NOAA's 47th Climate Diagnostics and Prediction Workshop Special Issue

Climate Diagnostics and Prediction Workshop Digest

https://doi.org/10.25923/ggwg-0b54

July 2023

NOAA's National Weather Service

Office of Science and Technology Integration
1325 East West Highway
Silver Spring, MD 20910

Climate Prediction Center
5830 University Research Court
College Park, MD 20740

Inside this Issue:

Although the skill of current operational climate prediction is limited and the research on the topic presents many challenges, there are promises of improvement on the horizon. To accelerate advancement in climate services, an effective mechanism of science and technology infusion from research to operation for application is much needed. Published since 2008, this Digest has been established to clarify science-related problems and relevant issues identified in operation, inviting our partners in the research community to work together on improvement of national climate prediction services. You are invited to access archived issues in the NOAA Institutional Repository.

Office of Science and Technology Integration
Research to Operations Team

National Weather Service
National Oceanic and Atmospheric Administration
U.S. Department of Commerce

PREFACE

It is with great pleasure that the Climate Prediction Center (CPC) and the Office of Science and Technology Integration (STI) offer you this synthesis of the 47th Climate Diagnostics and Prediction Workshop (CDPW). The CDPW remains a must attend workshop for the climate monitoring and prediction community. As is clearly evident in this digest, considerable progress is being made both in our ability to monitor and predict climate. The purpose of this digest is to ensure that climate research advances are shared with the broader community and also transitioned into operations. This is especially important as NOAA works to enhance climate services both across the agency and with external partners. We hope you find this digest to be useful and stimulating. And please drop me a note if you have suggestions to improve the digest.

I would like to thank Nicole Kurkowski and Sarah Champion of the Office of Science and Technology Integration, for their efforts in assembling and editing the digest. This partnership between STI and CPC is an essential element of NOAA climate services.

David G. DeWitt
Director, Climate Prediction Center National Centers for Environmental Prediction
NOAA's National Weather Service

Table of Contents

OVERVIEW

NOAA's 47th Climate Diagnostics and Prediction Workshop (CDPW) was hosted by the Utah Climate Center at Utah State University in Logan, Utah, from October 25-27, 2023. The CDPW was co-sponsored by the Climate Prediction Center (CPC), the National Centers for Environmental Prediction (NCEP), and the Climate Services Branch (CSB) of the National Weather Service (NWS).

The CDPW focused on four major themes, with an emphasis on climate prediction, monitoring, attribution, diagnostics, and service delivery related to:

1. Predictions and predictability of climate variability across spatial and temporal scales, and forecasts verifications.
2. Applications of modern technologies such as machine learning / artificial intelligence (ML/AI) on the development of improved predictions at Sub-seasonal to Seasonal (S2S) time scales.
3. Tools and techniques that bridge climate forecasts with climate applications and service delivery, including methods to address service equity, to foster resilience to impactful weather, water, and climate hazards.
4. Prediction and monitoring of hydroclimate in the Western U.S., including drought, snowfall and snowpack, pluvial flooding, extreme precipitation, and other variables related to water resources.

The CDPW featured oral and poster presentations, lightning talks, invited speakers, and group discussions. This Digest is a collection of select extended abstracts contributed by CDPW presenters.

The workshop continues to grow and provides a platform for discussing improvements in climate monitoring, diagnostics, prediction, applications and services. The next CDPW will be hosted by the Department of Earth, Ocean and Atmospheric Science (EOAS) and the Center for Ocean-Atmospheric Prediction Studies (COAPS) at Florida State University in Tallahassee, Florida from October 24-26, 2023.

Recommended article citation: Authors, 2023: Article Title. *Extended Summary, NOAA's 47th Climate Diagnostics and Prediction Workshop,* Logan, UT, DOC/NOAA, page range. https://doi.org/10.25923/ggwg-0b54.

A Historical Perspective of the La Niña Event in 2020/21

Xiaofan Li[1,2], Zeng-Zhen Hu[3], Yu-heng Tseng[4], Yunyun Liu[5], and Ping Liang[6]

[1]Key Laboratory of Geoscience Big Data and Deep Resource of Zhejiang Province, School of Earth Sciences, Zhejiang University, Hangzhou, Zhejiang 310027, China

[2]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China

[3]Climate Prediction Center, NCEP/NWS/NOAA, College Park, MD 20740, USA

[4]Institute of Oceanography, National Taiwan University, Taipei 10617, Taiwan

[5]Laboratory of Climate Studies, National Climate Center, China Meteorological Administration, Beijing 100041, China

[6]Key Laboratory of Cities' Mitigation and Adaptation to Climate Change in Shanghai, Shanghai Regional Climate Center, China Meteorological Administration, Shanghai, 200030, China

ABSTRACT

1. Introduction

El Niño–Southern Oscillation is the strongest interannual variability in the tropical oceans and the major source of global climate predictability. In this work, we examine the evolution of oceanic and atmospheric anomalies in the tropical Pacific during 2020/21 La Niña and compare it with the historical strong La Niña events since 1982, identify the contributions of different time scale components, and assess the predictions and the impact on extra-tropical climate.

2020/21 La Niña emerged in August 2020 and dissipated in May 2021. 2020/21 La Niña was uniquely preceded by a borderline El Niño instead of an El Niño and a weak equatorial-heat discharge process (Fig. 1). That resulted in the weakest event among strong La Niña events since 1982, although there were strong upwelling Kelvin wave activities. Moreover, compared with other strong La Niña events, the surface easterly wind anomalies and the warm pool extended further eastward in 2020/21 La Niña, linking to a relatively weaker dipole-like pattern of the subsurface ocean temperature anomalies. The strength of all the strong La Niña events is determined by the in-phase amplification of all-time scale variations. Their decay in the boreal spring and early summer is mainly controlled by the intraseasonal-interseasonal variation (Fig. 2). 2020/21 La Niña was successfully predicted (Fig. 3), however, the North American climate anomalies didn't match the typical La Niña response, leading to low prediction skill in the extra-tropics during its mature phase.

An edited version of this paper was published by AGU. Copyright (2022) American Geophysical Union. Li, X., Z.-Z. Hu, Y.-h. Tseng, Y. Liu, and P. Liang: 2022: A Historical Perspective of the La Niña Event in 2020/21. J. Geophys. Res., 127 (7), e2021JD035546. https://doi.org/10.1029/2021JD035546.

Correspondence to: Zeng-Zhen Hu, Climate Prediction Center, NCEP/NWS/NOAA, College Park, Maryland, zeng-zhen.hu@noaa.gov

2. Figures

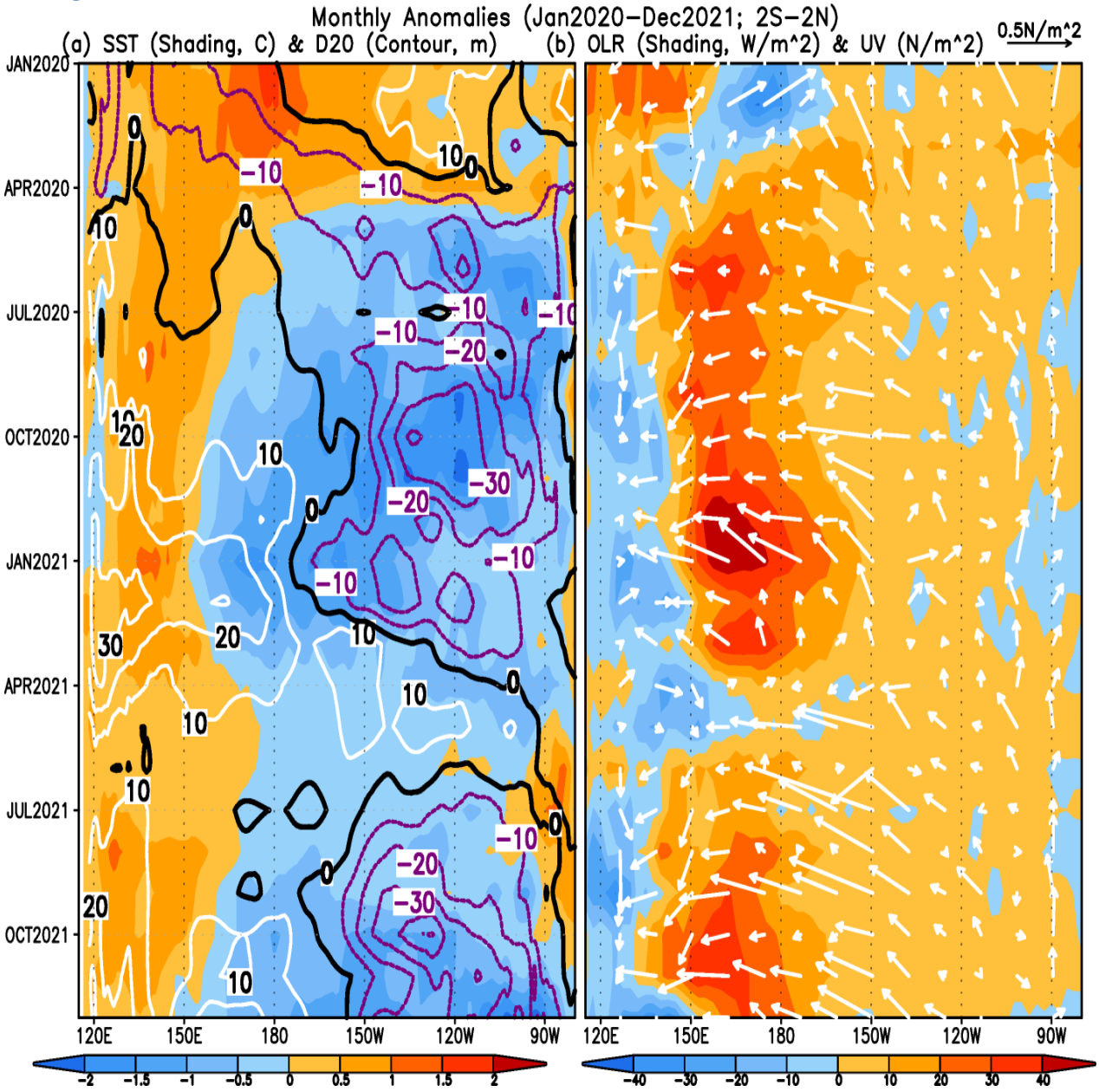

Figure 1: Hovmöller diagrams of the monthly mean of (a) sea surface temperature (SST; shading) and the depth of 20 degrees Celsius isotherm (D20; contours) anomalies, and (b) Outgoing longwave radiation (OLR; shading) and surface wind stress (vector) anomalies averaged in 2 degrees South to 2 degrees North during (a) January 2020-December 2021. The units are degrees Celsius for SST, meters for D20, Watts per square meter for OLR, and Newtons per square meter for wind stress. SST, D20, OLR, and surface wind stress data are from Huang et al. (2021), Behringer (2007), Liebmann and Smith (1996), Kanamitsu et al. (2002), respectively.

Correspondence to: Zeng-Zhen Hu, Climate Prediction Center, NCEP/NWS/NOAA, College Park, Maryland, zeng-zhen.hu@noaa.gov

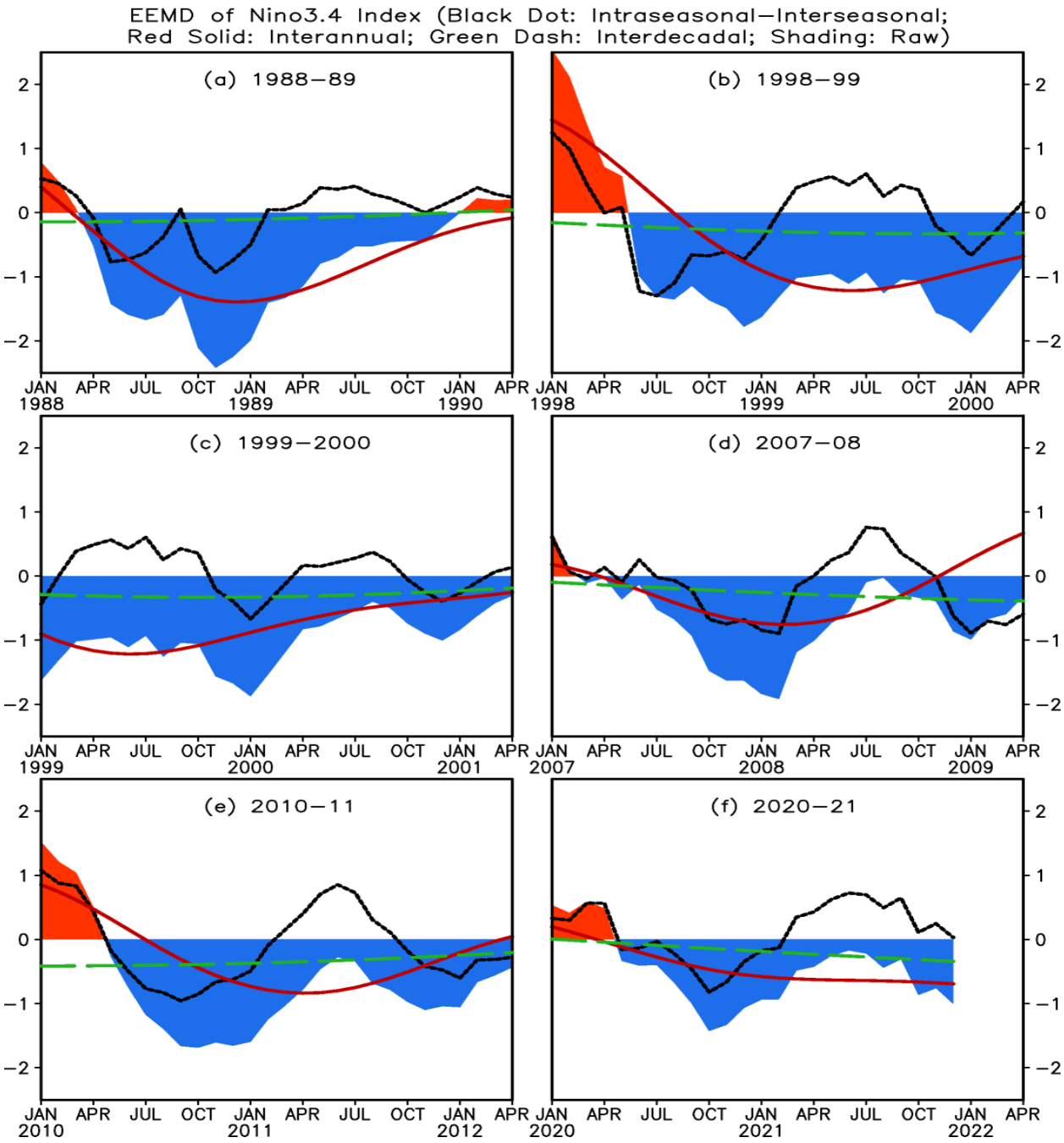

Figure 2: Monthly mean Niño3.4 index during (a) January 1988-April 1990, (b) January 1998-April 2000, (c) January 1999-April 2001, (d) January 2007-April 2009, (e) January 2010-April 2012, and (f) January 2020-December 2021. The shading, black dot, red solid, and green dash lines represent raw data, Ensemble Empirical Mode Decomposition (EEMD) components at intraseasonal-interseasonal, interannual, and interdecadal and longer time scales, respectively. The unit is degrees Celsius. SST data are from Huang et al. (2021).

Correspondence to: Zeng-Zhen Hu, Climate Prediction Center, NCEP/NWS/NOAA, College Park, Maryland, zeng-zhen.hu@noaa.gov

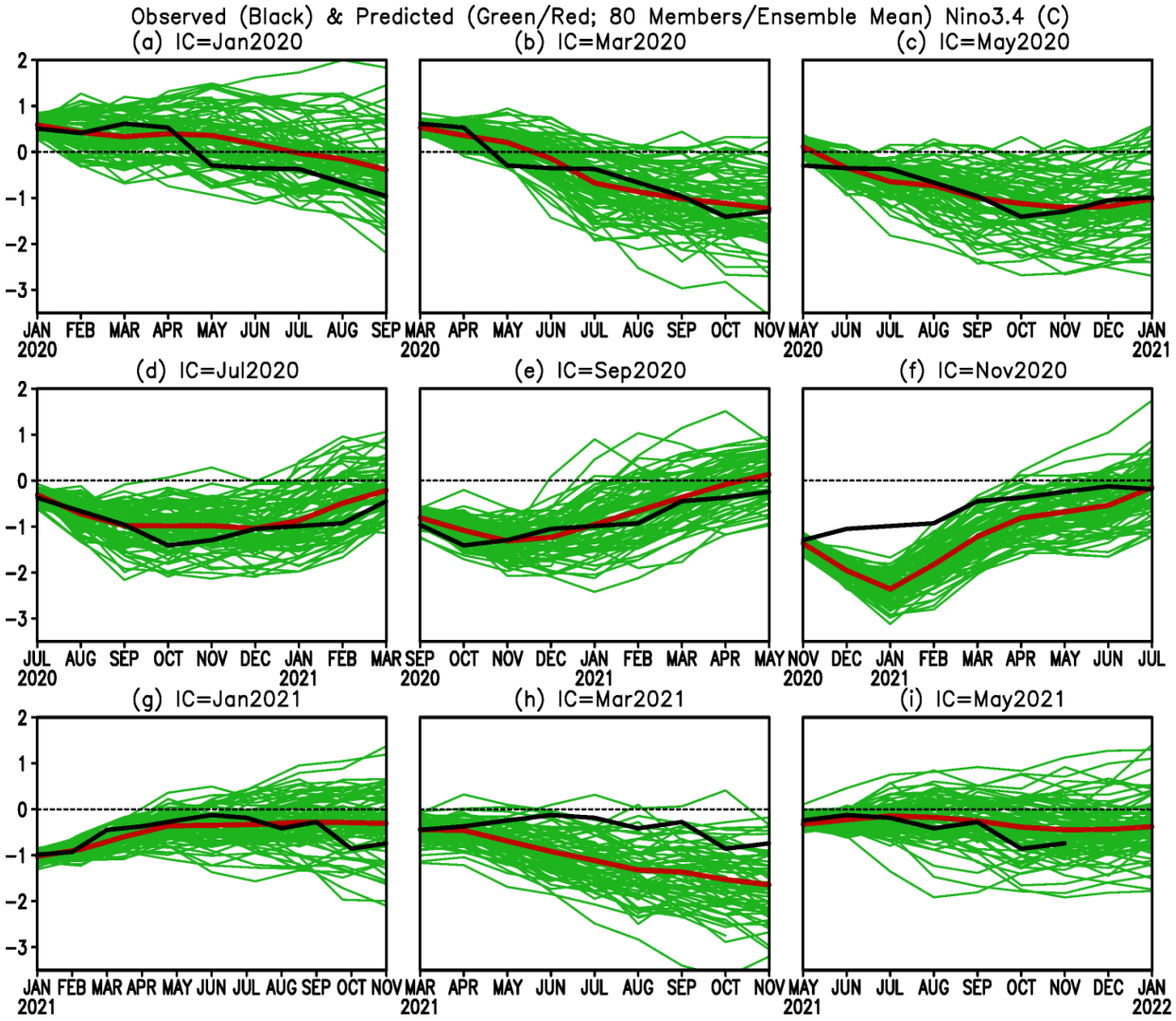

Figure 3: Observed (black line) and Climate Forecast System version 2 (CFSv2) predicted monthly mean Niño3.4 index with initial conditions in (a) January 2020, (b) March 2020, (c) May 2020, (d) July 2020, (e) September 2020, (f) November 2020, (g) January 2021, (h) March 2021, and (i) May 2021. The green lines denote the 80 individual members, and the red line represents the ensemble mean of 80 members. The unit is degrees Celsius. CFSV2 data are from Xue et al. (2013) and Saha et al. (2014).

3. Data and Code Availability Statement

The data used in this study can be downloaded from https://www.esrl.noaa.gov/psd/data/gridded/data.godas.html, https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html, https://psl.noaa.gov/data/gridded/data.interp_OLR.html, https://rda.ucar.edu/datasets/ds094.2/, and https://www.cpc.ncep.noaa.gov/products/predictions/long_range/tools/briefing/seas_veri.grid.php, respectively. The calculations and plots are made with http://cola.gmu.edu/grads/grads.php.

Correspondence to: Zeng-Zhen Hu, Climate Prediction Center, NCEP/NWS/NOAA, College Park, Maryland, zeng-zhen.hu@noaa.gov

4. References

1. Behringer, D.W., 2007: The Global Ocean Data Assimilation System (GODAS) at NCEP. Preprints, 11th Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface, San Antonio, TX, Amer. Meteor. Soc., 3.3. [Available online at http://ams.confex.com/ams/87ANNUAL/techprogram/paper_119541.htm.
2. Liebmann B., and C.A. Smith, 1996: Description of A Complete (Interpolated) Outgoing Long Wave Radiation Dataset. Bull. Amer. Meteor. Soc., **77**, 1275-1277, https://doi.org/10.1175/1520-0477-77.6.1274.
3. Huang, B., C. Liu, V. Banzon, E. Freeman, G. Graham, B. Hankins, T. Smith, and H.-M. Zhang, 2021: Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. J. Climate, **34**(8), 2923-2939, https://doi.org/10.1175/JCLI-D-20-0166.1.
4. Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J.J. Hnilo, M. Fiorino, and G.L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). Bull. Amer. Meteor. Soc., **83**, 1631−1643, https://doi.org/10.1175/BAMS-83-11-1631.
5. Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System Version 2. J. Climate, **27**(6), 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1.

   Xue, Y., M. Chen, A. Kumar, Z.-Z. Hu, and W. Wang, 2013: Prediction Skill and Bias of Tropical Pacific Sea Surface Temperatures in The NCEP Climate Forecast System Version 2. J. Climate, **26** (15), 5358-5378, https://doi.org/10.1175/JCLI-D-12-006.

Correspondence to: Zeng-Zhen Hu, Climate Prediction Center, NCEP/NWS/NOAA, College Park, Maryland, zeng-zhen.hu@noaa.gov

Week 3-4 Multi-Model Ensemble Subsampling: A Real-Time Verification

Cory F. Baggett[1], Steven Simon[1,2], Michael Halpert[1]
[1] NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland
[2] Earth Resources Technology, Inc.; Laurel, Maryland

ABSTRACT

1. Introduction

Forecasters at the Climate Prediction Center (CPC) produce probabilistic Week 3-4 outlooks for above or below normal temperature and precipitation. To produce these outlooks, a variety of experimental tools are at their disposal. These experimental tools usually undergo a few years of real-time testing before they are moved into operational status. Adding skill over existing operational tools is a key test an experimental tool must pass, otherwise it would have little chance of improving CPC's official Week 3-4 outlook. The Ensemble Subsampling Tool is one such experimental tool that forecasters have been utilizing since November 2020, following positive results during its research and development phase at CPC during 2020 (Baggett et al. 2021). The goal of the Ensemble Subsampling Tool is to create a subsample forecast that improves upon the multi-model all-member forecast by objectively subsampling ensemble members in real-time.

2. Data and Methods

The Ensemble Subsampling Tool subsamples from 185 ensemble members available from the extended-range forecasts from the European Centre for Medium-range Weather Forecasting (ECMWF), the Global Ensemble Forecasting System Version 12 (GEFSv12), the Climate Forecast System Version 2 (CFSv2), the Environment and Climate Change Canada (ECCC), and the Japan Meteorological Agency (JMA; Vitart et al., 2017; Zhou et al. 2022). Each model undergoes simple bias correction by subtracting its respective reforecast climatology from its ensemble output. Individual members are subsampled by choosing the members that have the highest pattern correlation between their Week 2 500-hectopascal height anomaly forecast and CPC's Week 2 500-hectopascal Autoblend height anomaly forecast (Baggett et al. 2021). The Autoblend is a simple weighted-mean of the ensemble means of the following models: ECMWF (50%), ECCC (25%), and GEFS (25%).

Temperature, precipitation, and 500- hectopascal height forecasts produced by the Ensemble Subsampling Tool are verified against CPC's Global Temperature dataset (Fan et al. 2008), CPC's Global Unified Gauge-Based Analysis of Daily Precipitation dataset (Chen et al. 2008), and the NCEP/NCAR Reanalysis Version 1's 500- hectopascal heights (Kalnay et al. 1996), respectively.

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

3. Analysis

First, a demonstration of the Ensemble Subsampling Tool is provided for the forecast issued 02-Dec-2022 for the Week 3-4 period, valid for 17-Dec to 30-Dec-2022. Figure 1a displays the multi-model, 185-member (all-member) probabilistic temperature forecast for this period. There are numerous regions where below normal temperatures are weakly favored throughout the contiguous U.S. (CONUS). In order to create a subsample forecast, the Week 2 500- hectopascal height anomaly forecast from each of the individual 185 members is compared to CPC's Week 2 500- hectopascal Autoblend, valid 10-Dec to 16-Dec-2022 (Fig. 1b) via pattern correlation analysis and are subsequently ranked from worst to best matching members. In this example, the subsampling process will favor members with anomalously high height centers over the Aleutian Islands and the Labrador Sea. Figure 1c shows the subsample forecast that retains the top 67% of the best matching members. Figure 1d highlights the difference in Week 3-4 temperature probabilities between the all-member forecast (Fig. 1a) and the 67% subsample forecast (Fig. 1c). Here, the 67% subsample forecast has higher probabilities for below normal temperature across most of CONUS, such that its categorical forecast favors below normal in more locations than the all-member forecast. Figure 2 provides the observed anomalous temperature for the forecast valid period, 17-Dec to 30-Dec-2022. Much of north-central and southeastern CONUS experienced two-week averaged temperatures more than 3 degrees Celsius below normal. To measure the improvement of the subsample forecast over the all-member forecast, the Heidke Skill Score (HSS) is employed to measure how well the probabilistic forecast correctly favored the observed categories of below or above normal. For reference, an HSS of 100 represents a correct forecast everywhere; an HSS of 0 is a forecast with no skill with respect to a climatological forecast; while an HSS of -100 is a forecast that is wrong everywhere. For this particular forecast, the 67% subsample had an HSS of 44.6 in comparison to the all-member forecast having a worse HSS of 27.5.

Now, verification statistics from the Ensemble Subsampling Tool are provided for 248 real-time Week 3-4 forecasts, issued on Tuesdays and Fridays, from 03-Nov-2020 through 21-Mar-2023. Figure 3a provides the HSS for temperature as a function of the percentage of best members retained in the subsample. The all-member forecast has an average HSS of 28.3, both the 67% and 50% subsample forecasts have an HSS of 28.7, and the best performing subsample forecast occurred at 40%, with an HSS of 29.5. It should be noted here that it is not known in real-time which, if any, percentage subsample will have the most improvement over the all-member forecast. However, Baggett et al. (2021) showed through reforecast verifications that subsamples that retain about half to two thirds of the original members provided the greatest improvement over the all-member forecast. Figure 3b assesses the statistical significance of the subsample forecasts versus the all-member forecasts via a sign test, displayed as a random walk. A random walk counts across forecast issuances the number of times the subsample forecast has a higher HSS or lower HSS versus the all-member forecast, assigning a +1 when it wins and a -1 when it loses. If the subsample forecast "walks" outside the gray areas, it is considered statistically distinguishable from the all-member forecast. While the subsample forecasts performed significantly better early during the real-time period, they have not shown much of an advantage recently. However, the 67% subsample forecast appears to offer the most value during extended winter, as indicated by its positive slope during 2020-21, 2021-22, and, to a lesser extent 2022-23.

Figures 4a and 4b provide similar random walk plots for 500- hectopascal heights and precipitation, respectively. For precipitation, the 82% subsample forecast has had the best performance, although it shows no significant improvement over the all member forecast. For 500- hectopascal heights,

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

the 69% subsample forecast has outperformed the all-member forecast, having won 23 more times than losing.

4. Conclusion

While the real-time verifications are not entirely conclusive that the subsample forecast outperforms the all-member forecast, the Ensemble Subsampling Tool is nonetheless moving into operational status during 2023. Week 3-4 forecasters at CPC have regularly used the experimental tool to inform their forecasts, as it often provides an important link between CPC's Week 2 and Week 3-4 products. Further, it seemingly adds the most value during extended winter, such as was the case for the forecast initialized on 02-Dec-2022. Extended winter is often a challenging period to forecast due to high, intrinsic variability. Also, it is worth noting that the recent real-time period has been dominated by La Niña conditions in the tropical Pacific. Thus, it would be worthwhile to continue to monitor the tool during a variety of background El Niño-Southern Oscillation states.

The Ensemble Subsampling Tool will also be maintained in an experimental mode 1) to serve as a back-up should the operational tool fail, 2) to convert the tool from 2-category to 3-category forecasts, 3) to determine the optimal subsample size, and 4) to test additional methods to achieve subsampling.

5. Data Availability Statement

Model data from the ECMWF, CFSv2, JMA, and ECCC are available from the S2S Project Database (https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/). Model data from the GEFSv12 is available from NCEP Central Operations (https://www.nco.ncep.noaa.gov/pmb/products/gens/ and Amazon Web Services (https://registry.opendata.aws/noaa-gefs-reforecast/). Reanalysis data, including CPC global temperature, CPC global precipitation, and NCEP/NCAR Reanalysis V1 500- hectopascal heights are available from NOAA's Physical Science Laboratory (https://psl.noaa.gov/data/). Code can be made available upon request to the corresponding author.

6. References

1. Baggett, C., and Coauthors, 2021: Ensemble subsampling to improve week 3-4 temperature and precipitation outlooks. Extended Summary, *Climate Prediction S&T Digest, 45th NOAA Climate Diagnostics and Prediction Workshop*, Virtual Online, DOC/NOAA, 60-64, https://doi.org/10.25923/tpfe-4n87.
2. Chen, M., W. Shi, P. Xie, V.B.S. Silva, V.E. Kousky, R.W. Higgins, and J.E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research*, **113**, D04110, https://doi.org/10.1029/2007JD009132.

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

3. Fan, Y., and H. van den Dool, 2008: A global monthly land surface air temperature analysis for 1948-present. *Journal of Geophysical Research*, **113**, D01103, https://doi.org/10.1029/2007JD008470.

4. Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, **77**, 437-472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

5. Vitart, F., and Coauthors, 2017: The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bulletin of the American Meteorological Society*, **98,** 163-173, https://doi.org/10.1175/BAMS-D-16-0017.1.

6. Zhou, X., and Coauthors, 2022: The development of the NCEP Global Ensemble Forecast System Version 12. *Weather and Forecasting*, **37**, 1069-1084, https://doi.org/10.1175/WAF-D-21-0112.1.

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

7. Figures

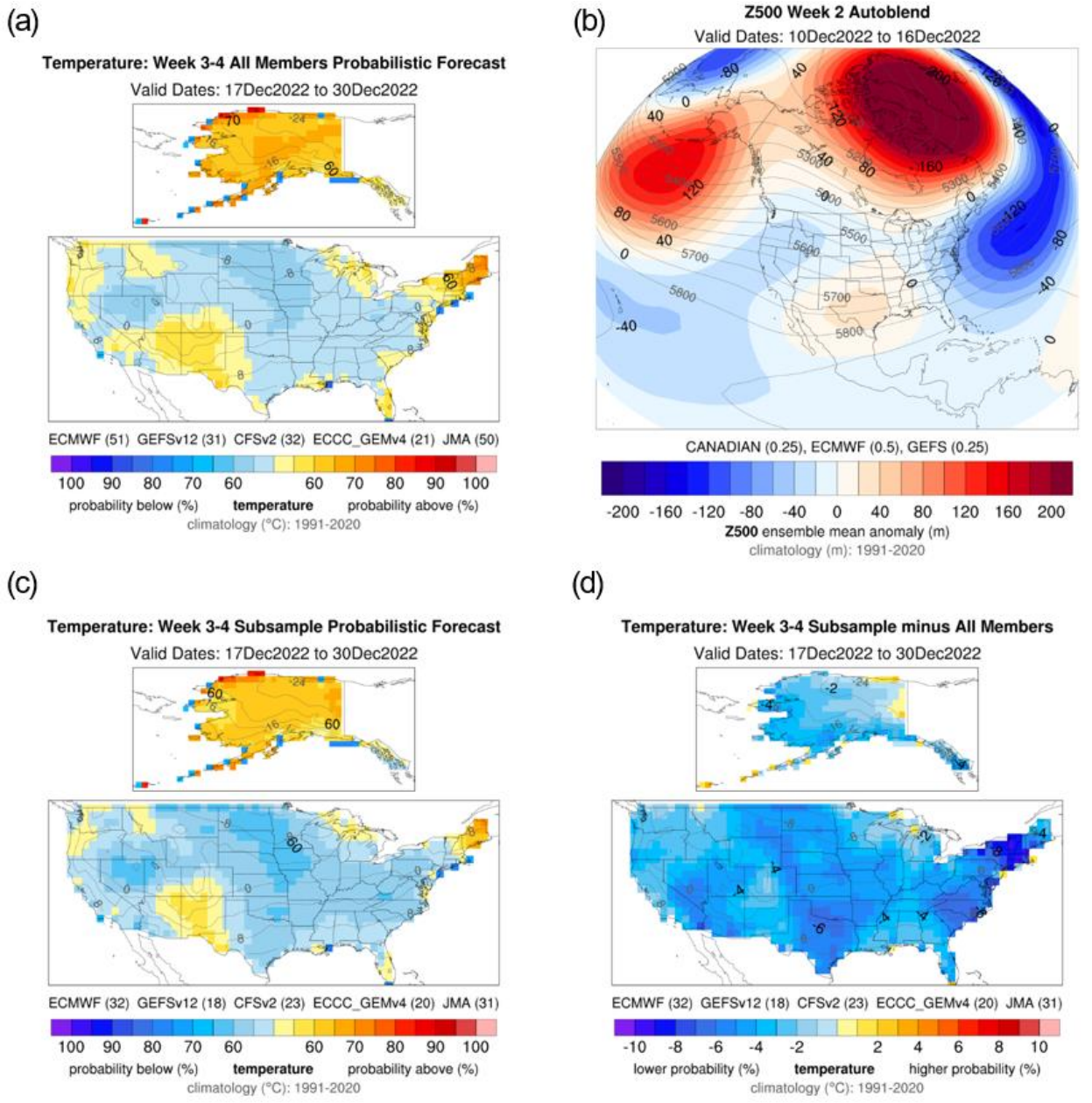

Figure 1: The Ensemble Subsampling Tool produced the following forecasts on 02-Dec-2022: (a) multi-model, 185-member Week 3-4 probabilistic temperature all-member forecast for the Week 3-4 period valid 17-Dec to 30-Dec-2022; (b) CPC's Week 2 500-hectopascal Autoblend height anomaly forecast for the Week 2 period valid 10-Dec to 16-Dec-2022; (c) multi-model, 124-member Week 3-4 probabilistic temperature 67% subsample forecast for the Week 3-4 period valid 17-Dec to 30-Dec-2022; and (d) the difference in probabilities between the all-member and 67% subsample forecasts. Positive values indicate that the 67% subsample forecast has a higher probability of above normal temperatures than the all-member forecast, and vice versa (Fan et al. 2008; Kalnay et al. 1996; Vitart et al. 2017; and Zhou et al 2022).

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

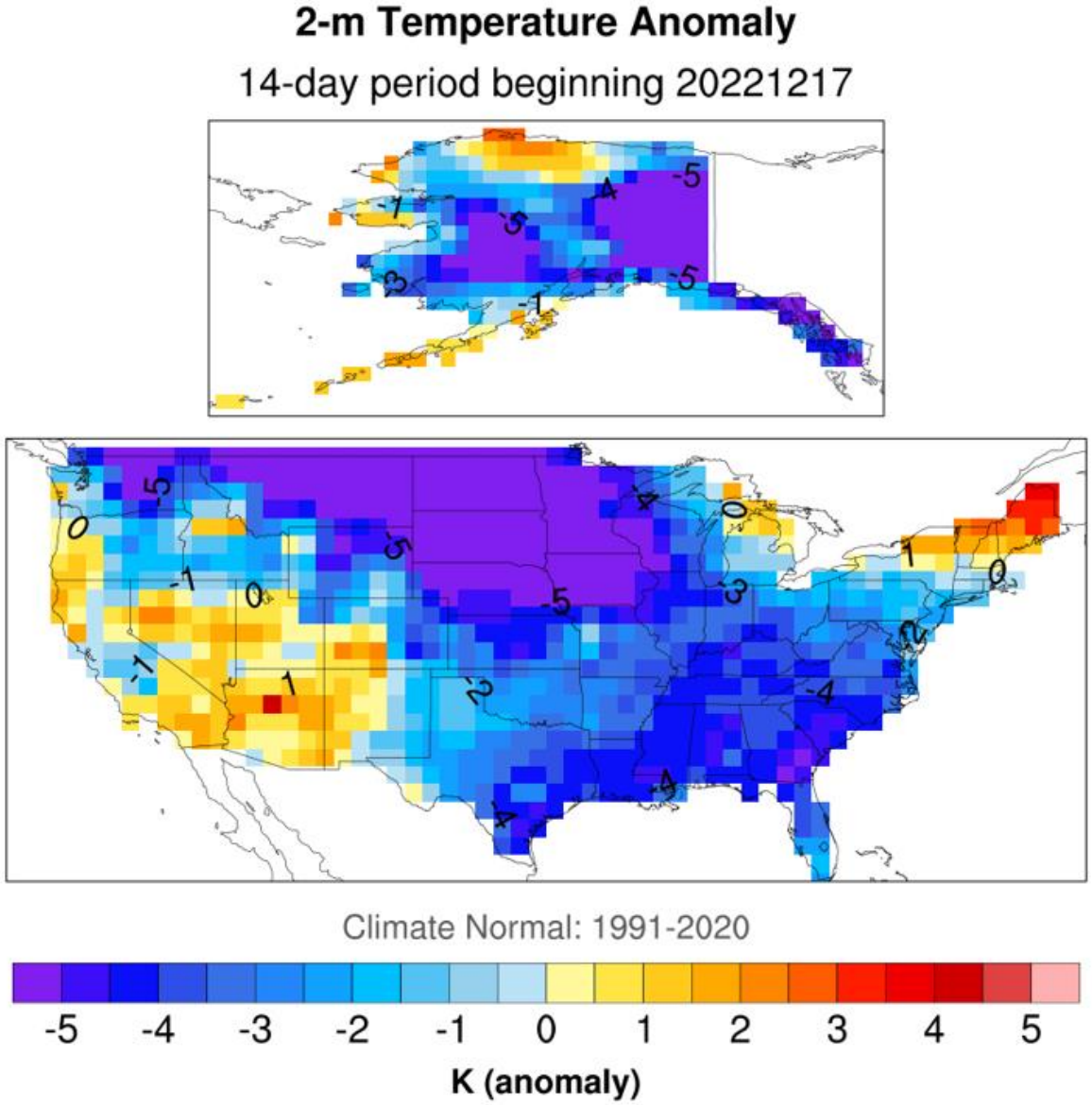

Figure 2: Observed anomalous temperature for the two-week period of 17-Dec to 30-Dec-2022, corresponding to the Week 3-4 forecast issued 02-Dec-2022 (Fan et al. 2008).

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

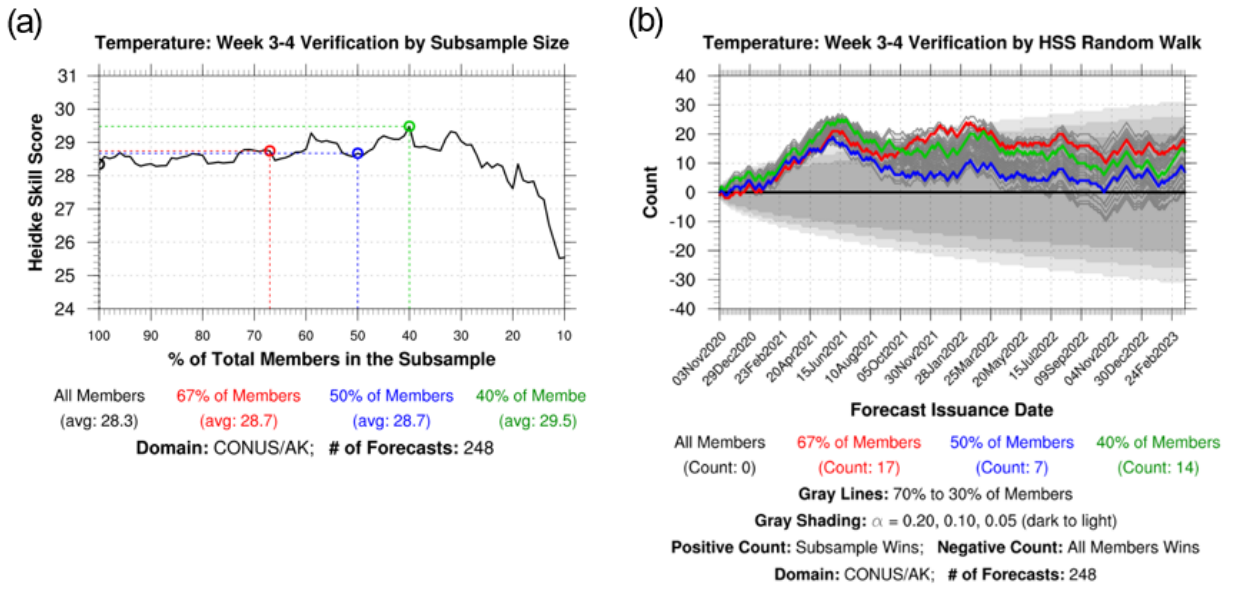

Figure 3: Verification statistics of temperature forecasts from the Ensemble Subsampling Tool are shown as: (a) averaged HSSs as a function of the percentage of members retained in the subsample forecast, where 100% represents the all-member forecast and (b) a sign test, or random walk, of the Week 3-4 temperature subsample forecasts versus the all-member forecast (Fan et al. 2008; Kalnay et al. 1996; Vitart et al. 2017; and Zhou et al. 2022).

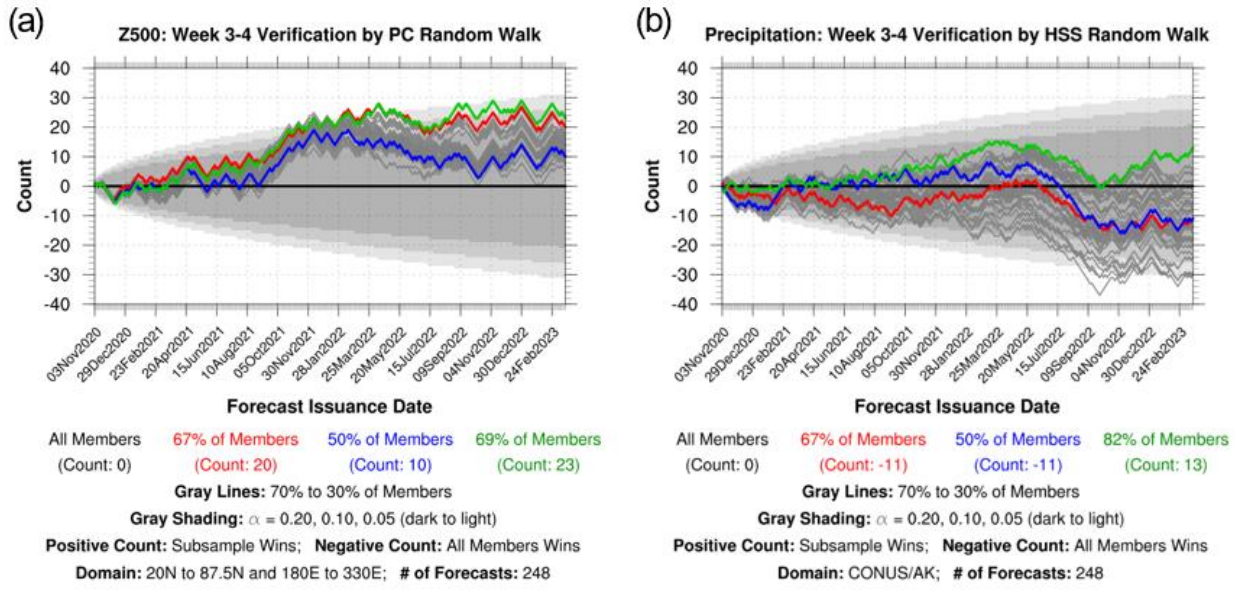

Figure 4: As in Fig. 3b, but for (a) 500-hectopascal heights and (b) precipitation (Chen et al. 2008; Kalnay et al. 1996; Vitart et al. 2017; Zhou et al. 2022)

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

Transition of CPC's Global Tropics Hazards Outlook to Weeks 2 and 3

Lindsey N. Long[1,2], Nicholas Novella[1], and Jon Gottschalck[1]
[1]Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD[2]ERT, Inc, Laurel, MD

ABSTRACT

1. Introduction

The Global Tropics Hazards Outlook (GTH) is a weekly product released on Tuesdays by the Climate Prediction Center (CPC) providing weekly forecasts of above- or below-average precipitation, above- or below-average temperature, and tropical cyclogenesis. On September 13, 2022, the GTH transitioned from a subjective, moderate-to-high-confidence format into a probabilistic format. The forecast period also shifted from Weeks 1 and 2 to Weeks 2 and 3 to better align with the CPC mission at the weeks to years timescale. As with the previous confidence-based format, the probabilistic format of the Tuesday product includes a graphic (Fig. 1), a detailed discussion, and a live briefing; however, a previously released Friday update associated with Week 1 has now been discontinued. Although Week 3 cyclogenesis is included in the full product, it is removed from the publicly released version and only internally released within NOAA. The following paper describes the process to generate the new experimental GTH product and the new tools created to support forecast development.

