GEFSv12 reforecast dataset for supporting subseasonal and hydrometeorological applications

Hong Guan[1], Yuejian Zhu[2], Eric Sinsky[3], Bing Fu[3], Wei Li[3], Xiaqiong Zhou[4], Xianwu Xue[1], Dingchen Hou[2], Jiayi Peng[3], M. M. Nageswararao[5], Vijay Tallapragada[2], Thomas M. Hamill[6], Jeffrey S. Whitaker[6], Gary Bates[7], Philip Pegion[6], Sherrie Frederick[6,7], Matthew Rosencrans[8], Arun Kumar[8]

[1]SRG at NOAA/NWS/NCEP/EMC, College Park, Maryland
[2]NOAA/NWS/NCEP/EMC, College Park, Maryland
[3]IMSG at NOAA/NWS/NCEP/EMC, College Park, Maryland
[4]CPAESS, University Corporation for Atmospheric Research at NOAA/NWS/NCEP/EMC and NOAA/OAR/GFDL, Princeton, New Jersey
[5]CPAESS, University Corporation for Atmospheric Research at NOAA/NWS/NCEP/EMC, College Park, Maryland
[6]NOAA Physical Sciences Laboratory, Boulder, Colorado
[7]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder
[8]NOAA/NWS/NCEP/CPC, College Park, Maryland

Submitted to *Monthly Weather Review*
17 December 2021

Corresponding author:

Dr. Hong Guan
5830 University Research Ct,
College Park, MD 20740
Email: Hong.Guan@noaa.gov

Abstract

For the newly implemented Global Ensemble Forecast System version 12 (GEFSv12), a 31-year (1989-2019) ensemble reforecast dataset has been generated at the National Centers for Environmental Prediction (NCEP). The reforecast system is based on NCEP's Global Forecast System version 15.1 and GEFSv12, which uses the Finite Volume 3 dynamical core. The resolution of the forecast system is ~25 km with 64 vertical hybrid levels. The Climate Forecast System (CFS) reanalysis and GEFSv12 reanalysis serve as initial conditions for the Phase 1 (1989–1999) and Phase 2 (2000–2019) reforecasts, respectively. The perturbations were produced using breeding vectors and ensemble transforms with a rescaling technique for Phase 1 and ensemble Kalman filter 6-h forecasts for Phase 2. The reforecasts were initialized at 0000 (0300) UTC once per day out to 16 days with 5 ensemble members for Phase 1 (Phase 2), except on Wednesdays when the integrations were extended to 35 days with 11 members. The reforecast data set was produced on NOAA's Weather and Climate Operational Supercomputing System at NCEP.

This study summarizes the configuration and dataset of the GEFSv12 reforecast and presents some preliminary evaluations of 500hPa geopotential height, tropical storm track, precipitation, 2-meter temperature, and MJO forecasts. The results were also compared with GEFSv10 or GEFS Subseasonal Experiment reforecasts. In addition to supporting calibration and validation for the National Water Center, NCEP Climate Prediction Center, and other National Weather Service stakeholders, this high-resolution subseasonal dataset also serves as a useful tool for the broader research community in different applications.

1. Introduction

The important role of a reforecast in validating and calibrating weather and climate model forecasts (Hamill et al. 2004, 2006, 2013, 2015; Hamill and Whitaker 2006; Wilks and Hamill 2007; Hagedorn et al. 2008, 2012; Hagedorn 2008; Hamill 2012; Hamill and Kiladis 2013; Baxter et al. 2014; Scheuerer and Hamill 2015; Ou et al. 2016; Guan et al. 2015, 2019; Gascon et al. 2019), diagnosing model errors (Hamill et al. 2013), and predicting extreme or rare events (Hagedorn 2008; Hamill et al. 2008, 2013; Guan and Zhu 2017; Nardi et al. 2018; and Li et al. 2019) has been widely recognized. Currently, reforecast datasets are utilized operationally at several weather-climate centers worldwide. For instance, a reforecast dataset is used to calibrate forecasts at the Canadian Meteorological Center (CMC), the National Centers for Environmental Prediction (NCEP), and European Centre for Medium-Range Weather Forecasts (ECMWF) to improve numerical weather guidance for a variety of forecast timescales. In combination with an analysis climatology, a reforecast (i.e., model) climatology is also employed to provide real-time extreme weather forecasts for some common concern weather elements at NCEP (Guan and Zhu 2017) and ECMWF (LaLaurette 2003; Hagedorn 2008). Reforecasts are used extensively in conjunction with hydrologic prediction (DeMargne et al. 2014; Scheuerer and Hamill 2018; Emerton et al. 2018). More recently, as part of the Subseasonal Experiment (SubX; Pegion et al. 2019), seven modeling groups from the U.S. and Canada generated reforecast datasets, separately. The combined datasets provide a foundation for employing current best practice methods for real-time weeks 3 and 4 outlooks of hazardous and extreme events at the NCEP Climate Prediction Center (CPC).

Ideally, creating a reforecast dataset requires a set of consistent reanalysis data as initial conditions. Both reforecast and reanalysis should also employ the same model system that is used in the actual real-time forecast, ideally at the same resolution. However, generating a full dataset for a reanalysis and reforecast, usually from 10 years to several decades of data, is an extremely time- and labor-intensive procedure and impractical in operational forecasting. Therefore, an inconsistent initial analysis had been used for the GEFSv11 (Guan and Zhu 2017) and GEFS-SubX reforecasts. For example, the 17 years (1999–2015) of GEFS_SubX reforecasts (Zhu et al. 2018; Li et al. 2019; Guan et al. 2019) used the Climate Forecast System

Reanalysis (CFSR) and Global Data Assimilation System (GDAS) as the initial conditions for 1999–2010 and 2011–2016, respectively. In addition to the inconsistency of the analysis itself, the forecast systems generating the reanalysis are also quite different from the reforecast and real-time forecast systems. This inconsistency in reanalysis has resulted in a difference in the 2-m temperature bias characteristics (Hamill 2017; Guan et al. 2019), especially for short lead times when initial conditions play a critical role in the forecast. This further confirms the strong desirability of simultaneously generating reanalysis and reforecast data in the operational implementation.

On September 23, 2020, the FV3 (Finite-Volume)-based Global Ensemble Forecast System version 12 (GEFSv12) was implemented at the National Oceanic and Atmospheric Administration (NOAA). To provide seamless numerical guidance to a broad range of users and partners, the integration time of the GEFSv12 was extended from week 1 (weather forecasts) and week 2 (extended forecasts) to weeks 3-5 (subseasonal forecasts). Accompanying the GEFSv12 implementation, 20-year reanalysis and 31-year reforecast datasets were also simultaneously produced by NOAA's Physical Science Laboratory (PSL) and Environmental Modelling Center (EMC), respectively, to support stakeholders CPC and the National Water Center (NWC) for subseasonal and hydrological applications. This marks the first official generation of a reanalysis/reforecast as an integral part of an implementation of the GEFS at NOAA. In addition, North American Ensemble Forecast System (NAEFS; Candille 2009; Candille et al. 2010) products have been updated based on the GEFSv12 Phase 2 reforecast.

The reforecast system configuration is summarized in Section 2. The reforecast dataset, public access, and data corrections are introduced in Section 3. The statistical characteristics of the raw forecasts are described in Section 4. In Section 5, an example of the reforecast application is discussed. Summary and conclusions are given in Section 6.

2. Reforecast system configuration

The GEFSv12 reforecast system is based on the current operational Global Forecast System version 15.1 (GFSv15.1; EMC website, 2019) which uses the Geophysical Fluid Dynamics Laboratory (GFDL) FV3 Cubed-Sphere dynamical core (Lin and Rood 1997; Lin

2004; Putman and Lin 2007; Harris and Lin 2013). The resolution of the forecast system is ~25 km (C384 grid) in the horizontal with 64 vertical hybrid levels with the top layer centered around 0.27 hPa (~55 km).

The convection scheme used in the GEFSv12 is the Simplified Arakawa-Schubert (SAS) shallow and deep convection schemes (Han and Pan 2011) updated with a scale-aware parameterization (Han et al. 2017). The scheme was also further modified to reduce excessive cloud top cooling for the model stabilization. The cloud microphysics scheme is from GFDL, which includes five predicted cloud species (cloud water, cloud ice, rain, snow and graupel; Zhou et al. 2019, 2021). The vertical mixing process of the planetary boundary is based on the hybrid Eddy-diffusivity Mass-flux (EDMF) scheme (Han et al. 2016). The shortwave and longwave radiative fluxes are calculated using the rapid radiative transfer model (RRTM) developed at Atmospheric and Environmental Research (Clough et al. 2005). The GFS orographic gravity wave drag and mountain blocking schemes follows Alpert (1988), while convective gravity wave drag employs the scheme developed by Chun and Baik (1998). The GFS Noah land surface model (Chen et al., 1996; Koren et al. 1999; Ek et al. 2003; Michell et al. 2005) are used to simulate the land-surface processes. The surface layer parameterization follows Long (1984; 1986) and Zheng et al. (2012; 2017).

The SST boundary condition is derived from a two-tiered Sea Surface Temperature (SST) and Near Sea Surface Temperature (NSST) approach that accounts for the day-to-day variability and diurnal variation of SST, respectively (Zhu et al. 2017, 2018; Li et al. 2019). A modern ensemble forecast system should include initial perturbations to approximate analysis/observation uncertainty and model perturbations to approximate the forecast uncertainty from model imperfections, such as the finite resolution of the prediction system and the use of deterministic parameterizations of sub-grid phenomena (Buizza et al. 1999; Palmer 2001, 2012; Berner et al. 2017). To improve the model's uncertainty representation, stochastic kinetic energy backscatter (SKEB; Shutts and Palmer 2004; Shutts 2005) and stochastically perturbed parameterization tendencies (SPPTs; Buizza et al. 1999; Palmer et al. 2009) are applied. More details on the GEFSv12 forecast system can be found in Zhou et al. (2019; 2021).

The reforecast was integrated once per day out to 16 days, except on Wednesdays when the forecast was extended to 35 days. In contrast to the real-time forecast system (31 members), the reforecast system has a smaller ensemble size to minimize computational expense: 5 and 11 members for the 16-day and 35-day runs, respectively. As illustrated in Table 1, the reforecast utilizes two sets of analysis data because a consistent 31-year reanalysis is unavailable.

Reforecast characteristic	**1989-1999**	**2000-2019**
Reanalysis states for initial conditions	CFSR (Saha et al. 2010) + bred vectors (Wei et al. 2008)	GEFSv12 (Hamill et al. 2021)
SST initial states	OI (Reynolds et al. 2002)	OI (Reynolds et al. 2002)
SST forecast	NSST (Zhu et al. 2017, 2018; Li et al. 2019)	NSST (Zhu et al. 2017, 2018; Li et al. 2019)
Soil moisture and vegetation classification for initial states	Following Zobler 1986, 1999, Dorman and Sellers 1989).	Following Ek et al. (2016)

Table 1. The summary of initial and boundary conditions for the GEFSv12 reforecasts.

For the Phase 1 reforecast (GEFSv12_p1, 1989–1999), the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) was used as the initial control analysis. The breeding vector and ensemble transform with rescaling (BV-ETR) cycling perturbations (Wei et al. 2008), generated for the NOAA's 2nd generation GEFS reforecast (Hamill et al., 2013), was used as initial conditions for the perturbed members. The new 16 State Soil Geographic (STATSGO) soil classification and 20 International Geosphere Biosphere Programme (IGBP) vegetation classification (Ek et al. 2016) were applied to characterize soil and vegetation in the reforecast runs, although the CFSR used the old 9 soil texture classes (Zobler 1986, 1999) and 13 vegetation catalogues (Dorman and Sellers 1989).

For the Phase 2 reforecasts (GEFSv12_p2, 2000–2019), initial conditions were GEFSv12 reanalyses (Hamill et al. 2021). The reanalyses were generated from the FV3 GFS/Ensemble Kalman Filter (EnKF) hybrid analyses and EnKF 6-h forecasts with the Incremental Analysis Update (IAU; Bloom et al. 1996) replay process, which distributes the analysis increments over each time step within a fixed time window (currently 2100–0300 UTC). During this replay procedure, the climatological snow depths at 0000, 0600, and 1200 UTC (affected by a bug in data assimilation, see Hamill et al. 2021) were replayed to corresponding snow analyses to adjust reanalysis states to be more consistent with the snow analyses at these times. The GEFSv12_p2 reforecast was initiated from the data at the end of the replay IAU window (i.e., 0300 UTC). For both the GEFSv12 reanalysis and GEFSv12_p2 reforecast, soil moisture and vegetation were sorted based on the 16 soil-moisture and 20 vegetation types (Ek et al. 2016).

The GEFSv12 reanalysis also has several differences compared to the current operational analysis. First, the IAU process was applied to reduce noise and improve accuracy. Second, the NSST was replaced by Optimum Interpolation Sea Surface Temperature (OISST; Reynolds et al. 2002) to avoid an observed large SST bias in climatologically cloudy regions for the earlier assimilation years. Third, to reduce the computation resources required, the horizontal resolutions of the control and perturbed members were decreased from C768 (~13 km) and C384 (~25 km) to C384 and C192 (~50 km), respectively. A detailed description of the GEFSv12 reanalysis can be found in Hamill et al. (2021).

