

Impact of Sea Surface Temperature Forcing on Weeks 3 and 4 Forecast Skill in the NCEP Global Ensemble Forecasting System

YUEJIAN ZHU

NOAA/NWS/NCEP/EMC, College Park, Maryland

XIAQIONG ZHOU

IMG, NOAA/NWS/NCEP/EMC, College Park, Maryland

MALAQUIAS PEÑA

Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut

WEI LI AND CHRISTOPHER MELHAUSER

IMG, NOAA/NWS/NCEP/EMC, College Park, Maryland

DINGCHEN HOU

NOAA/NWS/NCEP/EMC, College Park, Maryland

(Manuscript received 6 July 2017, in final form 17 October 2017)

ABSTRACT

The Global Ensemble Forecasting System (GEFS) is being extended from 16 to 35 days to cover the subseasonal period, bridging weather and seasonal forecasts. In this study, the impact of SST forcing on the extended-range land-only global 2-m temperature, continental United States (CONUS) accumulated precipitation, and MJO skill are explored with version 11 of the GEFS (GEFSv11) under various SST forcing configurations. The configurations consist of 1) the operational GEFS 90-day *e*-folding time of the observed real-time global SST (RTG-SST) anomaly relaxed to climatology, 2) an optimal AMIP configuration using the observed daily RTG-SST analysis, 3) a two-tier approach using the CFSv2-predicted daily SST, and 4) a two-tier approach using bias-corrected CFSv2-predicted SST, updated every 24 h. The experimental period covers the fall of 2013 and the winter of 2013/14. The results indicate that there are small differences in the ranked probability skill scores (RPSSs) between the various SST forcing experiments. The improvements in forecast skill of the Northern Hemisphere 2-m temperature and precipitation for weeks 3 and 4 are marginal, especially for North America. The bias-corrected CFSv2-predicted SST experiment generally delivers superior performance with statistically significant improvement in spatially and temporally aggregated 2-m temperature RPSSs over North America. Improved representation of the SST forcing (AMIP) increased the forecast skill for MJO indices up through week 2, but there is no significant improvement of the MJO forecast skill for weeks 3 and 4. These results are obtained over a short period with weak MJO activity and are also subject to internal model weaknesses in representing the MJO. Additional studies covering longer periods with upgraded model physics are warranted.

1. Introduction

Recently, the need for numerical guidance covering the weeks 3 and 4 period has been increasing, driven primarily by economic requirements to support decision-makers (e.g., the management of water supplies) and for preparedness to changes in climate. The National Oceanic and Atmospheric

Administration (NOAA) is accelerating its efforts to improve its numerical guidance and prediction capability for the extended range: the weeks 3 and 4 period that bridges the gap between weather and climate. Covering the extended-range period will enable NOAA to provide seamless numerical guidance to the public, protecting life and property.

Global efforts have been pursued to provide extended-range forecast guidance to the public, helping

Corresponding author: Yuejian Zhu, yuejian.zhu@noaa.gov

DOI: 10.1175/WAF-D-17-0093.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](http://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

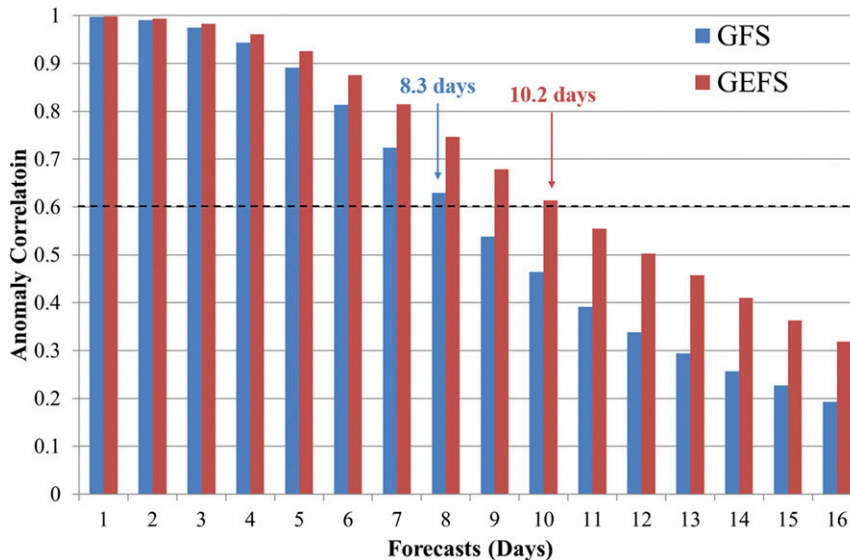


FIG. 1. Average anomaly correlation by lead day for 500-hPa geopotential heights over the NH covering the period of 1 Sep 2013–28 Feb 2014 for the deterministic GFS (blue) and the GEFS ensemble mean (red).

to reduce the effects of high-impact weather and extreme events. One such effort is the Subseasonal to Seasonal (S2S) project, a legacy project of The Observing System Research and Predictability Experiment (THORPEX). This project was endorsed in 2012 by the World Meteorological Organization’s (WMO) World Weather Research Program (WWRP) and World Climate Research Program (WCRP; Vitart et al. 2017). In the United States, NOAA is pursuing parallel efforts to “develop an intraseasonal to interannual prediction system that builds on the currently experimental real-time National Multi-Model Ensemble system and incorporates advances in statistical methodologies and forecast initialization” to provide weeks 3 and 4 forecast guidance (NOAA 5-yr research and development plan: 2013–17; <http://nrc.noaa.gov/CouncilProducts/ResearchPlans/5YearRDPlan/NOAA5YRPHome.aspx>). Since 2011, the National Weather Service (NWS) has been furthering the Weather-Ready Nation (WRN) strategic plan to “create a seamless suite of forecasts that look out beyond two weeks to support response and preparedness to changes in climate that incorporate research advances from within NOAA and other partners, including the commercial weather and climate industries” (Weather-Ready Nation—NWS Strategic Plan 2011, http://www.nws.noaa.gov/com/weatherreadynation/files/strategic_plan.pdf).

Past studies using dynamical models, statistical models, empirical methods, and other tools have examined the weeks 3 and 4, subseasonal, and/or intraseasonal time periods. The seminal studies by Lorenz (1969a,b, 1982) set the foundation for understanding predictability.

Subsequent studies attempted to find and explain key phenomena that impact predictability across temporal scales. In the tropics, the Madden–Julian oscillation (MJO; Madden and Julian 1971, 1972) was found to be a key phenomenon for extended-range prediction due to its 40–50-day oscillation time scale. In the Northern Hemisphere (NH), the Pacific–North American (PNA) and North Atlantic Oscillation (NAO) patterns in the mid- to high latitudes have been found to be sources of extended-range predictability (Wallace and Gutzler 1981; Barnston and Livezey 1987). In particular, specific blocking patterns can be identified in the extended range that can result in drought and heat waves in summer and produce conditions conducive for severe weather in the winter (Rex 1950). Several notable studies have explored various means to increase forecast skill on the subseasonal-to-seasonal time scales, with emphasis placed on high-impact weather events (Kirtman et al. 2014).

The studies using numerical models have focused on the scientific issues and relationships of various key phenomena, including the impact of SSTs and MJO (Pegion and Sardeshmukh 2011; Fu et al. 2013; Xiang et al. 2015), teleconnections (e.g., PNA, NAO; Van den Dool et al. 2000; Chen and Van den Dool 2003), monsoons (Adams and Comrie 1997; Chang et al. 2000; Luo et al. 2016), extreme rainfall events (Luo et al. 2016), sea ice (Hunke et al. 2010), and the interaction of tropospheric and stratospheric processes (Lindzen 1987). These studies raise important issues for extended-range numerical model prediction such as the relationship

