

Abstract

Numerical weather prediction (NWP) models provide essential weather forecasts but still exhibit systematic errors due to uncertainties in physical and dynamical processes. The NCEP EMC recently developed a U-Net-based bias correction model that incorporates a convolutional block attention module to enhance feature representation through attention mechanisms. Initially, the method was applied to reduce bias in operational GFSv16 2-m temperature (T2m) over the CONUS domain and promising results were achieved. This study aims to extend the bias correction framework to multiple GFSv16 variables over the global domain, including T2m, 2-m dew point temperature (DPT2m), maximum/minimum 2-m temperature (Tmax/Tmin), the eastward/northward component of the 10-meter wind speed (U10/V10) and 10-m wind speed (WSPD) and direction (WDIR). In addition, the study refined variable-specific model architectures to enhance their robustness and generalization capabilities.

Evaluation indicated that the models significantly improve the annual and seasonal mean GFS forecasts in terms of mean bias, mean absolute bias (MAE) and root-mean-square error (RMSE) over the global domain and different sub-regions. Notably, models trained on limited lead times (e.g., 72 h for T2m) generalized effectively across the full range of 6-384 h. Overall, the models achieved reductions in global-averaged annual mean RMSE by 20.3% for T2m, 26.3% for DPT2m, 8.3% for Tmax, 8.9% for Tmin, 12.8% for U10, 14.4% for V10, 20.3% for WSPD and 18.6% for WDIR at the 72 h lead time, with variable improvements across different lead times. The results highlighted the effectiveness of the proposed deep learning-based bias correction for global NWP forecasts.

