Correction of NWP ocean surface wind biases with machine learning and scatterometer data

Evgeniia Makarova[1] (makarova@icm.csic.es), Marcos Portabella[1], Ad Stoffelen[2], Ana Trindade[1]

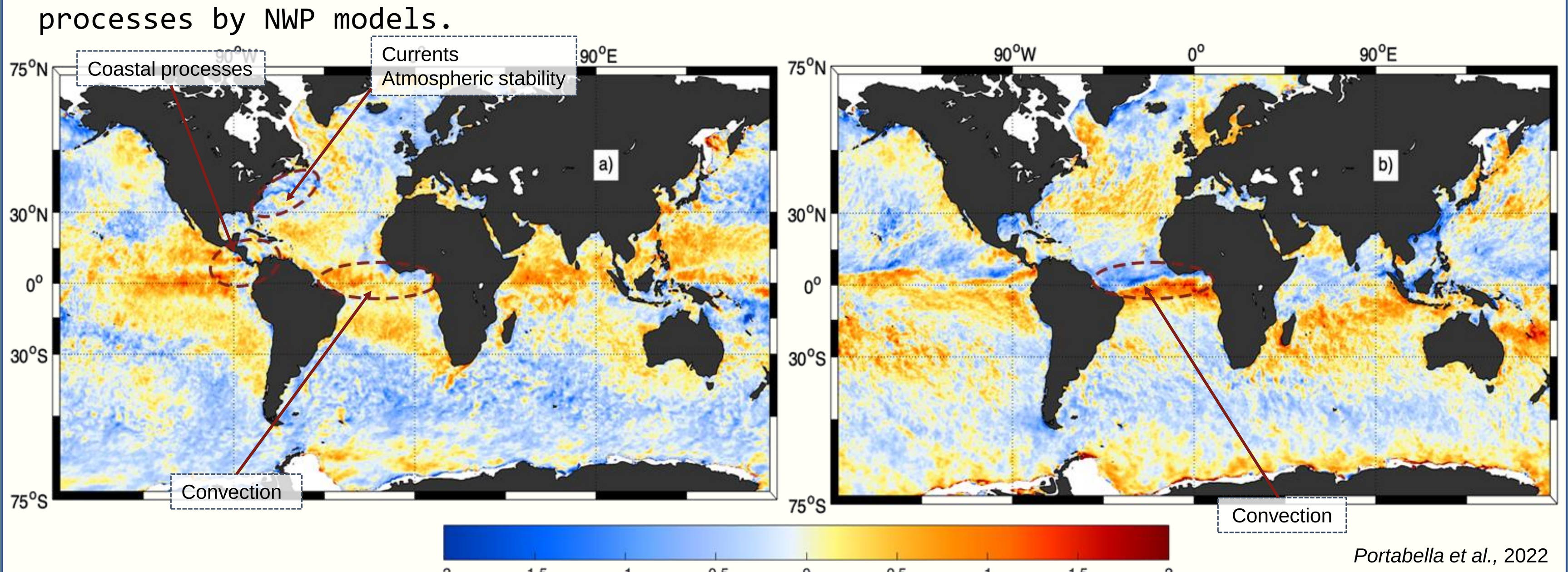
[1] Barcelona Expert Centre (BEC), Institute of Marine Sciences (ICM-CSIC) [2] Royal Netherlands Meteorological Institute (KNMI)