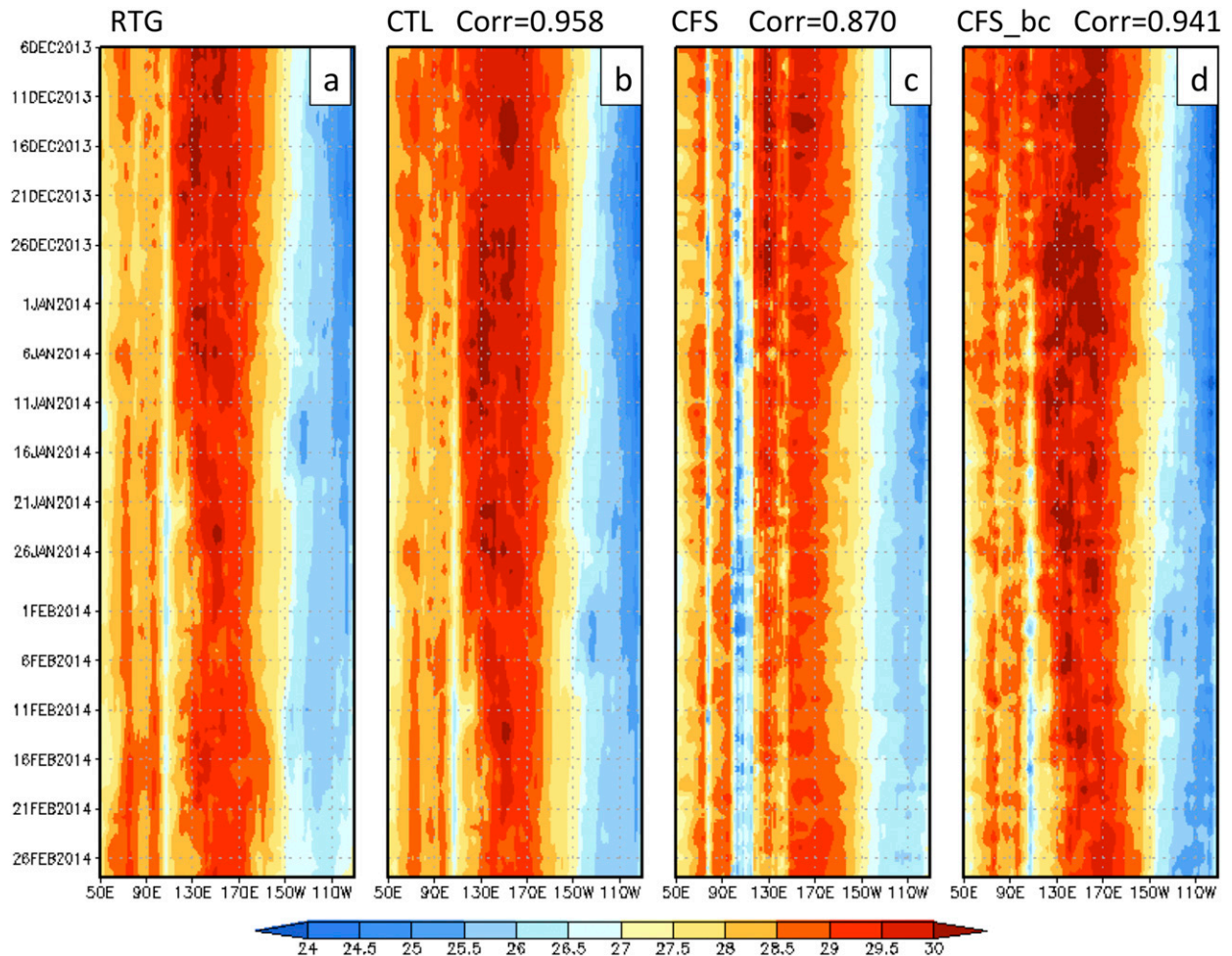


FIG. 2. Hovmöller diagrams of area-average SST (K) over the 15°S–15°N bands for the (a) RTG analysis, (b) CTL initial conditions, and (c) CFS and (d) CFS_BC SST forecast at lead day 20. The three panels on the right verify with the dates of the RTG on the left. At the top, the time–longitude correlation is given for each of the SST forecast panels.

between model resolution and physical parameterizations for coupled ocean–atmosphere models, initialization strategies for subseasonal prediction, ensemble generation, model systematic errors, and the representation of forecast uncertainties. Model systematic errors continue to plague medium- and extended-range forecasts, but retrospective forecasts can be implemented to reduce their impact. The additional resources required for retrospective forecasts make it more expensive to implement a numerical modeling system for extended-range prediction.

Operational global numerical guidance for weeks 3 and 4 and monthly predictions are available from several operational forecasting centers. NOAA’s National Centers for Environmental Prediction (NCEP) Climate Forecasting System (CFS) version 2 (CFSv2) is a coupled (ocean, sea ice, land, and atmosphere) model (Saha et al. 2006, 2010, 2014) that combines four forecasts initialized four times

daily into a daily 16-member time-lagged ensemble integrated out to 45 days with retrospective hindcasts for bias correction. The European Centre for Medium-Range Weather Forecasts (ECMWF) runs a 51-member global coupled (ocean, sea ice, land, and atmosphere) Ensemble Prediction System (EPS; Vitart et al. 2014) out to 46 days. The ECMWF EPS is initialized twice per week with a real-time hindcast for forecast calibration. Recently, Environment Canada (now known as Environment and Climate Change Canada) extended their 21-ensemble-member uncoupled Global Ensemble Prediction System (GEPS; Côté et al. 1998; Buizza et al. 2005) to 32 days once per week with a real-time reforecast (or hindcast) for forecast calibration.

The NCEP Global Ensemble Forecast system (GEFS) has been designed to incorporate forecast uncertainty, which results in improved forecast reliability (Toth and Kalnay 1993, 1997; Buizza et al. 2005; Wei et al. 2008;

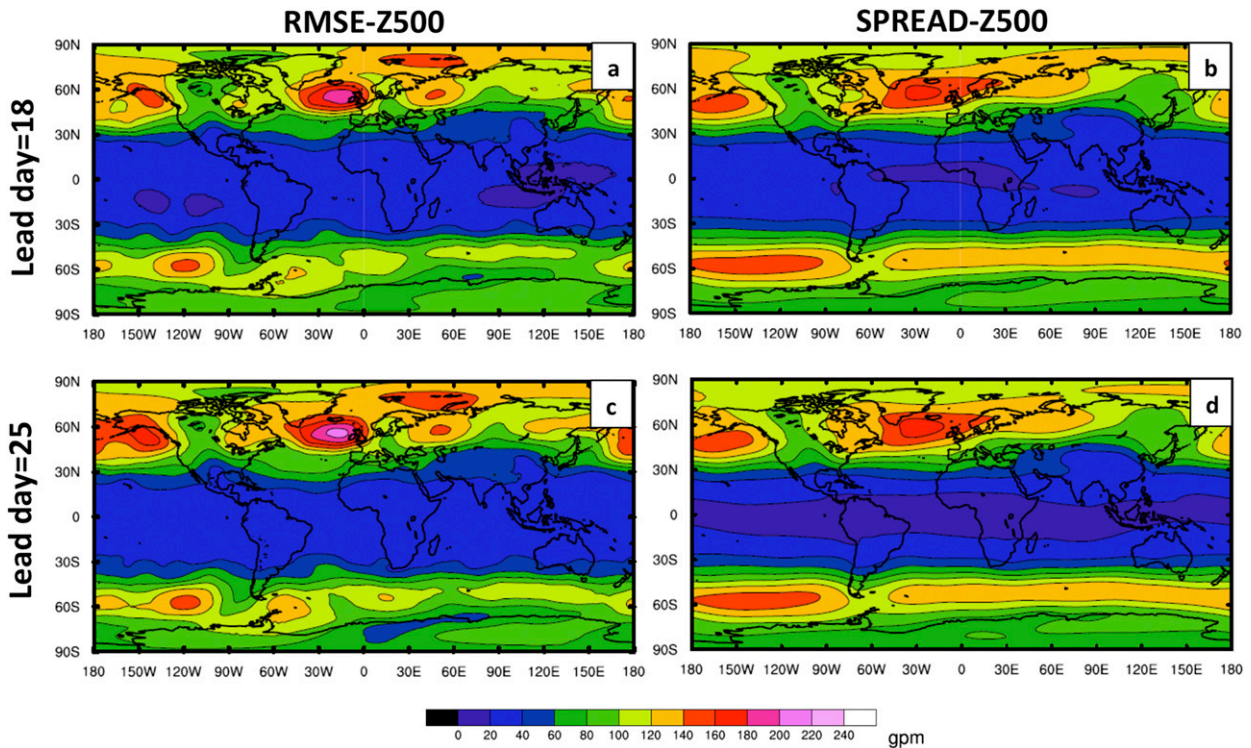


FIG. 3. Spatial distributions of (left) 5-day running mean RMS error and (right) ensemble spread of 500-hPa geopotential heights (gpm) for CTL over the 6-month experiment period for lead days (a),(b) 18 and (c), (d) 25.

Ma et al. 2014) in the medium-range. In recent years, GEFS has yielded excellent day-to-day forecast skill. The GEFS ensemble mean has consistently demonstrated similar or improved forecast skill compared to the deterministic Global Forecast System (GFS), which is pronounced at longer lead times. The NH 500-hPa geopotential height anomaly correlation out to 16 days for the experimental period in this manuscript (fall 2013 and winter 2013/14) is shown in Fig. 1. Unlike the GFS, the GEFS produces a probabilistic forecast, providing a measure of forecast uncertainty (Toth et al. 2001; Zhu et al. 2002; Zhu 2005) that can aid in forecasting extreme weather events (Guan and Zhu 2016). Extending the GEFS (currently run to 16 days) to cover the weeks 3 and 4 period provides additional benefits over the CFSv2, including a more frequent model upgrade cycle, higher model resolution, state-of-the-art flow-dependent initial perturbations from a hybrid four-dimensional ensemble-variational data assimilation (4DEnVar) system, stochastic physics, and larger ensemble membership (84 members for every 24-h cycle), all providing an improved sampling of forecast uncertainty.

In this study, the operational GEFS v11 configuration is extended to 35 days, and the forecast skill is evaluated (Melhauser et al. 2016). Various SST forcing experiments are performed to examine the impact of SST

forcing on the extended-range forecast skill of global 2-m temperature, accumulated precipitation over the continental United States (CONUS), and MJO indices. Section 2 describes the GEFS configuration for the SST forcing experiments, experiment forecast period, and aspects of the verification methodology. Section 3 provides results and offers discussion of the forecast skill for global 2-m temperature, CONUS accumulated precipitation, and MJO indices. Section 4 provides concluding remarks and future steps.

