Nonlinear Wave Ensemble Averaging using Neural Networks

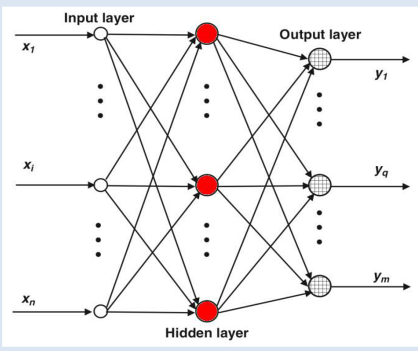
Ricardo M. Campos[IST], Jose-Henrique Alves[SRG-NCEP], Vladimir Krasnopolsky[NCEP], Stephen G. Penny[CIRES]

[IST] Centre for Marine Technology and Ocean Engineering (CENTEC), Instituto Superior Técnico, University of Lisbon
[NCEP] EMC/NCEP/ NOAA Center for Weather and Climate Prediction
[SRG-NCEP] SRG/EMC/NCEP/ NOAA Center for Weather and Climate Prediction
[CIRES] Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder

NA16NWS4680011 - NWS Office of Science and Technology (OST)

1. Introduction

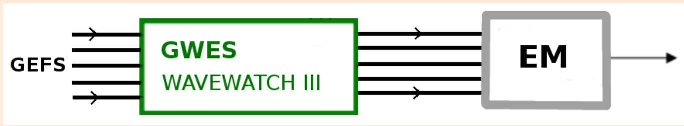
- Post-processing models for non-linear ensemble averaging are developed to reduce the error of wave forecasts, especially at longer forecast ranges.
- 20 ensemble members plus a control member compose the NCEP **Global Wave Ensemble System (GWES)**, forced by the Global Ensemble Forecast System (GEFS) winds on WAVEWATCH III.
- **Neural network models** (NNs) with different architectures are developed to reduce the GWES error by training the NNs using quality-controlled observations from buoys and altimeters.
- Target variables are **10-m wind speed** (U10) from GEFS and **significant wave height** (Hs) from GWES.



2. Neural Network Model

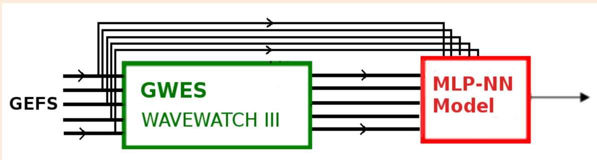
A conservative ensemble approach is currently used to calculate the arithmetic **ensemble mean** (EM) for a variable p in GWES, where n is the number of ensemble members and p_i is the i-th ensemble member. It assumes a linear relationship between the EM and ensemble members

1
$$EM = \frac{1}{n}\sum_{i=1}^{n} p_i$$



However, in reality this relationship may be strongly nonlinear so a neural network model is proposed:

2
$$NEM = NN(p_1, p_2, \cdots, p_n)$$

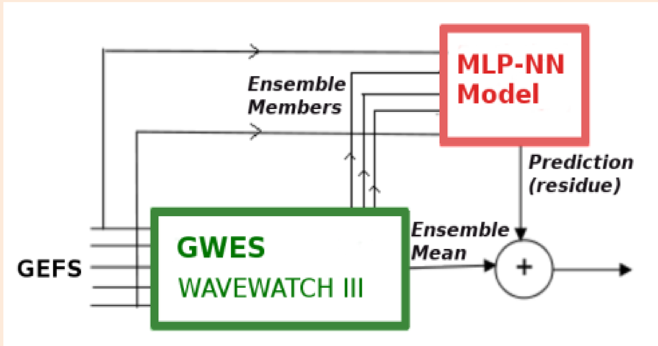


The combination of both methodologies (1 and 2) present better results, using the NN model to simulate the non-linear part of the signal together with the model error, appended to the arithmetic EM. The target variable to simulate is the error signal, or residue, of the arithmetic mean compared to observations:

3
$$\boldsymbol{NEM = EM + NN_r(p_1, p_2, \cdots, p_n)}$$

Our work is divided in three steps:

- NNs applied to single locations (two buoys)
- NNs in the Gulf of Mexico (six buoys)
- **Global NNs trained with altimeter data**



A **multilayer perceptron model (MLP-NN)** with hyperbolic tangent as the activation function is considered with backpropagation (gradient descent) training (Haykin, 1999); where p_i is the input and y_q the output, a and b are the NN weights, n and m are the numbers of inputs and outputs, and k is the number of nonlinear basis functions (hyperbolic tangents)

$$\boldsymbol{NN(p_1, p_2, \cdots, p_n; a, b) = y_q = a_{q0} + \sum_{j=1}^{k} a_{qj}.tanh\left(b_{j0} + \sum_{i=1}^{n} b_{ji}.p_i\right); \quad q = 1, 2, \ldots, m}$$

Sensitivity tests were performed, modifying the number of neurons of the intermediate layer and excluding/including variables, applied to different NN architectures.

Input variables (total of 133):

- Control run plus 20 ensemble members of: **10-m wind intensity** (U10), **significant wave height** (Hs), **peak wave period** (Tp), **mean wave period** (Tm), **significant wave height of wind-sea** (WsH), and **period of wind-sea** (Tws);
- **Latitude** and **longitude** (sin/cos);
- **Time** $\sin(2\pi t/T)$ and $\cos(2\pi t/T)$; and **Forecast time** (0 to 10 days).

Output variables: residue of **U10** and **Hs**

3. Neural Networks training and tests

- NN model evaluation: Cross-validation with 3 cycles. One year (2017) of data divided in training (2/3 of dataset) and test set (1/3). Further assessments of hybrid modeling results analyzing the scatter and systematic components of error (Mentaschi et al., 2013), and the multivariate distribution of it (Campos et al., 2018).

$$NBias = \frac{\sum_{i=1}^{n}(y_i - x_i)}{\sum_{i=1}^{n} x_i} \qquad SI = \sqrt{\frac{\sum_{i=1}^{n}[(y_i - \bar{y}) - (x_i - \bar{x})]^2}{\sum_{i=1}^{n} {x_i}^2}}$$

First test at **single locations**: two pairs of NDBC buoys. Compared to the arithmetic ensemble mean, the NN ensemble averaging allowed an improvement on the 5-day forecast of: 64% in bias; 29% in RMSE and SI; 11% in correlation coefficient. Further benefit of reducing the error at higher percentiles.

Second test in the **Gulf of Mexico** with six buoys. Total of **105,600 NNs**:

- 12 different numbers of neurons [2, 5, 10, 15, 20, 25, 30, 35, 40, 50, 80, 200];
- 8 different filtering windows to reduce noise [0, 24, 48, 96, 144, 192, 288, 480] hours;
- 100 seeds for random initialization;

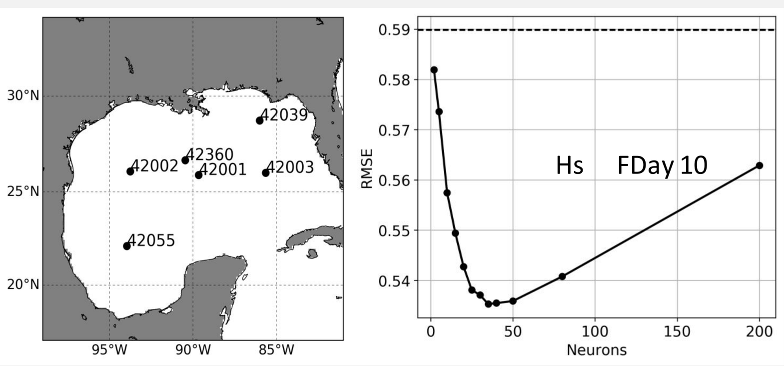


Figure 1: Position of NDBC buoys (left); and RMSE as a function of the number of neurons (right).

