DDM-Former: Transformer Networks for GNSS Reflectometry Global Ocean Wind Speed Estimation

Daixin Zhao[a,b], Konrad Heidler[a], Milad Asgarimehr[c,d], Caroline Arnold[e], Tianqi Xiao[c,d], Jens Wickert[c,d], Xiao Xiang Zhu[a], Lichao Mou[b,a,*]

[a] *Technische Universität München, Germany*
[b] *German Aerospace Center (DLR), Oberpfaffenhofen, Germany*
[c] *German Research Centre for Geosciences (GFZ), Potsdam, Germany*
[d] *Technische Universität Berlin, Germany*
[e] *German Climate Computing Center (DKRZ), Hamburg, Germany*

Abstract

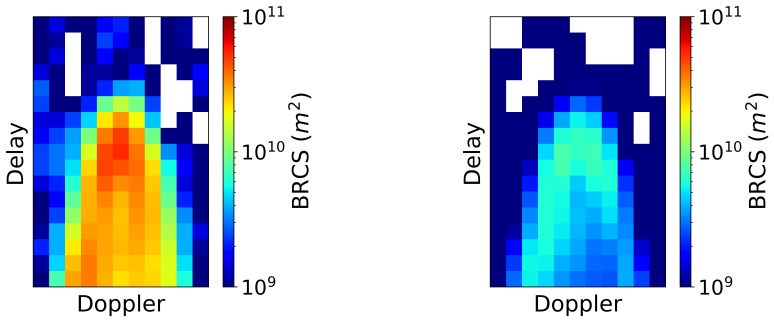
Global navigation satellite system reflectometry (GNSS-R) has shown a capability in recent years to be applied as a novel remote sensing technique to retrieve ocean wind speeds. The combination of GNSS-R observable delay-Doppler maps (DDMs) and deep learning algorithms provides the possibility to build an end-to-end pipeline for improving wind speed estimations. Recent studies have proven that data-driven approaches can be applied to generate enhanced estimation products. However, these are usually trained with convolutional neural networks (CNNs), which include inductive bias throughout the models. The inbuilt translation equivariance in CNNs represents an inexactitude for the feature extraction on DDMs. To address this issue, we propose Transformer-based models, named DDM-Former and DDM-Sequence-Former (DDM-Seq-Former), to exploit delay-Doppler correlation within and between DDMs, respectively. The advantages of our methods over conventional retrieval algorithms and other deep learning-based approaches are presented based on the Cyclone GNSS (CYGNSS) version 3.0 dataset. In addition, a comprehensive study on the attention mechanism for our models is demonstrated. The proposed DDM-Former yields the best overall performance with a root mean square error (RMSE) of 1.43 m/s and a bias of −0.02 m/s over the nine months test period. Moreover, with an RMSE of 2.89 m/s and a bias of −1.88 m/s, the proposed DDM-Seq-Former can promisingly improve the estimations in the cases with wind speeds higher than 12 m/s. There are still opportunities for further enhancements in creating more robust models that could perform well in all wind regimes given a non-uniform wind distribution. We will make our code publicly available.

Keywords: Deep learning, GNSS reflectometry, ocean wind speed, Transformer networks

1. Introduction

Ocean surface wind is one of the key parameters for monitoring ocean dynamics and climate change [1, 2]. Many human activities, including maritime operations, ship routing, and offshore fishing, are heavily

*Corresponding author
Email address: lichao.mou@tum.de (Lichao Mou)



(a) (b)

Figure 1: Illustration of CYGNSS Level-1 bistatic radar cross section (BRCS) DDMs at different wind speeds: (a) below 3 m/s, (b) above 20 m/s.

influenced by the surface wind state [3]. Having accurate knowledge of the ocean wind climatology also plays a vital role in planning and maintaining offshore wind energy facilities, which can help reduce the worldwide dependence on fossil fuels and help alleviate global warming [4]. For all these reasons, it is important to be able to reliably assess ocean wind speed variability on a global scale.

As a novel remote sensing technique, global navigation satellite system reflectometry (GNSS-R) can be applied to sense various types of the Earth surface parameters, such as the ocean wind speed. By exploiting scattered signals of opportunity from the GNSS, the surface state in a glistening zone can be derived [5, 6, 7, 8]. With the dense GNSS signal coverage nowadays, GNSS-R features with its significant improvements in spatial-temporal resolution and global coverage [9]. Additionally, cost-efficient and power-efficient receivers allow developing constellations, such as the UK TechDemoSat-1 mission (TDS-1) [10], NASA's Cyclone GNSS (CYGNSS) [11], and the Chinese Bufeng-1 A/B [12].

One of the most important GNSS-R observables is the so-called delay-Doppler map (DDM), which is a two-dimensional (2D) image of scattering coefficients as a function of propagation delay and Doppler shift [13]. Diametrically different from natural images, the abscissa and ordinate of DDMs do not correspond to spatial dimensions. Instead, the Doppler dimension records the frequency shift of scattered signals while the delay dimension represents the difference in traveling path length upon arrival at the GNSS-R receiver [9]. As the ocean surface roughness is altering due to ocean winds, the processed DDMs also change accordingly [14], as shown in Figure 1. Therefore, DDMs carrying fine-grained delay-Doppler correlations can be used for retrieving the ocean wind speeds.

Many previous works have proved the feasibility of estimating ocean wind speeds using DDMs [3, 8, 15, 16]. With the rise of deep learning in the field of remote sensing and with the increasing amount of GNSS-R data collected in recent years [17, 18], several studies have shown how data-driven approaches are opening up novel prospects for generating enhanced ocean wind speed products and address potential limitations in conventional retrieval methods. A few studies start by investigating the use of multilayer perceptrons (MLPs). In [13], the authors have proposed an MLP with two hidden layers. By combining DDM, DDM

average, leading edge slope (LES), and incidence angle in a feature set, an overall root mean square error (RMSE) of 1.79 m/s is achieved with true wind speeds ranging from 3 to 20 m/s. Asgarimehr *et al.* [19] further enrich the feature set of the MLP model to eight parameters by including features like σ^0 in dB, specular point elevation in the orbital reference frame in degrees, and antenna gain toward specular point in dB. In addition, they perform a feature selection procedure based on variables that decrease the RMSE values for a validation set. The advantages of utilizing MLPs are also demonstrated in [20, 21, 22]. Furthermore, a hybrid neural network model, named FA-RDN, achieves enhanced wind speed estimation performance in [23]. The model in that study utilizes a recurrent deep neural network to extract DDM features, spatial-temporal features, and spacecraft attitudes. The obtained features are then weighted by a feature attention mechanism to improve the retrieval accuracy.

More recently, convolutional neural network (CNN) models, which are well-known in computer vision, have shown a strong capability to estimate ocean wind speeds with DDMs. Rather than perform feature engineering empirically on DDMs, CNNs handle the DDMs as images and aggregate information with sliding convolutional kernels. Chu *et al.* [24] utilize a heterogeneous data fusion architecture to incorporate DDM data with all the satellite receiver status parameters. Eight convolutional layers and three fully connected layers are implemented to extract features from the input DDMs. In [25], the authors propose a two-branch network named CyGNSSnet, which results in a significant improvement in performance compared to conventional retrieval algorithms. Guo *et al.* [26] employ a CNN corrected with a cumulative distribution function (CDF) matching approach that demonstrate a robust performance with the test data, with an RMSE of 1.53 m/s with 0–25 m/s wind speed intervals. These studies prove that learning features by treating DDMs as images is indeed beneficial for the wind speed retrieval. Besides, the effectiveness of coupling DDMs with CNNs has been further demonstrated in other tasks using GNSS-R techniques, such as soil moisture estimation [27, 28] and sea ice detection [29, 30].

While CNNs are designed for extracting features from natural images, their inbuilt inductive bias can be misleading for perceiving DDMs. We thus turn to a more recent network with a weaker inductive bias named Transformer [31]. Transformer was initially designed for machine translation and now serves as the reference model for many natural language processing (NLP) tasks. By directly applying the Transformer architecture to natural images, the pioneering work on vision Transformer (ViT) [32] leads a trailblazing trend of employing Transformers in computer vision. Most recently, Transformer-based models have become almost ubiquitous in the computer vision community and have achieved state-of-the-art performance on multiple computer vision benchmarks, such as object detection [33, 34], semantic segmentation [35], and video understanding [36].

Differing from CNNs, ViT separates an input image into non-overlapping image patches, and uses multi-head self-attention (MSA) [32] sublayers to model the relationships among the patches explicitly. This is of great interest to our task due to the existence of delay-Doppler correlation in DDMs, which can be better captured by the attention mechanism globally rather than with local kernels. Moreover, rather than

including translation equivariance that is beneficial for processing natural images, Transformers can learn to exploit and perceive DDMs independently. We hypothesize that networks with these characteristics can better understand the fine-grained delay-Doppler correlation in DDM data. Inspired by ViT, we propose a Transformer-based model, termed DDM-Former, for performing wind speed estimations with DDMs. Further, considering that measurements made by a satellite are essentially spatial-temporal coherent, we formulate a variant of DDM-Former with sequential input DDMs, named DDM-Sequence-Former (DDM-Seq-Former). Instead of focusing on individual DDMs, the sequential setup is designed to learn distinguishable features from a DDM sequence. We show that both the proposed models are indeed beneficial for performing the task, but in different aspects.

The remainder of the paper is organized as follows: Section 2 presents the used CYGNSS dataset for performing ocean wind speed estimation and quality control. In Section 3, details of the proposed Transformer-based models, namely DDM-Former and DDM-Seq-Former, are provided. A comprehensive study on the attention mechanism of our models is also presented. The evaluation metrics and training details are provided in Section 4. Section 5 presents the numerical and geographical results, and spacecraft, spatial resolution, and data related discussion. Finally, we provide the overall summary of this work and a future perspective for further work in this field in Section 6.

2. Dataset and quality control

2.1. Data description

To verify the performance of our proposed methods for estimating global ocean wind speeds, we choose the CYGNSS version 3.0 dataset [37] with records from July 2019 to April 2021. Our training data contain 318 days of measurements, with validation data from May 2020 to August 2020, followed by 266 days of test data. More specifically, we use DDM data that include DDM bistatic radar cross section (BRCS), the corresponding effective scattering area, analog power, and raw counts. DDM BRCS represents the true surface scattering area (m^2) that contributes power to each DDM bin. The corresponding effective scattering area (m^2) is calculated by convolving the GPS ambiguity function with the surface area that contributes power to a given DDM bin, as determined by its delay-Doppler cross-correlation and the measurement geometry. The analog power represents the true power that would have been measured by an analog power sensor, and the raw counts are the uncalibrated power values measured by the DDM instrument.

