NeurOCAST: A Neural Operator-Based Model for Bias Correction in Forecast Guidance

Atieh Alipour[1,2] (atieh.alipour@noaa.gov), Fariborz Daneshvar[1,2], Soroosh Mani[1], Gregory Seroka[1], Lei Shi[1], Zizang Yang[1], Yuji Funakoshi[1], Georgios Britzolakis[1], Edward Payson Myers[1], Kamyar Azizzadenesheli[3], and Saeed Moghimi[1]

[1]National Oceanic and Atmospheric Administration/National Ocean Service (NOAA/NOS) - Office of Coast Survey (OCS); [2]Ocean Associates, Inc (OAI); [3]NVIDIA Research, USA

Abstract

- Numerical models used for weather and ocean forecasting often produce inaccurate results due to missing physics in the model and errors and biases from their inputs.
- NeurOCAST is a deep learning model designed to correct these biases; we apply it to NOAA's STOFS-2D-Global water level forecast guidance.
- It excels at learning from limited data and applying that knowledge to a wider grid, a new approach for ocean forecasting.
- NeurOCAST is more than 20 times faster than similar implementation of the neural operator and reduces forecast bias by over 80%, leading to better water level predictions for coastal safety and navigation.

Neural Operators

- Classical neural networks focus on learning mappings between finite-dimensional Euclidean spaces or finite sets.
- Neural operators map between infinite-dimensional function spaces, extending the scope of classical neural networks.
- Neural operators are formulated as compositions of linear integral operators and nonlinear activation functions.
- A key advantage of neural operators is their discretization-invariance, allowing input and output functions to be discretized on arbitrary meshes without retraining the model.
- Neural operators replace finite-dimensional linear layers in neural networks with linear operators in function spaces.

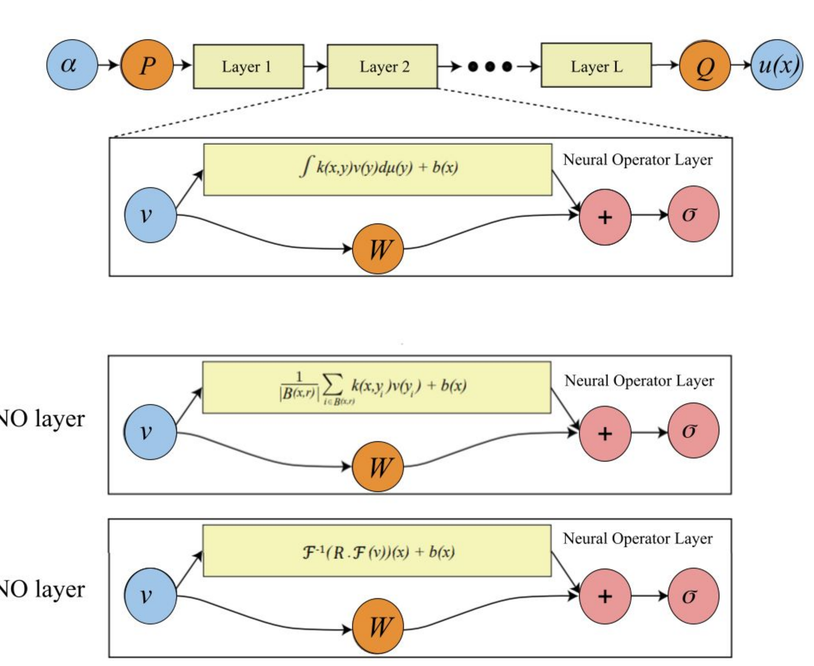

Figure 1. Schematic of the Neural Operator Architecture. The input function a a is first processed by a pointwise lifting operator P, followed by L layers consisting of integral operators and pointwise non-linearity operators σ. Finally, the pointwise projection operator Q produces the output function u.

STOFS-2D-Global

Figure 2. Locations of water level observation stations across the Northeastern United States. Blue stations were used for testing the model, while red stations were used for training.

A difference has been observed between the STOFS-2D-Global simulated total water level and the observed one. The observed total water level results from several factors, including rainfall, storm surge, river discharge, groundwater, waves, tides and other influences. Accurately simulating the total water level requires considering a variety of factors and processes, along with model coupling. However, these factors and processes are often simplified in the model. As a result, the simulated total water level may benefit from bias correction to improve its accuracy.