- ✓ Small number of neurons are sufficient to reduce the bias, while 35 to 50 neurons produce the greatest reduction in both the scatter and systematic errors.
- ✓ Main advantage of NNs occurs at longer forecast ranges beyond four days.
- ✓ The NNs were able to improve the correlation coefficient on forecast day 10 from 0.39 to 0.61 for U10, and from 0.50 to 0.76 for Hs.

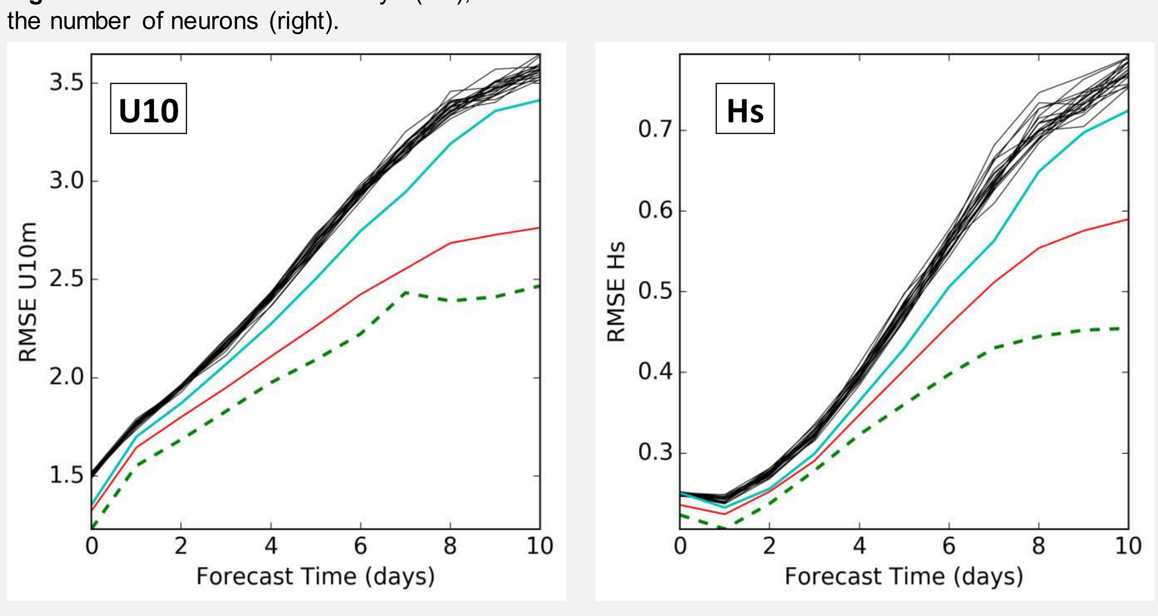


Figure 2: Results of simulations in the **Gulf of Mexico**. Error metrics as a function of the forecast range.

