

On the use of machine learning to correct NWP model sea surface wind forecasts with scatterometer data input

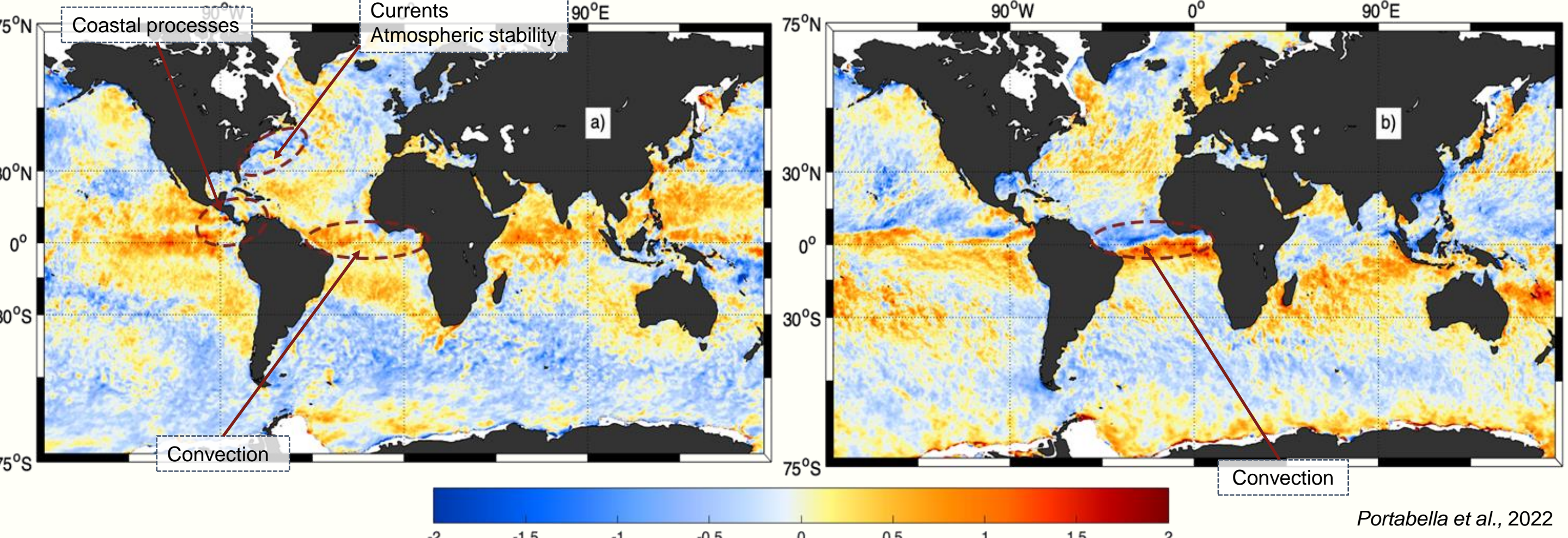
Evgeniia Makarova¹ (makarova@icm.csic.es), Marcos Portabella¹, Ad Stoffelen², Ana Trindade¹

