Sub-seasonal Prediction of Summer Surface Maximum Temperature Extremes over CONUS through a Hybrid Post-Processing Technique

M. M. Nageswararao[1] and Vijay Tallapragada[2]

[1]CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, MD-20740, USA.

[2]NOAA/NWS/NCEP/EMC, College Park, MD-20740, USA.

murali.n.malasala@noaa.gov

Background and Motivation

Temperature significantly affects critical sectors like energy, agriculture, and public health. Global warming has intensified heatwaves over the Contiguous United States (CONUS) since the mid-20th century, posing severe societal risks. Predicting maximum temperature (T max) extremes on a sub-seasonal scale is challenging due to diminishing atmospheric initial conditions memory in weather models. NOAA's GEFSv12 has improved extended-range forecasts but often requires post-processing to correct biases, especially at longer forecast lead times (Zhou et al., 2022). This study introduces a Hybrid Post-processing Technique (HPPT) (Nageswararao et al., 2023) that combines Artificial Neural Networks (ANN) with Detrended Quantile Mapping (QQ) to enhance T max predictions over CONUS, aiming to improve forecast accuracy and support sectors in managing extreme heat events.

Data and Methodology

This study used NCEP's GEFSv12 T_{max} data over CONUS for the 2000-2019 summer reforecast period, with 35-day lead times and 11 ensemble members (Guan et al., 2022). The dataset, initialized weekly, had a 0.5° spatial resolution and a temporal resolution of 3 hours for the first 10 days and 6 hours thereafter. GEFSv12 features the GFDL FV3 dynamical core and advanced model physics (Zhou et al., 2022). The CPC T_{max} dataset (2000-2019) was used to assess GEFSv12's T_{max} prediction performance (https://psl.noaa.gov/data/gridded/data.cpc.globaltemp.html). The HPPT addresses biases using QQ, while ANNs capture non-linear patterns for improved accuracy. The ANN model developed for this study features two hidden layers, each containing 16 neurons, with the first layer utilizing the ReLU activation function and the second layer using Tanh. The model was trained using the Levenberg-Marquardt algorithm, with a learning rate of 0.001, and was validated through double cross-validation to prevent overfitting. The training process involved 1000 iterations, with an error tolerance of 1e-14, using 70% of the data for training and 30% for validation, selected randomly. The performance of the neural network was assessed using the mean squared error performance function. A 31-day moving window was applied over a 20-year reforecast period, and a leave-one-out procedure ensured rigorous calibration. The study compared the raw GEFSv12 reforecasts and three calibration techniques—HPPT, QQ, and ANN—against CPC data to assess their effectiveness in predicting summer T_{max} and extreme T_{max} events at various lead times across the CONUS.

Results and Discussion:

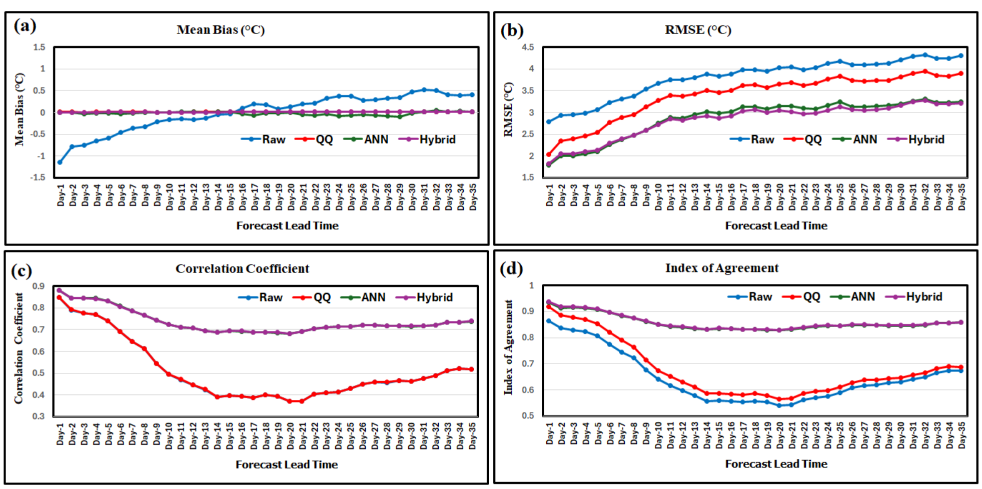

Fig. 1: Statistical skill scores (a) Mean Bias (b) Root Mean Square Error, (c) Correlation Coefficient and (d) Index of Agreement of Raw, QQ, ANN, and Hybrid methods against CPC in depicting Summer (JJA) T_{max} over CONUS with forecast lead time Day-1 to 35 for the reforecast period 2000-2019.

Figure 1 compares the performance of the Raw GEFSv12 model with three calibration methods—QQ, ANN, and Hybrid—in predicting summer daily T_{max} over CONUS from Day 1 to Day 35. The Raw model initially exhibits a cold bias, which shifts to a warm bias by the second week. The QQ method effectively neutralizes bias across all lead times, while ANN reduces the early cold bias but encounters slight bias issues at longer leads. In contrast, the Hybrid method maintains near-zero bias throughout the period. Regarding RMSE, the Raw model's error increases with time, while QQ and ANN both reduce RMSE, though ANN performs better initially. The Hybrid method consistently delivers the lowest RMSE, particularly at extended lead times. For the Correlation Coefficient, the Raw model's correlation declines sharply from 0.9

on Day 1 to below 0.5 by Day 35. ANN and Hybrid maintain higher correlations, with ANN showing strength in the early days, while the Hybrid method excels overall. QQ, however, mirrors the Raw model's correlation pattern. Finally, the Index of Agreement shows a decline for the Raw model from 0.95 to 0.6 over time. Both ANN and Hybrid exhibit better agreement, with the Hybrid method consistently performing the best, especially at longer lead times. Overall, the Hybrid method demonstrates superior skill in bias correction and error reduction across all metrics, making it the most reliable option for extended T_{max} forecasts.