Ground truth wind speeds are labeled with European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis data [38]. The ERA5 wind speed products have a temporal resolution of 1 h with $0.5^\circ \times 0.5^\circ$ grid resolution. We use nearest neighbors to match the CYGNSS measurements with the ERA5 wind speed estimates. Note that DDM-Seq-Former only takes spatial-temporal coherent measurements as sequential inputs; we therefore only utilize the data acquired by CYGNSS spacecraft 1 for comparison in our experiments.

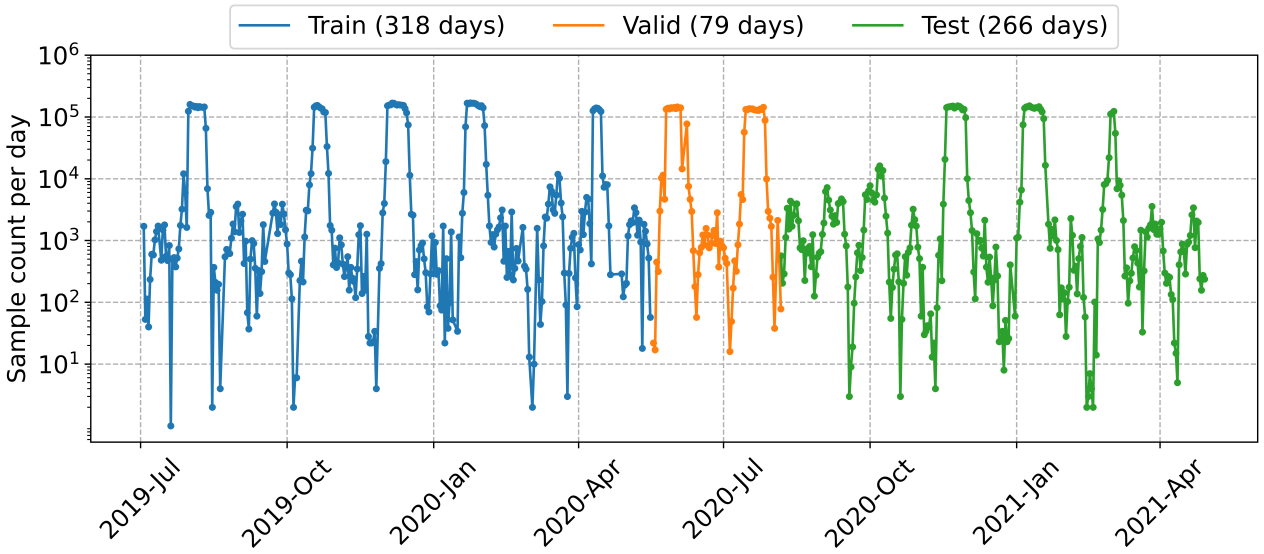


Figure 2: Temporal distribution of the training, validation, and test sets acquired by CYGNSS spacecraft 1 after eliminating the low-quality measurements.

2.2. Quality control

As deep learning methods are data-driven, DDMs of good quality are essential to give the model a correct intuition of how to map different ocean wind speeds with their corresponding DDMs. Several quality control procedures are carried out to eliminate low-quality measurements from the data [25]. We first remove samples with the LES and range corrected gain (RCG) figure of merit (FOM) for the DDMs that are smaller than zero. Data with the receive antenna gain in the direction of the specular point and a direct signal-to-noise ratio that are smaller than 0 dB are filtered out. Further, data with an uncertainty of the BRCS below one and a nano star tracker attitude status of "OK" are retained. We also keep samples with an absolute value of the spacecraft roll between 1° and 30°, pitch between 1° and 10°, or yaw between 1° and 5°, as indicated by the quality flags in the L1 science data record. Moreover, as CYGNSS measurements or ERA5 estimations may also be contaminated by unexpected errors which normally do not affect the major population, samples that lie outside the confidence interval of 95% are removed. Finally, due to the ambiguity caused by scattering mechanism changes at low wind speed [39], samples below 2.5 m/s are eliminated.

After filtering out the low-quality measurements, 8.0×10^6 training samples, 3.3×10^6 validation samples, and 4.5×10^6 test samples are still retained. Temporally clustered training, validation, and test sets allow us to evaluate the model generality and robustness with unseen samples. Figure 2 illustrates the temporal distribution of the three sets. The preprocessed training/validation/test sets are normalized to zero mean and unit variance to bring four channels of the input DDMs into the same scale and to stabilize the training.

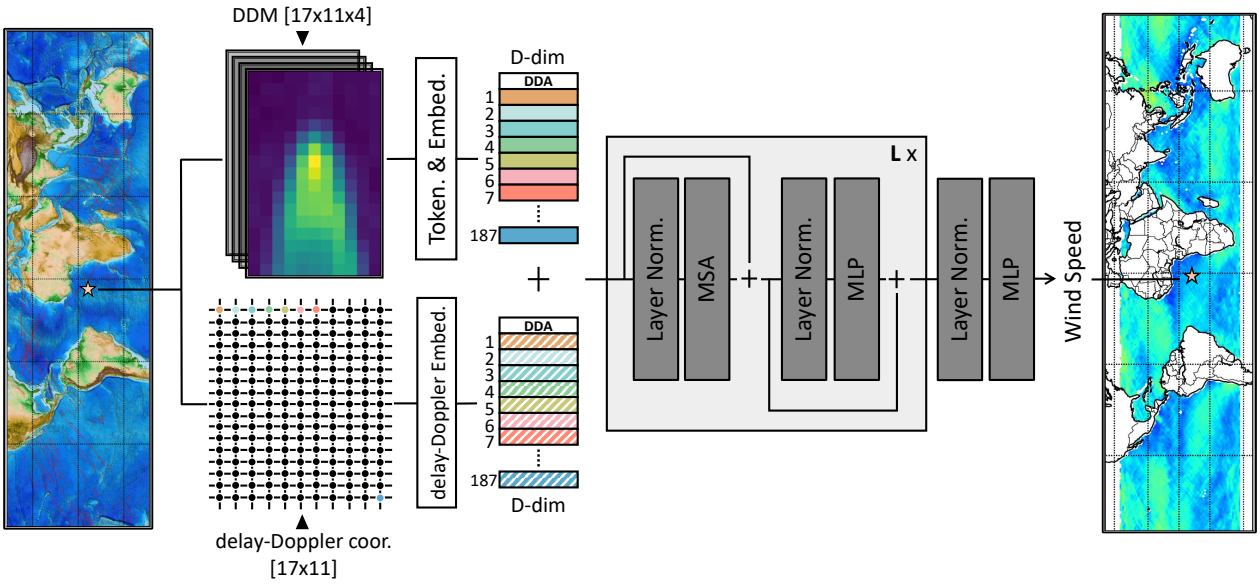


Figure 3: Schematic of the proposed DDM-Former. A DDM is embedded with delay-Doppler coordinates to generate an intermediate feature map with a shape of $188 \times D$ and passed through L Transformer encoder layers. The feature is normalized by layer normalization (LN) sublayers, and the multi-head self-attention (MSA) sublayer exploits delay-Doppler correlation. The output of the Transformer encoder layers is then normalized, flattened, and passed through an MLP head to predict the ocean wind speed with the given DDM.

3. Methodology

3.1. DDM-Former

Deep learning is well applied in remote sensing, and CNNs have already been adopted in several studies for ocean wind speed estimation [24, 25, 26]. These networks take DDMs as images to predict ocean wind speeds at the corresponding specular points. On the other hand, Transformer-based networks, e.g., ViT, yield great results in a variety of computer vision tasks. The unique features of ViT that differ from CNNs encourage us to try to exploit Transformers for retrieving ocean wind speeds.

Many empirical studies have proved that the translation equivariance of CNNs is ideal for dealing with natural images. Networks with a CNN backbone can use moving convolutional kernels to extract contextual information from the input images regardless of their locality. As a result, the corners and edges of objects are considered the same or similar in different locations. However, this inbuilt inductive bias might be misleading when we look at the formation of DDMs. A DDM is a relatively small 2D image (17×11 pixels) with fine-grained delay-Doppler correlation, and each pixel contains its own cross-correlating scattered signal with regard to a different time delay and carrier frequency offset [7, 40]. In terms of contextual information, DDMs immensely differ from natural images that we normally apply to CNNs. Extracting "corners" and "edges" in DDMs with weight-shared filters can lead to ambiguity in understanding crucial information that lies in the data. Thus, introducing translation equivariance to networks is an inexactitude for feature learning on DDMs.

Additionally, the subsampling or pooling which normally comes with CNNs tends to only propagate the most dominant features to the following layers. This is typically required for networks to reduce the spatial size of feature maps and enable handling of larger spatial contexts. However, applying these techniques to DDMs inevitably mixes small yet exquisite information with neighboring pixels, which might lead to performance degradation. On the contrary, ViT features the ability to aggregate information globally in each Transformer encoder layer. In order to address the aforementioned issues and put our hypotheses into practice, we devise a network, termed DDM-Former, to learn feature representations from DDMs.

An overview of the proposed DDM-Former is depicted in Figure 3. Let us assume that we have an input DDM $\boldsymbol{x} \in \mathbb{R}^{H \times W \times C}$, where H and W are the height and width of the DDM, and C represents the number of channels (namely, DDM BRCS, the corresponding effective scattering area, analog power, and raw counts). In the setup of DDM-Former, H, W, and C are 17, 11, and 4, respectively.

Since each pixel in the input DDM carries its own distinct frequency delay and Doppler shift, we perform a pixel-wise tokenization over the DDM instead of splitting it into non-overlapping small patches. Concretely, we reshape the input DDM $\boldsymbol{x}$, and add an extra learnable vector delay-Doppler aggregator (DDA) to formulate a pixel sequence $\boldsymbol{x}_p$, represented as:

$$\boldsymbol{x}_p = [\boldsymbol{x}_{\mathrm{DDA}};\, \boldsymbol{x}_p^1;\, \boldsymbol{x}_p^2;\, \cdots;\, \boldsymbol{x}_p^N]\,, \tag{1}$$

in which $N = H \times W$ is the number of pixels, and therefore the dimension of the sequence $\boldsymbol{x}_p$ is $(N+1) \times 4$. The DDA is randomly initialized and concatenated to the beginning of the pixel sequence. It can be fine-tuned during the training process and works as an ancillary vector to aggregate the global features from each DDM pixel. Afterwards, we apply a linear projection to embed the tokenized DDM $\boldsymbol{x}_p$ into a sequence of D-dimensional features, expressed as follows:

$$\boldsymbol{x}_e = \boldsymbol{x}_p \boldsymbol{E}_0\,, \tag{2}$$

where $\boldsymbol{x}_e \in \mathbb{R}^{(N+1) \times D}$, and $\boldsymbol{E}_0$ is a weight matrix that turns the pixel sequence $\boldsymbol{x}_p$ from four to D dimensions through the linear projection. This represents a learnable mapping, and its parameters are randomly initialized and optimized in the training. Moreover, the abscissa and ordinate of the DDM hold abundant cross-correlation information for the scattered signals over certain surface contours, e.g., pixels of the same abscissa are essentially on an equi-Doppler line. Therefore, we incorporate delay-Doppler coordinates to reinforce the model's learnability for preserving contextual information in a holistic way. The input token for the Transformer encoder layers can be expressed as follows:

$$\boldsymbol{x}_0 = \boldsymbol{x}_e + \boldsymbol{x}_{coor}\,, \tag{3}$$

where $\boldsymbol{x}_{coor} \in \mathbb{R}^{(N+1) \times D}$ is a delay-Doppler embedding and can be updated during the network training.

