep learning, GNSS reflectometry, ocean wind speed, Transformer network

1. INTRODUCTION

By exploiting scattered signals of opportunity from GNSS, spaceborne GNSS reflectometry (GNSS-R) has emerged as a novel remote sensing technique to retrieve various Earth surface parameters, e.g., ocean wind speed [2] and soil moisture [3]. With cost-efficient nanosatellites, abundant measurements are made available with significant improvements in both spatial and temporal resolutions. Cyclone GNSS (CYGNSS), launched in December 2016 as the first small satellite constellation fully dedicated to GNSS-R, can provide a spatial resolution of approximately $0.5\,\mathrm{km}$ to $25\,\mathrm{km}$ for smooth to rough surfaces with a mean revisit time of $7\,\mathrm{h}$ [4].

One of the fundamental observables of GNSS-R is the delay-Doppler map (DDM), which refers to a two-dimensional map of scattering power at a range of signal delays and Doppler frequency shifts. As ocean surface roughness is altering due to ocean winds, the distribution of ocean slopes is described by power changes in delay and Doppler dimensions

An extended version of this conference report is available at [1].

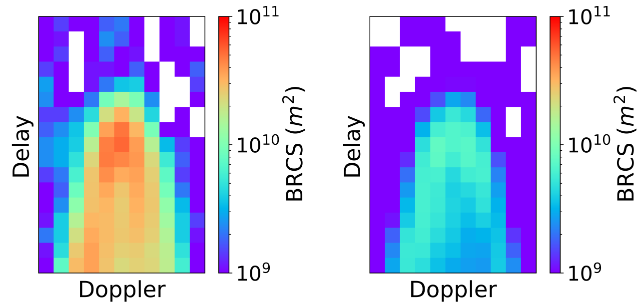

Fig. 1. CYGNSS Level-1 bistatic radar cross section (BRCS) DDMs at low (left) and high (right) wind speeds.

[5]. DDMs that carry fine-grained delay-Doppler correlation can thus be used to estimate ocean wind speeds. Examples of CYGNSS Level-1 DDMs are shown in Fig. 1.

Many previous works have proved the feasibility of estimating ocean wind speeds using DDMs [6, 7]. As deep learning is taking off in remote sensing [8] and an increasing amount of GNSS-R data is collected in recent years, several studies show that data-driven approaches could offer an alternative way to address potential limitations of conventional retrieval methods and enhance ocean wind speed estimations. Early works utilize multilayer perceptron (MLP) to incorporate GNSS-R parametric data and improve estimation performance [9, 10]. More recently, convolutional neural networks (CNNs) have demonstrated their strong capability to aggregate critical information from DDMs and facilitate ocean wind speed retrieval [11, 12].

Considering the fact that DDMs are diametrically distinctive from natural images, the inbuilt inductive bias of CNNs can be misleading in perceiving these observables. We thus turn to a more recent network with a weaker inductive bias named Transformer [13]. Transformers are widely employed in many natural language processing and computer vision tasks, and have achieved state-of-the-art performance