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



Motivation

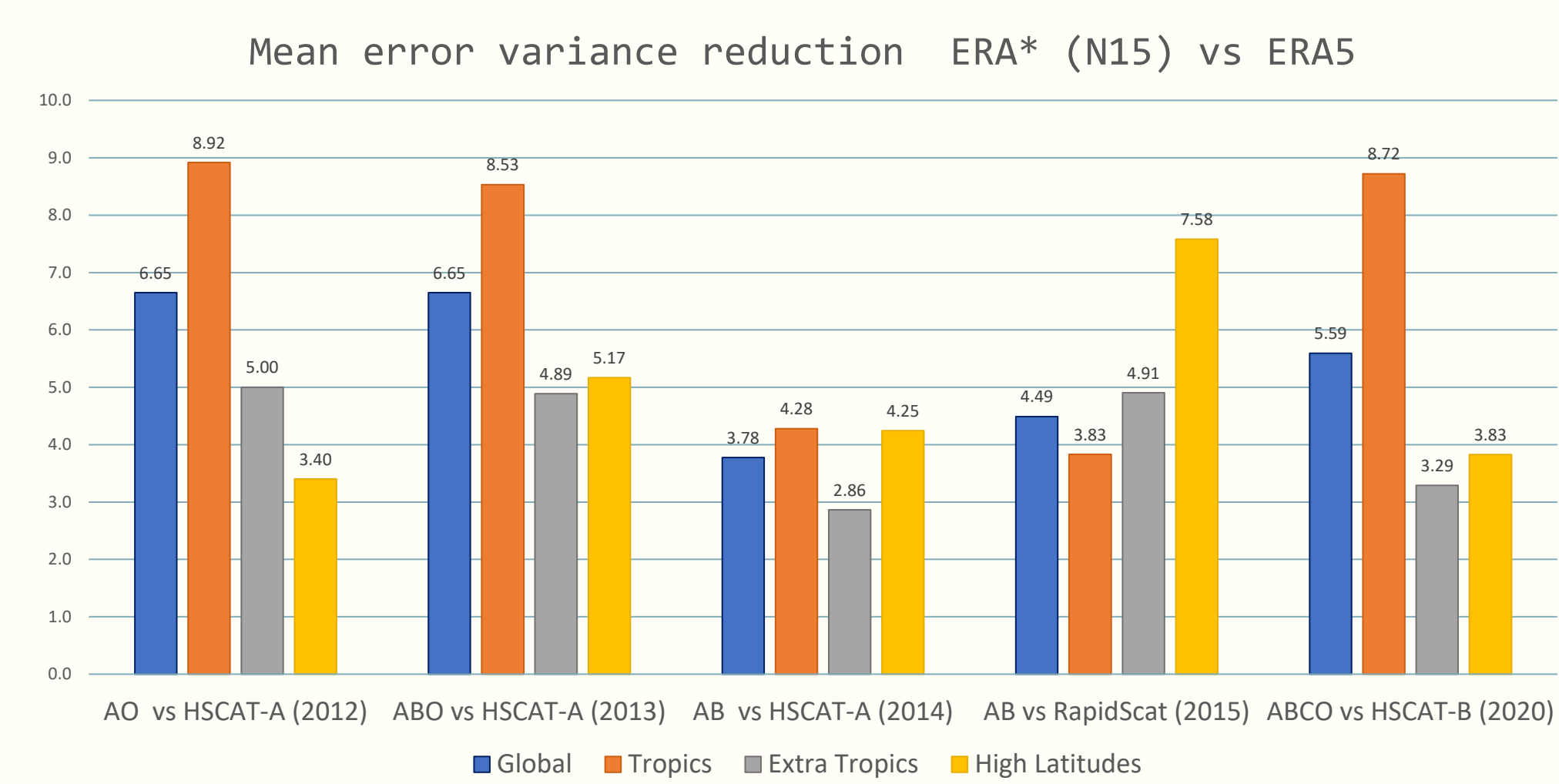
Global Numerical Weather Prediction (NWP) model sea-surface wind output is commonly used to force ocean models due to their time and space continuity. However, the output of the NWP models presents local biases, with one of the most systematic and longstanding biases in the sea surface wind direction [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).

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.