Subsequently, the embedded pixels are fed into several Transformer encoder layers. Each Transformer encoder layer contains MSA [32] and MLP sublayers. Self-attention calculates attention weights based on

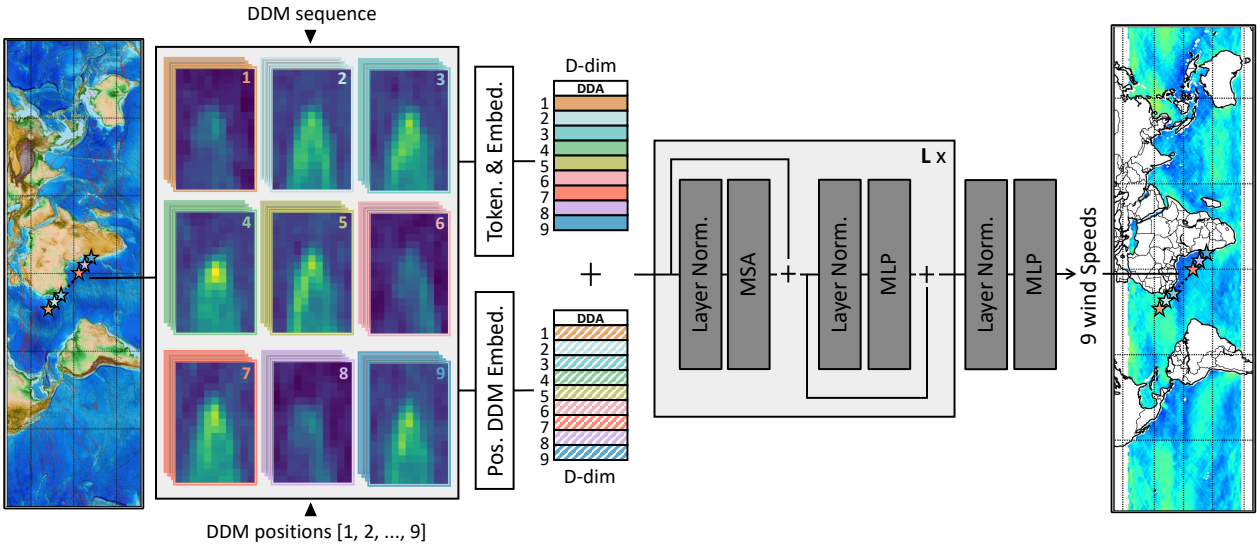


Figure 4: Demonstration of the proposed DDM-Seq-Former, a variant of DDM-Former. Nine spatial-temporal coherent DDMs are tokenized and embedded along with a positional DDM embedding. A feature map with the shape $10 \times D$ is processed by L layers of Transformer encoder layers. As a result, the final prediction is an ensemble of nine wind speeds corresponding to the input DDM sequence.

the pairwise similarity between two input sequences, and MSA is an extension of self-attention that repeats its computation multiple times in parallel. Each of these self-attention operations is called an attention head. In addition, dropout is applied following each fully connected layer within the MLP block to prevent overfitting [41]. Layer normalization (LN) is added before each block [42]. This is done by normalizing each training sample along feature dimensions with its mean and standard deviation. LN is thus invariant to per sample feature shifting and scaling, and it helps to speed-up the training and smoothing gradients [43, 44]. L Transformer encoder layers can then be described as follows:

$$\begin{aligned} \boldsymbol{x}'_l &= \mathrm{MSA}(\mathrm{LN}(\boldsymbol{x}_{l-1})) + \boldsymbol{x}_{l-1}\,, \\ \boldsymbol{x}_l &= \mathrm{MLP}(\mathrm{LN}(\boldsymbol{x}'_l)) + \boldsymbol{x}'_l\,, \end{aligned} \tag{4}$$

where $l = 1, 2, \cdots, L$. The MSA sublayer in each Transformer encoder layer encodes multiple relationships among the embedded pixels explicitly and allows the model to learn fine-grained delay-Doppler correlation within a DDM. Finally, the output representation from the Transformer encoder layers is normalized by an LN layer, flattened, and passed through an MLP head to generate an ocean wind speed estimation:

$$y = \mathrm{MLP}(\mathrm{Flatten}(\mathrm{LN}(\boldsymbol{x}_L)))\,. \tag{5}$$

The advantage of this learning strategy is to encourage DDM-Former to adaptively pay attention to its "regions of interest (ROIs)" in DDMs, which strengthens the network's ability to distinguish critical information from the inputs independently.

3.2. DDM-Seq-Former

As GNSS-R satellites receive signals following distinct tracks, it is clear that successive observations already exhibit a strong spatial-temporal correlation. However, training models with individual DDMs regrettably means losing this valuable information. Another consideration for training models with multiple DDMs to estimate ocean wind speeds is that the total distribution of labeled samples is extremely unbalanced. Moderate wind speeds (e.g., 4–10 m/s) are major candidates and account for more than 75% of the data. Models trained with individual DDMs with an unbalanced distribution are consequently not exposed enough to extreme cases. It is a notorious problem that the performance for rare cases often suffers from overfitting to moderate winds and leads to the performance dropping in extreme wind regimes. Therefore, to enhance network awareness with extreme wind speed samples, we take advantage of the attention mechanism over spatial-temporal coherent DDMs to formulate a variant of DDM-Former, termed DDM-Seq-Former.

Figure 4 demonstrates the architecture of DDM-Seq-Former. Divergent from DDM-Former, our sequential setup takes a sequence of nine spatial-temporal coherent DDMs as an input $\boldsymbol{x} \in \mathbb{R}^{9\times H\times W\times C}$, where the height H, width W, and the number of channels C of a component DDM in the sequence stay the same with those in DDM-Former. Here, our goal is to learn a network that can distinguish among distinct wind speeds in a DDM sequence. Consequently, we tokenize each component DDM as a whole, and compact them into a sequence of DDMs with DDA, represented as follows:

$$\boldsymbol{x}_p = [\boldsymbol{x}_{\mathrm{DDA}};\ \boldsymbol{x}_p^1;\ \boldsymbol{x}_p^2;\ \cdots\ ;\ \boldsymbol{x}_p^9]\,. \tag{6}$$

With a trainable linear projection, the DDM sequence is mapped to a D-dimension space as $\boldsymbol{x}_e \in \mathbb{R}^{(9+1)\times D}$. The DDM position that carries coherent information in the temporal domain is embedded and added to formulate an intermediate feature map:

$$\boldsymbol{x}_0 = \boldsymbol{x}_e + \boldsymbol{x}_{pos}\,, \tag{7}$$

where the positional DDM embedding $\boldsymbol{x}_{pos} \in \mathbb{R}^{(9+1)\times D}$. The feature map is then fed into the following Transformer encoder layers, and MSA helps the model to learn distinguishable features between different DDMs from the input DDM sequence. Finally, the output representation from the Transformer encoder layers $\boldsymbol{x}_L$ is normalized, flattened, and passed through an MLP head. Unlike DDM-Former, DDM-Seq-Former can predict nine wind speeds at one-time, which can be expressed as:

$$[y_1,\ y_2,\ \cdots,\ y_9] = \mathrm{MLP}(\mathrm{Flatten}(\mathrm{LN}(\boldsymbol{x}_L)))\,. \tag{8}$$

For a fair comparison with other methods, we still employ the same training samples, but compact coherent DDMs so that the model is "seeing" nine DDMs at once. This approach can be interpreted as the model concentrating more on perceiving distinguishable features between different DDMs to aggregate information from spatial-temporal coherent data.

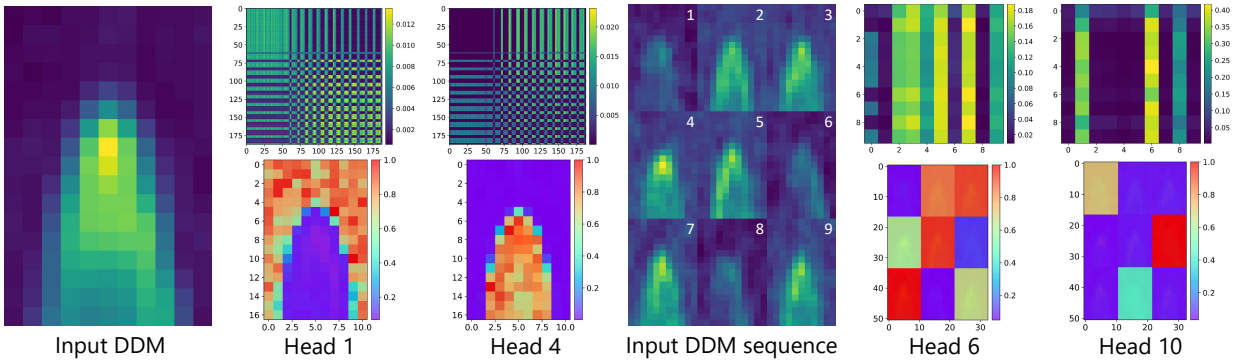


Figure 5: Attention visualizations for the proposed DDM-Former (a) and DDM-Seq-Former (b). For each model, on the left is the input DDM (sequence), while the top-right is the attention grid, and the bottom-right is the attention map. Note that DDM-Former uses pixel-wise attention, while DDM-Seq-Former has attention over each individual DDM in a sequence.

3.3. Attention map and explainability

A non-trivial aspect of data-driven approaches is that one cannot directly interpret how a supervised model learns the mapping from the given inputs to the desired outputs. A so-called black-box algorithm seems less attractive if we are unable to explain the underlying logic unambiguously. In our work, this shortcoming is resolved by inspecting intermediate attention layers of the Transformer architecture.