Model Architecture

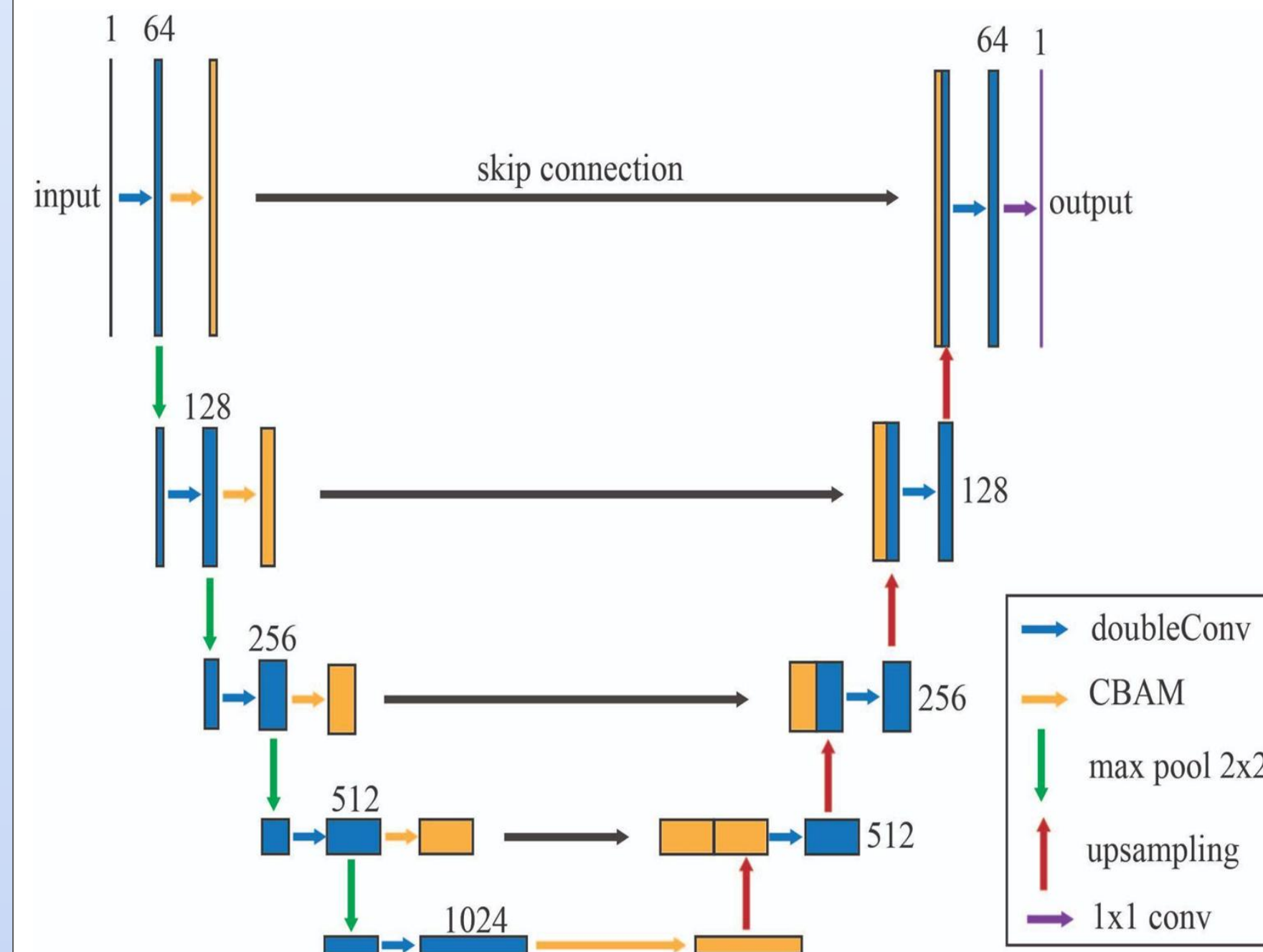


Fig.1 The architecture of BC-Unet. Arrows represent the operation pass. Blue boxes are two 3-by-3 convolutional layers. Numbers above blue boxes are the number of convolutional channels. The black arrows represent layer concatenations.

- **U-Net** is a convolutional neural network (CNN) architecture
- **Encoder**: extracts multi-scale spatial features
- **Decoder**: facilitates precise localization through successive upsampling and convolution operations
- **Skip connections**: directly transfer high-resolution feature maps from the encoder to the decoder
- **CBAM**: enhance the network's representational capacity by dynamically focusing on relevant features

Model Training & Evaluation

- **Data Period**:
 - Training: 23 March 2021 to 28 February 2024 (80% for training and 20% for validation)
 - Testing: 1 March 2024 to 31 March 2025
- **Target Value**:
 - T2m, DPT2m, U10, V10, WSPD - ERA5
 - Tmax and Tmin - GDAS
- **Settings**:
 - 3×3 convolutional kernels with batch normalization
 - activation function: rectified linear unit (ReLU)
 - adam optimizer with default parameters
 - loss function: mean squared error (MSE)
 - batch size: 8
 - maximum epochs are 100 with an early stopping strategy (30 epochs)
 - initial learning rate was set to 0.001 and was reduced by a factor of 0.1 if the validation loss did not decrease for four epochs
 - varying depths: 4, 5, and 6 downsampling/upsampling blocks
 - model training was initially conducted using a single forecast hour (FH72), and at other lead times if necessary
 - for the temperature-related variables, single-task U-Net models were trained individually; for the wind-related variables, single-task, two-task, and four-task U-Net models were explored.
- **Evaluation Metrics**:
 - Bias, mean absolute error (MAE) and root-mean-square error (RMSE)
 - Skill score (SS) as measured in terms of the RMSE