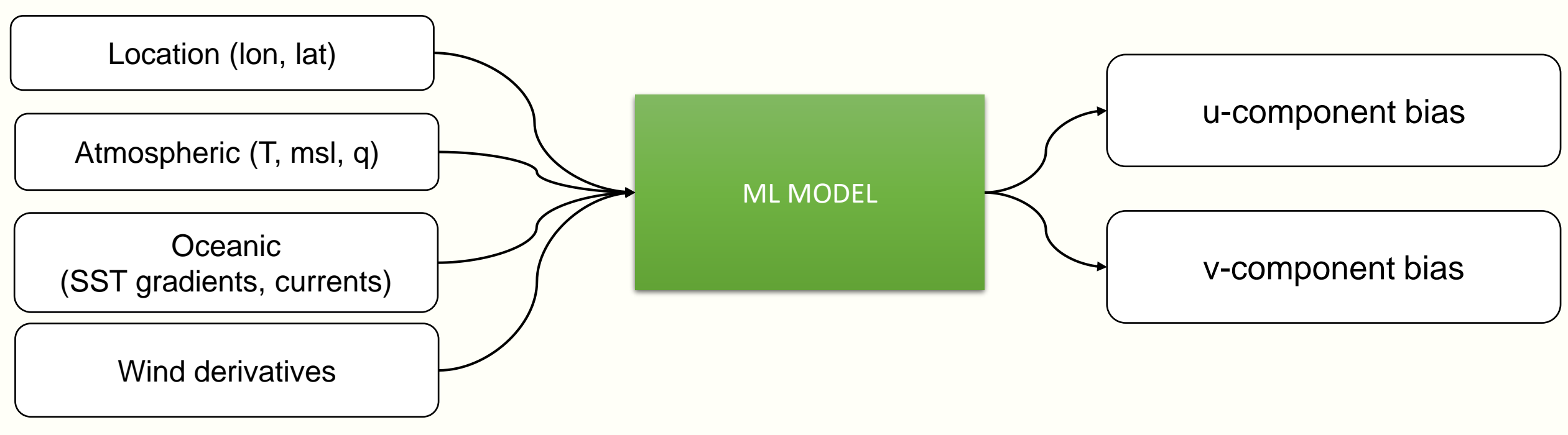
Default configuration: 15-day time window

- 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).

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.



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.