2. Methodology

a. Operational NCEP GEFS

The current operational configuration of GEFS uses version 12 of the GFS Global Spectral Model (GSM v12.0.0) for integration four times per day (0000, 0600, 1200, and 1800 UTC) out to 16 days (Sela 1980; Han and Pan 2011; Han et al. 2016). For days 0–8 the GEFS has a spectral resolution of T_L574 (semi Lagrangian with a linear grid; approximately 34 km) with 64 hybrid vertical levels and the horizontal resolution is reduced for days 8–16 to T_L384 (approximately 52 km). The 20-member ensemble initial condition perturbations are selected from the operational hybrid NCEP Global

Data Assimilation System (GDAS) 80-member ensemble Kalman filter (EnKF; Wu et al. 2002; Whitaker et al. 2008; Kleist et al. 2009; X. Wang et al. 2013; Kleist and Ide 2015; Zhou et al. 2016, 2017). If tropical cyclones (TCs) are present in the initial conditions, TC perturbations are calculated after TCs are separated from the environment (Kurihara et al. 1995) and are relocated to the same location (Liu et al. 2006). GEFS accounts for model errors by perturbing the total tendencies using the stochastic total tendency perturbation scheme (STTP; Hou et al. 2008). The GEFS has the same GFS SST forcing, which is initialized with the real-time global (RTG) analysis and damped to climatology (90-day *e*-folding time) during model integration. The sea ice concentration is initialized from the daily 0000 UTC sea ice and snow analysis from the Interactive Multisensor Snow and Ice Mapping System (Ramsay 1998) and held constant throughout the model integration. (Additional information on GSM v12.0.0 settings used in the operational GEFS is available online: <http://www.emc.ncep.noaa.gov/GFS/impl.php>.) For this study, the operational GEFS configuration is modified as follows: 1) the forecast is extended to 35 days with the horizontal resolution reduced to T_{1254} (approximately 78 km) for 16–35 days, 2) the SST is updated with various SST forcing schemes, and 3) the forecast is only initialized once per day at 0000 UTC because of resource constraints.

b. SST forcing experiments

The SST configurations consist of the operational GEFS 90-day *e*-folding time of the observed RTG-SST anomaly to climatology (CTL), an optimal Atmospheric Model Intercomparison Project (AMIP) configuration using the observed RTG-SST analysis updated every 24 h during model integration (RTG), a two-tier approach using the CFSv2-predicted SST updated every 24 h during model integration (CFS), and a two-tier approach using bias-corrected CFSv2-predicted SST updated every 24 h during model integration (CFS_BC). Detailed formulations for CTL and CFS_BC can be found in [appendix A](#).

The operational GEFS is an uncoupled system with the SST prescribed using NCEP's RTG-SST analysis persisted and damped to climatology during the forecast. Model boundary conditions, including the underlying SST, are known to influence prediction skill in the extended range. Therefore, it is important to assess the impact of SST forcing on extended-range forecast skill before fully coupling the GEFS to an ocean model. We show the area-average SST over the 15°S–15°N band for the RTG (Fig. 2a) analysis and lead day 20 forecasts valid at the corresponding analysis verification date for CTL (Fig. 2b), CFS (Fig. 2c), and CFS_BC (Fig. 2d). Comparing the two-tiered CFS experiment

to CTL, CFS provides additional multiscale information. Removing systematic biases in the CFS_BC experiment (Fig. 2d) improves the correlation between the RTG analysis and the lead day 20 forecast output.

c. Experiment period

All experiments in this study span the fall of 2013 and winter of 2013/14 and are initialized every 24 h starting on 1 September 2013 and ending 28 February 2014. For 2-m temperature and accumulated precipitation, only the lead times corresponding to 0000 UTC (e.g., 24, 48, 72 h, etc.) are verified to control for diurnal variability. Over the experiment period, the MJO was weak or nonexistent [Climate Prediction Center (CPC); <http://www.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/whindex.shtml>] and ENSO-neutral conditions persisted (Earth System Research Laboratory; <http://www.esrl.noaa.gov/psd/enso/mei>).

For the fall of 2013, parts of the CONUS including the northern Rockies and northern plains experienced wetter-than-normal conditions, with precipitation totals in the northern plains states, Colorado, and New Mexico ranking within their 10 wettest (since 1895). California remained extremely dry, with autumn 2013 ranking its 10th driest (since 1895), with below-normal precipitation also observed in the Southeast and Northeast. Eurasia experienced above-normal temperatures, having its record warmest November and December (since 1900). Over Europe, the beginning of fall was also anomalously warm, with Finland, Spain, and Norway experiencing above-normal temperatures for September (National Climatic Data Center Climate Global Analysis; <https://www.ncdc.noaa.gov/sotc/global>).

For the winter of 2013/14, the Northern Hemisphere was plagued with persistent dips in the jet stream that brought cold air into North America and central Russia and warm air into northern Europe. Environment Canada reported its coldest winter since 1996 and coldest November–March (since 1948). Across the CONUS, below-average temperatures were experienced east of the Rockies, but California had its warmest winter on record and above-normal conditions were experienced by the surrounding southwestern states. Over the western United States and the Great Plains, drier-than-normal conditions persisted (National Climatic Data Center Climate Global Analysis; <https://www.ncdc.noaa.gov/sotc/global>).

d. Verification procedure

1) RPSS: 2-M TEMPERATURE AND ACCUMULATED PRECIPITATION

The forecast skill for 2-m temperature and accumulated precipitation are evaluated using a tercile (below

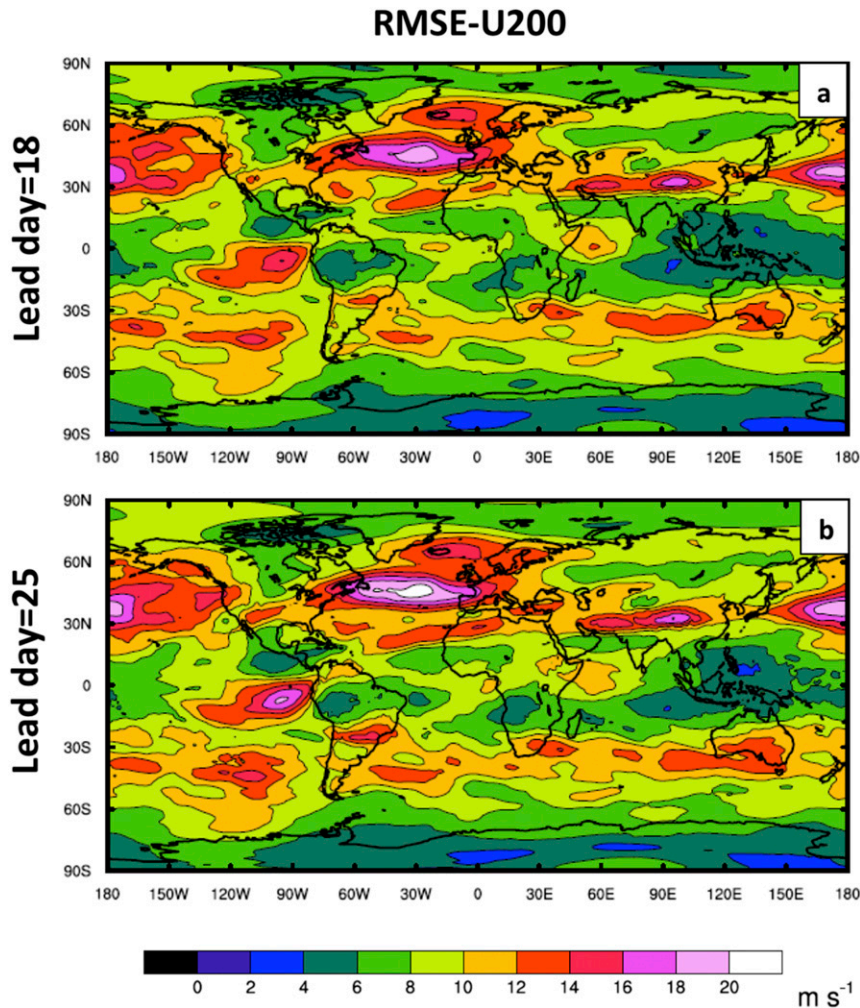


FIG. 4. Spatial distribution of RMS error of U200 (m s^{-1}) for CTL over the 6-month experiment for lead days (a) 18 and (b) 25.

normal, normal, or above normal) ranked probability skill score (RPSS; e.g., Wilks 2011); see appendix B for additional details. The 2-m temperature is verified for land-only cases against the 0000 UTC GDAS analysis, and the accumulated precipitation is verified for land-only cases against the 0000 UTC NCEP climatologically calibrated precipitation analysis (CCPA; Hou et al. 2014). The GEFS 2-m temperature is averaged and the accumulated precipitation accumulated over the lead times of interest (week 2, days 8–14; weeks 3 and 4, days 15–28); the results are verified against the corresponding GDAS and CCPA data averaged or accumulated over the same lead times. Different methods and period lengths can be defined, which can have a direct impact on forecast skill; generally, longer averaging periods produce higher RPSS values (not shown). Week 2 and weeks 3 and 4 were chosen in this investigation to match

the operational CPC week 2 and experimental weeks 3 and 4 forecasts.