Figure 5(a) and Figure 5(b) visualize the attention mechanism of the proposed DDM-Former and DDM-Seq-Former, respectively. Aside from the input DDM (sequence) for the proposed methods, visualizations of attention grids and attention maps are depicted. With the chosen attention head, the attention grid represents a weight matrix from an intermediate attention map, which indicates the probability distribution of the sequence alignment score. Whereas an attention map is a scalar matrix that demonstrates the relative importance of layer activation at a given DDM pixel.

By deliberately using pixel-level attention in DDM-Former, the network tries to find its own "ROIs" to differentiate between the horseshoe-shaped signal and background in an input DDM. For instance, the attention grid of head 4 in Figure 5(a) is not illuminated in the top-left corner. Moving from left to right, each pixel belonging to the background stays darkened until the very first pixel of the horseshoe-shaped signal is met in the input DDM. Henceforth, the illuminated columns in the attention grid are expanding as more pixels now belong to the horseshoe-shaped signal. The attention map further validates that head 4 pays less attention to the background of the input DDM. Head 1 works in a similar manner, but the other way around.

On the other hand, DDM-Seq-Former takes advantage of spatial-temporal coherent measurements to exploit the differences between discrepant DDMs. Rather than having a pixel-level attention, the sequential setup exhibits attention at the DDM-level. As shown in Figure 5(b), an input DDM sequence is combined with strong winds ($\sim$18 m/s, at positions 1, 6, and 8) and moderate winds ($\sim$10 m/s, others). It turns out that our model learns to find similarities between comparable winds and distinguishes the others on its own.

As a result, DDMs for moderate wind speeds are clustered in head 6, while those associated with strong winds are assorted together in head 10.

These visualizations give a clear indication that Transformer-based models are fully capable of learning critical features from different setups of DDMs. In addition, unlike CNNs with limited receptive fields, our models perceive the inputs as a whole and are able to capture long-range dependencies, thus resulting in a better learnability of the underlying mappings. With enough training samples, both methods can find out the influence of different contextual features on estimating ocean wind speeds and interpret the input DDMs or DDM sequences independently.

4. Experimental setup

4.1. Model evaluation

Evaluation is performed over a test period from August 2020 to April 2021. We choose CyGNSSnet [25] as our baseline model to validate the effectiveness of the proposed DDM-Former and DDM-Seq-Former with different wind speeds intervals. In order to compare models only with 2D input DDMs, the ancillary parameter branch of CyGNSSnet is dropped. Moreover, a conventional retrieval algorithm minimum variance estimator (MVE) [3] is also evaluated for reference purposes. MVE wind speed estimation is constructed from normalized BRCS (NBRCS) and LES. These observables derived from DDMs are sensitive to ocean surface variations and therefore can be utilized to retrieve wind speed changes.

The performance of the aforementioned models is mainly quantified with the RMSE formalized by Eq. 9 and the bias is calculated from Eq. 10:

$$\mathrm{RMSE}(v,\hat{v}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{v}_i - v_i)^2}\,, \tag{9}$$

$$\mathrm{Bias}(v,\hat{v}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{v}_i - v_i)\,, \tag{10}$$

where $\hat{v}$ is a predicted wind speed, and v is the reference wind speed from ERA5. The overall RMSE and bias are determined across a total of n samples for all models.

Additionally, the mean absolute percentage error (MAPE) and the coefficient of determination (R^2 score) are computed via Eq. 11 and Eq. 12 as follows:

$$\mathrm{MAPE}(v,\hat{v}) = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{v}_i - v_i}{v_i}\right|\,, \tag{11}$$

$$R^2(v,\hat{v}) = 1 - \frac{\sum_{i=1}^{n}(\hat{v}_i - v_i)^2}{\sum_{i=1}^{n}(\bar{v} - v_i)^2}\,, \tag{12}$$

where $\bar{v}$ is the average wind speed of the ERA5 labels.

4.2. Training details

DDM-Former and DDM-Seq-Former are implemented with Tensorflow [45]. Both models are trained on the training set within 8 hours on a single NVIDIA GeForce RTX 3090 GPU. The optimizer is set as AdamW with a learning rate of 10^{-5} and a weight decay of 10^{-5}, which yields a lower training loss, and the models are generalized better compared to a model trained with Adam [46]. We apply Gaussian error linear unit (GELU) as the activation function, which prevents neurons dying off and provides bounded activations in a negative gradient regime [47]. The commonly chosen mean squared error (MSE) is used as the loss function. We employ 512 samples per batch and use early-stopping with a patience of 6 epochs.

5. Results and discussion

To evaluate the performance and generality of the proposed models in comparison to the baseline model CyGNSSnet and conventional method MVE, in this section, we demonstrate the numerical results for different wind speed intervals. Then we present the global visualizations to further analyze the spatial distributions with regard to the RMSE and bias. Finally, we briefly discuss the findings and challenges with regard to the spacecraft, spatial resolution, and data aspects.

5.1. Model comparison

Wind speeds retrieved from the test set with different models are compared to ERA5 wind speeds. The RMSE for all test samples and with different wind speed intervals are reported in Table 1. The percentage improvements for DDM-Former and DDM-Seq-Former are calculated in comparison with the baseline network CyGNSSnet. In general, the results prove that the deep learning-based methods are fairly competitive in the task of wind speed estimation. Although these models only take DDMs as inputs without other ancillary parameters, they still achieve lower RMSE values compared to the conventional method. This demonstrates that data-driven approaches can learn the underlying mapping from DDMs to wind speeds just by giving them enough training samples with an appropriate network design. Note that the samples typically follow a long-tailed distribution, i.e., the performance measurement computed for each column is based on non-uniform sample counts.

Table 1: RMSE statistics for the conventional retrieval algorithm MVE, baseline network CyGNSSnet, and the proposed DDM-Former and DDM-Seq-Former with different wind speed intervals.

	All samples	$2.5\,\mathrm{m/s} < v \leq 4\,\mathrm{m/s}$	$4\,\mathrm{m/s} < v \leq 8\,\mathrm{m/s}$	$8\,\mathrm{m/s} < v \leq 12\,\mathrm{m/s}$	$12\,\mathrm{m/s} < v \leq 16\,\mathrm{m/s}$	$16\,\mathrm{m/s} < v \leq 20\,\mathrm{m/s}$	$v > 20\,\mathrm{m/s}$
Architecture	RMSE (m/s)	RMSE (m/s)	RMSE (m/s)	RMSE (m/s)	RMSE (m/s)	RMSE (m/s)	RMSE (m/s)
MVE	1.92	**1.23**	1.39	2.54	4.75	7.24	10.22
CyGNSSnet	1.55	1.63	1.35	1.69	3.07	4.67	7.87
DDM-Former	**1.43** (**+7.7%**)	1.49 (+8.6%)	**1.18** (**+12.6%**)	**1.63** (**+3.6%**)	3.15 (-2.6%)	4.65 (+0.4%)	7.77 (+1.3%)
DDM-Seq-Former	1.65	1.56 (+4.3%)	1.50	1.84	**2.83** (**+7.8%**)	**3.82** (**+18.2%**)	**5.82** (**+26.1%**)

Among all the listed models, the proposed DDM-Former yields the best overall performance with an RMSE of 1.43 m/s over a nine months test period, outperforming the respectable baseline model CyGNSSnet by 7.7% and conventional retrieval algorithm MVE by 25.5%. At the wind speed interval of 4–8 m/s, DDM-Former reveals a significant improvement in the lowest RMSE of 1.18 m/s. A similar performance could also be observed with wind speed intervals of less than 4 m/s and 8–12 m/s.

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

Architecture	RMSE (m/s)	Bias (m/s)	MAPE (%)	R^2 score
MVE	1.92	−0.98	20.8	0.29
CyGNSSnet	1.55	0.14	18.4	0.55
DDM-Former	**1.43**	**−0.02**	**16.9**	**0.61**
DDM-Seq-Former	1.65	0.31	19.1	0.48

What's more, in regimes with very low or extremely high wind speeds, our DDM-Seq-Former achieves considerable improvements. By incorporating an attention mechanism across multiple DDMs with spatial-temporal adjacency, our model yields an improvement of 4.3% at the wind speed interval of 0–4 m/s and a lower RMSE on average of 2.89 m/s for winds higher than 12 m/s compared to the baseline network. Since typical deep learning methods tend to have better estimations with the major candidates (e.g., moderate winds at 4–10 m/s), a promising result in extreme wind regimes is a strong indicator that the proposed model can competently deal with wind speed-imbalanced distributions. Alleviating the influence of imbalanced data further proves that the learning strategy of DDM-Seq-Former encourages the model to pay more attention to the distinguishable features between different DDMs. As a result, the model can achieve satisfactory predictions with minor wind speed candidates with the same amount of training samples.

The statistical results with different evaluation metrics for each model are reported in Table 2. All the metrics agree regarding the consistent increasing performance of DDM-Former; in particular, the average prediction bias close to 0. Essentially, DDM-Former achieves the lowest estimation residuals, and a better MAPE and R^2 score compared to the other competitors. These results confirm our assumption that a Transformer-based model equipped with MSA can efficiently exploit delay-Doppler correlation from DDMs, thus leading to enhanced estimation accuracy.