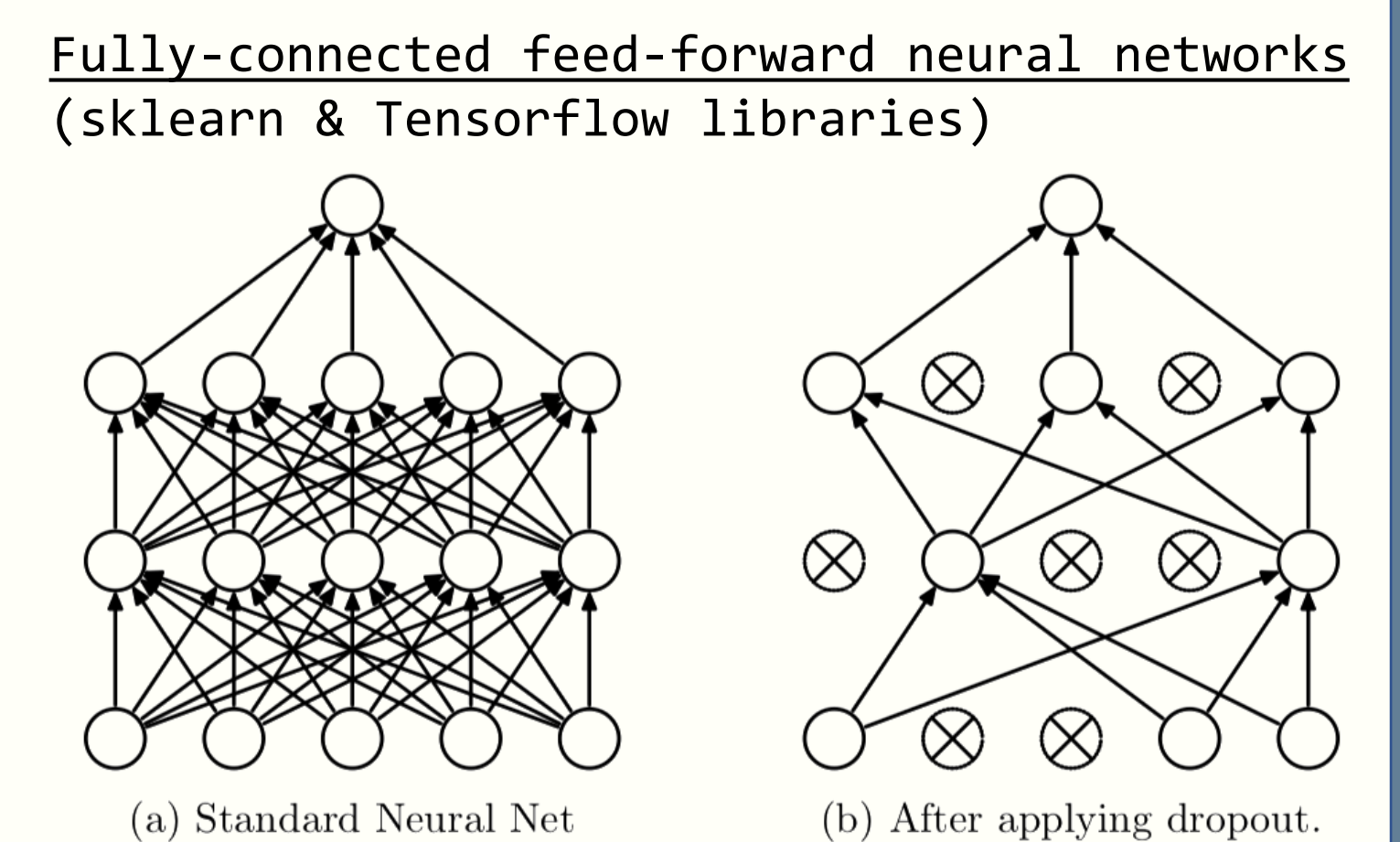