2. Data and Methods

Four global forecasting model systems are included in the dynamical tools at CPC: the European Centre for Medium-Range Weather Forecasts' (ECMWF, Vitart et al. 2008) Integrated Forecasting System, the National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System Version 12 (GEFS, Guan et al. 2022), and Climate Forecast System Version 2 (CFS, Saha et al. 2014), and the Environment and Climate Change Canada (ECCC, Gagnon et al. 2015) Global Ensemble Prediction System. In addition to the real-time data forecasts, the model reforecasts are utilized for bias-correction and filtering. Results are compared to observations from the best-track datasets from the National Hurricane Center (NHC) (Landsea and Franklin 2013) and the Joint Typhoon Warning Center (JTWC) (Chu et al. 2002). Tropical Cyclones (TCs) are tracked using the method described by Camargo and Zebiak (2002). The symmetric extreme dependency score (SEDS) is used as a skill measure for TCs (Hogan et al. 2009). It is defined as:

$$SEDS = \frac{\ln[(a+b)/n] + \ln[(a+c)/n]}{ln\,[a/n]} - 1$$

2.1. ENSO, MJO, and Other Tropical Modes

Forecasters use a variety of subseasonal-to-seasonal coherent tropical variability, as well as statistical and dynamical tools to create the GTH forecast. The forecast begins by assessing the state of the El Niño Southern Oscillation (ENSO) and the associated rainfall patterns. For example, from July 2020 through 2022, there has been a strong La Niña causing dryness in the

Correspondence to: Lindsey Long, NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland; lindsey.long@noaa.gov

central Pacific Ocean and increased precipitation near the Maritime continent. Although ENSO is the dominant mode of tropical variability at the seasonal timescale and provides a starting point for the GTH forecast, the Madden-Julian Oscillation (MJO) gives a better indication of shifting weather patterns at the subseasonal timescale. This 30- to 60-day propagation of tropical convection often begins over the Indian Ocean and moves eastward over the Pacific Ocean. It helps identify shifts in rainfall patterns and regions of probable tropical cyclone formation. A number of new and improved tools to predict the phase, strength, and impacts of the MJO have been created to assist CPC forecasters. These include bias-corrected Real-time Multivariate MJO (RMM) diagrams, which monitor MJO propagation extended out to Week 3 for the ECMWF, GEFS, and CFS and updated MJO composites for a number of atmospheric fields (200 millibar velocity potential, precipitation, surface temperature, tropical cyclogenesis, etc.) by MJO phase and season.

Other tropical modes that provide predictive information at the subseasonal scale include atmospheric Kelvin Waves, Equatorial Rossby Waves, and African Easterly Waves. Dr. Carl Schreck of the North Carolina Institute for Climate Studies (NCICS) first created a wave filtering tool using CFS that was heavily utilized at CPC to forecast these waves and predict both intense rainfall and tropical cyclogenesis (Schreck et al. 2012; Janiga et al. 2018). For this transition, the tool was brought into operations at CPC and expanded to include ECWMF and GEFS. This tool provides both spatial and temporal (Hovmöller) maps showing observed wave activity for the past weeks to months and forecasted wave activity for the following month. This wave filtering is provided for a number of variables including velocity potential, outgoing longwave radiation (OLR), and zonal wind.

2.2. Dynamical and Statistical Forecasting Tools

Once the strength and location of the key tropical modes have been explored, dynamical and statistical model tools are referenced. For tropical cyclogenesis, a number of new tools were created. For each of the four available global models, forecasters can view raw TC tracks for each forecast week, filtered tracks that remove false alarms, and TC probabilities in the same 20%/40%/60% chance format used in the final GTH product. Also, probabilities using a 4-model ensemble, consisting of the models described above, are available. Maps of TC genesis points observed since 1979-2021 with a rolling seven-day window are also available on the tools page. In addition to dynamical tools, a dynamical-statistical hybrid TC tool initially created by Wang et al (2009) to predict seasonal Atlantic TC activity in the CFS was adopted for subseasonal prediction. It was expanded to include all ocean basins and produce forecasts from all four global models at a weekly time period. This hybrid model is based on three TC predictors: sea surface temperatures (SST), wind shear, and sea-level pressure.

For precipitation and temperature, tools showing total and anomalous forecasts for each of the individual models and an equal-weighted ensemble are made available. Also, probabilities for rainfall and temperatures in the lower and upper thirds, or terciles, consistent with the new GTH format, were created for each model, along with a 4-model ensemble weighted by each model's reforecast correlations to observations. A map of the weights used at each grid point is also made available. For temperature, the GTH is only concerned with forecasts over land where temperatures become hazardous. For both variables, GTH forecasters coordinate with forecasters

Correspondence to: Lindsey Long, NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland; lindsey.long@noaa.gov

at the CPC International Desk and those performing Week 2 and Week 3-4 Outlooks to leverage their knowledge and insure consistency among products.

2.3. Dynamical Model Skill

To provide an overview of the model skill for these variables, a number of skill metrics were applied. As tropical activity is very rare over a single 1-degree by 1-degree grid point, maybe only hitting that grid space once a year except for a few points in the main region of development, it is best to use a skill metric made for extremes or often rare events. The SEDS does not degenerate at the extreme tails and removes skill received from correctly predicting that no TC will occur since this can cause inflated scores in other metrics. Figure 2a-d shows the SEDS values for Week 3 forecasts made by each of the models. Although not shown, there is an expected drop in skill from Weeks 1 to 2, but less so from Weeks 2 to 3. By Week 3, skill remains in large regions of the globe, with the Pacific Ocean retaining the most skill. Overall, the ECMWF model is the most skillful system at all lead times.

For precipitation and temperature, anomaly correlations for each season are calculated for each model. An example showing positive anomaly correlations for ECMWF during December-January-February (DJF) is shown in Figure 2e-f. As with tropical activity, there is an expected drop in skill from Week 1 to 2 (not shown), but skill remains during Week 3. For precipitation, the highest skill with most coverage is evident in the ENSO region of the tropics; however, skill extends into the subtropics, attributed to the MJO (Fig 2e). For temperature, anomaly correlations remain high at week 3 in areas of most interest for GTH stakeholders including subtropical regions, India, and northern South America (Fig 2f). Low correlations seen in Africa are most likely due to a lack of observations in the region. For both variables, a real-time anomaly correlation time-series for the area from 30 degrees South to 30 degrees North is available to gauge each model's current performance. Finally, Brier Skill Scores (BSS) are calculated for precipitation probabilities and show that although the ECMWF outperforms the other models, the skill-weighted model ensemble has the highest scores. This shows that the ensemble does add value at all leads. Corresponding reliability diagrams show that models tend to over-forecast probabilities greater than 30%.

3. Conclusion

The GTH transitioned to an experimental, probabilistic format forecasting above and below-normal precipitation and temperature for Weeks 2 and 3 on September 13, 2022. Week 2 tropical cyclogenesis is also released publicly while week 3 cyclogenesis is currently only released internally at NOAA. Many tools and internal websites were created or updated for this transition. These include forecasts of MJO phase and strength, MJO phase composites, wave filtering to highlight equatorial atmospheric Kelvin and Rossby waves, and dynamical and statistical model guidance for TC formation, precipitation and temperature. Skill scores calculated for these variables based on model reforecasts show skill remains at Week 3 in the vicinity of the Tropics and subtropics.

Correspondence to: Lindsey Long, NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland; lindsey.long@noaa.gov

4. Data Availability Statement

Observed best-track datasets are available on the web at: https://www.nhc.noaa.gov/data/#hurdat (Landsea and Franklin 2013) and https://www.metoc.navy.mil/jtwc/jtwc.html?best-tracks (Chu et al. 2002). Model data can be found at the following websites:

- CFSv2, Saha et al. (2014) -
  - Forecast: https://www.ncei.noaa.gov/thredds/catalog/model-cfs_v2_for_6h_flxf/catalog.html
  - Reforecast: https://www.ncei.noaa.gov/thredds/catalog/model-cfs_refor_6h_90d_flx/catalog.html
- GFSv12, Guan et al. (2022) –
  - Forecast: https://nomads.ncep.noaa.gov/pub/data/nccf/com/gfs/prod/gfs.YYYYMMDD/CC/atmos/gfs.tCCz.pgrb2.0p50.fFFF
  - Reforecast: https://noaa-gefs-retrospective.s3.amazonaws.com/index.html
- ECMWF, Vitart et al. (2008) – https://www.ecmwf.int/en/forecasts/accessing-forecasts. These forecasts are private and paid for by the U.S. government. For access, please contact ECMWF and CPC.
- ECCC, Gagnon et al. (2015) – Forecast: https://www.weather.gc.ca/grib/grib2_ens_geps_e.html
- Reforecast: https://collaboration.cmc.ec.gc.ca/cmc/ensemble/reforecast_data/

5. References

1. Camargo, S. J. and S. E. Zebiak, 2002: Improving the Detection and Tracking of Tropical Cyclones in Atmospheric General Circulation Models. *Wea. Forecasting*, **17** (6), 1152-1162, https://doi.org/10.1175/1520-0434(2002)017<1152:ITDATO>2.0.CO;2.
2. Chu, J.-H., C.R. Sampson, A.S. Levine, and E. Fukada, 2002: The Joint Typhoon Warning Center tropical cyclone best-tracks, 1945-2000, Technical report NRL/MR/7540-02-16. Naval Research Laboratory: Washington, DC.
3. Climate Prediction Center, 2023: Climate News. Accessed 27 April 2023, https://www.cpc.ncep.noaa.gov/.
4. Gagnon, N., X-X. Deng, P.L. Houtekamer, A. Erfani, M. Charron, S. Beauregard, R. Frenette, D. Racette, and R. Lahlou, 2015: Improvement to the Global Ensemble Prediction System (GEPS) reforecast system from version 3.1.0 to version 4.0.0. Canadian Meteorological Centre Technical Note. Environment Canada. 36 pp. [Available online at http://collaboration.cmc.ec.gc.ca/cmc/CMOI/product_guide/docs/lib/technote_geps-411_20151215_e.pdf.]
5. Guan, H., and Coauthors, 2022: GEFSv12 Reforecast Dataset for Supporting Subseasonal and Hydrometeorological Applications. *Mon. Wea. Rev.*, **150** (3), 647-665, https://doi.org/10.1175/MWR-D-21-0245.1.
6. Hogan, R. J., E. J. O'Connor, and A. J. Illingworth, 2009: Verification of cloud-fraction forecasts. *Quart. J. Roy. Meteor. Soc.*, **135**, 1494–1511, https://doi.org/10.1002/qj.481.
7. Janiga, M. A., C.J. Schreck, J.A. Ridout, M. Flatau, N.P. Barton, E.J. Metzger, C.A. Reynolds, 2018: Subseasonal Forecasts of Convectively Coupled Equatorial Waves and the MJO: Activity and Predictive Skill. *Mon. Wea. Rev.*, **146** (8), 2337–2360, https://doi.org/10.1175/MWR-D-17-0261.1.

Correspondence to: Lindsey Long, NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland; lindsey.long@noaa.gov

8. Landsea, C.W. and J.L. Franklin, 2013: Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format. *Mon. Wea. Rev.*, **141**, 3576-3592, https://doi.org/10.1175/MWR-D-12-00254.1.
9. Schreck, C.J., J. Molinari, and A. Aiyyer, 2012: A Global View of Equatorial Waves and Tropical Cyclogenesis. *Mon. Wea. Rev*., **140** (3), 774–788, https://doi.org/10.1175/MWR-D-11-00110.1.
10. Saha, S. and Coauthors, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27**(6), 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1.
11. Vitart, F., R. Buizza, M. A. Balmaseda, G. Balsamo, J.-R. Bidlot, A. Bonet, M. Fuentes, A. Hofstadler, F. Moleni, and T. Palmer, 2008: The new VAREPS-monthly forecasting system: A first step towards seamless prediction, *Q. J. R. Meteorol. Soc*., **134**, 1789–1799.
12. Wang, H., J-K. E. Schemm, A. Kumar, W. Wang, L. Long., M. Chelliah, G.D. Bell, P. Peng, 2009: A Statistical Forecast Model for Atlantic Seasonal Hurricane Activity Based on the NCEP Dynamical Seasonal Forecast. *J. Climate*, **22** (17), 4481-4500, https://doi.org/10.1175/2009JCLI2753.1.

Correspondence to: Lindsey Long, NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland; lindsey.long@noaa.gov

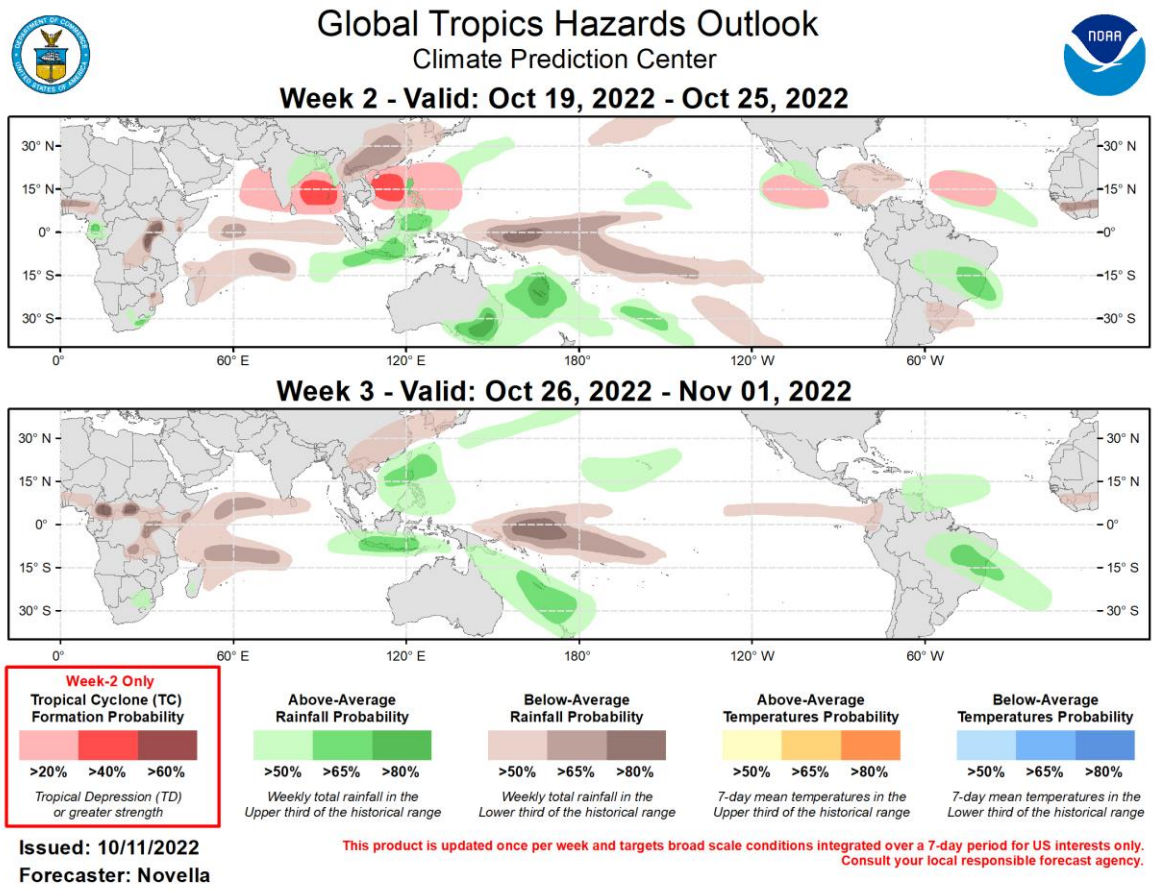

Figure 5: Example of the new and experimental GTH product. Precipitation and temperature probabilities are for 50%,65%, and 80% chances of above- or below-normal of the historical tercile. Tropical Cyclogenesis probabilities are for 20%, 40%, and 60% chances of above- or below-normal of the historical tercile. Tropical Cyclogenesis probabilities are for 20%, 40%, and 60% chances of formation. Source: CPC.

Correspondence to: Lindsey Long, NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland; lindsey.long@noaa.gov

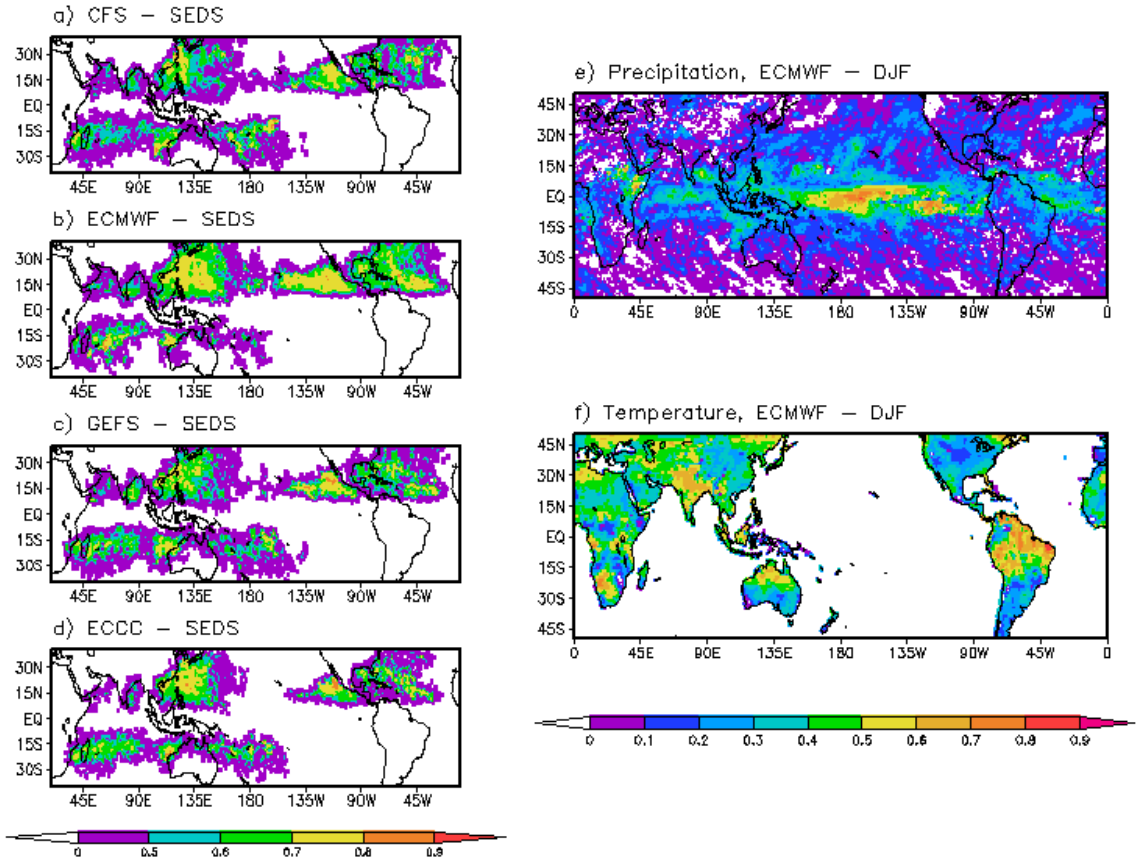

Figure 6: Summary of Week 3 model reforecast skill scores. Week 3 annual SEDS for a) CFS, b) ECMWF, c) GEFS, d) ECCC. e) Week 3 precipitation anomaly correlation for ECMWF during DJF. f) Week 3 temperature anomaly correlation for ECMWF during DJF. Data Sources: Saha et al. (2014); Guan et al. (2022); Vitart et al. (2008); Gagnon et al. (2015); Landsea and Franklin (2013); and Chu et al. (2002

Correspondence to: Lindsey Long, NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland; lindsey.long@noaa.gov

Diurnal Variability of Upper Ocean Simulated by a Climate Mode

J.E. Jack Reeves Eyre[1,2], Jieshun Zhu[2], Arun Kumar[2], Wanqiu Wang[2]
[1] ERT Inc., Laurel, MD, USA
[2] Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD, USA

1. Introduction

The diurnal cycle of sea surface temperature (SST), while small compared to diurnal variations of land surface temperature, still has significant effects on the overlying atmosphere and the dynamics and thermodynamics of the near-surface ocean. Modeling the upper ocean diurnal variability is made difficult by the limited depth (less than or equal to 5 meters) of diurnal warming compared to typical layer thickness (greater than or equal to 10 meters) in the global ocean models used for weather and climate prediction applications. Model parameterizations (e.g., Zeng & Beljaars 2005) have been developed to simulate the diurnal cycle at the surface, with the aim of capturing the diurnal cycle of surface fluxes and their atmospheric response in general circulation models (GCMs). However, a dynamic ocean model with high vertical resolution is needed for simulation of the full spatiotemporal structure of oceanic diurnal variations.

The aim of this paper is to evaluate one such model — a version of the NOAA Climate Forecast System (CFS) with 1-meter vertical resolution in the upper ocean. We first verify the model's ability to simulate the diurnal cycle of SST, which can be compared with global observations. In addition, we compare subsurface temperature and current with observations from a few comprehensively-instrumented moorings. Based on those findings, the self-consistent data provided by model simulations can be used to further advance global understanding of the subsurface diurnal cycle.

2. Data & Methods

The model simulation used in this study is the same as the CFSm501 experiment in Zhu et al. (2020). It uses the Climate Forecast System version 2 (CFSv2) model, modified to have increased vertical resolution in the upper ocean, including ten 1-meter layers at the surface. The simulation is for 26 years but we analyze the final four years, which had hourly output from both ocean and atmosphere models.

We analyze the monthly mean diurnal cycle via four metrics: range (maximum minus minimum), phase (time of maximum), threshold depth (the depth at which the range falls below a fixed threshold of 0.1 Kelvin for temperature and 2 centimeters per second for current), and the threshold delay (between the times of the surface diurnal maximum and the diurnal maximum at the threshold depth).

We use the Met Office Hadley Center's diurnal sea surface temperature range dataset (HadDTR; Kennedy et al., 2007) to assess the simulated diurnal cycle of SST. We also use the results of Masich et al. (2021; hereafter M21) in the tropical Pacific to qualitatively assess the diurnal variability of subsurface temperature and current.

3. Analysis

Comparison with the HadDTR observation-based dataset (Figure 1) shows generally good agreement, with key features being large values on the equator and under the summer hemisphere subtropical highs. There are some systematic differences, however, caused by model bias and/or sampling uncertainty. Overall though, CFSm501 seems to simulate reasonably well the SST monthly mean diurnal cycle, its spatial patterns and seasonal variability.

Correspondence to: J. E. Jack Reeves Eyre, Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD, USA; jack.reeveseyre@noaa.gov

The time-depth evolution of ocean temperature and current at locations in the tropical Pacific agrees well with observations from M21. There is an east-west split in the diurnal evolution. In the west, warming and diurnal jets are mostly limited to the upper 10 meters. In the east, the main feature is the downward propagation, over the day, of the temperature and current diurnal cycles. Significant warming and current variation are seen below 20 or even 30 meters in depth. These results, and the HadDTR comparisons, suggest that the model can realistically simulate surface and subsurface ocean diurnal variations.

The wider spatial patterns of diurnal variations at the surface (shown by diurnal range) and subsurface (shown by threshold depth) show imprints of a number of controlling factors: radiation fluxes, depending on latitude and season; turbulent fluxes, depending on winds and currents; and a number of localized effects depending on proximity to coastlines and features like the equatorial undercurrent. One surprising result is the quite widespread occurrence of significant current diurnal cycles below 30 meters in depth. We interpret this to mean that deep cycle turbulence occurs quite broadly across the tropical oceans and is not dependent on the large background shear of the equatorial undercurrent.

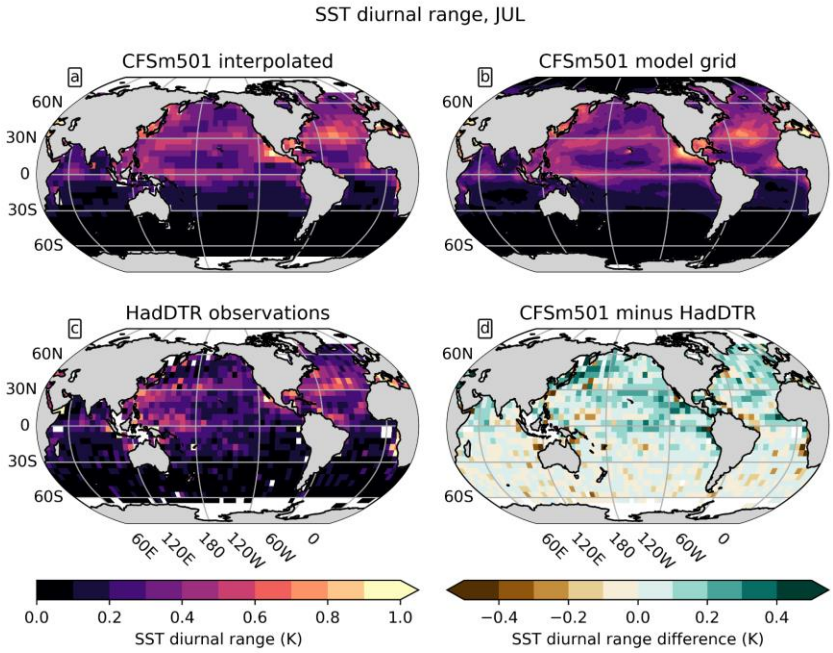

Figure 7: July mean sea surface temperature diurnal range from (a) CFSm501 interpolated to the HadDTR grid, (b) CFSm501 on the native model grid, (c) HadDTR observational data, and (d) the difference between CFSm501 and HadDTR. Note that CFSm501 data are nominally the four-year average while HadDTR data are the 1990-2004 average.

Correspondence to: J. E. Jack Reeves Eyre, Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD, USA; jack.reeveseyre@noaa.gov

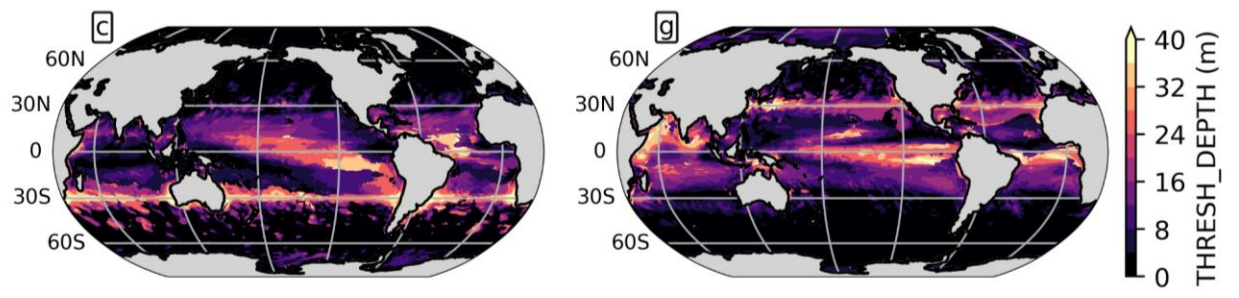

Figure 8: Ocean current diurnal cycle threshold depth (with threshold value of 2 centimeters per second) for January (left) and July (right).

4. Conclusion

In this study we have looked at the upper ocean mean diurnal cycles of temperature and current in a modified version of CFSv2. Comparison with a global observational dataset for SST diurnal range shows the simulations are accurate and realistic in terms of magnitude and spatial patterns. Examination of model profiles at locations used in M21 reveals the sub-surface temperature and current diurnal cycles are also realistic. The accuracy and realism revealed in these observational verifications suggests that the ocean model's high vertical resolution has the desired effect. This means that we can place reasonable trust even in the aspects of simulated upper ocean diurnal cycle that do not have global verification datasets. Subsurface temperature diurnal cycles reach below approximately10 meters in depth across quite broad stretches of the global oceans. This includes equatorial regions, as in mooring data (including M21), but also, eastern subtropical oceans and western boundary currents have relatively deep temperature diurnal cycles, at least in summer. The diurnal range of surface currents commonly exceed 20 centimeters per second in tropical oceans. This is significantly more than the diurnal range at 10 meters seen in observations in M21. These simulated surface currents are hard to verify with acoustic Doppler current profiler and standard in situ current measurements.

5. Open Research

This study has been submitted for publication (Reeves Eyre et al. 2023). The software and data used to create the plots will be freely and publicly available on GitHub and Zenodo after publication. Raw simulation output data is available through (Zhu et al., 2020). The HadDTR data (Kennedy et al., 2007b) are freely and publicly available online.

Correspondence to: J. E. Jack Reeves Eyre, Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD, USA; jack.reeveseyre@noaa.gov

6. References

1. Kennedy, J.J., P. Brohan, and S.F.B. Tett, 2007: A global climatology of the diurnal variations in sea-surface temperature and implications for MSU temperature trends. *Geophysical Research Letters*, **34**, https://doi.org/10.1029/2006GL028920.
2. Masich, J., W. S. Kessler, M. F. Cronin, and K. R. Grissom, 2021: Diurnal Cycles of Near-Surface Currents Across the Tropical Pacific. *Journal of Geophysical Research: Oceans*, **126**, e2020JC016982, https://doi.org/10.1029/2020JC016982.
3. Reeves Eyre, J. E. J., J. Zhu, A. Kumar and W. Wang, 2023: Diurnal variability of the upper ocean simulated by a climate model. Submitted to *Geophysical Research Letters, submitted.*
4. Zeng, X., and A. Beljaars, 2005: A prognostic scheme of sea surface skin temperature for modeling and data assimilation. *Geophysical Research Letters*, **32**, https://doi.org/10.1029/2005GL023030.
5. Zhu, J., A. Kumar, and W. Wang, 2020: Intraseasonal Surface Salinity Variability and the MJO in a Climate Model. *Geophysical Research Letters*, **47**, e2020GL088997, https://doi.org/10.1029/2020GL088997

Correspondence to: J. E. Jack Reeves Eyre, Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD, USA; jack.reeveseyre@noaa.gov

Implications of AMOC Base-State to the Mode of Variability of the U.S. Summer Precipitation

Sala, C. M.[1], Y. Fan[1], and L. Li[1]

[1]Pennsylvania State University, College of Earth and Mineral Sciences, Department of Meteorology and Atmospheric Science, University Park, Pennsylvania

ABSTRACT

1. Introduction

The Atlantic Meridional Overturning Circulation (AMOC) transports warm water towards the subpolar North Atlantic, where it sinks, and is transported back south (Buckley and Marshall, 2016). AMOC variability efficiently redistributes heat horizontally and vertically, thus resulting in multifaceted impacts on the sea surface temperature anomalies and pan-Atlantic climates. AMOC has been linked with Atlantic Multidecadal Variability (Zhang *et al*., 2019) and European temperature and precipitation variability (Buckley and Marshall, 2016). However, links between AMOC and North America are uncertain and leave a major knowledge gap in scientific literature.

The major challenges to AMOC research are the lack of strong observational data and inconsistent proxy data requiring models to be used for real-world inferences on AMOC. Thus, we use models to connect Contiguous United States (CONUS) summertime (June-July-August) precipitation variability to AMOC. Previous studies have shown that CONUS precipitation variability has been increasing in the 20th century caused by a westward extension of the North Atlantic Subtropical High (NASH; Li *et al*., 2012; 2013), a semi-permanent high-pressure system located over the North Atlantic Ocean. A NASH Western Ridge (WR) northwestern (southwestern) extension relative to the climatology will typically lead to increased (decreased) moisture advection into the Midwest (MW) and decreased (increased) moisture advection into the Southeast (SE) CONUS (Li *et al*., 2013). Conversely, an eastward recession yields moisture advection over the SE as the NASH western flank remains stationary over the East Coast (Li *et al*., 2013). Eastward recessions result in a decoupling of the moisture advection from the Gulf of Mexico to the Great Plains Low Level Jet, thus resulting in decreased moisture advection into the central CONUS (Li *et al*., 2013). Using the NASH WR-CONUS relationship, the goal of this study is to use model simulations to quantify impacts of the AMOC base-state on the NASH and the associated impacts on extreme summertime CONUS precipitation.

2. Data and Methods

We define the dominant mode of summertime precipitation variability by the first mode of empirical orthogonal function (EOF), which is calculated using the *eofs* Python library (Dawson, 2016). We project the EOFs as a correlation, which are calculated by performing a linear regression of the first principal component onto the input data. The Climate Prediction Center (CPC) rain-gauge based precipitation dataset (Higgins *et al*., 2000) serves as the baseline for model precipitation evaluations. We use model outputs from the Coupled-Model Intercomparison Project Phase 6 (CMIP6) Historical experiment. Specifically, the Community Earth System Model Version 2 (CESM2; Danabasoglu, 2019), Australian Community Climate and Earth System Simulator Earth System Model Version 1.5 (ACCESS-ESM1.5; Ziehn *et al*., 2019), and Canadian Earth System Model Version 5 (CanESM5; Swart *et al*., 2019). Model precipitation output is linearly detrended and spatially interpolated to match the CPC grid cells. The temporal coverage is confined to summers in the years 1948 to 2014, given the overlapping time between the CPC and models.

Correspondence to: Christopher M. Sala, [1]Pennsylvania State University, College of Earth and Mineral Sciences, Department of Meteorology and Atmospheric Science, University Park, Pennsylvania; cms8477@psu.edu

Given that each model ensemble member is initialized with different initial conditions, it is expected that individual members will capture the CPC EOF to varying degrees of success, thus serving as a proxy for real world natural variation. The internal variability of a model can thus provide insights to AMOC influence on CONUS precipitation variability by comparing groups of ensemble members that well capture the CPC EOF to those that do not. These two groups of ensemble members are quantified by correlating the ensemble member EOFs to the CPC EOF. The ensemble member EOFs with the highest (lowest) 20 percentile of correlation coefficients are averaged together to form the H20A (L20A) ensemble.

3. Analysis

Preliminary analysis determined the model members comprising H20A and L20A ensembles for each model. The CPC EOF is shown in Figure 1a and reveals a dipole mode of variability with two centers of action (CE) formed over the MW and SE CONUS. We define this EOF as the Midwestern-Southeastern (MISO) Precipitation Dipole. MISO is a pattern of opposing hydrologic extremes between the MW and SE CONUS, where flood-like conditions in the MW are opposed by drought-like conditions in the SE, and vice versa (Fig. 1a). The CPC MISO has an explained variance percentage of 13.7% (Fig. 1a), which is lower than the other variance percentages found in the model ensembles (Fig. 1b-1g). CESM2 H20A reveals a weak MISO-like pattern with the SE CE located further North than in CPC (Fig. 1b). Comparatively, CESM2 L20A indicates a dipole pattern between the eastern CONUS and the North American Monsoon (NAM) region over Arizona (Fig. 1c). Similarly, CanESM5 H20A reveals a MISO-like pattern, however the signal is weaker than CPC and CESM2 H20A. Instead CanESM5 H20A shows a more southern CE in the SE and an opposing CE along the northern Pacific coast (Fig. 1d-1e). Like CESM2 L20A, CanESM5 L20A also shows a dipole pattern between the NAM and eastern CONUS (Fig. 1e). ACCESS-ESM1.5 performs the least consistently revealing no dipole patterns in either ensemble (Fig. 1f-1g). The major difference between H20A and L20A is that the SE CE is pushed further northward in L20A than in H20A (Fig. 1f-1g). The consistency of the large SE CE between models indicates major NASH WR influence on CONUS precipitation.

Correspondence to: Christopher M. Sala, [1]Pennsylvania State University, College of Earth and Mineral Sciences, Department of Meteorology and Atmospheric Science, University Park, Pennsylvania; cms8477@psu.edu

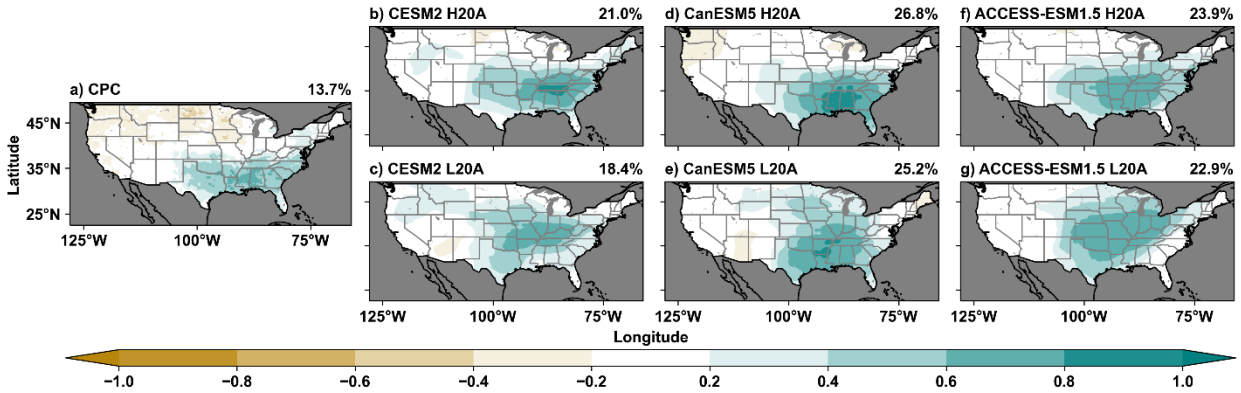

Figure 9: EOF analysis of the first principal component calculated over CONUS using summertime means from 1948 to 2014 using precipitation data from the CPC (a), CESM2 H20A (b), CESM2 L20A (c), CanESM5 H20A (d), CanESM5 L20A (e), ACCESS-ESM1.5 (f), and ACCESS-ESM1.5 (g). Numbers in the top right corner correspond to the explained variance percentage. Data Sources: Higgins et al. 2000, Danabasoglu 2019, Swart et. al. 2019, and Ziehn et al. 2019.

Model NASH WR analysis begins with the summertime 1948 to 2014 climatological ensemble-member mean NASH WR located for each model. Then the NASH WR is found for the H20A and L20A ensembles using the individual model's climatological geopotential height value of the NASH WR. Following the methodology of Li *et al*. (2013), the NASH WR is the intersection of the 850 hectopascal geopotential height contour that straddles the 86 degrees West longitude and the 0 meter per second zonal isotach. Results from this methodology indicate that there is little meridional movement of the NASH WR in these models, inconsistent with observations (Li *et al*., 2012; 2013; Fig. 2). CESM2 and CanESM5 H20As are extended westward relative to the individual multi-ensemble mean NASH WR position (Fig. 2a-2b), whereas the L20As exhibit eastward extensions (Fig. 2a-2b). The NASH WR westward extension is expected, however the location over land is not. Other studies indicate that the NASH WR is typically over the Gulf of Mexico (Li *et al*., 2012; 2013), as seen in ACCESS-ESM1.5 (Fig. 2c)**.** However, there is little to no zonal variability between ACCESS-ESM1.5 H20A and L20A, and both are extended west of the climatological NASH WR with L20A slightly more west than H20A (Fig. 2c).

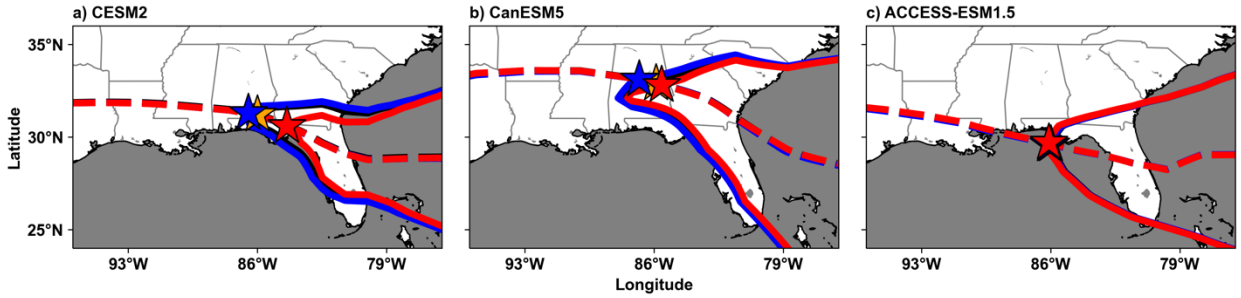

Figure 10: NASH WR summertime from 1948 to 2014 climatological locations indicated by stars for CESM2 (a), CanESM5 (b), and ACCESS-ESM1.5 (c). Solid contours represent the 850 hectopascal geopotential height that defines the NASH ridge and dashed contours represent the 0 meters per second zonal isotach. Black and gold indicate the multi-ensemble member mean, blue indicates H20A, and red indicates the L20A for a given model. Data Sources: Danabasoglu 2019, Swart et al. 2019, and Ziehn et al. 2019.

Correspondence to: Christopher M. Sala, [1]Pennsylvania State University, College of Earth and Mineral Sciences, Department of Meteorology and Atmospheric Science, University Park, Pennsylvania; cms8477@psu.edu

The NASH WR westward extension indicates NASH circulation changes caused by stronger tropical easterlies and mid-latitude westerlies in CESM2 and CanESM5, however there are minor changes between ensembles in ACCESS-ESM1.5 (not shown). Given the changes in NASH circulation strength between ensembles, we examine the differences between the AMOC base-states within the ensembles in Figure 3. CESM2 indicates that the AMOC base-state is weaker in H20A than L20A, specifically around 40 degrees North latitude corresponding with the NASH northern flank (Fig. 3a). CanESM5 H20A also indicates a weaker AMOC base-state, with slight strengthening north of 60 degrees North latitude and the weakest AMOC circulation occurring between 40 degrees North to 60 degrees North latitude. While not as clear and significant, ACCESS-ESM1.5 H20A shows a weaker AMOC base-state throughout the ocean (Fig. 3c). Although, Figure 3c shows that AMOC H20A is stronger from 40 degrees North to -60 degrees North at 1.5 kilometers depth.

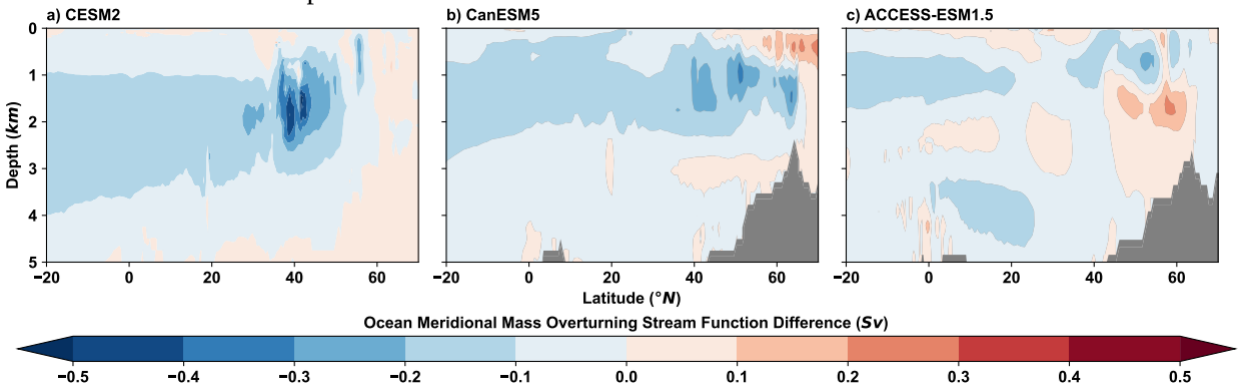

Figure 11: The difference (H20A – L20A) in annual 1948 to 2014 climatologies of the Atlantic Ocean meridional overturning stream function expressed as Sverdrup (Sv) for CESM2 (a), CanESM5 (b), and ACCESS-ESM1.5 (c). Data sources: Danabasoglu 2019, Swart et al. 2019, and Ziehn et al. 2019.

4. Conclusion

We hypothesized that a weakened AMOC base-state in CMIP6 historical model simulations may be responsible for increased hydrologic variability over CONUS by way of Bjerknes Compensation theory (BC; Bjerknes, 1964). BC states that atmospheric heat transport anomalies are balanced by oceanic heat transport anomalies (Bjerknes, 1964), thus the weaker AMOC base-state shown in H20A ensembles correspond with stronger NASH circulations which push the NASH WR further westward than those seen in L20A. This is most clearly seen in the CESM2 model which shows the farthest westward extension and the strongest AMOC base-state strength difference. A similar pattern is seen in the CanESM5 model, although the weaker AMOC base-state difference results in a less NASH WR westward extension. ACCESS-ESM1.5 shows the weakest difference in AMOC strength and yields almost no difference in NASH WR location between ensembles. Consequently, the changes in NASH WR locations are echoed in EOFs as CESM2 shows the most MISO-like pattern and ACCESS-ESM1.5 shows no MISO pattern at all. Therefore, we conclude that variations in MISO-like patterns in models indicate that model simulated MISO is a product of internal variation, and MISO-capturing ensemble members correspond with a NASH WR westward extension in association with a weakened AMOC base-state.

5. Data Availability Statement

All datasets used in this analysis are freely available online through the CPC climate data website (https://psl.noaa.gov/data/gridded/data.unified.daily.conus.html) and World Climate Research Programme model output website (https://esgf-node.llnl.gov/projects/cmip6/). The scripts used for the analyses are available to those who request it.