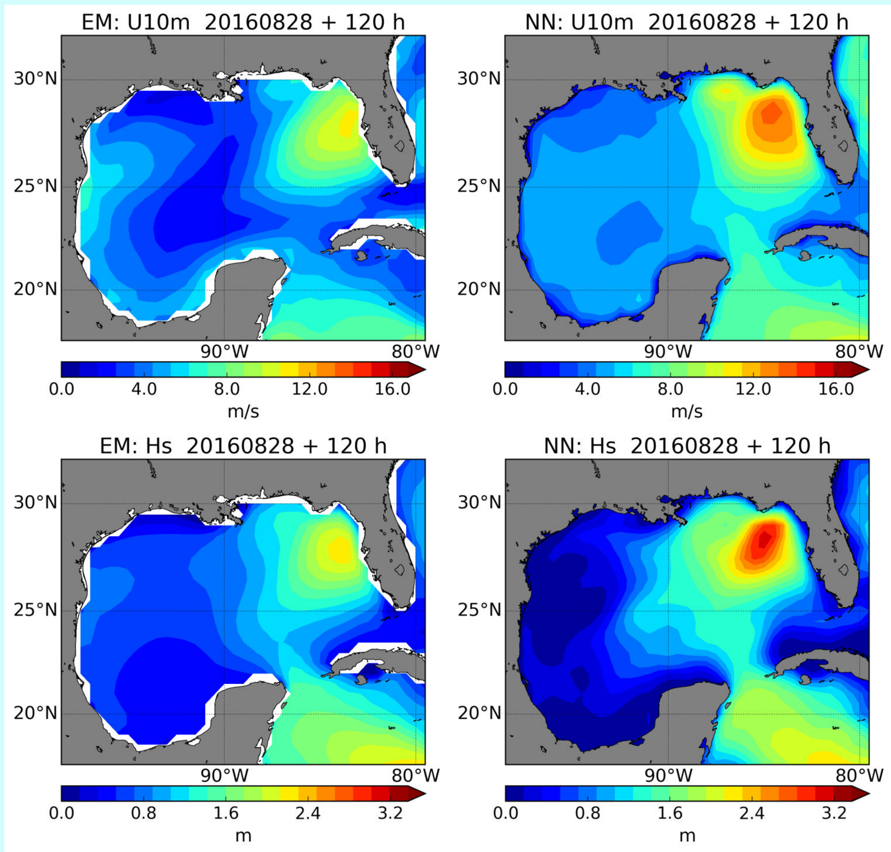
Black curves are the ensemble members,

cyan is the control run,

red is the arithmetic mean of the ensemble members,

dashed-green line is the **nonlinear ensemble average using NNs**.

Applied simulation of NNs in the Gulf of Mexico.



Hurricane Hermine: extreme winds up to 35 m/s and waves up to 6 m high. Extreme winds of the EM dropped from 25 m/s to 10 m/s (severe underestimation) while the NNs better capture the peak of the storms.

Figure 3: Practical implementation of the trained neural networks applied to the Gulf of Mexico in 2016 Comparison between the arithmetic ensemble mean (EM) with the nonlinear ensemble average using NNs. All plots represent the fifth forecast day for **September 2nd, 2016**.

Third test, **Global NNs**, trained with one year (2017) of altimeter data (JASON2, JASON3, SARAL, CRYOSAT2)

- The along-track data are collocated into the regular grid of GWES (weighted average: maximum distance of 25 km and time distance of 0.5 hours for each grid point [Lat/Lon/Time]).
- Total count of 7,521,298 altimeter/GWES matchups;
- For each matchup altimeter/GWES: matrix of 21 ensemble members per 41 forecast instants.

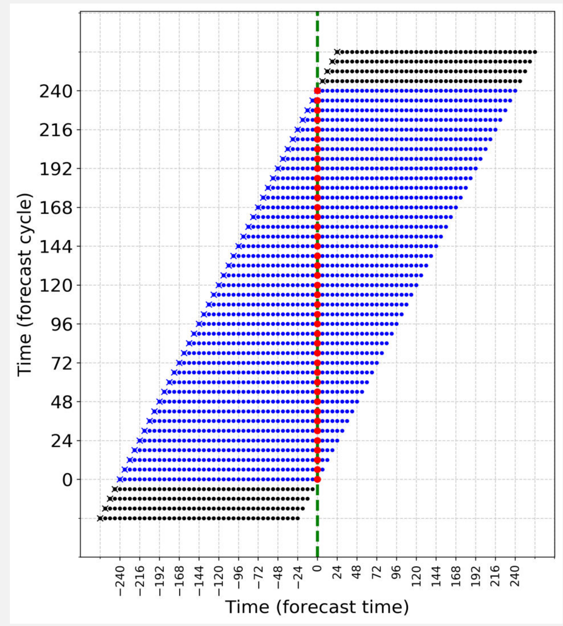


Figure 4: Schematic of time and forecast cycle data selection (both in hours), for a specific time and location of the observation, centered at the satellite time (green dashed line).