$$SS = (RMSE_{GFS} - RMSE_{BC}) / RMSE_{GFS} \times 100\%$$

Model Evaluation

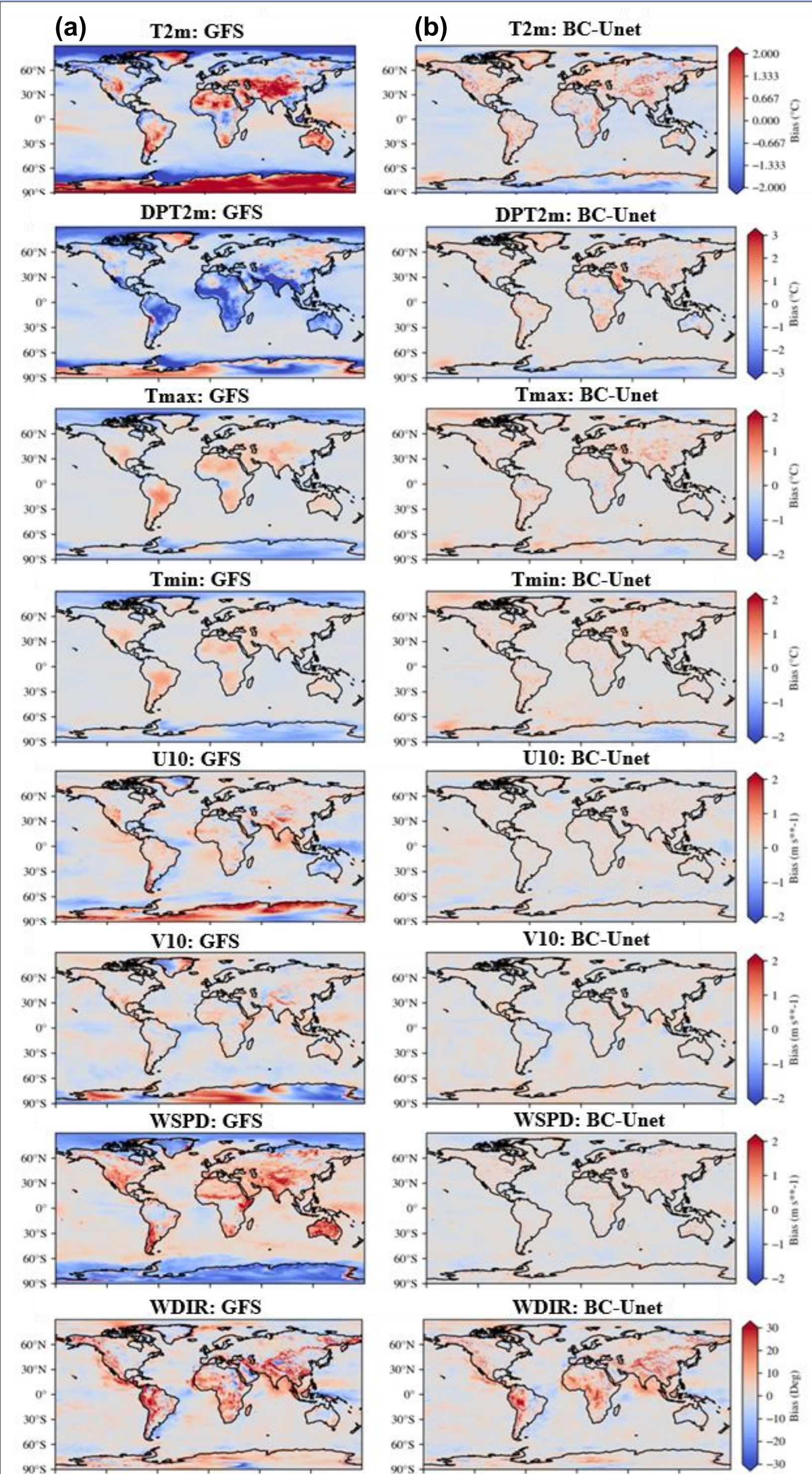


Fig.2 Annual mean bias @ FH72 for (a) original