Correspondence to: Christopher M. Sala, [1]Pennsylvania State University, College of Earth and Mineral Sciences, Department of Meteorology and Atmospheric Science, University Park, Pennsylvania; cms8477@psu.edu

6. References

1. Bjerknes, J., 1964: Atlantic Air-Sea Interaction. *Advances in Geophysics*, **10**, 1-82, https://doi.org/10.1016/S0065-2687(08)60005-9.
2. Buckley, M.W., and J. Marshall, 2016: Observations, inferences, and mechanisms of the Atlantic Meridional Overturning Circulation: A review. *Rev. of Geophys.*, **54**, 5-63, https://doi.org/10.1002/2015RG000493.
3. Danabasoglu, G., 2019: NCAR CESM2 model output prepared for CMIP6 CMIP historical. Version 20190313. *Earth System Grid Federation*, https://doi.org/10.22033/ESGF/CMIP6.7627.
4. Dawson, A., 2016: eofs: A Library for EOF Analysis of Meteorological, Oceanographic, And Climate Data. *J. Open Res. Softw.*, **4**(1), e14, https://doi.org/10.5334/jors.122.
5. Higgins, R.W., W. Shi, E. Yarosh, and R. Joyce, 2000: Improved United States Precipitation Quality Control System and Analysis. NOAA, National Weather Service, National Centers for Environmental Prediction, *Climate Prediction Center ATLAS No. 7*, Camp Springs, MD 20746, USA.
6. Li, L., W. Li, and Y. Deng, 2013: Summer rainfall variability over the Southeastern United States and its intensification in the 21st century as assessed by CMIP5 models. *J. Geophys. Res: Atmospheres*, **118**(2), 340-354, https://doi.org/10.1002/jgrd.50136.
7. Li, L., W. Li, and Y. Kushnir, 2012: Variation of the North Atlantic subtropical high western ridge and its implication to the Southeastern US summer precipitation, *Clim. Dyn.*, **39**, 1401-1412, https://doi.org/10.1007/s00382-011-1214-y.
8. Swart, N. C., and Coauthors, 2019: CCCma CanESM5 model output prepared for CMIP6 CMIP historical. Version 20190306. *Earth System Grid Federation*, https://doi.org/10.22033/ESGF/CMIP6.3610.
9. Zhang, R., R. Sutton, G. Danabasoglu, G., Y.-O. Kwon, R. Marsh, S.G. Yeager, D.E. Amrhein, and C.M. Little, 2019: A Review of The Role of The Atlantic Meridional Overturning Circulation in Atlantic Multidecadal Variability and Associated Climate Impacts. *Reviews of Geophysics*, **57**(2), 316-375, https://doi.org/10.1029/2019RG000644.
10. Ziehn, T. and Coauthors, 2019: CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 CMIP historical. Version 20191115, *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.4272

Correspondence to: Christopher M. Sala, [1]Pennsylvania State University, College of Earth and Mineral Sciences, Department of Meteorology and Atmospheric Science, University Park, Pennsylvania; cms8477@psu.edu

Circulation Study: On the Relationship of Temperature and Precipitation with 500-Hectopascal Heights at the Weeks 3 to 4 Timescale

Cory F. Baggett[1], Emerson LaJoie[1], Johnna Infanti[1], Michael Goss[1,2]
[1]*Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland*
[2] Earth Resources Technology, Inc.; Laurel, Maryland

ABSTRACT

1. Introduction

Each Friday, forecasters at the Climate Prediction Center (CPC) are tasked with producing a Week 3-4 outlook for temperature and precipitation. Skillful Week 3-4 outlooks are highly desired by stakeholders across many fields such as hydrology, energy, and agriculture. However, forecast skill at this time range remains modest, which can limit the usefulness of these outlooks to stakeholders. Thus, there is a need to continually improve our forecast tools, such as guidance from dynamical and statistical models. Forecasters primarily consider three variables from these tools: 500-hectopascal geopotential heights, 2-meter temperature, and surface precipitation. Oftentimes, the forecasters will scrutinize fields of 500-hectoPascal heights to provide physical justification as to how the circulation during the forecast period will impact temperature and precipitation. Here, we wish to assist the forecasters in this process by conducting a thorough study on the observed relationship between temperature and precipitation with 500-hectopascal heights on two-week timescales, and how this relationship seasonally evolves. Further, we examine the relationship between observed temperature and precipitation anomalies conditioned on the Global Ensemble Forecast System Version 12's (GEFSv12; Zhou et al. 2022) Week 3-4 forecast of 500 hectopascal height anomalies.

2. Data and Methods

We calculated three sets of temporal correlation coefficients between two-week averaged observed 2-meter temperature anomalies with 1) 500- hectopascal height anomalies, 2) the meridional gradient of 500- hectopascal height anomalies (westerly/easterly winds), and 3) the zonal gradient of 500-hectopascal height anomalies (southerly/northerly winds). This analysis is conducted across the 1991-2020 climate normal period for each grid point across the contiguous U.S. and Alaska (CONUS/AK) and for overlapping 3-month seasons. A similar analysis is conducted for two-week summed precipitation. Temperature and 500- hectopascal height anomalies are defined against their climatological mean, whereas precipitation anomalies are defined against their climatological median. 2-meter temperature derives from CPC's Global Temperature dataset (Fan et al. 2008); precipitation derives from CPC's Global Unified Gauge-Based Analysis of Daily Precipitation dataset (Chen et al. 2008); and the 500-hectopascal heights derive from NCEP/NCAR Reanalysis Version 1 (Kalnay et al. 1996).

Next, we calculated 3x3 contingency tables of observed temperature anomalies versus observed 500- hectopascal height anomalies. These tables indicate the frequency at which temperatures are below, near, or above normal during periods of below, near, or above normal 500- hectopascal heights. Here, the thresholds for below, near, and above normal are defined as the lower, middle, and upper thirds of the climatological distribution, respectively. We also produced 3x3 contingency tables of observed temperature anomalies versus Week 3-4 forecasts of 500- hectopascal height anomalies by the GEFSv12. This analysis is expected to inform the forecaster of how confidently one can apply the observed

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

relationship to forecasted values of 500- hectopascal heights. Again, a similar analysis is conducted for precipitation.

Below, we will highlight some results of the correlation analysis, along with a brief result that is derived from the contingency table statistics. The full correlation analysis and contingency table statistics are provided to the forecaster as an internal, experimental web-based tool comprised of a large database of figures; see Data Availability Statement.

3. Analysis

Results of the correlation analysis show a large positive relationship between 500- hectopascal heights and temperatures across all regions and seasons (Fig. 1a). Thus, above normal temperatures are associated with above normal heights, and below normal temperatures are associated with below normal heights. However, both the Pacific Northwest and Southern Plains exhibit weaker relationships, particularly during the winter (Fig. 1b). Furthermore, we find that anomalous southerlies are normally associated with above normal temperatures across most of CONUS/AK (Fig. 2a), with notable exceptions being the Southwest and the Southeast during the summer months (Fig. 2b), when moisture transport presumably tends to produce below normal temperatures. Anomalous westerlies are generally associated with below normal temperatures across most of CONUS, indicative of a jet stream displaced to the south (Fig. 3a). In contrast, the Pacific Northwest and Intermountain West often experience above normal temperatures with anomalous westerlies during winter, due to onshore flow immediately along the coast and downsloping off the north-south-oriented mountain ranges further east (Fig. 3b).

For precipitation, we find that its correlation with the gradients of 500-hectopascala heights is typically as large as or larger than its direct relationship with 500-hectopascal heights. Typically, above normal heights are weakly associated with below normal precipitation across most of CONUS/AK (Fig. 4a). In comparison, anomalous southerlies are associated with above normal precipitation (Fig. 4b), while anomalous westerlies have a subtler relationship due to the rain shadows produced by CONUS's mountain ranges (Fig. 4c).

Examining the 3x3 contingency tables of observed temperature anomalies versus Week 3-4 forecasts of 500-hectopascal height anomalies by the GEFSv12, we find that the model does not have as strong of a relationship as seen in observations. For example, while above normal temperatures generally occur more than twice their climatological frequency of 33.3% when above normal 500-hectopascal heights are observed (Fig. 5a), they are observed much less frequently when the GEFSv12 forecasts above normal 500-hectopascal heights at Week 3-4 (Fig. 5b). This result is expected because model biases, resolution limitations, and errors in initial conditions cause model errors to grow over time, and thus their Week 3-4 500-hectopascal height forecasts are often wrong.

4. Conclusion

We conclude by saying that while most of this study's results may be considered intuitive to a seasoned forecaster at CPC, its more nuanced findings will likely improve our fundamental understanding of the role that the circulation should play in the creation of our Week 3 -4 temperature and precipitation outlooks. The observed relationship between 500-hectopascal heights and its implied circulation provides an objective, physical reasoning for the expected temperature and precipitation anomalies in a given region and season. We can use this reasoning to gain confidence in the expected temperature and precipitation outcome at Week 3-4 given a model's forecasted depiction of the circulation.

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

5. Data Availability Statement

Model data from the GEFSv12 is available from NCEP Central Operations (https://www.nco.ncep.noaa.gov/pmb/products/gens/ and Amazon Web Services (https://registry.opendata.aws/noaa-gefs-reforecast/). Reanalysis data, including CPC global temperature, CPC global precipitation, and NCEP/NCAR Reanalysis V1 500-hPa heights are available from NOAA's Physical Science Laboratory (https://psl.noaa.gov/data/). Code and the database of figures can be made available upon request to the corresponding author.

6. References

1. Chen, M., W. Shei, P. Xie, V.S.S. Silva, V.E. Kousky, R.W. Higgins, and J.E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research*, **113**, D04110, https://doi.org/10.1029/2007JD009132.
2. Fan, Y., and H. van den Dool, 2008: A global monthly land surface air temperature analysis for 1948-present. *Journal of Geophysical Research*, **113**, D01103, https://doi.org/10.1029/2007JD008470.
3. Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, **77**, 437-472, https://doi.org/10.1175/1520-0477(1996)077<0437: TNYRP>2.0.CO;2.
4. Zhou, X., and Coauthors, 2022: The development of the NCEP Global Ensemble Forecast System Version 12. *Weather and Forecasting*, **37**, 1069-1084, https://doi.org/10.1175/WAF-D-21-0112.1.

7. Figures

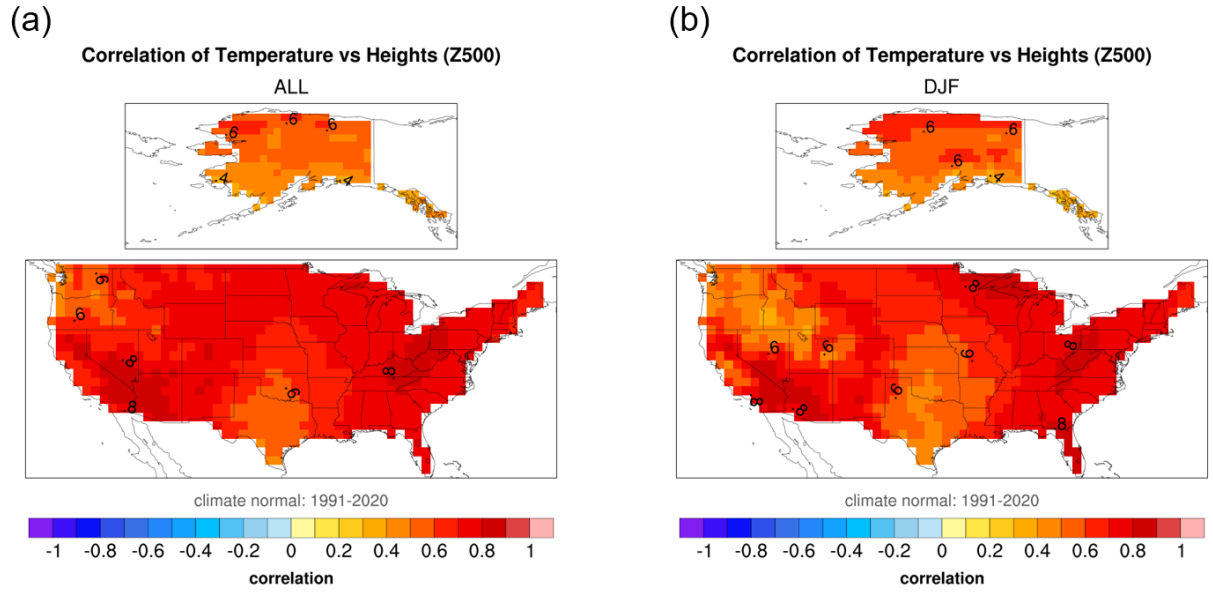

Figure 12:. Correlations of two-week averaged 2-meter temperature with two-week averaged 500-hectopascal height anomalies are shown over CONUS/AK for (a) all seasons and (b) December through February (DJF; Fan et al. 2008; Kalnay et al. 1996).

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

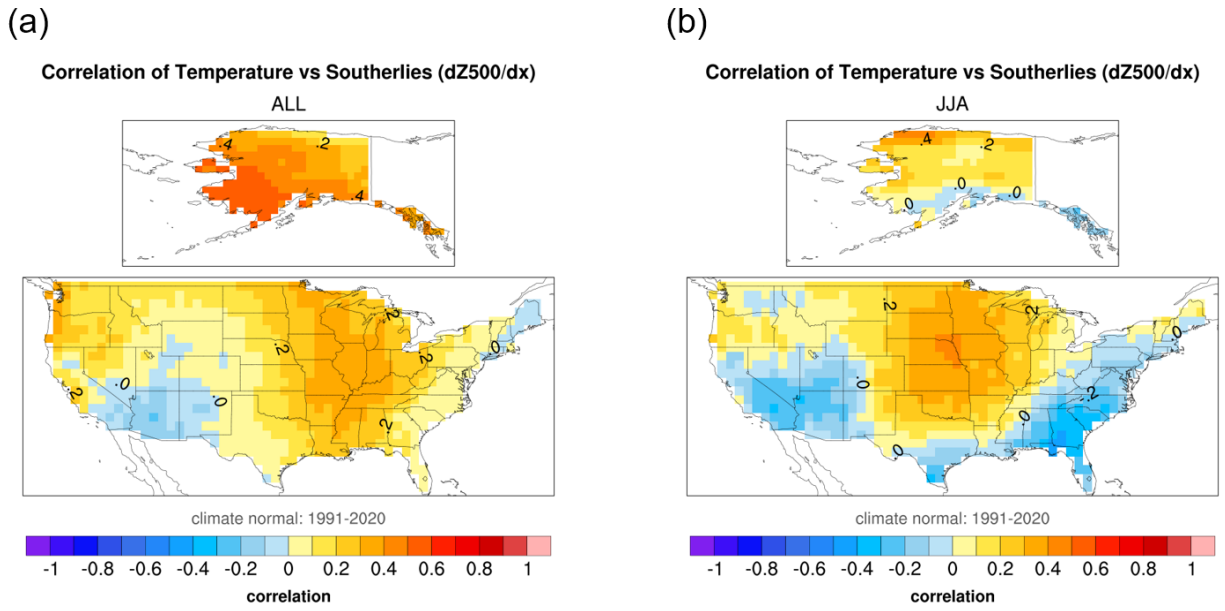

Figure 13: Correlations of two-week averaged temperature with two-week averaged meridional wind, as derived from the zonal gradient in 500-hectopascal heights (dZ500/dx), are shown over CONUS/AK for (a) all seasons and (b) June through July (JJA; Fan et al. 2008; Kalnay et al. 1996).

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

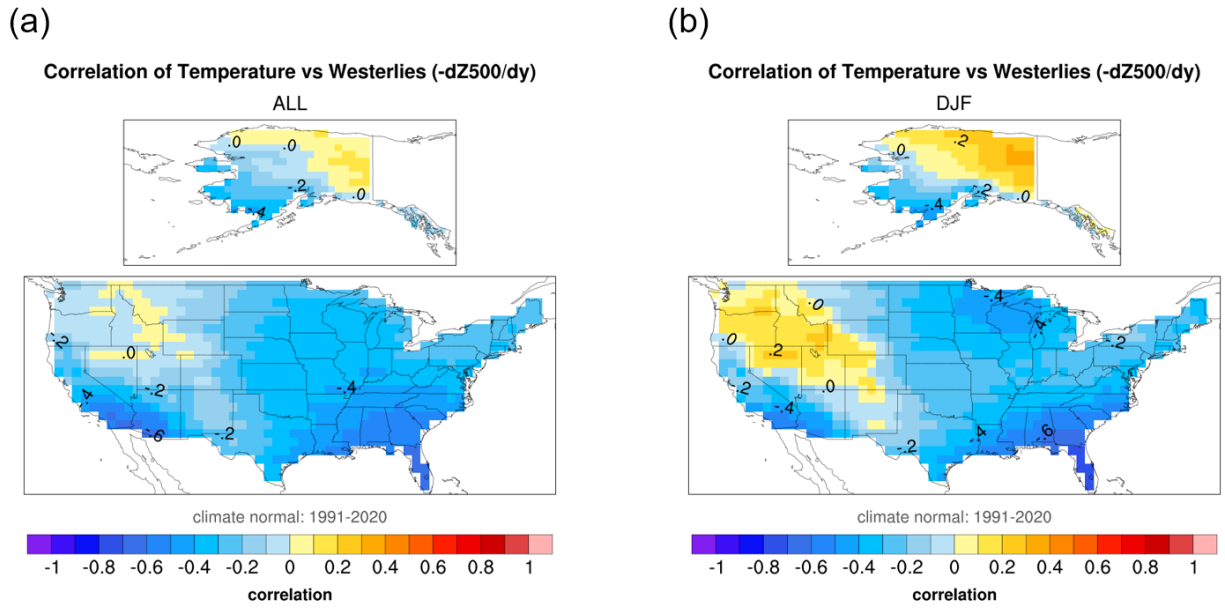

Figure 14:. Correlations of two-week averaged temperature with two-week averaged zonal wind, as derived from the meridional gradient in 500-hectopascal heights (-dZ500/dy), are shown over CONUS/AK for (a) all seasons and (b) December through February (DJF; Fan et al. 2008; Kalnay et al. 1996).

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

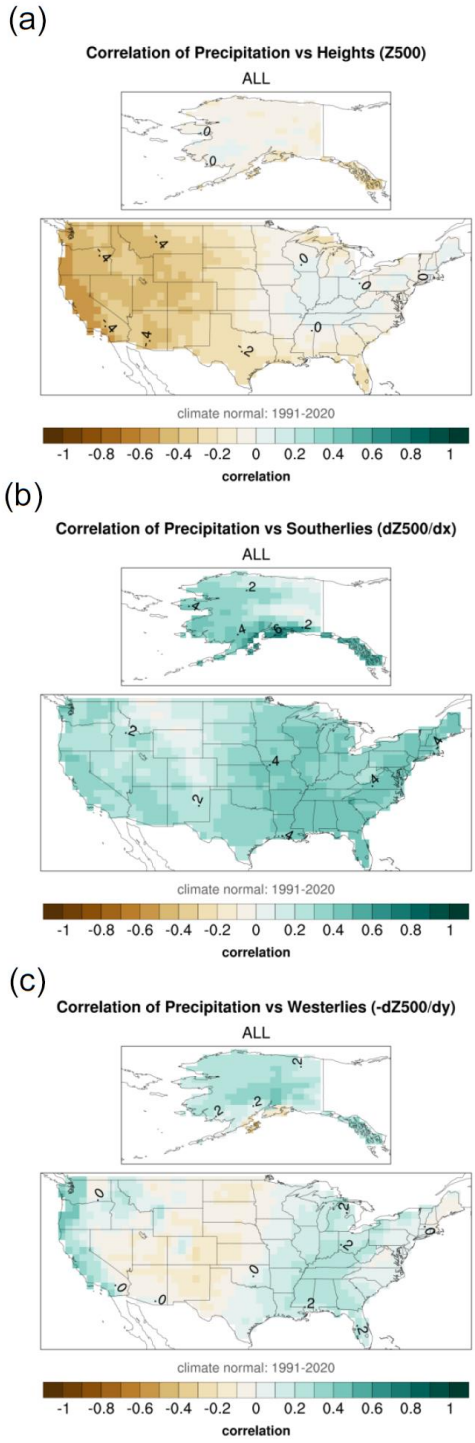

Figure 15: Correlations of two-week summed precipitation with two-week averaged (a) 500-hectopascal heights, (b) meridional gradient in 500-hectopascal heights (dZ500/dx), and (c) zonal gradient in 500-hectopascal heights (-dZ500/dy) are shown over CONUS/AK for all seasons (Chen et al. 2008; Kalnay et al. 1996).

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

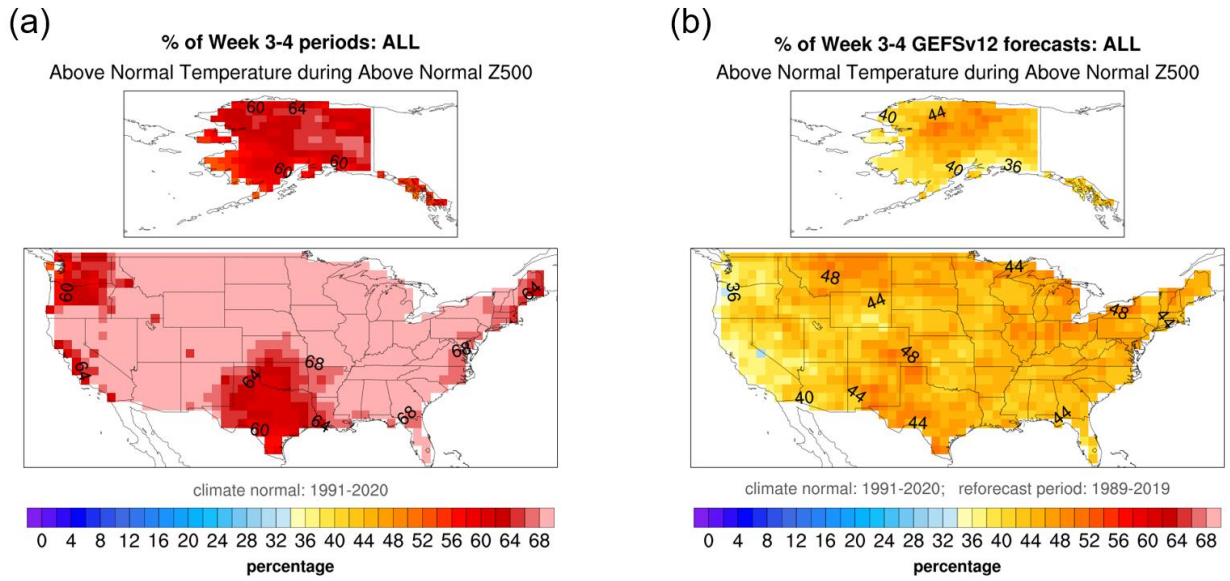

Figure 16: Panel (a) depicts the frequency at which temperatures are observed to be in the upper third of the climatological distribution conditioned on when 500-hPa heights are observed to be in the upper third of the climatological distribution during all seasons. Panel (b) depicts the frequency at which temperatures are observed to be in the upper third of the climatological distribution conditioned on when 500 hectopascal heights are forecasted to be in the upper third of the climatological distribution by the GEFSv12 at Week 3-4 during all seasons (Fan et al. 2008; Kalnay et al. 1996; Zhou et al. 2022)

Correspondence to: Cory F. Baggett, Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland; cory.baggett@noaa.gov

Trends from Reanalyses: Progress over the last 10 years

Wesley Ebisuzaki[1], Li Zhang[1,4], Arun Kumar[1], Jeffrey Whitaker[2], Jack Woollen[3,5]
[1]NOAA CPC, College Park, Maryland
[2]NOAA Physical Sciences Division, Boulder, Colorado,
[3]NOAA EMC, College Park, Maryland,
[4]ERT, Laurel, Maryland,
[5]LYNKER, Leesburg, Virginia

ABSTRACT

1. Introduction

The current global atmospheric reanalyses of 2022 from European Centre for Medium-Range Weather Forecasts (ECMWF), Japan Meteorological Agency (JMA), NASA, and the National Centers for Environmental Prediction (NCEP) are compared with reanalyses from systems that were available in 2012. We found that three of the modern reanalyses demonstrated similar trends in many of the large-scale averages. The consistency of the 2012 reanalyses was weaker, and could not resolve the expected anthropogenic forcing.

2. Data and Methods

The monthly means were analyzed from:

- CFSR (Climate Forecast System Reanalysis): Saha et al (2010)
- CORe (Conventional Observation Reanalysis): Ebisuzaki et al (2021)
- ERA-5 (ECMWF Reanalysis version 5): Hersbach et al (2020)
- ERA-interim (ECMWF Reanalysis Interim): Dee et al (2011)
- JRA-25 (Japanese 25-year Reanalysis): Kazutoshi et al (2007)
- JRA-55 (Japanese 55-year Reanalysis): Kobayashi et al (2015)
- MERRA (Modern-Era Retrospective Analysis for Research and Applications): Rienecker et al (2011)
- MERRA2 (Modern-Era Retrospective Analysis for Research and Applications Version 2): Galero et al (2017)

NOAA's Conventional Observational Reanalysis (CORe) (Ebisuzaki, et al, 2021), has completed processing data for the years 1950-2020, and we are working towards making it an operational product. During the evaluation of CORe, we found significant improvements in how the reanalyses systems from 2022 compared with reanalysis systems that were current in 2012.

Correspondence to: Wesley Ebisuzaki, NOAA CPC, College Park, Maryland;
wesley.ebisuzaki@noaa.gov

Table 1: The leading global reanalyses for the 1979-2012 period, as of 2012.

Reanalysis System	Period of Record	Model Members
CFSR	1979-ongoing	NCEP, CFSv2 for 2010+
ERA-interim	1979-2019	ECMWF, replaced by ERA-5
JRA-25	1979-2014	JMA and CRIEPI, replaced by JRA-55
MERRA	1979-2016	NASA, replaced by MERRA-2

Table 2: The leading global reanalyses for the 1980-2021 period, as of 2022.

Reanalysis System	Period of Record	Model Members
CORe	1950-ongoing	NCEP, working to make it operational
ERA-5	1950-ongoing	ECMWF, 1950-1957 is preliminary
JRA-55	1958-ongoing	JMA, JMA is in production of a replacement
CFSR	1979-ongoing	NCEP, CFSv2 for 2010+
MERRA2	1980-ongoing	NASA

A researcher from 2012 could have produced Figure 1 using data available in 2012. The top plot shows the 500 hectopascal global temperature anomaly (degrees Celsius) for CFSR (red), ERA-interim (black), JRA-25 (green) and MERRA (blue) removing the 1981-2020 climatology. The bottom plot of figure 1 shows the same except it shows the difference from ERA-interim. Assuming the anthropogenic forcing is O (0. Order(0.1C) degrees Celsius per decade), then you can expect an anthropogenic forcing of O (0.3/33years) over the period covered by the plot. Since the reanalyses (bottom plot, fig. 1) show differences greater than 0.3, the individual reanalyses are not resolving the trend well.

Correspondence to: Wesley Ebisuzaki, NOAA CPC, College Park, Maryland;
wesley.ebisuzaki@noaa.gov

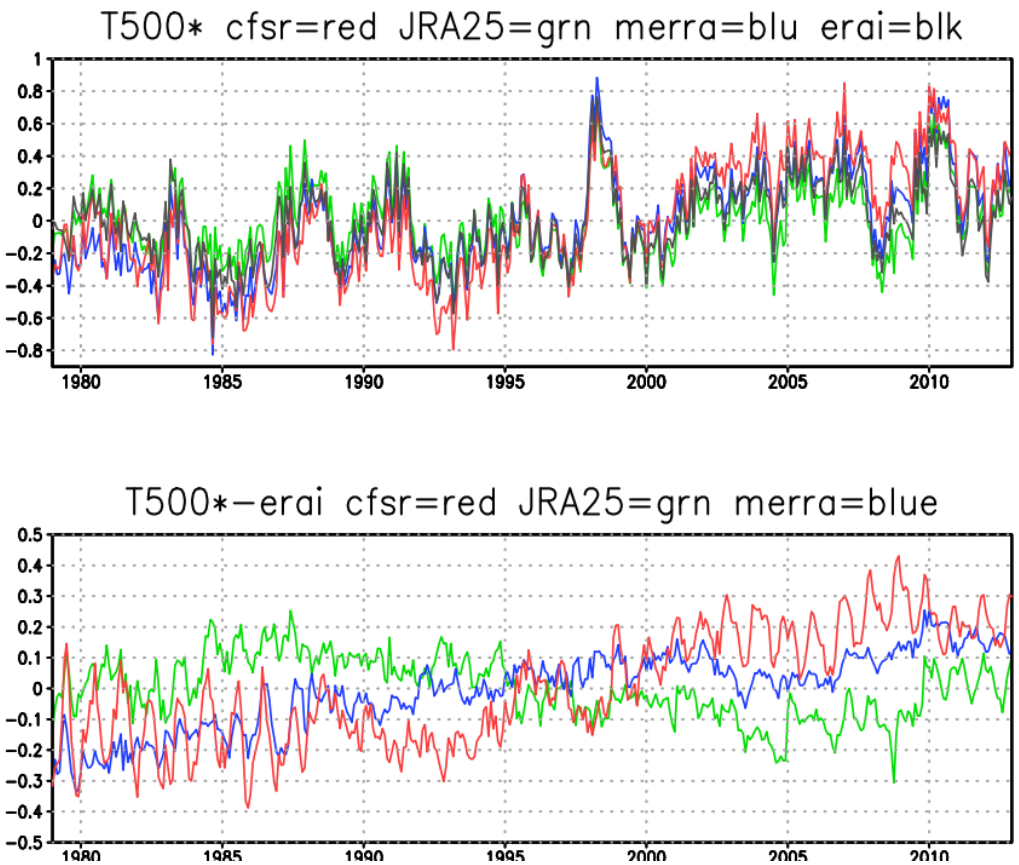

Figure 17: Top: Global mean 500 hectopascal temperature anomaly (degrees Kelvin) for CFSR (red), JRA-25 (green), MERRA (blue) and ERA-interim (black) using 1981-2010 climatologies. Bottom: similar to top except the time series are the deviations from ERA-interim anomalies.

Correspondence to: Wesley Ebisuzaki, NOAA CPC, College Park, Maryland;
wesley.ebisuzaki@noaa.gov

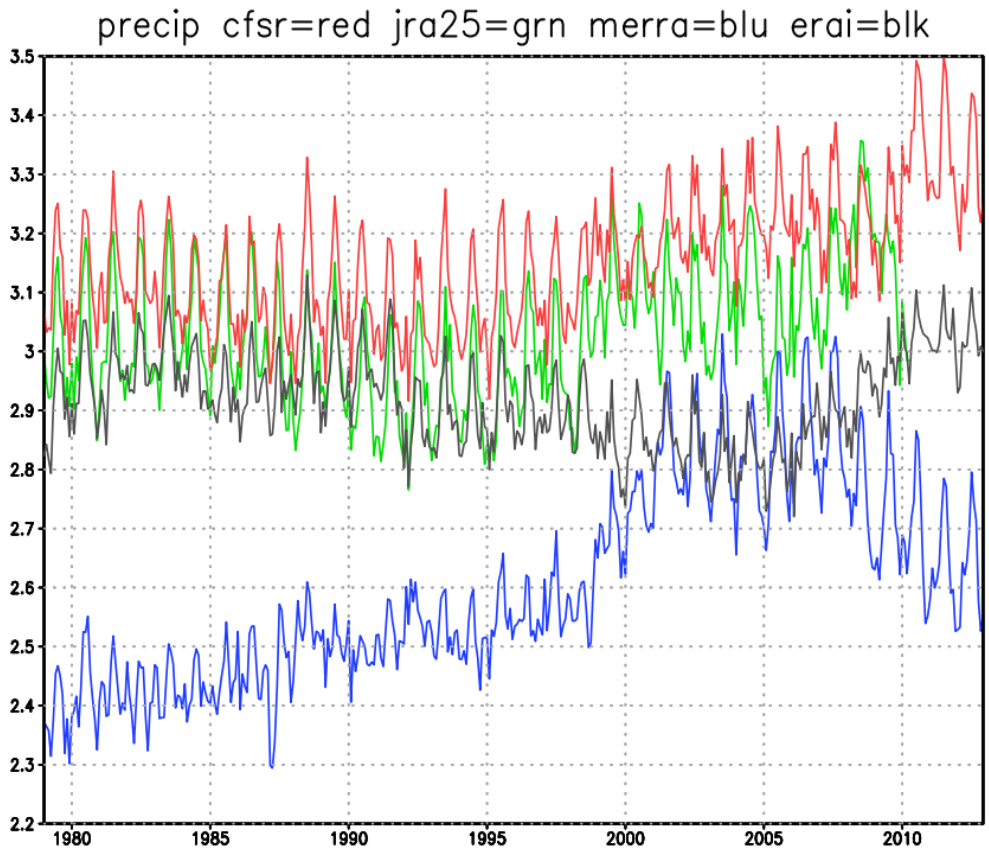

Figure 18: The global mean precipitation anomaly (millimeters/day) from CFSR (red), ERA-interim (black), JRA-25 (green) and MERRA (blue), removing 1981-2010 climatologies. The reanalyses from 2012 showed no consistency.

In 2022, the situation looks different. Figure 3 shows the difference in the 500 hectopascal temperature anomalies of CORe and ERA-5 (top plot, red), JRA-55 and ERA-5 (top plot, green), CFSR and ERA-5 (bottom plot, black) and MERRA-2 and ERA-5 (bottom plot, purple). As seen in the top plot of Figure 3, CORe and JRA-55 are within 0.1 degree Celsius of ERA-5. From the bottom plot of Figure 3, CFSR is within 0.5 degrees Celsius of ERA-5, and MERRA2 is within 0.3 degrees Celsius of ERA-5. The top plot shows that CORe, ERA-5 and JRA-55 are very consistent with each other. The bottom plot shows that older CFSR and MERRA-2 show more differences from the CORe/ERA-5/JRA-55 set of reanalyses as measured by the root mean square differences.

Correspondence to: Wesley Ebisuzaki, NOAA CPC, College Park, Maryland;
wesley.ebisuzaki@noaa.gov

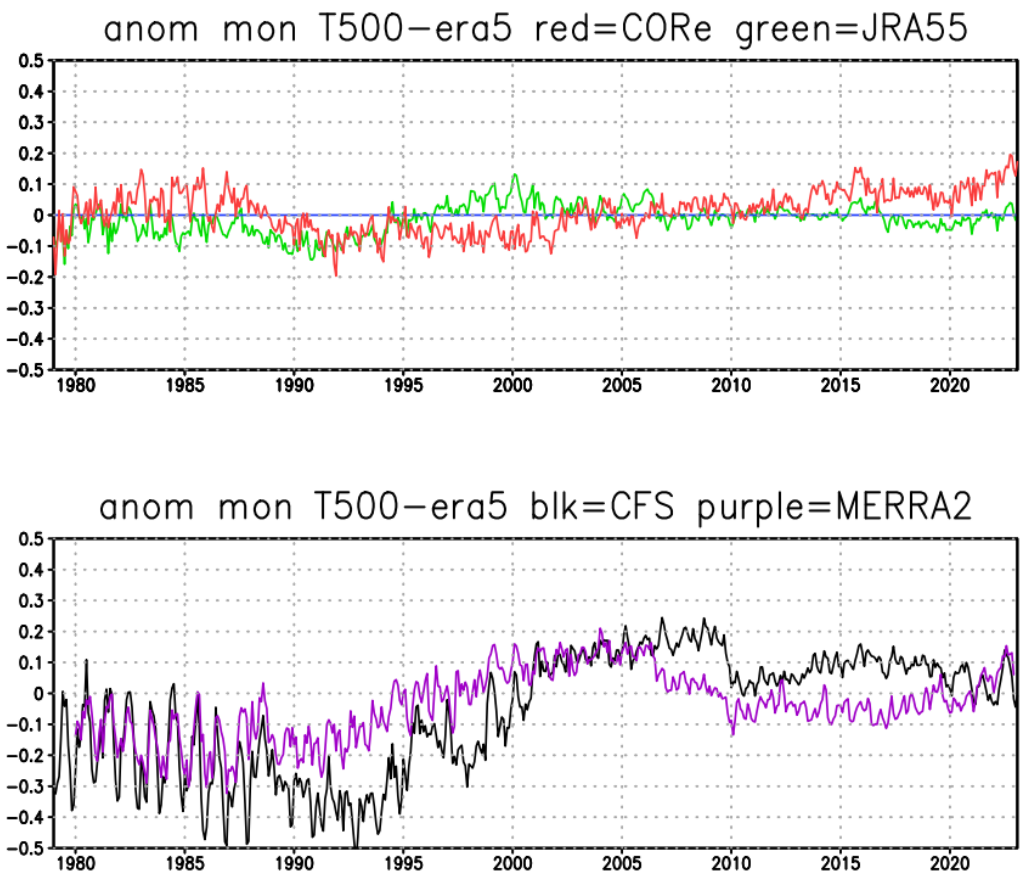

Figure 19: Top: global mean 500 hectopascal temperature anomaly for CORe (red) and JRA-55 (green) minus the corresponding ERA-55 values using 1991-2020 climatologies. Bottom: Similar to top except for CFSR (black) and MERRA2 (purple).

Figure 4 is similar to Figure 3 except it shows the global precipitation anomaly (millimeters per day). CORe (red), ERA-55 (blue) and JRA-55 (green) agree to 0.1 millimeters per day and show similar time series (top plot, Figure 4). The bottom plot shows that precipitation from CFSR (black) and MERRA2 (purple) are not as consistent as the newer reanalyses. For the rest of the figures, we will only consider CORe, ERA-5 and JRA-55.

Correspondence to: Wesley Ebisuzaki, NOAA CPC, College Park, Maryland;
wesley.ebisuzaki@noaa.gov

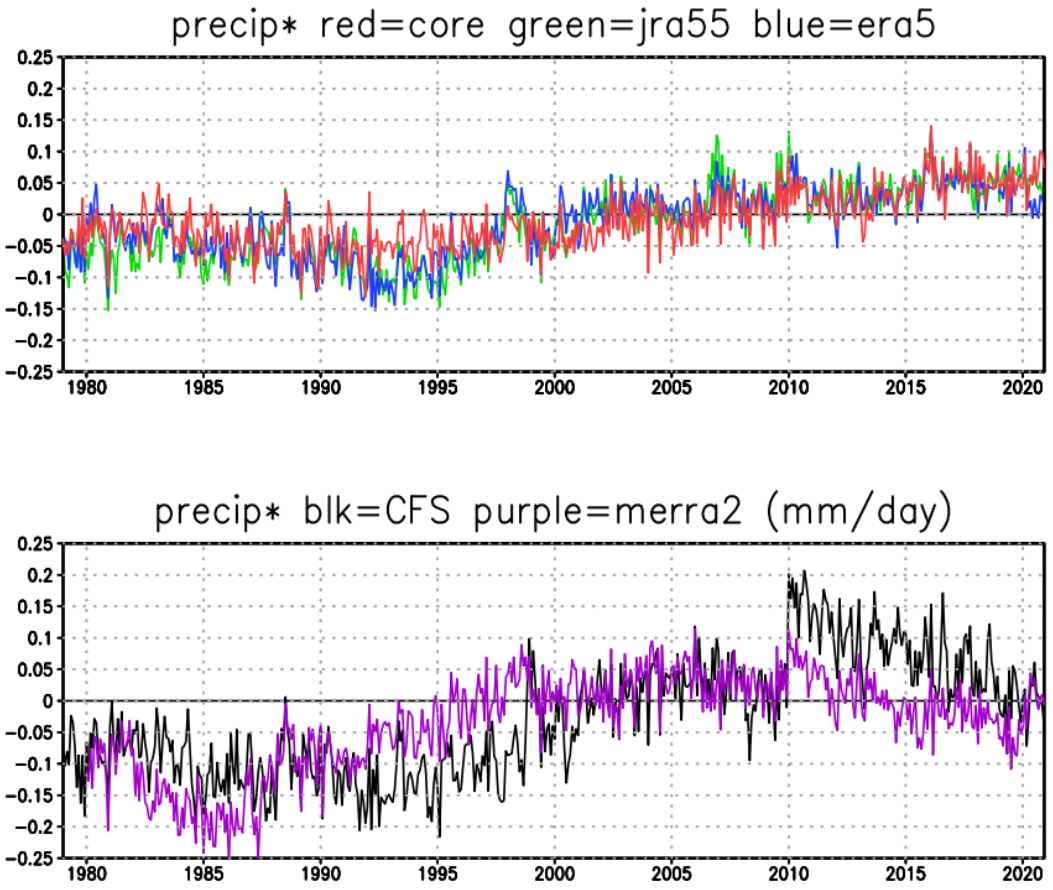

Figure 20: Top: global-mean precipitation anomaly for CORe (red), JRA-55 (green) and ERA-5 (blue) using 1991-2020 climatologies. Bottom: same as top except for CFSR (black), and MERRA2 (purple).

Figure 5 shows a smoothed global 500 hectopascal temperature anomaly from 1950-2020. For smoothing, a one year running mean was used. Note that the preliminary ERA-5 was used for 1950-1957. CORe (red), ERA-5 (black) and JRA-55 (green) agree very well in the later period. CORe and ERA-5 have difference up to 0.1 degrees Celsius, which occurred in the 1950's, so the three reanalyses have a consistent 500 hectopascal global temperature trend with differences that are much smaller than the expected anthropogenic forcing.

Correspondence to: Wesley Ebisuzaki, NOAA CPC, College Park, Maryland;
wesley.ebisuzaki@noaa.gov

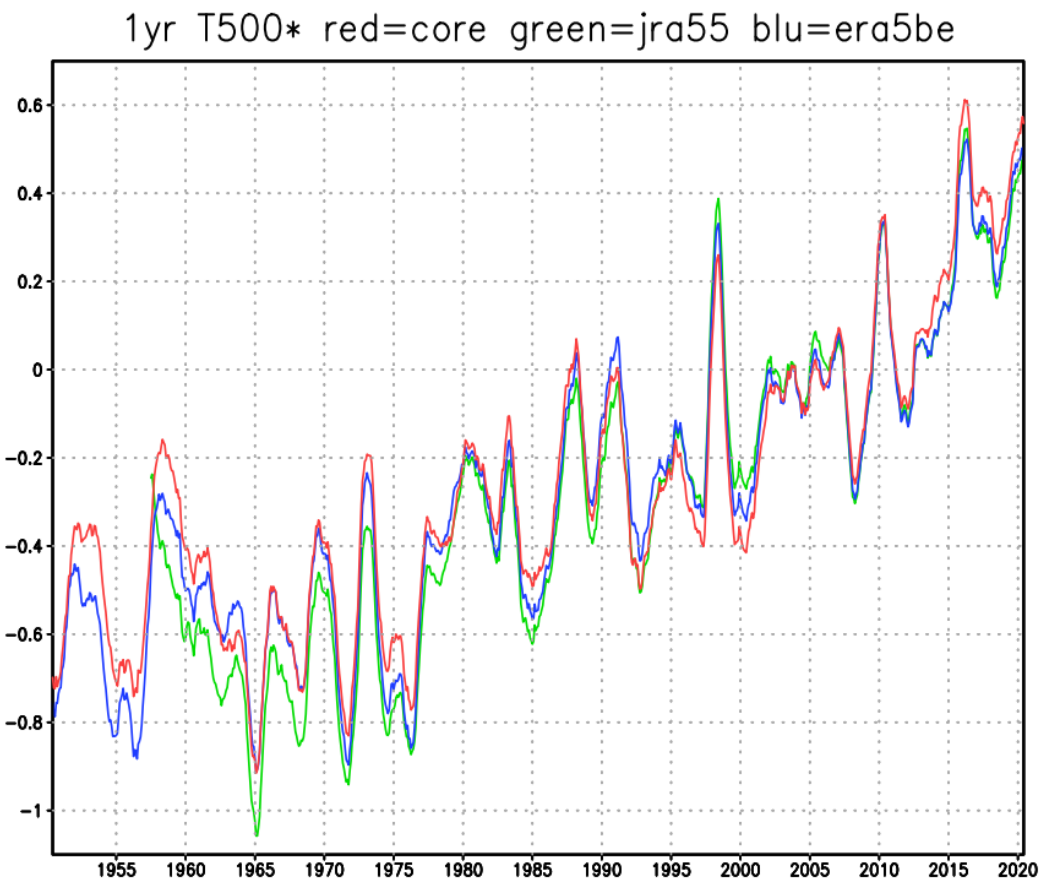

Figure 21: Global-mean 500 hectopascal temperature anomaly with a 1-year running mean smoothing for CORe (red), JRA-55 (green) and ERA-5 (blue). Used 1991-2020 climatologies, and ERA-5 used 1950-1957 preliminary analyses.

Correspondence to: Wesley Ebisuzaki, NOAA CPC, College Park, Maryland;
wesley.ebisuzaki@noaa.gov

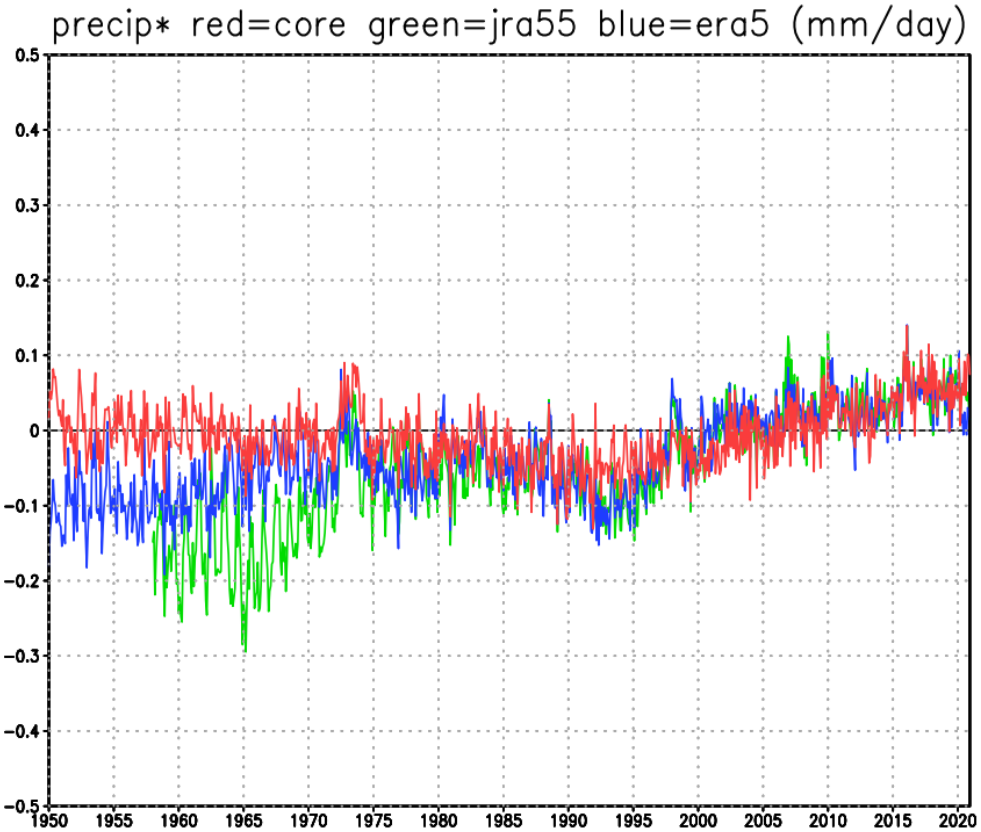

Figure 22: Global-mean precipitation anomaly for CORe (red), JRA-55 (green) and ERA-5 (blue) using 1991-2020 climatologies.