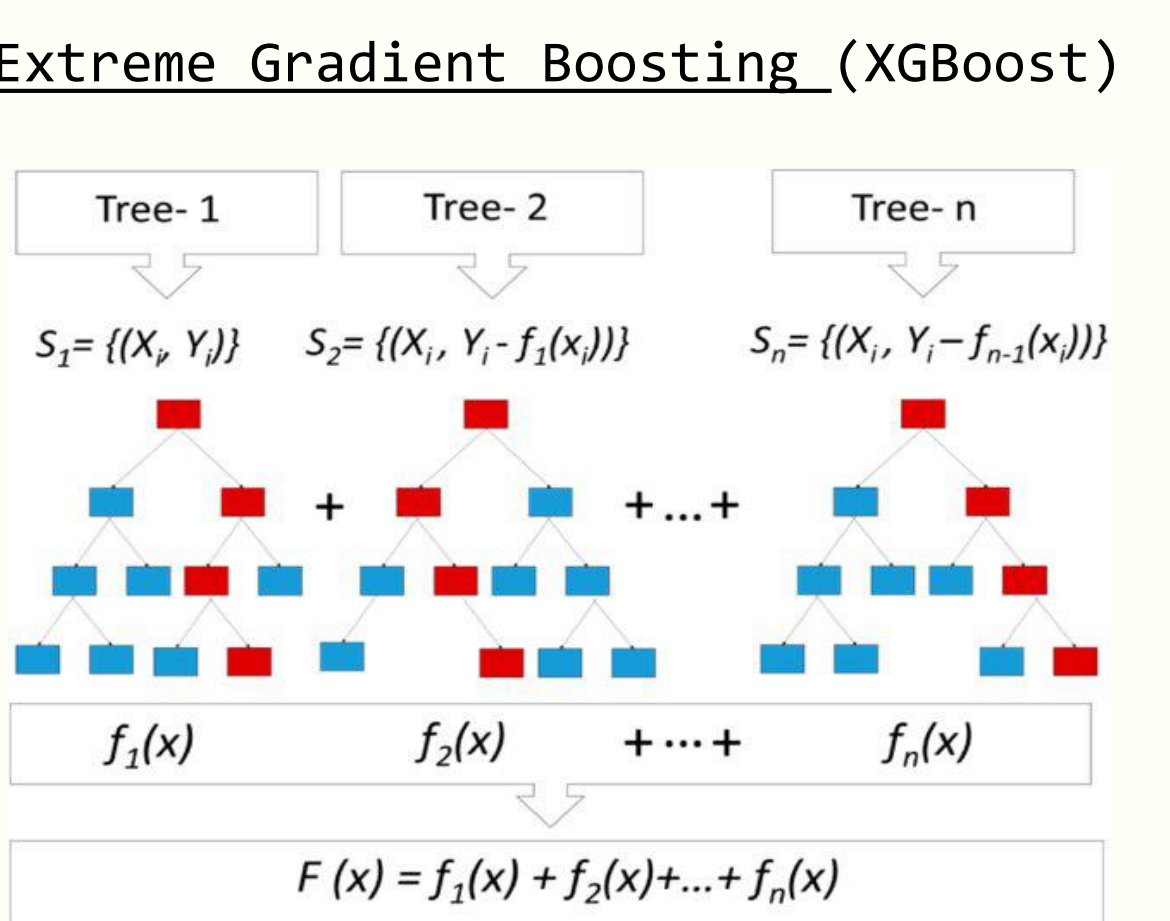
Datasets and algorithms