BEC Barcelona Expert Center

EUMETSAT OSI SAF OCEAN AND SEA ICE

Motivation

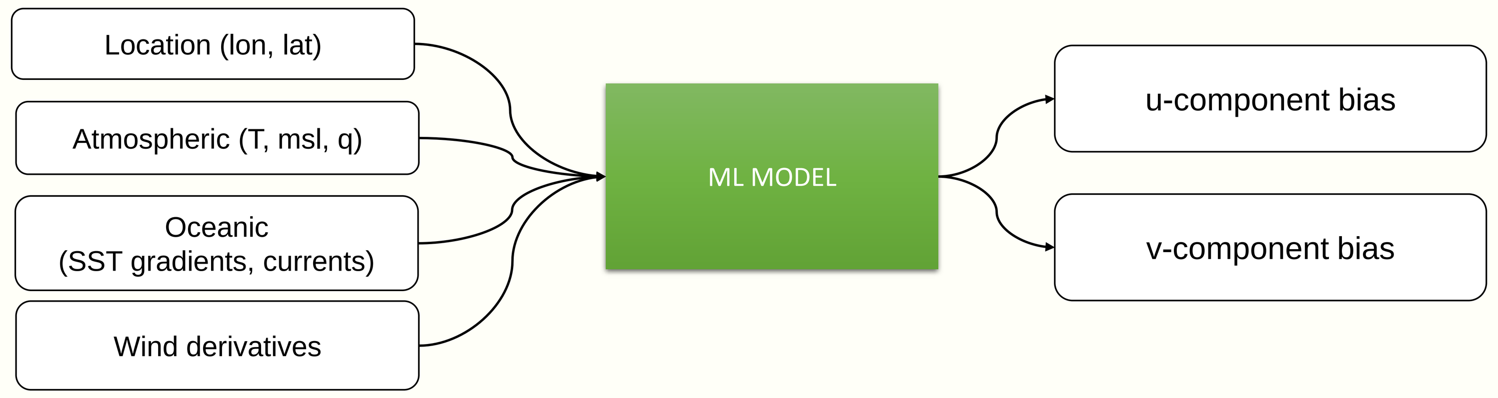
Global Numerical Weather Prediction (NWP) model sea-surface wind output is commonly used to force ocean models due to their time and space continuity. However, the output of the NWP models presents local biases, with one of the most systematic and longstanding biases in the sea surface wind direction [1]. After the assimilation of the stress-equivalent winds measured by scatterometers, the European Centre for Medium-Range Weather Forecasts (ECMWF) model output still presents the mentioned biases, which need to be corrected since they mostly represent unresolved geophysical processes by NWP models.



Persistent local wind biases globally for the zonal (a) and the meridional (b) components, i.e., scatterometer vs NWP differences accumulated over 30 days (February 2019).

Objectives

This work aims at creating a machine learning (ML) model for correcting the ECMWF ERA5 reanalysis stress-equivalent local wind biases. Several ML setups are evaluated, which look for the functional relationship between several oceanic and atmospheric variables and the persistent NWP biases as observed in the scatterometer-NWP differences. Such variables include ECMWF model parameters, such as stress-equivalent winds and their derivatives (curl and divergence), atmospheric stability related parameters, i.e., sea-surface temperature (SST), air temperature (Ta), relative humidity (rh), surface pressure (sp), as well as SST gradients and ocean currents [4].

