

Ocean Sea and Ice SAF

OSI_VSA24_01

**ON THE USE OF MACHINE LEARNING TO CORRECT NWP MODEL
SEA SURFACE WIND FORECASTS WITH SCATTEROMETER
(ERASTAR_AI2)**

Evgeniia Makarova¹
Marcos Portabella¹
Ad Stoffelen²

¹ Barcelona Expert Center (BEC ICM-CSIC)

² Koninklijk Nederlands Meteorologisch Instituut (KNMI)

Visiting Scientist Activity
Technical Report

9 February 2026

Acronyms

ASCAT Advanced Scatterometer. 1–4, 7, 9–18, 20

AWDP ASCAT Wind Data Processor. 3, 5

BUFR Binary Universal Form for the Representation of meteorological data. 5

CMEMS Copernicus Marine Environment Monitoring Service. 1

CNN Convolutional Neural Network. 18, 19

ECMWF European Centre for Medium-Range Weather Forecasts. 1, 3, 20

ERA5 ECMWF Reanalysis v5. 1–5, 7–12, 14–18, 20

ESA European Space Agency. 1

FCNN Fully-connected Feed-forward Neural Network. 1, 2, 6, 20

GBDT Gradient Boosting Decision Trees. 1

genscat generic scatterometer. 5

GLOBCURRENT Global Total, Ekman and Geostrophic currents at the Surface and 15m. 3, 5, 8

GNN Graph Neural Network. 20

GPU graphics processing unit. 5

GRIB General Regularly distributed Information in Binary form. 8

HSCAT Haiyang scatterometer. 1, 2, 4, 7, 9–12, 15–18, 20

HY-2B Haiyang-2B. 9

HY-2C Haiyang-2C. 9

IFS Integrated Forecasting System. 1, 20

ITCZ Intertropical Convergence Zone. 1, 11, 12, 15

LightGBM Light Gradient Boosting Machine. 1

MAE Mean Absolute Error. 19

MetOp Meteorological Operational satellite. 3, 9

ML Machine Learning. 1, 2

NetCDF Network Common Data Form. 5, 8

NN neural network. 3, 6, 12–20

NWP Numerical Weather Prediction. 1, 2, 20

OSI SAF EUMETSAT Ocean and Sea Ice Satellite Application Facility. 1, 3

PenWP Pencil beam Wind Processor. 4

QC quality control. 9–11

RMSE Root Mean Square Error. 6, 19

SST sea surface temperature. 1, 3–5, 8, 21, 22

TCZ Tropical Convergence Zone. 2

U10S stress-equivalent 10-meter wind. 2–5, 8, 10, 11, 18, 20–22

VRMS Vector Root Mean Square difference. 9, 10, 12, 15, 16, 19

WIND TAC Surface Wind Thematic Assembly Center. 1

WOC World Ocean Circulation. 1

XGBoost eXtreme Gradient Boosting. 1

Contents

1	Introduction	1
2	Datasets	3
2.1	Data coverage	3
2.2	ML model inputs	3
2.3	Training Targets	3
2.4	Datasets used for evaluation	4
3	Methodology	6
3.1	Data preprocessing	6
3.2	Model training	6
3.3	NN configurations and hyperparameters	7
3.4	Monthly and Yearly Models	7
3.5	Inference	9
3.6	Validation	10
4	Results and discussion	11
4.1	Differences between ERA5 biases	11
4.2	All-year and monthly models	12
4.3	Monthly statistics	17
4.4	Spatial Variance of Corrected Outputs	20
5	Conclusions and future work	22
A	Appendix	24

Abstract

Accurate ocean surface winds are essential for ocean modeling, yet widely used reanalyses such as ECMWF Reanalysis v5 (ERA5) exhibit geographically structured and persistent biases. This study evaluates the use of Machine Learning (ML) techniques to correct persistent stress-equivalent wind biases in ERA5 forecasts, aiming to improve the accuracy of Numerical Weather Prediction (NWP) outputs without relying on scatterometer observations during inference.

Persistent biases in ERA5 winds have been widely documented and are linked to unresolved air–sea interactions, sea surface temperature (SST) gradient effects, and the influence of ocean currents. Existing bias correction methods, such as ERA* and Copernicus Marine Environment Monitoring Service (CMEMS) Surface Wind Thematic Assembly Center (WIND TAC) products, depend on the availability of scatterometer data, which limits their applicability in forecasting and during periods of sparse or no scatterometer coverage.

To address these limitations, Fully-connected Feed-forward Neural Networks (FCNNs) were trained on five years of ERA5 atmospheric and oceanic variables, combined with surface current data from CMEMS, to predict wind bias corrections. The models were trained using Advanced Scatterometer (ASCAT) scatterometer observations as reference and validated against both C-band (ASCAT-B/C) and Ku-band (HSCAT-B/C) instruments for an independent test year (2023).

Results demonstrate that ML-based corrections significantly reduce ERA5 wind biases while preserving the spatial variance of wind fields. On a global scale, error variance reductions reached approximately 13% against ASCAT and 9% against HSCAT, with the highest improvements (up to 16%) observed in extra-tropical regions. Seasonal models outperformed all-year models, particularly in the tropics, where seasonal variability strongly influences bias patterns. Although residual discrepancies remain when validated against Ku-band scatterometers, the corrected outputs substantially mitigate large-scale bias structures.

1 Introduction

Numerical Weather Prediction (NWP) sea-surface wind outputs are commonly used in ocean modeling, climate research, and operational marine applications. These models are being continuously improved. However, they tend to exhibit recurrent and persistent biases compared to the observational data.

The European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) is assimilating the in-situ and satellite wind measurements. However, when compared to scatterometer data, after the assimilation of the scatterometer winds the biases quickly reappear.

Several studies have reported systematic biases in ECMWF surface winds, particularly in terms of wind direction. Sandu et al. [1] confirmed systematic rotation of surface winds, clockwise in the Northern Hemisphere and counter-clockwise in the Southern Hemisphere, when compared with scatterometer data. Belmonte Rivas and Stoffelen [2] conducted a comparative analysis of the ECMWF Reanalysis v5 (ERA5) and its predecessor ERA-Interim, revealing that local wind vector biases and wind variability biases in ERA are remarkably persistent over time. Their study highlights several characteristic patterns, including excessive westerlies in the mid-latitudes and easterlies in the tropics, and insufficient meridional convergence toward the Intertropical Convergence Zone (ITCZ). These biases are linked to a range of factors such as unresolved air–sea interaction processes including the influence of sea surface temperature (SST) gradients and ocean currents.

To address these issues, the ERA* dataset [3] was developed within the European Space Agency (ESA) World Ocean Circulation (WOC) project. This product applies temporally averaged corrections derived from scatterometer and model wind differences to ERA5 fields, thereby reducing local persistent biases [4]. However, the performance of ERA* strongly depends on the sampling and the number of the scatterometers operational during each time period [4, 5].

The Copernicus Marine Environment Monitoring Service (CMEMS) Surface Wind Thematic Assembly Center (WIND TAC) Level 4 (L4) sea surface wind products are based on a similar approach to that of ERA* to correct the biases both of the ECMWF operational forecasts and ERA5 (available through the Copernicus Marine Service [6, 7]). These products use the Advanced Scatterometer (ASCAT) – NWP outputs differences accumulated over 20 days with the centered time window for the reanalysis and a time window shifted backwards in time for the operational version.

To overcome the limitations caused by the insufficient spatial and temporal coverage of scatterometer observations, an alternative approach was proposed in the previous EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI SAF) study OSI_VSA22.01 [8]. That study focused on the use of Machine Learning (ML) models trained to predict wind stress biases using other atmospheric and ocean-related variables available directly from the NWP model and using scatterometer wind vector fields as reference. The goal of this approach is to enable the learning of systematic relationships between physical parameters and the resulting ECMWF wind errors, to allow bias correction of the forecast outputs, when the satellite observations are not available.

The study assessed the feasibility of this ML-based correction method using a reduced dataset covering the period January to March 2020. Two families of ML algorithms were evaluated: Gradient Boosting Decision Trees (GBDT), including eXtreme Gradient Boosting (XGBoost) and Light

Gradient Boosting Machine (LightGBM), and simple Fully-connected Feed-forward Neural Networks (FCNNs). The models were trained to predict the biases in both the zonal and meridional components of the ERA5 stress-equivalent 10-meter wind (U10S). Results demonstrated that both model families were capable of reducing wind stress bias for the test subset when validated against ASCAT-A, also used to derive the target biases on which the models were trained, and against the independent HSCAT-B scatterometer. The best-performing configurations achieved error variance reductions of up to 10.5% globally against ASCAT-A w.r.t ERA5 and 5.5% against HSCAT-B.

Further regional analysis revealed that the models performed best in the extra-tropical regions, achieving an error variance reduction of approximately 13% when validated against ASCAT-A, and 10.6% when validated against HSCAT-B [9]. In contrast, performance in the Tropics varied significantly depending on the scatterometer used for validation. When compared with ASCAT-A, the models showed up to a 6.3% reduction in error variance. However, validation against HSCAT-B indicated little to no improvement, and in some cases, even a degradation in performance. This was particularly pronounced in the Tropical Convergence Zone (TCZ), where extensive quality control filtering was applied due to rain-induced contamination affecting the Ku-band scatterometer measurements [10].

While the initial results were promising, the limited temporal extent of the training dataset imposed constraints on the model’s ability to generalize across different seasons and meteorological conditions. The three-month training period, focused on early 2020, captured a narrow range of atmospheric regimes, which limited the representativeness of the training samples. As shown in [9], model performance tended to be better during months that were included in the training period, while results degraded for months outside of that window. This seasonal sensitivity suggests that the model was learning patterns specific to the training conditions. These findings highlighted the importance of extending the training period to better account for seasonal variability and to improve the model’s ability to generalize.

To overcome the limitations of the initial study, the current work extends the training dataset to five years to include a broader range of atmospheric and oceanic conditions. The goal is to improve the robustness and generalization of the ML models by capturing seasonal variability and enhancing performance across different regions and time periods.