- Model input variables:**
- ERA5 stress-equivalent wind components, wind speed and direction;
 - ERA5 mean sea-level pressure, air temperature, specific humidity, SST;
 - Derivatives of ERA5 stress-equivalent wind components and SST gradients;
 - CMEMS global total surface current components.

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

- Validation:**
- Resulting model output collocated against independent scatterometer HSCAT-B.

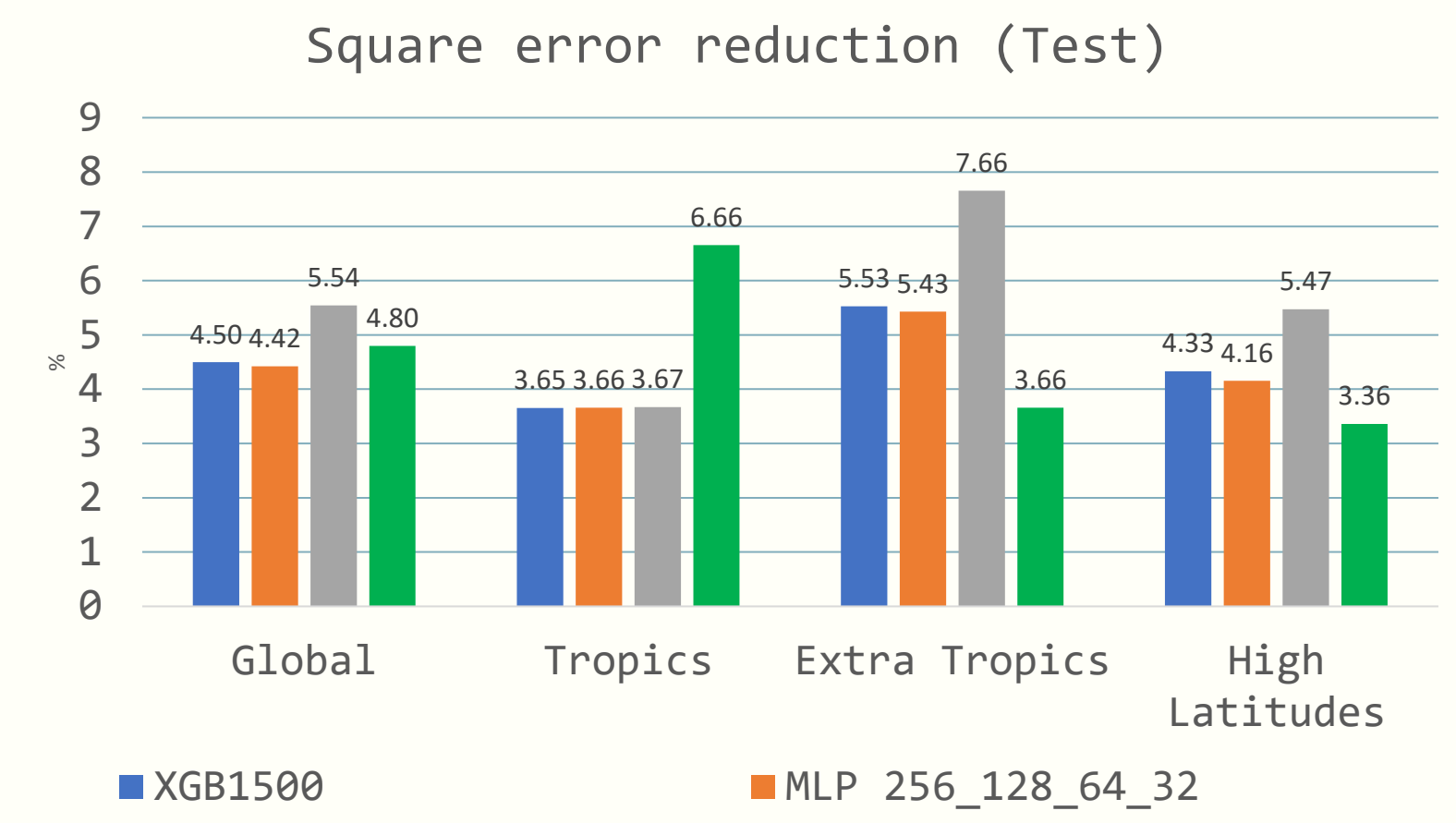
Machine learning algorithms



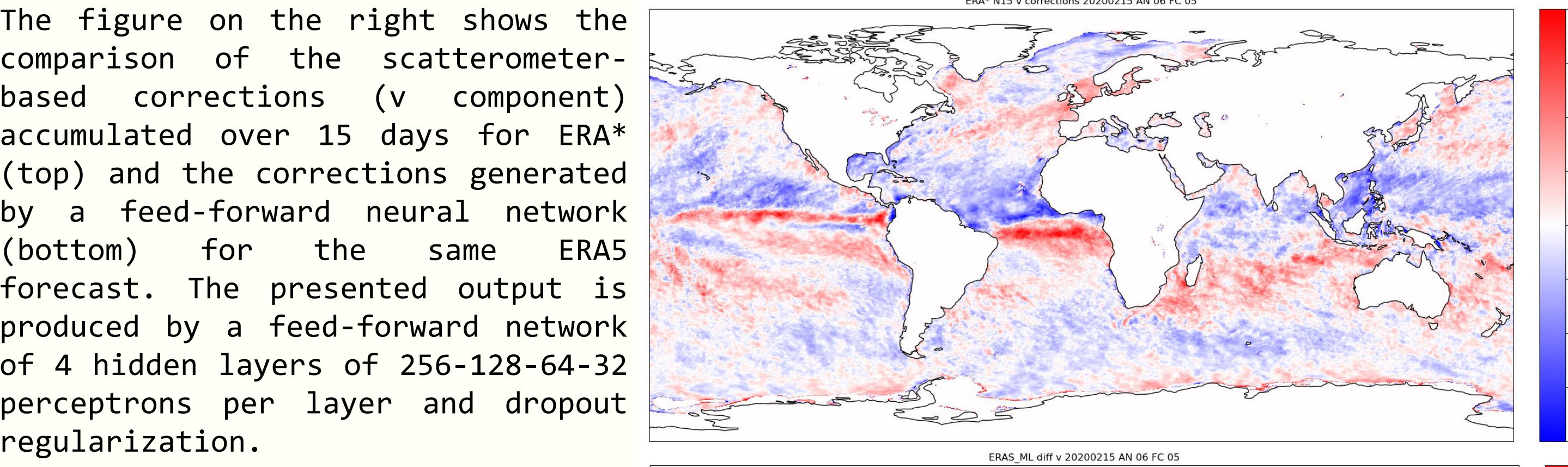
- No normalisation required
- Built-in support for model input feature importance
- Library supports various hyperparameters to avoid the overfitting of the model
- Slow at inference
- Sklearn implementation is very fast but not flexible enough for larger datasets
- Tensorflow allows custom pipelines for training over larger datasets
- Tensorflow supports implementation of custom architectures of neural networks and regularization layers such as Dropout
- Faster at inference than gradient boosting ensembles

Results and discussion

Several ML models are trained over 02/01/2020 - 06/03/2020 period and validated against HSCAT-B. The results are shown for the test period 10/03/2020 - 29/04/2020.