2) MJO SKILL SCORE

In this study, the MJO is evaluated using the traditional real-time multivariate MJO (i.e., RMM) index [the Wheeler–Hendon index (WH index); Wheeler and Hendon (2004); Gottschalck et al. (2010)]. The MJO forecast skill is defined as the bivariate anomaly correlation between the analysis and forecast RMM1 and RMM2 over the fall of 2013 and winter of 2013/14, calculated at each lead time. The GEFS ensemble mean outgoing longwave radiation (OLR), 850-hPa u -wind component (U850), and 200-hPa u -wind component (U200) are verified against the same variables from GDAS. The long-term climatology is calculated from the NCEP/

NCAR reanalysis (R-1; <http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html>) for U200 and U850 and from the NCAR Interpolated Outgoing Longwave Radiation dataset (http://www.esrl.noaa.gov/psd/data/gridded/data.interp_OLR.html; Liebmann and Smith 1996) for the OLR, both for the period 1981–2010. The long-term mean and average of the previous 120 days are removed from the climatology to eliminate long-term trends and seasonal variability.

3. Results and discussion

An ensemble prediction system is performing well if it can produce an accurate estimate of its lead-time-specific forecast errors (error) through its ensemble dispersion (spread). If this is the case, the benefit of an ensemble predicting its own forecast errors can be utilized. The ensemble root-mean-square (RMS) error and spread for GEFS 500-hPa geopotential heights over the fall of 2013 and the winter of 2013/14 for lead day 18 (Figs. 3a,b) and lead day 25 (Figs. 3c,d) generally supports the notion that GEFS is performing well during the weeks 3 and 4 period, although deviations occur for other variables, lead times, and locations. For both lead days 18 and 25, the spread and error over the NH polar latitudes show similar spatial patterns and magnitudes, although the results are slightly underdispersive over the NH storm tracks (Buizza et al. 2005). In the Southern Hemisphere (SH), the GEFS appears to be slightly overdispersive over a large swath of the SH Arctic Circle.

Locating the sources of uncertainty in the large-scale circulation is a necessary step toward an accurate forecast for the weeks 3 and 4 period. During the fall–winter in the NH, the subtropical jet is one of the major large-scale circulations that modulate North American (NA) weather. As such, demonstrating the uncertainty associated with the upper-level circulation is helpful for a model developer when evaluating jet stream forecasts.

The 6-month experiment period average of 200-hPa RMS error (Fig. 4) shows a similar magnitude and spatial distribution between lead days 18 and 25. As expected, the largest errors reside in the NH storm tracks given the time frame of the experiment period. Most of the larger errors reside in the midlatitudes south of 30°S and north of 30°N. This suggests that for weeks 3 and 4 forecasts, improving the skill of the large-scale circulation, especially over the subtropical jet region, should not be ignored.

a. 2-m temperature forecast skill

Over the experiment period, the global land-only 2-m temperature RPSS is regionally and lead time dependent. The tropics (TR) have the highest RPSSs for both week 2

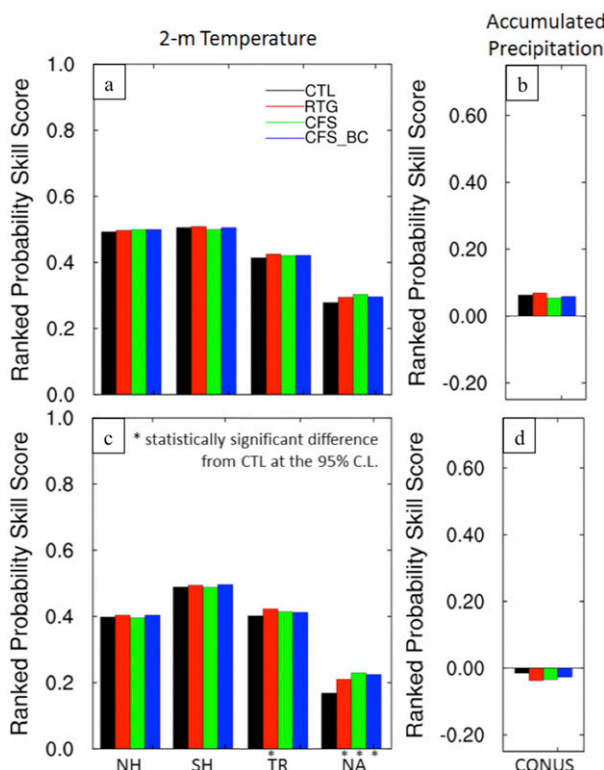


FIG. 5. RPSS for CTL (black), RTG (red), CFS (green), and CFS_BC (blue) calculated for (top) week 2 and (bottom) weeks 3 and 4 for (a), (c) 2-m temperature and (b), (d) accumulated precipitation averaged over the 6-month experiment period. Asterisks beneath the respective experiment column scores indicate that the difference of that experiment from CTL is statistically significant at the 95% confidence level.

(Fig. 5a) and weeks 3 and 4 (Fig. 5b) with NA having the lowest. Comparing between week 2 and weeks 3 and 4, the RPSSs remain similar for the tropics and SH, with the NH and NA dropping ~0.1–0.3. Within each region, the levels of forecast skill for the SST forcing experiments are generally statistically indifferent from CTL for both week 2 and weeks 3 and 4. RTG, CFS, and CFS_BC show a statistically significant improvement during weeks 3 and 4 over NA with RTG showing statistically significant improvements over TR. It is interesting that RTG does not have a more robust improvement compared to the other experiments, given this experiment is being forced with the observed SST forcing. During weeks 3 and 4 over NA, CFS and CFS_BC actually outperform the RTG experiment in terms of RPSS. It should be restated that the period of this experiment does occur over an inactive MJO period with ENSO neutral conditions; thus, the tropical forcing and correlations with global weather may have a low signal-to-noise ratio.

The global weeks 3 and 4 spatial 2-m temperature RPSS score for CTL (Fig. 6a) indicates the highest skill over land extending from the western Sahara into the

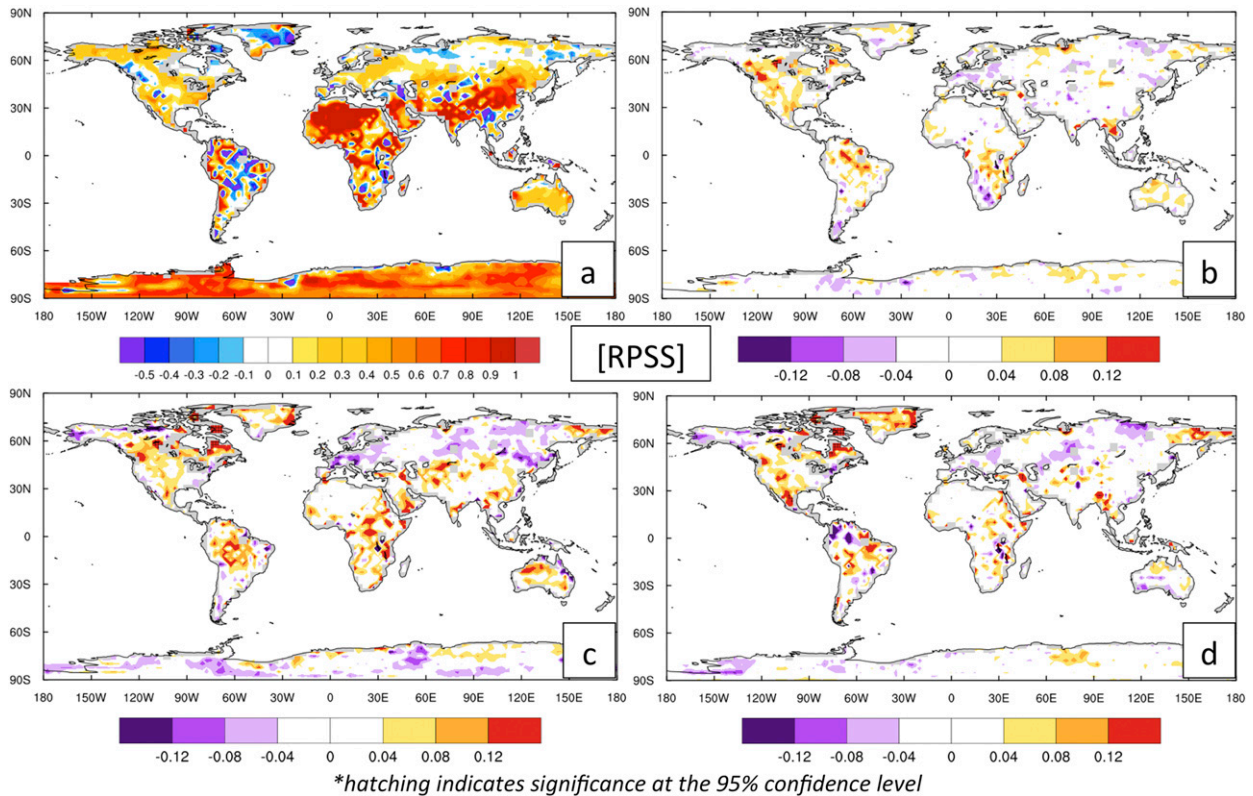


FIG. 6. Land-only 2.5° global 2-m temperature RPSSs averaged over the 6-month experimental period for weeks 3 and 4 for (a) CTL and the difference between (b) RTG, (c) CFS, and (d) CFS_BC and CTL. Hatching in (b)–(d) indicates the difference is statistically significant at the 95% confidence level.