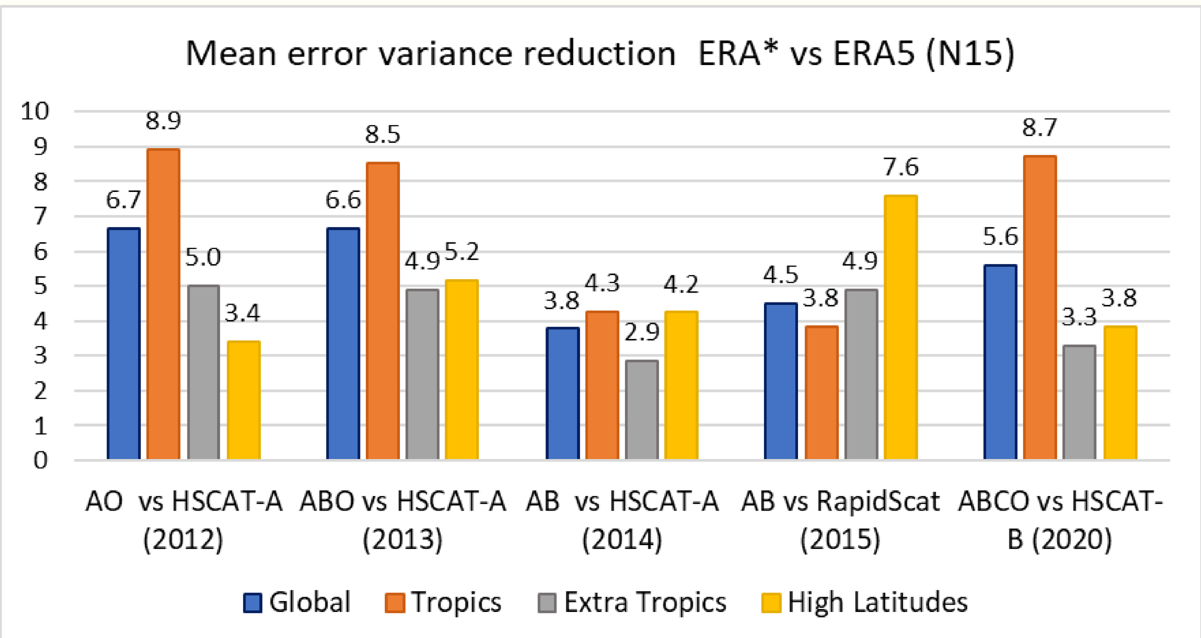


The trained model doesn't require scatterometer observations to produce the corrections and:

- It can be used in operational forecasting;
- It enhances reanalysis stress-equivalent wind products for the periods when scatterometer observations were not available.

State of art: ERA* and SC corrections

A data series correcting for local, persistent NWP stress-equivalent wind biases was produced in the framework of the World Ocean Circulation (WOC) project, which led to the generation of the so-called ERA* dataset [2], for the period 2010-2020. The ERA* product aims to correct persistent, local systematic errors of ERA5 reanalysis with the use of the varying scatterometer constellation. The rationale of the method is that when the scatterometer-NWP wind differences are accumulated over certain periods of time and used to correct for NWP local biases, it is possible to overcome sampling errors and maintain some of the scatterometers most beneficial features, i.e., those related to relatively small-scale ocean processes, such as wind-SST interaction and ocean-current relative winds, and furthermore, correct for the other small- and large-scale NWP parameterization and dynamical errors.



Default configuration: 15-day time window

- Best performance in the tropics (reduction up to 8.9% error variance) [3]
- Globally 3.8 – 6.7% error variance reduction, depending on the available constellation

ERA* method limitations:

- It only corrects local biases persistent over several days.
- It is very sensitive to scatterometer sampling, especially over shorter time windows.
- It doesn't directly show NWP error dependence on both atmospheric and ocean state conditions.
- It has limitations in operational use: computationally expensive and need to shift temporal window (which in turn degrades performance).

Datasets and algorithms

Model input variables:

- ERA5 stress-equivalent winds
- ERA5 mean sea-level pressure, air temperature, specific humidity, SST
- Derivatives wind components and SST gradients
- CMEMS global total surface current

Targets:

- Differences between ASCAT-A 12.5 km winds and ERA5 stress-equivalent wind components.

ML algorithms:

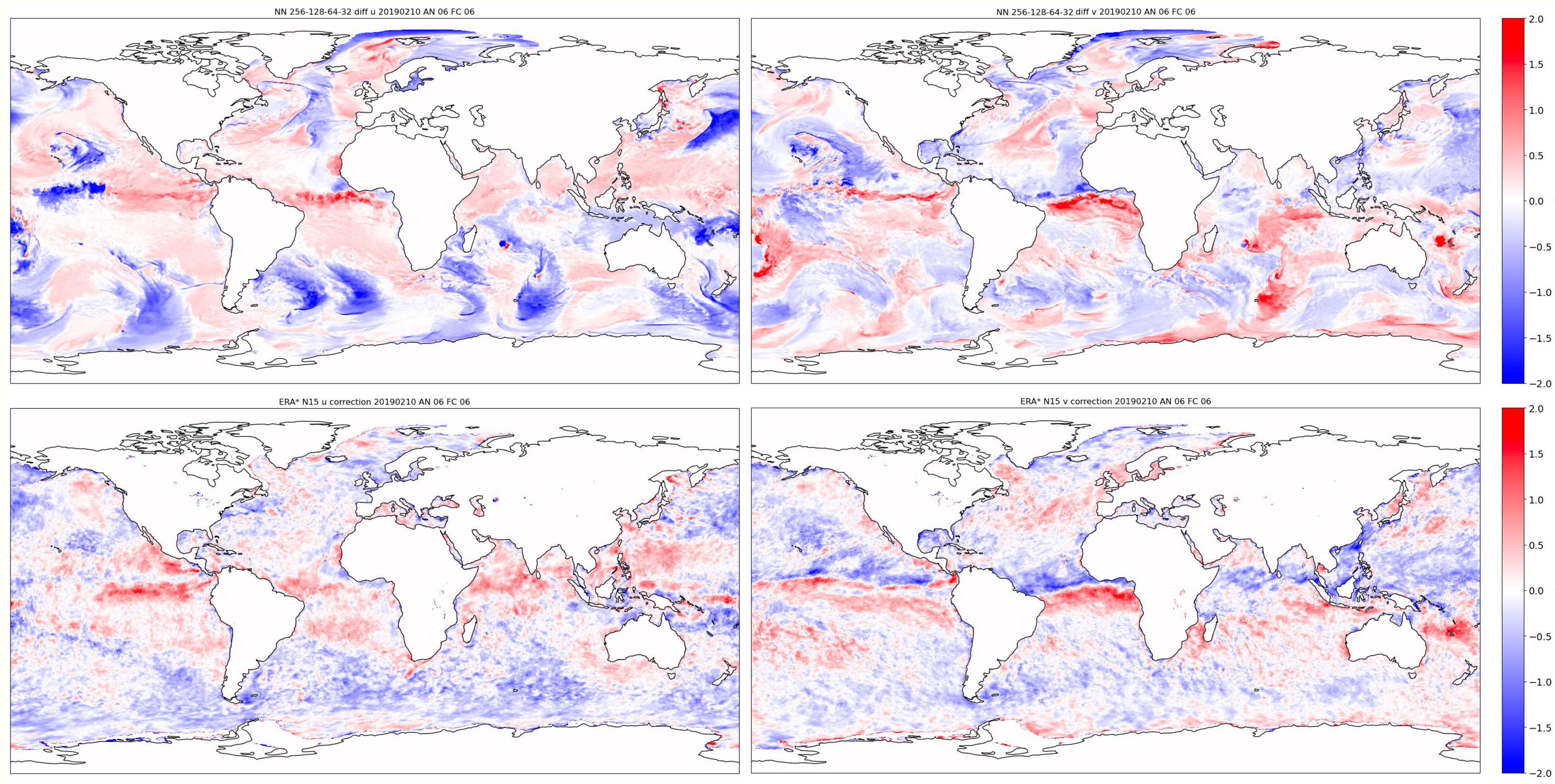
- Extreme Gradient Boosting (XGBoost)
  - → Fast training on GPUs
- Fully-connected neural networks
  - → Several hidden layers, dropout

Training:

- Preliminary models trained on a small subset **02/01 - 06/03 2020**
- Training data was subsampled to **10%** to accelerate the training.

Validation:

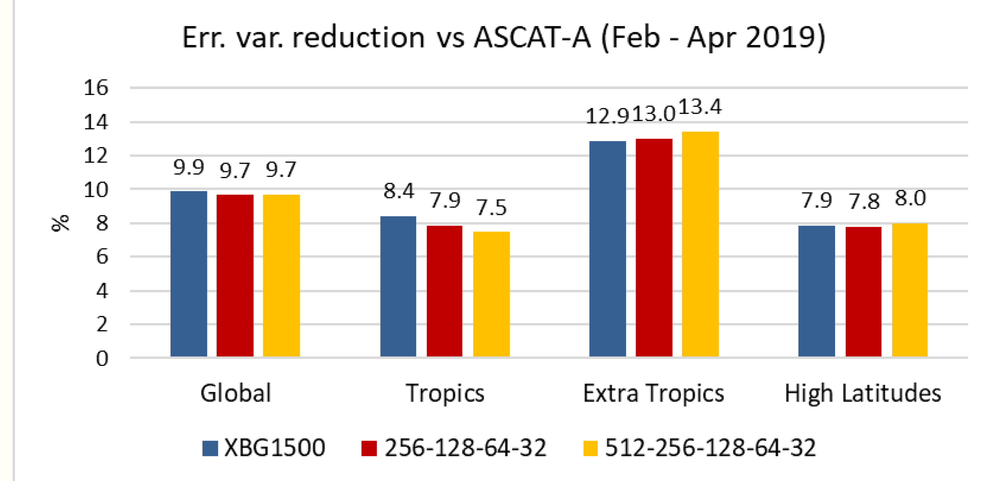
- Similar period but for previous year (**01/02 – 30/04 2019**)
- Against **ASCAT-A** scatterometer
- Against KU-band scatterometer **HSCAT-B**
  - → Temporal coverage **3.5 hours** apart
  - → Sensible to **rain** and higher QC rejection rate, especially in the tropics
- Vector root mean square error (**VRMS**) and **error variance reduction** metrics.