and (b) corrected GFS

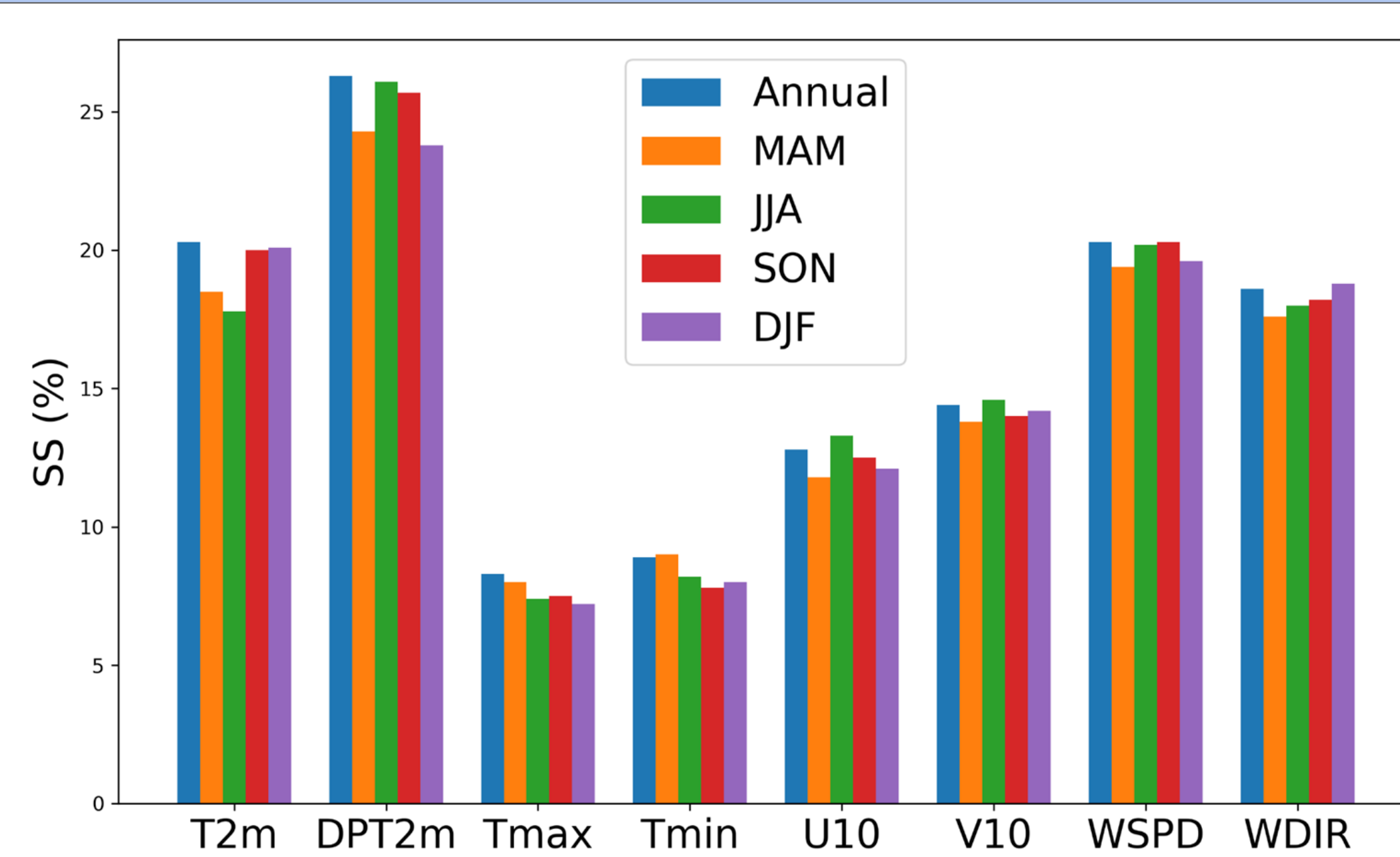


Fig.3 Skill score (SS) for annual and seasonal mean corrected GFS variables @ FH72 over global