The following sections of this report provide a detailed overview of the study. The Dataset section 2 describes the scatterometer ground truth data and the input NWP model data used for training and inference of the ML models, as well as for subsequent validation. The Methodology section 3 outlines the data preprocessing steps and presents an overview of the neural network architectures and training strategies. These approaches are initially tested on a smaller subset of data before applying the selected configuration to a five-year dataset. The Results section 4 presents the validation of the ML model performance against several scatterometers available during the test period and discusses both the strengths and limitations of the ML-based approach. Finally, the Conclusions section 5 summarizes the key findings and provides an outlook on future work.

2 Datasets

2.1 Data coverage

To ensure the dataset covers a broad range of meteorological and oceanic conditions, five years of data (2019–2022 and 2024) are used for training. The year 2018 serves as a validation period for hyperparameter tuning and model selection, while 2023 is reserved as an independent test set to evaluate the performance of the ML models.

2.2 ML model inputs

The ML model inputs include atmospheric and oceanic variables obtained from two datasets: the ECMWF ERA5 reanalysis [11] and the Copernicus Globa Total, Ekman and Geostrophic currents at the Surface and 15m (GLOBCURRENT) [12, 13].

The ERA5 dataset used in this work originates from the full dataset on a reduced Gaussian grid, containing two analyses per day, at 6:00 and 18:00 UTC. The forecast steps include a range from +3 to +18 hours with hourly temporal resolution. The variables that are obtained from ERA5 include eastward (u10s) and northward (v10s) stress-equivalent wind components derived from 10-meter neutral winds according to Equation 1 [14], specific humidity (q), surface temperature (t), mean sea level pressure (msl), and sea surface temperature (SST).

$$U10S = U10N \sqrt{\frac{\rho_{air}}{< \rho_{air} >}} \quad (1)$$

The Copernicus Global Total Surface Current dataset is used to obtain the eastward (uo) and northward (vo) components of sea water velocity at surface level. Daily averages centered at 12:00 UTC are used, with a spatial resolution of 0.25°.

Several additional input variables are calculated such as stress-equivalent wind speed and direction, vorticity and divergence and SST gradients (sst_dx, sst_dy). An overview of the input variables is provided in Table 1.

It is important to note that any inference based on these inputs may propagate uncertainties present in the original datasets. While explicit error estimates for all variables are beyond the scope of this work, wind component uncertainties have been documented in previous studies [15], and GLOBCURRENT uncertainties are known to be considerable in relative terms [16]. These limitations should be kept in mind when interpreting the results.

2.3 Training Targets

The neural network models are trained to predict the differences between scatterometer observations and ERA5 U10S for both the eastward (u) and northward (v) components. The scatterometer data used to derive these differences were obtained from the ASCAT instruments onboard the Meteorological Operational satellite (MetOp) series. The EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI SAF) Level 2 (L2, swath-gridded) 12.5 km data are further reprocessed using the ASCAT Wind Data Processor (AWDP) [17], incorporating ERA5 U10S as background.

Table 1: Input variables and their corresponding source datasets.

Source dataset	Variable	Units
-	Longitude	°
-	Latitude	°
ERA5	Eastward U10S component	m/s
ERA5	Northward U10S component	m/s
ERA5	U10S speed	m/s
ERA5	U10S direction	°
Derived from ERA5 U10S	U10S curl	N/m ³
Derived from ERA5 U10S	U10S divergence	N/m ³
ERA5	Mean sea level pressure	Pa
ERA5	Surface air temperature	K
ERA5	Surface specific humidity	kg/kg
ERA5	Sea surface temperature	K
Derived from ERA5 SST	Eastward SST gradient	K/m
Derived from ERA5 SST	Northward SST gradient	K/m
GLOBCURRENT	Eastward sea water velocity	m/s
GLOBCURRENT	Northward sea water velocity	m/s

Table 2 presents an overview of the ASCAT scatterometer data available for each year during the training, validation, and test periods.

Table 2: ASCAT data availability.

ASCAT-A, ASCAT-B, ASCAT-C	January 2018 – November 2021
ASCAT-B, ASCAT-C	November 2021 - December 2024

2.4 Datasets used for evaluation

While the neural network models are trained exclusively on ASCAT data, the validation for the test period (2023) is performed using both C-band scatterometers, such as ASCAT-B and ASCAT-C and Ku-band scatterometers, such as HSCAT-B and HSCAT-C onboard Hai Yang 2B (HY-2B) and 2C (HY-2C). The Ku-band scatterometer data are reprocessed using the Pencil beam Wind Processor (PenWP) [18], performing the inversion with ERA5 U10S as background.

Note that AWDP and PenWP are co-developed using plentiful close collocations between both products [10], reassuring similar wind vector field characteristics. However, both sensitivity to heavy rain and limited measurement azimuth diversity hamper Ku-band products to achieve ASCAT wind vector quality [15]. One compelling difference exists in the representation of wind gradients near deep moist convection, such as in the ITCZ [19]. It is anticipated that these have potentially a large influence on the representation of tropical air-sea interaction [20]. Where ASCAT validates

well with tropical winds, both HY2 and ECMWF have a weaker representation of the wind fields associated with deep moist convection.

We finally note that ERA5 winds above 15 m/s are weaker than those from the operational ECMWF model, from scatterometers and those from moored buoys. Potentially such biases are corrected by the approach followed here, while the effect on the validation sets might be mixed due to the different QC applied.

3 Methodology

3.1 Data preprocessing

The first step in data preprocessing involves the collocation in space and time of scatterometer data with the input variables obtained from ERA5 and GLOBCURRENT, as well as the calculation of the additional input variables, such as U10S derivatives and SST gradients, described in Section 2.2.

The collocation of scatterometer data with ERA5 U10S is performed during reprocessing using the AWDP, which includes wind inversion. To facilitate further processing, scatterometer data are converted from Binary Universal Form for the Representation of meteorological data (BUFR) to Network Common Data Form (NetCDF) format using the generic scatterometer (genscat) bufr2nc tools [17]. During this conversion, vorticity and divergence for the collocated ERA5 stress-equivalent winds are also calculated, as described in [21].

Additional collocations are carried out using a standalone tool developed in Python, which utilizes the LinearNDInterpolator library from SciPy [22]. For variables with hourly resolution, temporal interpolation is applied following the approach used in the AWDP processor, i.e., using two forecast steps preceding and one following the measurement time. For SST, the analysis file closest to each forecast time is selected, as SST data is only available every 12 hours. Note, that the SST data present in the ECMWF GRIB files at 6:00 UTC and 18:00 UTC (analysis times) do not present significant differences resulting in an unresolved diurnal SST cycle, which affects the performance of the neural network. For mean daily currents, data corresponding to the selected day are used.

To reduce training time across five years of data, the training and validation datasets are subsampled. The models presented in this report are trained using 15% of the original dataset. Subsampling of the training data is conducted in an ordered fashion: considering the relatively narrow swath width, which results in a two-dimensional matrix of 82 columns and a large number of rows, the matrix is first flattened (i.e., converted) into a one-dimensional array by columns rather than by rows. The validation subset is randomly subsampled using a fixed random seed (123) at a rate of 5%.

The next step involves normalization of the data. The type of normalization applied depends on the variable. Coordinates and directional variables are normalized using sine and cosine transformations, while other variables are normalized using their respective means and standard deviations. Details on the normalization types, means, and standard deviations used for each variable are provided in Appendix Table 1.

Finally, the pairs of normalized inputs and targets are converted into Torch dataset format and stored in 2 GB files.

3.2 Model training

The models are trained on NVIDIA Tesla v100-SXM2 32GB GPU. The number of training epochs varies between 20 and 40, depending on the size of the dataset. Training and validation metrics are recorded at each epoch using the Weights & Biases tracking service. All seasonal models

discussed in the Results section are trained for a fixed duration of 30 epochs.

3.3 NN configurations and hyperparameters

The optimization of the model hyperparameters is performed on a neural network trained exclusively on January data from the five-year training dataset.

The neural network architecture is based on a Fully-connected Feed-forward Neural Network (FCNN), with normalization implemented via Dropout layers between hidden layers and regularization through weight decay. The Root Mean Square Error (RMSE) loss function is used for both training and validation, and the Adam optimizer is applied with a fixed learning rate of 10^{-4} .

Table 3 presents the best validation loss achieved during training over 30 epochs for a range of network configurations, including variations in the number of hidden layers, the number of perceptrons per layer, and regularization hyperparameters such as dropout rate and weight decay.

Table 3: Evaluated neural network configurations with corresponding training and best validation losses.

Hidden layers	Number of parameters	Dropout	Weight Decay	Train Loss	Validation Loss
256-128-64-32	48162	0.25	1.E-05	1.158	1.162
512-256-128-64	182338	0.25	5.E-05	1.143	1.151
512-256-128-64-32	184354	0.25	1.E-05	1.142	1.151
1024-512-256-64-16	693106	0.15	5.E-05	1.120	1.143
1024-512-256-128-64	716866	0.25	5.E-05	1.124	1.144
1024-512-256-128-64	716866	0.15	5.E-05	1.113	1.142
1024-512-256-128-64-32	718882	0.25	1.E-04	1.132	1.145
1024-1024-1024-1024	3170306	0.15	5.E-05	1.091	1.144

The FCNN configuration with five hidden layers and a decreasing structure of 1024–512–256–128–64 yielded the best performance in terms of validation loss. Networks with a higher number of trainable parameters produce similar results; however, the flat architecture with four hidden layers of 1024–1024–1024–1024 showed signs of overfitting (see Figure 1) and a larger number of trainable parameters didn’t improve the performance.

3.4 Monthly and Yearly Models

A previous preliminary study [9] showed that model performance was higher when validated on the same month (but in a different year) that it was trained on. This improvement may be attributed to better adjustment to the seasonal variability of the biases. In the present work, we validate this hypothesis by training both an all-year model using the complete five-year dataset and individual monthly models.