3. Reforecast dataset, public access, and data corrections

a. Reforecast dataset and public access

The full 31 years of reforecast data are currently archived on the High Performance Storage System (HPSS). All 590 variables in grib2 format are saved at 3-hour intervals at 0.25° resolution for the first 10 days and 6-hour intervals at 0.5° beyond 10 days of the forecast. By request, 77 of the 590 variables were stored on the WCOSS disk for quick access by the internal NOAA stakeholders. The 219 selected variables for the Phase 2 reforecasts are saved on dedicated disks mounted on NOAA/NWS/NCEP's ftp server (ftp://ftp.emc.ncep.noaa.gov/GEFSv12/reforecast) and Amazon Web Services (AWS, https://noaa-gefs-retrospective.s3.amazonaws.com/index.html), which are accessible by the broader community. These 176 upper-air and 43 surface or single-level publicly accessible variables are separately listed in Tables 2 and 3, respectively. For pressure-level data above 700 hPa (Table 2), the Phase 2 data are also saved at 0.5-degree grid spacing, even during the first 10 days of the forecast, to conserve space.

Vertical Level	**U**	**V**	**W**	**T**	**Height (P)**	**Q (RH)**	**PV**
1 hPa	X	X	X	X	X		
2 hPa	X	X	X	X	X		
3 hPa	X	X	X	X	X		
5 hPa	X	X	X	X	X		
10 hPa	X	X	X	X	X		
20 hPa	X	X	X	X	X		
30 hPa	X	X	X	X	X		
50 hPa	X	X	X	X	X		
70 hPa	X	X	X	X	X		
100 hPa	X	X	X	X	X	X	
150 hPa	X	X	X	X	X	X	
200 hPa	X	X	X	X	X	X	
250 hPa	X	X	X	X	X	X	
300 hPa	X	X	X	X	X	X	
400 hPa	X	X	X	X	X	X	
500 hPa	X	X	X	X	X	X	
600 hPa	X	X	X	X	X	X	
700 hPa	X	X	X	X	X	X	
800 hPa	X	X	X	X	X	X	
850 hPa	X	X	X	X	X	X	
900 hPa	X	X	X	X	X	X	
925 hPa	X	X	X	X	X	X	
950 hPa	X	X	X	X	X	X	
975 hPa	X	X	X	X	X	X	
1000 hPa	X	X	X	X	X	X	
1(hybrid)	X	X	X	X	X	(X)	
2(hybrid)	X	X	X	X	X	(X)	
3(hybrid)	X	X	X	X	X	(X)	
4(hybrid)	X	X	X	X	X	(X)	
$2x10^{-6}$ (PV)	X	X		X	(X)		

$310x10^{-6}$ K m^2 kg^{-1} s^{-1} (Isentropic)							X
$320x10^{-6}$ K m^2 kg^{-1} s^{-1} (Isentropic)							X
$350x10^{-6}$ K m^2 kg^{-1} s^{-1} (Isentropic)							X
10m (AGL)	X	X					
100m (AGL)	X	X					

Table 2. One-hundred seventy-six upper air variables**.**

Variables	**total**
Mean sea-level pressure	1
Surface pressure	1
Surface height	1
Skin temperature	1
Soil temperature at 0.0-0.1, 0.1-0.4, 0.4-1.0 and 1.-2. m depth	4
Volumetric soil content at 0.0-0.1, 0.1-0.4, 0.4-1.0 and 1.-2. m depth	4
Water equivalent of accumulated snow depth	1
2-m temperature	1
2-m specific humidity	1
Maximum temperature in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Minimum temperature in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Surface wind gust	1
Surface wind stress, u-component	1
Surface wind stress, v-component	1
Surface roughness	1
Total precipitation in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Convective precipitation in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Non-convective precipitation in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Boundary layer height	1
Average surface latent heat net flux average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Average surface sensible net heat flux average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Average ground heat net flux average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Convective available potential energy	1

Convective inhibition	1
0-3 km Storm relative helicity	1
Perceptible water	1
Total ozone	1
Total cloud cover average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Downward shortwave radiation flux at the surface average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Downward longwave radiation flux at the surface average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Upward shortwave radiation flux at the surface average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Upward longwave radiation flux at the surface average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Upward longwave radiation flux at the top of the atmosphere average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Momentum Flux, U-Component Average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Momentum Flux, V-Component average in last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1
Cloud ceiling	1
Water runoff sum over the last 6-h period (00, 06, 12, 18 UTC) or in last 3-h period (03, 09, 15, 21 UTC)	1

Table 3. Forty-three surface and other single-level variables.

b. Data corrections

The integrations for the Phase 2 reforecasts were initiated from the 0300 UTC restart data files. Thus, the model outputs for the 41 0000-0300 UTC and 0000-0600 UTC accumulated / minimum / maximum / average variables are incorrect since they were actually calculated based on the values from the beginning of integration (i.e., 0300 UTC) to the first time-step and to 0600 UTC, respectively. These 41 variables were post-processed by combining the control NEMSIO (NOAA Environmental Modeling System Input/Output) replay reanalysis at 0300 UTC and the reforecast data at 0600 UTC. Note that the replay process was only applied to the control members so that for 0300 UTC, the reforecast data for each member was simply replaced by the corresponding control-member replay data. For 0600 UTC, the minimum and maximum are the smaller and larger of the two values, respectively, while the accumulated values are the sum of the two. The 6-h average fields were processed in a more complicated manner. The raw reforecast average field at 0600 UTC is actually the 0300–0600 UTC accumulation divided by a 6-h time period, while in reality the accumulations take place over a 3-h period. This was corrected to a 3-h average and then averaged with the reanalysis data at 0300 UTC. But for some variables and conditions such an average is not suitable and special processing is needed. For cloud-base/cloud-top pressures and cloud-top temperatures, the 0000–0600 averages were set to be the same as those at 0300 UTC when clouds do not exist in the 0600 UTC, while the corresponding averages were set to be the same as those at 0600 UTC when clouds do not present in the 0300 UTC forecasts. Such a special rule was also applied to snow melting flux.

4. Reforecast evaluation

In addition to the GEFSv12 reforecast and corresponding reanalyses used for initialization described in Section 3, there are also six sets of additional data being used for the current evaluations and comparisons. These additional datasets are as follows:

1. CFS reanalysis (1979–Mar 2011) at T382L64 (~34 km horizontal) resolution. The documentation of the system, including the configurations, can be found in Saha et al. 2010. The dataset was used as the initial condition for NOAA's second-generation of reforecasts (or GEFSv10 reforecast; Hamill et al. 2013) and GEFS_SubX reforecast (Zhu et al. 2018).
2. NCEP's operational analysis from the GDAS (NCEP hybrid Global Data Assimilation System) (2011–Present). The documentation of the GDAS upgrade, including the changes in configurations, can be tracked through the EMC web-page: https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs.php

These data served as the initial condition for the GEFSv10 and GEFS_SubX reforecats for the periods 2011–present and 2011–2016, respectively.

3. The European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis version 5 (ERA5) (1950–Present) data approximately 30 km horizontal resolution with 137 hybrid vertical levels, up to an 80 km model top. The documentation of the ERA5 system, including the configurations, can be found through ECMWF's web-page: https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation. These data were used to evaluate the 2-m temperature forecast for the GEFSv12 and GEFS_SubX reforecast.

4. NCEP's Climate Calibrated Precipitation Analysis (CCPA; 2002–Present) version 4 (v4) for Continental United States (CONUS). The documentation can be found in Hou et al. (2014) and Luo et al. (2018). These data were used to evaluate precipitation forecasts for the GEFS_SubX and GEFSv12 reforecasts and to calibrate the GEFSv12 reforecast.

5. GEFSv10 reforecast (1985–2011) and forecast (2012–2019). The documentation on this system and configurations can be found through Zhu et al. (2012) and Hamill et al. (2013). These data were used for the comparison with the GEFSv12 reforecast for hurricane track forecasts.

6. GEFS_SubX reforecast (1999–2016) and forecast (2017–2018) at TL574L64 (day 0–8; ~34 km horizontal resolution) and TL382L64 (day 8–35; ~52 km horizontal resolution). The documentation of the GEFS_SubX system and the configurations can be found in Zhu et al. (2018). The GEFS_SubX reforecast is considered a benchmark dataset to measure the ability of the GEFSv12 reforecast to predict 500-mb geopotential height, 2-m temperature, precipitation, and Madden Julian Oscillation (MJO).

a. 500-hPa geopotential height

The anomaly correlation of 500-hPa geopotential height is widely used as an essential metric to estimate the skill of weather forecasts, especially for mid- and high-latitude weather systems. Here, 500-hPa geopotential height for the GEFS_SubX and GEFSv12_p2 reforecasts are evaluated against their own analyses (i.e., CFSR and GEFSv12 reanalysis). CDAS2 is the analysis climatology used to calculate analysis anomalies as well as forecast anomalies for both GEFSv12_p2 and GEFS_SubX. Over the Northern Hemisphere (NH, Fig. 1), the GEFSv12_p2 outperforms the GEFS_SubX with improvements in average anomaly correlation (AC) of 1.5%, 5.5%, and 2.5% for week 1, week 2, and weeks 3 and 4 forecasts, respectively. Like Zhu et al.'s work (2018), the anomaly correlations for week 1, week 2, and weeks 3 and 4 are calculated by averaging forecast lead days 1–7, 8–14, and 15–28, respectively, and the corresponding analysis valid at 0000 and 0012 UTC. Over the Southern Hemisphere (SH, Fig. 2), the average AC scores are slightly lower than over the NH, which is consistent with the previous finding in Zhu et al. (2018) for the evaluation of the 16-year GEFS_SubX reforecast. Relative to the GEFS_SubX, the GEFSv12_p2 shows

1.3% and 3.0% improvements for week 1 and week 2 forecasts and a 3.3% degradation for the weeks 3 and 4 forecasts. The significant tests indicate that the Week -1 and Week-2 GEFSv12_p2 AC are significantly higher than GEFS_SubX for both NH and SH, while the corresponding AC values are not significantly different between the GEFSv12_p2 and GEFS_SubX for Weeks 3 and 4. The figures also reveal higher AC scores in the second decade (2010–2019) than the first decade (2000–2009) of the reforecast, and the corresponding calculations indicate that the weeks 3 and 4 scores for the NH in the second decade increase by 0.074 (or 25%) and 0.077 (or 26%) for the GEFS_SubX and GEFSv12_p2, respectively. The enhanced observation system (Noh et. al. 2020) may be an explanation for the better performances of 500-hpa forecasts in the most recent decade.

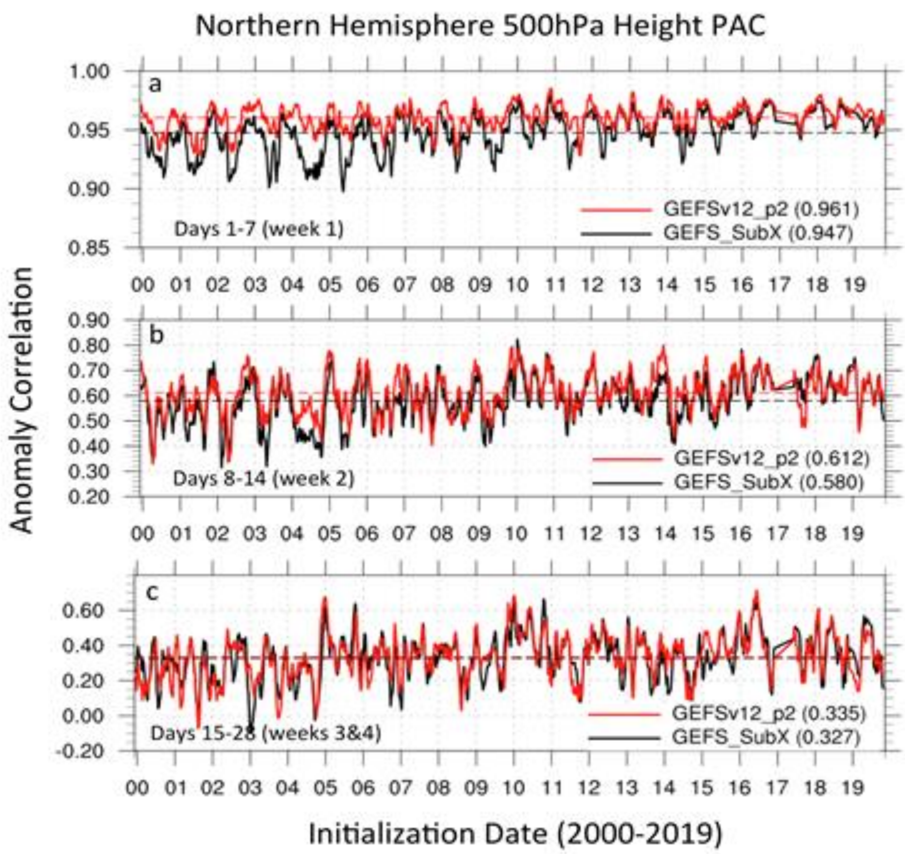

Figure 1. Ensemble-mean anomaly correlation for Northern Hemisphere (NH; 20°N–80°N) 500-hPa geopotential height for week 1 (a), week 2 (b), and weeks 3&4 (c) forecasts. The black and red colors denote the GEFS_SubX and GEFSv12_p2. The average scores for the two sets of reforecasts are indicated by the dashed lines and shown in parentheses. Note there is a data

gap from Dec 2016 to May 2017, corresponding to the period between the GEFS_SubX reforecast and corresponding real-time forecast. A 6-case moving average is applied to the time series. Since the forecasts are initialized every 7 days, the moving average spans over 42 calendar days.