New batch of NN tests (**780** NNs trained in high-performance computer; DT2,UMD):

- 26 different number of neurons (2 to 500);10 seeds (initialization); and 3 equally-divided datasets.

Main Goal: evaluate the capacity (and complexity) of a single MLP-NN model to improve the GWES ensemble average for the whole globe and forecast ranges (0 to day-10).

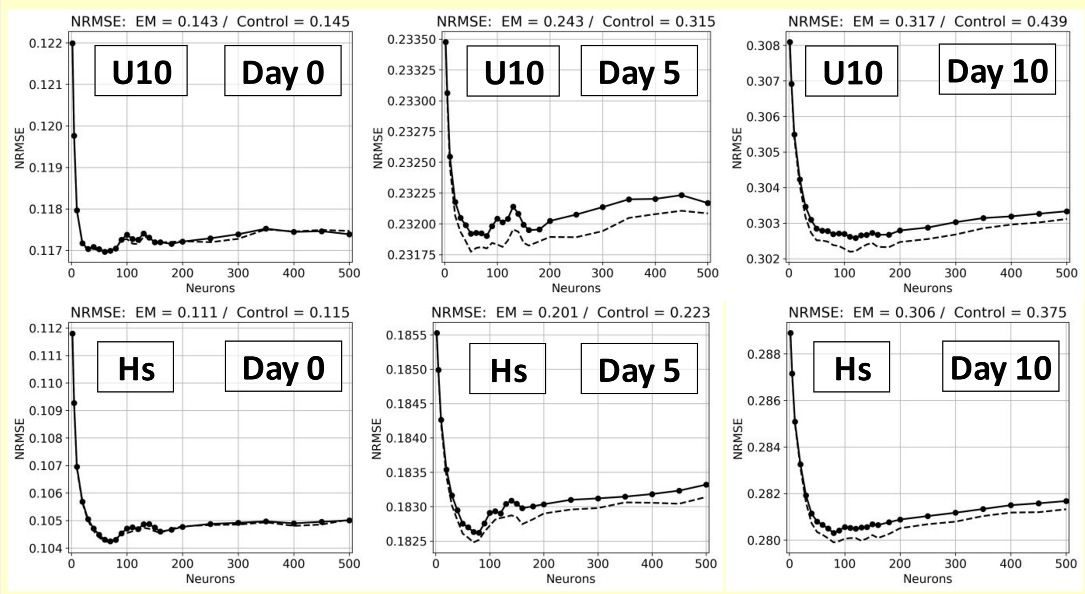
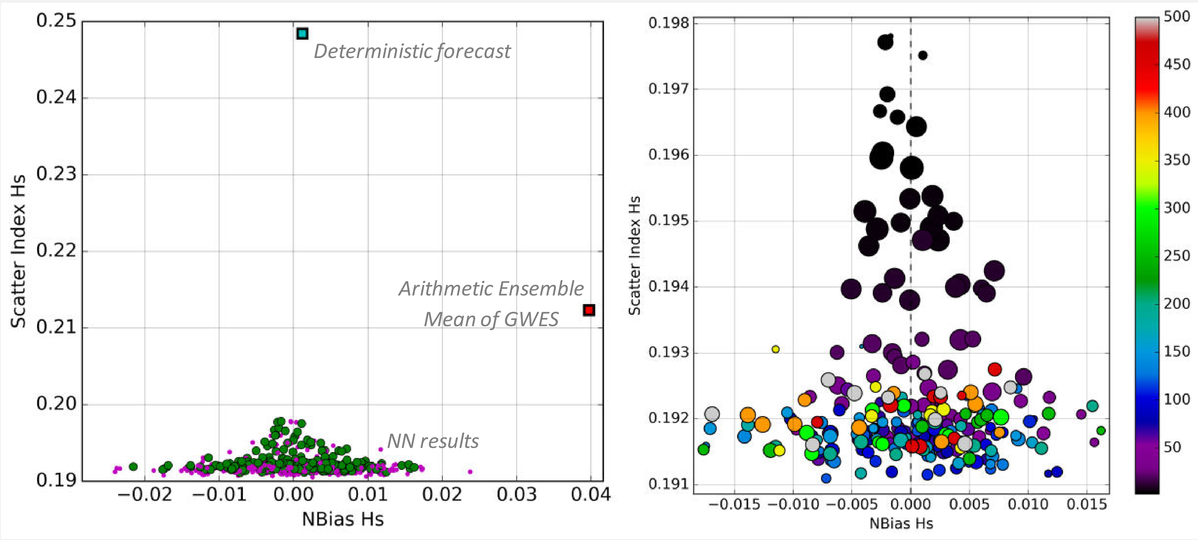
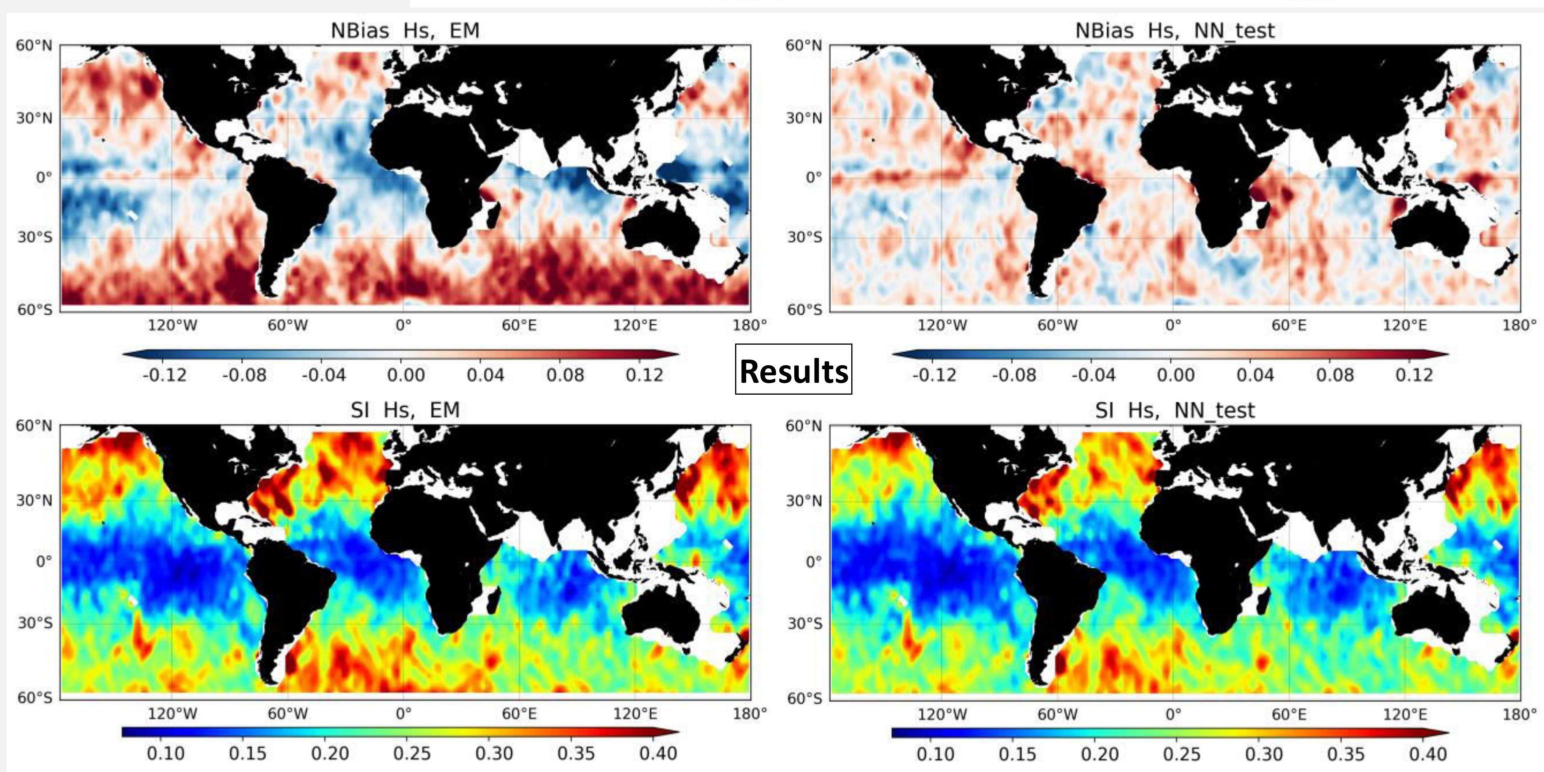