Middle East and northern China. Generally, the lowest relative skill is found over Europe, central South America, and the northern portions of Asia. Comparing RTG to CTL (Fig. 6b), no gridpoint statistical significance is found anywhere over land, but some general hints at coherent areas of improvement in RPSS can be found over central South America, North America, and Australia. Differences over land can also be found comparing CTL to the CFS (Fig. 6c) and CFS_BC (Fig. 6d) experiments. In general, the experiments forced with the CFSv2 (CFS and CFS_BC) hinted at larger improvements in the same areas except for a generally coherent degradation over north-central Asia. Over the ocean (not shown), the CFS experiment shows degradation in RPSS over the northern high latitudes. Along the western portion of South America and extending to the eastern equatorial region, applying a bias correction in CFS_BC significantly improves the degradations found in CFS.

A warm bias exists in CTL across central NA (Fig. 7a), extending north into Greenland. This suggests the GEFS had a hard time capturing the unusually cold conditions across the central and eastern United States

and Canada that were observed during the experiment period. This warm bias is reduced in RTG (Fig. 7b), CFS (Fig. 7c), and CFS_BC (Fig. 7d), and corresponds to improved RPSSs in central NA. Both CFS and CFS_BC indicate large regions of the northern high latitudes that were cooler than CTL, which may be partially explained by differences in how sea ice is handled in the CFS SST forcing compared to climatological SST forcing in CTL.

Specifically comparing CFS (Fig. 7c) and CFS_BC (Fig. 7d), the bias correction in CFS_BC does little to reduce the 2-m temperature forecast bias over the northern latitudes, indicating a clear systematic difference between dynamically evolving sea ice and prescribing the SST. However, the CFS_BC clearly reduces the bias over the western United States extending into northern Mexico, improving the RPSS (Fig. 6d). Additionally, CFS_BC significantly reduces the warm bias in CFS along the west coast and southern portion of South America. Focusing on Asia, a cold bias in CFS is present over Siberia. This is not present in CFS_BC, with it being slightly warmer in this area.

The weeks 3 and 4 time frame falls within the gray zone between weather and climate; thus, one way to

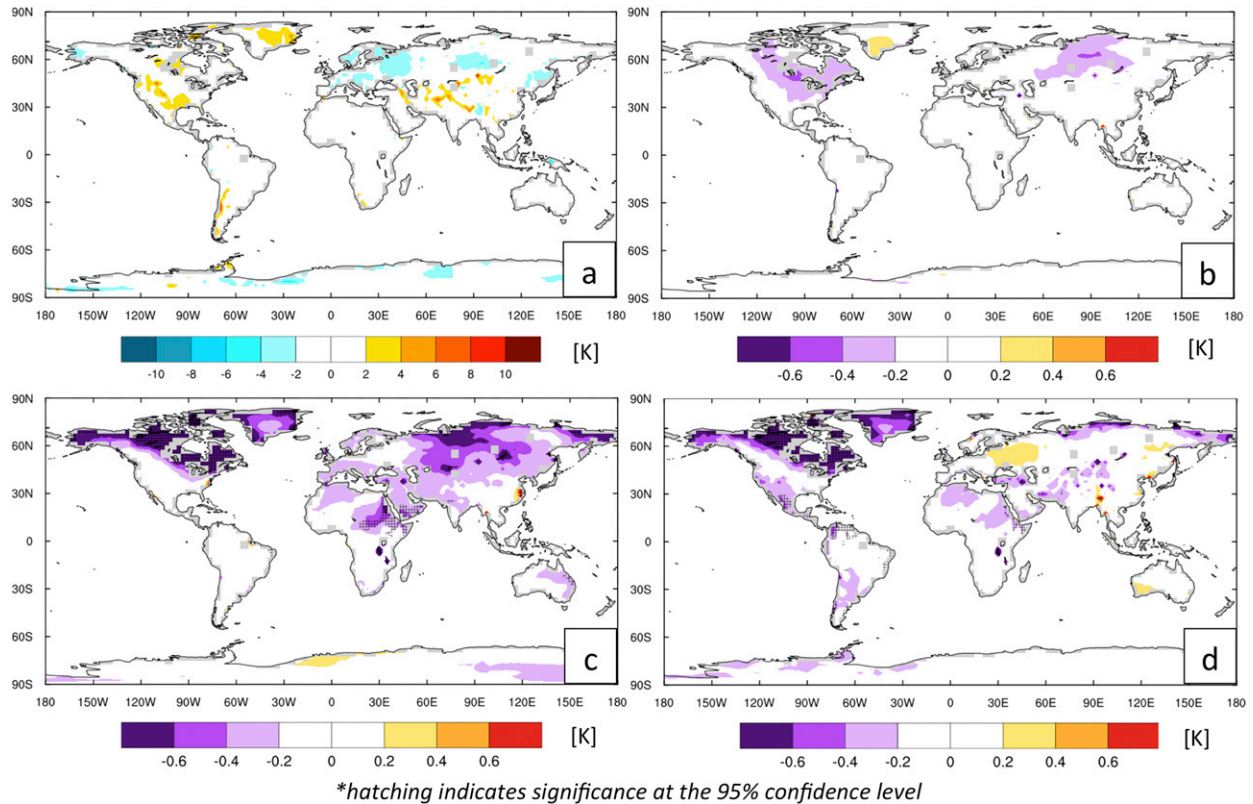


FIG. 7. Land-only 2.5° global 2-m temperature bias (forecast – analysis) (K) averaged over the 6-month experiment period for weeks 3 and 4 for (a) CTL and the difference between (b) RTG, (c) CFS, and (d) CFS_BC and CTL.

highlight the subseasonal time scale and increase the prediction capability is to remove the short-term noise associated with the synoptic weather using a 5-day running mean. The 5-day running mean RMS error for 2-m temperature shows the largest error over central and western NA and central Siberia, extending across Asia (Fig. 8a). RTG (Fig. 8b) reduces the error across NA, while increasing the error over Siberia and across continental Asia. CFS (Fig. 8c) has areas of error reduction around the Great Lakes in NA, but areas of increased error are found along the West Coast and extending into Alaska and across the central United States. Similar large increases in error were found across Siberia. Interestingly, CFS_BC (Fig. 8d) has an almost opposite impact across NA and NH, with increased error across central NA and a reduction in error across Siberia.

While RMS error provides the forecast error, the signal-to-noise ratio (SNR, which is defined as the ensemble mean divided by the ensemble spread for each model grid) indicates the predictability for a certain forecast variable (B. Wang et al. 2013; Zhang et al. 2016). For 2-m temperature, the SNR mainly occurs over the tropical regions (Fig. 9). Over the western CONUS, there is higher predictability compared with the central

United States, but overall the predictability is low. It should be noted the GEFS is underdispersive in 2-m temperature, especially in the tropics.

The small improvement in 2-m temperature RPSS in RTG over land using a “perfect” SST setup indicates there are deficiencies that need to be addressed in the forecast model. The GEFS in its current configuration may not effectively propagate the information contained in the tropical SSTs to land regions around the globe. This is not simply an issue of low forecast skill over weeks 3 and 4 (Fig. 5c), as this was also evident during week 1 (not shown) and week 2 (Fig. 5a). It should be noted again that the experiment period is only 6 months and occurred during a period of weak MJOs and ENSO-neutral conditions. It is interesting that CFS_BC performs as well or better than RTG (statistically significant over NA) for the 2-m temperature RPSS. Further investigation needs to be performed to determine if this trend holds over other forecast variables and verification metrics.

b. Accumulated precipitation forecast skill: CONUS

Over the fall of 2013 and the winter of 2013/14, the CONUS accumulated precipitation RPSS shows no

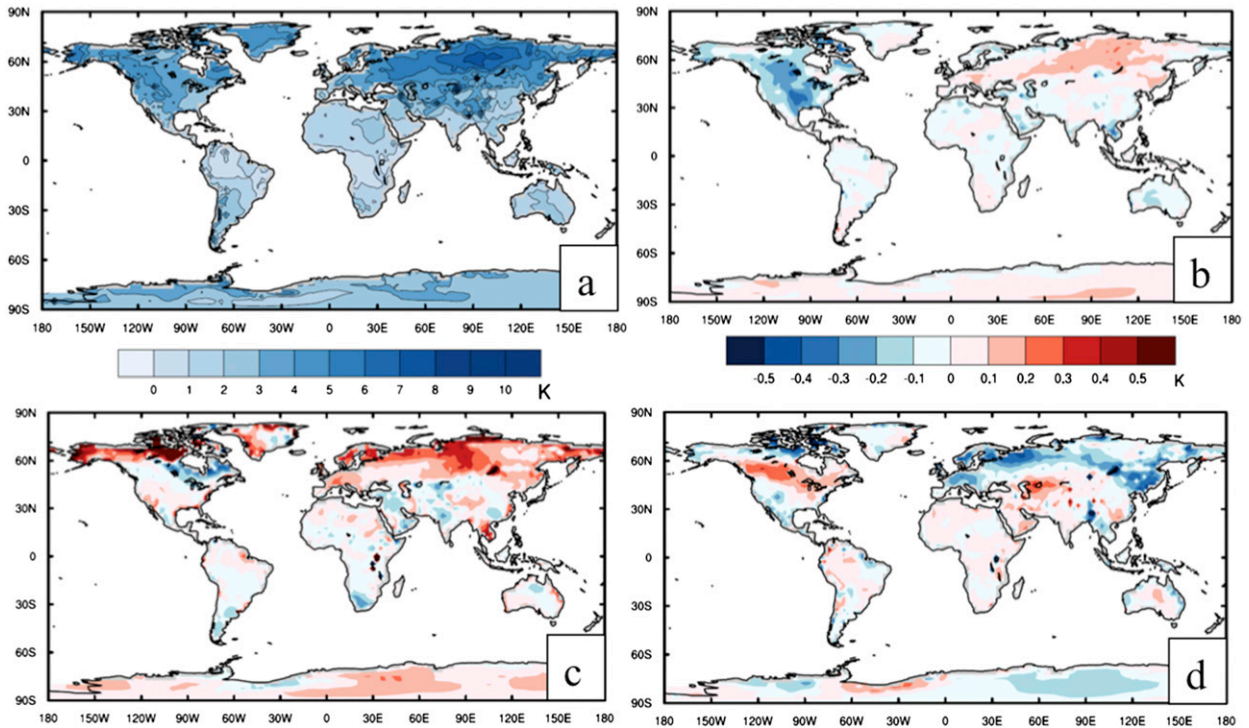


FIG. 8. Land-only 2.5°2-m temperature RMS error (K) for (a) CTL and the difference between (b) RTG, (c) CFS, and (d) CFS_BC and CTL, averaged over the 6-month experiment period.