The NN configuration defined in section 3.3 was used to train both the seasonal model for January and the model trained on the full five-year dataset. An exhaustive hyperparameter search for the all-year model was not performed due to computational limitations.

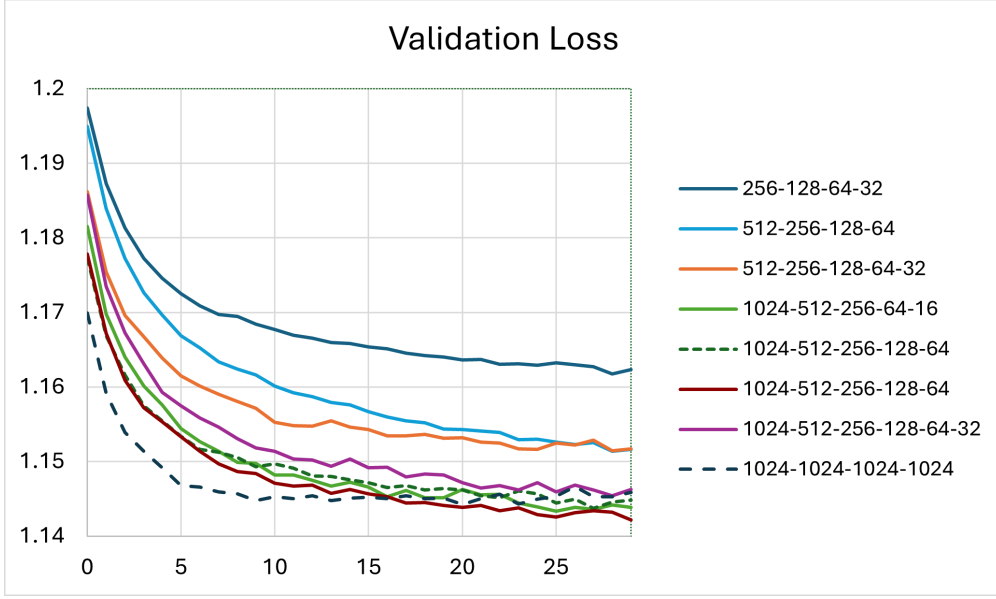


Figure 1: Evolution of validation loss over 30 epochs of training.

For the all-year model, three different versions were trained and validated using data from the test period of January 2023. The first version was trained using only the input parameters listed in Table 1. Given the presence of seasonal variability that the model should capture, a second version included an additional input: the day of the year, normalized by the total number of days in a year. However, this linear normalization between 0 and 1, resulted in poor performance for January as it didn't capture well the transition between December and January (see Table 4).

The third version, which is further evaluated in Section 4 (Results and Discussion), incorporated sine and cosine transformations for the day-of-year input parameter. This cyclical encoding shows improved performance compared to versions 1 and 2.

The neural network configuration for the second and the third versions remained the same as defined in Section 3.3, as the primary goal was to isolate the effect of adding seasonal information rather than introducing architectural changes. While adding an extra input could theoretically increase the risk of overfitting, the results indicate that the models generalize well.

Table 4: Mean error variance reduction (in %) for January 2023 for the all-year models validated against ASCAT-B, ASCAT-C, HSCAT-B and HSCAT-C w.r.t ERA5.

Global	Version1: All-year	Version 2: Date norm	Version 3: Sincos
ASCAT-B	13.08	12.49	13.35
ASCAT-C	13.49	12.81	13.76
HSCAT-B	9.11	7.62	9.33
HSCAT-C	10.54	8.84	10.55

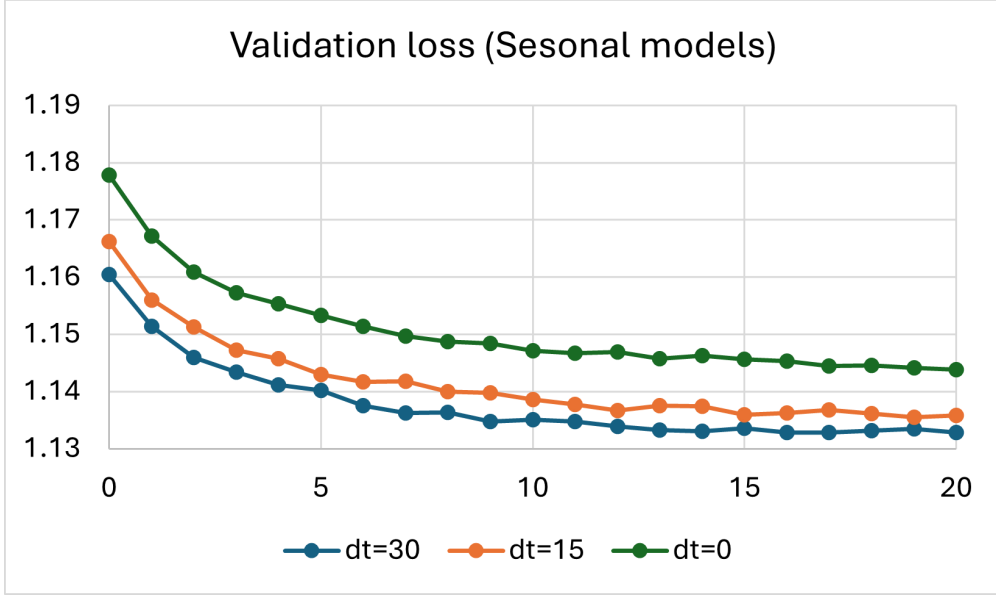


Figure 2: Validation loss for the January monthly models including only January data ($dt=0$), 15 days from the end of December and 15 days from the beginning of February ($dt=15$) and December-January-February data ($dt=30$).

Three preliminary tests were conducted for the January seasonal model. Unlike the all-year models, the monthly models did not include the day of the year input parameter. The first version of the January model was trained on all the January data extracted from the five-year training period. The second version extended the training window by including an additional 15 days from the second half of December and 15 days from the beginning of February. The third version was trained on the full winter season (December–January–February).

Figure 2 shows the validation loss over 20 training epochs, where a notable improvement can be observed for the model trained with the ± 15 -day extension. A further, though smaller, improvement is achieved when training on the entire season; however, this gain is less significant when weighed against the additional training time and computational resources required.

In Section 4, results are presented for the seasonal models trained with the ± 15 -day extension.

3.5 Inference

To prepare the inputs for the inference, ERA5 forecasts and GLOBCURRENT surface currents were linearly interpolated to a 0.125° grid and collocated for each ERA5 analysis time and forecast step. sea surface temperature data were taken from the same analysis time.

SST gradients, as well as U10S curl and divergence, were calculated by first interpolating the SST and wind fields to a staggered grid, ensuring that the resulting grid matches the grid of the other input parameters. Further details on the interpolation applied to each input variable can be found in Appendix Table 2.

The collocated input variables were subsequently normalized and flattened for inference. The predicted biases were then reshaped back to the 0.125° regular grid, applied to ERA5 U10S forecasts as corrections, and stored in both NetCDF and GRIB formats.

3.6 Validation

The corrected ERA5 forecasts for the test period (2023) were collocated with the same scatterometers used in the training dataset, specifically ASCAT-B and ASCAT-C. This validation was performed at the same local times as those observed during training.

Additional validation was conducted using independent scatterometers: HSCAT-B onboard the sun-synchronous Haiyang-2B (HY-2B) satellite, which has an equatorial crossing time of 6am for the descending orbit (approximately 3.5 hours before from MetOp’s passing time), and HSCAT-C onboard Haiyang-2C (HY-2C), which operates in a 66° inclined orbit and samples at varying local times.

The metrics presented in Section 4 include the Vector Root Mean Square difference (VRMS) (Equation 2) and error variance reduction (Equation 3). These metrics were calculated excluding data points flagged by the following quality control (QC) flags available in the product: land_qc, var_qc, knmi_qc, and the product monitoring flag.

$$VRMS = \sqrt{\frac{1}{N} \sum_i (u_i^{scat} - u_i^{model})^2 + (v_i^{scat} - v_i^{model})^2} \quad (2)$$

$$Var_red(\%) = \frac{VRMS_{ERA5}^2 - VRMS_{ML}^2}{VRMS_{ERA5}^2} * 100\% \quad (3)$$

4 Results and discussion

4.1 Differences between ERA5 biases

ERA5 biases and VRMS differences vary depending on the scatterometer instrument used for validation. Table 5 presents the mean yearly VRMS values of ERA5 compared against four scatterometers: the C-band instruments ASCAT-B and ASCAT-C, which show global mean VRMS values around 2 m/s, and the Ku-band instruments, which generally exhibit lower VRMS values, particularly HSCAT-B. Tropical regions (latitudes $<30^\circ$) exhibit the lowest VRMS values, primarily due to the lower mean wind intensity. The highest errors are observed in high-latitude regions ($>55^\circ$), while the extra-tropics show intermediate values.

Recall that triple collocation validation involving in-situ data (moored buoys), persistently shows that scatterometer winds have the lowest VRMS difference with respect to buoys, where ASCAT wind errors are clearly the smallest [15]. Moreover, recall the relatively low spatial resolution of PenWP and ERA5 winds, favoring a low VRMS in table 5.

Table 5: Mean yearly VRMS of ERA5 against ASCAT-B and -C and HSCAT-B and -C scatterometers.

ERA5 VRMS (m/s)	Global	Tropics	Extra-Tropics	High Latitudes
ASCAT-B	2.036	1.999	2.026	2.143
ASCAT-C	2.030	1.996	2.020	2.130
HSCAT-B	1.602	1.550	1.593	1.756
HSCAT-C	1.682	1.615	1.687	1.804

Figure 3 illustrates the spatial distribution of ERA5 biases in the zonal and meridional wind components, as estimated using ASCAT-B and HSCAT-B U10S as reference. Recall that due to QC differences and sampling, the weather sample at each location may be different for the two instrument collocations and hence the ERA5 errors might differ too. While the overall patterns are generally similar across instruments, HSCAT-B related biases are less pronounced in the zonal component, particularly in the tropical regions, and a broader band of meridional bias associated with the ITCZ. In the extra-tropical regions, the magnitude of the HSCAT-B related biases is also lower than that of the ASCAT-B related biases.