Figure 6 illustrates the wind speeds retrieved with different models compared to ERA5 winds displayed in log-density plots. Consistent with the findings in Table 1 and Table 2, the conventional method presents a larger dispersion compared to the data-driven methods, especially when the wind speed is higher than 8 m/s. We also observe a bias toward an underestimation of the wind speed in Figure 6(a), in which the majority of the predictions are below the 1:1 line. From CyGNSSnet to DDM-Former, the bump area in MVE estimation is dissolved gradually, and more predictions are symmetrically centered along the 1:1 line. The wind speeds retrieved from DDM-Former appear apparently less spread around, which indicates its lowest error standard

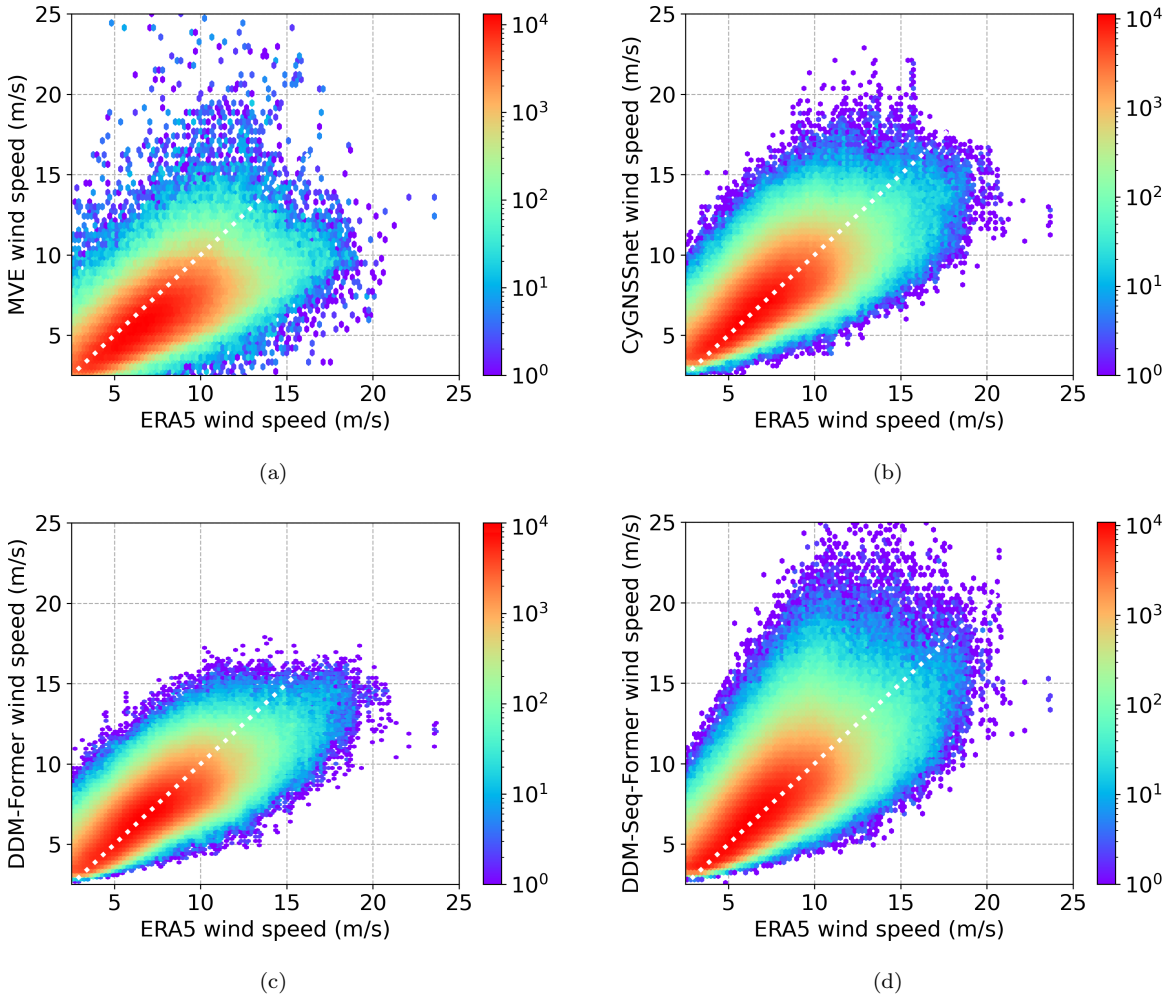


Figure 6: Density plots in log-scale for four methods against ERA5 wind speeds. The color bar measures samples per bin. Results of (a) MVE, (b) CyGNSSnet, (c) DDM-Former, and (d) DDM-Seq-Former contain nine months of test data. A 1:1 diagonal is shown as a white dotted line for reference purposes.

deviation among all the models tested. However, all the aforementioned models encounter an increasingly underestimation as the wind speeds increase (v>16 m/s). In contrast, the DDM-Seq-Former wind speeds are more symmetrically distributed along the 1:1 line compared to the other methods in strong wind regimes with reduced bias. Although we recognize the performance trade-off between moderate and extreme winds, DDM-Seq-Former still attains promising improvements with certain wind speed intervals without additional samples.

The RMSE values and bias for different wind speed intervals as a function of ERA5 wind speed are shown in Figures 7(a) and 7(b). In addition, the ERA5 sample distribution and probability density functions of the wind speeds retrieved by different models are depicted in Figure 7(c). For wind speeds lower than 12 m/s, all the models achieve satisfactory performance with an RMSE under 2 m/s, while the proposed DDM-Former obtains the lowest RMSE. For the especially challenging strong wind regimes, DDM-Seq-Former displays

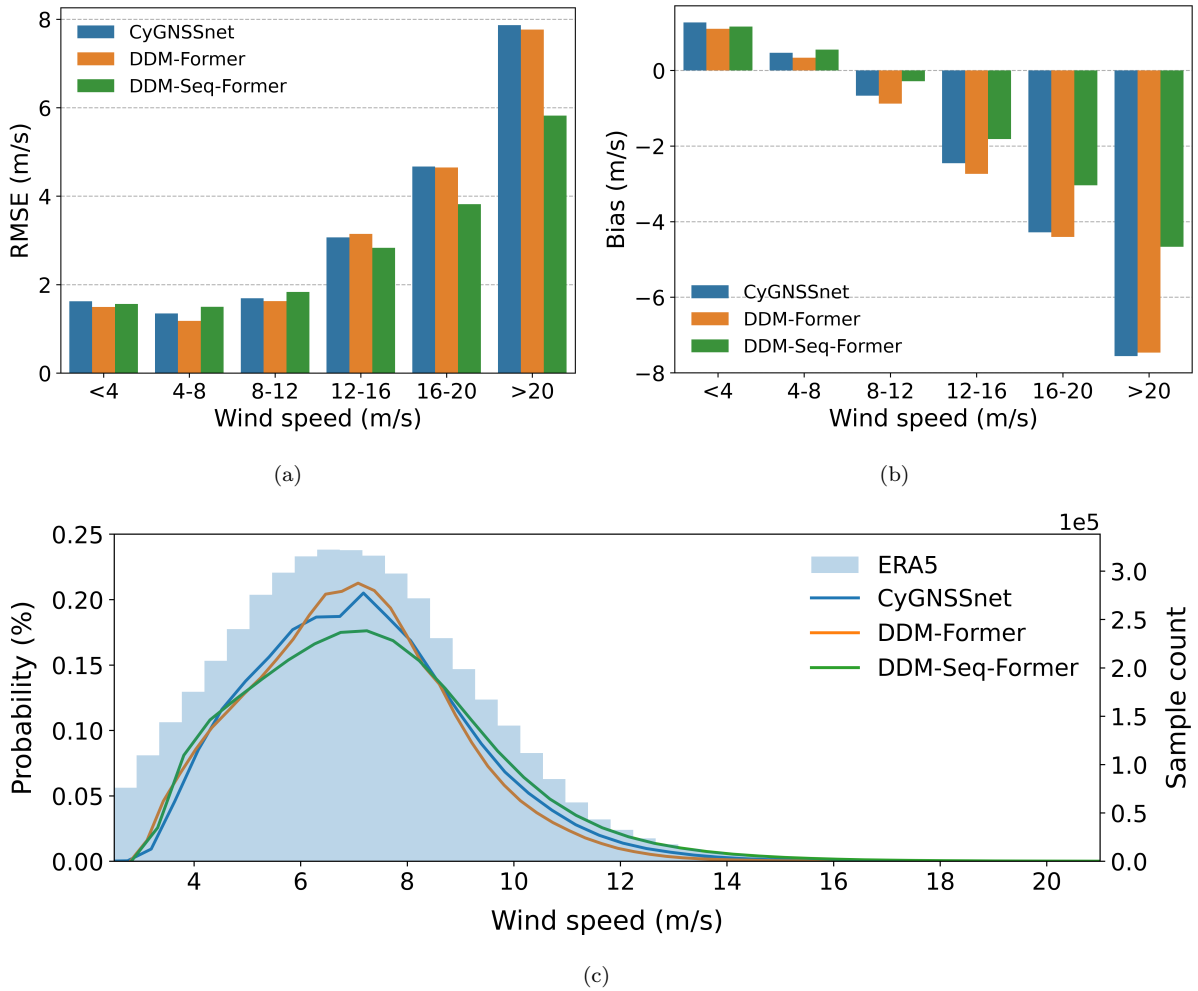


Figure 7: RMSE values for different ranges as a function of ERA5 wind speed (a), bias for different ranges as a function of ERA5 wind speed (b), and probability density functions of wind speeds from compared models with ERA5 sample distribution (c).

considerably lower errors in terms of both the RMSE and bias. As can be seen in Figure 7(c), the samples display a non-uniform distribution, with the majority of the data accounting for wind speeds ranging from 4 m/s to 10 m/s. This fact demonstrates that a slightly worse estimation in the major population will harm the overall performance under the condition of having imbalanced data. Further, the probability density function slope of DDM-Seq-Former shifts slightly toward higher wind speeds compared to the other models. Such an effect proves that imprecise estimations of high winds will be penalized more compared to their moderate neighbors in the sequential setup. In other words, the model tends to grant higher weights to extreme winds and can achieve enhanced estimations in such rare cases. In contrast to models trained with individual DDMs, DDM-Seq-Former mitigates the model's tendency to fit to moderate wind speed samples.