Figure 6 shows the global precipitation anomaly from 1950-2020 for CORe (red), ERA-5 (blue) and JRA-55 (green). From 1973 onwards, the three reanalyses show good consistency, with a slight increase in precipitation with time.

Figure 7 shows the globally averaged anomalous temperature as a function of time and pressure. There is much agreement in the troposphere. CORe is warmer in the early stratosphere; however, much of the warmth is coming from the poorly observed southern hemisphere.

Correspondence to: Wesley Ebisuzaki, NOAA CPC, College Park, Maryland;
wesley.ebisuzaki@noaa.gov

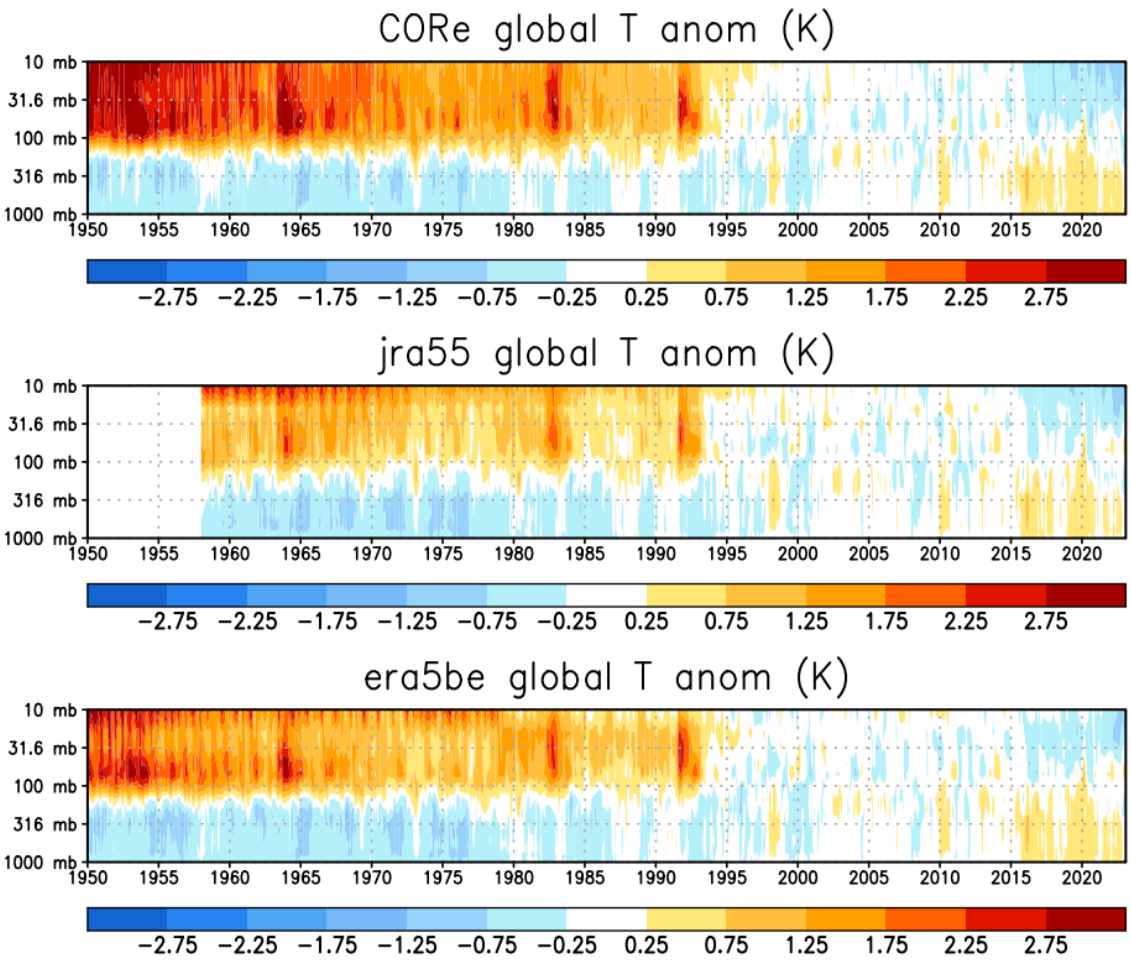

Figure 23: Global-mean temperature anomaly as function of pressure and time for CORe (top), JRA-55 (middle) and ERA-5 (bottom). Used 1991-2020 climatologies and ERA-5 1950-1957 preliminary analyses.

Figure 8 is similar to figure 7 but for 30 degrees North-60 degrees North, and from 20-2 hectopascals. 30 degrees North-60 degrees North was chosen because of there are more radiosondes in that latitude band. This plot shows that these reanalyses have some notable differences. For example, JRA-55 shows a warmth in the 2-4 hectopascal region in the late pre-satellite period. ERA-5 shows a cold anomaly in the 6-2 hectopascal region from 1967-1973 that is not seen in the other reanalyses. In the 10-20 hectopascal levels, the reanalyses have smaller root mean square differences.

Correspondence to: Wesley Ebisuzaki, NOAA CPC, College Park, Maryland;
wesley.ebisuzaki@noaa.gov

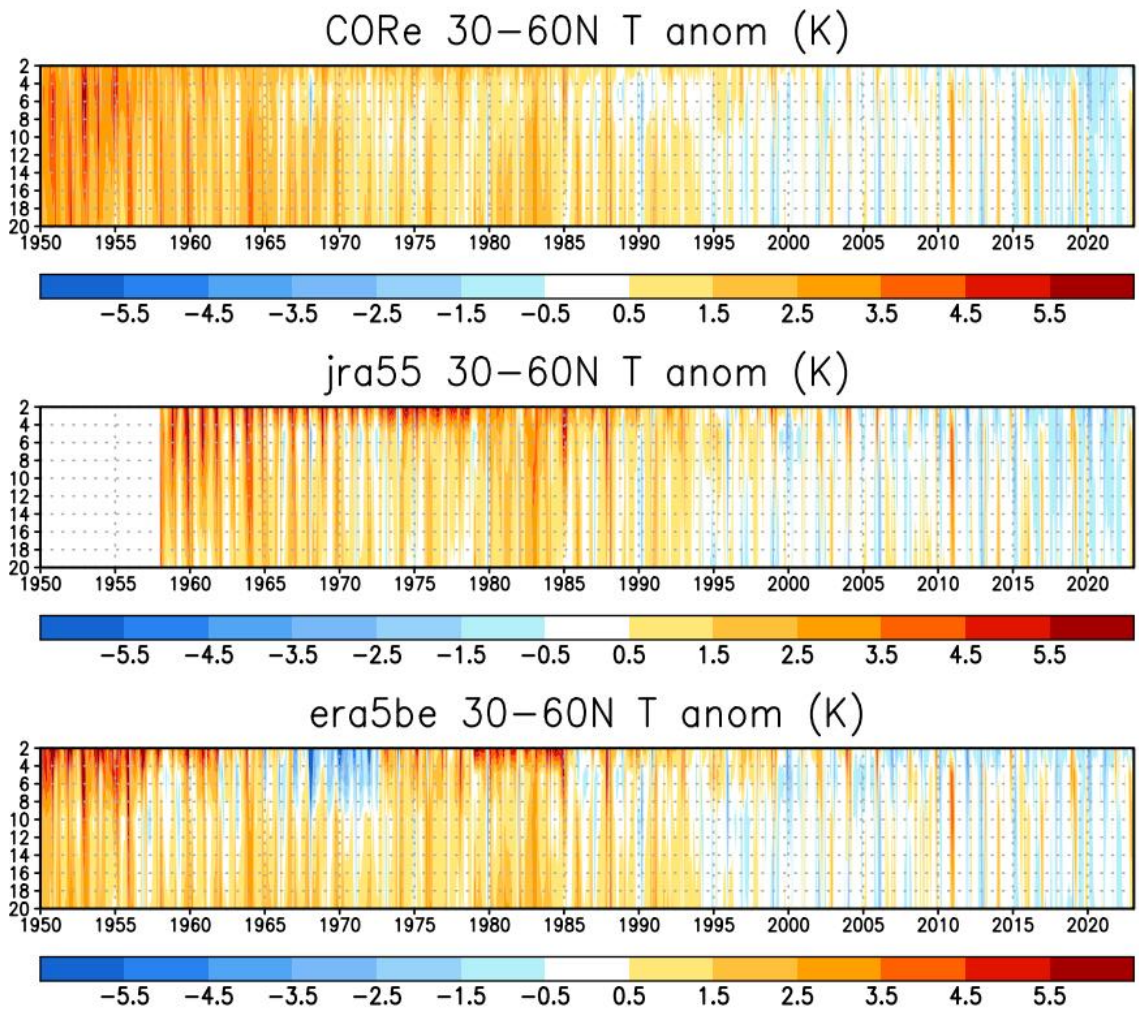

Figure 24: Similar to Figure 7, except for 30 degrees North-60 degrees North means, and 20-1 hectopascal levels.

3. Conclusion

Three out of 5 modern reanalyses (CORe, ERA-5, JRA-55) have a consistent trend of large-regional averages for many variables. They have different data assimilation systems (CORe: Ensemble Kalman Filter, ERA-5: hybrid 4D-var, JRA-55: 4D-var) and different treatment of satellite data (CORe: only AMV (Atmospheric Motion Vectors), ERA-5: all sky+AMV, JRA-55: clear sky+AMV). The three systems have different models, and quality control (QC) procedures. So, these newer reanalysis systems are quite varied. This suggests that the reanalyses have improved and are converging to a common trend. This improves the utility of the newer reanalyses to examine trends.

Correspondence to: Wesley Ebisuzaki, NOAA CPC, College Park, Maryland;
wesley.ebisuzaki@noaa.gov

4. Data Availability

CORe is in the public domain and data availability is being arranged. CFSR is in the public domain and is available from National Centers for Environmental Information (NCEI). ERA-interim and ERA-5 are available from the European Center from Medium Range Weather Forecasting (ECMWF) through Copernicus for ERA-5. JRA-25 and JRA-55 are available from the Japan Meteorological Agency (JMA). MERRA and MERRA-2 are available from the National Aeronautics and Space Administration, Goddard Space Flight Center (NASA/GSFC) by https://gmao.gsfc.nasa.gov/reanalysis/MERRA, and https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2.

5. References

1. Dee, D. Pand Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. R. Meteorol. Soc.*, 137, 553-597, https://doi.org/10.1002/qj.828.
2. Ebisuzaki, W., et al, 2021: A Conventional Observation Reanalysis (CORe) for Climate Monitoring. Extended Summary, Climate Prediction S&T Digest, 45th NOAA Climate Diagnostics and Prediction Workshop, Virtual Online, DOC/NOAA, 56-59, https://doi.org/10.25923/tpfe-4n87.
3. Galero, R and Coauthors. 2017: MERRA-2 Overview: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Clim., 30 (14), https://doi.org/10.1175/JCLI-D-16-0758.1.
4. Hersbach, H. and Coauthors, 2020: The ERA5 Global Reanalysis, Quarterly Journal of the Royal Meteorological Society 146 (730), https://doi.org/ 10.1002/qj.3803.
5. Kazutoshi, O and Coauthors,2007: The JRA-25 Reanalysis, Journal of the Meteorological Society of Japan. Ser. II,85 (3), 369-432.
6. Kobayashi, S. and Coauthors, 2015: The JRA-55 Reanalysis: General specifications and basic characteristics, J. Meteor. Soc. Japan, 93(1)-5-48.
7. Rienecker, M.M and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*,24 (14), 3624-3648, https://doi.org/10.1175/JCLI-D-11-00015.1.
8. Saha, S and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. Bull. Amer. Meteor. Soc.,91 (8), 1015–1058, https://doi.org/10.1175/2010BAMS3001.1.

Correspondence to: Wesley Ebisuzaki, NOAA CPC, College Park, Maryland;
wesley.ebisuzaki@noaa.gov

Trends in Global Tropical Cyclone Activity: 1990-2021

Philip J. Klotzbach[1], Kimberly M. Wood[2*], Carl J. Schreck III[3], Steven G. Bowen[4], Christina M. Patricola[5], and Michael M. Bell[1]

[1]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA.
[2]Department of Geosciences, Mississippi State University, Mississippi State, MS, USA.
[3]Cooperative Institute for Satellite Earth System Studies (CISESS), North Carolina State University, Asheville, NC, USA.
[4]Gallagher Re, Chicago, IL, USA.
[5]Department of Geological and Atmospheric Sciences, Iowa State University, Ames, IA, USA
*Klotzbach and Wood contributed equally to this work.

ABSTRACT

This paper examines trends in global tropical cyclone activity from 1900-2021. We found an increasing trend in short-lived (2 days or less) named storms, likely due to improved satellite technology. Tropical cyclone-related damage has also increased, with most of this increase attributed to growth of population and infrastructure (e.g., exposure) in coastal regions. We also found an increase in tropical cyclones intensifying by greater than or equal to 50 knots per 24 hours, which is likely fueled by warmer sea surface temperatures due in part to climate change.

As a counterpoint to these increases, we found decreasing trends in global hurricane-strength (greater than or equal to 64 knots) tropical cyclones as well as Accumulated Cyclone Energy – an integrated metric accounting for storm intensity, frequency and duration (Figure 1). While North Atlantic hurricane activity has increased during this period, Pacific hurricane/typhoon activity has decreased. Since the Pacific Ocean is a much larger ocean with higher oceanic heat content, decreases in Pacific activity swamp the signal from increases in North Atlantic activity. We attribute this decrease in Pacific activity (and increase in North Atlantic activity) to a long-term trend towards a more La Niña-like basic state. La Niña typically increases North Atlantic activity and decreases Pacific activity due to alterations in large-scale conditions such as vertical wind shear, low-level vorticity and mid-level moisture.

This study was published in *Geophysical Research Letters* in 2022[1].

Correspondence to: Philip Klotzbach, Colorado State University, Fort Collins, CO; Email: philk@atmos.colostate.edu

Data Availability Statement

Tropical cyclone data from US warning agencies were taken from the International Best Track Archive for Climate Stewardship version 4[2]: https://www.ncei.noaa.gov/products/international-best-track-archive

References

1. Klotzbach, P. J., K. M. Wood, C. J. Schreck III, S. G. Bowen, C. M. Patricola, and M. M. Bell, 2022: Trends in global tropical cyclone activity: 1990-2021. *Geophys. Res. Lett.*, **49,** e2021GL095774, doi:10.1029/2021GL095774.
2. NOAA's International Best Track Archive for Climate Stewardship (IBTrACS) data, accessed on 3 January 2022.

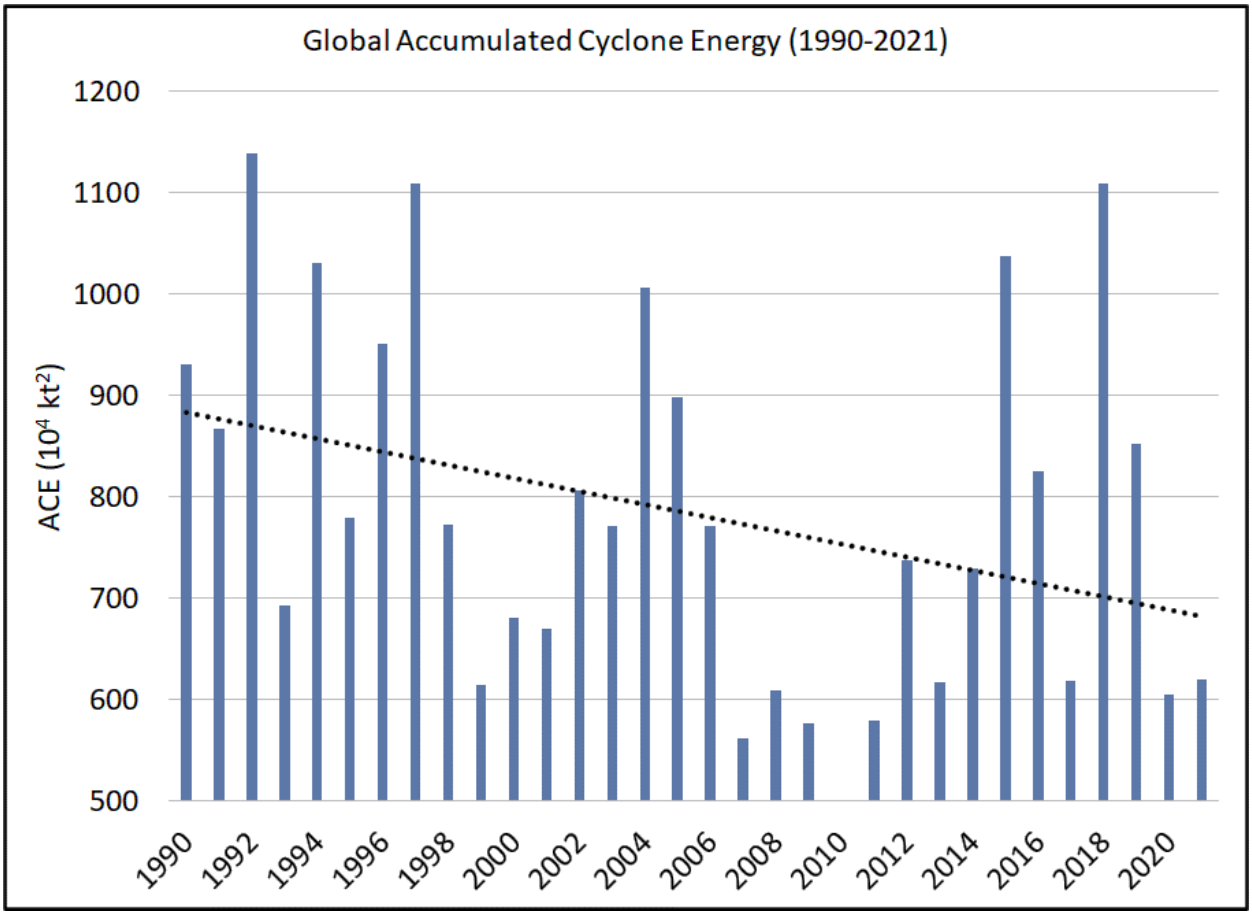

Figure 25: Global Accumulated Cyclone Energy (ACE) from 1990-2021. There has been a statistically significant (at the 10% level) decreasing trend in ACE during this 32-year period.

Correspondence to: Philip Klotzbach, Colorado State University, Fort Collins, CO; Email: philk@atmos.colostate.edu

Global Tropics Hazards Outlook Tropical Cyclone Verification

Lindsey N. Long[1,2], Nicholas Novella[1], and Jon Gottschalck[1]
[1]Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD
[2]ERT, Inc, Laurel, MD

ABSTRACT

1. Introduction

The Climate Prediction Center (CPC) currently releases a weekly product every Tuesday called the Global Tropics Hazards Outlook (GTH). This product highlights areas with increased changes of Tropical Cyclone (TC) formation and above- and below-average rainfall and temperatures across the global tropics. In September 2022, the product officially switched from the Week 1-2 forecast period to the Week 2-3 forecast period. Week 3 TC forecasts are experimentally produced and only shared internally within NOAA. This research delves into the skill of the TC genesis forecasts at Weeks 1 and 2 across all basins.

2. Data and Methods

Observed global TC genesis points are defined using the Best-Track Datasets from National Hurricane Center (NHC) (Landsea and Franklin 2013) and the Joint Typhoon Warning Center (JTWC) (Chu et al. 2002). The GTH is produced using ArcGIS software with forecast shapes for each variable archived each week as shape files. The dataset archive begins in 2010 and continues to the present; however, data for 2013 are missing from the archive due to data loss in the archiving computer. For the remainder of the paper, the 12-year period from 2010-2021 will be referenced with the understanding that 2013 is excluded, leading to an 11-year average for the period. Verification of the global TC forecasts are broken up into six ocean basins: the Atlantic (ATL), Eastern North Pacific (ENP), Western North Pacific (WNP), North Indian (NI), South Indian (SI), and South Pacific (SP).

Both spatial and temporal buffering are incorporated into the forecast verification. Spatially, observed genesis points are expanded out to a circle with a radius of 3 degrees. This allows inclusion of near misses, which have large impacts on skill scores, while maintaining the goal of forecasting relative risk at extended periods. For these extended range forecasts, correctly highlighting potential development in a general region is more important than pinpointing the exact location. Sensitivity tests with different buffer sizes were performed (not shown), and the 3-degree buffer encompassed the most near-misses without covering too much extra area. Temporal buffering accounts for storms that occur just before or after the valid forecast period. The buffer includes one day before and one day after the forecast week.

Skill scores are calculated using a 2x2 contingency table (Table 1) of hits, *a*, false alarms, *b*, misses, *c*, and correct nulls, *d*. The total number of cases, *n*, equals the sum of these four terms. The following skill scores are then calculated and discussed below: Hit Rate, $a/(a+c)$, False Alarm Ratio (FAR), $b/(a+b)$, Bias Ratio, $(a+b)/(a+c)$, Critical Success Index (CSI) or Threat Score, $a/(a+b+c)$, and Heidke Skill Score (HSS),

$$HSS = \frac{(a+d)-EC}{n-EC} \text{ where } EC = \frac{(a+c)(a+b)+(d+c)(d+b)}{n}$$

Correspondence to: Lindsey Long, NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland; lindsey.long@noaa.gov

Two methods are used for verification. The first method uses shapes from each week within the entire basin to count each term in the contingency table. Genesis points are converted into a point shape and expanded to a 3-degree radius circle. The Python package ArcPy can then be used to find overlaps between the forecasted shapes and the observed shapes. In the second method, the GIS forecast shapes are converted to a binary, 1-degree by 1-degree grid file, and the observed genesis points are expanded outward from their center point into a 3- degree by 3-degree grid box. This method allows for a better spatial view of the forecast skill within the basins.

2.1. Method 1: Basin-wide Verification

For the basin-wide verification method, HSSs show a large variability from year-to-year in each basin, especially in Week 2 (not shown). For example, Week 2 HSSs range from 0.47 to 0.04, with a mean of 0.21, in the Atlantic from 2010-2021. With this caveat, Figure 1a shows the 11-year averages for HSS and CSI by basin. The CSI is an important measure of skill when compared to the HSS because it removes the dependency on the "d" term and highlights how many observed and/or forecasted TCs were correctly predicted. Since TCs are considered an extreme event, the "d" term can inflate some skill metrics, but when looking at an entire basin, this is tempered. This is shown in Figure 1 by the comparison of HSS and CSI, with the two metrics having similar results. CSI is actually higher than HSS in most instances, especially in the WNP basin. The ENP basin shows the highest skill with the Week 2 ENP basin skill being close to the Week 1 WNP skill. The SP basin has the largest drop in skill from Week 1 to Week 2 (approximately 0.35) while the Atlantic has the smallest drop (approximately 0.15). Although not shown, the Hit Rate is also similar to the HSS and CSI values.

Figure 1b shows the 11-year average bias ratio among the basins. The bias ratio is a comparison of false alarms to missed forecasts. Values above (below) one indicate over- (under-) forecasting, consisting of too many false alarms (misses). The GTH verification indicates the over-forecasting of TCs in the Southern Hemisphere and the NI basins during Week 1 and under-forecasting in all basins during Week 2. This drop in bias is expected during Week 2 because forecasters have less confidence at the later lead. Again, as with CSI and HSS, forecasts in the ENP basin have the most reliability with a bias ratio close to one at both leads. FARs increase at Week 2 by about 0.1 points; however, a larger increase is seen in the SP basin (approximately 0.25) which is the cause for the large drop in the HSS and CSI at this lead.

To further investigate the ATL basin, the basin is divided up into three regions – the North Atlantic (NATL, 25 to 35 degrees North, 80 to 20 degrees West), the Main Development Region (MDR, 10 to 25 degrees North, 6 to 20 degrees West), and the Gulf of Mexico and Caribbean (Gulf, the remaining region from 10 to 35 degrees North to the ENP basin). HSSs calculated for these subregions show that the MDR is more skillful than the ATL basin as a whole with scores increasing 0.11 at Week 1 and 0.21 at Week 2. While the Gulf region has scores similar to the basin average, the NATL region has a large drop in skill to 0.13 in Week 1 and 0.05 in Week 2. This demonstrates that the GTH performs poorly in the NATL region where storms tend to have different drivers initiating genesis compared to the other two regions. Also, there has been a noticeable increase in misses with time (not shown) that is likely attributed to the recent increase in short-lived storms in the region which are harder to predict.

Correspondence to: Lindsey Long, NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland; lindsey.long@noaa.gov

2.2. Method 2: Spatial Verification

For the 1-degree by 1-degree spatial verification method, the HSS spatial patterns are consistent with the basin-wide verification (Fig. 2a-b), although more nuances can be seen. In the ATL basin, the NATL regional skill is low with minimal skill in Week 2, even with a number of cases occurring (not shown). Note that the high scores (reds) seen along the basin edges are due to a low number of cases which all verify but are not statistically significant. In the Pacific basins, the scores are very consistent throughout the region at both leads while, in the NI, SI, and SP basins, skill becomes spottier in Week 2.

In Figure 2c-d, the bias ratio shows the most active areas of each basin tend to have a ratio close to one, meaning little bias is present in the forecasts. There is a tendency to over-forecast TCs during Week 1 in certain regions, specifically in the Caribbean, the southern genesis region of the ENP around 95^{o}W, and throughout the SI (warm colors). In Week 2, more blues are seen, indicating a tendency to under-forecast. During both leads, the NATL region is consistently showing under-forecasting which is also consistent with the basin-wide verification method.

3. Conclusion

Two methods are utilized to provide automated verification of GTH TC forecasts with both methods accounting for near-misses by using both a spatial and temporal buffer. While the two methods are consistent with each other, the basin-wide verification provides a better overall summary while the spatial verification method provides more detail within the basins. Among the six ocean basins, the ENP basin consistently has the highest skill for all skill metrics, including bias ratio which is close to one at both leads. Breaking up the Atlantic into subregions shows that the majority of skill comes from the MDR. There is little skill in the NATL region where it is harder to anticipate formation based on tropical modes and variability. With the new GTH format, future research includes incorporating the new probabilistic format into the verification methods.

4. Data Availability Statement

GTH shape files are available upon request via the journal authors or the CPC. Observed best-track datasets are available on the web at: https://www.nhc.noaa.gov/data/#hurdat (Landsea and Franklin 2013) and https://www.metoc.navy.mil/jtwc/jtwc.html?best-tracks (Chu et al. 2002).

5. References

1. Chu, J.-H., C.R. Sampson, A.S. Levine, and E. Fukada, 2002: The Joint Typhoon Warning Center tropical cyclone best-tracks, 1945-2000, Technical report NRL/MR/7540-02-16. Naval Research Laboratory: Washington, DC.
2. Landsea, C.W. and J.L. Franklin, 2013: Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format. Mon. Wea. Rev., **141**(10), 3576-3592, https://doi.oeg/10.1175/MWR-D-12-00254.1.

Correspondence to: Lindsey Long, NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland; lindsey.long@noaa.gov

Table 3: A 2x2 Contingency Table, defining terms a, b, c, and d. The total number of cases, n, is defined as n=a + b + c + d.

Model/Observations	**Yes**	**No**
Yes	**"a" Hit**	**"c" Miss**
No	**"b" False Alarm**	**"d" Correct Null**

Correspondence to: Lindsey Long, NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland; lindsey.long@noaa.gov

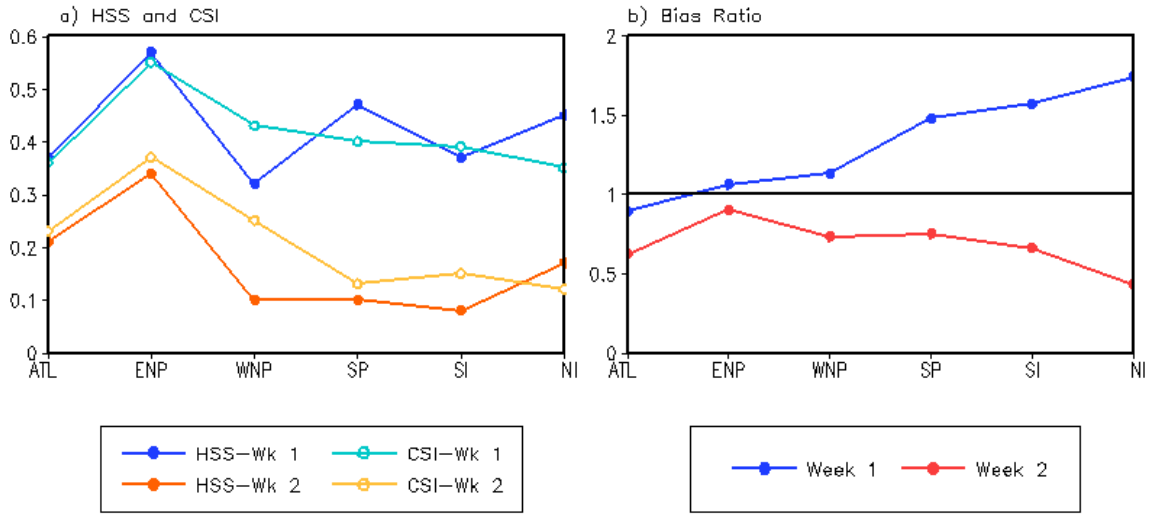

Figure 26: 11-Year averages of a) HSS, CSI and b) Bias by basin for Weeks 1 and 2. The black line indicates a perfectly unbiased forecast of 1. Data Sources: Landsea and Franklin, 2013; Chu et al., 2002, and CPC GIS shape files.

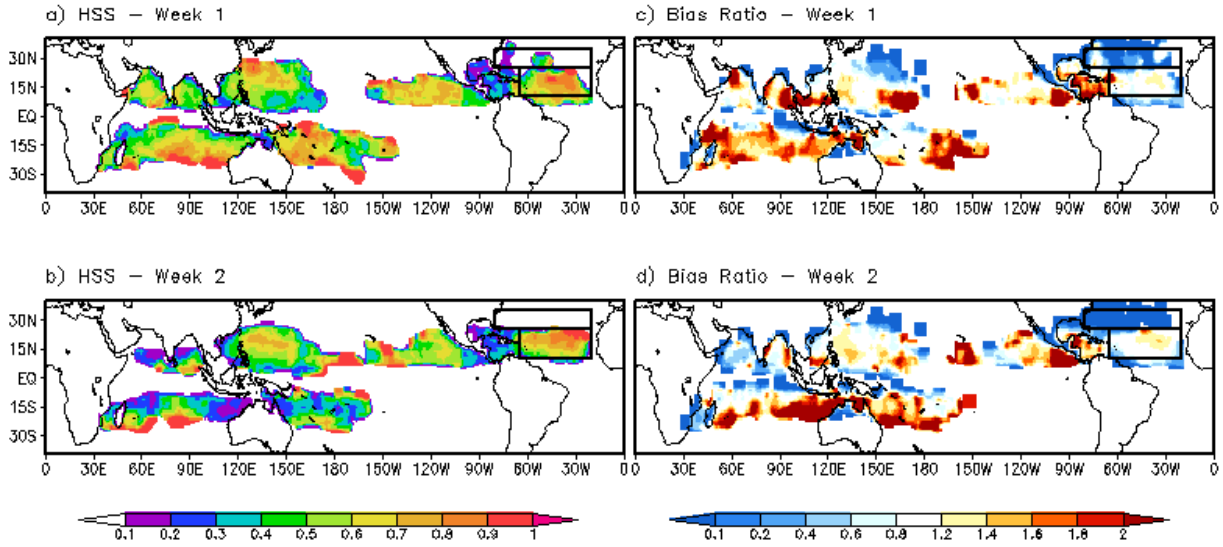

Figure 27: Spatial map of HSS for a) Week 1 and b) Week 2. c) and d) same as a) and b) but for Bias Ratio. Subregions for NATL and MDR are identified by black boxes. Data sources: Landsea and Franklin, 2013; Chu et al., 2002; CPC GIS shape files.

Correspondence to: Lindsey Long, NOAA/NWS/NCEP/Climate Prediction Center; College Park, Maryland; lindsey.long@noaa.gov

Predictability of Summer Extreme Maximum Temperatures over Taiwan by using NOAA NCEP GEFSv12 Reforecast Products

M. M. Nageswararao[1], Yuejian Zhu[2], Vijay Tallapragada[2] and Meng-Shih Chen[3]
[1]CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, MD-20740, USA.
[2]NOAA/NWS/NCEP/EMC, College Park, MD-20740, USA.
[3]Central Weather Bureau, Taipei, Taiwan

ABSTRACT

1. Introduction

Temperature, a key element of the weather, is a measure of the warmth or chill of the environment and has a major effect on activities in industries such as energy, aviation, communication pollution dispersal, and agriculture. This century has dramatically increased global temperatures, making climate change one of the most pressing scientific and social challenges. According to the Intergovernmental Panel on Climate Change (IPCC, 2013), the average global temperature has increased by 0.13 degrees Celsius per decade over the last 50 years, almost double the rate of the previous century. This has resulted in more frequent and intense heat waves, which have been observed across the globe in recent decades. These heat waves can have devastating consequences, particularly for vulnerable populations such as the elderly, children, and those with chronic illnesses, leading to increased deaths and emergency hospital admissions. Taiwan, as an East Asian subtropical island, is particularly susceptible to the extreme weather and climatic changes brought on by global warming (Lin et al., 2015). The IPCC assessment (IPCC 2007) found that the air temperature in Taiwan has risen by 1.4 degrees Celsius from 1911 to 2005, which is near twice the increase (0.7 degrees Celsius) observed in the Northern Hemisphere. A study conducted in Taipei from 1994 to 2003 found that each 1-degree Celsius increase in surface air temperature above 31.5 degrees Celsius increases respiratory mortality by approximately 9.3% (range, 4.1 to 14.8%), while each 1-degree Celsius increase above 25.2 degrees Celsius increases cardiovascular mortality by around 1.1% (range, 0.3 to 1.9%) (Chung et al., 2009). To mitigate the impacts of extreme weather and climate, it is essential to develop extended-range or sub-seasonal weather forecasts that can predict weather and climate extremes such as heatwaves, cold waves, droughts, and floods on extended range time scale. Such forecasts can provide relevant information on the timing of the onset of a rainy season, the risk of extreme rainfall events, heat waves, etc. However, there is still a well-known gap in current numerical prediction systems for short, medium-range and extended range time scale forecasts. This gap exists between medium-range weather forecasts (up to 10 days) and long-term climate predictions (longer than one month). The initial conditions of the atmosphere play a role in determining the accuracy of medium-range weather forecasts. In contrast, seasonal climate predictions are more sensitive to slowly changing surface boundary conditions, such as sea surface temperature and soil moisture content (Mariotti et al., 2018). Despite this, progress has been made in some regions and seasons in predicting sub-seasonal timescale events (Li et., 2015; DelSole et al., 2017). In particular, the impressive advancements in numerical modeling in recent decades have led to a remarkable improvement in the skill of short and medium-range weather forecasts in many parts of the globe, particularly in extra-tropical regions. However, the prediction skill is still not up to mark over tropical and monsoon regions (Vitartet al. 2017; Nageswararao et al. 2022a,2022b). This is mainly due to the complexity of the numerical modeling of tropical processes, which are influenced by interactions among ocean-land-atmosphere feedback, atmospheric circulation, organized convection, radiation, precipitation, moisture, aerosols, and clouds on different space and time scales. Therefore, there is still a need to work on realistic simulations of extended range time over small regions like Taiwan Island. Several studies (Wang et al. 1999; Xu 1999; Sperber et al. 2001; Kang et al. 2002) have suggested that the relatively coarse resolution of global models and the lack of representation

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

of land-sea contrast and topography may be the main reasons for the low prediction skill over small regions. To improve the extended-range forecasts over small regions, it is essential to incorporate suitable post-processing techniques on GCM products.

In September 2020, the National Centers for Environmental Prediction (NCEP) implemented the Global Ensemble Forecast System version 12 (GEFSv12) model to provide stakeholders with sub-seasonal forecasts, hydrological, and other meteorological applications (Zhou et al., 2019, 2022; Guan et al., 2022; Hamill et al., 2022; Nageswararao et al. 2022a, 2022b). The consistent reforecast products of this model for the period 2000 through2019 are available on Amazon Web Services (AWS, https://registry.opendata.aws/noaa-gefs/) for the benefit of the broader community. In this research, a Hybrid Post-processing (ANN-QQ) approach, which combines Artificial Neural Network (ANN) and Quantile mapping (QQ) techniques, is used to predict summer (June through September; JJAS) surface air maximum temperature (T_{max}) and its related extremes on an extended-range time scale over Taiwan. This approach is applied to the NOAA NCEP GEFSv12 reforecast raw products. The performance of the hybrid technique is evaluated against ECMWF Reanalysis v5 (ERA5) and compared to the Raw, Mean Bias removal (MB), Multi-linear regression (MLR), QQ, and ANN techniques using standard skill metrics.

2. Data and Methodology

2.1. Data Used

This study analyzed the NCEP GEFSv12 T_{max} products over Taiwan for the reforecast period 2000 through 2019. The data were based on 00-hour UTC initial conditions and had a forecast lead time of up to 16 days with 5 members during the JJAS season. The data were obtained from Amazon Web Services (AWS) in grib2 format at 3 (6) hour intervals at 0.25-degree (0.5 degree) resolution for the first 10 days (beyond 10 days) of the forecast. For uniformity, day-1 to 10 forecasts were also considered on the same grid points of day-11 to 16 forecasts. The reforecast products were generated using the current operational Global Forecast System version 15.1 (GFSv15.1) with the Geophysical Fluid Dynamics Laboratory (GFDL) FV3 Cubed-Sphere dynamical core. The model had a horizontal resolution of approximately 25 kilometers (C384 grid) with 64 vertical hybrid levels, with the top layer centered around 0.27 hectopascals (approximately 55 kilometers). The scale-aware parameterization convection scheme was used instead of the Simplified Arakawa-Schubert (SAS) shallow and deep convection scheme. This scheme was further modified to reduce excessive cloud-top cooling for model stabilization. The hybrid Eddy-Diffusivity Mass-flux (EDMF) scheme was used for the vertical mixing process of the planetary boundary. The cloud microphysics scheme used from GFDL includes five predicted cloud species (cloud water, cloud ice, rain, snow, and graupel), and the short and longwave radiative fluxes were estimated from the rapid radiative transfer model (RRTM) developed at Atmospheric and Environmental Research (Clough et al. 2005). The convective gravity wave drag estimation was based on the schemes developed by Chun and Baik (1998), while the GFS orographic gravity wave drag and mountain blocking schemes followed those developed by Alpert (1988). A two-tiered Sea Surface Temperature (SST) and Near Sea Surface Temperature (NSST) approach (Zhu et al. 2017, 2018) was employed to estimate the SST boundary condition, which accounted for the day-to-day variability and diurnal variation of SST, respectively. The stochastic kinetic energy backscatter (SKEB) and stochastically perturbed parameterization tendencies (SPPTs) were also used to improve the model's uncertainty. A description of the GEFSv12 forecast system at NOAA NCEP is provided by Zhou et al. (2019; 2022). The ERA5 maximum 2-meter air temperature (T_{max}) over Taiwan for the period 2000 through 2019 has been used as a reference to evaluate the performance of GEFSv12 for summer T_{max} and associated T_{max} extreme events

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

(daily T_{max}, greater than the 90th percentile) over Taiwan with day-1 to 16 forecast lead times for the reforecast period (2000 through 2019) (Hersbach et al., 2018).

2.2. Calibration Methods

Statistical post-processing is necessary to improve the accuracy of the numerical weather models' raw products at extended range scales and to make them more useful for forecasting. Various post-processing models with varying levels of complexity and statistical basis have been developed to calibrate the raw GCM forecasts (Wilks, 2018). These include the rank histogram recalibration method (Hamill & Colucci, 1998), "poor man's ensemble" (Ebert, 2001), analog method (Hamill & Whitaker, 2006), frequency match method (FMM; Zhu & Luo, 2015) and the quantile mapping method (Nageswararao et al., 2022a, 2022b). While these approaches are computationally efficient and easy to implement, their reliability and skill in calibrated forecasts may not ensure. Research has demonstrated that the use of linear or non-linear regression-based models with a few parameters can significantly improve raw forecasts by linking predictors and predictands (Verkade et al., 2013). Recent studies have compared the performance of various post-processing models, such as machine-learning models, for forecast post-processing (Mert et al., 2020). However, Wilks (2018) found that the performance of these models varied depending on the study area, GCMs, and evaluation metrics used. As such, there is no consensus on which post-processing model best suits a given application. Therefore, there is a need for an effective post-processing approach that can provide calibrated forecasts that are unbiased, reliable in ensemble spread, as skillful as possible, and no worse than the climatology reference forecasts. In this study, a hybrid post-processing approach that combines ANN and QQ techniques is applied to calibrate the NOAA NCEP GEFSv12 reforecast raw products for summer (June through September; JJAS) T_{max} and its related extremes on an extended-range time scale over Taiwan. The performance of the hybrid technique is evaluated against ERA5 reanalysis, and the results are compared with Raw, MB, MLR, QQ, and ANN techniques using standard skill metrics. A brief description of these methods is given in the following sub-sections.

2.2.1. Mean Bias (MB) -Remove Technique (U)

MB is defined as the difference between running monthly climatological mean from ERA5 and GEFSv12:

$$b_t = \bar{Y} - \bar{\bar{F}}_t$$

For each day, this difference (b_t) is calculated in the leave one out cross-validation manner and adds this mean bias in the 'test' (t) day's model forecast.

$$U_t = F_t + b_t$$

2.2.2. Multiple linear Regression (MLR)

In this MLR method, Singular Value Decomposition (SVD) has been employed for the computation of the regression coefficients ($b_0, b_1, b_2, \ldots\ldots, b_p$). The advantage of this method is that it removes the singular matrix problem while calculating covariance among ensemble members, which cannot be entirely solved with the Gauss-Jordan elimination method.

$$E(y) = b_0 + b_1 x_1 + b_2 x_2 + \ \ldots\ldots\ldots\ldots + b_p x_p$$

2.2.3. Quintile Mapping Method (QQ)

This technique uses the empirical probability distributions of ERA5 and GEFSv12 T_{max} values to produce a calibrated output. This output is the inverse of the cumulative distribution function (CDF) of

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

ERA5 values at the probability corresponding to the GEFSv12 model output CDF at a particular value. The bias-corrected value for the T_{max} forecast of the GEFSv12 model is as follows:

$$Q = {CDF_{ERA5}}^{-1}(CDF_{GEFSv12}(F_t))$$

Here, CDF^{-1} is an inverse of CDF. Thus, the technique of QQ is a transformation between two CDFs of the ERA5 and GEFSv12 model.

2.2.4. Artificial Neural Network (ANN)

Predicting temperature under climate change is becoming increasingly challenging for the meteorological scientific community due to its complexity, non-linearity, and significant fluctuations. A non-linear technique can help bring out the irregularities in the current system. Saxena et al. (2013) suggested that an ANN-based post-processing technique can be used more accurately to predict temperature than the conventional technique, as the ANN is better at handling complex non-linearity. The ANN can model non-linear atmospheric variables without making assumptions, which is more advantageous than statistical approaches for temperature predictions. However, Bani-Ahmad et al. (2014) cautioned that making assumptions in temperature prediction can lead to more confusion than the predicted effects. The ANN is advantageous due to its rapid information processing based on the principle of parallel, massively machine-based processing that simplified models of biological neurons inspire. ANN is a powerful information processing system that is composed of highly interconnected neurons, allowing it to solve complex problems with greater accuracy than traditional statistical techniques (Saxena et al., 2013). Furthermore, ANNs require fewer data and shorter training times than other statistical methods (Feng & Lu, 2019). The ability to adapt and learn from existing data is one of the most important characteristics of ANN, which sets it apart from other methods. This feature enables ANNs to develop a relationship between inputs and outputs, allowing them to make predictions based on given inputs (Abbot & Marohasy, 2014). The ANN is a data-driven approach that can identify non-linear relationships between input and output parameters without the need to solve complex partial differential equations (Yilmaz et al., 2011). This makes it an ideal tool for temperature predictions, which has been demonstrated in several studies involving atmospheric time series data, including temperature predictions (Ahmad et al., 2014). Therefore, this study implements the ANN calibration method on GEFSv12 reforecast products for summer T_{max} and associated T_{max} extremes over Taiwan with forecast lead times ranging from day-1 to 16.