In general, HSCAT-B related bias maps present less granularity than ASCAT-B related bias maps. This is due to the fact that PenWP Ku-band pencil-beam scatterometer winds are in general of lower spatial resolution than AWDP C-band fan-beam scatterometer winds. Moreover, Ku-band scatterometers are substantially more affected by rain contamination effects than C-band scatterometers. As such, Ku-band QC is more active in regions such as the ITCZ than C-band QC, which leads to differences in weather sampling and therefore differences in estimated biases. In particular, note that ERA5 winds near rain have rather large wind errors [23] and will generally not be validated in HY2 data set due to QC. Moreover, as mentioned in Section 3.6, the (temporal) sampling is also rather different between C-band and Ku-band systems, leading once again to different weather sampling and thus potentially different biases.

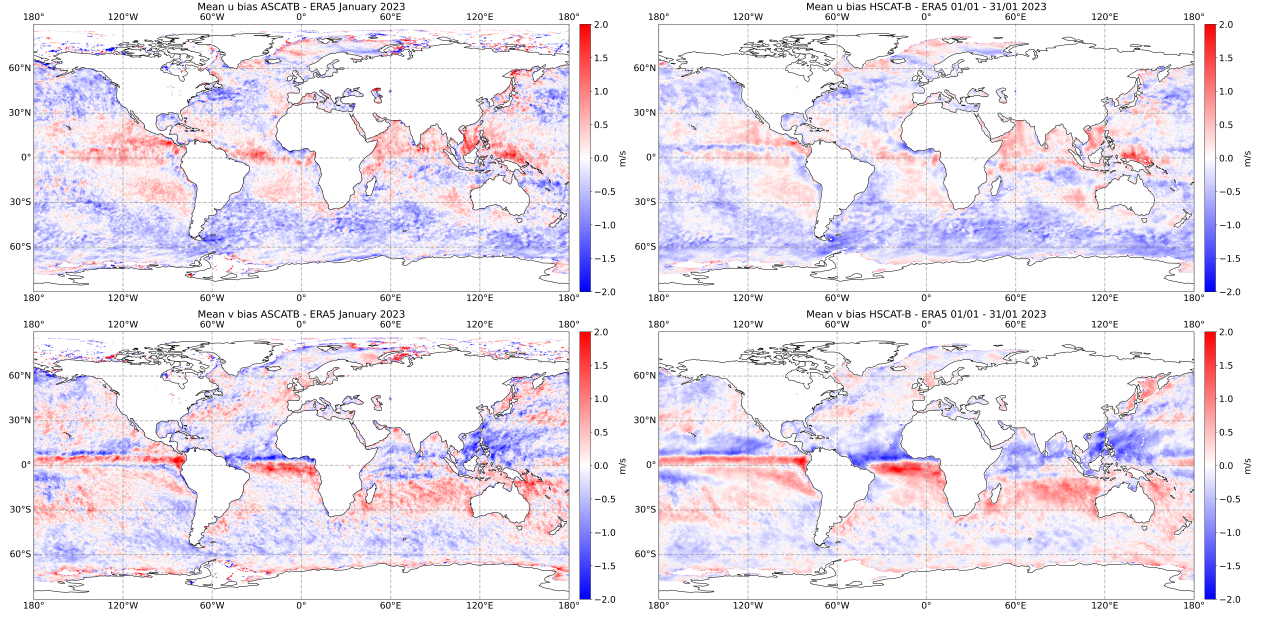


Figure 3: Top: Mean bias (m/s) of ERA5 against ASCAT-B (left) and ERA5 against HSCAT-B (right) for January 2023 (with own QC filtering), for both the zonal (top) and the meridional (bottom) U10S components.

Since the neural networks were trained to replicate ASCAT data, these differences in ERA5 sample bias characteristics may influence the validation metrics when evaluated against independent Ku-band scatterometers such as HSCAT-B and HSCAT-C. Nevertheless, since ASCAT winds are of higher wind vector quality and available over a broader bandwidth of weather conditions, the ERA5 corrections should be optimal if well sampled (over a long time period).

4.2 All-year and monthly models

Table 6 presents the performance of the all-year and monthly neural network outputs for the year 2023, measured in terms of error variance reduction relative to that of ERA5. Overall, the monthly models slightly outperform the all-year model, achieving approximately 13% error variance reduction against ASCAT scatterometers, compared to the 12% reduction for the all-year model and 8.6–9% against the HSCAT instruments. When verified against HSCAT data, the error reduction is also larger for the monthly models, i.e., up to 8.6–9

In tropical regions (latitudes $<30^\circ$), the monthly models achieve an error variance reduction of 11–12% against ASCAT scatterometers. This represents a significant improvement over the 8% reduction reported in the preliminary study [9]. However, the error reduction is more moderate when validated against Ku-band scatterometers, particularly HSCAT-B. In this region, the monthly models show the most notable improvement compared to the all-year model.

In the extra-tropics (latitudes between 30° and 55° N/S), the neural network models demon-

Table 6: Mean error variance reduction for all-year and monthly models over the 2023 test period.

%	Global		Tropics		Extra Tropics		High Latitudes	
	All-year	Monthly	All-year	Monthly	All-year	Monthly	All-year	Monthly
ASCAT-B	12.19	12.97	10.24	11.49	15.44	15.83	10.39	10.88
ASCAT-C	12.44	13.22	10.48	11.82	15.77	16.07	10.51	10.95
HSCAT-B	7.88	8.60	3.81	4.90	11.71	12.03	9.35	10.00
HSCAT-C	8.77	9.05	5.10	6.09	11.29	11.22	10.60	10.29

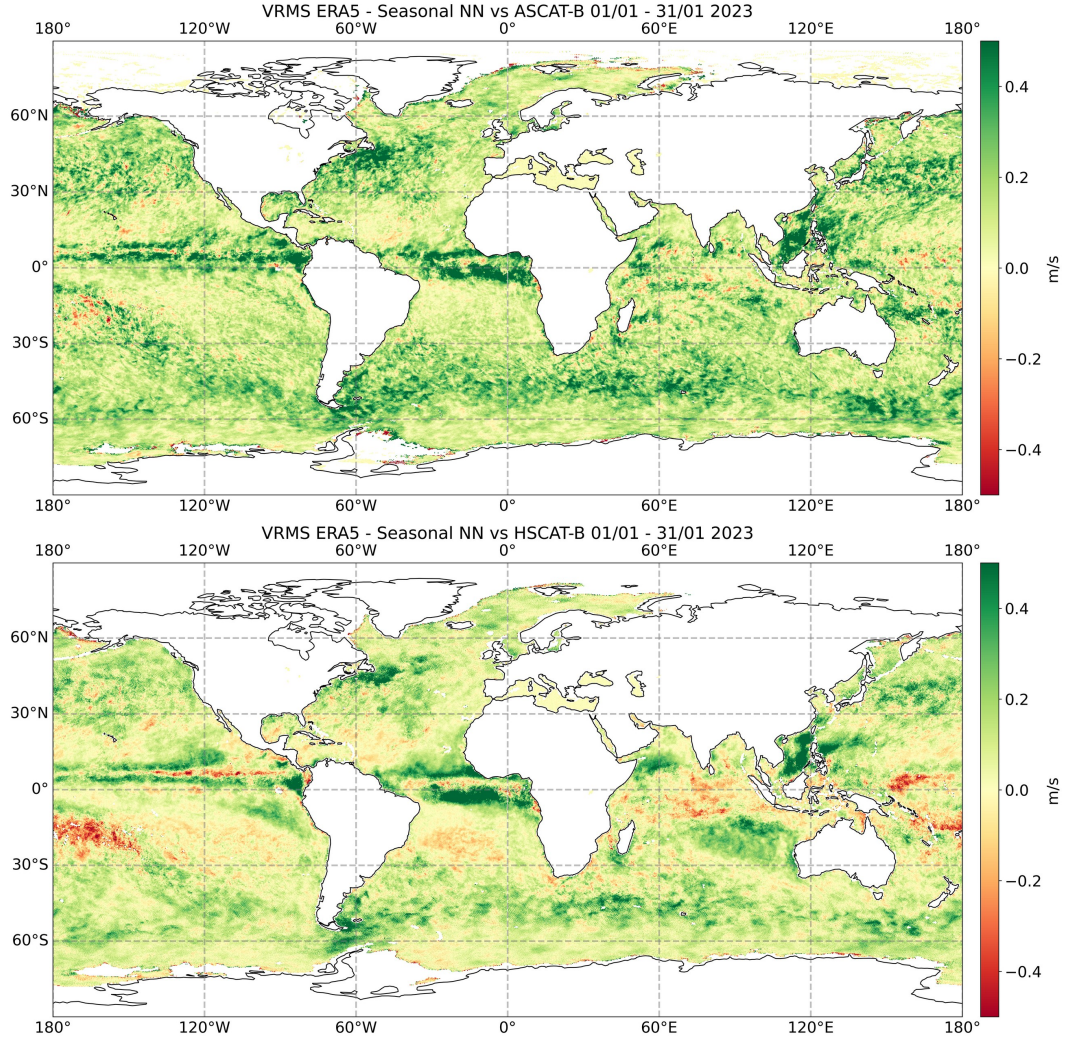


Figure 4: Mean VRMS reduction (green) or increase (red) of monthly NN output (compared to ERA5), when verified against ASCAT-B (top) and HSCAT-B (bottom).