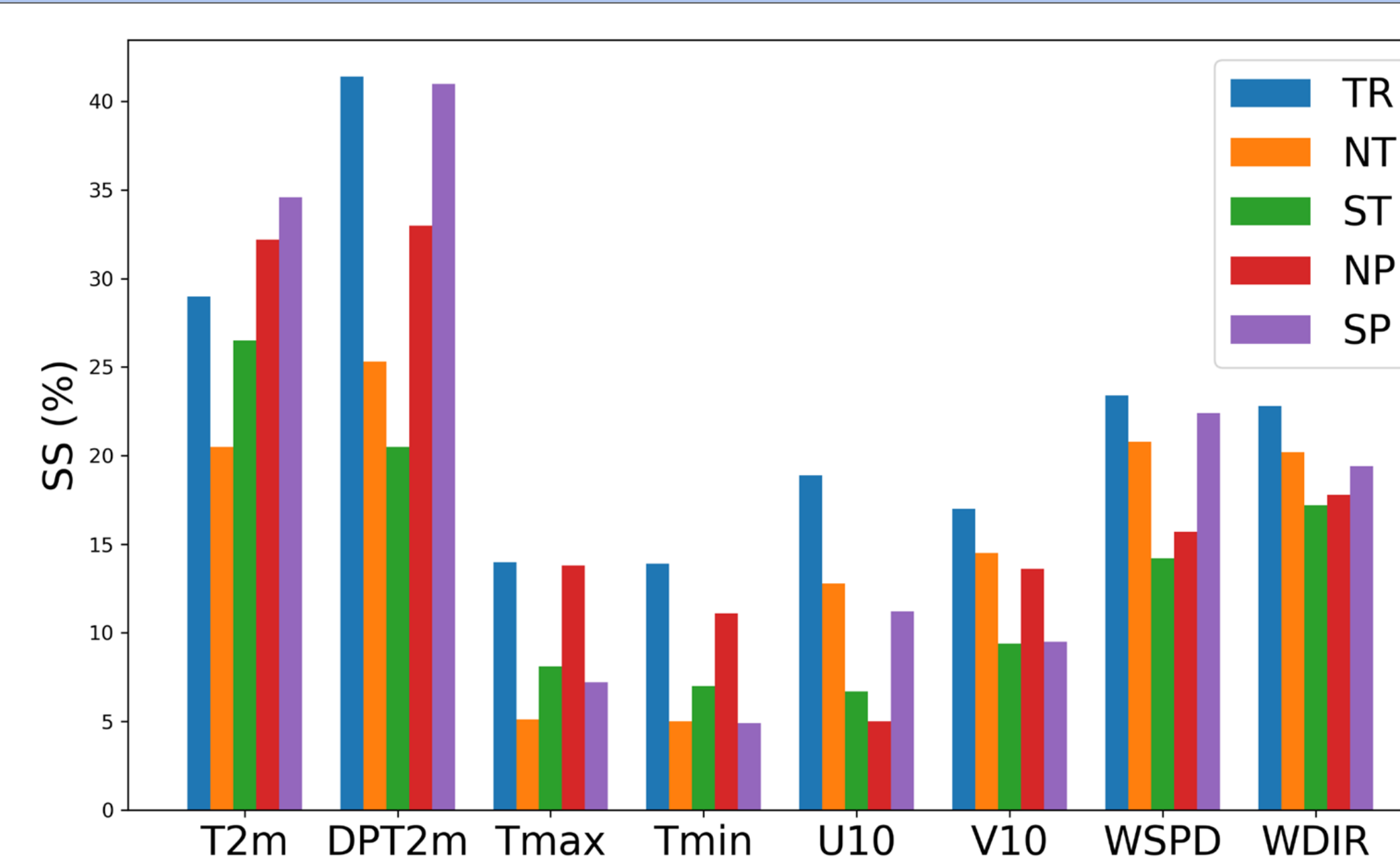


Fig.4 Skill score (SS) for annual mean corrected GFS variables @ FH72 over different region: (TR: tropical; NT/ST: north/south temperate; NP/SP: north/south polar)