Back Propagation Neural Network (Pal et al., 2002; Singh & Panda, 2015) is a popular ANN that uses a gradient descent algorithm to reduce the difference between the output values and the desired targets. This error calculation method works by propagating the errors from the output layer to the hidden layers. The initial weights are set close to zero, and the difference between the actual and desired outputs is then calculated. After the error is determined, the weights are adjusted with a bias correction applied. The network is then supplied with new input values, resulting in different output values and errors. The weights and biases are revised by adjusting the parameters until the desired accuracy is achieved, with the error minimized. Hecht-Nielsen (1990) conducted research suggesting that a single hidden layer of neurons is sufficient to accurately model any practical solution surface. A double-cross-validation procedure was used to determine the optimal number of hidden neurons for the single hidden layer architecture to ensure the neural network does not overfit during the training process. The performance of an ANN is highly dependent on the data used for training, and the characteristics of the datasets determine the ANN's structure. To evaluate and test the accuracy of the constructed ANN, it is important to determine the appropriate values for parameters such as initial weights, learning rate, momentum, epoch, and activation functions. In order to ensure successful data analysis, pre-processing is necessary to identify and eliminate outliers, reduce noise, and normalize the range of inputs and outputs. To achieve

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

this, the Neural Network Toolbox in MATLAB (http://www.mathworks.in/help/nnet/index.html) was used in this study to train, visualize, and simulate artificial neural networks for the comparison of daily maximum temperature from the 5 ensemble members of GEFSv12 to the same. To determine the optimal ANN structure for a particular day forecast, a double cross-validation procedure was employed. This procedure involved leaving out one-time step from the 20-time steps and fitting the ANN model to the remaining time steps in a cross-validation mode. The performance of each iteration was then monitored using metrics such as Mean Square Error (MSE) and Root Mean Squared Error (RMSE) while increasing the number of hidden neurons from 1 to 10. The results of this study showed that when the number of hidden neurons was increased, the MSE and RMSE decreased for both training and testing data. However, after a certain point, the MSE and RMSE decreased for training but increased for testing, indicating that after a certain point, the error in testing data will continue to increase without much change in the training data. The optimal number of neurons for the ANN model was determined to be 3, which provided the best balance between training and testing errors. A simple randomization technique was employed to prevent the algorithm from quickly converging to local minima. This technique involved randomly selecting the weights and biases of the neurons in the ANN model before each iteration of the training process. This randomization technique helped to introduce oscillations in the ANN model, which prevented it from quickly converging to local minima. The Min–Max transformation is used in ANN development to speed up the training process, avoid saturation, and reduce the chances of getting stuck in local optima by transforming the input and output values (Johnstone and Sulungu,2021). This study implemented a feed-forward backpropagation neural network with 3 hidden neurons and a Hyperbolic tangent sigmoid transfer function using the MATLAB ANN toolbox and the Levenberg Marquardt training algorithm for deep learning of summer T_{max} over Taiwan. The hyperbolic tangent sigmoid function was chosen due to its non-linear, differentiable, and monotonic properties, which have been found to yield better training performance for multilayer neural networks. In this study, a basic ANN is constructed using the components outlined in Table 1.

Table 4: The following are considered to develop a simple ANN model to improve the GEFSv12 prediction skill in depicting Summer daily Tmax and Associate Tmax extremes over Taiwan.

Skill Criteria Item	ANN Specification
No. Hidden layers:	1
No. of nodes/neurons in the hidden layer	3
Neural Network used	Feed forward network
Neural Network Processing Functions:	Map matrix row minimum and maximum values to [-1 1]
Data divided function	70% data for training and 30% data for validation
Learning rate	: 0.001
Max number of iterations/epochs used	: 1000
Error tolerance for stopping criterion	: 1e-14
Training function used	: Supervised weight/bias training function with Sequential order weight/bias training (trains)
Neural Network Performance Functions used	: Mean squared error performance function

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

2.2.5. Hybrid Post-Processing

The Hybrid Statistical Post-processing method, which combines ANN with QQ techniques, was applied to the GEFSv12 output in order to enhance summer daily T_{max} and related T_{max} extremes in Taiwan. The entire process of all the calibration methods has incorporated the leave-one-out cross-validation procedure. The predictive accuracy of five calibration techniques (MB, MLR, QQ, ANN, and Hybrid) for forecasting summer daily T_{max} and related T_{max} extremes over Taiwan for Day-1 to 16 forecast lead times has been evaluated in comparison to Raw using ERA5 as a reference. Standard skill metrics have been used to assess the precision of the various calibration methods.

3. Results and Discussion

The spatial patterns of the climatological mean of summer (JJAS) daily T_{max} over Taiwan from 2000 through 2019 have been analyzed using ERA5, Raw, MB, MLR, QQ, ANN, and Hybrid calibration methods with Days-1, 5, 10, and 15 forecast lead times to evaluate the performance of the raw and five calibration methods (Fig. 1). The results show that the spatial patterns of summer daily T_{max} over Taiwan captured by Raw-GEFSv12 are similar to those from ERA5 for all forecast lead times. However, the magnitude of the forecasts is limited, and a warm bias is observed in most parts of the country for all forecast lead times. The highest summer T_{max} is observed in the southernmost region of Taiwan, which is also reflected in the GEFSv12 forecasts for all lead times. The temperature maximum is lower in the eastern parts of the country and increases towards the west. All the calibration methods were successful in reducing the warm bias in most parts of Taiwan, resulting in a climatological mean of T_{max} over Taiwan that was comparable to ERA5 for all forecast lead times (Fig.1). The spatial patterns of interannual variability (IAV) of summer T_{max} over Taiwan from GEFSv12 and ERA5 are similar for all forecast lead times (figure not provided). However, GEFSv12 tends to overestimate the IAV of summer T_{max} in most parts of the country for all forecast lead times, particularly in the northeastern part. Most of the calibration methods used in this study effectively reduced the overestimation of the IAV of T_{max} over Taiwan. The MB calibration method, however, still has a large overestimation of IAV and is more or less similar to the raw product. The QQ method has the advantage of being able to adjust the T_{max} probability distribution to the observed data, particularly in the extreme tails, thereby considering the IAV. However, while it improves the spatial patterns, the temporal patterns remain unchanged. The ANN method is better at capturing temporal patterns, but it slightly underestimates the IAV of T_{max}. The Hybrid method is superior to all four other calibration methods as it can effectively capture the temporal as well as spatial patterns of T_{max} over Taiwan, including the IAV of T_{max}.

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

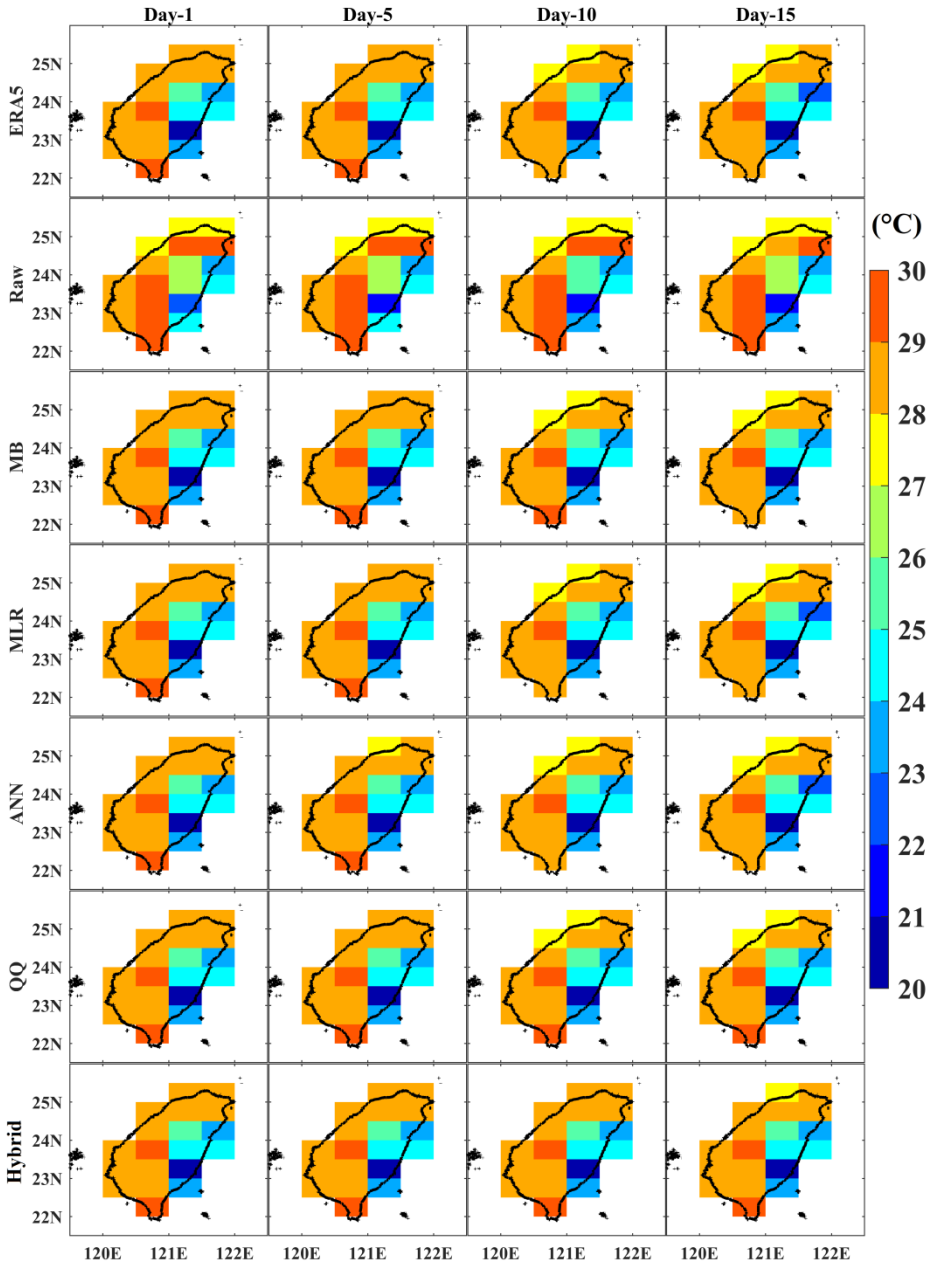

*Figure1: Climatological mean of summer (JJAS) surface air maximum Temperature (T_{max}) over Taiwan from ERA5, Raw, MB, MLR, QQ, ANN, and Hybrid methods with Day-1, 5, 10, and 15 forecast lead times for the period 2000-2019. [Data Source: ERA5 (*Hersbach et al. 2018), GEFSv12 (Guan et al. 2022)]

The performance of the Raw, MB, MLR, QQ, ANN, and Hybrid methods in predicting T_{max} over Taiwan from Day-1 to 16 forecast lead time for the reforecast period was evaluated using the Average RMSE, Correlation Coefficient (CC), and Index of Agreement (IOA). The results of this evaluation are presented in Figure 2 to compare the errors and prediction skills of the raw and all five calibration methods with increasing lead time. The RMSE of the Raw forecast for summer T_{max} over Taiwan was remarkably high for all forecast lead times (ranging from 1.5 to 1.6 degrees Celsius). However, the application of calibration methods remarkably reduced the RMSE for all forecast lead times. The comparison between

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

methods reveals that MB had the highest RMSE (1.0 to 1.3 degrees Celsius), followed by QQ (0.7 to 1.2 degrees Celsius), MLR (0.7 to 1 degrees Celsius), ANN (0.6 to 0.8 degrees Celsius), and Hybrid (0.6 to 0.8 degrees Celsius). The ANN and Hybrid methods results show similar RMSE values for all forecast lead times (Fig. 2a).

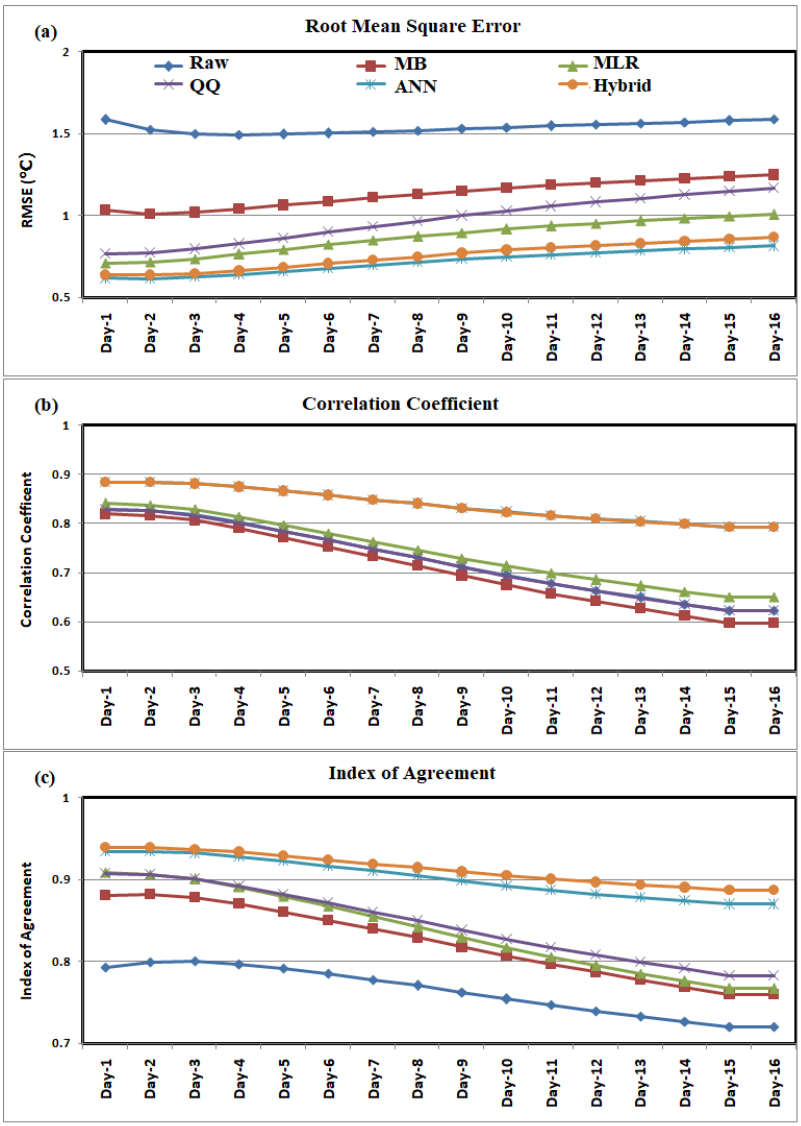

*Figure 2: Statistical skill scores (a) Root Mean Square Error, (b) Correlation Coefficient and (c) Index of Agreement of Raw, MB, MLR, QQ, ANN, and Hybrid methods against ERA5 in depicting Summer (JJAS) T_{max} over Taiwan with forecast lead time Day-1 to 16 for the reforecast period 2000-2019. [Data Source: ERA5 (*Hersbach et al. 2018), GEFSv12 (Guan et al. 2022)]

The CC of GEFSv12 in predicting summer daily T_{max} over Taiwan is notably high for the Day-1 lead time forecast, with a value of more than 0.8, and decreases with increasing forecast lead time, ranging from 0.8 to 0.6 (Fig. 2b). The QQ method did not show any improvement in the CC for predicting summer daily T_{max} over Taiwan compared to the Raw products for all forecast lead times. The MLR slightly improved the CC for all forecast lead times, while the MB slightly decreased the CC for all forecast lead times.

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

However, ANN and Hybrid calibration methods demonstrated a marked improvement in the CC (greater than 0.8) for predicting summer daily T_{max} over Taiwan for all forecast lead times. The CC values obtained from the Hybrid method were identical to those obtained from the ANN for all forecast lead times. The IOA of GEFSv12 in predicting T_{max} over Taiwan is notably higher for shorter lead time forecasts, decreasing from 0.8 to 0.7 as the forecast lead time increases (Fig. 2c). All the calibration methods showed a remarkable improvement in IOA (0.78 to 0.92) for summer daily T_{max} over Taiwan for all forecast lead times. The Hybrid method was found to be the most effective, with IOA values ranging from 0.89 to 0.95, which were higher than those obtained from the ANN (0.88 to 0.95) and other methods such as QQ (0.78 to 0.91), MLR (0.78 to 0.91) and MB (0.74 to 0.88). The Hybrid method of predicting summer daily T_{max} over Taiwan has been found to be more effective than other calibration methods for all forecast lead times. This improvement, particularly for longer lead time forecasts, can be highly beneficial for climate management in various sectors across Taiwan.

For the period of 2000 through 2019, various statistical categorical skill scores, such as Probability of Detection (POD), Frequency Bias, False Alarm Rate (FAR), Accuracy (ACC), Success Rate (SR), Threat Score (TS), and Equitable Threat Score (ETS), were computed using the contingency table for summer daily T_{max} extreme days (T_{max} greater than 90th percentile of annual T_{max}) over Taiwan for day-1 to 16. An analysis of summer T_{max} extreme days using the ETS from GEFSv12 reveals that coastal areas of Taiwan have higher ETS values than interior regions for all forecast lead times (Fig. 3). The ETS of GEFSv12 for summer daily T_{max} extremes over Taiwan decreases with increasing forecast lead time, with the highest values observed for shorter lead times. All calibration methods tested show an improvement in the ETS score for summer daily T_{max} extremes over Taiwan for all forecast lead times, except for MLR for longer lead time forecasts. All five calibration methods used to predict summer daily T_{max} extremes over Taiwan showed a decrease in ETS scores with increasing forecast lead times. The comparison between the five methods revealed that ANN and Hybrid methods had relatively higher ETS scores than MB, MLR, and QQ for all forecast lead times. The Hybrid method of forecasting summer daily T_{max} extremes in Taiwan produces a higher ETS score than the ETS score from ANN calibration methods for all forecast lead times (Fig. 3). This indicates that the Hybrid method is more accurate in predicting summer daily T_{max} extremes in Taiwan than the ANN calibration method. The ETS analysis was further extended to the Week-1, Week-2, and Week-1 to 2 scales. The ETS score for predicting summer daily T_{max} extremes over Taiwan from GEFSv12 was found to be significantly higher for Week-1 than Week-2. Moreover, the ETS score for the 2-week scale (Week-1 to 2) was even higher than that of Week-1 and Week-2. All five calibration methods (CDF, PDF, MLR, SVR, and ANN) improved the ETS score for summer daily T_{max} extremes for Week-1, Week-2, and Week-1 to 2, except MLR for Week-2. Comparing the ETS scores for summer daily T_{max} extremes in Taiwan for Week-1, 2, and 1 to 2 reveals that the ANN and Hybrid methods yielded higher scores than MB, MLR, and QQ methods. Notably, the Hybrid method produced the highest ETS scores in most parts of Taiwan for summer daily T_{max} extremes for Week-1, 2, and 1 to 2.

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

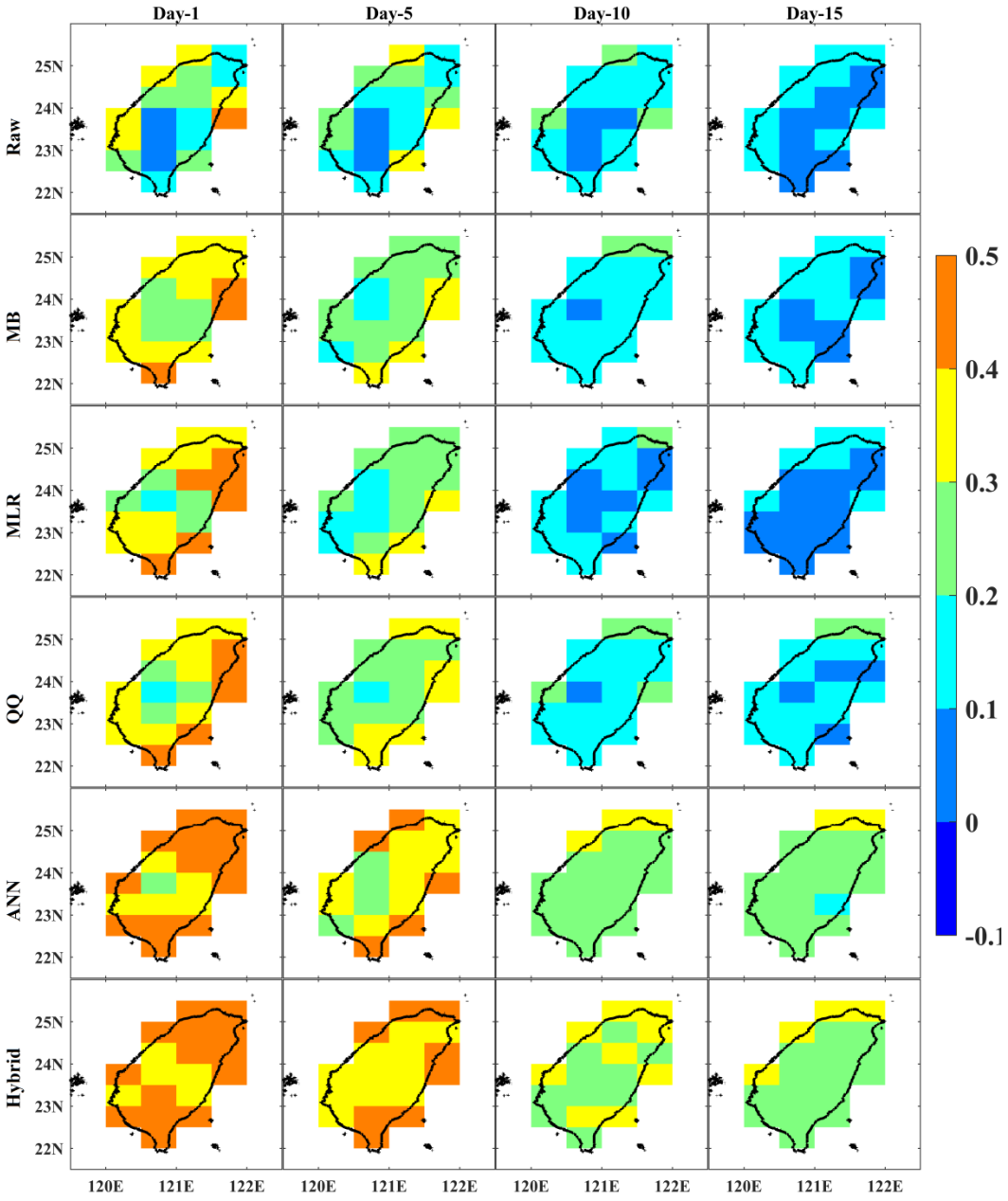

*Figure 3: Equitable Threat Score (ETS) of Raw, MB, MLR, QQ, ANN, and Hybrid calibration methods with Day-1, 5, 10, and 15 forecast lead times against ERA5 in depicting Heatwave days over Taiwan for the period 2000-2019. [Data Source: ERA5 (*Hersbach et al. 2018), GEFSv12 (Guan et al. 2022)]

A performance diagram is a helpful visual representation for displaying multiple skill scores, such as POD, Frequency Bias, TS, and SR (1-FAR), in a single graph (Roebber 2009; Huang and Luo 2017). The solid contour lines in the performance diagram as shown in Figure 4 indicate the TS, while the dashed lines illustrate the frequency Bias with additional labels on the X and Y axes' upper (2nd) sections.

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

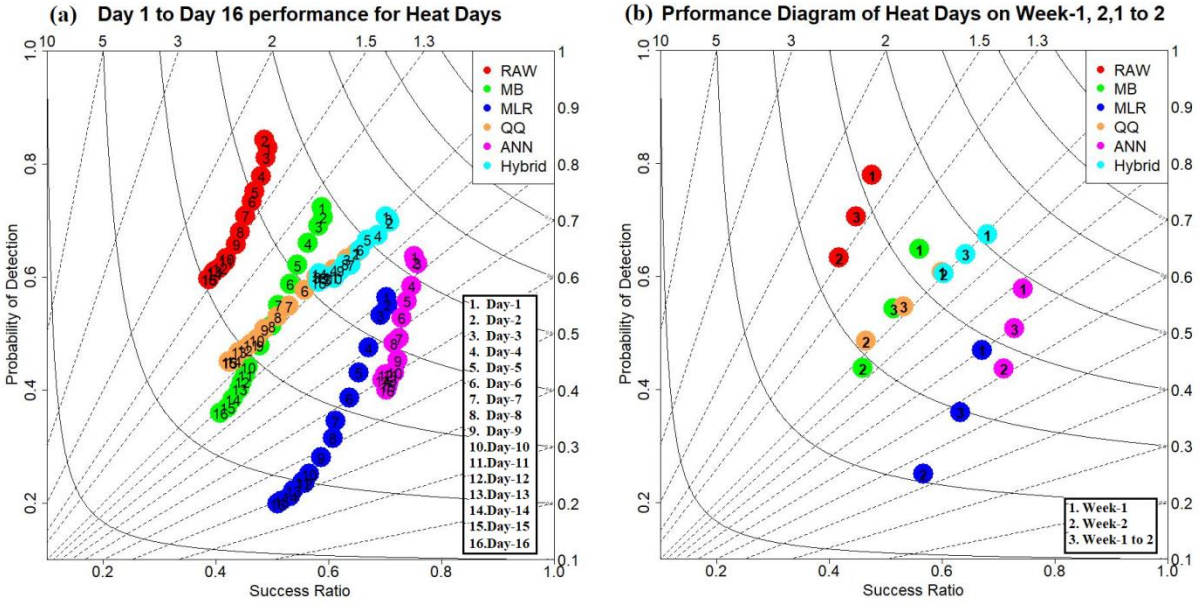

*Figure 4: Performance diagram summarizing the SR, POD, Frequency Bias, and TS statistical categorical skill scores of Raw, MB, MLR, QQ, ANN, and Hybrid calibration methods against ERA5 for (a) Extreme Heat days (Tmax > 90th percentile of annual Tmax) with Day-1 to 16 forecast lead times, and (b) Extreme Heat days on Week-1, Week-2 and Week-1 to 2-time scale over Taiwan for the period 2000-2019. The solid and dashed lines represent TS and Frequency Bias scores, respectively. [Data Source: ERA5 (*Hersbach et al. 2018), GEFSv12 (Guan et al. 2022)]

The GEFSv12 model shows an overestimation of summer daily T_{max} extreme days over Taiwan for all forecast lead times, with a Frequency Bias of more than 1.5. Despite this, the model's Probability of Detection (POD) for this is remarkably high, ranging from 0.6 to 0.85 (Fig. 4a). Similarly, the model's Skill Scores (SR and TS) for predicting the daily T_{max} extremes in Taiwan during the summer season are highest for shorter lead times and decrease with increasing forecast lead time (Fig. 4a). The results of the five calibration methods show a significant reduction in the overestimation of daily T_{max} extremes over Taiwan for all forecast lead times (Fig. 4a). However, the Probability of Detection (POD) for summer daily T_{max} extremes over Taiwan has decreased from all five calibration methods for all forecast lead times. The POD from MLR and ANN is notably lower than that of the other three methods, MB, QQ, and Hybrid, for all forecast lead times. The results from ANN show higher Skill Ratio (SR) and Threat Score (TS) values than those obtained from MLR, MB, and QQ methods for all forecast lead times. Both the QQ and Hybrid calibration methods were able to accurately reproduce the same number of summer T_{max} extreme days as observed in ERA5 for all forecast lead times. However, the Hybrid method yielded notably higher POD, SR, and TS skill scores than the other four calibration methods for all forecast lead times. This suggests that the Hybrid method is more effective in predicting summer T_{max} extreme days over Taiwan, particularly for longer lead-time forecasts which could be beneficial for extended-range timescale predictions (Fig. 4a). The comparison of GEFSv12 with the five calibration methods for Week-1, 2, and 1 to 2 revealed a significant overestimation of summer T_{max} extreme days (Fig. 4b). All five calibration methods were successful in mitigating this overestimation. However, the Hybrid method demonstrated the highest statistical categorical skill scores for all three weeks, indicating its superior performance in predicting summer T_{max} extreme days. This was especially evident for Week-1 and Week-1 to 2, where the majority of the categorical skill scores from Raw, MB, MLR, QQ, ANN, and Hybrid

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

calibration methods were higher than for Week-2. Therefore, it can be concluded that the raw data from GEFSv12 for summer T_{max} extreme days are not reliable without calibration. The Hybrid calibration method was found to be the most effective in improving the statistical categorical skill scores for day-to-day, Week-1, Week-2, and Week-1 to 2 scale forecasts of summer T_{max} extremes over Taiwan. This makes it a valuable tool for climate risk management sectors in the region.

4. Conclusion

- The GEFSv12 is able to capture the T_{max} over Taiwan for all lead time forecasts, however, it has a warm bias and an overestimation of the IAV of T_{max} in the south and interior parts of Taiwan.
- The prediction skill (CC and IOA) of GEFSv12 decreases with increasing lead time.
- The Hybrid and ANN were found to be highly effective in reducing the warm bias over Taiwan, similar to other traditional calibration methods. Furthermore, the ANN was observed to have a relatively higher prediction skill (CC and IOA) improvement compared to other methods. The hybrid method is relatively better than ANN for all forecast lead times.
- The GEFSv12 showed a large overestimation of heatwave days over Taiwan, but all calibration methods successfully reduced this overestimation.
- The statistical categorical skill scores (ETS, TS, SR, and Frequency Bias) for heat wave days are generally higher when using a Hybrid method compared to other calibration methods.
- The Hybrid post-processing technique for calibrating the GEFSv12 T_{max} forecasts is a beneficial tool for providing extended-range forecast guidance of T_{max} and associated extremes over Taiwan, which can be advantageous for managing climate risk in various sectors.

5. Acknowledgments

This research was made possible by generous funding from the NCEP Visiting Scientist Program, managed by the University Corporation for Atmospheric Research (UCAR) Cooperative Programs for the Advancement of Earth System Science (CPAESS). The authors are thankful to the Ensemble Team at the NCEP Environmental Modeling Center (EMC) for providing access to the datasets used in this study.

6. Data Availability Statement

The maximum 2 meter air temperature data for the period 2000- through 2019 over Taiwan island has been obtained from ERA5 (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form) while the GEFSv12 reforecast products for the same period have been retrieved from Amazon Web Services (AWS; https://noaa-gefs-retrospective.s3.amazonaws.com/index.html).

7. References

1. Abbot, J., and J. Marohasy, 2014: Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. *Atmos. Res.*, **138** (1),166-178, https://doi.org/10.1016/j.atmosres.2013.11.002.
2. Ahmad, R., N. M. Lazin, and S. F. M. Samsuri, 2014: Neural Network Modeling and Identification of Naturally Ventilated Tropical Greenhouse Climates. *Wseas. Trans. Syst. Control.*, **9**, 445-453.
3. Al-Matarneh, L., A. Sheta, S. Bani-Ahmad, J. Alshaer, and I. Al-oqily, 2014: Development of Temperature-Based Weather Forecasting Models Using Neural Networks and Fuzzy Logic. *Int J Multimed Ubiquitous Eng.* ,**9** (12), 343–366, http://dx.doi.org/10.14257/ijmue.2014.9.12.31.
4. Alpert, J. C., M. Kanamitsu, P. M. Caplan, J. G. Sela, G. H. White, and E. Kalnay, 1988: Mountain induced gravity wave drag parameterization in the NMC medium-range forecast model. *Eighth Conf. on Numerical Weather Prediction*, Baltimore, MD, Amer. Meteor. Soc., 726–733.

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

5. Chun, H.-Y., and J.-J. Baik, 1998: Momentum Flux by Thermally Induced Internal Gravity Waves and Its Approximation for Large-Scale Models. *J. Atmos. Sci.*, **55** (21), 3299–3310, https://doi.org/10.1175/1520-0469(1998)055<3299:MFBTII>2.0.CO;2.
6. Chung, J. Y., Y. Honda, Y. C. Hong, X. C. Pan, Y. L. Guo, and H. Kim, 2009: Ambient temperature and mortality: an international study in four capital cities of East Asia. *Sci, Total Environ.*, **408** (2), 390–396, https://doi.org/10.1016/j.scitotenv.2009.09.009.
7. Clough, S.A., M.W. Shephard, E. J. Mlawer, J.S. Delamere, M.J. Iacono, K. Cady-Pereira, S. Boukabara, and P.D. Brown, 2005: Atmospheric radiative transfer modeling: A summary of the AER codes. *J. Quant. Spectrosc. Radiat. Transfer*, **91** (2), 233–244, https://doi.org/10.1016/j.jqsrt.2004.05.058.
8. DelSole, T., L. Trenary, M.K. Tippett, and K. Pegion, 2017: Predictability of Week-3–4 Average Temperature and Precipitation Over the Contiguous United States. *J. Clim.*, **30** (10), 3499–3512, https://doi.org/10.1175/JCLI-D-16-0567.1
9. Ebert, E. E., 2001: Ability of A Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. *Mon. Wea. Rev.*, **129** (10), 2461-2480, https://doi.org/10.1175/1520-0493(2001)129%3C2461:AOAPMS%3E2.0.CO;2.
10. Feng, J., and S. Lu, 2019: Performance Analysis of Various Activation Functions in Artificial Neural Networks. *J Phys: Conf Ser.*, **1237**:022030, https://doi.org/10.1088/1742-6596/1237/2/022030.
11. Guan, H., and Co-authors, 2022: Gefsv12 Reforecast Dataset for Supporting Subseasonal And Hydrometeorological Applications. *Mon. Wea. Rev.*, **150** (3), 647–665, https://doi.org/10.1175/MWR-D-21-0245.1.
12. Hamill, T.M., and Co-authors, 2022: The Reanalysis for the Global Ensemble Forecast System, version 12. Monthly Weather Review, **150** (1), 59–79, https://doi.org/10.1175/MWR-D-21-0023.1.
13. Hamill, T.M., and J.S. Whitaker, 2006: Probabilistic Quantitative Precipitation Forecasts Based on Reforecast Analogs: Theory and application. *Mon. Wea. Rev.*, **134** (11), 3209-3229, https://doi.org/10.1175/MWR3237.1.
14. Hamill, T.M., and S.J. Colucci, 1998: Evaluation of Eta-RSM Ensemble Probabilistic Precipitation Forecasts. *Mon. Wea. Rev.*, **126** (3), 711–724, https://doi.org/10.1175/1520-0493(1998)126<0711:EOEREP>2.0.CO;2.
15. Hecht-Nielsen, R., 1990: *Neurocomputing.* Addison-Wesley, 433 pp.
16. Hersbach, H., and Co-authors, 2018: ERA5 hourly data on single levels from 1959 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, https://doi.org/10.24381/cds.adbb2d47.
17. IPCC, 2007: Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K and Reisinger, A. (eds.)]. IPCC, Geneva, Switzerland, 104 pp.
18. IPCC, 2013: Summary for Policymakers In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, T.F. Stocker, D. Quin, G.-K. Plattner, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, an P.M. Midgley, Eds., Cambridge: Cambridge University Press, 29 pp.
19. Johnstone, C. and E.D. Sulungu, 2021: Application of neural network in prediction of temperature: a review. *Neural Comput & Applic*., **33**, 11487–11498, https://doi.org/10.1007/s00521-020-05582-3

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

20. Kang, I.-S., and Co-authors, 2002: Intercomparison of the climatological variations of Asian summer monsoon precipitation simulated by 10 GCMs. *Clim. Dyn*., 19, 383–395, https://doi.org/10.1007/s00382-002-0245-9.
21. S. Li and A.W. Robertson, 2015: Evaluation of Submonthly Precipitation Forecast Skill from Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, **143** (7), 2871–2889, https://doi.org/10.1175/MWR-D-14-00277.1.
22. Lin, C.-Y., Y.-J. Chua, Y.-F. Sheng, H.-H. Hsu, C.-T. Cheng, and Y.-Y. Lin, 2015: Altitudinal and latitudinal dependence of future warming in Taiwan simulated by WRF nested with ECHAM5/MPIOM. *Int. J. Climatol*., **35**(8), 1800–1809. https://doi.org/10.1002/joc.4118.
23. Mariotti, A., P.M. Ruti, and M. Rixen, 2018: Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. *npj Clim. Atmos. Sci.*, **1**(4), https://doi.org/10.1038/s41612-018-0014-z.
24. Mert, G., D. Yang, and D. Srinivasan, 2020: Ensemble solar forecasting using data-driven models with probabilistic post-processing through GAMLSS. *Sol. Energy*, **208**, 612-622, https://doi.org/10.1016/j.solener.2020.07.040.
25. Nageswararao, M. M., U.C. Mohanty, S. Kiran Prasad, K.K. Osuri, and S.S.V.S. Ramakrishna, 2016: Performance evaluation of NCEP climate forecast system for the prediction of winter temperatures over India. *Theoretical and Applied Climatology*, **126**(3-4):1–15, https://doi.org/10.1007/s00704-015-1588-6.
26. Nageswararao, M. M., Y. Zhu, and V. Tallapragada, 2022a: Prediction Skill of GEFSv12 for Southwest Summer Monsoon Rainfall and Associated Extreme Rainfall Events on Extended Range scale over India. *Wea.Forecasting*, **37**(7), 1135–1156, https://doi.org/10.1175/WAF-D-21-0184.1.
27. Nageswararao, M.M., Y. Zhu, V. Tallapragada, and M.-S. Chen, 2022b: Prediction Skill of GEFSv12 in Depicting Monthly Rainfall and Associated Extreme Events over Taiwan during the Summer Monsoon. *Wea. Forecasting*, **37**(12), 2239–2262, https://doi.org/10.1175/WAF-D-22-0025.1.
28. Nageswararao, M.M., P. Sinha, U.C. Mohanty, and S. Mishra, 2020: Occurrence of More Heat Waves Over the Central East Coast of India in the Recent Warming Era. *Pure and Applied Geophysics*, **177**(2): 1143–1155, https://doi.org/10.1007/s00024-019-02304-2.
29. Pal, S., J. Das, P. Sengupta, and S.K. Banerjee, 2002: Short term prediction of atmospheric temperature using neural networks. *Mausam*, **53** (4), 471–480, https://doi.org/10.54302/mausam.v53i4.1662.
30. Saxena, A., N. Verma, and K.C. Tripathi, 2013: A review study of weather forecasting using artificial neural network approach. *Int J Eng Res Technol.,* **2**(11), 2029–2036.
31. Singh, G., and R.K. Panda, 2015: Bootstrap-based artificial neural network analysis for estimation of daily sediment yield from a small agricultural watershed. *International Journal of Hydrology Science and Technology.,* **5** (4), 333–348,https://doi.org/10.1504/IJHST.2015.072634.
32. Sperber, K.R., and Coauthors, 2001: Dynamical Seasonal Predictability of The Asian Summer Monsoon, *Mon. Weather Rev.,* **129** (9), 2226– 2247, https://doi.org/10.1175/1520-0493(2001)129<2226:DSPOTA>2.0.CO;2.
33. Verkade, J.S., J.D. Brown, P. Reggiani, and A.H. Weerts, 2013: Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *J. Hydrol.*, **501**, 73-91, https://doi.org/10.1016/j.jhydrol.2013.07.039.
34. Vitart, F., and Coauthors, 2017: The Sub-Seasonal to Seasonal Prediction Project (S2S) And the Prediction of Extreme Events. *Bull. Amer. Meteor. Soc.*, **98** (1),163-173, https://doi.org/10.1175/BAMS-D-16-0017.1.

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

35. Wang, B., and Z. Fan, 1999: Choice of South Asian Summer Monsoon Indices. *Bull. Amer. Meteor. Soc.*, **80** (4), 629–638, https://doi.org/10.1175/1520-0477(1999)080<0629:COSASM>2.0.CO;2.
36. Wilks, D.S., 2018: Chapter 3. Univariate ensemble forecasting. *Statistical Postprocessing of Ensemble Forecasts*, S. Vannitsem, D.S. Wilks, J.W. Messner, Eds., Elsevier, ,49-89, https://doi.org/10.1016/C2016-0-03244-8.
37. Xu, C.-Y., 1999: From GCMs to river flow: A review of downscaling methods and hydrologic modeling approaches. *Prog. Phys. Geogr.*, **23** (2), 229–249, https://doi.org/10.1177/030913339902300204.
38. Yilmaz, A. G., M.A. Imteaz, and G. Jenkins, 2011: Catchment flow estimation using artificial Neural Networks in the mountainous Euphrates Basin. *J. Hydrol* , **410** (1-2), 134–140, https://doi.org/10.1016/j.jhydrol.2011.09.031.
39. Zhou, X., Y. Zhu, B. Fu, D. Hou, J. Peng, Y. Luo, and W. Li, 2019: The development of the Next NCEP Global Ensemble Forecast System. *Science and Technology Infusion Climate Bulletin, NOAA's National Weather Service, 43rd NOAA Annual Climate Diagnostics and Prediction Workshop (CDPW),* 159-163.
40. Zhou, X., and Coauthors, 2022: The Development of the NCEP Global Ensemble Forecast System Version 12. *Wea. Forecasting*, **37** (6), 1069-1084, https://doi.org/10.1175/WAF-D-21-0112.1.
41. Zhu, Y., X. Zhou, M. Peña, W. Li, C. Melhauser, and D. Hou, 2017: Impact of Sea Surface Temperature Forcing on Weeks 3 And 4 Forecast Skill in The NCEP Global Ensemble Forecast System. *Wea. Forecasting*, **32**(6), 2159-2174, https://doi.org/10.1175/WAF-D-17-0093.1.
42. Zhu, Y., and Co-authors, 2018: Toward the improvement of sub-seasonal prediction in the National Centers for Environmental Prediction Global Ensemble Forecast System. *J. Geophys. Res.: Atmos.*, **123**(13), 6732-6745, https://doi.org/10.1029/2018JD028506

Correspondence to: Murali Nageswara Rao Malasala, CPAESS, UCAR at NOAA/NWS/NCEP/EMC, College Park, Maryland; murali.n.malasala@noaa.gov

Improving the NWS Subseasonal-to-Seasonal Forecast with the Unified Forecast System: Highlights of Modeling and Analysis Results

Yan Xue[1], Vijay Tallapragada[2], Avichal Mehra[2], Fanglin Yang[2], Michael Barlage[2], Yuejian Zhu[2], Cristiana Stan[3], Jim Kinter[3], Jeff Whitaker[4], Wanqiu Wang[5], Deepthi Achuthavarier[6,1], Kevin Garrett[1]

[1]Office of Science and Technology Integration, National Weather Service, Silver Spring, MD, [2]Environmental Modeling Center, National Centers for Environmental Prediction, National Weather Service, College Park, MD, [3]George Mason University, Fairfax, VA, [4]Physical Sciences Laboratory, Office of Atmospheric Research, National Oceanic and Atmospheric Administration, Boulder, CO, [5]Climate Prediction Center, National Centers for Environmental Prediction, National Weather Service, College Park, MD, [6]IBSS Corporation, Silver Spring, MD

ABSTRACT

1. Introduction

The Environmental Modeling Center (EMC) at the National Centers for Environmental Prediction (NCEP) within the National Weather Service (NWS) currently employs three different numerical models for operational global weather, subseasonal-to-seasonal (S2S) and seasonal applications, respectively. The Global Forecast System (GFS) is run for medium-range, high-resolution, deterministic weather prediction, the Global Ensemble Forecast System (GEFS) for probabilistic forecasts in the weather to subseasonal timescales and the Climate Forecast System (CFS) for seasonal timescales. The current operational versions of these models are: GFS version 16 (v16; NOAA 2021), GEFS version 12 (v12; NOAA 2020; Zhou et al. 2022) and CFS version 2 (v2; Saha et al. 2014). Both the GFSv16 and GEFSv12 are built within the framework of the Unified Forecast System (UFS) with the Geophysical Fluid Dynamics Laboratory (GFDL)'s Finite Volume Cubed Sphere (FV3) dynamical core (Harris et al. 2021), while the CFSv2 uses a spectral dynamical core (Sela. 1982). The GFSv16 and GEFSv12 do not have interactive ocean or sea-ice components and are run with one-way coupling from atmosphere to waves. A single member of the GEFSv12 ensemble is run with one-way coupling from atmosphere to aerosols.

Under the UFS Research to Operation Project (UFS-R2O), a global coupled UFS has been built with six components of the Earth system: the FV3 dynamical core and Common Community Physics Package (CCPP) (atmosphere), MOM6 (ocean), Noah-Multiparameterization (Noah-MP) (land surface), GOCART (aerosols), CICE6 (sea ice) and WW3 (ocean surface waves) with the Noah-MP connected through CCPP. The global coupled UFS will serve as the basis for developing the next generation Global Forecast System (GFSv17), the Global Ensemble Forecast System (GEFSv13) and the Seasonal Forecast System (SFSv1), which will serve as a replacement for CFSv2.

In this extended abstract, we provide a brief overview of the development of the coupled UFS (also referred to as the UFS S2S Application) including model specifications, enhancements, and a preliminary analysis of the model's ability in forecasting key S2S phenomena in its interim prototype versions that

Correspondence to: Yan Xue, Office of Science and Technology Integration, National Weather Service, Silver Spring, MD; yan.xue@noaa.gov

run with a deterministic member only. We start with a brief background of the UFS and UFS-R2O Project, followed by a summary of the UFS S2S Application, highlights of model performance and conclude with opportunities to engage and collaborate with the UFS S2S Application Team.

2. UFS and UFS-R2O Project

The UFS is a community-based modeling system that is primarily funded by NOAA presently, with an objective of modernizing and unifying the NCEP's operational production suite. The UFS is one of the outcomes of the University Corporation for Atmospheric Research (UCAR) Community Advisory Committee for NCEP (UCACN)'s review of the NCEP operational production suite where a key recommendation was to transition toward a unified and collaborative approach to accelerate research to operations in modeling and data assimilation.

The UFS-R2O Project is NOAA's largest investment in the UFS (Figure 1) with a total funding support of $13 million dollars per year in its first two years (July 2020 to June 2022) and approximately $9 million dollars in the third year (July 2022 to June 2023), that is jointly supported by NOAA's two-Line Offices, the NWS and the Office of Atmospheric Research (OAR). The support for UFS-R2O is leveraged through multiple Programs at the NWS, namely, the Next Generation Global Prediction System (NGGPS), Weeks 3-4, the Hurricane Forecast Improvement Program (HFIP) as well as OAR's Weather Program Office (WPO): The Joint Technology Transfer Initiative (JTTI), the Earth Prediction Innovation Center (EPIC) and the Sub-seasonal to Seasonal (S2S) Program. The UFS-R2O project is a broad collaboration among various organizations, including NCEP EMC, various research laboratories in OAR, the National Environmental Satellite, Data, and Information Service (NESDIS) and a number of Cooperative Institutes, as well as other national agencies and laboratories such as the National Center for Atmospheric Research (NCAR), Naval Research Laboratory (NRL) and university collaborators. Over the last two and a half years, the R2O project has established a unique collaboration where research and operational communities within and outside of NOAA directly work with EMC toward operational model development and implementation.