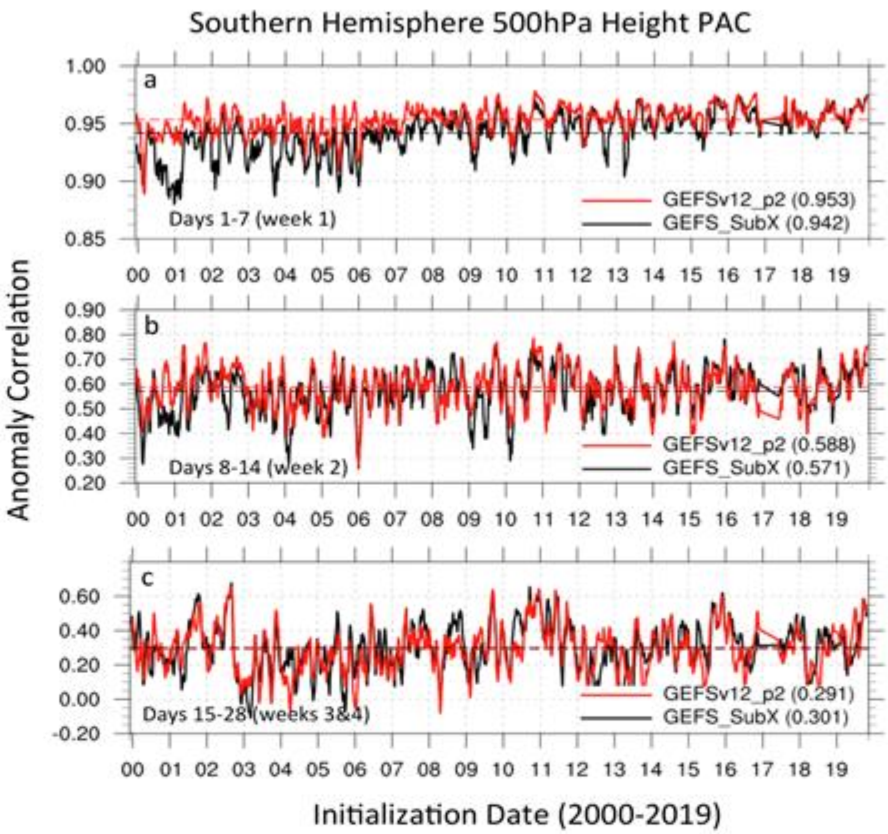

Figure 2. The same as Fig. 3 except for the Southern Hemisphere (SH; 20°S–80°S).

b. Tropical cyclone track

Tropical cyclone (TC) track forecasting has been challenging (Landsea and Cangialosi 2018), especially for the extended range (beyond day 5). To evaluate the ability of GEFSv12 to forecast tracks, track errors of the 5-member ensemble means of the GEFSv10 and GEFSv12 are compared for the 31-year reforecast period. The GEFSv10 was selected because it has a large sample data size like the GEFSv12 does. For consistency, in addition to the 5-member runs of the GEFSv12 reforecast, only the first five members of the GEFSv10 and of the 11-member runs of GEFSv12 are used in this comparison. The National Hurricane Center

(NHC)/Joint Typhoon Warning Center (JTWC) best (or observed) tracks were used as a reference for evaluating the two datasets.

The GEFSv12 skill in forecasting TC tracks has improved from the GEFSv10. Figure 3 shows the three-basin (Atlantic, East Pacific and West Pacific) averaged track errors from both forecast systems, binned by decade. For all three decades, the GEFSv12 reduces the track errors with the maximum reduction during the 2000-2010 period, when the reductions reach approximately 25% and 10% for 1-day and 7-day forecasts, respectively. For the GEFSv10, the track errors decline with decade (Fig. 3), which is qualitatively consistent with the finding in Hamill et al. (2013), based on the 1985-2011 reforecast. This evolution of the track errors is attributed to the improvement in the initial analysis over the multi-decade period, implying the important impact of initial conditions on the TC track forecast. For the shorter lead times, the decline in error from the 2000–2010 to 2011–2019 period is more evident than that from the 1989–1999 to 2000–2010 period. For example, the error reduction is 11.6 nm (or 29.8%) between the two later periods, while the corresponding reduction is 5.7 nm (or 12.8%) between the two earlier periods. In addition to the observation data increase with decade, the analysis system upgrade from CFSR to GFS/GDAS and the perturbation method change from BV-ETR to EnKF during the 2011–2019 period may be a reason for the observed sharper error reduction. The impact of initial conditions is also further confirmed in the current GEFSv12 reforecast. The track errors in the two GEFSv12 reanalysis time periods (2000–2010 and 2011–2019) are more consistent with each other and much smaller compared to the CFSR period (1989–1999), showing the importance of initialization with modern assimilation methods. The consistent error characteristics during the Phase 2 reforecast provide a good potential for statistical post processing algorithms to improve the TC track forecast (Galarneau and Hamill 2015). In addition to the initial conditions, the reforecast model itself also plays a role in influencing the accuracy of the track forecast. This is illustrated by the comparison between the GEFSv10 and GEFSv12 during the 1989–1999 period, when both reforecasts used the CFSR as the initial condition. As should be expected, the model's influence becomes more pronounced at longer lead times (> ~4-days). Compared to GEFSv10, the GEFSv12 reduces the track errors by 6.3% and 5.5% for the 6-day and 7-day forecasts, respectively.

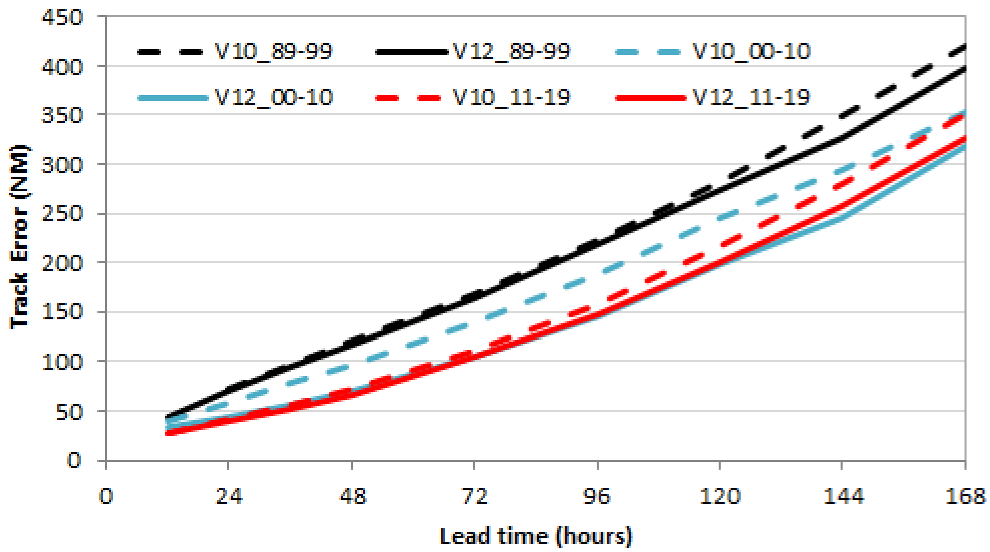

Figure 3. The TC track errors averaged over the Atlantic, East Pacific and West Pacific basins binned by decade during the 31-year reforecast for GEFSv10 (dashed lines) and GEFSv12 (solid lines). Black, blue, and red lines denote the 1989–1999, 2000–2010, and 2011–2019 periods, respectively.

c. Precipitation

The precipitation forecasts for the GEFS_Subx and GEFSv12 were estimated against the CCPAv4 for the 2002–2019 period when the reforecast and CCPA data overlapped. The CCPA climatology was calculated based on the 2002-2019 CCPA data. For this study, the 11-member reforecasts and CCPA data were interpolated to a 1°x1° grid over the Continental United States (CONUS), the only available analysis region. Figures 4a and 4b show the comparisons of Brier score (BS, Brier 1950) between the two sets of reforecasts for the 24-h accumulated precipitation greater than 1 mm and 5 mm, respectively. The BS, ranging between 0 and 1, is commonly used to verify the accuracy of a probability forecast. Clearly, the

GEFSv12 consistently displays the better (i.e., lower) Brier scores compared to the GEFS_SubX, with a more obvious improvement at lead times shorter than about 10 days. Forecast skill decreases with lead time and reaches saturated values at approximately day 13 for all situations. The precipitation probability forecast biases for 12–36 hrs and 60–84 hrs for amounts greater than 1-mm and 5-mm were measured by reliability diagrams (Fig. 5). The GEFSv12 and GEFS_SubX show very similar performance for the precipitation greater than 1.00 mm. For the heavier precipitation category (> 5 mm), the GEFSv12 slightly outperforms the GEFS_SubX with its curves being closer to the diagonal lines. Fig. 5 also shows the reliability curves are much closer to the diagonal at low probabilities but veering away for high probabilities. The Brier Skill Score (BSS, Wilks 1995) measures the improvement of the probability forecast over the reference climatology. Unlike BS, where lower is better, for BSS higher is better. In the heavier rain conditions, the BSS for the probabilistic precipitation forecast for the GEFSv12 are improved by about 16.1% and 20.1% for 12–36 hrs and 60–84 hrs, respectively (Fig. 5). The improvements are also observed for the other lead times (not shown). These improvements are attributed to the combined influence of better initial conditions, more advanced microphysics schemes, finer resolution and a new FV3 dynamic core. The impact that each of these factors has individually on the evaluation is not addressed in this study.

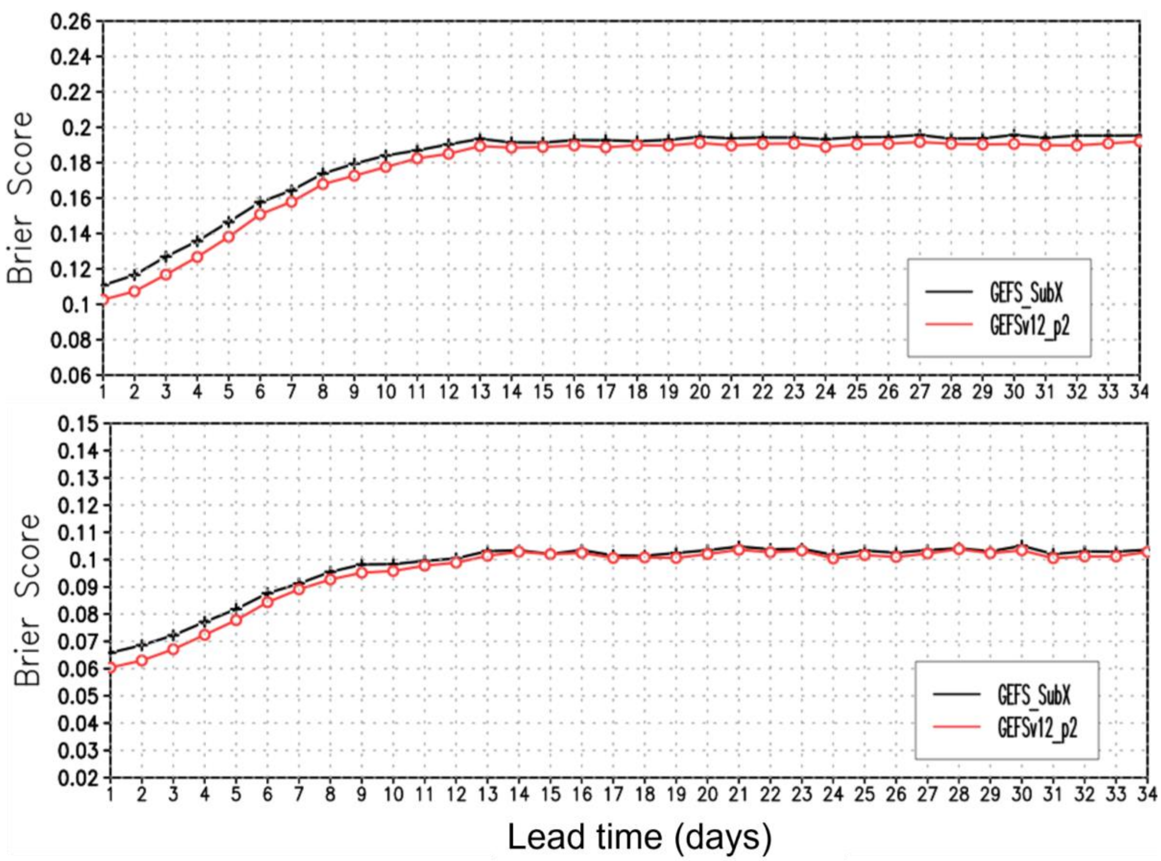

Figure 4. The daily average Brier Score of the CONUS probabilistic quantitative precipitation forecast (PQPF) from 2002 to 2019 for 24-h accumulated precipitation greater than or equal to 1.00mm (top) and 5.00mm (bottom). The comparison is for the GEFS_SubX reforecast (black) and GEFSv12_p2 reforecast (red) that were run once per week (Wednesday) with 11 members out to 35 days. The reference truth is CCPAv4.