statistically significant difference between CTL and RTG, CFS, or CFS_BC for week 1 (not shown), week 2 (Fig. 5b), or weeks 3 and 4 (Fig. 5d). The magnitude of the RPSS falls off drastically after week 1—approximately 0.55 at lead day 1 and 0 at lead day 7 (die-off curves not

shown)—leveling off around approximately 0 (no skill) for all experiments for the extended period. The aggregate accumulated week 2 RPSS is slightly higher than weeks 3 and 4, but overall the results indicate very little skill with the current model configurations, regardless of SST forcing.

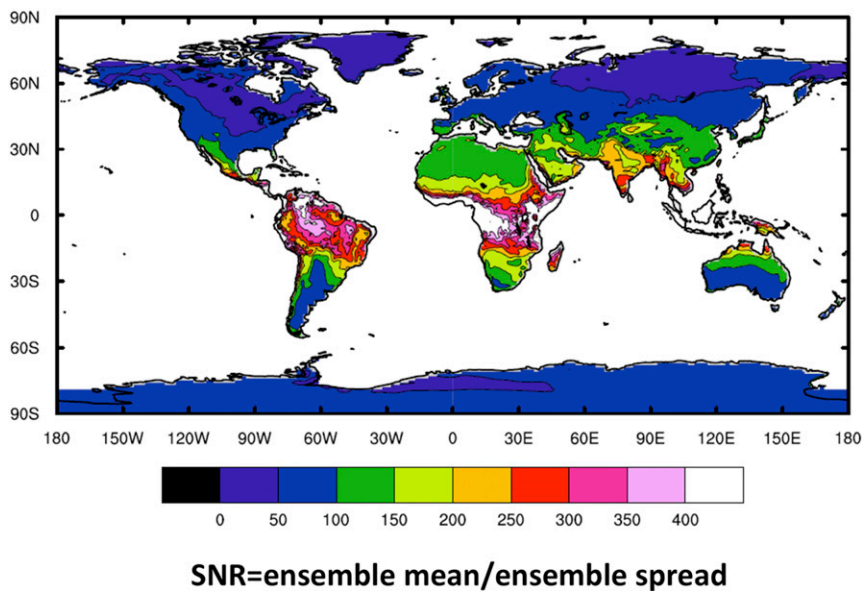


FIG. 9. The 2-m temperature SNR for CTL averaged over weeks 3 and 4.

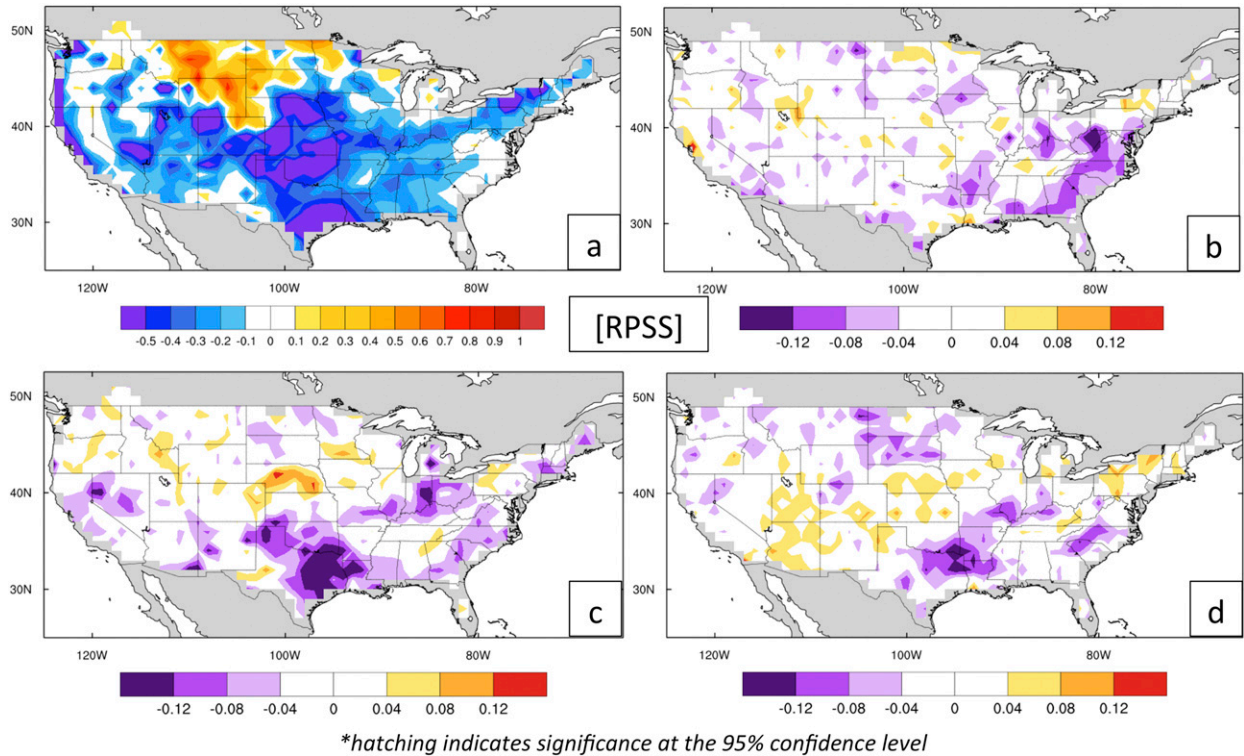


FIG. 10. Spatial weeks 3 and 4 accumulated precipitation RPSS over the CONUS averaged over the 6-month experimental period for (a) CTL and the difference between (b) RTG, (c) CFS, and (d) CFS_BC and CTL.

The distribution of the weeks 3 and 4 accumulated precipitation RPSS for CTL (Fig. 10) indicates the highest skill is over the northern plains with small positive or even negative skill across the Southwest, south-central plains, and Southeast. Comparing the RPSS differences from CTL, the RTG and CFS (SST forcing) experiments (Figs. 10c,d) generally show higher relative skill over the central plains into the Great Lakes, but less skill over northwest Texas. All SST forcing experiments have reduced RPSSs in the Southeast. The bias partially explains the RPSS distribution, with CTL too dry over the south-central plains extending into the Mississippi River valley and slightly too wet over the northern plains and far Southeast (Fig. 11a). There are coherent spatial bias differences between RTG (Fig. 11b), CFS (Fig. 11c), and CFS_BC (Fig. 11d) and the CTL, but none is large enough in magnitude to be statistically significant. The RTG and CFS_BC experiments have a reduced dry bias in coherent regions of the central and portions of the eastern United States. The small differences in bias between the SST forcing experiments suggests the systematic model errors from model parameterizations dominate the biases over the extended period.

c. MJO forecast skill and evolution

The MJO is one of the dominant sources of predictability at the subseasonal time scale. As such, the

forecast skill of MJO is a key metric when evaluating the capability of operational models for subseasonal forecasts (Kim et al. 2014; Shelly et al. 2014; Ling et al. 2014; Xiang et al. 2015). The MJO forecast skill in the operational GEFS during the experimental period (Fig. 12) is ~ 14.6 days—defined as the lead time when the bivariate anomaly correlation coefficient drops to 0.5. After week 2, MJO forecast skill quickly drops. Changing the prescribed SST to be closer to the observations (RTG), the MJO forecast skill was improved up to ~ 2 days. For the weeks 3 and 4 range, the most skillful SST forcing is RTG with the CFS_BC being the most skillful scheme that could be used in operations.

The MJO skill averaged for weeks 3 and 4 was improved by $\sim 10\%$ (figure not shown) for CFS_BC. This implies that the MJO prediction skill is related to the accuracy of the representation of the SST, which is consistent with other studies (e.g., Wang et al. 2015). Therefore, without changing the model, it is found that improving the SST results in an increase in MJO forecast skill.