Framework

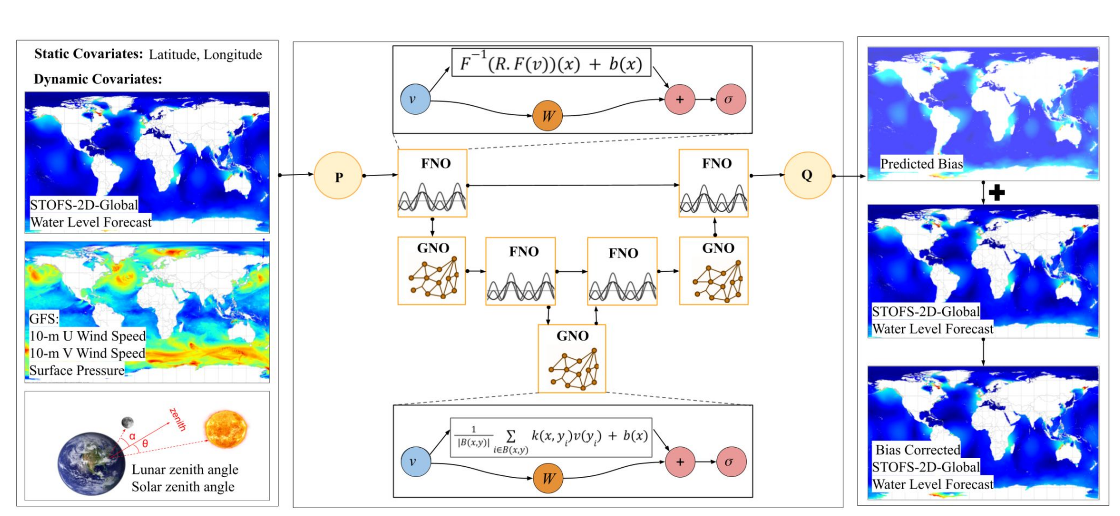

Figure 3. Neural Operator architecture. The model consists of multiple Fourier Neural Operators (FNO) and Graph Neural Operators (GNO) layers that are sequentially connected and repeated.

Results

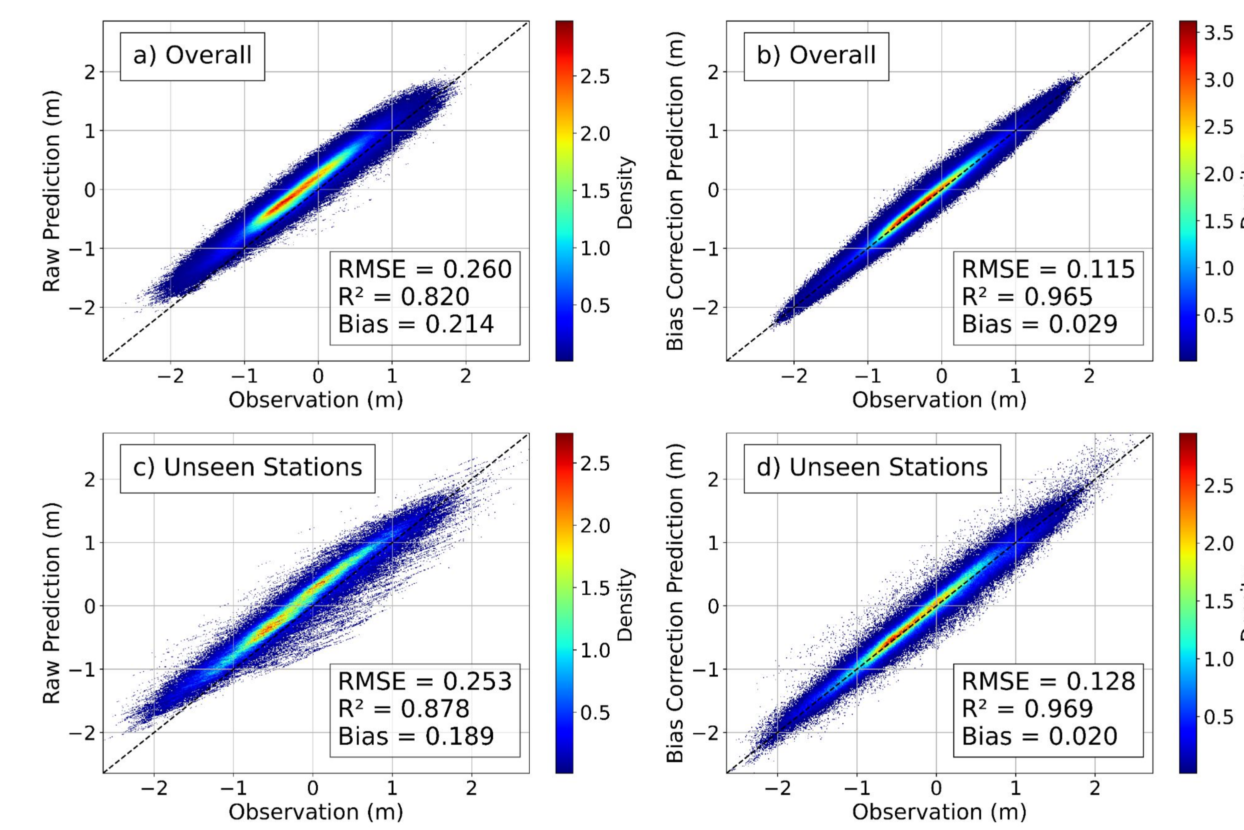

Figure 4. Scatter plots of the raw (left) and bias corrected (right) model outputs, against observations for the year of 2024, across overall (all stations) and at unseen stations (blue stations in Figure 2).