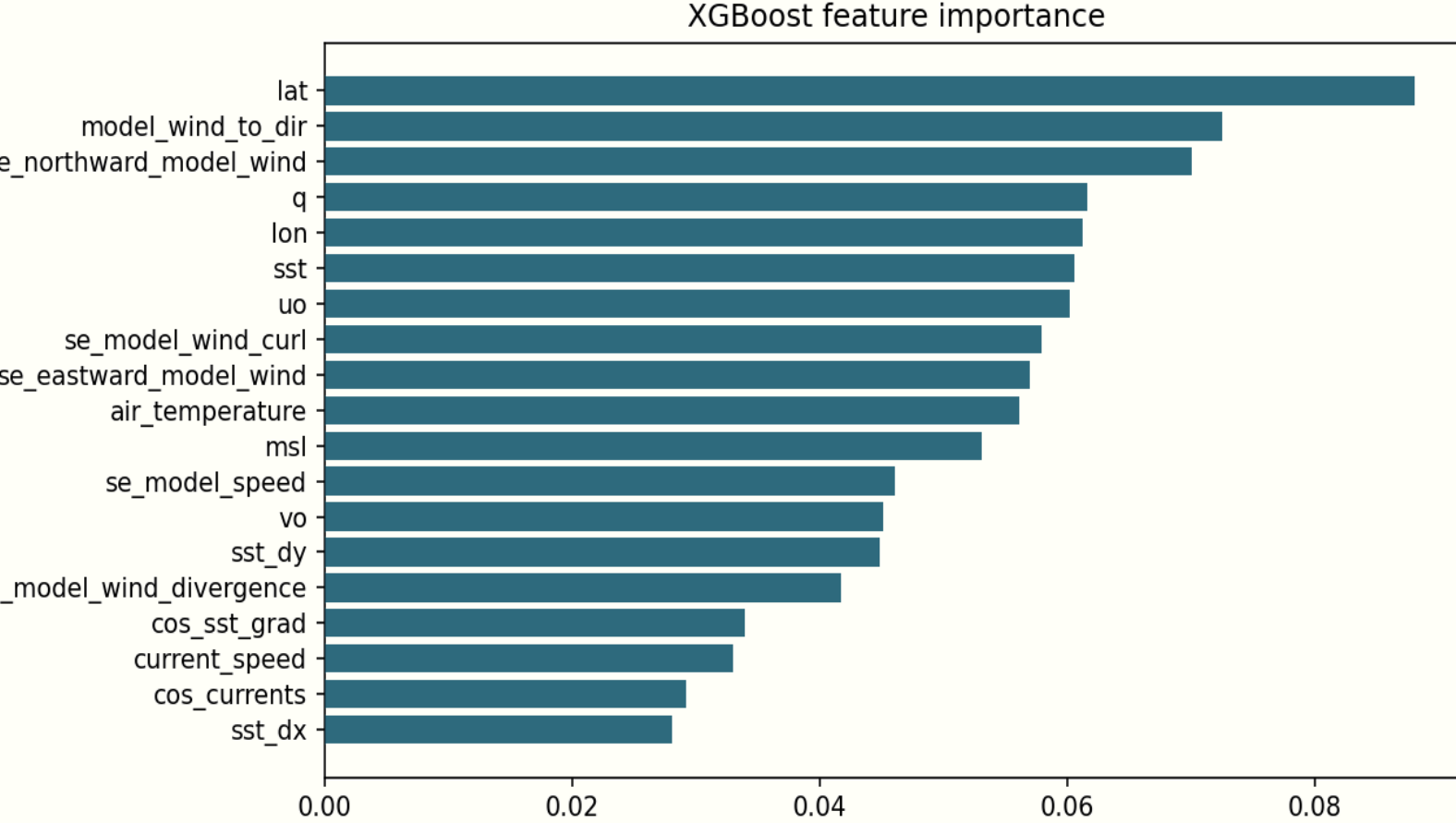


The obtained preliminary ML models, which are only trained on a small subset (5%) of a 70-day dataset, show a reduction of error variance with respect to ERA5 U10S which is globally comparable to the performance of the ERA* N15 baseline product and of significantly higher quality in the extra-Tropics for the 1.5 month period following the training period. When validating the models against the independent HSCAT-B scatterometer, a 5.5% error variance reduction is achieved globally for the test data set, and up to 7.66% in the extra-Tropics [4].



The ML model is able to reproduce similar correction patterns to those in ERA*, but by using only other model variables as input.

The forecasts corrected by ML models however are unable to show increased spatial variance at small scales, compared to ERA* (not shown).



The XGBoost library shows similar performance compared to the implemented feed-forward neural networks, but is several times slower during the inference and thus not suitable for generation of multiple forecasts over larger periods of time. However, it has a built-in feature that allows to assess the overall importance of the input features, which can help in discarding less important inputs and make the ML model more robust.

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. This work shows that neural networks are more suitable for the generation of such predictions on larger temporal scales than decision tree ensembles. 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.

Authors:

Evgeniia Makarova¹ (makarova@icm.csic.es)

Marcos Portabella¹

Ad Stoffelen²

Ana Trindade¹

Affiliations:

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

On the use of machine learning to correct NWP model sea surface wind forecasts with scatterometer data input

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 and need to be corrected since they mostly represent unresolved geophysical processes by NWP models.