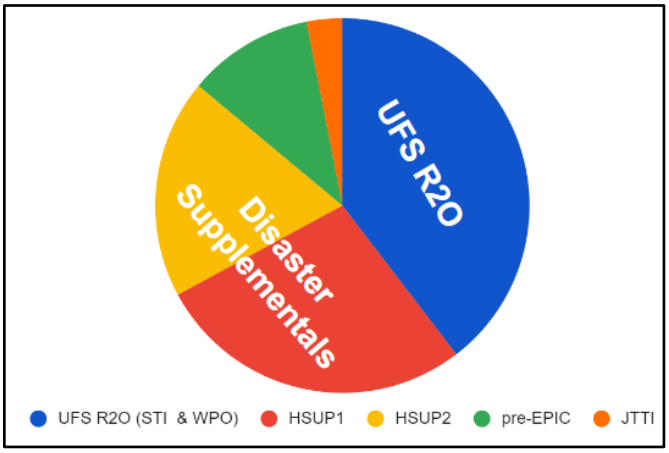

Figure 28: The NOAA's relative investments in the UFS, based on funding allocations in 2020 - 2022, including the UFS-R2O Project supported by NWS Science and Technology Integration (STI) Programs,

Correspondence to: Yan Xue, Office of Science and Technology Integration, National Weather Service, Silver Spring, MD; yan.xue@noaa.gov

and the OAR WPO, Hurricane Supplemental Programs 1 and 2 (HSUP 1 and 2), funds available to the Earth Prediction and Innovation Center (EPIC), prior to its formal establishment (pre-EPIC) and JTTI program. Data sources: STI and WPO budget allocations.

An overview of the UFS-R2O Project can be found in Whitaker et al. (2021). A major part of the UFS S2S Application development summarized here is carried out through the UFS-R2O project.

3. UFS S2S Application

The GEFSv13, currently planned for operational implementation in 2025, will be the first FV3-based global coupled model for operational ensemble sub-seasonal forecasts at NCEP. In addition to inclusion of atmosphere-land-ocean-sea ice-waves-aerosol coupling, this version will feature a few major enhancements compared to its predecessor (GEFSv12), including increased vertical resolutions, improved physics parameterizations, enhanced coupling infrastructure, a 30-year replay-based reanalysis and reforecast, and a weakly-coupled Global Data Assimilation System (GDAS) for initialization of real time forecasts.

The new coupled system is expected to improve probabilistic forecasts for surface temperature and precipitation over the continental United States (CONUS) for weeks 1 through 4, reduce tropical cyclone track and intensity errors, improve forecasts of tropical intraseasonal variability and its teleconnections, and improve and extend Northern Hemisphere 500 hectopascal height forecasts. Additional benefits are to enhance forecasts of atmospheric teleconnections, extreme weather frequencies, North American monsoons, fire weather, rapid drought onset, atmospheric river, and Arctic sea ice.

Correspondence to: Yan Xue, Office of Science and Technology Integration, National Weather Service, Silver Spring, MD; yan.xue@noaa.gov

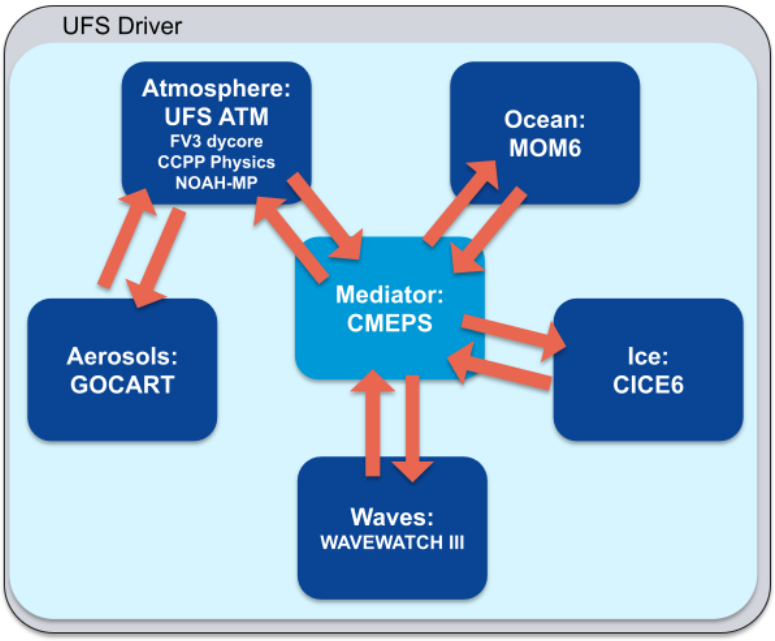

Figure 2: The UFS S2S Application configuration with component models and Community Mediator for Earth Prediction Systems (CMEPS).

The UFS S2S Application (Figure 2) will consist of the GFDL's FV3 dynamical core and Modular Ocean Model Version 6 (MOM6; Adcroft et al. 2019), the Los Alamos Sea ice model version 6 (CICE6; Hunke et al. 2017), the Noah-Multi Parameterization Land Surface Model (Noah-MP LSM; Niu et al. 2011), WAVEWATCH III (Tolman 1997; 1999; 2009) for ocean surface waves, and Goddard Chemistry Aerosol Radiation and Transport (GOCART; Chin et al. 2002) for atmospheric composition. Other key enhancements include replacement of the current Interoperable Physics Driver (IPD) with the Common Community Physics Package (CCPP) from the Developmental Testbed Center (DTC) and the introduction of the Community Mediator for Earth Prediction Systems (CMEPS) in place of the NOAA Environmental Modeling System (NEMS) mediator. The CCPP provides new functionalities at the physics-dynamics interface such as the ability to choose the order and frequency of parameterizations and group parametrizations allowing computations in between them. The CMEPS is a National Unified Operational Prediction Capability (NUOPC)-compliant mediator which uses Earth System Modeling Framework (ESMF) that provides data structures, interfaces, grid remapping, time management, model documentation and data communications (Theurich et al. 2016 and references therein).

Additionally, the physics package will have significant upgrades that include transition to a double moment Thompson microphysics (from single moment GFDL microphysics), updates in turbulent kinetic energy-based Moist Eddy-Diffusion Mass-Flux (EDMF) scheme for the Planetary Boundary Layer (PBL) including positive definite tracer advection and other optimizations, modifications in the Scale-aware Simplified Arakawa-Schubert (Sa-SAS; Han and Pan 2011) convection scheme for stochastic convective organization, better accuracy and computational efficiency in the radiation scheme, and adoption of the orographic gravity-wave drag parameterization (Kim and Doyle 2005) and non-orographic gravity-wave drag parameterization (Yudin et al. 2021).

Correspondence to: Yan Xue, Office of Science and Technology Integration, National Weather Service, Silver Spring, MD; yan.xue@noaa.gov

The atmospheric component in the cubed sphere grid is configured at horizontal resolution C384 (approximately25 kilometer grid spacing) and has 127 vertical layers with the model top at 0.01 hectopascals (80 kilometers). The ocean model with hybrid-coordinates and a tripolar grid has approximately 25 kilometers global resolution and 75 hybrid levels. The sea ice shares the same tripolar grid as the ocean and waves are run at 25 kilometers spatial resolution. The model will be initialized with a *weakly coupled* Global Data Assimilation System (GDAS) and will include 30-year reforecasts (1991 to2022) for calibration and validation purposes at various NCEP centers. The reforecasts will be initialized with a 30-year replay of the UFS coupled model to ERA5 atmosphere (Hersbach et al. 2020), Ocean ReAnalysis System 5 (ORAS5; Zuo et al. 2019) ocean and sea ice, Noah-MP spin-up and snow data assimilation. The reforecasts will be run up to a lead of 48 days and will be initialized every Monday and Thursday with 11 ensemble members. Additionally, daily initialization runs will be performed which will have a forecast lead of 16 days and 6 members.

4. UFS S2S Prototype Evaluations

While a finalized model version and reforecasts are expected toward 2025, here, we provide an interim update, leveraging evaluations completed and reported thus far using various prototype versions of the UFS coupled model. The prototypes were initialized on the 1st and 15th of each month for April 1, 2011 through March 15, 2018, producing a 35-day run for each initialization with a total of 168 hindcasts for each prototype. The prototype configurations P1 through P8 are summarized in Table 1. Although not yet at the anticipated configuration described in the Section 3, the prototype versions and their corresponding evaluations mark key milestones and are indicative of a number of potential improvements in the final operational version. For a comprehensive summary of prototypes P1-P7 evaluations, the reader is referred to Stefanova et al. (2022).

Correspondence to: Yan Xue, Office of Science and Technology Integration, National Weather Service, Silver Spring, MD; yan.xue@noaa.gov

Table 1: Summary of UFS S2S Prototype configurations, adapted from Stefanova et al. (2022).

<table>
<tr><th>Prototype</th><th colspan="6">Model Type</th><th>Mediator</th></tr>
<tr><td></td><td colspan="3">Atmospheric Model</td><td>Ocean Model</td><td>Wave Model</td><td>Ice Model</td><td></td></tr>
<tr><td>P1</td><td rowspan="5">FV3
64 layers, Non-Fractional grid (model top at 54 kilometers</td><td rowspan="2">GFSv15.2, IPD driver</td><td rowspan="4">Noah LSM</td><td rowspan="6">MOM6</td><td rowspan="2">N/A</td><td rowspan="2">CICES</td><td rowspan="2">NEMS</td></tr>
<tr><td>P2</td></tr>
<tr><td>P3.1</td><td rowspan="3">GFSv15.2, CCPP driver</td><td rowspan="5">Wavewatch III</td></tr>
<tr><td>P4</td></tr>
<tr><td>P5</td><td>CICE6 (Mushy thermodynamics not turned on)</td><td rowspan="4">CMEPS</td></tr>
<tr><td>P6</td><td rowspan="3">FV3
127 layers, Fractional grid (model top at 80 kilometers)</td><td>GFSv16</td><td rowspan="3">CICE6 (Mushy thermodynamics turned on)</td></tr>
<tr><td>P7</td><td>Modified GFSv16</td><td>Noah-MP</td></tr>
<tr><td>P8</td><td>Further Modified GFSv16</td><td>Modified Noah-MP</td></tr>
</table>

Overall, in the most recent prototypes (P6-P8), the benchmark anomaly correlation (AC) scores for 2-meter temperature, Sea Surface Temperature (SST), 500 hectopascal heights and precipitation over CONUS for Week 1 and Weeks 3-4 leads are generally comparable or superior to CFSv2. The Madden Julian Oscillation (MJO) forecast skills in P8 examined using Realtime Multivariate MJO (RMM) bivariate correlation coefficient (Wheeler and Hendon 2004) is greater than or equal to 0.6 through day 21, which is a noticeable improvement compared to the CFSv2 as well as earlier prototypes. In general, compared to the CFSv2, the MJO forecast skill is enhanced by a lead time of 3–5 days for RMM1 and RMM2 and 4–7 days for the combined RMM index in all prototypes.

While a detailed diagnosis of the MJO skill improvement would need targeted model sensitivity experiments, preliminary tests with uncoupled runs indicate that the improved skill can be attributed to enhancements in convective parameterization including Stochastic Cellular Automata (Bengtsson et al., 2021), changes in the PBL scheme and a more accurate representation of the diurnal sea-surface skin temperature variability (Stefanova et al. 2022). The land model was transitioned to the Noah-MP land model in P7 and incorporates sub-grid tiles for vegetation and bare soil, separate canopy structure, and groundwater transfer and storage. P7 shows improved land-atmosphere coupling and improved basin-scale water budget analysis over the majority of large river basins. The land-atmosphere coupling diagnostics show that the atmosphere responds more realistically to soil moisture anomalies in P7 compared to P6 and P5 that employed the Noah land model (Seo et al. 2021). The Arctic sea ice area prediction shows promising improvements with the activation of the mushy thermodynamics option in the ice model in P7 and onward. These improvements and other advancements have been maintained for P8, the final configuration of these prototypes.

While the above results are encouraging, there are outstanding challenges such as poor precipitation forecasts over CONUS in weeks 3 and 4 and increased model biases. Further tests and evaluations with a reasonably good ensemble size are also needed for a good understanding of probabilistic skill scores.

Correspondence to: Yan Xue, Office of Science and Technology Integration, National Weather Service, Silver Spring, MD; yan.xue@noaa.gov

These efforts are currently ongoing at EMC and in the UFS-R2O Project. A subset of the data from P5 through P8 that includes major weather variables for atmosphere, land, ocean, sea ice, and ocean waves is publicly hosted on Amazon Web Services[1], through the UFS-R2O Project and the NOAA Open Data Dissemination (NODD) Program.

5. Opportunities for Collaboration

While UFS S2S model development currently takes place largely within NCEP-EMC, through the UFS-R2O Project and other collaborative initiatives within the UFS, a growing community of S2S enthusiasts including researchers, model developers, evaluators, forecasters and stakeholders have come together to form a UFS S2S Application Team. The primary objectives of this team are: 1) to collect and prioritize forecast objectives working with NWS forecasters and model users in general, 2) establish scientific goals for the model development and ensure that they meet NWS forecast priorities, and 3) to promote or conduct model evaluations and comparisons in order to stay abreast on model performance and deficiencies. The UFS S2S Application Team in collaboration with EMC's coupled modeling group organizes monthly All-Hands meetings on S2S science topics, including prediction skill of UFS and other models in this timescale and new diagnostics designed to advance understanding of Earth system predictability in S2S timescale. These meetings are also leveraged to identify and attract projects that can be spun up to fill potential gaps in the UFS model evaluation. For more information on the S2S Application Team and All-Hands meetings the reader is referred to their website[2].

[1] https://registry.opendata.aws/noaa-ufs-s2s/

[2] https://vlab.noaa.gov/web/ufs-r2o/ufs-s2s-applications-team

Correspondence to: Yan Xue, Office of Science and Technology Integration, National Weather Service, Silver Spring, MD; yan.xue@noaa.gov

6. References

1. Adcroft, A., and Coauthors, 2019: The GFDL Global Ocean and Sea Ice Model OM4.0: Model Description and Simulation Features. J. Adv. Model. Earth Syst., **11**(10), 3167–3211, https://doi.org/10.1029/2019MS001726.
2. Bengtsson, L., J. Dias, S. Tulich, M. Gehne, and J.-W. Bao, 2021: A stochastic parameterization of organized tropical convection using cellular automata for global forecasts in NOAA's Unified Forecast System. J. Adv. Model. Earth Syst., **13**(1), https://doi.org/10.1029/2020MS002260.
3. Chin, M., and Coauthors, 2002: Tropospheric Aerosol Optical Thickness from The GOCART Model and Comparisons with Satellite and Sun Photometer Measurements. *J. Atmos. Sci*., **59**(3), 461–483, https://doi.org/10.1175/1520-0469(2002)059<0461: TAOTFT>2.0.CO;2.
4. Han, J., and H.-L. Pan, 2011: Revision of Convection and Vertical Diffusion Schemes in The NCEP Global Forecast System. *Wea. Forecasting*, **26**(4), 520–533, https://doi.org/10.1175/WAF-D-10-05038.1.
5. Harris, L., X. Chen, W. Putman, L. Zhou, and J-H. Chen, 2021: A Scientific Description of the GFDL Finite-Volume Cubed-Sphere Dynamical Core, NOAA technical memorandum OAR GFDL, 2021-001, https://doi.org/10.25923/6nhs-5897.
6. Hersbach, H. and Coauthors, 2020: The ERA5 global reanalysis, *Q. J. Roy. Meteorol. Soc*., **146**(730), 1999–2049, https://doi.org/10.1002/qj.3803.
7. Hunke, E., L. William, J. Philip, T. Adrian, J. Nicole, and E. Scott. CICE, The Los Alamos Sea Ice Model. Computer software. https://www.osti.gov//servlets/purl/1364126. Vers. 00. USDOE, (Accessed12 May. 2017).
8. Kim, Y-J. and J.D. Doyle, 2005: Extension of an orographic-drag parametrization scheme to incorporate orographic anisotropy and flow blocking. *Q. J. Roy. Meteorol. Soc*., **131**(609), 1893-1921, https://doi.org/10.1256/qj.04.160.
9. Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res*., **116**(D12), https://doi.org/10.1029/2010JD015139.
10. NOAA, 2020: Announcement of upgrade to the Global Ensemble Forecast System (GEFS), coupled with the Global Wave Ensemble System (GWES) and the NEMS GFS aerosol component (NGAC): Effective September 23, 2020 and request for comments. Service Change Notice 20-75. National Weather Service Headquarters, Silver Spring MD, https://www.weather.gov/media/notification/SCN_20-75_GEFSv12_Changes.pdf.
11. NOAA, 2021: Upgrade NCEP Global Forecast Systems (GFS) to v16: Effective March 17, 2021. Service Change Notice 21-20, Updated. National Weather Service Headquarters, Silver Spring MD, https://www.weather.gov/media/notification/scn_21-20_gfsv16.0_aaa_update.pdf.
12. Saha, S., S. and Coauthors, 2014: The NCEP Climate Forecast System Version 2, *J. Climate*, **27**(2), 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1.
13. Sela, J.G., 1982: The NMC Spectral Model. NOAA Tech Report, NWS, 30 pp. https://repository.library.noaa.gov/view/noaa/7044/noaa_7044_DS1.pdf.
14. Seo, E., P. Dirmeyer, M. Barlage, H. Wei, M. Ek, 2021: Investigation of land-atmosphere interaction in UFS and its influence on model mean bias, UFS Webinar Series (Virtual), https://www.ufscommunity.org/ufs-webinar-series-old/ufs-webinar-archive/#dirmeyer (*Accessed online, May 3, 2023*).
15. Stefanova, L. and Coauthors, 2022: Description and results from UFS coupled prototypes for future global, ensemble and seasonal forecasts at NCEP, NOAA/NCEP Office Note 510, 252 pp, https://doi.org/10.25923/knxm-kz26.

Correspondence to: Yan Xue, Office of Science and Technology Integration, National Weather Service, Silver Spring, MD; yan.xue@noaa.gov

16. Theurich, G, and Coauthors, 2016: The Earth System Prediction Suite: Toward A Coordinated U.S. Modeling Capability, *Bull Am Meteorol Soc*., **97**(7), 1229-1247, https://doi.org/10.1175/BAMS-D-14-00164.1.
17. Tolman, H. L., 1997: User manual and system documentation of WAVEWATCH-III version 1.15. NOAA / NWS / NCEP / OMB Technical Note 151, 97 pp.
18. Tolman, H. L., 1999: User manual and system documentation of WAVEWATCH-III version 1.18. NOAA / NWS / NCEP / OMB Technical Note 166, 110 pp.
19. Tolman, H. L., 2009: User manual and system documentation of WAVEWATCH III version 3.14. NOAA / NWS / NCEP / MMAB Technical Note 276, 194 pp.
20. Wheeler, M.C., and H.H. Hendon, 2004: An All-Season Real-Time Multivariate MJO Index: Development of An Index for Monitoring and Prediction. *Mon. Wea. Rev*., **132**(8), 1917-1932, https://doi.org/10.1175/1520-0493(2004)132%3C1917:AARMMI%3E2.0.CO;2.
21. Whitaker, J., J. Kinter and V. Tallapragada, 2021: Celebrating the first anniversary of the UFS R2O Project, Bulletin Of the UFS Community, https://ufscommunity.org/wp-content/uploads/2021/07/Bulletin_UFS_Community_April-June_2021_Full.pdf.
22. Yudin, V. A., F. Yang., S. Karol, T. Fuller-Rowell, A. Kubaryk, H. Juang, S. Kar, J. Alpert, and Z. Li, 2020: The Unified Gravity Wave physics in the vertically extended atmosphere models of NGGPS and UFS. *1st UFS Users' Workshop*, Boulder, Colorado.
23. Zhou, X. and Coauthors, 2022: The Development of the NCEP Global Ensemble Forecast System Version 12. *Wea. Forecasting*, **37**(6),https://doi.org/10.1175/WAF-D-21-0112.1.
24. Zuo, H., M.A. Balmaseda, S. Tietsche, K. Mogensen, and M. Mayer, 2019: The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment, *Ocean Sci*., **15**(3), 779–808, https://doi.org/10.5194/os-15-779-2019

Correspondence to: Yan Xue, Office of Science and Technology Integration, National Weather Service, Silver Spring, MD; yan.xue@noaa.gov

The Development of UFS Coupled GEFS for Subseasonal and Seasonal Forecasts

Yuejian Zhu[1], Bing Fu[2], Hong Guan[3], Eric Sinsky[2], Bo Yang[2]
[1]NOAA/NWS/NCEP/EMC, College Park, MD
[2]IMSG at NOAA/NWS/NCEP/EMC, College Park, MD
[3]SRG at NOAA/NWS/NCEP/EMC, College Park, MD

ABSTRACT

1. Introduction

Subseasonal to seasonal (S2S) prediction is a type of weather prediction that focuses on predicting weather patterns over periods of weeks to months. It is a relatively new area of research and is becoming increasingly important due to its potential to improve the accuracy of long-term forecasts. Seasonal prediction is the forecasting of weather patterns over periods of months to years. It is a complex and challenging task due to the difficulty in accurately predicting large-scale climate patterns, such as the El Niño-Southern Oscillation (ENSO). Challenges associated with S2S prediction and seasonal prediction include the need for better understanding of complex physical processes, improved data assimilation techniques to accurately capture initial conditions, and the development of better models for data analysis and prediction.

The Global Ensemble Forecast System version 12 (GEFS v12) was implemented in the National Centers for Environmental Prediction (NCEP) operations in September 2020, which is the first Unified Forecast System (UFS) application with coupling to Wave Watch 3 (WW3). Furthermore, GEFS v12 contains 31 ensemble members, a horizontal resolution of about 25 kilometers, and an integration time of 35 days to cover subseasonal prediction. GEFS v12 has demonstrated forecast capability and excellent performance for subseasonal time scales including Madden-Julian Oscillation (MJO) predictions (*Zhou et al, 2022; Guan et al, 2022*). The Climate Forecast System version 2 (CFSv2) was implemented in NCEP operations in 2010 (*Saha et al, 2012*), which is a full coupled system to cover subseasonal and seasonal prediction, which has very comparable skills with other S2S prediction systems around the world.

In collaboration with Unified Forecast System Research to Operations Medium-Range Weather/Subseasonal to Seasonal (UFS-R2O MRW/S2S) applications, the focus of this effort is to improve NCEP's subseasonal to seasonal predictions targeting GEFS v13 for operational implementation in FY24, and extending a similar system for seasonal prediction toward SFSv1 (Seasonal Forecast System version 1) implementation (2026).

The improvements of MJO and ENSO predictions are the keys of subseasonal to seasonal prediction. The UFS fully coupled GEFS experiments have been carried out to demonstrate the capability and improvement to the subseasonal prediction (GEFSv12 and CFSv2) and seasonal prediction (CFSv2). In this study, a coupled GEFS experiment based on UFS prototype version 5 (P5; *Stefanova et al, 2022*) has been carried out with optimum atmospheric model perturbations. The experiments are performed during a 2-year period (initialized once per week at 00 UTC time from October 2017 - September 2019) with 10 perturbed members and 1 unperturbed member, out to 35 days for high resolution (C384, about 25 kilometers for atmosphere and 0.25 degree for MOM6 and CICE6) and low resolution (C192, about 50 kilometers for atmosphere and 0.5 degree for MOM6 and CICE6; Zhu et al. 2023). A low resolution (C192, 0.5d MOM6 and CICE6) configuration will be extended to 9 months forecasts initialized on June

Correspondence to: Yuejian Zhu, NOAA/NWS/NCEP/EMC, College Park, Maryland; yuejian.zhu@noaa.gov

1 2015 (El Niño case), June 1 2017 (La Niña case) and June 1 2012 (neutral case) respectively with 40 unperturbed forecasts and 1 unperturbed forecast (*Zhu et al, 2023*).

2. Experiments Configurations

The NCEP Global Ensemble Forecast System (GEFS), by using a set of initial perturbations generated from EnKF (ensemble Kalman filter) analysis (*Zhou et al. 2016; Zhou et al. 2017; Zhou et al. 2022*), has been in daily operations since 1992. It has been providing reliable weather and week-2 probabilistic forecast guidance that has translated into valuable information for the general public. With the science enhancements in past years, current GEFS, especially the SubX version, has reached an extraordinary performance across the scales from weather to subseason (*Zhu et al. 2017; 2018*). The SubX version of NCEP GEFS has been presented by *Zhu et al. (2018); Li et al. (2018); Zhu et al. (2019); Guan et al. (2019).* The FV3-GEFS (Global Ensemble Forecast System Finite-Volume-Cubed Sphere dynamical core) version has been summarized by *Zhou et al. (2022); Guan et al. (2022)*. The coupled GEFS version is also discussed by *Zhu et al. (2023)*.

Table 5: Model configurations of CGEFS-H, CGEFS-L and CGEFS-S. See the model description section for the details.

Model	**CGEFS-H (p5)**	**CGEFS-L (p5)**	**CGEFS-S (p5)**
ATM	GFSv15 Retrospective Anl and EnKF (f06) FV3 C384L64	GFSv15 Retrospective Anl and EnKF (f06) FV3 C192L64	GFSv15 Retrospective Anl and EnKF (f06) FV3 C192L64
OCN	CFSR Salinity and T MOM6 0.25dL75	CFSR Salinity and T MOM6 0.5dL75	CFSR Salinity and T MOM6 0.5dL75
ICE	CPC ice analysis CICE6 0.25d	CPC ice analysis CICE6 0.5d	CPC ice analysis CICE6 0.5d
WAV	WW3 0.5d Start from rest	WW3 0.5d GFSv15 wind/ice forcing	N/A
Physics	Hybrid-EDMF Sa-SAS GFDL-MP GWD (stationary oro) NOAH-LSM	Hybrid-EDMF Sa-SAS GFDL-MP GWD (stationary oro) NOAH-LSM	Hybrid-EDMF Sa-SAS GFDL-MP GWD (stationary oro) NOAH-LSM NSST
Stochastics (atmosphere)	SPPT (0.56,0.28,0.14,0.056,0.028) SKEB (0.7)	SPPT (0.56,0.28,0.14,0.056,0.028) SKEB (0.7)	SPPT (0.56,0.28,0.14,0.056,0.028) SKEB (0.7)
Stochastics (ocean)	N/A	N/A	oSPPT ePBL
Ensemble size	10 perturbed 1 unperturbed	10 perturbed 1 unperturbed	40 perturbed 1 unperturbed
Forecast leads	35 days	35 days	9 months

The configurations of the experiments have been summarized in table 1. There are two experiments with a similar configuration except that the horizontal resolutions of the atmosphere, ocean

Correspondence to: Yuejian Zhu, NOAA/NWS/NCEP/EMC, College Park, Maryland; yuejian.zhu@noaa.gov

and sea ice models are different. The low resolution coupled GEFS (CGEFS-L hereafter) is about 0.5 degree (approximately 50 kilometers) and the high resolution coupled GEFS (CGEFS-H hereafter) is about 0.25 degree (approximately 25 kilometers) for all model components. The SPPT (Stochastic Perturbed Physical Tendency) scheme and configurations are similar to the GEFSv12 reforecast, but the amplitudes (or coefficients) are reduced by 30% compared to the operational GEFSv12 (*Zhou et al. 2021; Zhu et al. 2019b*) in order to reduce tropical overdispersion. The adjusted coefficients are 0.56, 0.28, 0.14, 0.056 and 0.028 for 5 scales. The SKEB (Stochastic Kinetic Energy Backscatter) scheme is the same as GEFSv12 except for an increased amplitude from 0.6 to 0.7 (*Zhou et al. 2021; Zhu et al. 2019b; Zhu et al. 2023*).

For CGEFS-L and GEFS-H, (see table 1) the experiment period spanned from October 4, 2017 to September 25, 2019 (2 years) initialized every Wednesday at 00UTC (once per week; 104 total runs), and 11 (1 unperturbed and 10 perturbed) members with a uniform resolution out to 35 lead days. For CGEFS-S (seasonal prediction), the overall configurations are the same as CGEFS-L except for including NSST (Near Sea Surface Temperature); introducing ocean stochastics for sea temperature, salinity and layer thickness (named oSPPT), and perturbed ocean PBL (Planetary Boundary Layer) energy generation and dissipation (named ePBL); integrating to 9 months for 3 special cases (El Niño - 2015, La Niña - 2017 and neutral - 2012).

3. Evaluation Methodology

The GEFS v12 model performance has been presented in many studies (Zhou et al. 2022; Guan et al. 2022). Various metrics have been used to assess the prediction skill and predictability of the forecast system, with dependencies on forecast elements, different spatial/temporal scales, and different forecast regions. In this study, anomaly correlation (AC) has been used to assess forecast skill and potential skill for weather. The ratio of error-spread is used to quantify the forecast uncertainty. Meanwhile, the bivariate anomaly correlation (RMM1 and RMM2), a traditional real-time multivariate (RMM) MJO index (W*heeler and Hendon 2004; Gottschalck et al. 2010*), has been used to evaluate tropical forecast skill and potential forecast skill.

For seasonal prediction, the Sea Surface Temperature (SST) anomaly is presented for the Niño 3.4 domain (170^{o}W-120^{o}W; 5^{o}N-5^{o}S) for each case. The mean error, Root Mean Square (RMS) error and ensemble spread of SST (or bias) have been presented in the global and Niño 3.4 domain. Meanwhile, day-to-day RMS error and ensemble spread of the tropical 850hPa and 2-meter temperature are used to measure the ensemble performance of the tropical atmosphere.

4. Results

Two-year experiments (October 2017 - September 2019) have been carried out for weather to subseasonal prediction. The NCEP Global Data Assimilation System (GDAS) analyses and NCEP/NCAR (National Center for Atmospheric Research) 40 years reanalysis climatology have been used to calculate forecast skills. For three special cases of seasonal prediction, OSTIA (The Operational Sea Surface Temperature and Ice Analysis) is used as a reference SST. SST climatology is generated by a 7-year (2011-2018) OSTIA SST.

4.1. Overall error to subseasonal forecast skills

To summarize the performance of the global forecast, the average AC scores of week-1 (days 1-7); week-2 (days 8-14), weeks 3 & 4 (days 15-28) and month 1 (days 1-30) for Northern Hemisphere

Correspondence to: Yuejian Zhu, NOAA/NWS/NCEP/EMC, College Park, Maryland; yuejian.zhu@noaa.gov

(NH), Southern Hemisphere (SH), Tropical (TR) and Pacific North American (PNA) domains have been plotted (not shown). Overall, coupled ensembles (CGEFS-L and CGEFS-H) are better than the un-coupled ensemble (the GEFSv12 reforecast) for most weekly-averaged lead-times and all the domains. In the tropical domain, the coupled ensembles are significantly better than the un-coupled ensemble for all the weekly-averaged lead-times. These improvements mainly occur for two reasons. First, the AC score improvements are mainly due to the coupling of the atmosphere to the ocean and sea ice, which improves the lower boundary conditions essential to the prediction of the atmospheric circulation, e.g., Rossby waves and tropical convection. Second, the enhanced model performance benefits from the adjusted coefficients of the stochastic schemes (SPPT and SKEB), which improve the representation of forecast uncertainty and ensemble mean solution.

4.2. Error-Spread Ratio of Zonal Winds

To discuss the relationship of ensemble spread and RMS error, a ratio of them has been used to quantify the forecast uncertainty. An investigation of the ratio through a spatial distribution is generated from latitude-vertical (200hectopascals - 1000hectopascals) cross section maps of the zonal mean for the same forecast lead-time (144-hours) (Fig. 1). There is overdispersion over the tropics for all vertical levels in the GEFSv12 reforecast. After adjusting the SPPT and SKEB schemes, the spread-error ratio for zonal mean has remarkably improved for the two coupled GEFS experiments (CGEFS-H and CGEFS-L), but the resolution does not have an impact on the error-spread ratio.

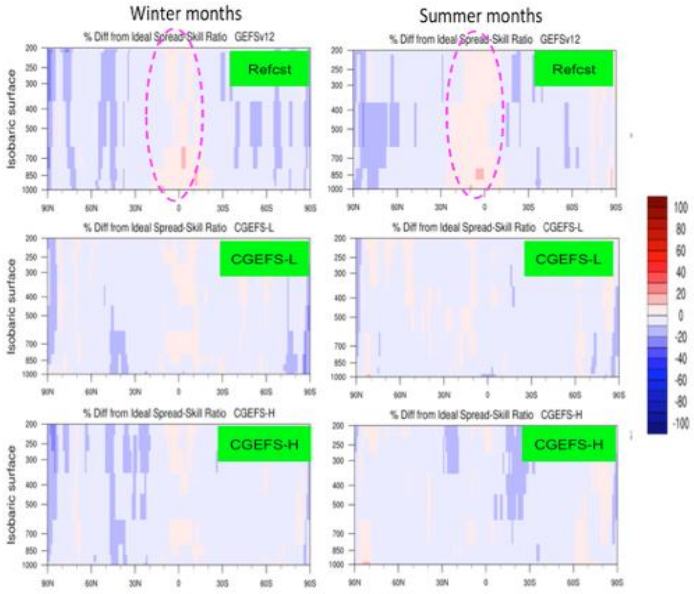

Figure 29: The vertical cross section of the spread-skill ratio for boreal two winter six months (left column) and boreal two summer six months (right column) of zonal wind from surface (1000 hectopascals) to 200 hectopascals in vertical, for 144 hours (6 days) forecasts of October 2017 – September 2019) and for the GEFSv12 reforecast (top), CGEFS-L (middle) and CGEFS-H (bottom).

Correspondence to: Yuejian Zhu, NOAA/NWS/NCEP/EMC, College Park, Maryland;
yuejian.zhu@noaa.gov

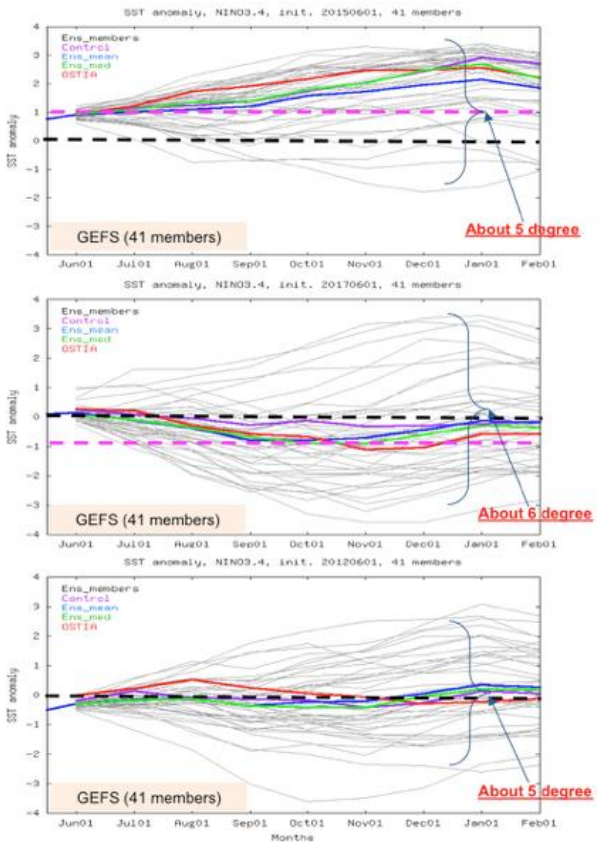

Figure 30: The SST forecast anomaly for Niño 3.4 domain average and monthly average initialized from June 1 2015 (top: El Niño case); June 1 2017 (middle: La Niña case) and June 1 2012 (bottom: neutral case). The individual ensemble members (40 perturbed) are in light gray, ensemble control (unperturbed) is in purple, ensemble mean is in blue and ensemble medium is in green. The OSTIA SST anomaly is in red.

4.3. ENSO Prediction of Seasonal Experiments

The El Niño-Southern Oscillation (ENSO) can be measured by a number of indices, such as the Southern Oscillation Index (SOI) and the Niño 3.4 Index. The SOI measures the difference in air pressure between Tahiti and Darwin, Australia, while the Niño 3.4 Index measures the sea surface temperature anomalies in an area of the Pacific Ocean near the equator. These indices are used to measure the strength of the El Niño and La Niña phases and to make predictions about the future state of ENSO. In this study (CGEFS-S), the forecast anomaly of SST (Fig. 2) has been calculated to compare the observed SST anomaly which could show us the relative closeness of the prediction and observation. Please note that there is no bias correction of SST since we do not have model climatology due to limited resources.

To discuss three special cases of seasonal prediction, we have seen the seasonal forecast capability before the forecast is calibrated, which will be discussed in section 4.4. There are several characteristics we should consider 1). Ensemble mean could capture the forecast anomaly in general; 2). Ensemble median should have a better representation of SST anomaly because the forecast distribution is not normally distributed; 3). The forecast ensemble is biased in terms of seasonal variability which needs more discussion; 4). Ensemble spread (or forecast variability) is much larger than climatological variability (and/or RMS error) which could be a potential issue.

Correspondence to: Yuejian Zhu, NOAA/NWS/NCEP/EMC, College Park, Maryland; yuejian.zhu@noaa.gov

4.4. SST Mean Error of Seasonal Prediction

Figure 3 presents the SST mean errors from three cases averaged for global domain (top left) and Niño 3.4 domain (top right), and shows that the SST prediction is very close to the observation from global average which should indicate there is a very reliable coupling model. However, the SST prediction of the Niño 3.4 domain is biased apparently. There is a cold bias around summer time, then a warm trend starts from winter time.

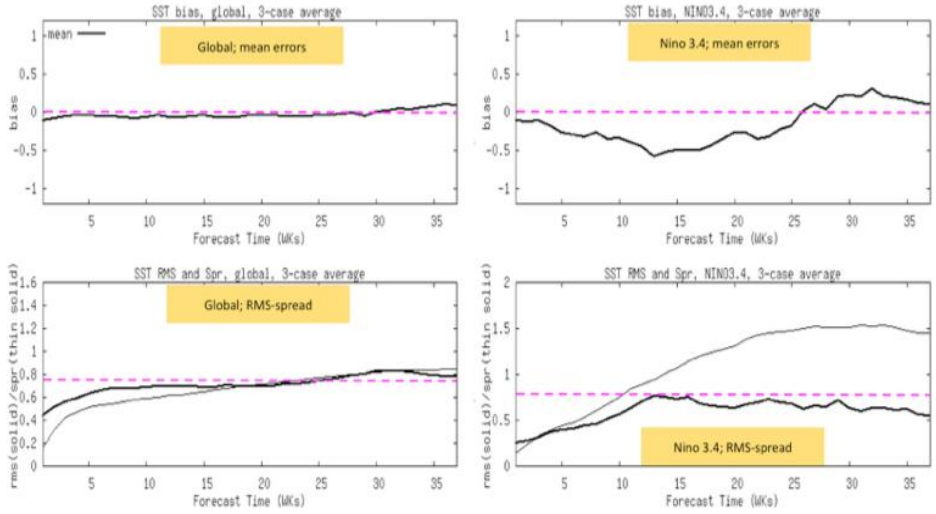

Figure 31: Three cases (El Niño case; La Niña case; Neutral case) average of mean errors for global domain (top left) and Niño 3.4 domain (top right) from OSTIA SST. The average of RMS errors and spread of ensemble mean for global domain (bottom left) and Niño 3. 4 domain (bottom right).

The SST RMS errors and spreads of the global domain (Fig. 3: bottom left) and Niño 3.4 domain (Fig 3: bottom right) are presented based on the weekly average. The ensemble spread of global SST is still under-dispersive for short lead-times which may indicate that it requires the ocean initial perturbation. For longer lead-times, the ratio of error-spreads is very close to each other, but we should understand that the RMS error may be overestimated due to forecast bias. For the small domain (Niño 3.4), the ensemble spread is much larger than the RMS error for longer lead-times which presents larger forecast uncertainty and non-Gaussian distribution as well. The cause of the over-dispersion and non-Gaussian distribution is unclear.

The SST mean error (or bias) are displayed from three cases (Fig. 4: top left), El Niño case (Fig. 4: top right), La Nina case (Fig. 4: bottom left), neutral case (Fig. 4: bottom right). Based on the three-case average, it shows the cold bias during summer time and tends to warm starting winter time. Meanwhile, three individual cases have demonstrated the same characteristics.

Correspondence to: Yuejian Zhu, NOAA/NWS/NCEP/EMC, College Park, Maryland; yuejian.zhu@noaa.gov

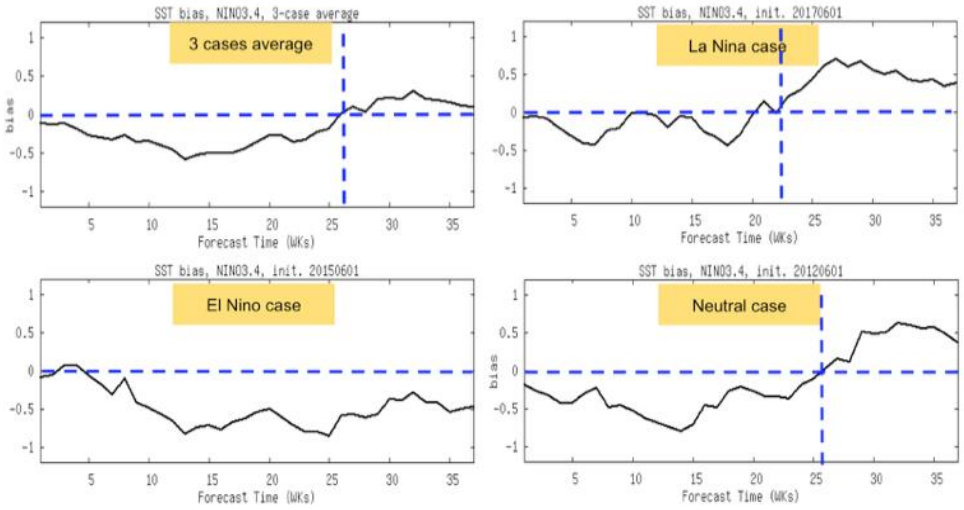

Figure 4: The time series of Niño 3.4 domain average SST errors between OSTIA SST and forecasts for three cases (El Niño, La Niña and Neutral) average (top left); El Niño case - top right; La Niña case - bottom left; Neutral case - bottom right.

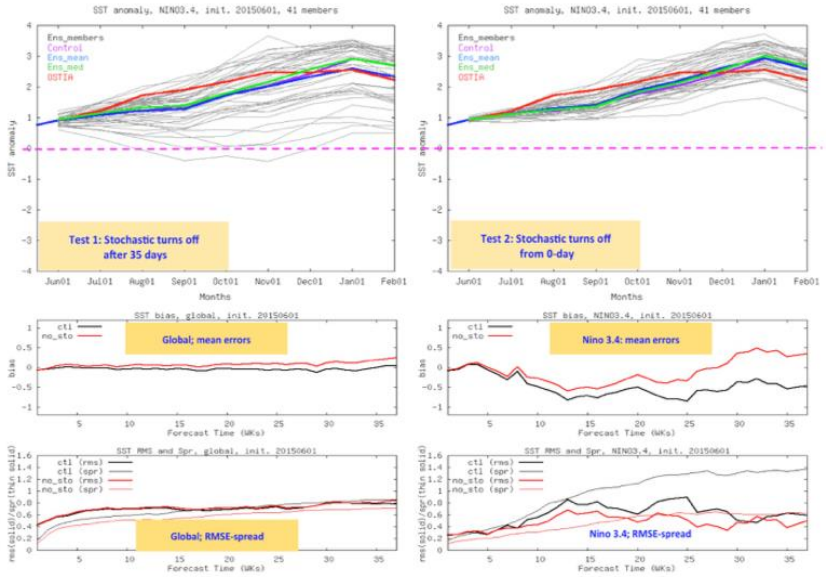

Figure 5: The top 2 plots are for El Niño case from two configurations (no stochastic perturbations for both atmosphere and ocean from day-35 - Test 1: top left; and no stochastic perturbations for both atmosphere and ocean from day-0 - Test 2: top right). The bottom four plots are comparison of control and Test 2 for global SST mean error (middle left); Niño 3.4 domain SST mean error (middle right); global average SST RMS error and spread (bottom left); and Niño 3.4 domain SST RMS error and spread (bottom right).

4.5. Tropical 850hectopascal and 2-Meter Temperature

In order to understand the large uncertainty (spread) for the Niño 3.4 domain (not global domain), the experiments for the El Niño case (initialized on June 1 2015) have been conducted to exclude ocean

Correspondence to: Yuejian Zhu, NOAA/NWS/NCEP/EMC, College Park, Maryland; yuejian.zhu@noaa.gov

and atmosphere stochastics from day-0 and day-35. By comparing Figure 2 (top) to Figure 5 (no stochastic perturbations from day-35; top left; and no stochastic perturbations from day-0; top right), the ensemble spreads are much reduced when stochastic schemes are turned off. If the stochastic perturbations are stopped from the beginning, the spread of longer lead-time forecasts is smaller which is close to the RMS error for the Niño 3.4 domain (not shown). There is another interesting characteristic (not shown) which shows the SST mean error differences with and without stochastics. For the global domain, the SST mean error is improved from warm mean error to near perfect by the stochastic perturbations. For the Niño 3.4 domain, the SST mean error is increased too cold by the stochastic perturbations. A further investigation and analysis is required to understand the reason for forecast mean error (and/or bias).

By contrast of SST prediction of the tropical and Niño 3.4 domains, the temperature of 850hectopascal and 2-meter have been presented for the 20 degrees North-20 degrees South tropical band (not shown). For both of the 2-meter and 850hectopascal temperature, there are less differences between the stochastics, which may indicate that there is less impact and contribution from the atmosphere to the tropical and Niño 3.4 domains.