The strength and variability of the MJO index are subject to forecast errors that increase with lead time. Over the experiment period, the MJO is predicted to be weaker in September, from late November to mid-December of 2013, and late January and late February 2014, but

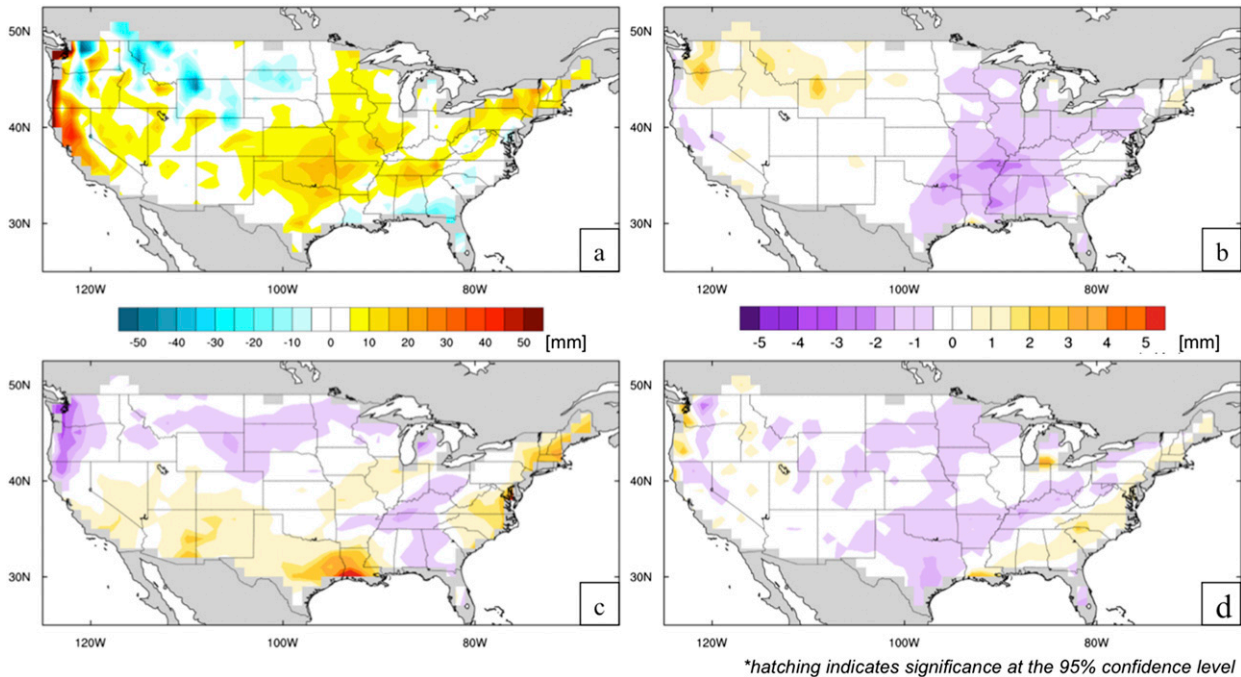


FIG. 11. Spatial weeks 3 and 4 accumulated precipitation bias (model – analysis) (mm) over the CONUS averaged over the 6-month experimental period for (a) CTL and the difference between (b) RTG, (c) CFS, and (d) CFS_BC and CTL.

stronger over all other periods (Fig. 13). The bias in MJO strength was consistent across lead times. For longer lead times (e.g., lead day 22), the forecast MJO indices tend to become stronger in most verification months except for December. Although the weeks 3 and 4 forecast MJO magnitude is generally too strong and slightly out of phase, there are some periods when GEFS performed well (e.g., from late November to early December for lead day 14). The investigation of this is left for future study.

4. Conclusions and future work

The NCEP GEFS is being extended from 16 to 35 days to cover the subseasonal forecast period. The impacts of SST forcing on the extended-range land-only global 2-m temperature, CONUS accumulated precipitation, and MJO indices forecast skill were examined using various SST forcing configurations. The SST configurations consisted of 1) the operational GFS and GEFS 90-day e -folding time of the observed RTG SST anomaly to climatology; 2) an optimal AMIP configuration using the observed RTG SST analysis updated every 24 h; 3) a two-tier approach using the CFSv2-predicted SST, updated every 24 h; and 4) a two-tier approach using bias-corrected CFSv2-predicted SST, updated every 24 h. The experiments are carried out over a 6-month period covering the fall and winter months of 2013–14. This

period was characterized by weak MJO events and neutral ENSO conditions.

There was little improvement in land-only 2-m temperature and accumulated precipitation forecasts over the extended weeks 3 and 4 period. Forcing the GEFS with an optimal SST did not show statistically significant improvements. This indicates there are deficiencies in the current GEFS configuration that need to be addressed. For accumulated precipitation over the CONUS, there are small differences in RPSS between experiments during the weeks 3 and 4 period.

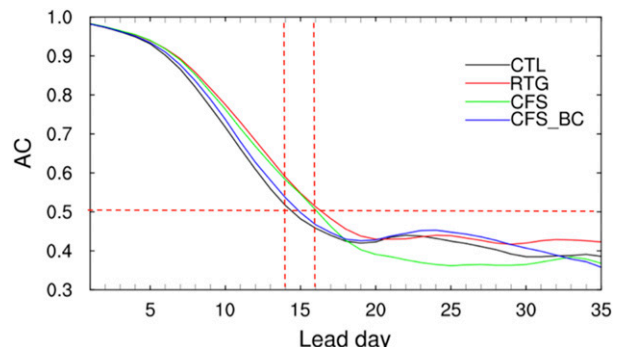


FIG. 12. MJO forecast skill (i.e., bivariate correlation between ensemble mean forecast and analysis data) as a function of lead time for the period 1 Sep 2013–28 Feb 2014. Climatology and previous 120-day mean are removed from the forecast and analysis data while calculating the RMMs.

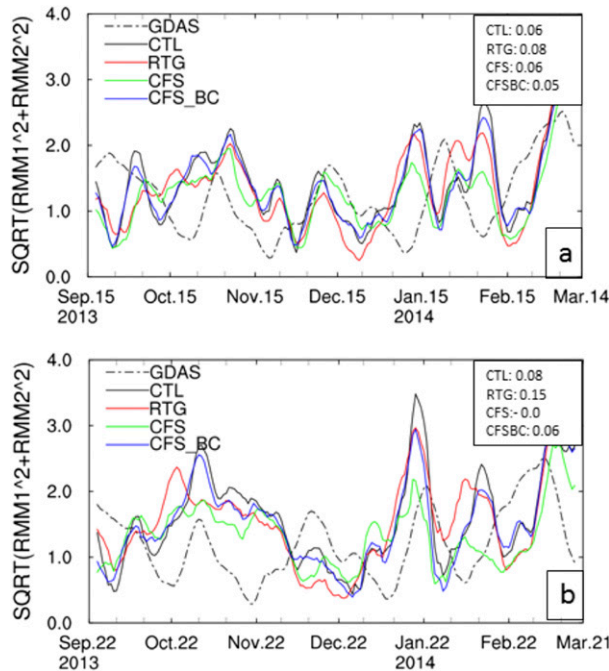


FIG. 13. MJO index for different lead times: (a) lead day 14 and (b) lead day 21. A seven-point running mean is applied on the time series to smooth the data. The numbers in the text box are the variance of each experiment from the analysis for all initial times.

Additionally, there are small differences in bias between the SST forcing experiments, which suggests that systematic model errors dominate the biases at the extended period with model boundary condition forcing having a secondary impact.

It was found that the MJO forecast skill during the experimental period for the operational GEFS is ~ 14.6 days. Using a more realistic SST increased the MJO skill by 10%. The strength and variability of the MJO index are subject to forecast errors, increasing with lead time. The bias in MJO strength was consistent across lead times. For longer lead times (e.g., lead day 22), the forecast MJO indices tend to become stronger in most verification months.

Overall, the one-way forcing of GEFS with more realistic SSTs slightly enhances the MJO skill. It is likely that the cumulus scheme used in this GEFS version underestimates the impacts of SST forcing on the MJO convection (Wang et al. 2015). Along with the relatively weak MJO activity in the observations, external intraseasonal SST forcing does not significantly improve NA weather (2-m temperature and accumulated precipitation) during this period. This implies 1) an inherent predictability issue for NA weather over the weeks 3 and 4 period and that future work needs to be performed 2) to improve the GEFS model as well as 3) to improve boundary forcing predictions, such as those of sea ice, snowpack, and soil moisture for a potential gain in weeks 3 and 4 forecast skill. Also, observations indicate that

the fall of 2013 and the winter of 2013/14 had a generally weak MJO. Future work will focus on a two-year span that covers a stronger MJO period covering 1 May 2014–31 May 2016. This will provide further insight into the predictability from strong MJO events and their relationship with 2-m temperature and CONUS accumulated precipitation from global teleconnections. In these experiments, the potential impacts of higher resolution, stochastic physics, and improving convection parameterization on the MJO in the upgraded GEFS will be examined.

Acknowledgments. The authors thank Dr. Jing Chen for her valuable contribution of precipitation analysis, Dr. Yan Luo for her contribution of Fig. 1, Drs. Qin Zhang and Ping Liu for MJO discussions, Drs. Xingren Wu and Wanqiu Wang for discussions of the SST configuration, Eric Sinsky for improvements to the figures, Drs. Shrinivas Moorthi and Suranjana Saha for their valuable comments through the EMC internal review process, and three anonymous reviewers. This work has been supported by NWS/NGGPS project and OAR/CPO/MAPP.