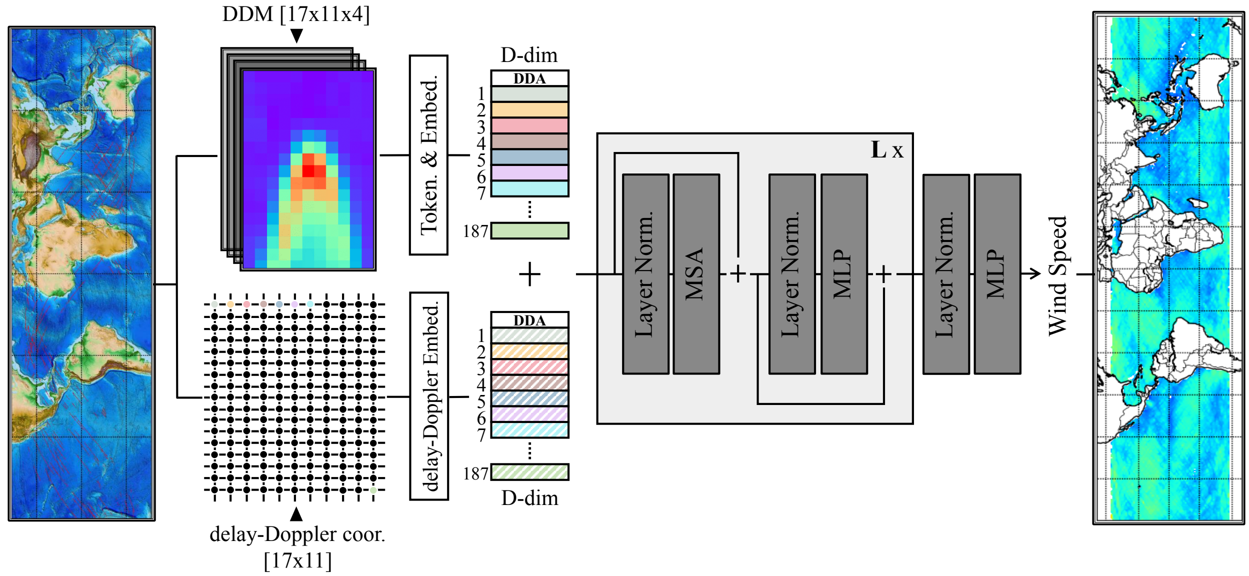

Fig. 2. Schematic of the proposed DDM-Former. L Transformer encoder layers exploit an embedded DDM with delay-Doppler coordinates to retrieve the ocean wind speed at a given location.

on multiple computer vision benchmarks. Differing from CNNs, Transformer-based models are able to capture long-range dependencies by their attention mechanism [14] rather than with local kernels. Additionally, instead of including translation equivariance that is beneficial for processing natural images, Transformers can learn to exploit and perceive DDMs independently.

With the hypothesis that networks with a global receptive field and a weaker inductive bias can better learn the delay-Doppler correlation in DDMs, in this study, we devise a Transformer-based model, termed DDM-Former, for ocean wind speed retrieval. Our model achieves promising results compared to the other retrieval methods.

2. METHOD

For CNN-based models, features extracted from homogeneous objects in a natural image are considered the same or similar, regardless of their spatial locations. However, the abscissa and ordinate of pixels in DDMs contain the corresponding Doppler frequency shifts and signal delays rather than locality information. The central idea of our method is to aggregate fine-grained delay-Doppler correlation globally and explore the underlying mapping from DDMs to the corresponding wind speeds with a weaker inductive bias.

The overall architecture of the proposed DDM-Former is depicted in Figure 2. It is trained with four channels of DDMs (namely, DDM bistatic radar cross section (BRCS), the corresponding effective scattering area, analog power, and raw counts) to estimate a wind speed over the glistening zone. Supposing that we have an input DDM $\boldsymbol{x}_D \in \mathbb{R}^{H\times W\times C}$, where H, W, C are height, width, and the number of channels, respectively. Instead of separating the input into non-overlapping small patches, we deliberately perform a pixel-wise tokenization over the DDM. Subsequently, in order to enhance the aggregation of global features, an extra learnable vector is added so that the pixel sequence $\boldsymbol{x}_p$ becomes a matrix of size $(M+1)\times C$, where $M = H \times W$. In addition, we incorporate DDM coordinates that hold rich cross-correlation for scattering signals, i.e., pixels of the same abscissa are essentially on an equi-Doppler line. This strategy helps the model to take into account the distinct delay and Doppler in DDMs.

Furthermore, the tokenized DDM that carries scattering power and its delay-Doppler correlation is fed into a linear transformation to produce a feature embedding. This embedding then served as the input of several Transformer encoder layers. For each Transformer encoder layer, it contains several vital components: multi-head self-attention (MSA) sublayers [14], MLP sublayers, and layer normalization (LN). As we calculate self-attentions of the embedded pixels independently over all attention heads, the model exploits delay-Doppler correlation within MSA sublayers through a reweighting process. This mechanism allows the model to focus on both global delay-Doppler correlation and contextual information adaptively. To perform normalization across feature dimension and smooth gradients during the training, LN is added before each MSA and MLP sublayers [15]. L Transformer encoder layers take embeddings as input and generate feature vector $\boldsymbol{x}_l$ as follows:

$$\boldsymbol{x}_l' = \mathrm{MSA}(\mathrm{LN}(\boldsymbol{x}_{l-1})) + \boldsymbol{x}_{l-1}\,, \tag{1}$$

$$\boldsymbol{x}_l = \mathrm{MLP}(\mathrm{LN}(\boldsymbol{x}_l')) + \boldsymbol{x}_l'\,, \tag{2}$$

Table 1. RMSE values of the proposed DDM-Former compared to the other retrieval methods for different wind speed intervals.