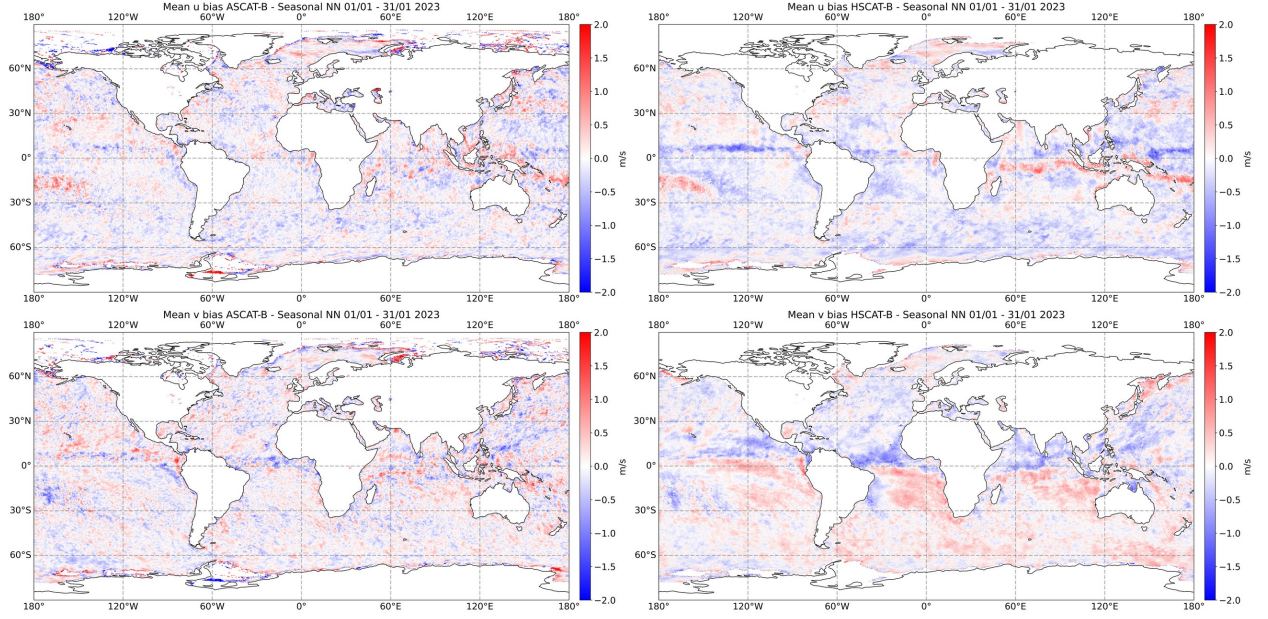


Figure 5: Mean u (top) and v (bottom) biases for the monthly NN outputs, validated against ASCAT-B (left) and HSCAT-B (right) for January 2023.

strate the highest performance, with error variance reductions of approximately 16% against ASCATs and 11–12% against HSCATs. At high latitudes ($>55^\circ$ N/S), performance is similar across both scatterometer types, with error variance reductions around 10–11%. It should be noted that HSCAT-C does not provide measurements beyond approximately 74° N/S [24].

Figure 4 shows the spatial distribution of VRMS reduction/increase for the monthly neural network outputs compared to ERA5, when validated against ASCAT-B and HSCAT-B. Yellow areas indicate neutral impact, while green (red) regions highlight areas where performance is improved (degraded) relative to that of the ERA5 baseline. Validation against the C-band scatterometers reveals notable improvements around the ITCZ. In the extra-tropics, the most significant improvements are observed in the western parts of the oceanic basins in the Northern Hemisphere, corresponding to regions of strong zonal surface currents. In the Southern Hemisphere, the improvement is more spatially homogeneous. The validation against HSCAT-B, however, shows more modest improvement and some degradation in the tropics, mainly in the ITCZ.

The left panels in Figure 5 shows the resulting biases in the zonal and meridional wind components from the monthly NN models, validated against ASCAT-B. Compared to Figure 3, the recurring bias structures largely disappear, leaving behind a more granular residual bias. However, this reduction in biases is not fully observed when validating against Ku-band scatterometers, where some bias structures persist (see right panels in Fig. 5). Distinct bias patterns remain, which can be attributed to the differences between C-band and Ku-band sampling discussed in Section 4.1 (and, specifically in Fig. 3). The NNs are trained to replicate the ERA5 biases against ASCAT in all weather conditions at 9:30 LST, which might not fully capture the ERA5 biases associated with

Ku-band systems at other times.

The differences may be caused by diurnal cycle related biases on ERA5, since the instruments have different local sampling times. Additionally, differences in the viewing geometry and the operating frequency (Ku-band being substantially more affected by rain than C-band), as discussed in Section 4.1, also can explain some of the differences seen in Fig. 5. This should be further investigated.

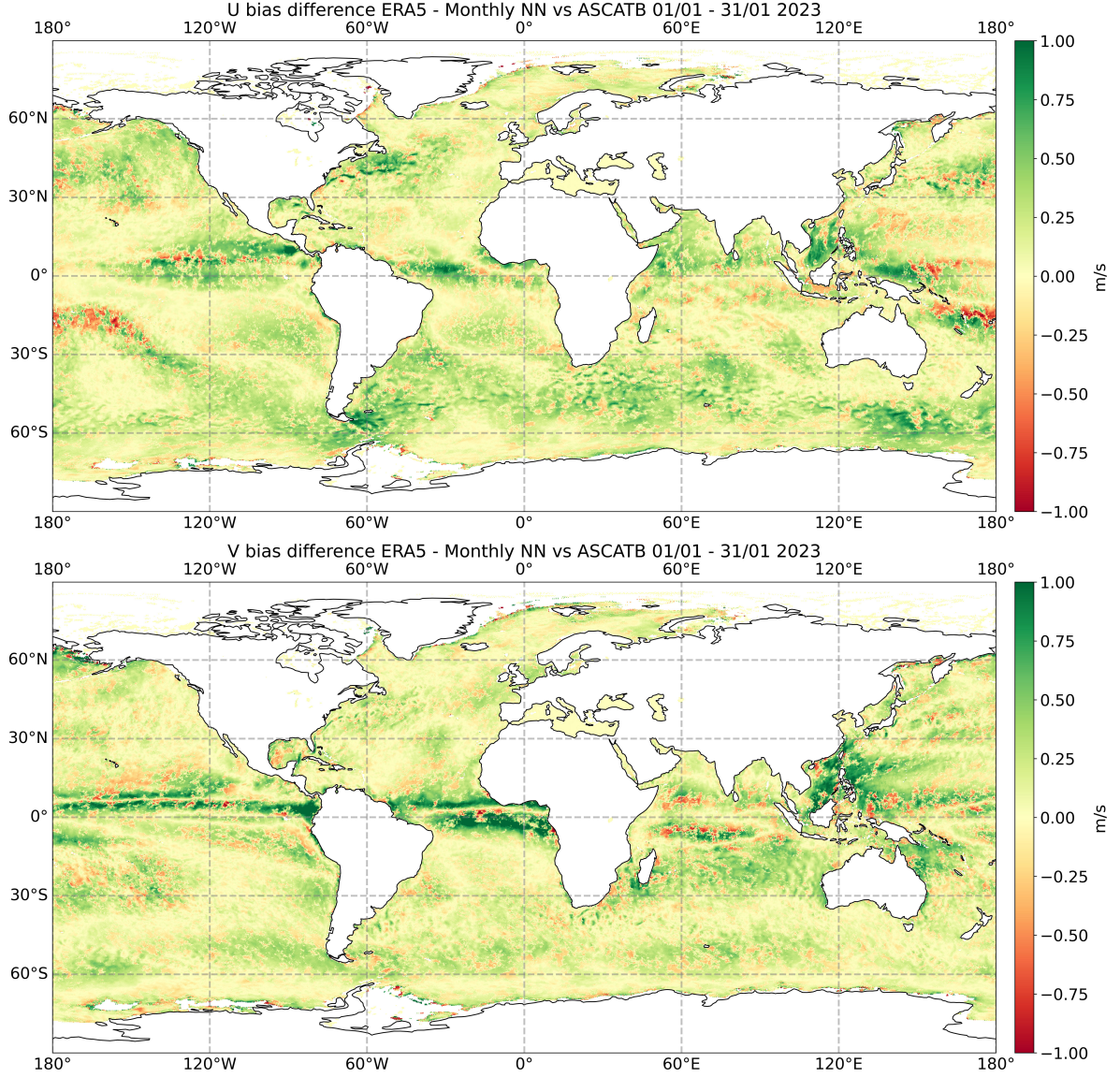


Figure 6: Mean absolute bias difference between ERA5 and monthly NN output for u (top) and v (bottom), validated against ASCAT-B for January 2023.