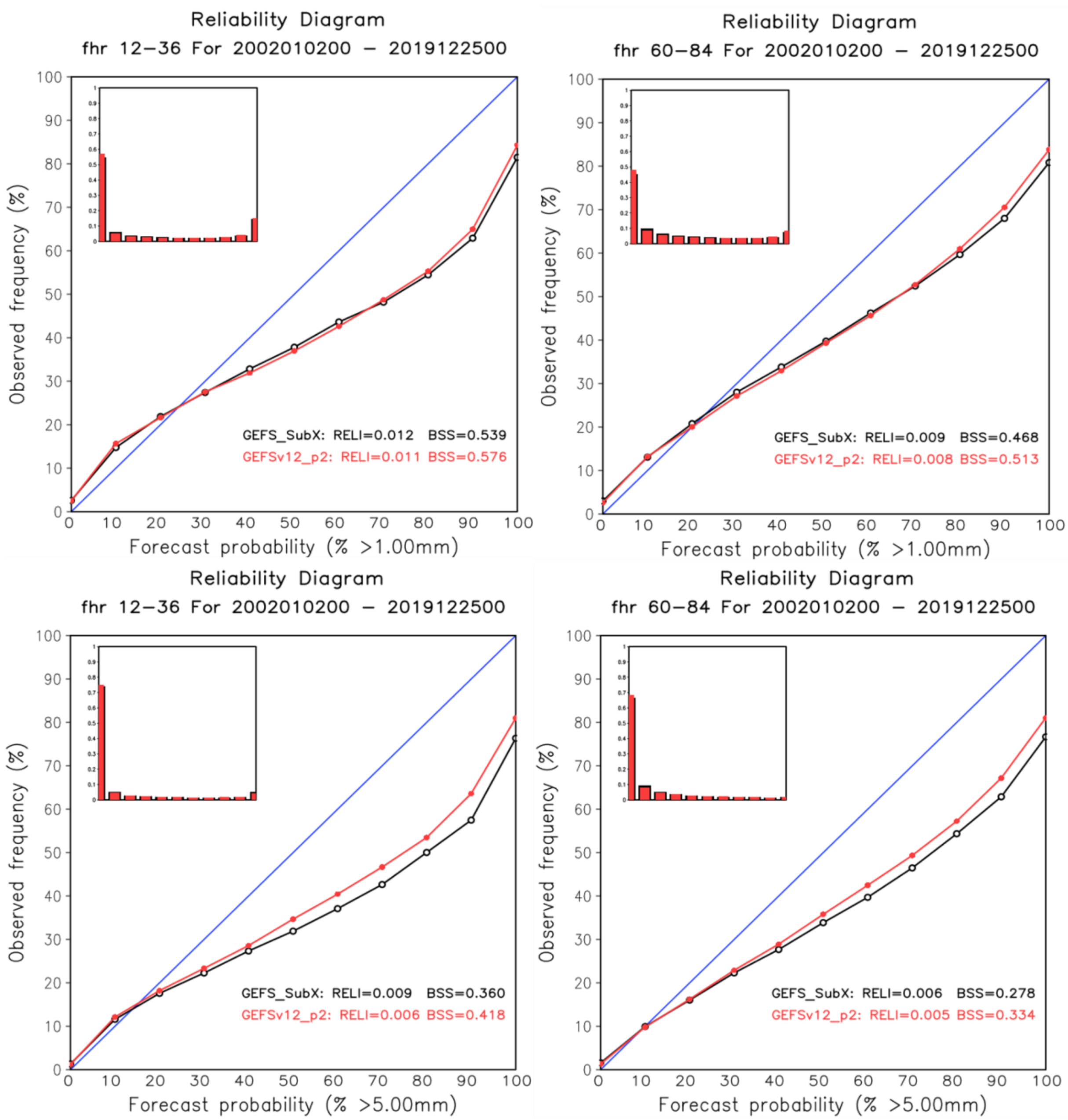

Figure 5. The reliability diagram of the CONUS probabilistic quantitative precipitation forecast (PQPF) from 2002 to 2019 for 24-h accumulated precipitation greater than or equal to 1.00mm (12-36 hours, top left; 60-84 hours, top right) and 5.00mm (12-36 hours, bottom left; 60-84 hours, bottom right). The comparison is for the GEFS SubX version reforecast (black) and GEFSv12_p2 reforecast (red) that run once per week (Wednesday) with 11 members out to 35 days. The reference truth is CCPAv4. The average reliability score (RELI) and Brier skill score (BSS) are also presented in each subplot. (Note: This is for a raw ensemble forecast with limited ensemble members (11) compared to the operational 31 members.)

d. MJO prediction skill

The newly operational GEFSv12 extended its output to +35 days lead to cover the sub-seasonal time scale. The MJO is one of the most important climate phenomena for sub-seasonal forecasts. Here we estimate MJO prediction skill using the real-time multivariate MJO (RMM) index (Wheeler and Hendon 2004) for the GEFS_SubX, GEFSv12_p1, and GEFSv12_p2 (Fig. 6). Skill is defined as the bivariate anomaly correlation between the analysis and forecast RMM1 and RMM2 index. For this comparison, the CFSR (GEFSv12 reanalysis) serves as the reference analysis for the GEFS_SubX and GEFSv12_p1 (GEFSv12_p2). In other words, the estimates are based on their own analysis data. Overall, the MJO forecast skill for the GEFSv12_p2 (~21.5 days) is similar to the GEFS_SubX and GEFSv12_p1 (~21 days) when using AC=0.5 as the threshold of useful skill. The SubX forecast skill for the 20-year sample in this study is also very comparable to the estimate (~21–22 days) that was made using a much smaller sample size (2 years) in Zhu et al. (2018) and Li et al. (2019). The GEFSv12_p2 also exhibits higher skill for shorter lead times (< ~18 days) than the GEFSv12_p1, possibly due to the benefit of the improved initial conditions for the Phase 2 reforecast. For lead times longer than 22 days, the forecast skill for all three sets of data is poor. A fully coupled atmosphere-ocean-wave-ice model, currently under development at NCEP, aims to improve the MJO forecast skill, especially for longer lead times. The reader is referred to Hamill and Kiladis (2013) for MJO verification on GEFSv10 reforecasts.

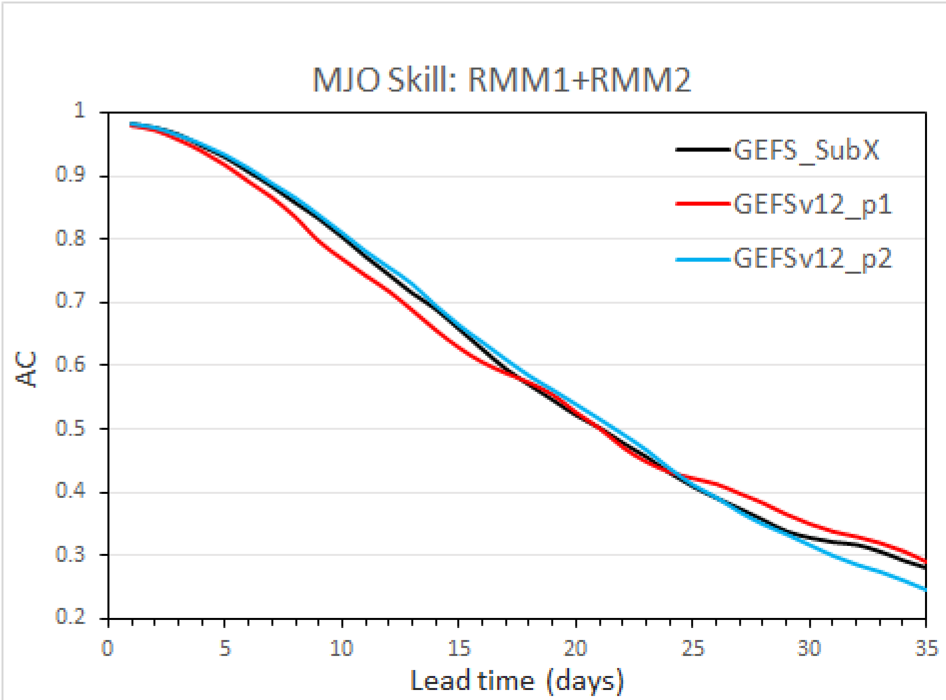

Figure 6. The real-time multivariate MJO (RMM) skill as a function of lead time for GEFS_SubX (black; 2000–2016), GEFSv12_p1 (red; 1989–1999), and GEFSv12_p2 (blue; 2000–2019) reforecasts.

e. 2-meter temperature errors

The January and July global 2-m temperature mean errors (or biases) for the 11-member runs were calculated for week 1, week 2, and weeks 3 and 4 during the GEFSv12_p1 (Fig. 7) and GEFSv12_p2 (Fig. 8) reforecast periods. The biases of week 1, week 2, and weeks 3 and 4 are the day 1–7, 8–14, and 15–28 averaged forecast errors over the corresponding forecast periods, respectively. Also displayed are the differences between CFSR and ERA5 (Fig. 7 a, b) and the differences between the GEFSv12 reanalysis and ERA5 (Fig. 8 a,b). The ERA5 was used as the reference for both phases to ensure a consistent comparison. A large warm bias over northern Asia is persistently seen in January (Figs. 7 a, c, e, and f and Figs. 8 a, c, e, and f) with a decreasing trend over increasing forecast lead time. In general, the error in 2-m temperature at the weeks 3 and 4 timescale is nearly saturated (Guan et al. 2019) and the impact from initial conditions decreases. At this timescale, the GEFSv12 generates a cold bias over North America (NA) in January (Figs. 7 g and 8 g). The cold bias locations are different, mostly over the eastern United States for GEFSv12_p1 and western Canada for GEFSv12_p2. A larger cold bias for the boreal winter season over the NA domain has been persistently observed in several generations of the NCEP GEFS (Guan et al. 2015, 2019) and was thought to be related to the imperfect parameterization of winter-associated physical processes (Guan et al. 2019).

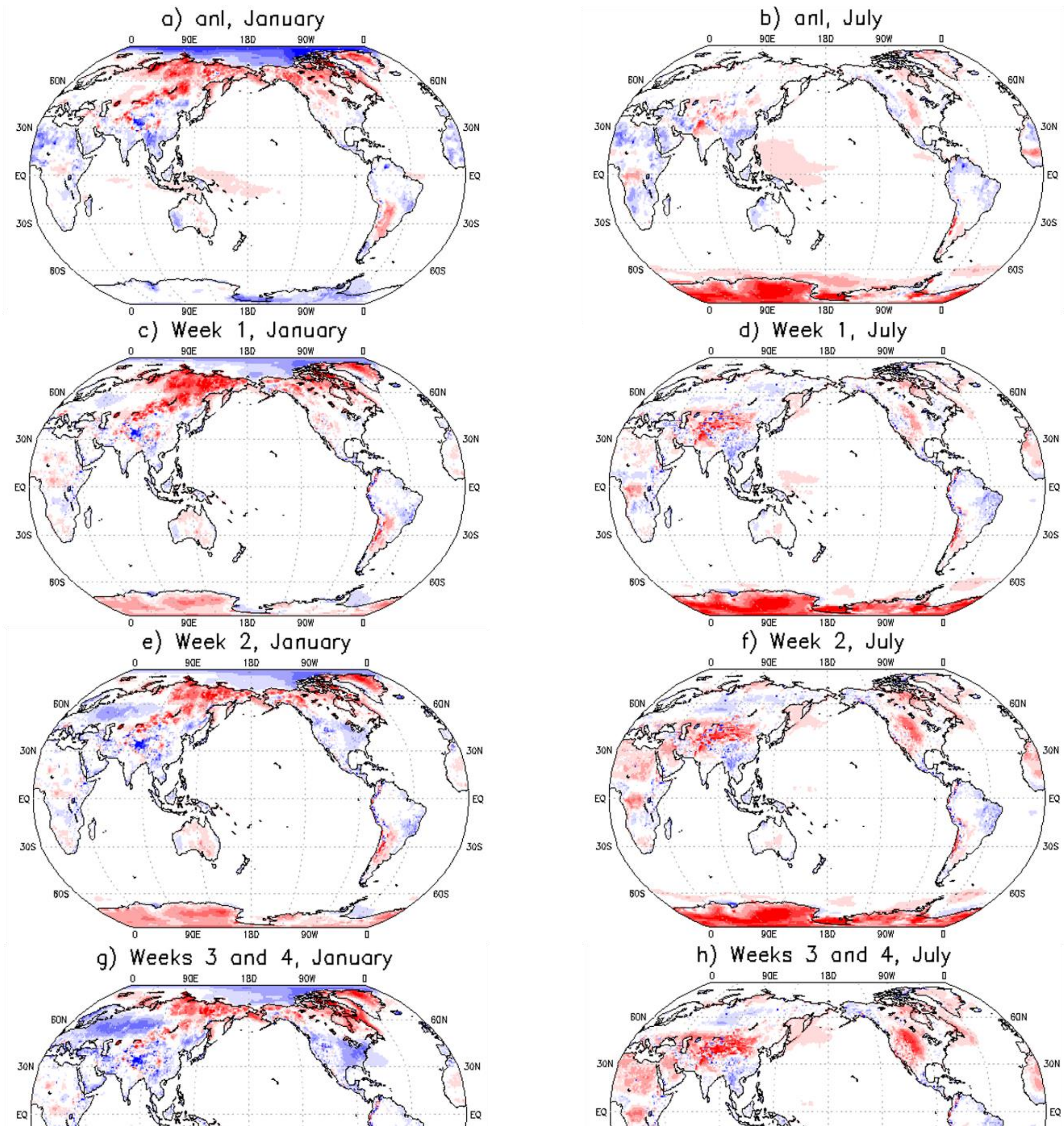

Figure 7. The difference in 2-m temperature (°C) between the CFSR and ERA5 for January (a) and July (b) over phase 1. Spatial distribution of 2-m temperature mean error (i.e., bias) over phase 1 for January during (c) week 1, (e) week 2, and (h) weeks 3 and 4 forecasts, and July during (d) week 1, (f) week 2, and (g) weeks 3 and 4 forecasts.