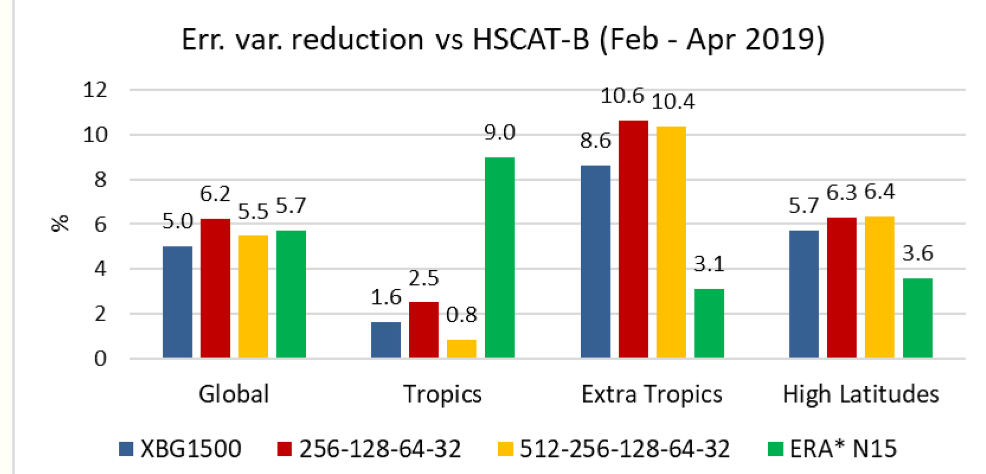
Corrections for u and v-components predicted by NN 256-128-64-32 (top) and ERA* N15 scatterometer corrections (bottom) for 10/02/2019 AN 06 FC 06

Results and discussion

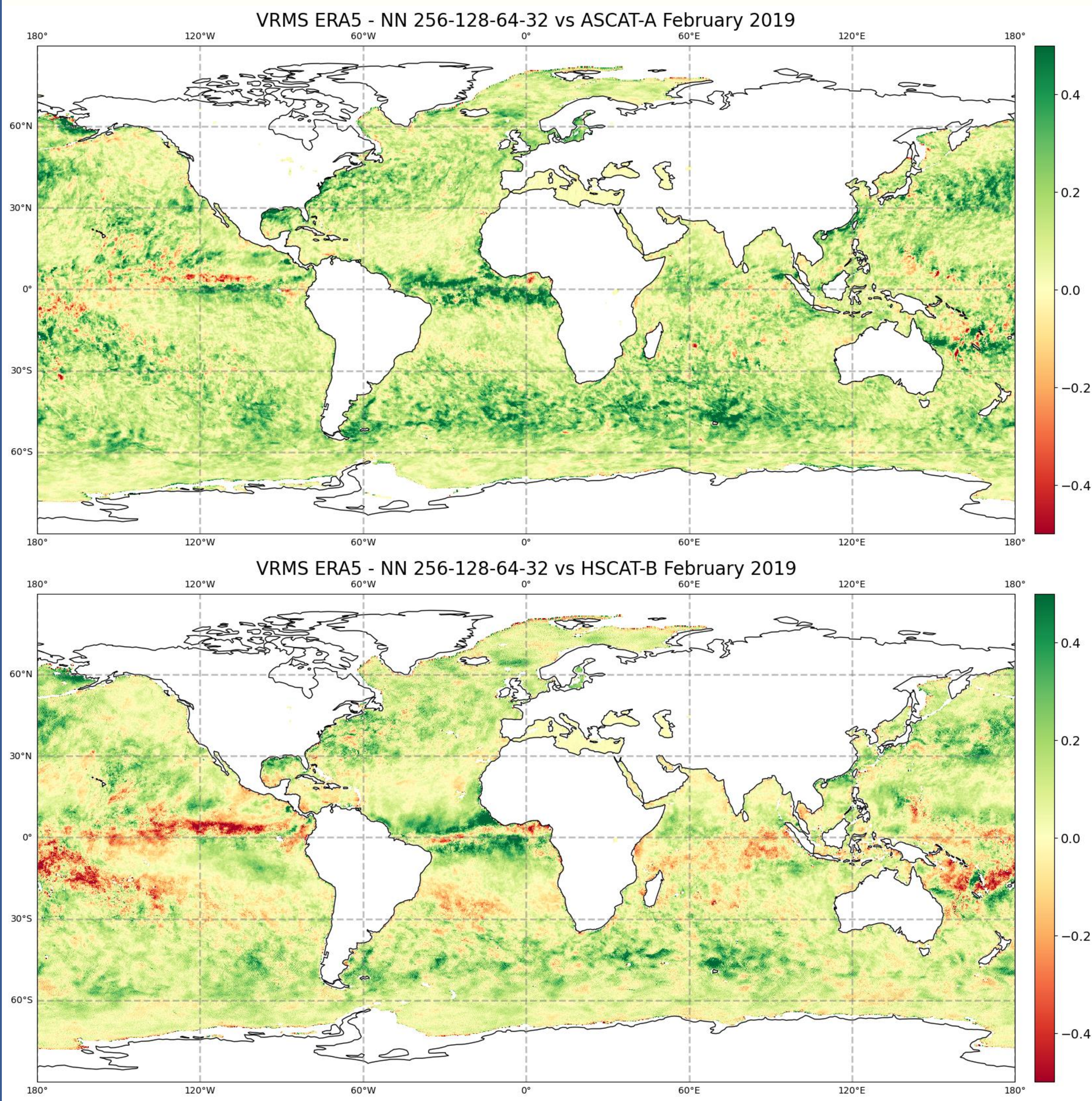
The obtained preliminary ML models, which are only trained on a small subset of data show 9.9% error variance reduction globally vs ASCAT-A and reaching 13% in extra-tropics. ASCAT-A validation shows similar performance in Tropics and High Latitudes with reduction around 8%.