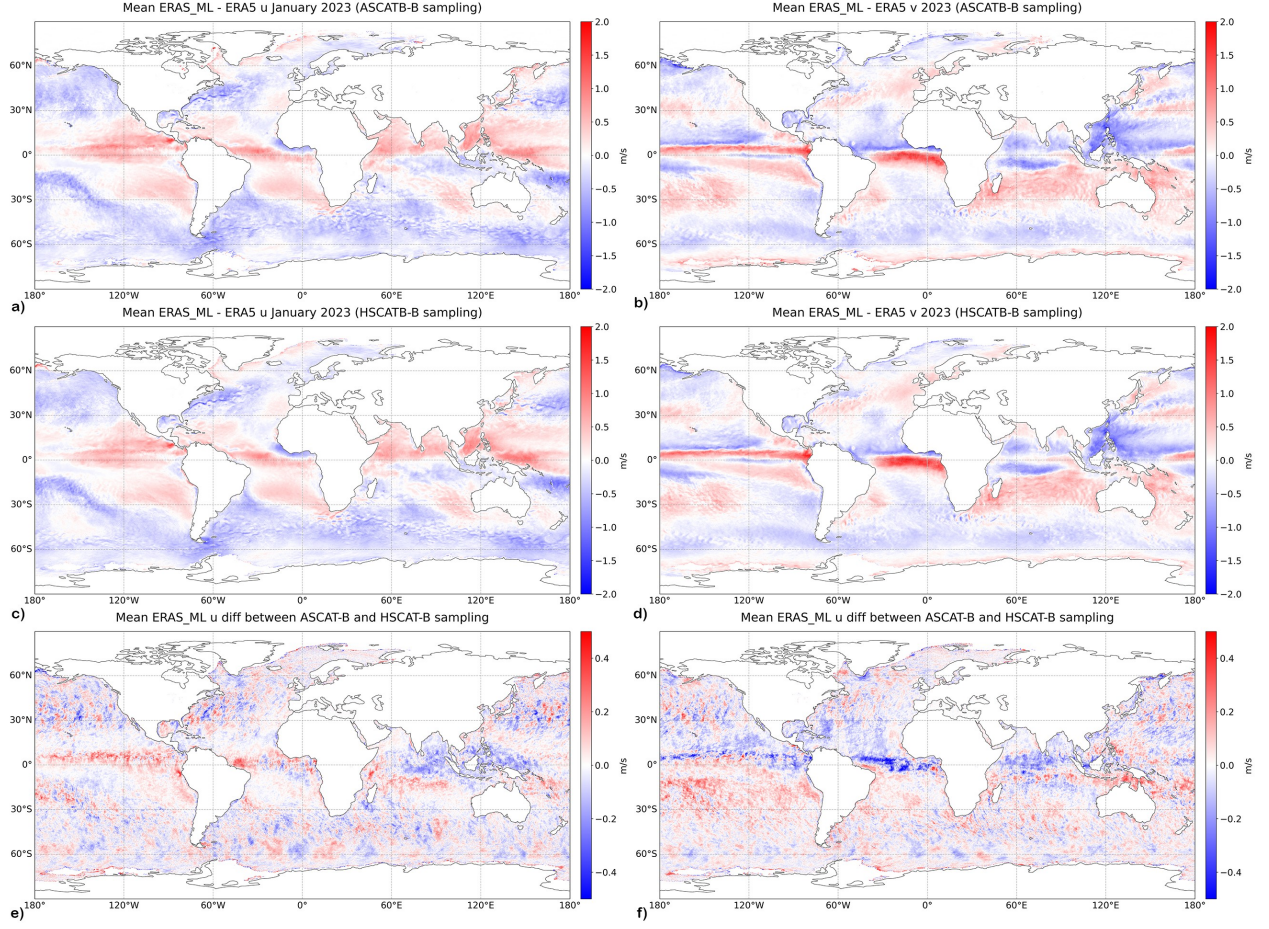


Figure 7: NN mean monthly output for January 2023 and for u (left) and v (right) components, computed at ASCAT-B (top) and HSCAT-B (middle) sampling. The differences between the top and the middle plots are shown in the bottom plots.

Figure 6 illustrates the difference between ERA5 and monthly NN mean absolute monthly biases for the zonal and meridional components against ASCAT-B. Green (red) areas indicate a reduction (increase) in the NN bias amplitude relative to that of ERA5, with the strongest improvements occurring in parts of the Tropics, particularly along the ITCZ for both components. For the zonal component, the reduction pattern closely follows ocean currents, especially in the extra-tropics.

Figure 7 shows the mean monthly NN output (correction) for January 2023, but computed only at times and locations corresponding to either ASCAT-B (top plots) or HSCAT-B (middle plots) overpasses. The overall structure of the mean corrections is very similar for both samplings, with more pronounced differences in the ITCZ, as shown in the bottom plots. The NN output partially simulates the diurnal cycle; however, it remains fully dependent on the representation of the diurnal cycle in the other ERA5 atmospheric variables used as input in the training. Since SST includes

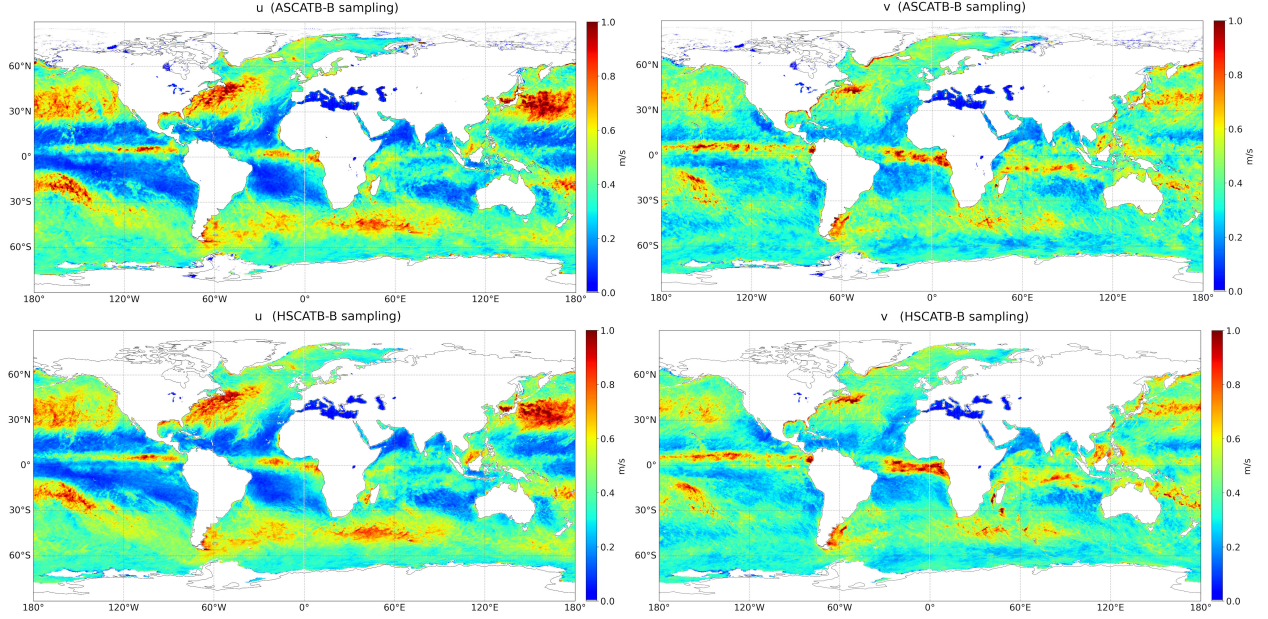


Figure 8: Standard deviation of the NN monthly output, for u (left) and v (right) components, computed only at ASCAT-B (top) and HSCAT-B (bottom) sampling.

no diurnal variability and daily mean values are used for the surface currents, these inputs do not contribute to any diurnal signal in the NN output.

Figure 8 shows the standard deviation of the monthly NN output, computed only at ASCAT-B (top) or HSCAT-B (bottom) sampling. The patterns are very similar for both sampling types, with the highest values occurring in the ITCZ and the extra-tropics, the latter corresponding to the passage of wintertime low-pressure systems. While the biases predicted by the NN are weather-dependent (see correspondence between Figures 7 and 8), the lack of diurnal variability on the ERA5 inputs may limit the NN’s ability to provide accurate corrections for local times not represented during training.

4.3 Monthly statistics

This section presents the metrics for the monthly NNs only, as they generally performs better than the all-year model. Figure 9 displays the monthly mean statistics for ERA5 VRMS when validated against four scatterometers, along with the monthly mean error variance reduction achieved by the monthly NNs models.

Globally, the ERA5 VRMS values compared to ASCAT-B and ASCAT-C are nearly identical, with slightly higher errors during winter and spring than in summer. In line with Table 5, ERA5 VRMS shows significantly lower values against Ku-band sensors than against C-band sensors. For the Ku-band verification, the largest errors occur in March and April, while the lowest are observed during the summer months. On a global scale, the NN performs best in January and February, with

a noticeable drop (mostly driven by the sharp drop in the tropics) in performance in March across all four satellites. The autumn months also show slightly lower error variance reduction compared to other periods.

In tropical regions, the highest ERA5 VRMS values against C-band systems are observed in May and June, while for Ku-band systems, the peak occurs in spring—particularly in March and April. The NN model’s performance declines during these months, especially in March when validated against HSCAT-B. The best results in the tropics are seen in January and February.

In the extra-tropics, ERA5 shows the highest VRMS values in January and February, and the lowest in May and October–November. In this region, the NN model performs consistently well throughout the year.

At high latitudes, the interannual variability is more pronounced, with higher VRMS values during spring and autumn. The NN model shows similar performance across all satellites. This is probably due to the fact that, at high latitudes, the temporal sampling is more spread throughout the day for all sun-synchronous systems, i.e., ASCAT-B, -C, and HSCAT-B. Indeed, when the temporal sampling is similar for C- and Ku-band systems, the model tends to perform similarly. This indicates that at lower latitudes, and most relevantly in the tropics, daily cycle errors in ERA5 may be relevant.

As HSCAT-C does not sample latitudes above 74° its geographical sampling at high latitudes is quite different from the other instruments, hence resulting in different temporal evolution of associated errors.