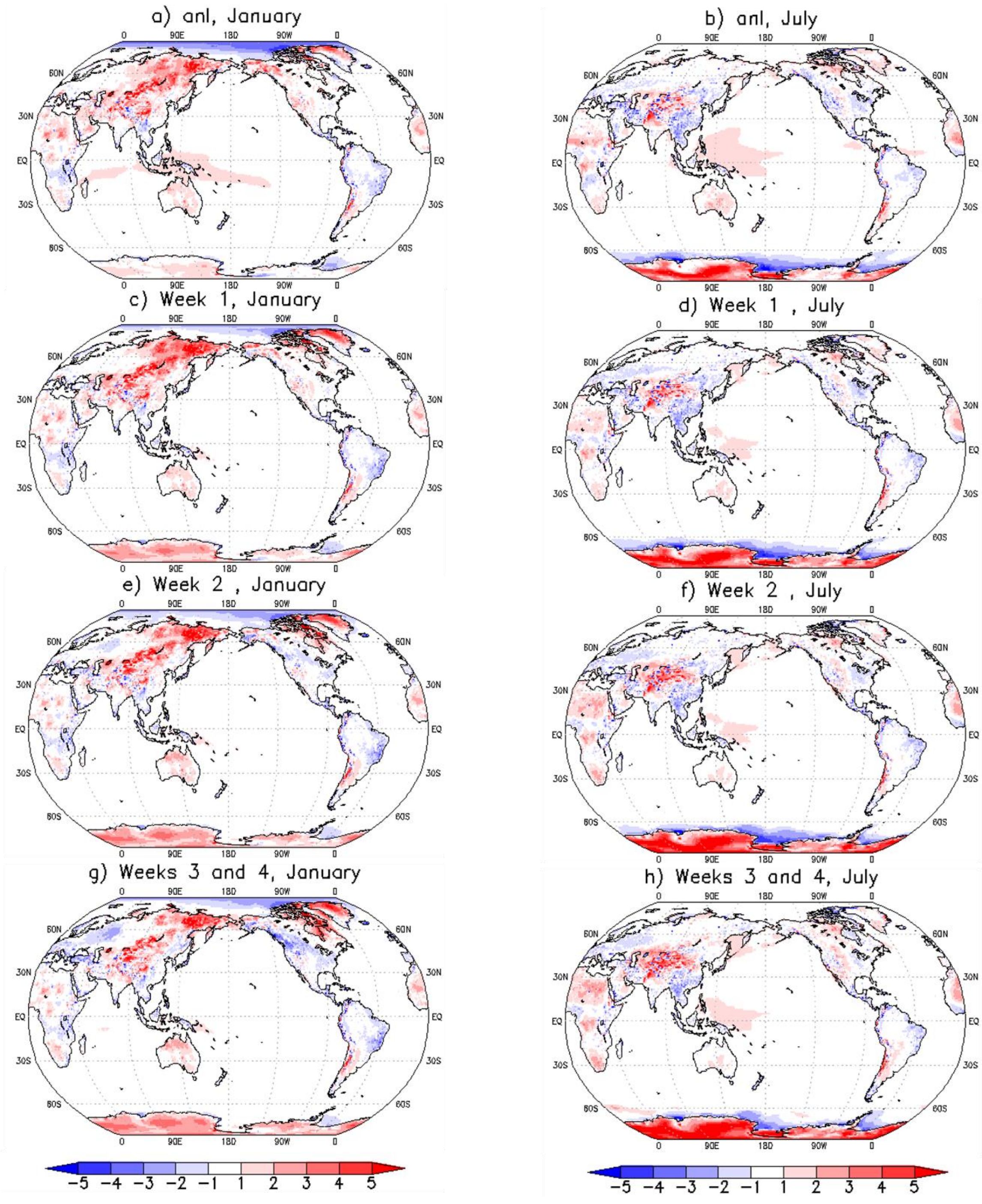

Figure 8. The difference in 2-m temperature (°C) between the GEFSv12 reanalysis and ERA5 for January (a) and July (b) over phase 2. Spatial distribution of 2-m temperature mean error (i.e., bias) over phase 2 for January during (c) week 1, (e) week 2, and (h) weeks 3 and 4 forecasts, and July during (d) week 1, (f) week 2, and (g) weeks 3 and 4 forecasts.

Snow is considered to be one of the most important wintertime land surface characteristics. To illustrate the influence of the snow forecast on bias characteristics, we compare the 2-m temperature bias over the NA domain for the 408h forecast (approximately the middle of week 3) with snow cover, without snow cover, and for all conditions (Fig. 9). The comparison was performed based on control members for the GEFSv12_p2 reforecast period. January–March is selected because those months show a consistently large cold bias (see red line in Fig.10) and the expected frequent occurrences of snow cover. The selection by individual members leads to a clear division between snow-covered and snow-free cases. The existence of forecast snow was inferred if the snow water equivalent is greater than or equal to 1mm. Clearly, the 2-m temperature bias characteristics are quite different between the two conditions (Fig. 9c and d). Figure 10 shows the time evolution of biases over a small region near the central US. A larger cold bias is dominant under the existence of snow cover with a domain-averaged value of -4.79°C during the GEFSv12_p2 period. In contrast, bias is much smaller under snow-free conditions where the average value is about -0.18°C. This indicates there is considerable room for improving the 2-m temperature forecast under snow-covered conditions. An improvement in modeling snow-associated physical processes would undoubtedly lead to a better 2-m temperature forecast. The large difference in bias characteristics between cases with and without snow cover also suggests that statistical calibration of 2-m temperature should be performed based on the existence of snow. It was noted that the bias correction using a unified 2-m temperature bias climatology for the NA cold season is much less efficient compared to the warm season (Guan et al., 2019). Apparently, the proposed snow dependent bias correction method should improve statistical post processing for the 2-m temperature forecast during the cold season. This will be confirmed in our future work.

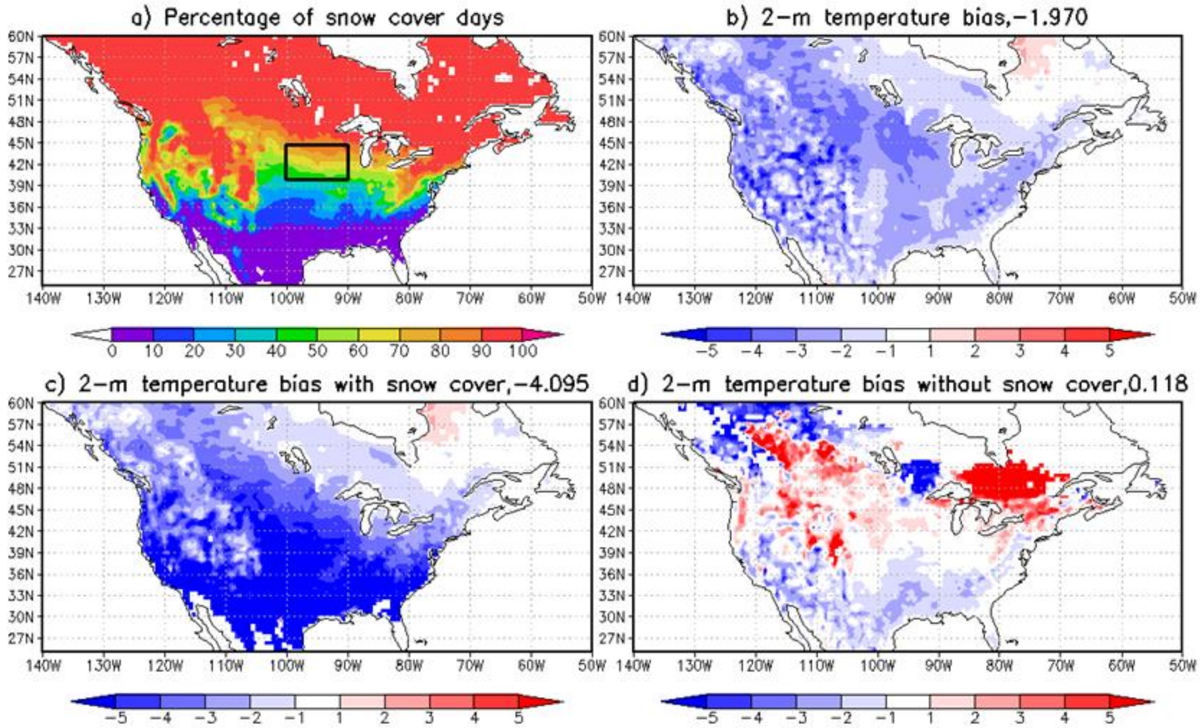

Figure 9. Percentage of snow cover days (a), 2-m temperature forecast bias under all conditions (b), bias with snow cover forecast (c), bias without snow cover forecast (d) for 408-h control-member forecast over NA. The results are based on the GEFSv12_p2 reforecast for January, February, and March.

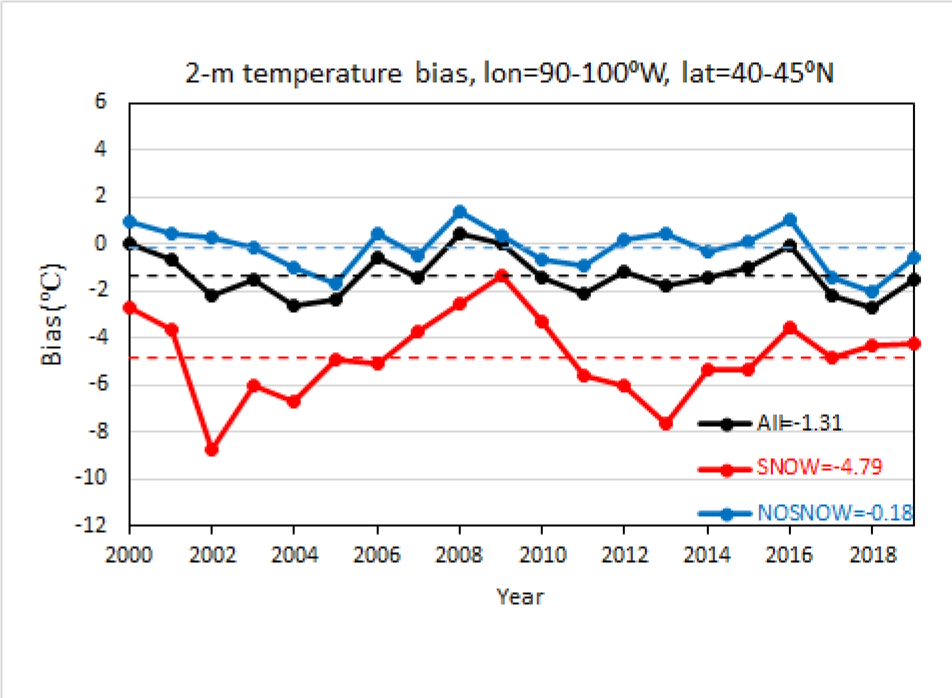

Figure 10. Time series of 2-m temperature forecast errors for 408-h control-member forecast over a small region (40°N-45°N, 90°W-100°W) near the central US (marked with the black rectangle in Fig. 9a). Black, red, and blue solid curves indicate the errors for January, February, and March forecasts under all conditions, with, and without snow cover, respectively. The corresponding dashed lines denote the averages over the entire period.

In contrast to January 2-m temperature biases in the initial state are relatively smaller in July (Figs. 7b and Figs. 8b). For weeks 3 and 4, the model showed a large warm bias over the central United States (US) during the GEFSv12_p1, which is consistent with the findings in Guan et al. (2019) though an earlier forecast system (i.e., GEFS_SubX) was used in that study. During the GEFSv12_p2, the model shows a bias pattern similar to the GEFSv12_p1, but the warm bias over the central US is reduced.

To better understand the impact of using different initial conditions and forecast systems to produce 2-m temperature forecasts, the seasonal variability of 2-m temperature bias is compared for the NA weeks 3 and 4 forecasts (land only) among the GEFS_SubX, GEFSv12_p1, and GEFSv12_p2 in Fig. 11. All three sets of reforecasts display a cold bias during the October-April and warm bias during the May-June period. The GEFS_SubX shows the strongest seasonal variability (or largest amplitudes) with a maximum cold bias of -1.8°C in March and warm bias of 1.5°C in June. When the forecast systems are the same (i.e., GEFSv12_p1 and GEFSv12_p2), the differences in 2-m temperature bias are relatively small. Overall, the GEFSv12_p1 is warmer than the GEFSv12_p2, except in December. The systematic difference during the July-September period is also noteworthy. Further diagnosis is needed to address this difference in the future.