Against HSCAT-B the models show lower performance with up to 6.2% reduction globally. The best performance is still observed in extra tropics with 10.6% reduction; however, tropics show poor results compared to ASCAT-A validation. If compared to ERA* 15-day configuration, which was only validated against HSCAT-B, ML models outperform it in extra-tropics and in high latitudes, but ERA* performs much better in the tropics.



The XGBoost shows slightly lower performance against HSCAT-B but is much faster when training the model.



The plot on the left shows the spatial distribution of the mean VRMS difference between ERA5 and the model output corrected by neural network against ASCAT-A (top) and HSCAT-B (bottom). Green colors show the areas where the NN corrections are reducing ERA5 errors and red colors mark where the NN model has higher VRMS than ERA5. There are significant differences in the validations against ASCAT-A and HSCAT-B, especially in the Inter-Tropical Convergence Zone (ITCZ). When compared to ASCAT-A the NNs generally reduce the errors of ERA5, however when using HSCAT-B as ground truth the performance of the model in the tropics is quite poor.

This can be possibly explained by the high QC rejection rate for HSCAT-B in the presence of rain and the differences due to the diurnal cycle.

HSCAT-B stress-equivalent winds are also much closer to the background model winds used during the inversion which makes the reduction of the errors smaller.

Conclusions and future work

In this preliminary work, we demonstrate that it is possible to reduce ERA5 stress-equivalent wind biases, based only on NWP atmospheric and oceanic output. At this stage, we only use the simplest fully-connected feed-forward neural networks and manually calculate the spatial gradients and derivatives, while future work will include the implementation of the convolutional neural networks architectures (CNNs) that will learn the filters required to extract the spatial relationships from the data.

References:

1. Sandu, I. et al., 2020, "On the causes of systematic forecast biases in near-surface wind direction over the oceans."
2. Trindade, A. et al., 2022 WOC ERA* Hourly Global Stress Equivalent Wind and Wind Stress (V.2.0) [Dataset]
3. Portabella, M. et al., 2022. World Ocean Circulation: Algorithm Technical Baseline Document for ERAstar (v2.0).
4. Makarova E. et al., Proposal and Evaluation of the Machine Learning Models for Correcting ERA5 Stress Equivalent Wind Forecasts as a Function of Atmospheric and Oceanic Conditions. Ocean and Sea Ice SAF : OSI VSA22 01: 1-52 (2022)