5. Conclusion

The ensemble prototypes (CGEFS-H and CGEFS-L) and seasonal experiments (CGEFS-S) have demonstrated the forecast capability for weather, subseasonal and seasonal predictions. The preliminary results can be summarized as follows:

1) Both CGEFS-H and CGEFS-L perform better than GEFSv12 for all lead-times and all domains out to weeks 3 & 4. It improves significantly for tropical areas which should be due to the benefits of using a coupled system.
2) It is unnecessary to have better forecast skills from a higher resolution model (CGEFS-H) than lower resolution (CGEFS-L). Based on the 2-year experiment, the lower resolution is slightly better than high resolution for NH week-2.
3) The tropical over-dispersion of GEFSv12 is much improved through a full coupled system with tuning stochastic schemes.
4) The CGEFS-S is able to catch up to the ENSO prediction (or SST trend). It could be much improved if the forecast bias could be removed.
5) The global SST mean error of CGEFS-S is near perfect. The RMS error and spread are very comparable on the global domain.
6) The SST mean error (or bias) of the Niño 3.4 domain is different from the global domain. It has a cold bias for summer, and turns warm for winter time.
7) The ensemble spread of the SST anomaly for Niño 3.4 is over-dispersive (or much larger than climate variability and forecast RMS error). The forecast distribution is non-Gaussian.
8) The ensemble spread of the SST anomaly for the Niño 3.4 domain is reduced when both atmosphere and ocean stochastics are turned off. However, the reasons are still under investigation.

Correspondence to: Yuejian Zhu, NOAA/NWS/NCEP/EMC, College Park, Maryland;
yuejian.zhu@noaa.gov

6. Acknowledgements

The authors would like to thank all of the EMC ensemble team members, model group and coupling group.

7. Data/Code Availability Statement

The climatology and analysis are from NCEP-NCAR Reanalysis (Kalnay et al., 1996; https://web.archive.org/web/20100527091106/http://nomad3.ncep.noaa.gov/ncep_data/). The reforecast and real-time operation data could be accessed https://noaa-gefs-retrospective.s3.amazonaws.com/index.html. The Operational Sea Surface Temperature and Ice Analysis: documentation (OSTIA) and data can be accessed through UKMet office website at https://ghrsst-pp.metoffice.gov.uk/ostia-website/index.html

8. References

1. Gottschalck, J., and Coauthors, 2010: A Framework For assessing operational Madden-Julian Oscillation Forecasts: A CLIVAR MJO Working Group Project. *Bull. Amer. Meteor. Soc.*, **91,** 1247-1258, https://doi.org/10.1175/2010BAMS2816.1.
2. Guan H., Y. Zhu, E. Sinsky, W. Li, X. Zhou, D. Hou, C. Melhauser, and R. Wobus., 2020: Systematic Error Analysis and Calibration of 2-m Temperature for the NCEP GEFS Reforecast of the Subseasonal Experiment (SubX) Project. Wea. Forecasting, **34**, 361-376, https://doi.org/10.1175/WAF-D-18-0100.1
3. Guan, and Coauthors, 2022: The Gefsv12 Reforecast Dataset for Supporting Subseasonal And Hydrometeorological Applications, Mon. Wea. Rev., **150** (3), 647-665, https://doi.org/10.1175/MWR-D-21-0245.1
4. Li, W., Y. Zhu, X. Zhou, D. Hou, E. Sinsky, C. Melhauser, M. Pena, H. Guan and R. Wobus, 2020: Evaluating the MJO Forecast Skill from Different Configurations of NCEP GEFS Extended Forecast. Clim Dyn **52**, 4923–4936. https://doi.org/10.1007/s00382-018-4423-9
5. Saha, S., S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, et al. 2014: The NCEP climate forecast system version 2. J. Climate, **27**(6), 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1
6. Stefanova and co-authors, 2022: Description and Results from UFS-Coupled Prototypes for future Global, Ensemble and Seasonal Forecasts at NCEP, NCEP Office Notes 510. https://doi.org/10.15923/knxm-kz26
7. Wheeler, M. C., and H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. Mon. Wea. Rev., **132**, 1917–1932. https://doi.org/10.1175/1520-0493(2004)132%3C1917:AARMMI%3E2.0.CO;2
8. Zhou, X., Y. Zhu, D. Hou, and D. Kleist 2016: Comparison of the Ensemble Transform and the Ensemble Kalman Filter in the NCEP Global Ensemble Forecast System. Wea. Forecasting, **31**, 2058-2074. https://doi.org/10.1175/WAF-D-16-0109.1
9. Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng and D. Wobus, 2017: Performance of the New NCEP Global Ensemble Forecast System in a Parallel Experiment. Wea. Forecasting, **32**, 1989-2004. https://doi.org/10.1175/WAF-D-17-0023.1
10. Zhou, X., Y. Zhu, B. Fu, D. Hou, J. Peng, Y. Luo and W. Li, 2019: The development of the next NCEP Global Ensemble Forecast System. STI Climate Bulletin. *https://www.nws.noaa.gov/ost/climate/STIP/43CDPW/43cdpw-XZhou.pdf*
11. Zhou and co-authors, 2022: The Introduction of the NCEP Global Ensemble Forecast System version 12, Wea. Forecasting, https://doi.org/10.1175/WAF-D-21-0112.1

Correspondence to: Yuejian Zhu, NOAA/NWS/NCEP/EMC, College Park, Maryland; yuejian.zhu@noaa.gov

12. Zhu and co-authors, 2017: Impact of Sea Surface Temperature Forcing on Weeks 3 and 4 Forecast Skill in the NCEP Global Ensemble Forecasting System. Wea. Forecasting, **32,** 2159-2174. https://doi.org/10.1175/WAF-D-17-0093.1
13. Zhu and co-authors, 2018: Toward the Improvement of Subseasonal Prediction in the National Centers for Environmental Prediction Global Ensemble Forecast System, JGR Atmosphere, **123** 6732-6745. https://doi.org/10.1029/2018JD028506
14. Zhu and co-authors, 2018: Towards the Improvement of Sub-Seasonal Prediction in the NCEP Global Ensemble Forecast System (GEFS). STI Climate Bulletin, https://doi.org/10.7289/V5/CDPW-NWS-42nd-2018
15. Zhu, Y., W. Li, X. Zhou, and D. Hou, 2019: Stochastic Representation of NCEP GEFS to Improve Subseasonal Forecast. In the book of Current trends in the Representation of Physical Processes in Weather and Climate Models, Editors: Dr. Randells et al; Springer Atmospheric Science, 317-328
16. Zhu and co-authors, 2023: Quantify the Coupled GEFS Forecast Uncertainty for the Weather and Subseasonal Prediction. JGR Atmosphere, **128** 1-19. https://doi.org/10.1029/2022JD037757

Correspondence to: Yuejian Zhu, NOAA/NWS/NCEP/EMC, College Park, Maryland; yuejian.zhu@noaa.gov

Machine Learning to Construct Probabilistic Subseasonal Precipitation Forecasting over California

[1,2]Nachiketa Acharya and [1,2]Kyle Hall

[1]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Boulder, CO, United States,

[2]National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Laboratory, Boulder, CO, United States

ABSTRACT

1. Introduction

Following prolonged droughts in California during the years of 2012–16, severe dry and warm conditions re-emerged in 2019. The years 2020 and 2021 were marked as the fifth and second driest years in the past 100 years, further stressing the state's already stressed water resources. Considering the challenges presented by these drought extremes, accurate prediction of precipitation in the Sub-Seasonal (S2S) scale becomes particularly important for disaster preparation and mitigation as well as for drought and water restriction policy-making purposes. Although S2S climate predictions have a comparative lack of skill beyond two-week lead times, there has been a substantial research effort to improve prediction skill via novel advanced statistical and Artificial Intelligence/Machine Learning (AI/ML) methods over the past decade, to include post-processing of the dynamical model output. The large size of S2S datasets, unfortunately, makes traditional ML approaches computationally expensive, and therefore mostly inaccessible to those without access to institutional computing resources. To address this problem, we proposed a novel ML method called Probabilistic Output Extreme Learning Machine (PO-ELM) to post-process the outputs of European Centre for Medium-Range Weather Forecasts (ECMWF) S2S forecast for weeks 3-4 over California during Dec-Jan-Feb.

2. Data and Methods

PO-ELM is a modified version of Extreme Learning Machine (ELM), which is a generalized single-hidden-layer feed- forward neural network that randomly assigns input weights and biases in the hidden layer and calculates the output weights to the output layer using a generalized inverse (Acharya et al, 2014). However, since the traditional ELM network only produces a deterministic outcome, PO-ELM modifies the original ELM model such that it optimizes a sigmoid objective function and produces bounded continuous output on the interval (0, 1) to make probabilistic predictions. The schematic structure of PO-ELM is shown in Figure 1 (Acharya and Hall 2023). The detailed mathematical description and implementation procedure of the PO-ELM approach can be found in Acharya and Hall, (2023). The ECMWF 2001-2022 reforecast initializations for Mondays and Thursdays in the months Dec-Jan-Feb are calibrated and verified against the daily accumulated precipitation dataset from the Santa Barbara Climate Hazards Group InfraRed Precipitation with Station data (Funk et al. 2015). This data is at the 0.25-degree spatial resolution, which allows for good spatial coverage of the entire state of California. XCast, a high-performance Python data science toolkit for climate forecasting, designed by Hall and Acharya (2022), is used to implement the PO-ELM.

Correspondence to: Nachiketa Acharya, Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Boulder, CO; nachiketa.acharya@noaa.gov

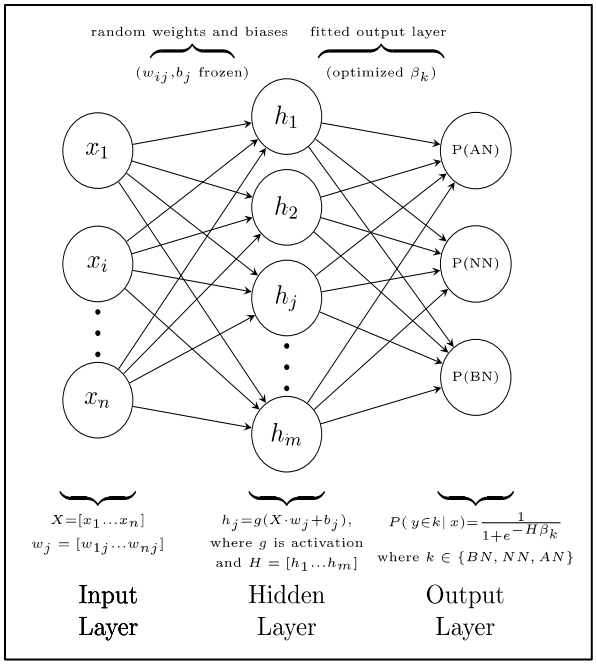

Figure 32: Schematic demonstration of the network of the Probabilistic Output Extreme Learning Machine (PO-ELM) (Acharya and Hall 2023; CC BY-NC 4.0).

3. Results and Discussions

The skill of the proposed methodology is demonstrated by the Generalized Receiver Operating Characteristic scores (GROC) and Rank Probability Skill scores (RPSS) in a leave-one out cross-validation manner. The resulting calibrated forecasts are successfully discriminated between events of occurrences and non-occurrences in terms of GROC score. However, the value of RPSS is not too high and is characterized by low reliability and sharpness. There is further room for improvement in skill by doing hyperparameter tuning of PO-ELM.

Correspondence to: Nachiketa Acharya, Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Boulder, CO; nachiketa.acharya@noaa.gov

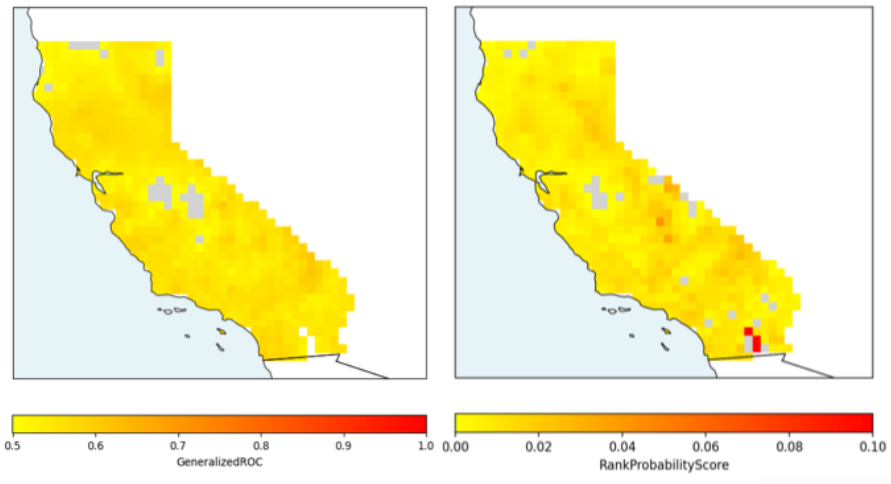

Figure 33: Generalized Receiver Operating Characteristic scores (left) and Rank Probability Skill scores (right) of PO-ELM during 2001-2022 for weeks 3-4 over California during Dec-Jan-Feb.

4. Conclusion

This study focuses on developing machine learning based post-processing to construct probabilistic S2S precipitation forecasting over California. The skill in weeks 3-4 is moderate. More skill can be found through hyperparameter tuning.

5. Data Availability Statement

The ECMWF reforecasts can be found in S2S data portal by ECMWF (https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/) and CHIRPS data can be found in data base of Climate Hazards Center UC Santa Barbara (https://www.chc.ucsb.edu/data/chirps).

6. References

1. Acharya N., and K. Hall, 2023: A Machine Learning Approach for Probabilistic Multi-Model Ensemble Predictions of Indian Summer Monsoon Rainfall, *MAUSAM*, **74** (2), 421–428. https://doi.org/10.54302/mausam.v74i2.5997

2. Acharya N., Srivastava N.A., Panigrahi B.K. and Mohanty U.C.,2014: Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine, *Climate Dynamics*, **43**(5), 1303-1310, https://doi.org/10.1007/s00382-013-1942-2

Correspondence to: Nachiketa Acharya, Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Boulder, CO; nachiketa.acharya@noaa.gov

3. Funk, C.C., and Coauthors., 2015: The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. *Scientific. Data*. **2** (150066), https://doi.org/10.1038/sdata.2015.66.

4. Hall, K., and N. Acharya,2022: XCast: A python climate forecasting toolkit, *Frontiers in Climate*, 4. https://doi.org/10.3389/fclim.2022.953262

5. Huang GB, M/B. Li, L. Chen, C.K. Siew, 2008: Incremental extreme learning machine with fully complex hidden nodes. *Neurocomputing* **71**(4-6), 576–583, https://doi.org/10.1016/j.neucom.2007.07.025.

Correspondence to: Nachiketa Acharya, Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Boulder, CO; nachiketa.acharya@noaa.gov

Post-Processing of Week 2 GEFSv12 Heat Forecasts via Neural Nets

Gregory Jennrich[1], Evan Oswald[1], Matt Rosencrans[2], and Li Xu[1]
[1]Earth Resources Technology, Inc., Laurel, MD.
[2]NOAA Climate Prediction Center, College Park, MD

ABSTRACT

1. Introduction

Hazardous heat has numerous societal impacts, yet very few forecasts explicitly for extreme heat extend beyond one week. Challenges such as relatively lower sample sizes to work with, greater sensitivity to differences in statistical distribution, along with a general focus on daily average temperatures, have thus far hampered skillful forecasting of Week 2 maximum temperature (Tmax) and maximum heat index (HI). In addition, hazardous heat is often forecast in months (April to October) with relatively low forecast skill on subseasonal timescales. Current Climate Prediction Center (CPC) efforts in this realm are largely based on GEFSv12 (Global Ensemble Forecast System version 12) and European Centre for Medium-Range Weather Forecasting (ECMWF) forecasts.

There are two types of extremely warm air temperatures: relative to historical (a function of location, daily summary variable, and time of year) and in an absolute sense (e.g. 100 degrees Fahrenheit). The former type of extreme heat forecasts does not need worry about the systematic forecast biases, especially if calibrated. The latter type of extreme heat forecasting would struggle to make skillful forecasts outside of the presence of strong seasonality. Absolute thresholds are only regionally relevant.

At CPC and in regards to extreme heat forecasting in an absolute sense, the current practice is distribution mapping/quantile mapping. Preliminary examination of this method indicated conspicuously large spatial pattern dissimilarity with raw forecasts and had relatively large computational resource needs. Ultimately, only a fairly simple bias correction has been retained over the long term. This process calculates and subsequently adjusts the weekly bias in the mean and the sub-weekly standard deviation of forecasted daily summaries (temperatures). Currently, this bias correction process is employed for weeks 3-4 but not week 2, primarily because these corrections are conspicuous and without stronger historical evidence of their benefit, the forecast team felt uneasy about them. Conversely, without employing these corrections for the week 3-4 period, the probabilities of many forecast metrics (e.g. probability of being over 90 degrees Fahrenheit) look completely inaccurate. This work seeks to fill some of the gap left by employing more advanced statistical corrections, including Machine Learning (ML) techniques as ML techniques are quickly being incorporated to enhance or replace more classical post-processing methods and may provide an advantage in these heat forecasts.

2. Data and Methods

In this study we explore the utility of ML for GEFSv12 heat forecasts. GEFSv12, week-2, 2-meter Tmax reforecasts from 2000-2020 (Guan et al. 2022) and the NCEP/NCAR Reanalysis (Kalnay et al. 1996) (here referred to as CDAS – Climate Data Assimilation System) 2-meter temperature are used to train a 2-layer neural network (NN). The NN is constructed using the TensorFlow software package (Abadi et al. 2015). More specifically, we use the Week-2 average maximum temperature as the target output. We do the same for a heat index calculated from the CDAS dataset. Data from ECMWF Reanalysis v5 (ERA5) (Hersbach et al. 2020), with analogous variables, time periods, and temporal resolutions were also used as validation, after identifying weaknesses in the CDAS analysis. Daily initialization dates from April 9 to

Correspondence to: Matt Rosencrans, NOAA/NWS Climate Prediction Center;
matthew.rosencrans@noaa.gov

September 20 were chosen in order to focus on the hazardous heat over the Continental United States (CONUS). The initial training of the neural networks used data from 2000-2020, with the validation set being 4 random years dropped out of the training dataset. Final testing used model and observational data from 2021 for a truly independent sample.

3. Analysis

As shown in Figure 1, the neural net uses each individual ensemble member (5 total) at forecast days 8-14 as an input. We use mean square error (MSE) as a loss function for the ML training. A comparison of the skill of the ML method and with both the raw GEFSv12 and bias corrected output is performed using the average root-mean-square error (RMSE) and MSE over the CONUS and surrounding North American points.

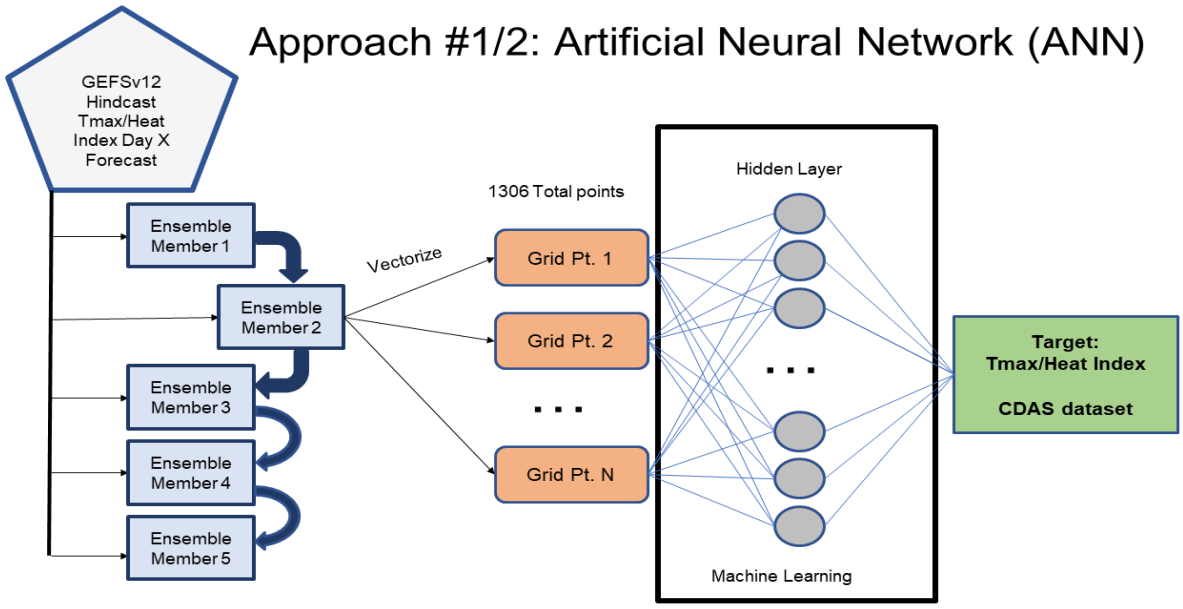

Figure 34: Artificial Neural Net Flowchart that depicts how the inputs are used to train the ML model. The individual ensemble members are the sole inputs for this model. Each member is vectorized into a 1D array of grid points and used as input. The actual machine learning portion is depicted by the black box, where a hidden layer contains weights that are adjusted for each epoch (training run). The CDAS data is the target dataset that the ML model is trying to predict.

We find the RMSE of the individual ensemble members of NN corrected Tmax and HI forecasts from our verification dataset to be around 3 degrees Fahrenheit less than from the original GEFSv12 model for our verification period (4 years). Using a simple linear bias correction on the same forecasts/observation pairs, the improvement was only around 1.5 degrees Fahrenheit for each of the variables. To test the model with real-time data, the summer of 2021 serves as the test dataset of the model. While this is a small sample size, we can investigate the utility of this model with a recent period, and a 124-ensemble member real-time model. In Figure 2, the skill maps of bias (plotted MSE) and RMSE from the 2021 summer period show improvement over the GEFSv12 forecasts. The average mean absolute error is noted in the bias plots to show the differences in bias. The NN correction average bias is a 50% reduction from the raw GEFSv12. Of note, the large warm bias in the Southeast U.S. is removed, along with much of the topographical bias and error in the western U.S. The NN corrected map has an RMSE reduction of about 1.5 degrees Fahrenheit and is particularly lowered in the Eastern U.S. Heat index skill scoring shows similar NN corrected improvements over the raw GESFv12 (not shown).

Correspondence to: Matt Rosencrans, NOAA/NWS Climate Prediction Center;
matthew.rosencrans@noaa.gov

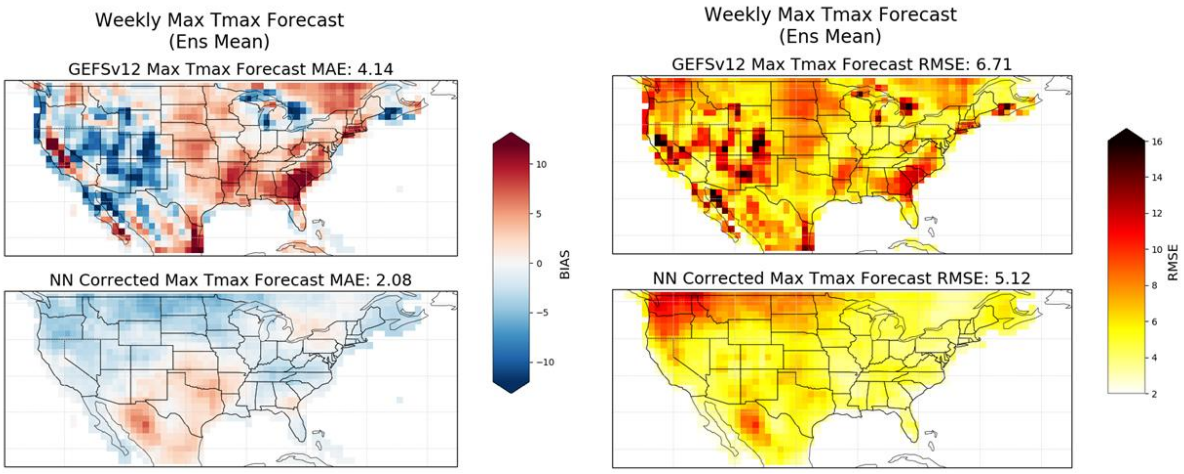

Figure 35: Skill score maps for the 2021 real-time period. The average MSE (right) for 2021 is shown for the NN corrected forecasts (bottom) and for the raw GEFSv12 (top). The spatial average mean absolute error (MAE) is noted on the top of each subfigure. The average RMSE (left) for 2021 is shown for the NN corrected forecasts (bottom) and the raw GESFv12 (top), with the spatial average RMSE noted on the top of each subfigure. Data Sources: GEFSv12 and Neural Net Corrected GEFSv12.

These results are encouraging on their face, but deeper examination illuminated some potential issues with this bias correction method. First, from the bias map in Figure 2, it is clear that there is a bit of a cool bias overall. Outside of the southern plains, most of the CONUS has a cool bias in the NN corrected forecasts. Second, this cool bias is exacerbated when considering the hottest observed days (figure 3), which are the goal of these forecast tools. The NN-based bias correction method often substantially reduces large anomalies, thereby imparting a cool bias on specifically warm days, as errors are reduced on days with typical air temperatures but less-so on days with notably hot air temperatures. Further examination of this mechanism of error, and the potential methods to minimize it, should be undertaken before operationally using this method. Subsequently, we attempted other methods, one of which included 1) targeting the final operational product rather than the equivalent CDAS variable and another, (2) which was a multi-layer model (also known as a deep-learning neural net (DLNN).

Correspondence to: Matt Rosencrans, NOAA/NWS Climate Prediction Center;
matthew.rosencrans@noaa.gov

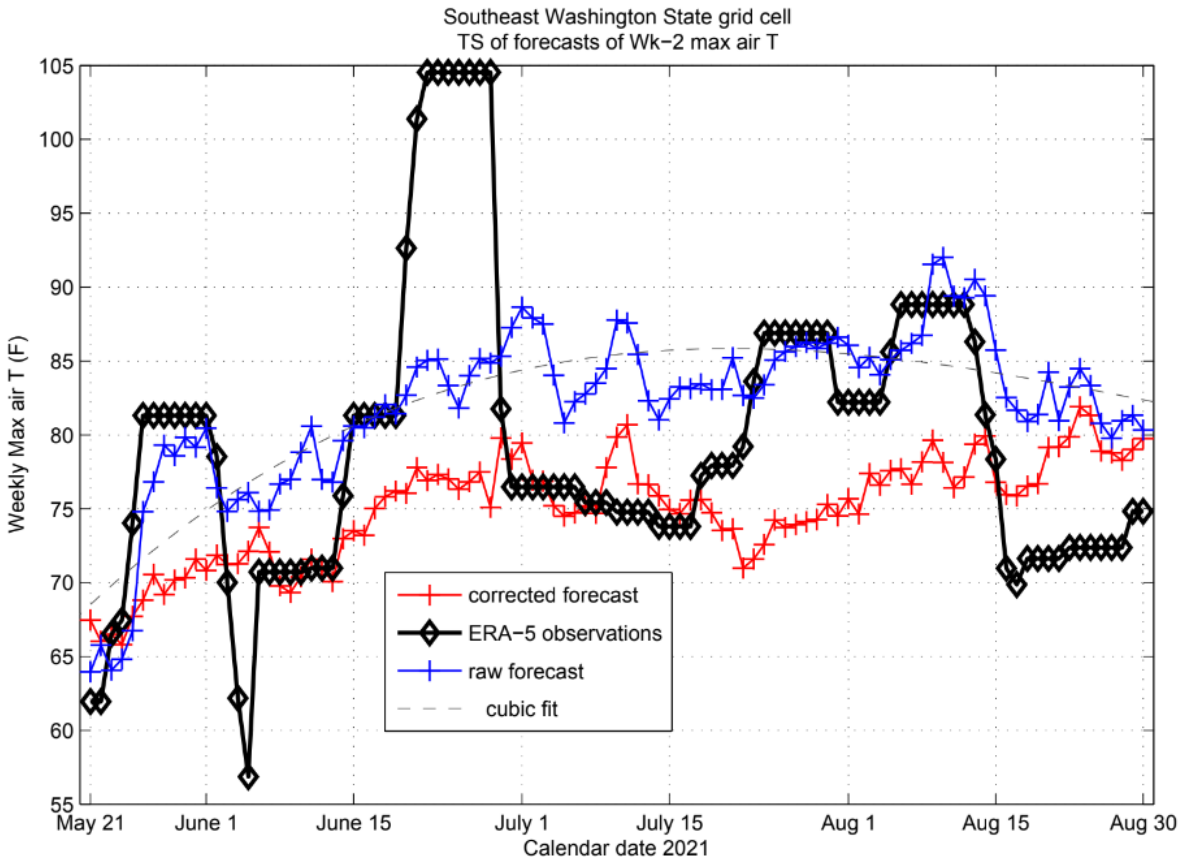

Figure 36: Time series of weekly maximum air temperature during the 2021 for the grid cell containing Sacramento, CA. Observations are from the ERA-5 reanalysis, along with raw and NN corrected forecasts. Data Sources: ECMWF Reanalysis v5, Neural Net Corrected GEFSv12, GEFSv12.

The aforementioned efforts operated on the daily temperature values (maximums) within the forecast period were subsequently used to create forecast metrics relevant to forecasting extremely hot weather. An example of such a forecast metric is the probabilities of reaching 100 degrees Fahrenheit during the forecast period. We examined how NN algorithms would impact the forecasts of these metrics directly, instead of applying the NN algorithm on the individual dates and subsequently calculating the forecast metrics. These attempts were more fruitful than the daily-scale counterparts, albeit potentially more difficult to implement in practice. Separate training of each model for each of the forecast metrics for extreme heat in an absolute sense would be required.

The AUC (Area Under the Curve) ROC (Receiver Operating Characteristics) (AUC-ROC) score for the 2021 testing year improved by 10% (Fig. 4) when this method was employed, as compared to creating probabilities directly from the bias corrected GEFSv12 reforecast. The results were not as straightforward because the area that is non-zero (probabilities of reaching 100 degrees Fahrenheit) changes somewhat.

Correspondence to: Matt Rosencrans, NOAA/NWS Climate Prediction Center;
matthew.rosencrans@noaa.gov

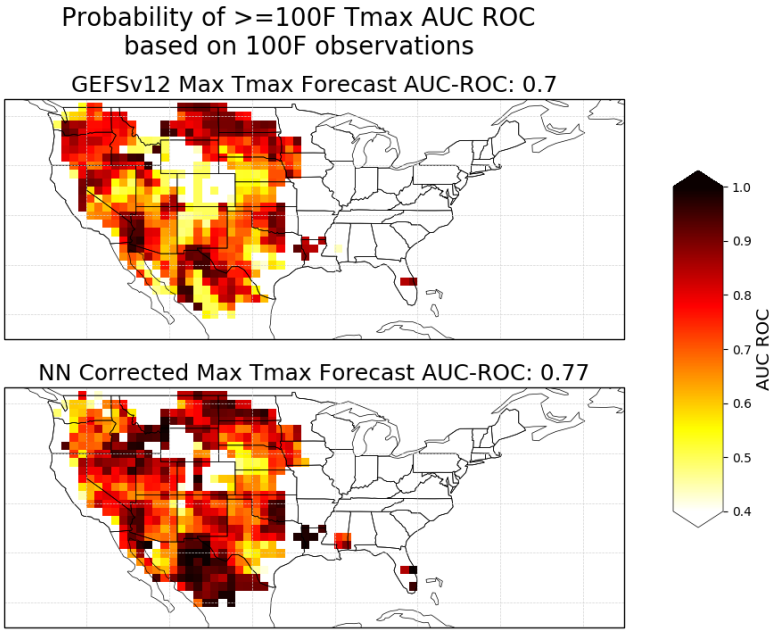

Figure 37: Maps of Area Under the Curve Receiver Operating Characteristics (AUC-ROC) curve for the 2021 real-time period. The AUC-ROC for 2021 is shown for the NN corrected forecasts (bottom) and for the raw GEFSv12 (top). The spatial average AUC-ROC is noted on the top of each figure panel. Data Source: NN corrected GEFSv12.

To further explore the datasets and potential of ML methods, the group also tried to employ a DLNN. DLNN can create highly specialized, non-linear relationships between multiple input datasets and target datasets. The group tried to enhance the information available to the DLNN by also including input point metadata (latitude, longitude, elevation, and day of year), soil moisture, sea surface temperature, and vegetation fraction. The DLNN provided some positive benefit in predicting the right-tail of the distribution (Fig. 4).

Correspondence to: Matt Rosencrans, NOAA/NWS Climate Prediction Center;
matthew.rosencrans@noaa.gov

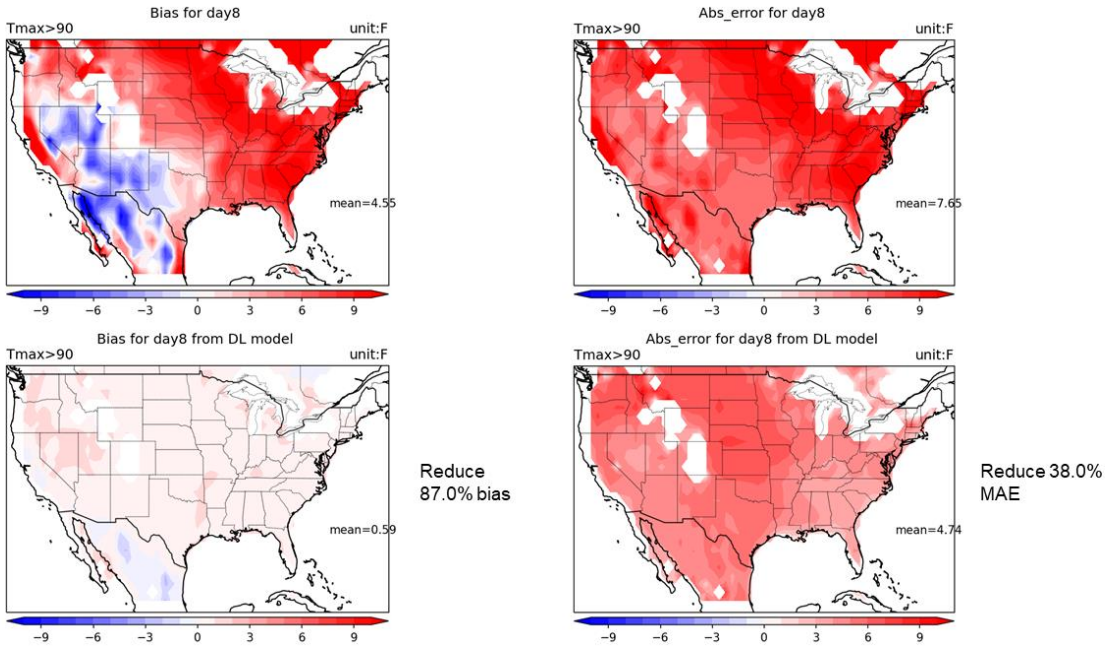

Figure 38: Maps of the bias (left column) and absolute error (right column) as derived from GEFSv12 reforecast (top row) and deep-learning neural network model (bottom row) for all of the forecasts with lead time of 8 days in the validation period (April 9 to September 20, 2017-2020). Data Source: GEFSv1.2

4. Conclusion

The use of artificial neural networks (ANN) show promise as potential solutions for increasing our ability to forecast extreme and hazardous heat. Feed-forward neural networks with 1 hidden layer may be complex enough to further our abilities, but more complex neural networks are likely to provide further benefit. Additionally, the quality of the dataset used for training and validation became a hindrance, so the group recommended the development of an observational based dataset specifically suited to the monitoring of extreme/hazardous heat.

Future work may include closer examination of the NN model that operated directly on daily temperature values, but specifically during times of high anomaly. Explicitly including anomaly value in the NN model as a second variable may be useful to better predict extreme heat. The NN correction models that operate on the forecast metrics directly may be expanded to 1) operate on other thresholds (90 and/or 110 degrees Fahrenheit, for example) or 2) work on the deterministic forecast metrics. Lastly our work on the Deep-Learning Model (DLM) demonstrated promise, and could be replicated with the multiple linear regression model for extreme heat that currently exists for the weeks 3-4 time period. We feel these preliminary results represent a step in understanding the utility of NN corrections in long-range heat forecasting.

5. Data Availability Statement

GEFSv12 reforecast and forecast data are freely available via NOAA's Open Data Dissemination (https://www.noaa.gov/information-technology/open-data-dissemination). NCEP/NCAR reanalysis data is available on CPC's FTP (https://ftp.cpc.ncep.noaa.gov/wd51we/reanalysis-1/6hr/rotating/).

Correspondence to: Matt Rosencrans, NOAA/NWS Climate Prediction Center;
matthew.rosencrans@noaa.gov

6. References

1. Abadi M. and Coauthors, 2015: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16),* Savannah, GA, USENIX Assoc. Software available from tensorflow.org. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf.
2. Guan, H., and Coauthors, 2022: GEFSv12 Reforecast Dataset for Supporting Subseasonal and Hydrometeorological Applications. Mon. Wea. Rev., **150**(3), 647–665, https://doi.org/10.1175/MWR-D-21-0245.1.
3. Hersbach, H and Coauthors, 2020: The ERA5 global reanalysis. Q J R Meteorol Soc. 2020; **146**(730), 1999– 2049, https://doi.org/10.1002/qj.3803.
4. Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**(3), 437–472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Correspondence to: Matt Rosencrans, NOAA/NWS Climate Prediction Center; matthew.rosencrans@noaa.gov

Using Seasonal Forecast Ensembles to Support Harmful Algal Bloom Forecasts at the National Weather Service Ohio River Forecast Center

Abram DaSilva
National Weather Service (OHRFC), Wilmington, OH

ABSTRACT

1. Introduction

Harmful algal blooms (HABs) are a growing environmental problem in the Great Lakes, and nowhere are these occurrences more problematic than Lake Erie. Each year, these HABs threaten millions of people living along or visiting the coast of Lake Erie. Toxins produced by these cyanobacterial blooms can be harmful to fish, wildlife, and humans, leading to significant impacts for both ecosystems and local economies. Economies dependent upon tourism can further be harmed when swimmers, boaters, and beachgoers are kept away by the foul-smelling green scum (Stumpf et al., 2016). To provide coastal managers and the public with an idea of what to expect for the coming HAB season in mid- to late-summer, NOAA's National Centers for Coastal Ocean Science (NCCOS) provides weekly projections of HAB severity, beginning in May and continuing through June, before releasing an official Lake Erie HAB seasonal forecast in early July. The projections for HAB severity use an index to estimate the bloom's maximum biomass over the peak 30-days in a season (Figure 1).

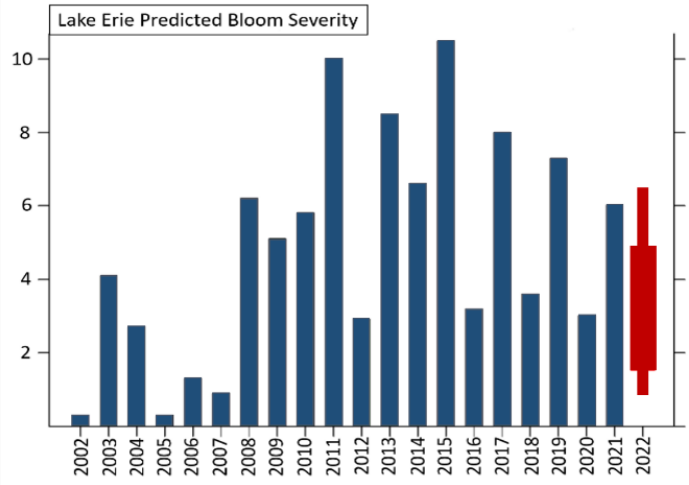

Figure 39: The Western Lake Erie Harmful Algal Bloom severity projection from June 14th, 2022. This chart shows the 2022 projected bloom intensity as compared to previous years. The wide 2022 bar is the likely range of severity (based on data to June 13th, 2022), while the narrow bar for 2022 shows the potential range of severity. Figure adapted from Stumpf et al., 2022.

To support these efforts, the NWS Ohio River Forecast Center (OHRFC) began providing daily Climate Forecasting System (CFS) 60-day flow forecasts for the Maumee River in 2016. These CFS flow forecasts are used to improve predictions of phosphorus loading, which is a key driver of HABs. The Maumee River is the primary input source of phosphorus to the western Lake Erie basin. The Maumee River watershed covers approximately 6,290 mi2 of Indiana, Michigan, and Ohio, and nearly 75% of it is

Correspondence to: Abram DaSilva, National Weather Service, Ohio River Forecast Center, Wilmington, OH; Email: abram.dasilva@noaa.gov

used for agriculture (Dewitz, 2021). The CFS flow forecasts are for the Maumee River at Waterville, OH, which is the approximate terminus for the river as it flows into the western Lake Erie basin (Figure 2).

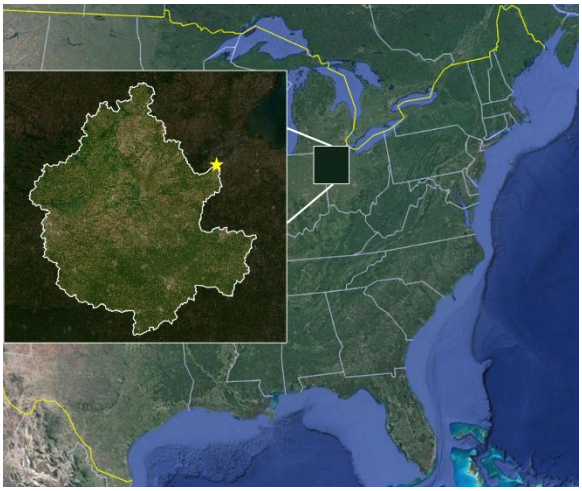

Figure 40: The Maumee River watershed drains parts of Indiana, Michigan, and Ohio, before emptying into the western Lake Erie basin just downstream of Waterville, Ohio (yellow star). The CFS forecasts were produced by the National Weather Service Ohio River Forecast Center, and the observed data was obtained from the US Geological Survey.

From 2016 until 2021, NCCOS and the OHRFC evaluated the benefit of using CFS forecasts in the HAB severity predictions by comparing severity forecasts made using this new method to the old (climatology), finding that the forecasts made using the CFS were closer to the measured severity each year. In 2021, a more quantitative analysis was conducted to compare CFS flow forecasts to observations to ensure NCCOS was using the daily forecasts in the best way possible, leading to the identification of a clear best use. This was followed by a more in-depth statistical analysis to assess the performance of the CFS flow forecast for the Maumee River. Finally, the Hydrologic Ensemble Forecasting System (HEFS), which provides bias-correction and a longer lead time (90+ days), was evaluated as a potential improvement over the CFS flow forecasts.

2. Data and Methods

Each daily flow forecast for the Maumee River at Waterville uses 43 days of rainfall, extends 60 days into the future, and includes 32 CFS ensembles along with the mean and median as well as chance of exceedance percentiles (5%, 10%, 25%, 75%, and 90%). Since NCCOS started receiving these daily forecasts in 2016, they have used the median ensemble trace for the phosphorus loading component of their HAB model, but until 2021 there had been no attempt to determine if this was the best way to predict the Maumee River flow. NCCOS has found that the total volume coming from the Maumee River in the late spring months is a primary driver of magnitude for the Lake Erie summer HABs.

The following three-part analysis was conducted to evaluate CFS forecast performance and investigate the potential for improving OHRFC support for NCCOS: 1) Five years (2017-2021) of US Geological Survey observed total monthly volume for the months of April and May were compared to monthly volumes provided by the 10% and 25% exceedance traces, the mean trace, and the median trace of the CFS forecast to determine the best ensemble trace for representing late spring total volume. 2) After identifying the best ensemble trace, the NWS National Water Center's Water Resource Evaluation

Correspondence to: Abram DaSilva, National Weather Service, Ohio River Forecast Center, Wilmington, OH; Email: abram.dasilva@noaa.gov

Service (WRES) was employed to run verification statistics (bias fraction and mean square error (MSE) skill score) on total daily volume of this CFS flow forecast trace to assess performance over the past few seasons (2020, 2021, and 2022). 3) Finally, the new HEFS flow forecasts were compared to the currently-provided CFS flow forecasts using WRES to investigate the potential for improving flow estimates for the Maumee River at Waterville, OH.

3. Analysis

The initial analysis compared total monthly volume predicted by the CFS flow traces (10% and 25% exceedance, mean, and median) for the Maumee River at Waterville, OH to the total monthly volume observed by the US Geological Survey (USGS) for April and May from 2017 - 2021, and gave ratios of predicted to observed for each (Table 1). A result of 1 would indicate predicted volume equals observed volume, >1 would indicate predicted volume is higher, and <1 would indicate a lower predicted volume.

Table 6: Ratio of predicted to observed total volume for April and May 2017 - 2021 for different traces of the CFS flow forecast for the Maumee River at Waterville, OH.

CFS Trace	10% Exceedance	25% Exceedance	Mean	Median
Total Volume (Predicted/Observed)	2.98	1.89	1.32	1.09

Of all the traces, the median resulted in better predictions of total April and May volume when looking at the time period 2017 - 2021. While 1.09 still indicates overprediction, the alternative CFS traces (10% exceedance, 25% exceedance, and the mean) all predicted even higher total volumes. Once the best trace was established, the second component of the analysis focused on verifying the performance of the CFS median flow forecast using bias fraction and MSE skill score.

Equation 1: WRES Bias Fraction.

$$Bias\ Fraction = \frac{\sum_{i=1}^{n}(predicted - observed)}{\sum_{i=1}^{n} observed}$$

Equation 2: MSE Forecast Skill Score, valuated against a baseline (either the observation average or another forecast for comparison). This metric returns values ranging between 1 and -infinity, with a value of zero indicating no skill over using the baseline, negative values meaning that using the baseline would be better than the forecast, and a value of 1 indicating perfect skill.

$$MSE\ Skill\ Score = 1 - \frac{MSE_{forecast}}{MSE_{baseline}}$$

The verification for the CFS median flow forecasts focused on the forecast total daily volumes compared to the USGS observed total daily volumes from March to September, which is the period of the year when flow is most impactful on the Lake Erie summertime HABs (Figure 3).