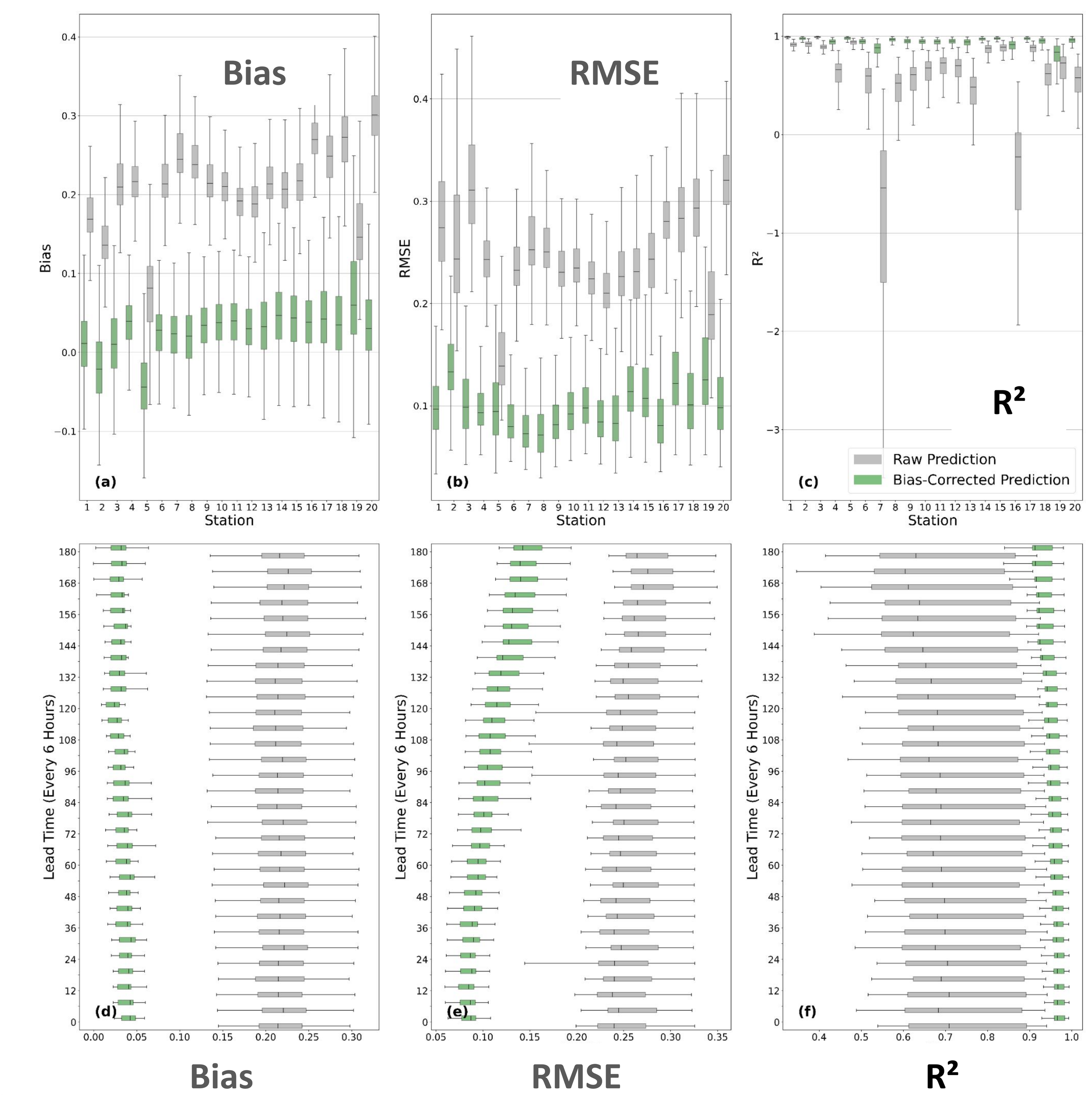

Figure 5. Box plots showing the performance of the bias correction model at different locations and forecast lead times. Subplots a-c show the distribution of bias, RMSE, and R² values across 20 stations for before (gray) and after (green) bias correction. Similarly, subplots d-f display the change in these metrics for different lead times (in 6-hour intervals up to 180 hours).