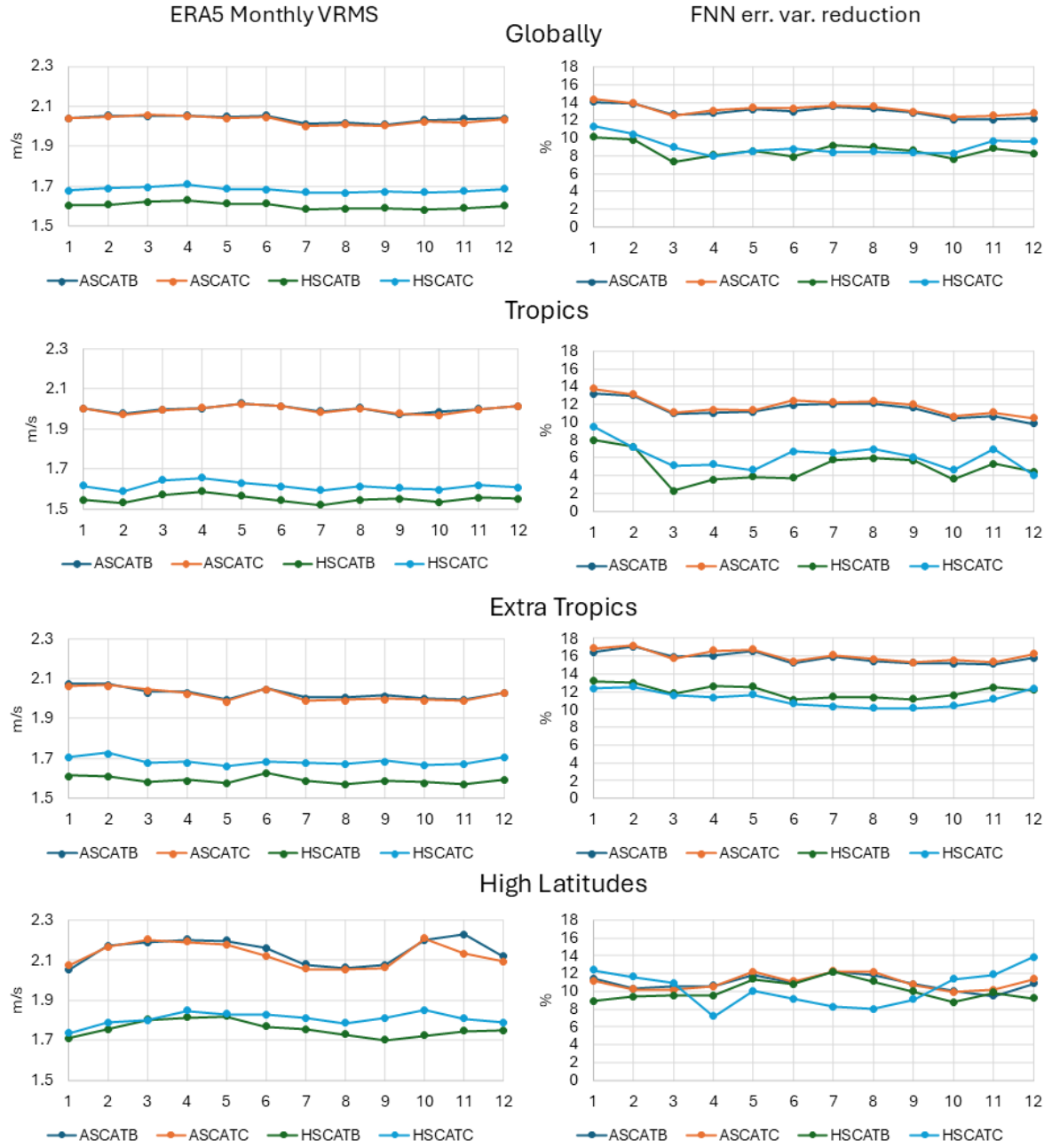


Figure 9: Mean monthly ERA5 (left) and monthly NN error variance reduction (right) against ASCAT-B, ASCAT-C, HSCAT-B and HSCAT-C (see legend), for global, tropical, extra-tropical and high latitude regions.

4.4 Spatial Variance of Corrected Outputs

The monthly NN models' output exhibits spatial variance patterns similar to those of the original ERA5 forecasts. Figure 10 shows the power spectral density of ERA5 and monthly NN models' U10S, collocated with ASCAT-B and HSCAT-B U10S in February 2023.

In the tropical regions, the NN output shows slightly reduced spatial variance at scales between 500 and 1000 km compared to that of ERA5. In the extra-tropics, however, the spectral slopes of the monthly NN models closely match those of ERA5, indicating that corrections do not reduce the spatial variance of the original wind fields.

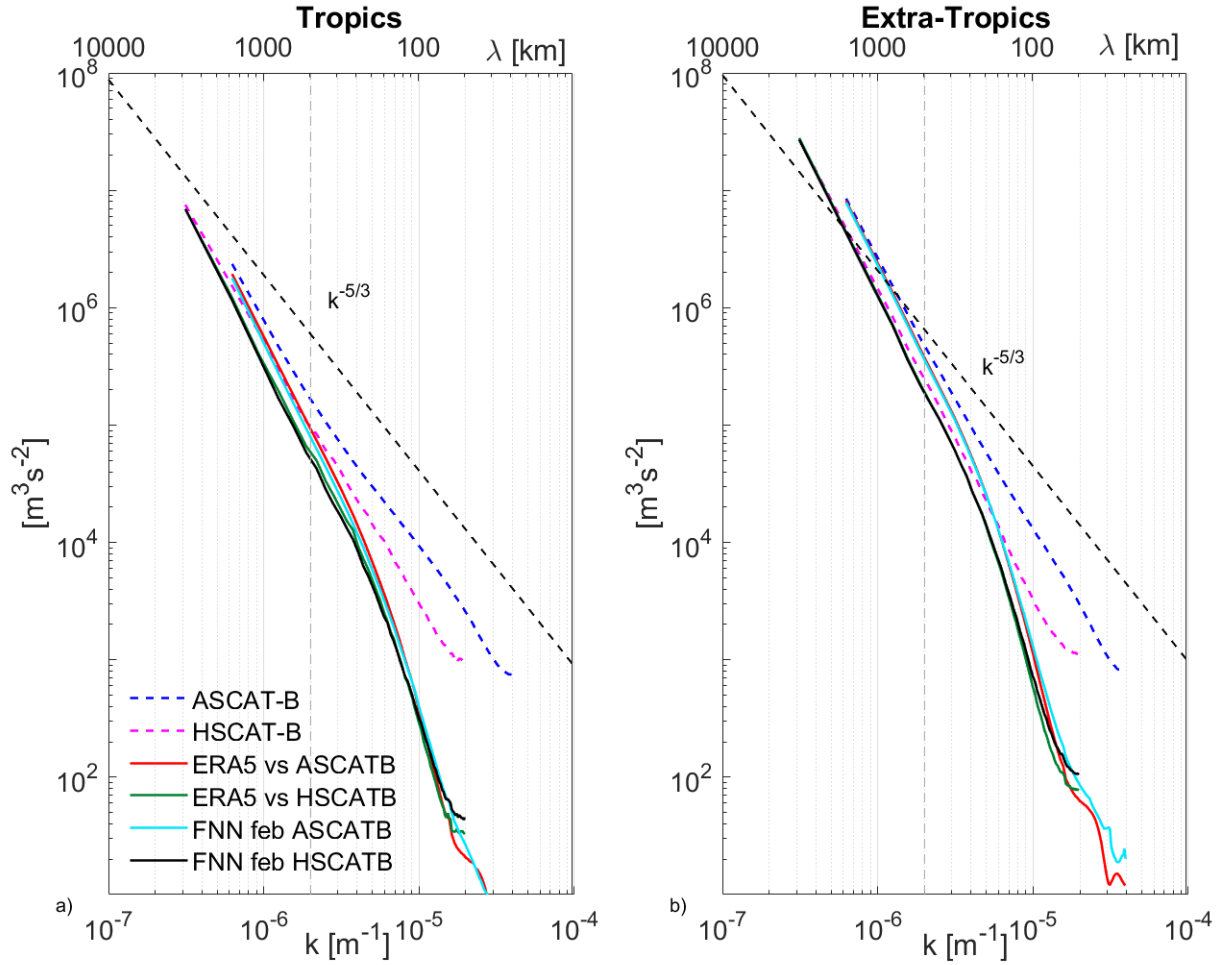


Figure 10: Spectral density plot of the zonal U10S over February 2023, for the tropics (left) and the extra-tropics (right), for: ASCAT-B, HSCAT-B, ERA5 and monthlyNN output collocated with ASCAT-B and HSCAT-B

Before this study, a preliminary work was done to evaluate the performance of the Convolutional

Neural Networks (CNNs), such as U-Net [25]. However, using RMSE/Mean Absolute Error (MAE) as loss functions the NN learned to smooth the wind fields to achieve better metrics resulting in the important loss of the spatial variance. Due to this issue, and despite showing better metrics in terms of VRMS, the CNN architecture was discarded and not used in the current study.

5 Conclusions and future work

This study demonstrates that Machine Learning (ML) models can effectively reduce persistent stress-equivalent wind biases in ERA5 forecasts without requiring scatterometer observations during inference. By training FCNNs on five years of atmospheric and oceanic data, the models successfully captured systematic relationships between NWP variables and wind bias patterns, leading to significant improvements over the ERA5 baseline.

The results show that both all-year and monthly models achieve notable bias reductions. Moreover, on a global scale, error variance reductions reached approximately 13% when validated against ASCATs and around 9% against HSCATs, with the highest performance observed in the extra-tropics, where improvements reached up to 16%. Monthly NN models outperform more generalized models trained on the entire five years of data, particularly in tropical regions where seasonal variability strongly influences bias patterns.

Regional analyses reveal that the extra-tropics benefits the most from the ML-based corrections, while tropical regions show mixed results, especially when validated against Ku-band scatterometers. This discrepancy reflects the differences in the ERA5 biases derived at C-band and Ku-band scatterometer sampling, which requires further investigation. At high latitudes, both monthly and all-year NN models achieve similar improvements, reducing the error variance by about 10–11%.

Although the NN-corrected U10S significantly reduced large-scale bias structures, some residual patterns persisted when validated against independent Ku-band scatterometers. These differences are possibly associated with different viewing geometry, temporal sampling (orbit characteristics) and operating frequency (rain effects) of both C-band and Ku-band systems, resulting in different weather sampling. It may imply that ERA5 errors are not equally well corrected in all different weather conditions and at all local times. The follow-up study will focus on the analysis of the diurnal variability of the ERA5 biases using inclined-orbit satellites, such as HY-2C and HY-2D, as well as assessing the capability of the neural networks to predict this variability based solely on the NWP input information. This assessment will help to determine whether Ku-band scatterometer data from inclined-orbit satellites should be included in the training dataset to extend the local temporal sampling for improved representation of the diurnal cycle, even if these data are more weather dependent.

The approach of using simple feed-forward neural networks trained on a point-by-point basis preserved the spatial variance of the corrected wind outputs, with only a slight reduction in U10S variance in the tropical region at scales between 500 and 1000 kilometers.