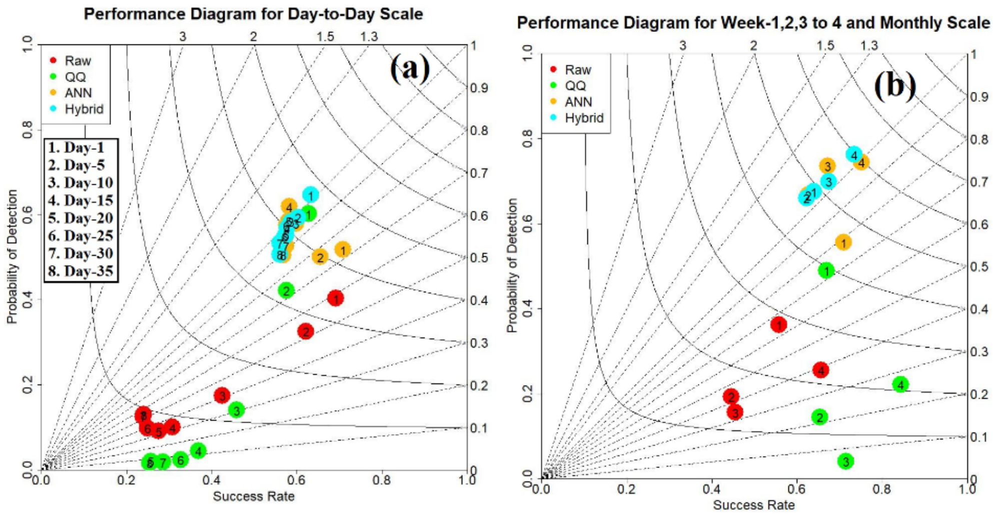

Fig. 2: Performance diagrams comparing Success Rate (SR), Probability of Detection (POD), Frequency Bias, and Critical Success Index (CSI) for Raw, QQ, ANN, and Hybrid methods in predicting summer T_{max} extremes over CONUS during 2000-2019. (a) Day-to-day scale, and (b) Week-1, Week-2, Weeks 3-4, and Monthly scales. Solid lines represent CSI, and dashed lines indicate Frequency Bias.

Figure 2 illustrates performance diagrams comparing the forecast accuracy of four methods—Raw, QQ, ANN, and Hybrid—for predicting T_{max} across various lead times. In day-to-day performance (Fig. 2a), the Raw GEFSv12 model performs poorly, especially beyond Day 15, with low Probability of Detection (POD) and Success Rate (SR) values. The QQ method improves accuracy for short lead times (Days 1-10) but declines at longer lead times. The ANN method consistently outperforms both Raw and QQ, particularly at extended lead times. The Hybrid approach is the top performer, achieving the highest POD and SR values across all lead times, especially for Days 1-10, and maintains superior accuracy at longer leads. For weekly and monthly performance (Fig. 2b), a similar trend is observed. The Raw model and QQ method struggle at longer lead times, while the ANN method shows significant improvement, particularly at longer scales. The Hybrid approach outperforms all methods across all scales, with the highest POD and SR values and a frequency bias close to 1, demonstrating its robustness and reliability. In summary, the Hybrid method consistently provides the most accurate and reliable T_{max} forecasts across all lead times, effectively combining the strengths of ANN and QQ.

Conclusions

The GEFSv12 model exhibits biases when predicting summer T_{max} over CONUS, showing a cold bias at shorter lead times and a warm bias at longer lead times. Calibration methods, particularly the Hybrid approach, effectively reduce these biases and significantly improve prediction skill across all lead times, outperforming both the raw model and QQ method. While the Raw model shows higher Correlation Coefficient (CC) and Index of Agreement (IOA) at shorter lead times, these decrease with lead time. The ANN and Hybrid methods enhance both CC and IOA, especially at longer lead times. The Raw model consistently underestimates Tmax extremes, but the ANN and Hybrid methods substantially improve skill scores, making the Hybrid approach the most effective for predicting T_{max} extremes across various forecast periods.

References

Guan, H., and et al., (2022). GEFSv12 reforecast dataset for supporting subseasonal and hydrometeorological applications. *Mon. Wea. Rev.*, **150**, 647–665, https://doi.org/10.1175/MWR-D-21-0245.1.

Nageswararao, M.M., Zhu, Y., Tallapragada, V. & Chen, M.-S. (2023). Hybrid Post-Processing on GEFSv12 Reforecast for Summer Maximum Temperature Ensemble Forecasts with an Extended-Range Time Scale over Taiwan. *Atmosphere*, *14*, 1620. https://doi.org/10.3390/atmos14111620

Zhou, X.; Zhu, Y.; Hou, D.; Fu, B.; Li, W.; Guan, H.; Sinsky, E.; Kolczynski, W.; Xue, X.; Luo, Y.; et al., (2022). The Development of the NCEP Global Ensemble Forecast System Version 12. *Weather Forecast.*, 37, 727, https://doi.org/10.1175/WAF-D-21-0112.1