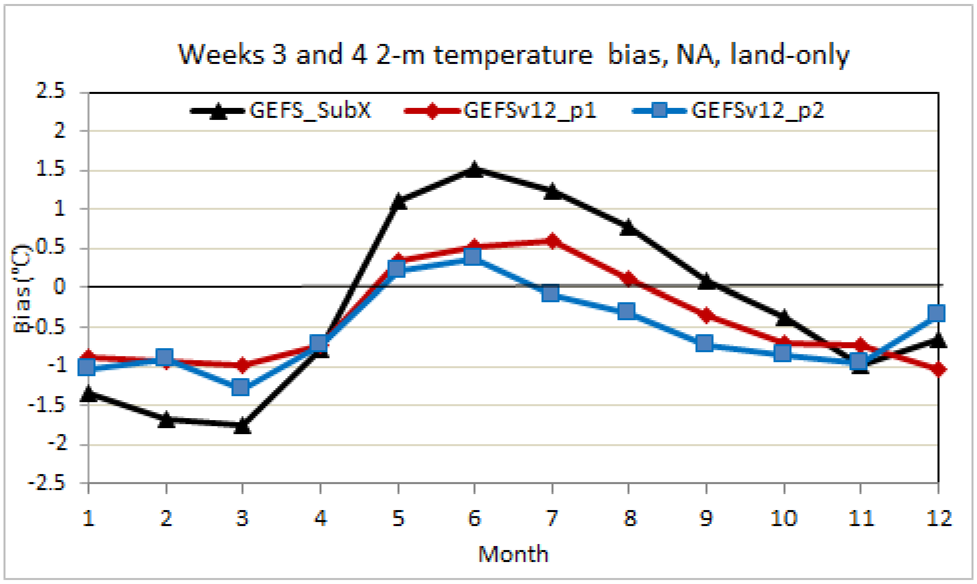

Figure 11. Weeks 3 and 4 biases in 2-m temperature forecasts averaged during the GEFS_SubX (black, 1999–2016), GEFSv12_p1 (red, 1989–1999), and GEFSv12_p2 (blue, 2000–2019) reforecast periods over NA, land-only.

5. Post-processing of reforecast (precipitation)

Calibration is one of the most common applications of a reforecast dataset. Precipitation is one of the most impactful weather elements (Hamill and Whitaker 2006; Hamill et al. 2008; Hamill 2012; Schmeits and Kok 2010; Hamill et al. 2015; Hamill and Scheuerer 2018; Scheuerer and Hamill 2018; Specq and Batté 2020). Here we demonstrate the impact of using reforecast data to improve precipitation forecasts.

a Methodology

We take advantage of long-term training data to calibrate precipitation through a quantile-mapping technique (Ines and Hansen 2006; Hamill and Scheuerer 2018). A 'quantile-based' bias correction approach, also referred to as 'histogram equalization' and/or 'rank matching' (Hamlet et al. 2002; Wood et al. 2004; Piani et al. 2010), is useful to statistically transform rainfall simulated by a model to bias corrected data.

In this study, the statistics of 24-hr accumulated rainfall for CCPA and GEFSv12 reforecasts were determined independently for each grid point and each lead times over CONUS. For simplicity, the 5-member ensemble means for Day-1, 5, 10, and 15 forecasts during the 2002-2019 period were used for this practice. The method can also be applied to the individual ensemble members. The corresponding sample size at each grid point and each lead time is 6574 days. The rainfall intensity distributions for both CCPA and GEFSv12 reforecasts are well approximated by the gamma distribution. The leave-one-out-cross-validation procedure has been implemented. For example, 2019 forecasts are trained using 2002-2018.

The bias-corrected procedure is to do a transformation between CCPA cumulative distribution function (CDF) and reforecast CDF, rather than explicitly to calculate bias. The formula for the calibration for a particular lead time (t) and grid (i, j) is expressed as follows:

$$Q_{bc}(i, j, t) = F_{CCPA}^{-1}(F_{GEFSv12}(Q_{raw}(i, j, t))) \quad (1)$$

The bias-corrected value (Q_{bc}) is the inverse of the CCPA CDF (F_{CCPA}^{-1}) at the probability corresponding to the reforecast CDF ($F_{GEFSv12}$) for a given raw forecast (Q_{raw}).

b. Application

Figures 12 and 13 demonstrate that both 24-h precipitation amounts and precipitation probability distributions in the calibrated forecast are more consistent with the CCPA than the raw forecasts. The bias correction dramatically reduces the wet bias over the entire CONUS (Fig. 12). For longer lead times (day 10 and day 15; Fig. 13), the raw forecast tends to underestimate the probability of precipitation less than ~7.5 mm/day and overestimate the corresponding value more than ~7.5 mm/day. After the calibration, the model curves overlap the observed curves for all lead times (Fig. 13). The calibration using long-term reforecast data is particularly important in improving the model climatology for the heavy precipitation events (> 50 mm) as illustrated in Figure 14. In the raw forecast, the model 24-h precipitation events exceeding 50 mm are substantially lower than the CCPA, especially for the longer lead times, when heavy (or extreme) precipitation events are completely missed for most of the domain. After the bias correction, both distributions and magnitudes in heavy precipitation events are much more consistent with the CCPA throughout all lead times.

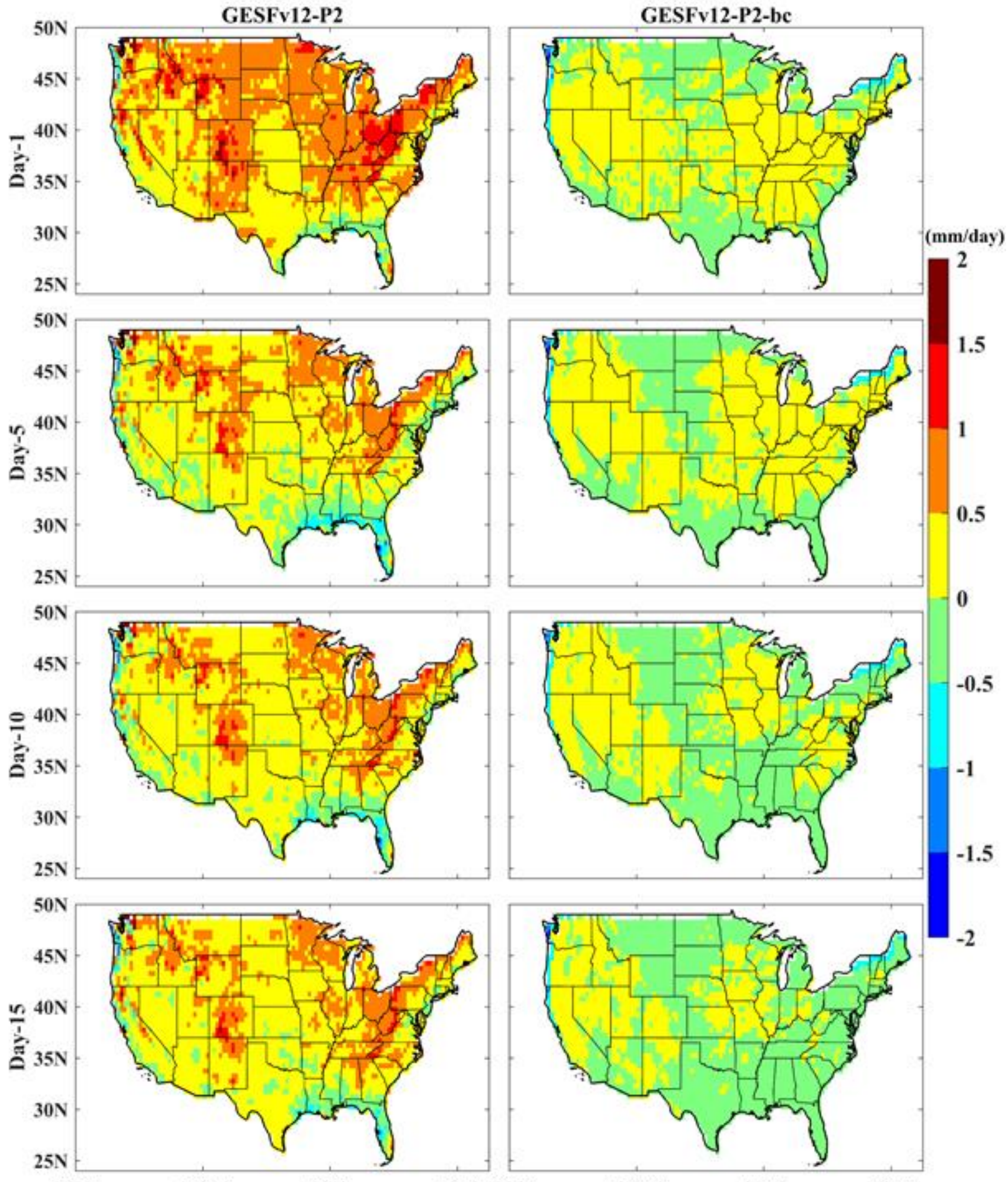

Figure 12. The Day-1, Day-5, Day-10, and Day-15 (Row-1 to 4, respectively) biases for 24-h precipitation from the (5 member) raw (GEFSv12_p2, left panels) and calibrated (GEFSv12_p2-bc, right panels) ensemble mean forecasts over the CONUS.

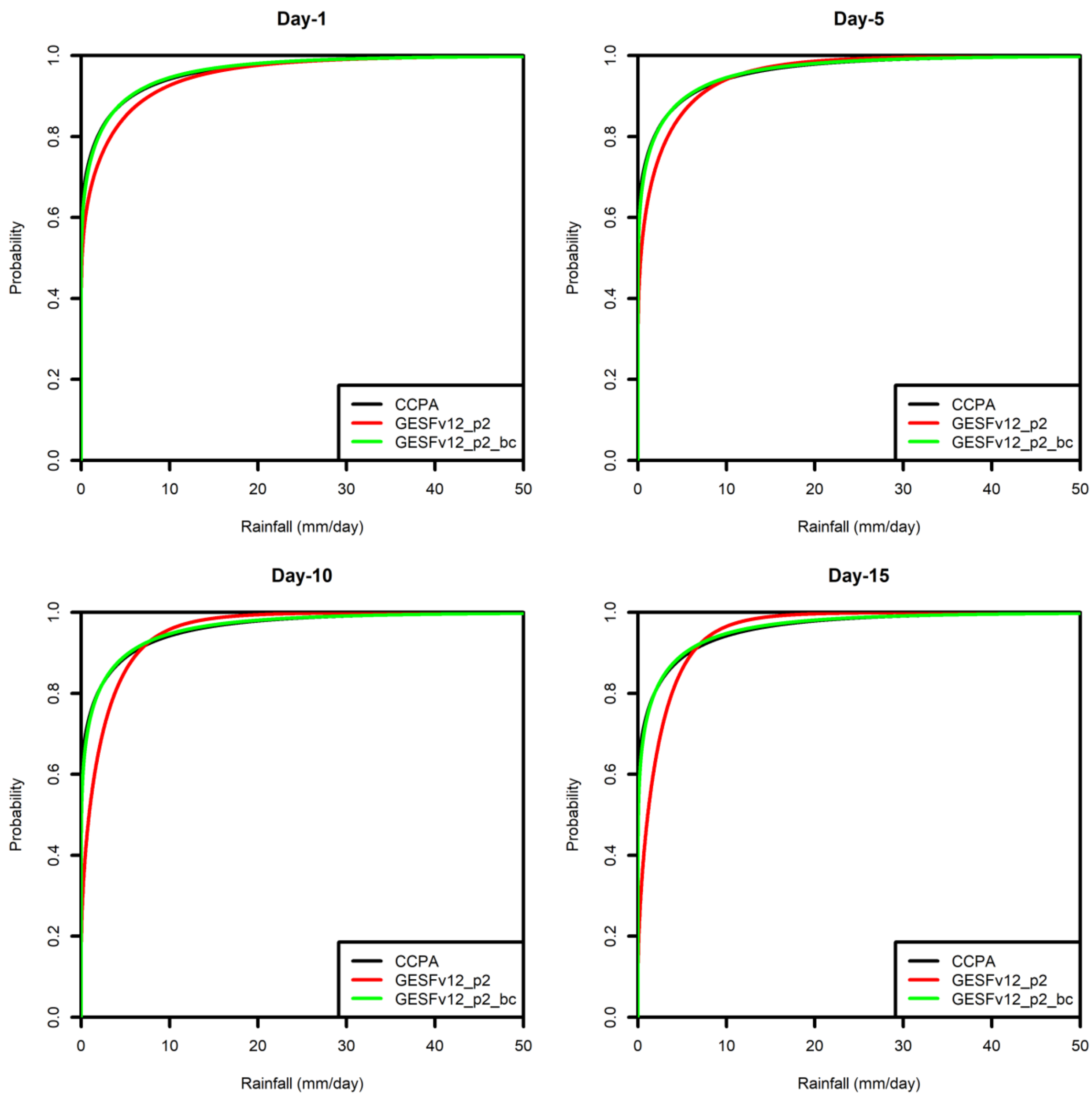

Figure 13. The Day-1, Day-5, Day-10, and Day-15 probability distributions of 24-h accumulated precipitation for CCPA (black lines), raw (red lines), and bias-corrected (green lines) 5-member ensemble mean forecasts over the full CONUS domain.