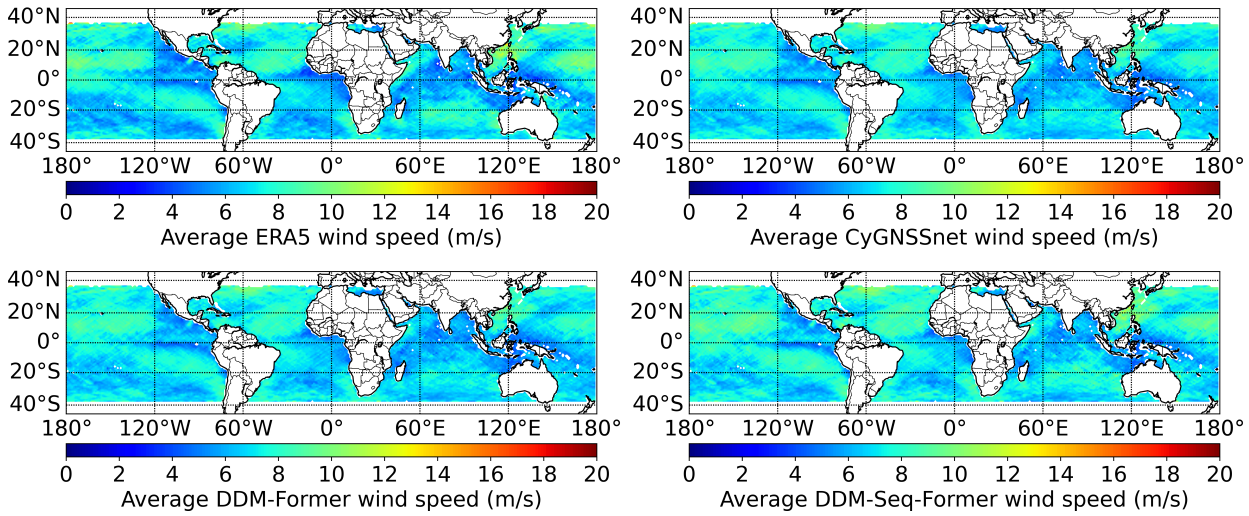
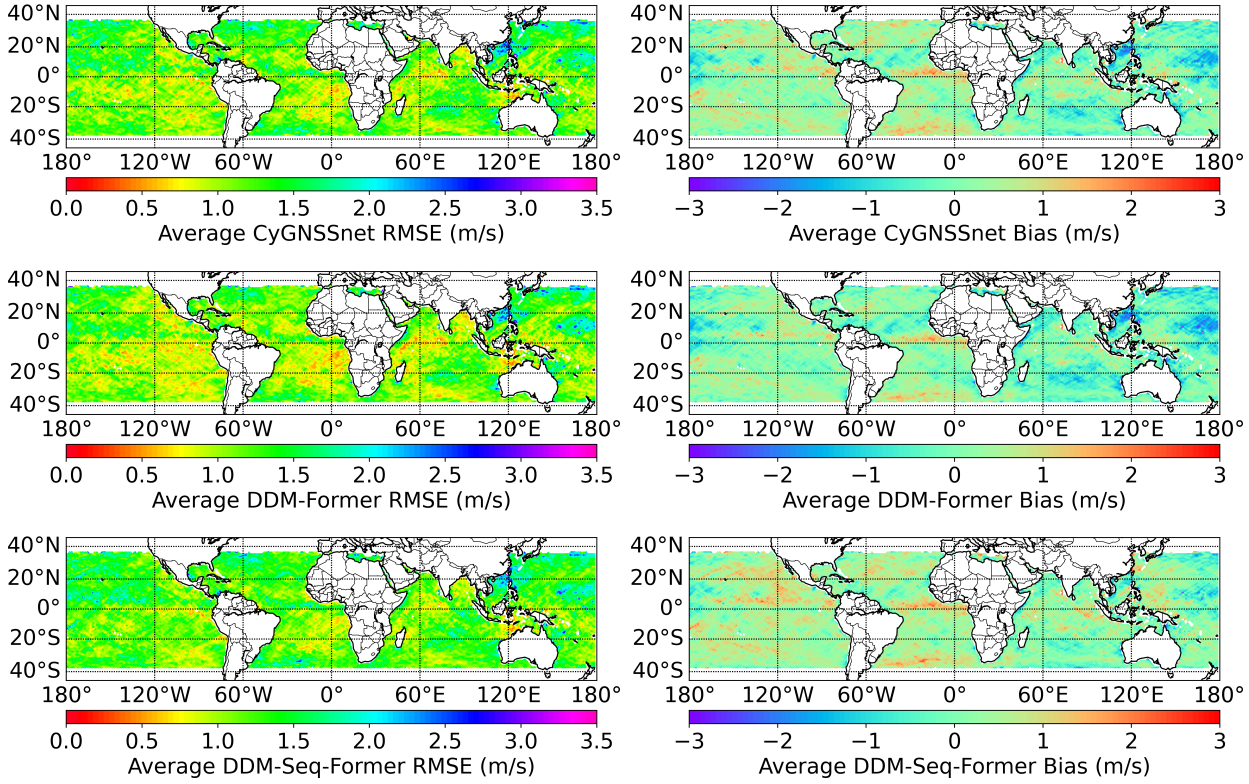


Figure 8: Geographical distributions of average ERA5 wind speeds as well as retrieved average wind speeds of CyGNSSnet, DDM-Former, and DDM-Seq-Former from July 2019 to April 2021.



(a) (b)

Figure 9: Average RMSE (a) and bias (b) spatial distributions of CyGNSSnet, DDM-Former, and DDM-Seq-Former compared with ERA5 true winds with a $1^\circ \times 1^\circ$ grid resolution on the test set.

5.2. Geographical analysis

Figure 8 shows the global distributions of the average wind speed estimations of CyGNSSnet, DDM-Former, and DDM-Seq-Former along with the ERA5 wind speed labels. When the wind grid is averaged with 1-degree resolution over the nine months test period, the averaged wind speeds in each grid are primarily lower than 14 m/s. A quantitative comparison in these wind speed maps demonstrates that the overall global trends agree well with the true winds. Both our models can reproduce the global wind speed distribution consistently. Notably, DDM-Seq-Former has fewer deviations for strong winds in the Pacific regions, whereas CyGNSSnet and DDM-Former display minor underestimations.

The global RMSE and bias distributions for DDM-Former, DDM-Seq-Former, and the baseline model against ERA5 winds are demonstrated in Figure 9. Again, both visualizations are calculated with a grid resolution of $1^\circ \times 1^\circ$. Generally, the RMSE distribution of DDM-Former holds more orange to yellow fields compared to the other models, which indicate small retrieval errors. An improved estimation of the wind speeds can be identified in the Atlantic, South Pacific, and Indian Ocean, where denser yellow areas are shown in DDM-Former's results. However, we still notice a performance degradation of CyGNSSnet and DDM-Former for strong winds near the South China Sea and the North Pacific. From August 2020 to November 2020, 19 tropical storms and typhoon events happened between 100°E and meridian 180°, with a maximum wind speed of above 60 m/s. With a relative dense coverage of strong winds in these regions, an RMSE spike is expected due to there not being enough training samples left after quality control, thus leading to inaccurate estimations. Other possible drivers of the spike in RMSE are a more complicated ocean dynamics close to land, and potential radio-frequency interference (RFI) caused by the Quasi-Zenith satellite system (QZSS) L-band signals [48]. Regarding the bias distribution, few geographical regions confront a deviation higher than 2.5 m/s. The overall bias of DDM-Former is evidently improved on a global scale. It can be again observed that both CyGNSSnet and DDM-Former suffer from the dilemma of underestimation in higher wind speed districts. On the other hand, the proposed DDM-Seq-Former overcomes the challenging strong wind estimation in most regions with a light blue to cyan color (bias of -1 to 0 m/s).

To specifically assess the different models' behaviors in strong wind regimes, RMSE maps of the Asia-Pacific region are exhibited in Figure 10. The ERA5 on the top-left hints at the average true winds in the given area. As discussed, deep learning-based models tend to favor overfitting to moderate winds and ignore minor cases in unevenly distributed data. Despite this commonly known issue, DDM-Seq-Former marginally inhibits failures in features with less blue and magenta distributions (RMSE of 2–3 m/s) over strong wind regions in comparison to the other methods. Although DDM observables' sensitivity is saturated at very high winds [49], and many extreme wind speed samples are therefore filtered out the during quality control, our model effectively learns the discrepant features between DDMs of low, medium, and high wind speed regimes.

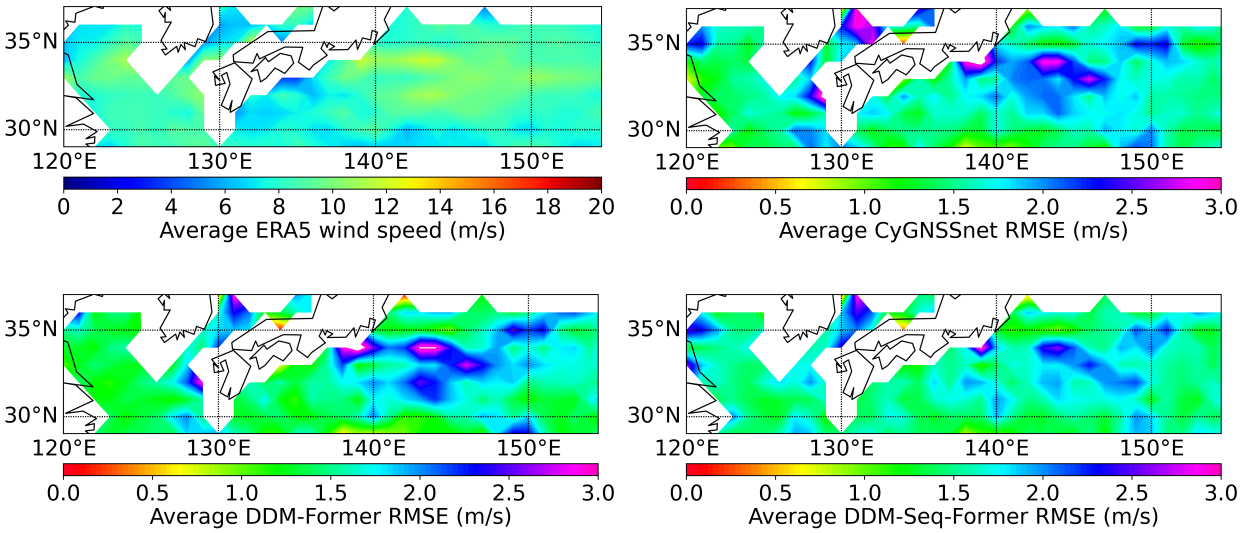
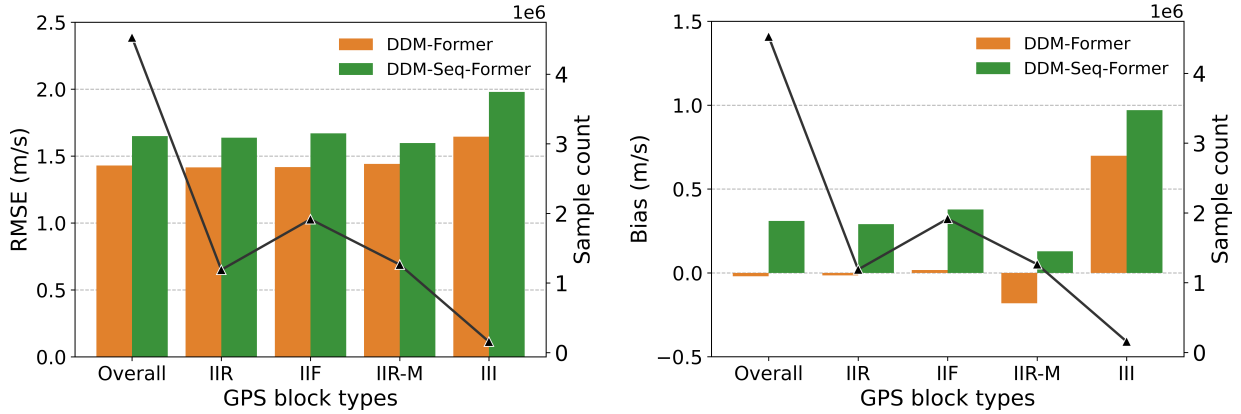


Figure 10: RMSE spatial distributions in strong wind regimes over the Asia-Pacific region with a 1-degree grid resolution. Average ERA5 wind speed map is shown for reference.