Correspondence to: Abram DaSilva, National Weather Service, Ohio River Forecast Center, Wilmington, OH; Email: abram.dasilva@noaa.gov

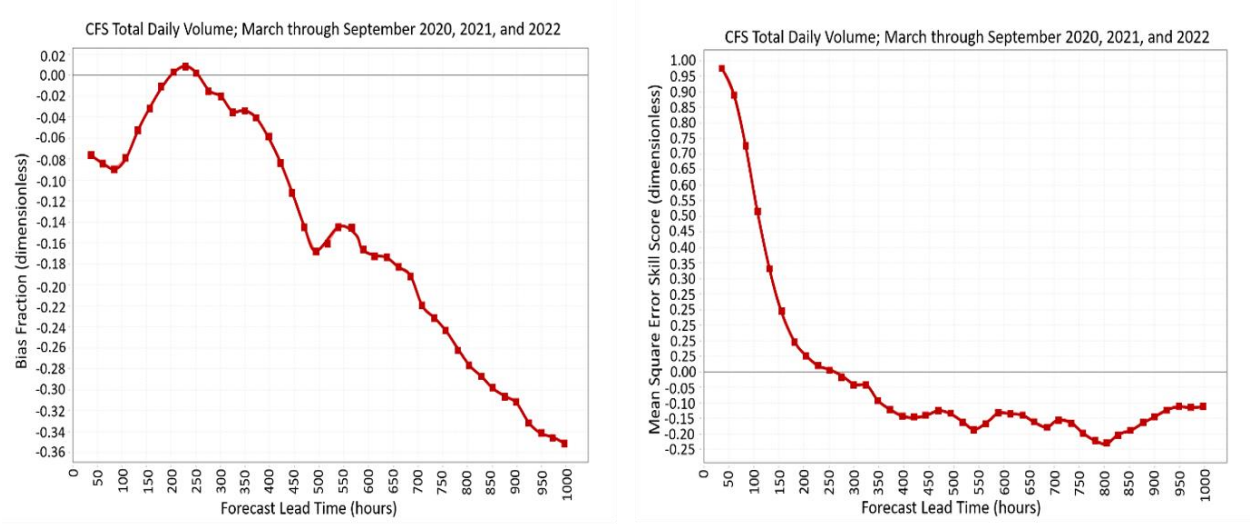

Figure 41: Bias fraction (left plot) and MSE skill score (right plot) for total daily volume given by the CFS median trace from March through September of 2020, 2021, and 2022. Pooled by forecast lead time (hours) to visualize performance change over time from forecast issuance. The CFS forecasts were produced by the National Weather Service Ohio River Forecast Center, and the observed data was obtained from the US Geological Survey.

The bias fraction results show the CFS median tracked closely with the observed total daily volume for most of the first three weeks of the forecast window (near neutral to ~10% underprediction until about 450 hours out from forecast issuance) before dropping to an underprediction of ~35% by six weeks. Results for the MSE skill score show the CFS median performed better than the average of observations for ~10 days out from forecast issuance, and after this point the average of the observations was better.

To evaluate the HEFS median as a potential replacement of the CFS median for NCCOS, we used the same metrics (bias fraction and MSE skill score; Figure 4) for total daily volume over the HAB seasons of 2021 and 2022, but in this case the CFS median was used as the baseline for MSE skill score to allow a direct comparison of HEFS median to the CFS median. Values greater than 0 indicate better performance by the HEFS median, values below 0 indicate the CFS median performed better when compared to observations.

Correspondence to: Abram DaSilva, National Weather Service, Ohio River Forecast Center, Wilmington, OH; Email: abram.dasilva@noaa.gov

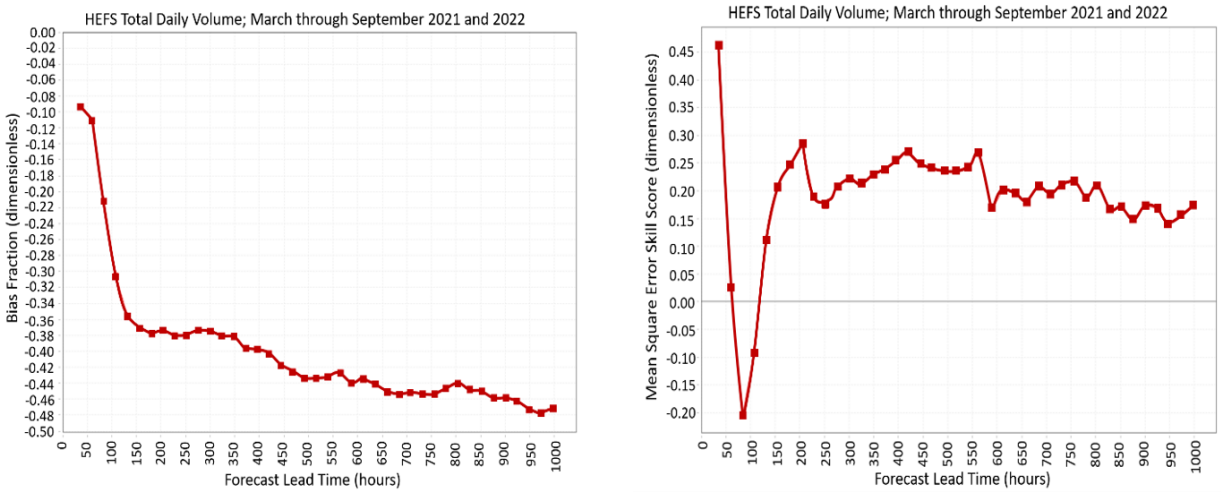

Figure 42: Bias fraction for the HEFS median trace (left plot) and MSE skill score for the HEFS median trace with a baseline of the CFS median trace (right plot) from March through September 2021 and 2022. Pooled by forecast lead time (hours) to visualize performance change over time from forecast issuance. The HEFS and CFS forecasts were produced by the National Weather Service Ohio River Forecast Center, and the observed data was obtained from the US Geological Survey.

The bias fraction results for the period of March through September 2021 and 2022 show the HEFS median consistently underpredicted flow, starting off about 10% below and moving to an underprediction of ~40% by one week out, before gradually moving lower to an underprediction of ~50% by six weeks out from forecast issuance. The results of the MSE skill score indicate that the HEFS median flow forecasts were more skillful than the CFS median flow forecasts for March-September 2021 and 2022 for all forecast lead times but two (84 and 108 hours). The dip down below zero represents this shift to the CFS performing better briefly before jumping back up to show HEFS had better skill for the remainder of forecast lead times.

4. Conclusion

Since 2016, the OHRFC has supported NCCOS in their efforts to provide the public with critical information on the Lake Erie HAB severity by providing daily CFS flow forecasts. The purpose of this analysis was to improve the utilization and interpretation of the CFS flow forecasts, and to explore the potential for replacing it with the HEFS forecasts, which could provide greater accuracy and longer lead time.

The initial investigation of how NCCOS could optimally use the CFS forecasts to predict phosphorus loading supported the current practice of using the CFS median, which yielded the best prediction of total volume from the Maumee River into Lake Erie.

The WRES verification results indicated that the CFS median consistently underpredicted the total daily volume and this trend became more pronounced as time from forecast issuance increased. Evaluating the HEFS median flow forecast as a potential replacement for the CFS showed promise, despite giving a greater underprediction of total daily volume. However, it is important to note that while the CFS evaluation included three years of data, the evaluation of HEFS included just two. The direct comparison (MSE skill score) showed that the HEFS median outperformed the CFS median in predicting total daily flow, but more data and further evaluation may be required before recommending a switch from the CFS median.

Correspondence to: Abram DaSilva, National Weather Service, Ohio River Forecast Center, Wilmington, OH; Email: abram.dasilva@noaa.gov

5. Data Availability Statement

The daily CFS and HEFS flow forecast ensembles used in this study were archived and compressed by forecast type and year (e.g., 2020_CFS.tgz, 2021_HEFS.tgz, etc.). These TGZ files are available upon request by contacting abram.dasilva@noaa.gov at the Ohio River Forecast Center. Observed Maumee River flows at Waterville, OH can be accessed at: https://waterdata.usgs.gov/monitoring-location/04193500.

6. References

1. Dewitz, J., 2021: National Land Cover Database (NLCD) 2019 Products Data set. U.S. Geological Survey. Accessed 01 March, 2023. https://doi.org/10.5066/P9KZCM54.
2. Google Earth, 2023. Accessed 01 March, 2023, https://earth.google.com/web/
3. Stroud Water Research Center, 2023: Model My Watershed. Accessed 01 March, 2023, https://modelmywatershed.org/.
4. Stumpf, Richard P., L.T. Johnson, T.T. Wynne, and D.B. Baker., 2016. Forecasting Annual Cyanobacterial Bloom Biomass to Inform Management Decisions in Lake Erie. Journal of Great Lakes Research, 42, 6, 1174–1183., https://doi.org/10.1016/j.jglr.2016.08.0

Correspondence to: Abram DaSilva, National Weather Service, Ohio River Forecast Center, Wilmington, OH; Email: abram.dasilva@noaa.gov

Understanding the Connection Between Soil Moisture and Safe Water Access Using Earth Observations

Farah Nusrat[1], Joseph Goodwill[1], Jason Parent[1], Kristin Johnson[1], Ashraful Islam Khan[2], Md. Taufiqul Islam[2], Firdausi Qadri[2], Ali Shafqat Akanda[1]

[1]University of Rhode Island, Rhode Island, USA,

[2]International Centre for Diarrhoeal Disease Research, Bangladesh

ABSTRACT

1. Introduction

More than 4 billion people across the world encounter acute water scarcity for at least one month each year (Mekonnen and Hoekstra 2016). This situation will worsen by 2040 when approximately 1 in every 4 children will live in extremely water-stressed areas, according to UNICEF (2017). The challenge further intensifies during times of natural calamities i.e., floods, drought, hurricanes, which make people vulnerable to waterborne diseases and sometimes lead to epidemic outbreaks. Water scarcity can be defined in terms of both quality and quantity; however, in general, the difference between water demand and availability can be described as water scarcity. According to Jiang (2009), lack of water availability to meet demand is referred to as quantity-based water scarcity, whereas poor water quality that cannot be used for economic purposes is referred to as quality-based water scarcity.

Although direct quantification of safe water access is challenging, hydroclimatic variables can help us understand the scarcity of total available water resources over a large geographic region and their impacts on water scarcity, water quality, and access to safe water and sanitation. Soil Moisture (SM), a measure of wetness in the topsoil, can be used as a proxy indicator to understand surface water availability. Low soil moisture can also be associated with available groundwater levels. Employing high spatiotemporal resolution remote sensing observations of soil moisture and other hydroclimatic variables, this study highlights the impact of different hydroclimatic processes on seasonal water availability in Bangladesh.

Based on diverse water availability and water management issues, Bangladesh is divided into seven terrestrial hydrological regions for this analysis, according to the National Water Management Plan by Water Resources Planning Organization (2001). This study takes a regional approach to analyze the connection between soil moisture and safe water access across the different hydrological settings in Bangladesh.

2. Data and Methods

Soil Moisture from NASA-USDA Enhanced SMAP Global Soil Moisture Data (Soil moisture profile), Land Surface Temperature (LST) from Moderate Resolution Imaging Spectroradiometer (MODIS), Rainfall from Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) and Liquid Water Equivalent thickness (LWE) from Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow-On (GRACE-FO) is used in this study. LWE is used as an estimation of groundwater. Maximum consecutive 5-day precipitation (Rx5day), Maximum daytime temperature (TXx) are used as a measurement of extreme events. For analysis, Spearman correlation is used to determine the intensity and direction of association between two variables as these datasets do not follow a normal distribution (Prion and Haerling 2014).

Correspondence to: Farah Nusrat, University of Rhode Island, Kingston, Rhode Island; farah_nusrat@uri.edu

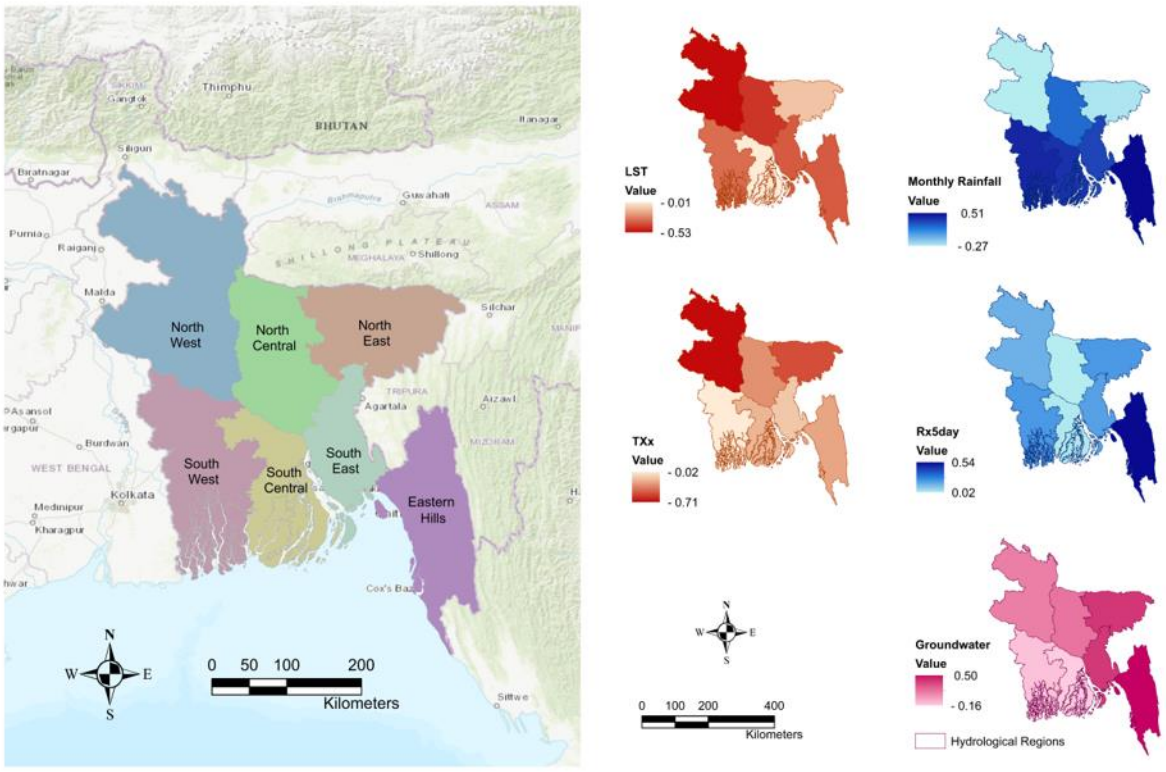

Figure 43: (a) Hydrological Regions of Bangladesh (b) Correlation between anomalies of soil moisture and anomalies of different hydroclimatic variables and extreme event variables during spring season. Data source: Water Resources Planning Organization.

3. Result and Discussion

In this study, we tried to identify hydroclimatic variables that are responsible for water scarcity using soil moisture as a proxy of water availability. According to Eltahir (1998), soil saturation plays a role in determining the hydraulic properties of soil, rate of evaporation, along with water availability and soil moisture is controlled by both the amount of rainfall and evaporation. A positive correlation is expected between rainfall and soil moisture as rainfall causes an increase in soil moisture (Li et al. 2016). Both 5 days of consecutive rainfall and monthly rainfall show a positive correlation with soil moisture (with few exceptions) implying that higher rainfall, after accumulating on the ground surface, increases the soil moisture. Rainfall intensity also plays a role in soil moisture patterns as the impact of single high-intensity rainfall and several small-intensity rainfall events will be different on soil moisture (Dai et al. 2022).

Land Surface Temperature (LST) and soil moisture share a negative correlation as an increase in LST causes more evapotranspiration, and that leads to less soil moisture (Seneviratne et al. 2010). Across all regions, a moderate negative correlation between LST and soil moisture is observed during the spring season. Also, maximum daytime temperature (TXx) shows a negative correlation with soil moisture as a high value of TXx causes a decrease in soil moisture.

Soil moisture can also offer terrestrial water storage information (surface water, groundwater) (Seneviratne et al. 2010). Groundwater and soil moisture typically follow similar patterns, low during winter and spring seasons and increases with the monsoon rainfall, so a positive association is expected between these two (Irfan et al. 2020). In our study, groundwater shows a positive correlation with soil

Correspondence to: Farah Nusrat, University of Rhode Island, Kingston, Rhode Island; farah_nusrat@uri.edu

moisture in all seven regions with few exceptions. During the water-scarce spring season, the eastern parts of the country (North East, South East, and Eastern Hills regions) show a higher positive correlation with groundwater than other regions.

4. Conclusion

As the behavior of soil moisture is associated with several other water sources, especially over large regions, it could be an indirect measurement of water availability and a useful inference to scarcity in water-stressed populations. Along with monthly rainfall, LST, and groundwater, representative extreme variables are also analyzed in this study to obtain a comprehensive overview of water scarcity in different hydrological regions. As water scarcity causes a shortage of safe water for drinking and sanitation purposes, which eventually contributes to higher diarrheal and other waterborne diseases, we hope this study will be beneficial to build understanding and resilience to hydroclimatic hazards and public health vulnerability.

5. Data Availability Statement

Using Google Earth Engine (GEE), MOD11A1.061 Terra Land Surface Temperature and Emissivity Daily Global 1 kilometer data are collected from https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD11A1 (Access Date 10/11/2022).

NASA-USDA Enhanced SMAP Global Soil Moisture Data are collected from https://developers.google.com/earth-engine/datasets/catalog/NASA_USDA_HSL_SMAP10KM_soil_moisture (Access Date 07/07/2022).

Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) data are collected from https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_daily/tifs/p05/ (Access Date 01/08/2020).

For GRACE and GRACE-FO data, a NetCDF file named "GRCTellus.JPL.200204_202206.GLO.RL06M.MSCNv02.nc4" is downloaded from https://grace.jpl.nasa.gov/data/get-data/jpl_global_mascons/ (Access Date 08/26/2022).

6. References

1. Dai, L., R. Fu, X. Guo, Y. Du, F. Zhang, and G. Cao, 2022: Soil Moisture Variations in Response to Precipitation Across Different Vegetation Types on the Northeastern Qinghai-Tibet Plateau. *Frontiers in Plant Science*, **13**, 854152, https://doi.org/10.3389%2Ffpls.2022.854152.
2. Eltahir, E. A. B., 1998: A soil moisture–rainfall feedback mechanism: 1. Theory and observations. *Water Resources Research*, **34**, 4, 765–776, https://doi.org/10.1029/97WR03499.
3. Funk, C.C., Peterson, P.J., Landsfeld, M.F., Pedreros, D.H., Verdin, J.P., Rowland, J.D., Romero, B.E., Husak, G.J., Michaelsen, J.C., and Verdin, A.P., 2014, A quasi-global precipitation time series for drought monitoring: U.S. Geological Survey Data Series 832, 4 p. http://pubs.usgs.gov/ds/832/
4. Irfan, M., N. Kurniawati, M. Ariani, A. Sulaiman, and I. Iskandar, 2020: Study of groundwater level and its correlation to soil moisture on peatlands in South Sumatra. *Journal of Physics: Conference Series*, **1568**, 12028, https://doi.org/10.1088/1742-6596/1568/1/012028.
5. Jiang, Y., 2009: China's water scarcity. *Journal of Environmental Management*, **90**, 11, 3185–3196, https://doi.org/10.1016/j.jenvman.2009.04.016.
6. Li, B., L. Wang, K. F. Kaseke, L. Li, and M. K. Seely, 2016: The impact of rainfall on soil moisture dynamics in a foggy desert. *PLoS One*, **11**, 10, e0164982, https://doi.org/10.1371/journal.pone.0164982.

Correspondence to: Farah Nusrat, University of Rhode Island, Kingston, Rhode Island; farah_nusrat@uri.edu

7. Mekonnen, M. M., and A. Y. Hoekstra, 2016: Four billion people facing severe water scarcity. *Science Advances*, **2**, 2, e1500323, https://doi.org/10.1126/sciadv.1500323.
8. Prion, S., and K. A. Haerling, 2014: Making sense of methods and measurement: Spearman-rho ranked-order correlation coefficient. *Clinical Simulation in Nursing*, **10**, 10, 535–536, https://doi.org/10.1016/j.ecns.2014.07.005.
9. Sazib, N., I. Mladenoca, and J. Bolten, 2018: Leveraging the Google Earth Engine for drought assessment using global soil moisture data. *Remote sensing*, **10** (8), 1265, https://doi.org/10.3390/rs10081265.
10. Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, A. J. Teuling, 2010: Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, **99**, 3–4, 125–161, https://doi.org/10.1016/j.earscirev.2010.02.004.
11. UNICEF, 2017: Nearly 600 million children will live in areas with extremely limited water resources by 2040 - UNICEF. Accessed 14 June 2023, https://www.unicef.org/press-releases/nearly-600-million-children-will-live-areas-extremely-limited-water-resources-2040.
12. Wan, Z., S. Hook, and G. Hully: 2021 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061. NASA EOSDIS Land Processes DAA. Accessed 11 October 2022, https://doi.org/10.5067/MODIS/MOD11A1.061.
13. Water Resources Planning Organization, 2001: National Water Management Plan. Main Report Volume 2, 212 pp, http://old.warpo.gov.bd/nwmp/nwmp_vol2.pdf.
14. Wiese, D.N., D.-N. Yuan, C. Boening, F.W. Landerer, M.M. Watkins, 2023: JPL GRACE and GRACE-FO Mascon Ocean, Ice, and Hydrology Equivalent Water Height CRI Filtered. Ver. RL06.1Mv03. PO.DAAC, CA, USA. Accessed 26 August 2022, https://doi.org/10.5067/TEMSC-3JC63.

Correspondence to: Farah Nusrat, University of Rhode Island, Kingston, Rhode Island; farah_nusrat@uri.edu

Verification of a Drought Prediction System over the Horn of Africa

Miliaritiana Robjhon[1,2], Li Xu[1,2], Wassila Thiaw[1], and Yun Fan[1]
[1]Climate Prediction Center, NCEP/NWS/NOAA, College Park, MD
[2]ERT, Inc., Laurel, Maryland

ABSTRACT

1. Introduction

Unlike drought monitoring, which has seen consistent progress over the past few decades, drought prediction remains a challenge due to the complexity of the nature of drought, limited prediction skills at longer lead times, and a gap in forecasting rapid onset and intensification of drought. While many developed countries are equipped with established coping mechanisms, including irrigation and water management systems to mitigate the impacts of droughts, reliable and timely drought prediction plays a crucial role for rain-fed dependent and vulnerable regions such as the Horn of Africa, where frequent and recurring droughts can push millions of people to face and cope with food insecurity.

The main objective of this paper is to assess the skills of a novel drought prediction system over the Horn of Africa and verify its performance during the case study of the recent March-May 2022 drought, which has resulted in substantial agricultural losses in the sub-region (FAO 2022).

2. Data and Methods

The drought forecasting system has four components: the version 12 of the Global Ensemble Forecast System (GEFSv12), Bias Correction and Spatial Downscaling (BCSD) (Yoon *et al.* 2012), Leaky Bucket land surface model (Fan and van den Dool 2004), and the integration of meteorological, agricultural, and hydrological drought indices forecasts. Precipitation and temperatures output from the GEFSv12, which extend up to 35 days, were corrected first using BCSD to remove model systematic biases, then were used to force the Leaky Bucket hydrologic model to produce agricultural and hydrological drought indices forecasts, such as soil moisture percentile, Standardized Runoff Index (SRI), Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano *et al.* 2010), and Evaporative Demand Drought Index (EDDI) (Hobbins *et al.* 2016) forecasts up to week 5. Bias corrected precipitation was appended to the CPC Unified Precipitation Analysis (Chen *et al.* 2008) to compute the Standardized Precipitation Index (SPI) (McKee *et al.* 1993) forecasts, which were integrated with the other agricultural and hydrological indices forecasts, and all indices forecasts were converted to percentile distribution to construct an objective and blended drought index with a probabilistic format.

Prior to the coupling between GEFSv12 and the Leaky Bucket model, the effectiveness of the BCSD was evaluated over the Horn of Africa, which was delimited by 12 degrees South to 18 degrees North and 25 degrees East to 52 degrees East, using standard metrics such as the Heidke Skill Score (HSS), root mean squared error (RMSE), mean absolute error (MAE), mean error (ME), and the coefficient of determination.

The skills of the system in predicting meteorological drought were assessed by comparing SPI forecast with SPI observed at the 3, 6, and 12 (SPI3, SPI6, and SPI12) month time scales up to a 4-week lead time using HSS and anomaly correlation coefficient (ACC). SPI forecasts were also contrasted with the Famine Early Warning Systems Network (FEWS NET) Africa hazard outlooks during the case study of the March-May 2022 drought.

Correspondence to: Miliaritiana Robjhon, NCWCP, College Park, MD 20740;
miliaritiana.robjhon@noaa.gov

3. Analysis

Verification of bias corrected temperatures and precipitation forecasts is shown in Figure 1. BCSD consistently reduced RMSE and also contributed to an increase in HSS, though at a small proportion, throughout all lead times for temperatures and precipitation, respectively. The drop in errors and increase in skills suggested that BCSD was effective in removing biases in the GEFSv12 at the sub seasonal time scale.

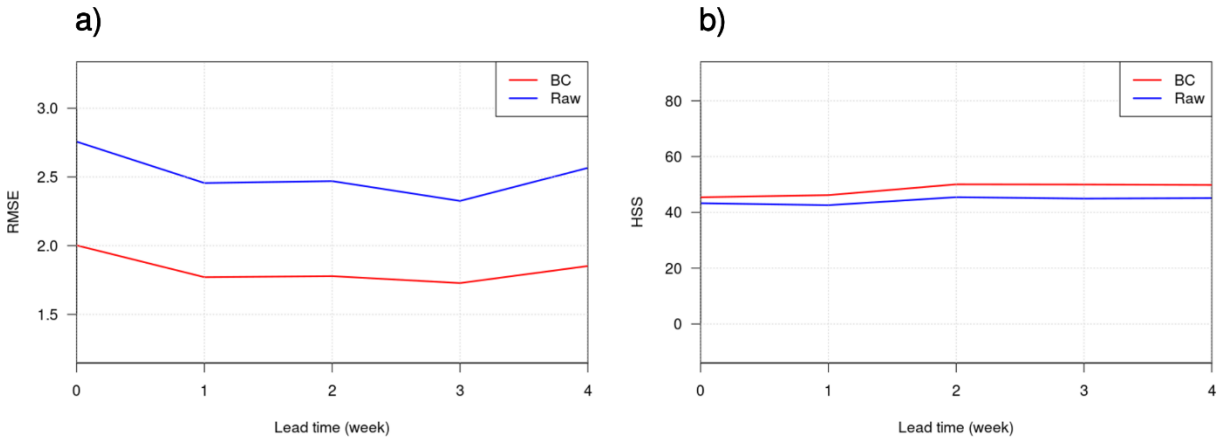

Figure 44: Evolution as a function of lead time for GEFSv12 raw and bias corrected forecasts for (a) temperatures RMSE in degrees Celsius and (b) precipitation HSS in %. Data source: GEFSv12.

Based on the values of SPI, droughts were classified into four categories: (i) abnormally dry to exceptional droughts, when SPI was less than -0.5 (D0-D4), (ii) moderate to exceptional droughts, when SPI was less than -0.8 (D1-D4), (iii) severe to exceptional droughts, when SPI was less than -1.3 (D2-D4), and (iv) extreme to exceptional droughts, when SPI fell below -1.6 (D3-D4). The spatial distribution of HSS over the Horn of Africa for drought at the D1-D4 category for SPI3 at 3-week lead time is shown in Figure 2a. High (greater than 70%) values of HSS were found over portions of Ethiopia, Kenya, Somalia, and Tanzania, which suggests that the location of observed droughts and severe meteorological droughts were captured by the prediction system. An analysis of the evolution of HSS for droughts at various severities as a function of lead time is shown in Figure 2b. As expected, HSS degraded as lead time also increased; however, the skill remained above 40% for D0-D4 and D1-D4 even up to a 4-week lead time, which suggests that the forecasting system could be skillful in predicting developing and or intensifying meteorological droughts at a sub seasonal time scale.

Correspondence to: Miliaritiana Robjhon, NCWCP, College Park, MD 20740;
miliaritiana.robjhon@noaa.gov

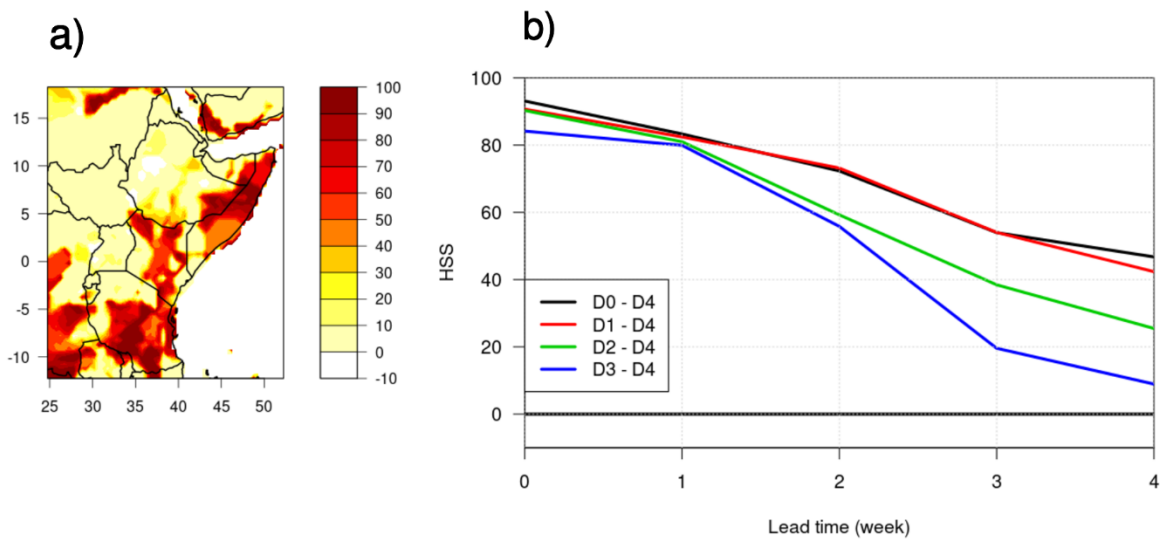

Figure 45: (a) Spatial distribution of HSS (%) over the Horn of Africa for drought at the D1-D4 classification for SPI3 at 3-week lead time, (b) Evolution of HSS (%) for drought at four different categories as a function of lead time. Data source: GEFSv12.

Due to its adverse impact on food production in eastern Africa, the March-May 2022 Horn of Africa drought was used to verify this capability in predicting rapidly intensifying droughts. A comparison of SPI3 forecasts, initialized on 25 April 2022 with SPI3 observed showed that while the system, overall, forecasted more severe meteorological droughts relative to the observations, the framework captured the rapid onset and intensification of droughts in Kenya, Somalia, and Tanzania at 2 to 3-week lead times. This predicted intensification of drought was also consistent with the observed change in agrometeorological conditions such as a one or more categories decrease in SPI, lower soil moisture status, and accompanying further degradation in vegetation conditions, which prompted the FEWS NET Africa regional hazards outlook to upgrade droughts to severe droughts over portions of the sub-region by late May 2022.

4. Conclusion

In this study, the performance of a novel drought prediction system, which primarily consisted of the GEFSv12, Bias Correction, Spatial Downscaling, and Leaky Bucket was assessed over the Horn of Africa. Results showed that the applied bias correction method was effective in reducing model systematic biases in temperatures and precipitation forecasts. The forecast system was skillful in predicting not only the locations of meteorological droughts over the Horn of Africa but also their overall severities from D0-D4 and D1-D4 up to 4-week lead time. The prediction system also captured the rapid intensification of droughts during the March-May 2022 Horn of Africa drought at 2 to 3-week lead times.

Future work includes a further evaluation of drought indices forecast, extension to global scale, and transition to operations.

5. Data availability statement

The GEFSv12 model data are accessible from the Amazon Web Services (https://registry.opendata.aws/noaa-gefs).

Correspondence to: Miliaritiana Robjhon, NCWCP, College Park, MD 20740; miliaritiana.robjhon@noaa.gov

6. References

1. Chen, M., W. Shi, P. Xie, V.B.S. Silva, V.E. Kousky, R. Wayne Higgins, and J.E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation, *J. Geophys. Res.*, **113** (D4), https://doi.org/doi:10.1029/2007JD009132.
2. Fan, Y., and H. van den Dool, 2004: Climate Prediction Center global monthly soil moisture data set at 0.5-degree resolution for 1948 to present, *J. Geophys. Res.*,**109** (D10), https://doi.org/10.1029/2003JD004345.
3. FAO, 2022: GIEWS Country Brief: Ethiopia, 24 August 2022, https://reliefweb.int/report/ethiopia/giews-country-brief-ethiopia-24-august-2022.
4. Hobbins, M.T., A. Wood, D.J. McEvoy, J.L. Huntington, C. Morton, M. Anderson, and C. Hain, 2016: The Evaporative Demand Drought Index. Part I: Linking Drought Evolution to Variations in Evaporative Demand, *J. Hydrometeor.*, **17(6)**, 1745–1761, https://doi.org/doi:10.1175/JHM-D-15-0121.1.
5. McKee, T.B., N.J. Doesken, and J. Kleist, 1993: The Relationship of Drought Frequency and Duration to Time Scales, *Proceedings of the Eighth Conference on Applied Climatology*, Anaheim, CA, Amer. Meteor. Soc., 179-184. https://www.droughtmanagement.info/literature/AMS_Relationship_Drought_Frequency_Duration_Time_Scales_1993.pdf.
6. Vicente-Serrano, S.M., S. Beguería, and J.I. López-Moreno, 2010: A Multi-Scalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index, *J. Climate*, **23**(7), 1696–1718, https://doi.org/10.1175/2009JCLI2909.1.
7. Yoon, J.-H., K. Mo, and E.F. Wood, 2012: Dynamic Model Based Seasonal Prediction of Meteorological Drought Over the Contiguous United States, *J. Hydrometeor.*, **13**(2), 463–482, doi:10.1175/JHM-D-11-038.

Correspondence to: Miliaritiana Robjhon, NCWCP, College Park, MD 20740; miliaritiana.robjhon@noaa.gov

Characterizing the Development and Drivers of 2021 Western U.S. Drought

Grace Affram[1], Wei Zhang[1,2], Lawrence Hipps[1], Cody Ratterman[1]
[1]Department of Plants, Soils & Climate, Utah State University, Logan, UT, USA
[2]Ecology Center, Utah State University, Logan, UT, USA

ABSTRACT

1. Introduction

While the effects of the failed 2020 North American monsoon and precipitation deficit over the southwestern U.S. have been examined (Hoell et al., 2022, Wang et al., 2022), the relative roles of drought drivers (i.e., precipitation, temperature, and evapotranspiration), and radiative forcings (e.g., anthropogenic and natural) in shaping the 2021 drought across the Western U.S. are not yet documented. Here we seek to understand the relative contributions of precipitation reductions and higher temperatures to the severe 2021 drought and the roles of natural and anthropogenic forcings in shaping this event.

2. Data and Methodology

Monthly gridded temperature data across the Western U.S. is obtained from the Global Historical Climatology Network version 2 and the Climate Anomaly Monitoring System (Fan & van den Dool, 2008), while precipitation data is provided by NOAA's Climate Prediction Center (Chen et al, 2008). Soil moisture estimates at 40 centimeters depth is obtained from the European Centre for Medium Range Weather Forecasts Reanalysis-5 (Muñoz-Sabater et al., 2021) and evapotranspiration from North American Land Data Assimilation System Phase 2 (Xia et al., 2012). We compared the anomalies of these variables from 2020 to 2021 for investigating the processes of drought development. Additionally, the Coupled Model Intercomparison Project Phase 6 (CMIP6) historical "all forcing" along with "natural forcing" runs from 1980-2014 are used to examine the potential role of external forcing in driving the extreme drought conditions using the risk ratio framework (Lloyd and Shepherd, 2021; Reisinger et al., 2020; Zhang et al., 2016). The 2021 extreme drought event is characterized by comparing observed values with the 5^{th} percentile – extremely low event of precipitation and soil moisture and the 95^{th} percentile of temperature. Four CMIP6 models (ACCESS-CM2, FGOALS–g3, IPSL-CM6A-LR, MRI–ESM2-0) are used to detect anthropogenic influence on temperature, precipitation, and soil moisture.

3. Results

Evapotranspiration (ET) undergoes a strong negative anomaly from 2020 to 2021 which means that changes in ET strongly depend on the availability of moisture at the onset of a drought as well as the severity and duration of the drought (Hanson, 2016). Similarly, our results show a significant positive correlation between ET and precipitation (r=0.658), contrary to that between temperature and ET (r=-0.170) during 2020 and 2021 (Figure 1a). Hence, the ideology that soil water gets depleted more rapidly under high temperatures may not necessarily be appropriate for regions where soil water is already under a dry condition. The dynamical impact of ET on temperature could explain why temperature and ET show opposite phases as seen in Figure 1a. A considerable portion of the 2021 dry soil in the western U.S.

Correspondence to: Grace Affram, Department of Plants, Soils & Climate, Utah State University, Logan, UT; g.affram@usu.edu

results from persistent soil moisture deficits established in late 2020 due to the subsequent failed summer monsoon in 2020 and extremely low precipitation (Figure 1b). The substantial positive precipitation anomaly in late 2021 and the strong summer monsoon in 2021 are used to recover soil moisture levels. Hence, little water remained for recovering streamflow and groundwater levels, causing major water reservoirs to decline (Van Loon, 2015).

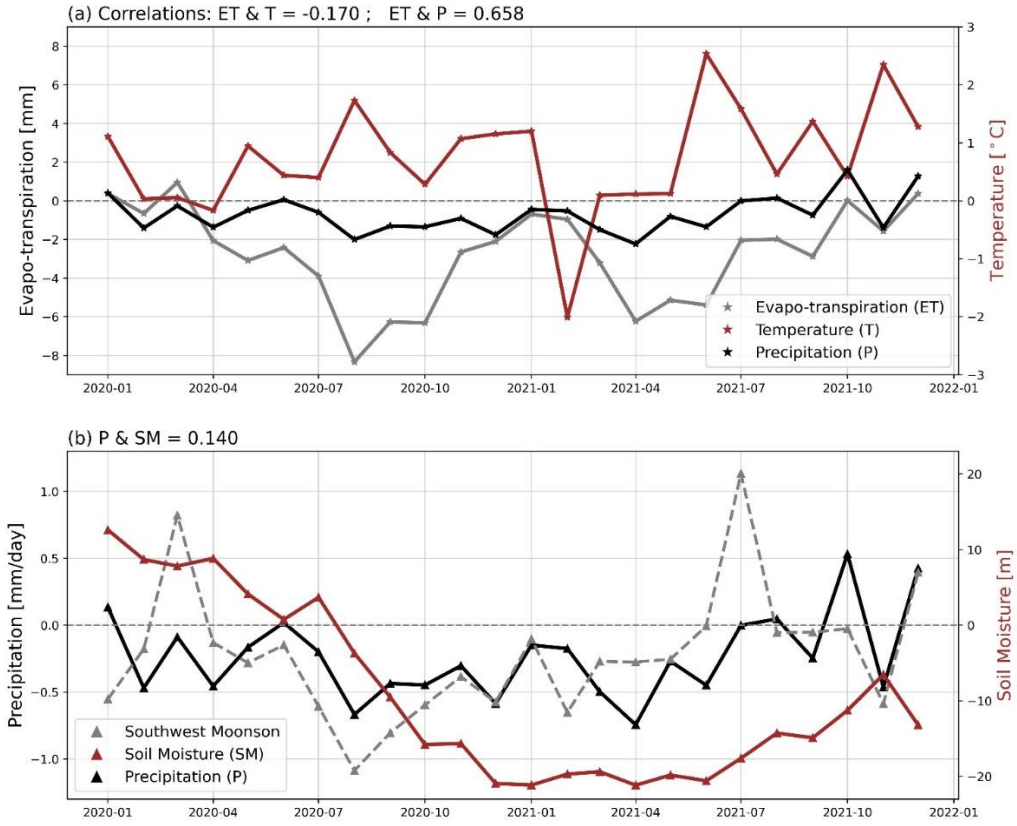

Figure 46: Monthly anomalies of (a) temperature (red), evapotranspiration (gray), (a, b) precipitation (black) and (b) soil moisture (red) across the western U.S. as well as precipitation across the North American monsoon region (dash-gray) from January 2020 to December 2021. Inscriptions at the top left corner indicate the correlations between the hydrometeorological variables (Affram et al. 2023).

Using CMIP6 model experiments under an attribution framework, we evaluate the relative roles of external forcing (including anthropogenic and natural forcing) in extreme temperature, precipitation, and soil moisture. Consistent with previous studies (Norris et al., 2021). The observed 2021 extreme events are slightly different from that in CMIP6 models, and this could be attributed to the models' mean biases. Overall, historical all forcing experiments exhibit an increased exceedance probability of the 2021 extreme event (Figure 2). For instance, the risk ratios (RR) for temperature, precipitation and soil moisture are 73.91, 12.78 and 25.81 respectively (Figure 2). These results suggest that anthropogenic forcing (i.e., greenhouse gases and land-use land-cover change; all forcing minus natural forcing) has dramatically increased the risk of the extreme dry conditions across the West.

Correspondence to: Grace Affram, Department of Plants, Soils & Climate, Utah State University, Logan, UT; g.affram@usu.edu

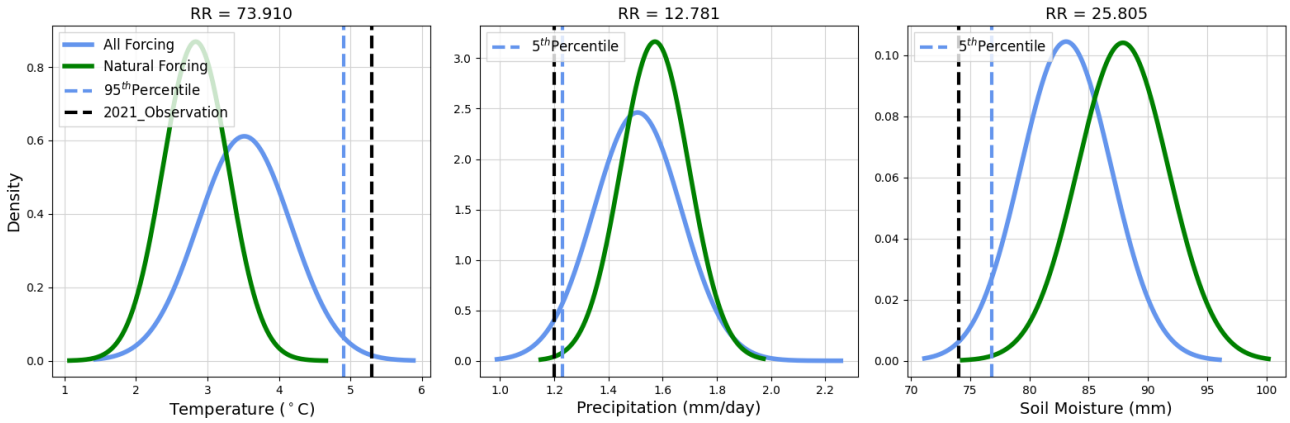

Figure 47: Attributing the extreme drought conditions using CMIP6's all forcing (blue) and natural forcing (green) experiments. The black dash lines represent the observed 2021 values, while the blue dash lines represent the extreme events in the models' all forcing experiments – 95th percentile for (a) temperature and 5th percentile for (b) precipitation and (c) soil moisture. RR represents risk ratio, and the y-axis represents the probability density (Affram et al. 2023.)

4. Conclusion

While recent studies indicated evapotranspiration as the driving force of global droughts (Wang et al., 2021), our analysis suggests that the suppressed precipitation may be the main driver of this extreme 2021 dry condition across the western U.S., as seen in a recent study (Affram et al. 2023; Hoell et al., 2022). The anomalous low soil moisture characterizes the severity of the extreme 2021 dryness in the western U.S. The lower-than-normal evapotranspiration under extremely high temperature during 2021 suggests that high temperatures may play a secondary role in driving the low soil moisture across the West and that increase in evapotranspiration in response to higher temperatures depends on the availability of water. Under an attribution framework, we have found that anthropogenic forcing significantly increased the risk of the extreme drought conditions in the western U.S., consistent with previous studies (NOAA, 2021; Chiang et al., 2021). Given precipitation deficit has been found as the major driver of the 2021 extreme drought event across the West, we encourage further investigation on its forcing mechanisms and the unusual, failed North American summer monsoon in 2020.

5. Data Availability Statement

The CMIP6 model outputs were obtained from the website, https://esgf-node.llnl.gov/search/cmip6/. Monthly temperature data set was obtained from the Global Historical Climatology Network version 2 and the Climate Anomaly Monitoring System via https://psl.noaa.gov/data/gridded/data.ghcncams.html. Precipitation data was provided by NOAA Climate Prediction Center https://psl.noaa.gov/data/gridded/data.unified.daily.conus.html. Soil moisture data was obtained from the European Centre for Medium Range Weather Forecasts (ECMWF) Reanalysis-5 (ERA5) via https://cds.climate.copernicus.eu as well as evapotranspiration data from the North American Land Data Assimilation System Phase 2 via https://disc.gsfc.nasa.gov/datasets.

Correspondence to: Grace Affram, Department of Plants, Soils & Climate, Utah State University, Logan, UT; g.affram@usu.edu

6. References

1. Affram, G., W. Zhang, L. Hipps, and C. Ratterman, 2023: Characterizing the development and drivers of 2021 Western US drought. Environmental Research Letters, **18**(4), 044040. https://doi.org/10.1088/1748-9326/acc95d
2. Chen, M., W. Shi, P. Xie, V.B.S. Silva, V.E. Kousky, R.W. Higgins, and J.E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research*, **113**, D04110, https://doi.org/10.1029/2007JD009132
3. Hanson, R.L., 2016: Evapotranspiration and Droughts, U.S. Geological Survey, https://geochange.er.usgs.gov/sw/changes/natural/et/

Correspondence to: Grace Affram, Department of Plants, Soils & Climate, Utah State University, Logan, UT; g.affram@usu.edu