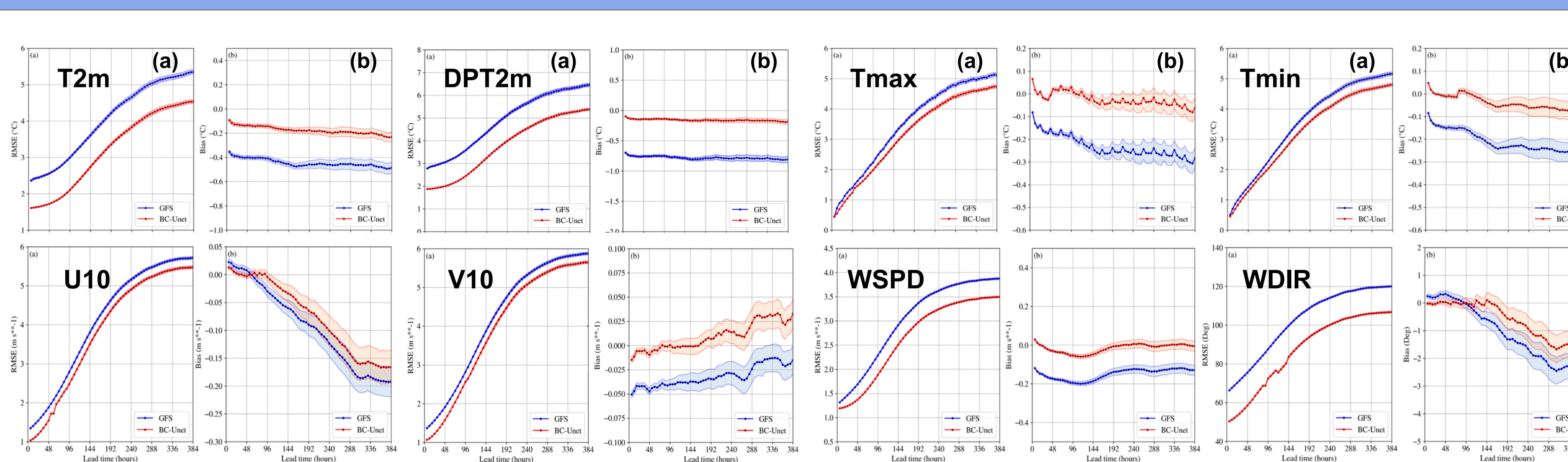


Fig.5 Global-averaged (a) RMSE and (right) bias of raw (b) and corrected (red) GFS T2m, DPT2m, Tmax, Tmin, U10, V10, WSPD and WDIR as a function of forecast lead time (0-384 h). The shade areas represent the 5th-95th percentile range.

Summary

- The variable-specific U-Net-based bias correction model was developed to correct multiple GFSv16 forecasts variables over the global domain.
- Varying depths with single-, two- and four-task U-Net models were explored.
- T2m, DPT2m, U10, V10, WSPD and WDIR models were trained with ERA5 data, while Tmax and Tmin with GDAS data for data consistency.
- The models significantly improve the annual and seasonal mean GFS forecasts in terms of mean bias, mean absolute bias (MAE) and root-mean-square error (RMSE) over the global and regional domain, especially on DPT2m.
- Although the models were trained on limited lead times, they generalized well across the full 6–384 h range, except V10 beyond 288 h.
- As forecast lead time increases, the global mean RMSE of all raw GFS variables increases significantly, while the global mean bias remains predominantly negative across lead times. Bias trends vary by variable: T2m, DPT2m, and WSPD show slight variations; Tmax, Tmin, and U10 exhibit increasing bias; and V10 shows a decreasing trend.

Supplementary Materials

- Reference:
 - Cui, L., J. Wang, S. S. Tabas, and J. R. Carley. (2025). A Machine Learning-Based Bias Correction Method for Global Forecast System Products. U.S. Dept. of Commerce, NOAA NCEP Office Note 520, 23 p. doi: 10.25923/6266-e822.
- GitHub repository:
 - <https://github.com/NOAA-EMC/ML4BC>