Method	All samples RMSE $(\mathrm{m/s})$	$2.5\,\mathrm{m/s} < v \leq 4\,\mathrm{m/s}$ RMSE $(\mathrm{m/s})$	$4\,\mathrm{m/s} < v \leq 8\,\mathrm{m/s}$ RMSE $(\mathrm{m/s})$	$8\,\mathrm{m/s} < v \leq 12\,\mathrm{m/s}$ RMSE $(\mathrm{m/s})$	$12\,\mathrm{m/s} < v \leq 16\,\mathrm{m/s}$ RMSE $(\mathrm{m/s})$	$16\,\mathrm{m/s} < v \leq 20\,\mathrm{m/s}$ RMSE $(\mathrm{m/s})$	$v > 20\,\mathrm{m/s}$ RMSE $(\mathrm{m/s})$
MVE [2]	1.92	**1.23**	1.39	2.54	4.75	7.24	10.22
CyGNSSnet [12]	1.55	1.63	1.35	1.69	**3.07**	4.67	7.87
DDM-Former	**1.43**	1.49	**1.18**	**1.63**	3.15	**4.65**	**7.77**

in which $l = 1, 2, \cdots, L$. Finally, the output representation from Transformer encoder layers is normalized by an LN layer, flattened, and passed through an MLP head to generate an ocean wind speed estimation.

Compared with previous CNN-based algorithms, the advantage of our method is to encourage the model to adaptively pay attention to its "regions of interest" in DDMs, which strengthens the network's ability to distinguish critical information from the inputs independently.

3. EXPERIMENTAL SETUP

We evaluate the proposed DDM-Former with the CYGNSS version 3.0 dataset [16] to verify retrieval performance. Our training data contain 318 days of measurements, with validation data from May 2020 to August 2020, followed by nine months of test data. Temporally clustered training, validation, and test sets allow us to evaluate model generality and robustness with unseen samples. In addition, ground truth wind speeds are labeled with nearest neighbor wind speed estimates from ERA5 data. To validate our method's effectiveness, the proposed model is compared with the conventional retrieval algorithm minimum variance estimator (MVE) [2], and baseline model CyGNSSnet [12] over 266 days of the test period.

Before model training, we carry out quality control procedures to eliminate low-quality data as in [12], e.g., we remove any observation with a receiver antenna gain in the direction of the specular point and a direct signal-to-noise ratio that are less than $0\,\mathrm{dB}$. After quality control, 8.0×10^6 training samples, 3.3×10^6 validation samples, and 4.5×10^6 test samples are retained. In addition, to bring four channels of input DDMs into the same scale and stabilize the training, processed datasets are normalized with zero mean and unit variance. We apply Gaussian error linear unit (GELU) as the activation function, and use mean squared error (MSE) as the loss function.

4. RESULTS & DISCUSSION

Table 1 quantifies the performance of the evaluated models over the nine months test period for different wind speed intervals. Generally speaking, these results demonstrate a reasonably competitive performance of deep learning-based methods in wind speed estimation. Although only DDMs are used for model inputs without other auxiliary parameters, such models still show smaller RMSE values than the conventional method. It proves that data-driven approaches are able to learn an underlying mapping from DDMs corresponding to wind speeds by giving sufficient training samples with appropriate network design.

Table 2. RMSE, bias, MAPE, and R^2 score for different methods.

Method	RMSE $(\mathrm{m/s})$	Bias $(\mathrm{m/s})$	MAPE (%)	R^2 score
MVE [2]	1.92	-0.98	20.8	0.29
CyGNSSnet [12]	1.55	0.14	18.4	0.55
DDM-Former	$\mathbf{1.43}$	$\mathbf{-0.02}$	$\mathbf{16.9}$	$\mathbf{0.61}$

By exploiting the delay-Doppler correlation independently, the proposed DDM-Former achieves the best overall performance with a root mean square error (RMSE) of $1.43\,\mathrm{m/s}$, outperforming the respectable baseline model CyGNSSnet by 7.7% and the conventional retrieval algorithm MVE by 25.5%. In addition, our model reveals a significant improvement in the lowest RMSE of $1.18\,\mathrm{m/s}$ at the wind speed interval of 4–8 $\mathrm{m/s}$. A similar performance could also be observed with wind speed intervals of 8–12 $\mathrm{m/s}$ and larger than $20\,\mathrm{m/s}$.