(a) (b)

Figure 11: RMSE and bias distribution over different GPS block types. A black curve depicts the sample count of each type.

5.3. Discussion and challenges

Even though the proposed models outperform the baseline network and conventional retrieval method in different aspects, it is worth briefly discussing the effects of the GPS block types, per-DDM spatial resolution, intersatellite calibration, strong wind performance, quality control, and their related challenges.

5.3.1. GPS block types

GPS is built in several blocks that transmit different levels of the effective isotropic radiated power (EIRP). The EIRP is also age dependent, as older satellites tend to have lower EIRP due to their energy loss over time. Sometimes this parameter is also affected by intentional changes at a specific time and area, known as "flex power". The EIRP variation is known as a technical source of error for GNSS-R wind speed retrievals. Figure

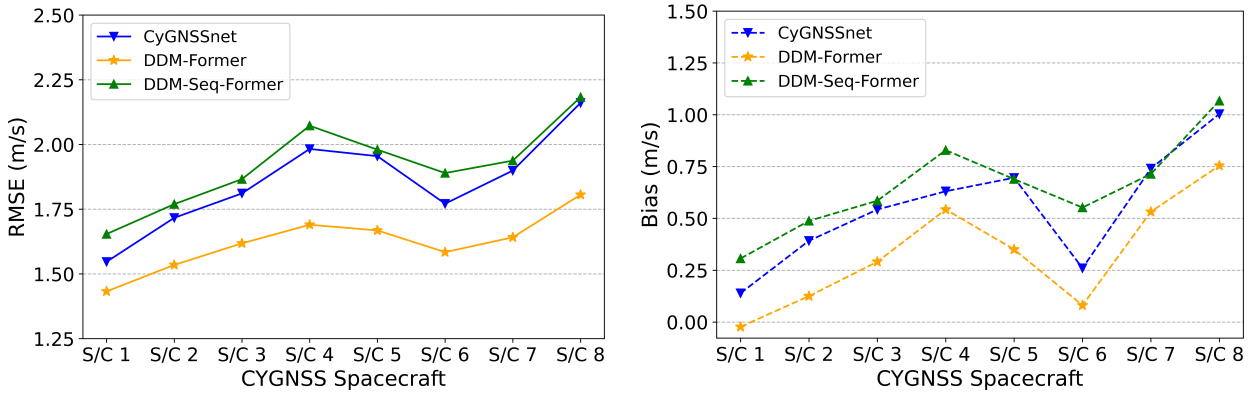


Figure 12: Overall RMSE and bias results for all CYGNSS spacecrafts. Models are trained with spacecraft 1 data only.

11 therefore investigates the RMSE and bias of each GPS block. Accordingly, the measurements from blocks IIR, IIF, and IIR-M are of comparable accuracies, whereas that of block III is significantly degraded. This difference is due to the lower number of observations from this block during the model training. However, future networks can incorporate the computed EIRP, available in CYGNSS L1 data, as an additional input as an immunity against such changes.

5.3.2. Per-DDM spatial resolution

The L2 CYGNSS spatial resolution is estimated at around $25 - 40$ km, depending on the incidence angle, in which the resolution is essentially bound by the 3×5 "box" set around the specular point DDM bin. In our study, the entire DDMs are used as inputs, and commenting on which bins have been used to extract features for each individual prediction is not an easy task. However, as the DDMs are used to train models to predict wind speeds from ERA5 estimations, it can be assumed that the predicted wind speeds have the same resolution, as the models are forced to predict the equivalent wind speeds using any extracted features from the DDMs.

5.3.3. Intersatellite calibration

Since all the models are trained with data from a single CYGNSS spacecraft for ensuring a fair comparison, potential intersatellite errors (e.g., sigma0 bias) may not be learned by the networks. We therefore examine the performance of these models with data other than those from spacecraft 1 over the same test period. The RMSE and bias results are depicted in Figure 12. Not surprisingly, we see a slight performance drop with the other spacecrafts compared with the spacecraft 1 model, which could be caused by any intersatellite deviations, including sigma0 biases. Still, the proposed DDM-Former generally works well with an RMSE lower than 2 m/s. Further studies can formulate an ensemble of the models trained with individual spacecraft data that could potentially enhance the intersatellite calibration.

5.3.4. Performance in strong wind regimes

Table 3: RMSE values of different models with wind speed samples over 20 m/s.

Spacecraft	S/C 1	S/C 2	S/C 3	S/C 4	S/C 5	S/C 6	S/C 7	S/C 8
Sample counts	59	6	9	32	211	149	8	99
CyGNSSnet	7.87	9.50	5.32	5.59	8.06	4.89	8.32	6.78
DDM-Seq-Former	**5.82**	**9.26**	**3.86**	**4.10**	**5.17**	**3.87**	**7.25**	**4.43**

Compared with moderate wind speed samples, high wind speed samples are much fewer among the total wind speed distribution. On the other hand, DDM observables' sensitivity is saturated at very high winds [49]. The challenging determination of the wind speed in strong wind regimes, combined with the limited amount of data available after quality control and the tendency for data-driven methods to regress to moderate wind speeds, results in performance degradation. To address this issue, the proposed DDM-Seq-Former takes DDM sequences as inputs and highlights the discrepant features between moderate and strong wind speeds. Table 3 shows the RMSE results of different models with wind speed samples over 20 m/s, with the samples coming from the nine months test period. Although there is still room to meet the CYGNSS requirement (RMSE $\leq$ 2 m/s), it is clear that the proposed method achieves a promising improvement with strong wind regimes.

5.3.5. Necessity of quality control

Table 4: Overall RMSE and bias results of different models using raw data without quality control.

Architecture	RMSE (m/s)	Bias (m/s)
CyGNSSnet	54.99	1.23
DDM-Former	8.24	-3.26
DDM-Seq-Former	5.10	-1.53

Since deep learning methods are data-driven, high-quality DDM samples are essential to give the models an opportunity for learning a correct intuition of how to map different DDMs to their corresponding wind speeds. Still, it is important to assess the tolerance of networks with poor samples. We therefore investigate the performance of all the methods with 10^6 CYGNSS unfiltered data from November 2020. Table 4 demonstrates that the models fail to learn the correct mapping from DDMs for wind speeds with contaminated measurements. Hence, quality control is a critical step for applying GNSS-R data to the task of wind speed retrieval and cannot be omitted.

6. Conclusion

In this paper, we introduced two Transformer-based models, DDM-Former and DDM-Seq-Former, for global ocean wind speed estimation. With low-cost GNSS-R constellation data, the proposed DDM-Former

outshines the conventional retrieval algorithm and CNN-based baseline network CyGNSSnet, achieving an average RMSE of 1.43 m/s and the lowest bias of −0.02 m/s in comparison to ERA5 wind speed labels. Evaluation statistics prove that our model can be well applied for the task of wind speed retrieval. In addition, by utilizing spatial-temporal coherent DDMs, the proposed DDM-Seq-Former shows strong improvements for extreme winds without additional data samples. Further, we demonstrate that Transformer-based models with MSA are indeed beneficial for exploiting underlying mappings from DDMs to the corresponding wind speeds.

Since both models belong to data-driven approaches without physical interpretation, we also consider model explainability in this work. To better understand how our models perceive the input DDMs and DDM sequences, an extensive investigation of intermediate features with attention visualizations is presented. For DDM-Former, we show it tends to pay attention to distinguishing between DDM background and the horseshoe-shaped signal in the middle, which exploits a finer delay-Doppler correlation, thus improving estimation performance; whereas DDM-Seq-Former is more interested in differentiating between distinct wind speeds and finding similarities between comparable wind speeds. For a fair comparison without changing the data sample distribution, DDM-Seq-Former achieves the desirable results with lower RMSE values in extreme wind regimes. These investigations clearly prove that both the proposed methods can learn their ways of aggregating critical features from the inputs independently.

Furthermore, there are still opportunities for further enhancements in creating more generalized models to improve estimation performance. With the increasing amount of constellations and data, GNSS-R incorporated with deep learning could be further developed to generate enhanced products. In addition, because the proposed models have an end-to-end characteristic that requires no feature engineering, a direct migration to other applications with DDMs as inputs has great potential. Considering the features of DDM signals are abundant and alter dramatically over different Earth surfaces, it is expected that the same process could be applied to sense soil moisture, inland water bodies, or sea ice globally.

Acknowledgements

This work is jointly supported by the Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren e.V. project “Artificial Intelligence for GNSS Reflectometry: Novel Remote Sensing of Ocean and Atmosphere (AI4GNSSR)” [grant number: ZT-I-PF-5-091], by the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab “AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond” [grant number: 01DD20001] and by the German Federal Ministry of Economics and Technology in the framework of the “national center of excellence ML4Earth” [grant number: 50EE2201C]. CA was funded by Helmholtz Association’s Initiative and Networking Fund through Helmholtz AI [grant number: ZT-I-PF-5-01]. This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project ID AIM and was supported by Helmholtz AI computing resources (HAICORE) of the Helmholtz Association’s Ini-

tiative and Networking Fund through Helmholtz AI. The datasets used in this study are available free of charge and we thank the scientific teams associated with the CYGNSS mission at NASA and the University of Michigan, and ERA5 reanalysis estimates at the European Centre for Medium-Range Weather Forecasts (ECMWF).

References

[1] H. Dankert, J. Horstmann, W. Rosenthal, Ocean wind fields retrieved from radar-image sequences, Journal of Geophysical Research: Oceans 108 (11) (2003) 2150–2152. doi:10.1029/2003JC002056.

[2] E. Rodríguez, M. Bourassa, D. Chelton, J. Thomas Farrar, D. Long, D. Perkovic-Martin, R. Samelson, The Winds and Currents Mission Concept, Frontiers in Marine Science 6 (JUL) (2019) 1–8. doi:10.3389/FMARS.2019.00438.

[3] M. P. Clarizia, C. S. Ruf, P. Jales, C. Gommenginger, Spaceborne GNSS-R Minimum Variance Wind Speed Estimator, IEEE Transactions on Geoscience and Remote Sensing 52 (11) (2014) 6829–6843. doi:10.1109/TGRS.2014.2303831.