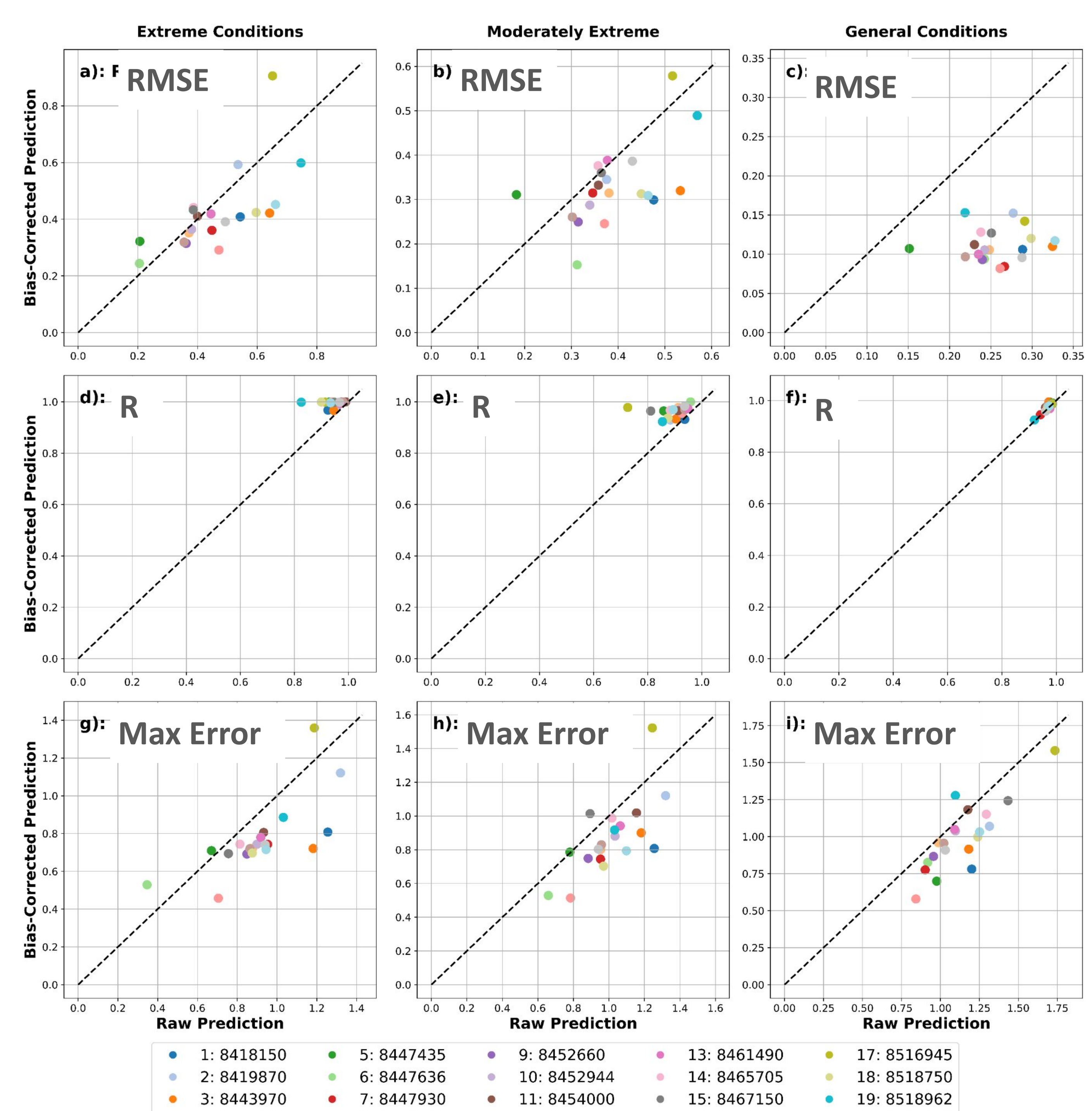

Figure 6. Scatter plots showing the performance of raw (x-axis) and bias-corrected (y-axis) predictions for general conditions and extreme and moderately extreme event categories (the 99.9th percentile and the 99.99th percentile of water level forecast guidance). Each point represents the aggregated performance for one station.

- We trained the model for a three-years timespan of 2021 to 2023 and then tested it on over 3 million data points from 2024.
- We used data from seventeen observational stations (shown in red in Figure 2) to train the model and held out three stations (shown in blue in Figure 2) for testing. These held-out stations, here referred to as unseen stations.
- To evaluate the model's performance under both general conditions and extreme scenarios, we selected events with water levels exceeding the 99.9th and 99.99th percentiles as an example of moderate extreme, and extreme water level/weather conditions, respectively.