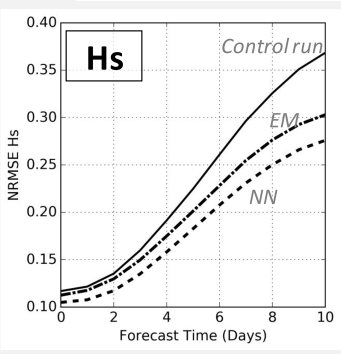
Figure 5: NRMSE as functions of the number of neurons, for 3 different forecast horizons. The dashed line is the average (over different seeds) result for training set while solid line is the results for test set. On top of each plot are the same error metrics for the GWES control run and the arithmetic ensemble mean (EM).

Figure 6: Results of NN tests in terms of the scatter error (y-axis) and systematic error (x-axis). The left plot present the results (training set in magenta and test set in green) compared to the control run (red square) and the arithmetic EM (cyan square). The right plot is a magnification of the clouds of NN results on the test set, where the color indicates the number of neurons and the size of the dots indicates the normalized standard deviation of scatter error throughout different forecast ranges.

- NN improvement and required complexity (k) are different over the variables (U10, Hs) and forecast ranges.
- Despite different performances, all NN models provide better results than the deterministic forecast and the arithmetic ensemble mean (EM).







Nonlinear ensemble averages using NNs are better than the arithmetic EM for the whole globe and range of forecast, including U10 and Hs.

- NN models with few neurons are able to reduce the systematic bias for short-range forecasts, while NNs with more neurons are required to minimize the scatter error at longer forecast ranges.
- 60 to 150 neurons produce the best results for both the scatter systematic errors.

The RMSE of NN for forecast day10 is similar to the EM of day8 and the control run of day6.

4. Conclusions

- Multilayer perceptron neural networks are able to calculate nonlinear ensemble averages with lower systematic and scatter errors than the traditional arithmetic ensemble mean.
- The novel method shows that one single NN model with 140 neurons is able to improve the error metrics for the whole globe while covering all forecast ranges analyzed.
- The RMSE of day-10 forecasts from the NN simulations indicated a gain of two days in predictability when compared to the arithmetic ensemble mean of GWES (EM).
- The time-consuming step is the NN training only. The operational calculation of the nonlinear ensemble averages takes a few seconds. Results of NN training consists of two matrices of weights and two vectors of bias, and it is easily re-trained when necessary.

- Campos, R.M., Krasnopolsky, V., Alves, J.H.G.M, Penny, S.G., 2019. Nonlinear Wave Ensemble Averaging in the Gulf of Mexico using Neural Networks. Journal of Atmospheric and Oceanic Technology, 36, 113-127.
- Campos, R.M., Krasnopolsky, V., Alves, J.H.G.M, Penny, S.G., 2019. Improving NCEP's Global-Scale Wave Ensemble Averages Using Neural Networks. Submitted to Ocean Modeling.