[4] C. W. Zheng, C. Y. Li, J. Pan, M. Y. Liu, L. L. Xia, An overview of global ocean wind energy resource evaluations, Renewable and Sustainable Energy Reviews 53 (2016) 1240–1251. doi:10.1016/J.RSER.2015.09.063.

[5] M. Martin-Neira, M. Caparrini, J. Font-Rossello, S. Lannelongue, C. S. Vallmitjana, The PARIS concept: an experimental demonstration of sea surface altimetry using GPS reflected signals, IEEE Transactions on Geoscience and Remote Sensing 39 (1) (2001) 142–150. doi:10.1109/36.898676.

[6] J. L. Garrison, S. J. Katzberg, M. I. Hill, Effect of sea roughness on bistatically scattered range coded signals from the Global Positioning System, Geophysical Research Letters 25 (13) (1998) 2257–2260. doi:10.1029/98GL51615.

[7] V. U. Zavorotny, A. G. Voronovich, Scattering of GPS signals from the ocean with wind remote sensing application, IEEE Transactions on Geoscience and Remote Sensing 38 (2) (2000) 951–964. doi:10.1109/36.841977.

[8] J. L. Garrison, A. Komjathy, V. U. Zavorotny, S. J. Katzberg, Wind speed measurement using forward scattered GPS signals, IEEE Transactions on Geoscience and Remote Sensing 40 (1) (2002) 50–65. doi:10.1109/36.981349.

[9] V. U. Zavorotny, S. Gleason, E. Cardellach, A. Camps, Tutorial on Remote Sensing Using GNSS Bistatic Radar of Opportunity, IEEE Geoscience and Remote Sensing Magazine 2 (4) (2014) 8–45. doi:10.1109/MGRS.2014.2374220.

[10] G. Foti, C. Gommenginger, P. Jales, M. Unwin, A. Shaw, C. Robertson, J. Roselló, Spaceborne GNSS reflectometry for ocean winds: First results from the UK TechDemoSat-1 mission, Geophysical Research Letters 42 (13) (2015) 5435–5441. doi:10.1002/2015GL064204.

[11] C. S. Ruf, C. Chew, T. Lang, M. G. Morris, K. Nave, A. Ridley, R. Balasubramaniam, A New Paradigm in Earth Environmental Monitoring with the CYGNSS Small Satellite Constellation, Scientific Reports 8 (1) (2018) 1–13. doi:10.1038/S41598-018-27127-4.

[12] C. Jing, X. Niu, C. Duan, F. Lu, G. Di, X. Yang, Sea Surface Wind Speed Retrieval from the First Chinese GNSS-R Mission: Technique and Preliminary Results, Remote Sensing 11 (24) (2019) 3013. doi:10.3390/RS11243013.

[13] Y. Liu, I. Collett, Y. J. Morton, Application of Neural Network to GNSS-R Wind Speed Retrieval, IEEE Transactions on Geoscience and Remote Sensing 57 (12) (2019) 9756–9766. doi:10.1109/TGRS.2019.2929002.

[14] N. Rodriguez-Alvarez, D. M. Akos, V. U. Zavorotny, J. A. Smith, A. Camps, C. W. Fairall, Airborne GNSS-R Wind Retrievals Using Delay-Doppler Maps, IEEE Transactions on Geoscience and Remote Sensing 51 (1) (2013) 626–641. doi:10.1109/TGRS.2012.2196437.

[15] M. P. Clarizia, C. S. Ruf, Wind Speed Retrieval Algorithm for the Cyclone Global Navigation Satellite System (CYGNSS) Mission, IEEE Transactions on Geoscience and Remote Sensing 54 (8) (2016) 4419–4432. doi:10.1109/TGRS.2016.2541343.

[16] C. S. Ruf, R. Balasubramaniam, Development of the CYGNSS Geophysical Model Function for Wind Speed, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12 (1) (2019) 66–77. doi:10.1109/JSTARS.2018.2833075.

[17] X. X. Zhu, D. Tuia, L. Mou, G. S. Xia, L. Zhang, F. Xu, F. Fraundorfer, Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources, IEEE Geoscience and Remote Sensing Magazine 5 (4) (2017) 8–36. doi:10.1109/MGRS.2017.2762307.

[18] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, B. A. Johnson, Deep learning in remote sensing applications: A meta-analysis and review, ISPRS Journal of Photogrammetry and Remote Sensing 152 (2019) 166–177. doi:10.1016/J.ISPRSJPRS.2019.04.015.

[19] M. Asgarimehr, I. Zhelavskaya, G. Foti, S. Reich, J. Wickert, A GNSS-R Geophysical Model Function: Machine Learning for Wind Speed Retrievals, IEEE Geoscience and Remote Sensing Letters 17 (8) (2020) 1333–1337. doi:10.1109/LGRS.2019.2948566.

[20] J. F. Munoz-Martin, A. Camps, Sea Surface Salinity and Wind Speed Retrievals Using GNSS-R and L-Band Microwave Radiometry Data from FMPL-2 Onboard the FSSCat Mission, Remote Sensing 13 (16) (2021) 3224. doi:10.3390/RS13163224.

[21] J. Reynolds, M. P. Clarizia, E. Santi, Wind Speed Estimation From CYGNSS Using Artificial Neural Networks, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13 (2020) 708–716. doi:10.1109/JSTARS.2020.2968156.

[22] X. Li, D. Yang, J. Yang, G. Zheng, G. Han, Y. Nan, W. Li, Analysis of coastal wind speed retrieval from CYGNSS mission using artificial neural network, Remote Sensing of Environment 260 (2021) 112454. doi:10.1016/J.RSE.2021.112454.

[23] X. Liu, W. Bai, J. Xia, F. Huang, C. Yin, Y. Sun, Q. Du, X. Meng, C. Liu, P. Hu, G. Tan, FA-RDN: A Hybrid Neural Network on GNSS-R Sea Surface Wind Speed Retrieval, Remote Sensing 13 (23) (2021) 4820. doi:10.3390/RS13234820.

[24] X. Chu, J. He, H. Song, Y. Qi, Y. Sun, W. Bai, W. Li, Q. 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 13 (2020) 5971–5981. doi:10.1109/JSTARS.2020.3010879.

[25] M. Asgarimehr, C. Arnold, T. Weigel, C. Ruf, J. Wickert, GNSS reflectometry global ocean wind speed using deep learning: Development and assessment of CyGNSSnet, Remote Sensing of Environment 269 (2022) 112801. doi:10.1016/J.RSE.2021.112801.

[26] W. Guo, H. Du, C. Guo, B. J. Southwell, J. W. Cheong, A. G. Dempster, Information fusion for GNSS-R wind speed retrieval using statistically modified convolutional neural network, Remote Sensing of Environment 272 (2022) 112934. doi:10.1016/J.RSE.2022.112934.

[27] T. M. Roberts, I. Colwell, C. Chew, S. Lowe, R. Shah, A Deep-Learning Approach to Soil Moisture Estimation with GNSS-R, Remote Sensing 14 (14) (2022) 3299. doi:10.3390/RS14143299.

[28] M. M. Nabi, V. Senyurek, A. C. Gurbuz, M. Kurum, Deep Learning-Based Soil Moisture Retrieval in CONUS Using CYGNSS Delay–Doppler Maps, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 15 (2022) 6867–6881. doi:10.1109/JSTARS.2022.3196658.

[29] Q. Yan, W. Huang, Sea Ice Sensing From GNSS-R Data Using Convolutional Neural Networks, IEEE Geoscience and Remote Sensing Letters 15 (10) (2018) 1510–1514. doi:10.1109/LGRS.2018.2852143.

[30] H. Xie, S. He, X. Cheng, A Convolution Neural Network-based Method for Sea Ice Remote Sensing using GNSS-R Data, in: 2022 4th International Conference on Communications, Information System and Computer Engineering, 2022, pp. 284–289. doi:10.1109/CISCE55963.2022.9851159.

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, 2017, pp. 5998–6008.

[32] 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, in: 9th International Conference on Learning Representations, 2021.

[33] 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, 2021.

[34] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-End Object Detection with Transformers, in: 16th European Conference on Computer Vision, Vol. 12346, 2020, pp. 213–229. doi:10.1007/978-3-030-58452-8_13.

[35] S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, T. Xiang, P. H. Torr, L. Zhang, Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 6877–6886. doi:10.1109/CVPR46437.2021.00681.

[36] L. Zhou, Y. Zhou, J. J. Corso, R. Socher, C. Xiong, End-to-End Dense Video Captioning with Masked Transformer, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8739–8748. doi:10.1109/CVPR.2018.00911.

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

[38] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, J. N. Thépaut, The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society 146 (730) (2020) 1999–2049. doi:10.1002/QJ.3803.

[39] M. Asgarimehr, V. Zavorotny, J. Wickert, S. Reich, Can GNSS Reflectometry Detect Precipitation Over Oceans?, Geophysical Research Letters 45 (22) (2018) 12585–12592. doi:10.1029/2018GL079708.

[40] S. Jin, E. Cardellach, F. Xie, GNSS Remote Sensing: Theory, Methods and Applications, Springer Netherlands, Dordrecht, 2014. doi:10.1007/978-94-007-7482-7.

[41] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research 15 (56) (2014) 1929–1958. doi:10.5555/2627435.2670313.

[42] J. L. Ba, J. R. Kiros, G. E. Hinton, Layer Normalization, 2016. arXiv:http://arxiv.org/abs/1607.06450v1.

[43] Q. Wang, B. Li, T. Xiao, J. Zhu, C. Li, D. F. Wong, L. S. Chao, Learning Deep Transformer Models for Machine Translation, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 1810–1822. doi:10.18653/V1/P19-1176.

[44] A. Baevski, M. Auli, Adaptive Input Representations for Neural Language Modeling, in: 7th International Conference on Learning Representations, 2019.

[45] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, Software available from tensorflow.org (2015).

[46] I. Loshchilov, F. Hutter, Decoupled Weight Decay Regularization, in: 7th International Conference on Learning Representations, 2019.

[47] D. Hendrycks, K. Gimpel, Gaussian Error Linear Units (GELUs), 2016. arXiv:http://arxiv.org/abs/1606.08415v3.

[48] M. Asgarimehr, J. Wickert, S. Reich, TDS-1 GNSS Reflectometry: Development and Validation of Forward Scattering Winds, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11 (11) (2018) 4534–4541. doi:10.1109/JSTARS.2018.2873241.

[49] C. S. Ruf, S. Gleason, D. S. McKague, Assessment of CYGNSS Wind Speed Retrieval Uncertainty, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12 (1) (2019) 87–97. doi:10.1109/JSTARS.2018.2825948.

List of Figures