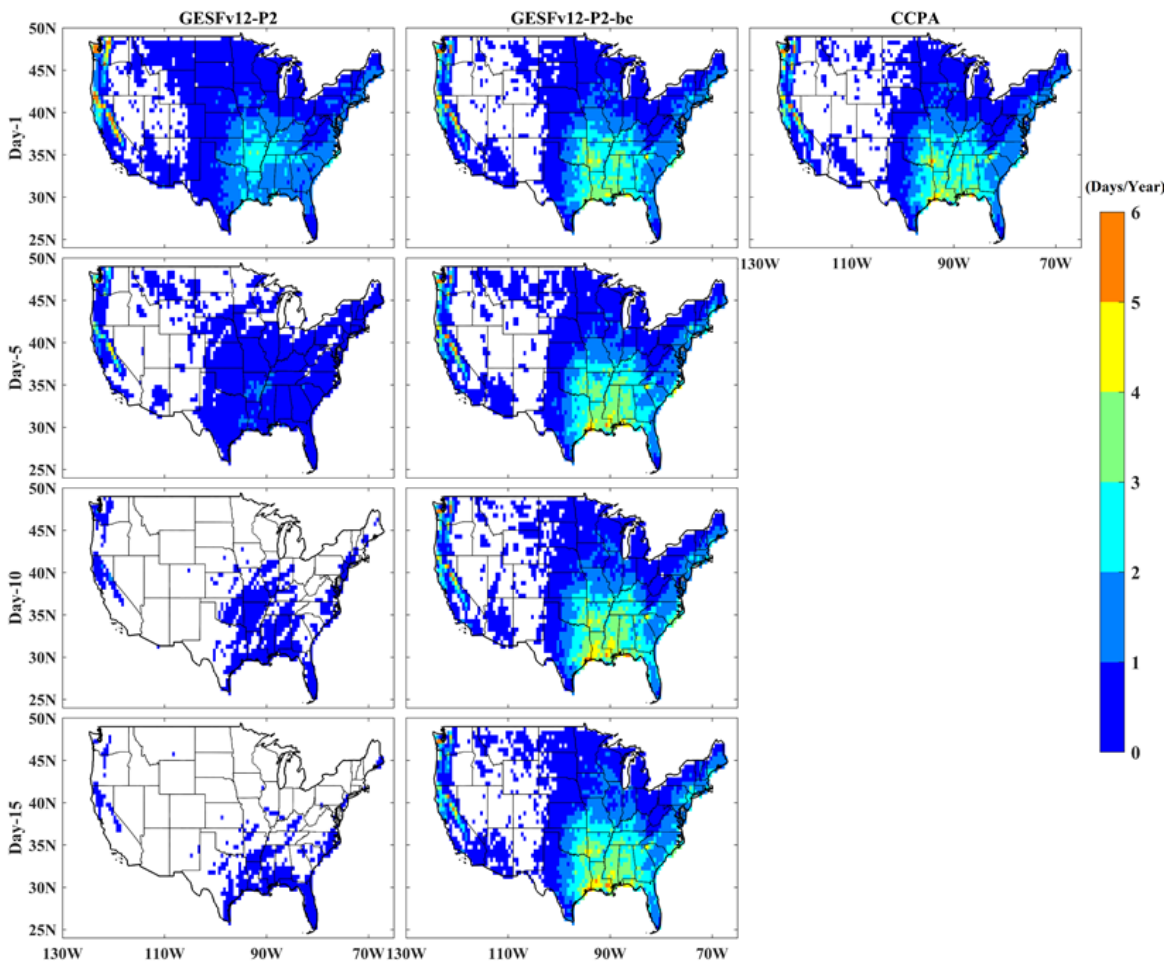

Figure 14. The days/year with 24-h precipitation exceeding 50 mm over the CONUS for raw (GEFSv12_p2, Column 1), bias-corrected (GEFSv12_p2-bc, Column 2) 5-member ensemble mean forecasts for Day-1, Day-5, Day-10 and Day-15 and CCPA (Column 3).

6. Summary

For the first time, the simultaneous generation of a multi-decade reanalysis and reforecast dataset became part of an operational GEFS implementation. The reforecast dataset is particularly important, considering the extension to subseasonal forecast time scale in the current GEFSv12. Statistical postprocessing with a long-term training sample of the reforecast has become a routine part of making subseasonal operational outlooks due to the larger forecast errors that exist at longer lead times. The dataset is being used to support several stakeholders in developing their operational products across many time scales. This large volume dataset is easily accessible by both the stakeholders and public users from the NCEP local machines and two public websites. Doubtlessly, this will further facilitate analysis and contributions to model developments.

The performance of several selected weather elements, hurricane track, and MJO in the GEFSv12 reforecast were compared with the GEFS_SubX and GEFSv10 reforecasts. The error characterization of the 2-m temperature forecast was analyzed. Overall, the forecast skill for the GEFSv12 is similar to or better than the GEFS_SubX in 500-hpa geopotential height, precipitation, and MJO forecasts. It is also worth mentioning that the degree of some of these improvements is less than those resulting from the change from the GEFSv11 to GEFS_SubX. It should be emphasized that when the GEFS_SubX was developed, considerable efforts were made to enhance the stochastic physics, surface boundary conditions and convection. These model enhancements resulted in substantial improvements in model performance compared to the GEFSv11 (Zhu et al. 2018; Li et al. 2019, Guan et al. 2019). Therefore, when using GEFS_SubX as a benchmark to evaluate GEFSv12_p2, it should be noted that the GEFS_SubX is a difficult model to outperform substantially. The two sets of nearly three decades of reforecast data (GEFSv10 and GEFSv12) provide a good opportunity to address the impacts of the model and analysis on hurricane track forecasts. The initial analysis plays an important role in the accuracy of the track forecast for lead times shorter than about 5 days. The improvement in the model itself may be a potential direction to take in reducing the track forecast error for lead times longer than 5 days, which is a persistent challenge for the NCEP GEFS.

In comparison with the GEFS_SubX, the GEFSv12 substantially reduces the warm (cold) bias over the NA domain during the boreal warm (cold) season. However, the cold bias for the cold season in the GEFSv12 is still considerable. Further analysis of the error characteristics demonstrates that this bias is snow-dependent, emphasizing the importance of 2-m temperature calibration for GEFSv12 based on the existence of snow cover. The multi-decadal reforecast dataset was also demonstrated to be very useful in calibrating the precipitation and capturing extreme precipitation events.

Acknowledgements

We thank Dr. Hui-Ya Zhuang and Wesley Ebisuzaki and Leigh Zhang for providing valuable helps on the NCEP post package and wgrib2 utility. We also acknowledge Dr. Yan Luo for supplying the CCPA data. We are grateful to Dr. Mark Fresch, James Ward, and Yuqiong Liu in the National Water Center for their careful reviews of the reforecast's precipitation dataset. Special appreciation goes to Mary Hart for editing the English and Lydia Stefanova and Weizhong Zheng for giving the good comments and suggestions through the internal review process in EMC. This study is partially supported through NOAA Weather Program Office grant T8R1CRS-PCR, NOAA Climate Program Office grants S8R1CWS-PRE and R8R1CWS-P01, NWS OSTI grants P8MWQNG-PTR, P8R1MP1-PPS, and M8M6H81-P10, and OAR/WPO-NWS-EMC-SLA. Jonathan Brannock of the North Carolina Institute for Climate Studies assisted with the transfer of reforecast data to Amazon Web Services under NOAA Big Data Program.

Data availability statement

The GEFSv12 Phase-2 reforecast data are openly available at NOAA/NWS/NCEP's ftp server (ftp://ftp.emc.ncep.noaa.gov/GEFSv12/reforecast) and Amazon Web Services (AWS, https://noaa-gefs-retrospective.s3.amazonaws.com/index.html). The 22 variables for the Phase 1 reforecast are also openly available at NOAA/NWS/NCEP's ftp server (ftp://ftp.emc.ncep.noaa.gov/GEFSv12/reforecast).

REFERENCES

Alpert, J., 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. p726–733, Baltimore, MD, Eighth Conf. on Numerical Weather Prediction, Amer. Meteor. Soc.

Baxter, M. A., G. M. Lackmann, K. M. Mahoney, T. E. Workoff, and T. M. Hamill, 2014: Verification of precipitation reforecasts over the Southeast United States. *Wea. Forecasting*, **29**, 1199-1207.

Berner, J., and others, 2017: Stochastic Parameterization: Toward a New View of Weather and Climate Models, *Bulletin of the American Meteorological Society*, *98*(3), 565-588. Retrieved Aug 31, 2021, from https://journals.ametsoc.org/view/journals/bams/98/3/bams-d-15-00268.1.xml

Bloom, S. C., L. L. Takacs, A. M. Da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256–1271, doi:10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3.

Buizza, R., M. Miller, and T. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, 125, 2887–2908, https://doi.org/10.1002/qj.49712556006.

Candille, G., 2009: The multiensemble approach: The NAEFS example. *Mon. Wea. Rev.*, 137, 1655–1665, https://doi.org/10.1175/2008MWR2682.1.

—, S. Beauregard, and N. Gagnon, 2010: Bias correction and multiensemble in the NAEFS context or how to get a “free calibration” through a multiensemble approach. *Mon. Wea. Rev.*, 138, 4268–4281, https:// doi.org/10.1175/2010MWR3349.1.

Chen, F., K. Mitchell, J. Schaake, Y. Xue, H.-L. Pan, V. Koren, Q.-Y. Duan, M. Ek, and A. Betts, 1996: Modeling of land-surface evaporation by four schemes and comparison with FIFE observations, *J. Geophys. Res.*, 101(D3), 7251– 7268, doi:10.1029/95JD02165.

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, 3299–3310.

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, 233–244, doi:10.1016/j.jqsrt.2004.05.058.

Demargne, J., Wu, L., Regonda, S. K., Brown, J. D., Lee, H., He, M., Seo, D., Hartman, R., Herr, H. D., Fresch, M., Schaake, J., & Zhu, Y., 2014: The Science of NOAA's Operational Hydrologic Ensemble Forecast Service, *Bulletin of the American Meteorological Society*, *95*(1), 79-98. Retrieved Aug 31, 2021, from https://journals.ametsoc.org/view/journals/bams/95/1/bams-d-12-00081.1.xml

Dorman, J. L., and P. J. Sellers, 1989: A global climatology of albedo, roughness length and stomatal resistance for atmospheric general circulation models as represented by the Simple Biosphere model (SiB), *J. Appl. Meteorol.*, 28, 833– 855, doi:10.1175/1520-0450(1989)028<0833:AGCOAR>2.0.CO;2.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land-surface model advances in the NCEP operational mesoscale Eta model. J. Geophys. Res., 108, 8851, doi:10.1029/2002JD003296.

Ek, M. B., Y. Xia, J. Meng, R. Shresha, H, Wei, J. Dong, Y. Wu, and W. Zheng, 2016; Land Prediction in NCEP Modeling Systems: Current Status and Future Plans, weather.gov/media/sti/nggps/10_Ek%20EMC-land-NGGPS-Feb2016.pdf.

EMCs, 2019: GFS v15 implementation in June 2019. Ref: https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs.php.

Emerton, R., E. Zsoter, L. Arnal, H. L. Cloke, D. Muraro, C. Prudhomme, E.M. Stephens, P. Salamon, and F. Pappenberger, 2018: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0. *Geosci. Model Dev.*, **11**, 3327–3346. DOI: https://doi.org/10.5194/gmd-11-3327-2018.

Galarneau, T. J., Jr., and Hamill, T. M., 2015: Diagnosis of track forecast errors for tropical cyclone Rita, 2005: Using GEFS reforecasts, *Weather and Forecasting*, **30**(5), 1334-1354. Retrieved Aug 31, 2021, from https://journals.ametsoc.org/view/journals/wefo/30/5/waf-d-15-0036_1.xml

Gascon, E., D. Lavers, D., T. M. Hamill, D. S. Richardson, Z. Ben Bouallegue, M. Leutbecher, and F. Pappenberger, 2019: Statistical post-processing of dual-resolution ensemble precipitation forecasts across Europe. *Quart. J. Royal Meteor. Soc.*, 1-18, https://doi.org/10.1002/qj.3615.

Guan, H., B. Cui, and Y. Zhu, 2015: Improvement of statistical postprocessing using GEFS reforecast information. *Wea. Forecasting*, **30**, 841–854, https://doi.org/10.1175/WAF-D-14-00126.1.

Guan, H., and Y. Zhu, 2017: Development of verification methodology for extreme weather forecasts. *Wea. Forecasting*, **32**, 479–491, https://doi.org/10.1175/WAF-D-16-0123.1.

Guan, H., Y. Zhu, E. Sinsky, W. Li, X. Zhou, D. Hou, C. Melhauser, and R. Wobus, 2019: 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.

Hagedorn, R., 2008: Using the ECMWF reforecast dataset to calibrate EPS forecasts. ECMWF Newsletter 117, 8-13.