The statistical results with multiple evaluation metrics of the models are shown in Table 2. The metrics all concur on the consistent improvement of DDM-Former's performance; in particular, an average prediction bias that is close to zero. Essentially, our model achieves a lower estimation residuals, and better mean absolute percentage error (MAPE) and R^2 scores compared to the other competitors. These results confirm our assumption that a Transformer-based model equipped with MSA can efficiently learn the complex delay-Doppler correlation from DDMs, thus leading to increased accuracy in ocean wind speed estimation.

5. CONCLUSION

With low-cost GNSS-R constellation data, we show that the proposed DDM-Former can be well applied to the task of wind speed retrieval. Compared to the conventional retrieval method and CNN-based algorithms, our method can adaptively explore and exploit delay-Doppler correlation in DDMs by utilizing Transformer-based models with the attention mechanism.

Moreover, there are still opportunities for further enhancements in creating more generalized models to improve estimation performance, especially in strong wind regimes. GNSS-R integrated with deep learning could be further developed to produce enhanced retrieval products with increasing constellations and data.

6. REFERENCES

[1] Daixin Zhao, Konrad Heidler, Milad Asgarimehr, Caroline Arnold, Tianqi Xiao, Jens Wickert, Xiao Xiang Zhu, and Lichao Mou, "DDM-Former: Transformer networks for GNSS reflectometry global ocean wind speed estimation," *Remote Sensing of Environment*, in press.

[2] Maria Paola Clarizia, Christopher S. Ruf, Philip Jales, and Christine Gommenginger, "Spaceborne GNSS-R Minimum Variance Wind Speed Estimator," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 6829–6843, 2014.

[3] Adriano Camps, Hyuk Park, Miriam Pablos, Giuseppe Foti, Christine P. Gommenginger, Pang Wei Liu, and Jasmeet Judge, "Sensitivity of GNSS-R Spaceborne Observations to Soil Moisture and Vegetation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 10, pp. 4730–4742, 2016.

[4] Christopher S. Ruf, Clara Chew, Timothy Lang, Mary G. Morris, Kyle Nave, Aaron Ridley, and Rajeswari Balasubramaniam, "A New Paradigm in Earth Environmental Monitoring with the CYGNSS Small Satellite Constellation," *Scientific Reports*, vol. 8, no. 1, pp. 1–13, 2018.

[5] Valery U. Zavorotny, Scott Gleason, Estel Cardellach, and Adriano Camps, "Tutorial on Remote Sensing Using GNSS Bistatic Radar of Opportunity," *IEEE Geoscience and Remote Sensing Magazine*, vol. 2, no. 4, pp. 8–45, 2014.

[6] James L. Garrison, Attila Komjathy, Valery U. Zavorotny, and Stephen J. Katzberg, "Wind speed measurement using forward scattered GPS signals," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 1, pp. 50–65, 2002.

[7] Maria Paola Clarizia and Christopher S. Ruf, "Wind Speed Retrieval Algorithm for the Cyclone Global Navigation Satellite System (CYGNSS) Mission," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4419–4432, 2016.

[8] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer, "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[9] Yunxiang Liu, Ian Collett, and Y. Jade Morton, "Application of Neural Network to GNSS-R Wind Speed Retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9756–9766, 2019.

[10] Milad Asgarimehr, Irina Zhelavskaya, Giuseppe Foti, Sebastian Reich, and Jens Wickert, "A GNSS-R Geophysical Model Function: Machine Learning for Wind Speed Retrievals," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 8, pp. 1333–1337, 2020.

[11] Xiaohan Chu, Jie He, Hongqing Song, Yue Qi, Yueqiang Sun, Weihua Bai, Wei Li, and Qiwu Wu, "Multimodal Deep Learning for Heterogeneous GNSS-R Data Fusion and Ocean Wind Speed Retrieval," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5971–5981, 2020.

[12] Milad Asgarimehr, Caroline Arnold, Tobias Weigel, Chris Ruf, and Jens Wickert, "GNSS reflectometry global ocean wind speed using deep learning: Development and assessment of CyGNSSnet," *Remote Sensing of Environment*, vol. 269, pp. 112801, 2022.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.

[14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *9th International Conference on Learning Representations*, 2021.

[15] Qiang Wang, Bei Li, Tong Xiao, Jingbo Zhu, Changliang Li, Derek F. Wong, and Lidia S. Chao, "Learning Deep Transformer Models for Machine Translation," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019.

[16] CYGNSS. 2020. CYGNSS Level 1 Science Data Record Version 3.0. Ver. 3.0. PO.DAAC, CA, USA. Dataset accessed [2021-11-06] at https://doi.org/10.5067/CYGNS-L1X30.