The demonstrated ability of neural networks to reduce wind biases provides a foundation for further improvements. Future work will focus on several key aspects. One of the lines of work is the evaluation of customized loss functions that, in addition to reducing biases, also minimize differences in wind field derivatives such as curl and divergence.

The present study employed relatively simple feed-forward neural networks, while next efforts will explore more advanced architectures, potentially based on Graph Neural Networks (GNNs), which have recently demonstrated promising results in weather forecasting [26, 27].

Further analysis will address feature importance at the regional scale, as the current assessment was limited to global analysis in the initial phase of this study [8]. This will help determine whether

the current combination of NWP input variables is optimal or if additional predictors could enhance the accuracy of the U10S bias estimation.

Another line of research will involve updating model weights for operational ECMWF IFS data and assessing the added value of bias correction for real-time wind forecasting. Additionally, the impact of corrected wind forcing on wave modeling will be evaluated to quantify the broader benefits of this approach for ocean and climate applications.

A Appendix

Table 1: Normalization of input features and targets.

Variable	type	mean	std
Longitude	sincos	-	-
Latitude	sin	-	-
U10S speed	norm	8.102557	3.654
U10S direction	sincos	-	-
U10S Eastward component	norm	0	6.887
U10S Nothward component	norm	0	5.591
Mean sea level pressure	norm	100796.7	1471.285
Air temperature	norm	287.205	10.506
Specific humidity	norm	0.009231	0.0055
SST	norm	288.5919	10.328
Eastward sea water velocity	norm	0	0.209
Northward sea water velocity	norm	0	0.158
U10S wind curl	norm	0	3.00E-05
U10S wind divergence	norm	0	2.00E-05
Eastward SST gradient	norm	0	7.20E-06
Nothward SST gradient	norm	0	9.00E-06
Day of year	sincos	-	-
Target u difference	norm	0	1.364
Target v difference	norm	0	1.461

Table 2: Normalization of input features and targets.

Variable	Interpolation type
Longitude	0.125 regular mesh
Latitude	0.125 regular mesh
U10S speed	Linear
U10S direction	Nearest
U10S Eastward component	Linear
U10S Nothward component	Linear
Mean sea level pressure	Linear
Air temperature	Linear
Specific humidity	Linear
SST	Nearest
Eastward sea water velocity	Nearest
Northward sea water velocity	Nearest
U10S wind curl	Linear half grid
U10S wind divergence	Linear half grid
Eastward SST gradient	Linear half grid
Nothward SST gradient	Linear half grid
Day of year	Fill

References

- [1] I. Sandu, P. Bechtold, L. Nuijens, A. Beljaars, and A. Brown, “On the causes of systematic forecast biases in near-surface wind direction over the oceans,” Jun. 2020, doi: 10.21957/wggbl43u. [Online]. Available: <https://www.ecmwf.int/node/19545>
- [2] M. Belmonte Rivas and A. Stoffelen, “Characterizing ERA-Interim and ERA5 surface wind biases using ASCAT,” *Ocean Science*, vol. 15, no. 3, pp. 831–852, Jun. 2019, doi: 10.5194/os-15-831-2019.
- [3] A. Trindade, G. Grieco, E. Makarova, and M. Portabella, “WOC ERA* Hourly Global Stress Equivalent Wind and Wind Stress (V.2.0) [dataset],” Apr. 2022, doi: 10.13039/501100000844.
- [4] A. Trindade, M. Portabella, A. Stoffelen, W. Lin, and A. Verhoef, “ERASTAR: A High-Resolution Ocean Forcing Product,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1337–1347, Feb. 2020, doi: 10.1109/TGRS.2019.2946019.
- [5] M. Portabella, A. Trindade, G. Grieco, and E. Makarova, “World Ocean Circulation : Algorithm Technical Baseline Document for ERAstar (v2.0),” Jun. 2022, Publisher: European Space Agency, doi: 10.13039/501100011033.
- [6] E.U. Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS), “Global Ocean Hourly Sea Surface Wind and Stress from Scatterometer and Model [dataset],” 2023, doi: doi.org/10.48670/moi-00305. [Online]. Available: https://data.marine.copernicus.eu/product/WIND_GLO_PHY_L4_NRT_012_004/description
- [7] —, “Global Ocean Hourly Reprocessed Sea Surface Wind and Stress from Scatterometer and Model [dataset],” 2024, doi: doi.org/10.48670/moi-00185. [Online]. Available: https://data.marine.copernicus.eu/product/WIND_GLO_PHY_L4_MY_012_006/description
- [8] E. Makarova, M. Portabella, and A. Stoffelen, “Proposal and Evaluation of the Machine Learning Models for Correcting ERA5 Stress Equivalent Wind Forecasts as a Function of Atmospheric and Oceanic Conditions,” Nov. 2022, publisher: EUMETSAT. [Online]. Available: <https://digital.csic.es/handle/10261/81991>
- [9] —, “Reduction of persistent stress-equivalent wind biases with machine learning and scatterometer data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–11, 2025, doi: 10.1109/TGRS.2025.3586375.
- [10] K. Zhao, A. Stoffelen, J. Verspeek, A. Verhoef, and C. Zhao, “A Conceptual Rain Effect Model for Ku-Band Scatterometers,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–9, 2023, doi: 10.1109/TGRS.2023.3264246.
- [11] 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, and J.-N. Thépaut, “The ERA5 global reanalysis,” *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803>
- [12] E.U. Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS), “Global Total (COPERNICUS-GLOBCURRENT), Ekman and Geostrophic currents at the Surface and 15m [dataset],” 2023, doi: doi.org/10.48670/mds-00327. [Online]. Available: https://data.marine.copernicus.eu/product/MULTIOBS_GLO_PHY_MYNRT_015_003/description
- [13] H. Etienne, “Product User Manual for For Global Ocean Multi Observation Products MULTIOBS_glo_phy_mynrt_015_003,” Tech. Rep., Nov. 2023. [Online]. Available: <https://catalogue.marine.copernicus.eu/documents/PUM/CMEMS-MOB-PUM-015-003.pdf>
- [14] J. de Kloe, A. Stoffelen, and A. Verhoef, “Improved use of scatterometer measurements by using stress-equivalent reference winds,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 2340–2347, 2017.
- [15] J. Vogelzang and A. Stoffelen, “On the accuracy and consistency of quintuple collocation analysis of in situ, scatterometer, and nwp winds,” *Remote Sensing*, vol. 14, no. 18, 2022, 10.3390/rs14184552. [Online]. Available: <https://www.mdpi.com/2072-4292/14/18/4552>
- [16] H. Etienne, “Global ocean multi observation products: Multiobs_glo_phy_mynrt_015_003 quality information document,” Copernicus Marine Service, Tech. Rep. CMEMS-MOB-QUID-015-003, May 2024. [Online]. Available: <https://documentation.marine.copernicus.eu/QUID/CMEMS-MOB-QUID-015-003.pdf>
- [17] A. Verhoef, J. Vogelzang, J. Verspeek, and A. Stoffelen, “AWDP User Manual and Reference Guide,” 2020. [Online]. Available: https://nwp-saf.eumetsat.int/site/download/documentation/scatterometer/AWDP/NWPSAF-KN-UD-005_AWDP_User_Guide_v3.3.pdf
- [18] —, “PenWP User Manual and Reference Guide,” 2022. [Online]. Available: https://nwp-saf.eumetsat.int/site/download/documentation/scatterometer/penwp/NWPSAF-KN-UD-009_PenWP_User_Guide_v4.0.pdf
- [19] J. Miguel Fernández, “Ocean surface wind properties of different scatterometer observation types and numerical weather prediction model fields,” Master’s thesis, Utrecht University, the Netherlands, 2023.
- [20] A. Stoffelen, “How well do we understand the water cycle over the ocean and its role in the dynamical coupling of the atmosphere and the ocean?” *GEWEX Quarterly*, vol. Q3, 2025. [Online]. Available: <https://www.gewex.org/gewex-content/uploads/2025/11/Q32025.pdf>

- [21] E.U. Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS), “Global Ocean Daily Gridded Reprocessed L3 Sea Surface Winds from Scatterometer [dataset] Product ID WIND_GLO_PHY_L3_MY_012_005,” 2024, doi: doi.org/10.48670/moi-00183. [Online]. Available: https://data.marine.copernicus.eu/product/WIND_GLO_PHY_L3_MY_012_005/description
- [22] “LinearNDInterpolator — SciPy v1.16.1 Manual.” [Online]. Available: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.LinearNDInterpolator.html>
- [23] W. Lin, M. Portabella, A. Stoffelen, J. Vogelzang, and A. Verhoef, “Ascat wind quality under high subcell wind variability conditions,” *Journal of Geophysical Research: Oceans*, vol. 120, no. 8, pp. 5804–5819, doi: 10.1002/2015JC010861.
- [24] OSI SAF Winds Team, “Product User Manual (PUM) for the HSCAT/OSCAT winds,” 2024. [Online]. Available: https://nwp-saf.eumetsat.int/site/download/documentation/scatterometer/penwp/NWPSAF-KN-UD-009_PenWP_User_Guide_v4.0.pdf
- [25] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015. [Online]. Available: <https://arxiv.org/abs/1505.04597>
- [26] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, O. Vinyals, J. Stott, A. Pritzel, S. Mohamed, and P. Battaglia, “Learning skillful medium-range global weather forecasting,” *Science*, vol. 382, no. 6677, pp. 1416–1421, 2023, doi: 10.1126/science.adi2336. [Online]. Available: <https://www.science.org/doi/abs/10.1126/science.adi2336>
- [27] S. Lang, M. Alexe, M. Chantry, J. Dramsch, F. Pinault, B. Raoult, M. C. A. Clare, C. Lessig, M. Maier-Gerber, L. Magnusson, Z. B. Bouallègue, A. P. Nemesio, P. D. Dueben, A. Brown, F. Pappenberger, and F. Rabier, “AIFS – ECMWF’s data-driven forecasting system,” 2024. [Online]. Available: <https://arxiv.org/abs/2406.01465>