Hagedorn, R., R. Buizza, T. M. Hamill, M. Leutbecher, and T. N. Palmer, 2012: Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF ensemble forecasts. *Quart J. Royal Meteor Soc.*, **138**, 1814-1827.

Hagedorn, R., T. M. Hamill, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part I: 2-meter temperature. *Mon. Wea. Rev.*, **136**, 2608-2619.

Hamill, T. M., and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Mon. Wea. Rev.*, **134 ,** 3209–3229.

Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: Reforecasts, an important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**,33-46.

Hamill, T. M., J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**, 1434–1447, https://doi.org/10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2.

Hamill, T. M., R. Hagedorn, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: Precipitation. *Mon. Wea. Rev.*, **136**, 2620–2632.

Hamill, T. M., 2012: Verification of TIGGE Multi-model and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the conterminous US. *Mon. Wea. Rev.*, **140**, 2232-2252.

Hamill T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Jr., Y. Zhu, and W. Lapenta, 2013: NOAA's Second-Generation Global Medium-Range Ensemble Reforecast Data Set. *Bull. Amer. Meteor. Soc.*, **94**(10), 1553-1565.

Hamill, T. M., and G. N. Kiladis, 2013: Skill of the MJO and Northern Hemispheric blocking in GEFS medium-range reforecasts. *Mon. Wea. Rev*., **142**, 686-885.

Hamill, T. M., M. Scheuerer, and G. T. Bates, 2015: Analog probabilistic precipitation forecasts using GEFS Reforecasts and Climatology-Calibrated Precipitation Analyses._ *Mon. Wea. Rev.*, **143**, 3300-3309.

Hamill, T. M., 2017: Changes in the systematic errors of global reforecasts due to an evolving data assimilation system. *Mon. Wea. Rev.*, **145**, 2479-2485.

Hamill, T. M., and Scheuerer, M., 2018: Probabilistic precipitation forecast postprocessing using quantile mapping and rank-weighted best-member dressing. *Mon. Wea. Rev.*, 146, 4079-4098.

Hamill, T. M., and Coauthors, 2021: The Reanalysis for the Global Ensemble Forecast System, version 12. *Mon. Wea. Rev.*, conditionally accepted.

Hamlet, A. F., Huppert, D., and Lettenmaier, D. P., 2002: The economic value of long-lead streamflow forecasts for Columbia River hydropower. Journal of Water Resources Planning and Management, 128(2):91–101, https://doi.org/10.1061/(ASCE)0733-9496(2002)128:2(91).

Han, J., and Pan, H.-L., 2011: Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. *Wea. Forecasting*, 26(4), 520– 533.

Han, J., W. Wang, Y. C. Kwon, S.-Y. Hong, V. Tallapragada, and F. Yang, 2017: Updates in the NCEP GFS cumulus convection schemes with scale and aerosol awareness. *Wea. Forecasting*, 32, 2005–2017, https://doi.org/10.1175/WAF-D-17-0046.1.

Han, J., M. Witek, J. Teixeira, R. Sun, H.-L. Pan, J. K. Fletcher, and C. S. Bretherton, 2016: Implementation in the NCEP GFS of a Hybrid Eddy-Diffusivity Mass-Flux (EDMF) boundary layer parameterization with dissipative heating and modified stable boundary layer mixing. *Wea. Forecasting*, **31**, 341–352, https://doi.org/10.1175/WAF-D-15-0053.1.

Harris, L. M., & Lin, S. J. 2013: A two-way nested global-regional dynamical core on the cubed-sphere grid. *Mon. Wea. Rev., 141(1), 283– 306.*

Hou, D., M. Charles, Y. Luo, Z. Toth, Y. Zhu, R. Krzysztofowicz, Y. Lin, P. Xie, D-J. Seo, M. Pena and B. Cui, 2014: Climatology-Calibrated Precipitation Analysis at Fine Scales: Statistical Adjustment of Stage IV toward CPC Gauge-Based Analysis. *J. Hydrometeorology*, 15, 2542–2557.

Ines AVM and Hansen J W., 2006: Bias correction of daily GCM rainfall for crop simulation studies. *Agricultural and Forest Meteorology* 138: 44– 53.

Koren, V., J. Schaake, K. Mitchell, Q.-Y. Duan, F. Chen, and J. Baker, 1999: A parameterization of snowpack and frozen ground intended for NCEP weather and climate models, *J. Geophys. Res.*, 104(D16), 19,569– 19,585, doi:10.1029/1999JD900232.

Lalaurette, F., 2003: Early detection of abnormal weather conditions using a probabilistic extreme forecast index. Quart. J. Royal Meteorol. Soc., **129**, 3037-3057. https://doi.org/10.1256/qj.02.152

Landsea, C. W., and J. P. Cangialosi, 2018: Have we reached the limits of predictability for tropical cyclone track forecasting? *Bull. Meteor. Amer. Soc.,* 99, 2237–2243, https://doi.org/10.1175/ BAMS-D-17-0136.1.

Li, W., Y. Zhu, X. Zhou, D. Hou, E. Sinsky, C. Melhauser, M. Pena, H. Guan and R. Wobus, 2019: Evaluating the MJO forecast skill from different configurations of NCEP GEFS extended forecast. *Clim. Dyn.*, https://doi.org/10.1007/s00382-018-4423-9.

Lin, S.-J., 2004: A "vertically Lagrangian" finite-volume dynamical core for global models. *Mon. Wea. Rev.*, 132, 2293–2307.

Lin, S.-J., and R. B. Rood, 1997: An explicit flux-form semi-Lagrangian shallow-water model on the sphere, *Q. J. R. Meteor. Soc.*, 123:2477-2498.

Long, P. J., 1984: An general unified similarity theory for the calculation of turbulent fluxes in the numercal weather prediction models for unstable condition. Office Note 302, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, National Meteorological Center, 30 pp.

——, 1986: An economical and compatible scheme for parameterizing the stable surface layer in the medium-range forecast model. Office Note 321, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, National Meteorological Center, 24 pp.

Luo Y., Y. Zhu, D. Hou, Y. Lin, P. Xie, 2018: The NCEP's Climatology-Calibrated Precipitation Analysis (CCPA) Version 4. *In the 98th American Meteorological Society Annual Meeting, AMS.*

Mitchell, K. E., H. Wei, S. Lu, G. Gayno, and J. Meng, 2005: NCEP implements major upgrade to its medium-range global forecast system, including land-surface component, *GEWEX News* 15(4), 8–9. [Available at http://www.gewex.org/Nov2005.pdf.]

Nardi, K. M., E. A. Barnes and F. M. Ralph, 2018: Assessment of numerical weather prediction model reforecasts of the occurrence, intensity, and location of atmospheric Rivers along the west coast of North America. *Mon. Wea. Rev.*, 146, 3343– 3362. https://doi.org/10.1175/MWR-D-18-0060.1.

Noh, Y.-C.; Lim, A.H.N.; Huang, H.-L.; Goldberg, M. D., 2020: Global Forecast Impact of Low Data Latency Infrared and Microwave Sounders Observations from Polar Orbiting Satellites. *Remote Sens.* **2020**, *12*, 2193. https://doi.org/10.3390/rs12142193

Ou, M., M. Charles, and D. Collins, 2016: Sensitivity of calibrated week-2 probabilistic forecast skill to reforecast sampling of the NCEP Global Ensemble Forecast System. *Wea. Forecasting*, **31**, 1093–1107, https://doi.org/10.1175/WAF-D-15-0166.1.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parameterization in weather and climate prediction. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304, doi:10.1002/qj.49712757202

Palmer, T. N. Buizza, F. Doblas Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. *ECMWF Tech. Memo*. 598, 42 pp., http://www.ecmwf.int/publications/.

Palmer, T. N., 2012: Towards the probabilistic Earth-system simulator: A vision for the future of climate and weather prediction. *Quart. J. Roy. Meteor. Soc.*, **138**, 841–861, doi:10.1002/qj.1923.

Pegion, K., and Coauthors, 2019: The Subseasonal Experiment (SubX): A multi-model subseasonal prediction experiment. *Bull. Amer. Meteor. Soc.*, 100, 2043–2060. https://doi.org/10.1175/BAMS-D-18-0270.1

Piani, C., J. Haerter, and E. Coppola, 2010: Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.*, 99(1-2), 187-192, https://doi.org/10.1007/s00704-009-0134-9

Putman, W. M., S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. J. Computational Phys., 227, 55-78, https://doi.org/10.1016/j.jcp.2007.07.022.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D.C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609-1625.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, 91, 1015.1057. doi: 10.1175/2010BAMS3001.1.

Schmeits, M. J., K. J. Kok, 2010: A Comparison between Raw Ensemble Output, (Modified) Bayesian Model Averaging, and Extended Logistic Regression Using ECMWF Ensemble Precipitation Reforecasts, *Mon. Wea. Rev.*, vol. 138, 4199–4211.

Scheuerer, M., and T. M. Hamill, 2015: Statistical post-processing of ensemble precipitation forecasts by fitting censored, shifted Gamma distributions. *Mon. Wea. Rev.*, **143**, 4578-4596. Also appendix A and appendix B and appendix C.

Scheuerer, M., and Hamill, T.M., 2018: Generating calibrated ensembles of physically realistic, high-resolution precipitation forecast fields based on GEFS model output. *J. Hydrometeorology,* **19** (10), 1651-1670. https://journals.ametsoc.org/doi/abs/10.1175/JHM-D-18-0067.1

Scheuerer, M., M. B. Switanek, R. P. Worsnop, and T. M. Hamill, 2020: Using Artificial Neural Networks for Generating Probabilistic Subseasonal Precipitation Forecasts over California. *Mon. Wea. Rev.*, **148**, 3489-3506, https://doi.org/10.1175/MWR-D-20-0096.1.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Q.J. Roy. Meteor. Soc.*, 131, 3079–3102, https://doi.org/10.1256/qj.04.106.

Shutts, G. and T. N. Palmer, 2004: The use of high-resolution numerical simulations of tropical circulation to calibrate stochastic physics schemes. Proc. Workshop on Simulation and Prediction of Intra-Seasonal Variability with Emphasis on the MJO, Reading, United Kingdom, ECMWF, 83-102, https://www.ecmwf.int/en/learning/workshops-and-seminars/past-workshops/2003-simulation-prediction-intra-seasonal-variability.

Specq, D., L. Batté, 2020: Improving subseasonal precipitation forecasts through a statistical–dynamical approach: application to the southwest tropical Pacific. *Clim. Dyn.* 55, 1913–1927. https://doi.org/10.1007/s00382-020-05355-7.

Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus*, 60A, 62–79, https://doi.org/10.1111/j.1600-0870.2007.00273.x.

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<1917:AARMMI>2.0.CO;2.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press.

Wilks, D. S., and T. M. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev.*, **135**, 2379-2390.

Wood, A. W., L. R. Leung, V. Sridhar, and D. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic change*, 62(1-3), 189-216, https://doi.org/10.1023/B:CLIM.0000013685.99609.9e.

Zheng, W., H. Wei, Z. Wang, X. Zeng, J. Meng, M. Ek, K. Mitchell, and J. Derber, 2012: Improvement of daytime land surface skin temperature over arid regions in the NCEPGFS model and its impact on satellite data assimilation. J. Geophys. Res., 117, D06117, doi:10.1029/2011JD015901.

Zheng, W., Ek, M., Mitchell, K., Wei, H., and Meng, J., 2017: Improving the stable surface layer in the NCEP global forecast system. Monthly Weather Review, 145, 3969– 3987. https://doi.org/10.1175/mwr-d-16-0438.1.

Zhou, X., Y. Zhu, B. Fu, D. Hou, J. Peng, Y. Luo and W. Li, 2019: The development of Next NCEP Global Ensemble Forecast System. *STI Climate Bulletin*, 159-163.

Zhou, X., Y. Zhu, D. Hou, B. Fu, W. Li, H. Guan, E. Sinsky, W. Kolczynski, X. Xue, Y. Luo, J. Peng, B. Yang, V. Tallapragada and P. Pegion, 2021: The introduction of the NCEP Global Ensemble Forecast System Version 12. Submitted to *Wea. Forecasting*.

Zhu, Y, D. Hou, M. Wei, R. Wobus, J. Ma, B. Cui, and S. Moorthi, 2012: GEFS upgrade (v10) at NWS/NCEP/Environmental Modeling Center. [Available online at http://www.emc.ncep.noaa.gov/gmb/yzhu/html/imp/201109_imp.html.]

Zhu, Y., X. Zhou, M. Pena, 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, 2159–2173, https://doi.org/10.1175/WAF-D-17-0093.1.

Zhu, Y., and Coauthors, 2018: Toward the improvement of sub-seasonal prediction in the NCEP Global Ensemble Forecast System (GEFS). *J. Geophys. Res.: Atmos.*, **123**, 6732–6745, https://doi.org/10.1029/2018JD028506.

Zobler, L., 1986: A world soil file for global climate modelling, *NASA Tech. Memo.*, TM-87802, 33 pp.

Zobler, L., 1999: Global soil types, 1-degree grid (Zobler), http://www.daac.ornl.gov, Distrib. Active Arch. Cent. for Biogeochem. Dyn., Oak Ridge Natl. Lab., Oak Ridge, Tenn., doi:10.3334/ORNLDAAC/418.