The previous approach of correcting the biases was based on the accumulation of the differences between the scatterometer stress-equivalent winds and the NWP model outputs. The resulting 11-year reanalysis dataset ERA* [2] that aims to correct ERA5 biases, in its baseline configuration was able to reduce the error variance about 3-9% compared to ERA5, depending on the ocean region and the scatterometer combination [3].

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 NWP-scatterometer 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 (T_a), relative humidity (rh), surface pressure (sp), as well as SST gradients and ocean currents. The algorithms that are evaluated include two libraries based on Gradient Boosting Decision Trees (GBDT), such as XGBoost and LightGBM, and feed-forward neural networks (FNNs), implemented with the sklearn (MLP Regressor) and with Tensorflow.

Globally, the best performing model is a Tensorflow-based neural network with 4 hidden layers with 256, 128, 64, 32 neurons per layer. It shows a 5.54% of error variance reduction globally, and in particular up to 7.66% in the extra-tropics, when compared to the ERA5 performance (test period) [4]. In the tropics and high latitudes, the error variance reduction is of 3.67% and 5.47%, respectively.

Further improvements with the use of Deep Learning will be explored by adapting the neural network architectures used in computer vision, i.e., for semantic segmentation, super resolution and generative tasks. The best performing model can be further fine-tuned on the operational ECMWF data and used to correct the biases of the operational stress-equivalent wind forecasts.

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," 2020.

2. Trindade, A.; Grieco, G.; Makarova, E.; Portabella, M., 2022, WOC ERA* Hourly Global Stress Equivalent Wind and Wind Stress (V.2.0) [Dataset]; <https://doi.org/10.20350/digitalCSIC/15436>; <http://hdl.handle.net/10261/330890>
3. Portabella, M., Trindade, A., Grieco, G., Makarova, E., 2022. World Ocean Circulation : Algorithm Technical Baseline Document for ERAstar (v2.0). <https://digital.csic.es/handle/10261/287610>
4. Makarova E., Portabella M., Stoffelen A., 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) <http://hdl.handle.net/10261/81991>