APPENDIX A

SST Forcing Calculations

a. Operational GEFS SST forcing (CTL)

The GEFS v11 operational SST forcing uses a 90-day e -folding time of the RTG analysis at initialization, relaxed to climatology, calculated as

$$SST_f^t = (SST_a^{t_0} - SST_c^{t_0})e^{-(t-t_0)/90} + SST_c^{t_0},$$

where f is the forecast, a is the analysis, c is climatology, t is forecast lead time, and t_0 is the initial time (in days).

b. Bias-corrected CFSv2-predicted SST forcing (CFS_BC)

The CFS_BC SST forcing is a hybrid of a persisted RTG anomaly at short lead times and bias-corrected CFSv2-predicted SST at longer lead times. The CFSv2-predicted SST is bias corrected using both the CFSR climatology and CFSv2 model climatology. The persisted RTG anomaly is linearly combined with the bias-corrected CFSv2-predicted SST over the 35-day period, calculated as

$$SST_f^t = (1 - w)(SST_a^{t_0} - SST_{cfsrc}^{t_0} + SST_{cfsrc}^t) + w \left[SST_{cfs}^t - (SST_{cfs,c}^t - SST_{cfsrc}^t) \right],$$

where f is the forecast, a is the analysis, cfsrc is the CFSR reanalysis climatology, cfs is the CFS (24-h mean)

forecast SST, cfs_c is the CFSv2 model climatology, t is forecast lead time, t_0 is the initial time, and w is defined as

$$w = (t - t_0)/35.$$

APPENDIX B

Rank Probability Skill Score

The rank probability skill score (RPSS) measures the improvement of a multicategory ensemble forecast relative to a reference forecast. It ranges from $-\infty$ to 1 with a score of 0 indicating it is no better than chance. Since it is a squared error score, RPSS will penalize incorrect forecasts made with a higher forecast probability more severely than an incorrect forecast made with a lower forecast probability (the converse is true for correct forecasts).

For this study, three equal climatological bins (terciles) are defined for each variable. The RPSS is calculated as

$$RPSS = 1 - RPS_f/RPS_c,$$

where the forecast ranked probability score (RPS_f) is calculated as

$$RPS_f = \frac{1}{N} \times \sum_{k=0}^N \left[(\text{prob}B_n - \text{obs}B_n)^2 + (\text{prob}N_n - \text{obs}N_n)^2 + (\text{prob}A_n - \text{obs}A_n)^2 \right],$$

with n corresponding to each forecast–observation pair, N are the total number of forecast–observation pairs, $\text{prob}X_n$ is the *ranked* cumulative forecast probability for each bin X , and $\text{obs}X_n$ is the *ranked* cumulative observation probability for each bin X . The RPS_f forecast probability is the proportion of ensemble members in each bin. The reference RPS_c is calculated similarly, but the forecast probability is set to 1/3 since each forecast bin is defined as being climatologically equal. See Wilks 2011 or the Climate Prediction Center website (<http://www.cpc.ncep.noaa.gov/products/verification/summary/index.php?page=tutorial>) for more information.

REFERENCES

- Adams, D., and A. Comrie, 1997: The North American monsoon. *Bull. Amer. Meteor. Soc.*, **78**, 2197–2213, [https://doi.org/10.1175/1520-0477\(1997\)078<2197:TNAM>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2197:TNAM>2.0.CO;2).
- Barnston, A., and R. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115**, 1083–1126, [https://doi.org/10.1175/1520-0493\(1987\)115<1083:CSAPOL>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2).
- Buizza, R., P. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097, <https://doi.org/10.1175/MWR2905.1>.
- Chang, C.-P., Y. Zhang, and T. Li, 2000: Interannual and interdecadal variations of the East Asian summer monsoon and tropical Pacific SSTs. Part I: Roles of the subtropical ridge. *J. Climate*, **13**, 4310–4325, [https://doi.org/10.1175/1520-0442\(2000\)013<4310:IAIVOT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<4310:IAIVOT>2.0.CO;2).
- Chen, W., and H. Van den Dool, 2003: Sensitivity of teleconnection patterns to the sign of their primary action center. *Mon. Wea. Rev.*, **131**, 2885–2899, [https://doi.org/10.1175/1520-0493\(2003\)131<2885:SOTPTT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<2885:SOTPTT>2.0.CO;2).
- Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.*, **126**, 1373–1395, [https://doi.org/10.1175/1520-0493\(1998\)126<1373:TOCMGE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1373:TOCMGE>2.0.CO;2).
- Fu, X., J.-Y. Lee, P.-C. Hsu, H. Taniguchi, B. Wang, W. Wang, and S. Weaver, 2013: Multi-model MJO forecasting during DYNAMO/CINDY period. *Climate Dyn.*, **41**, 1067, <https://doi.org/10.1007/s00382-013-1859-9>.
- Gottschalk, J. M., 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>.
- Guan, H., and Y. Zhu, 2016: Development of verification methodology for extreme weather forecasts. *Wea. Forecasting*, **32**, 479–491, <https://doi.org/10.1175/WAF-D-16-0123.1>.
- Han, J., and H. Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. *Wea. Forecasting*, **26**, 520–533, <https://doi.org/10.1175/WAF-D-10-05038.1>.
- , M. L. 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**, 341–351, <https://doi.org/10.1175/WAF-D-15-0053.1>.
- Hou, D., Z. Toth, Y. Zhu, and W. Yang, 2008: Impact of a stochastic perturbation scheme on global ensemble forecast. *19th Conf. on Probability and Statistics*, New Orleans, LA, Amer. Meteor. Soc., 1.1, <https://ams.confex.com/ams/88Annual/webprogram/Paper134165.html>.
- , and Coauthors, 2014: Climatology-calibrated precipitation analysis at fine scales: Statistical adjustment of Stage IV toward CPC gauge-based analysis. *J. Hydrometeorol.*, **15**, 2542–2557, <https://doi.org/10.1175/JHM-D-11-0140.1>.
- Hunke, E. C., W. H. Lipscomb, and A. K. Turner, 2010: Sea-ice models for climate study: Retrospective and new directions. *J. Glaciol.*, **56**, 1162–1172, <https://doi.org/10.3189/002214311796406095>.
- Kim, H.-M., P. J. Webster, V. E. Toma, and D. Kim, 2014: Predictability and prediction skill of the MJO in two operational forecasting systems. *J. Climate*, **27**, 5364–5378, <https://doi.org/10.1175/JCLI-D-13-00480.1>.
- Kirtman, B. P., and Coauthors, 2014: The North American Multi-model Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, <https://doi.org/10.1175/BAMS-D-12-00050.1>.
- Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the

- NCEP GFS. Part II: 4DEnVar and hybrid variants. *Mon. Wea. Rev.*, **143**, 452–470, <https://doi.org/10.1175/MWR-D-13-00350.1>.
- , D. F. Parrish, J. C. Derber, R. Treadon, W. S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691–1705, <https://doi.org/10.1175/2009WAF2222201.1>.
- Kurihara, Y., M. A. Bender, R. E. Tuleya, and R. J. Ross, 1995: Improvements in the GFDL hurricane prediction system. *Mon. Wea. Rev.*, **123**, 2791–2801, [https://doi.org/10.1175/1520-0493\(1995\)123<2791:IITGHP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<2791:IITGHP>2.0.CO;2).
- Liebmann, B., and C. A. Smith, 1996: Description of a complete (interpolated) outgoing longwave radiation dataset. *Bull. Amer. Meteor. Soc.*, **77**, 1275–1277.
- Lindzen, R. S., 1987: On the development of the theory of the QBO. *Bull. Amer. Meteor. Soc.*, **68**, 329–337, [https://doi.org/10.1175/1520-0477\(1987\)068<0329:OTDOTT>2.0.CO;2](https://doi.org/10.1175/1520-0477(1987)068<0329:OTDOTT>2.0.CO;2).
- Ling, J., P. Bauer, P. Bechtold, A. Beljaars, R. Forbes, F. Vitart, M. Ulate, and C. Zhang, 2014: Global versus local MJO forecast skill of the ECMWF model during DYNAMO. *Mon. Wea. Rev.*, **142**, 2228–2247, <https://doi.org/10.1175/MWR-D-13-00292.1>.
- Liu, Q., S. J. Lord, N. Surgi, Y. Zhu, R. Wobus, Z. Toth, and T. Marchok, 2006: Hurricane relocation in global ensemble forecast system. *27th Conf. on Hurricanes and Tropical Meteorology*, Monterey, CA, Amer. Meteor. Soc., P5.13, <https://ams.confex.com/ams/pdfpapers/108503.pdf>.
- Lorenz, E., 1969a: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, <https://doi.org/10.3402/tellusa.v21i3.10086>.
- , 1969b: Three approaches to atmosphere predictability. *Bull. Amer. Meteor. Soc.*, **50**, 345–349.
- , 1982: Low-order models of atmospheric circulations. *J. Meteor. Soc. Japan*, **60**, 255–267, https://doi.org/10.2151/jmsj1965.60.1_255.
- Luo, Y. R., and Coauthors, 2016: The Southern China Monsoon Rainfall Experiment (SCMREX). *Bull. Amer. Meteor. Soc.*, **98**, 999–1013, <https://doi.org/10.1175/BAMS-D-15-00235.1>.
- Ma, J., Y. Zhu, D. Hou, X. Zhou, and M. Peña, 2014: Ensemble transform with 3D rescaling initialization method. *Mon. Wea. Rev.*, **142**, 4053–4073, <https://doi.org/10.1175/MWR-D-13-00367.1>.
- Madden, R., and P. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708, [https://doi.org/10.1175/1520-0469\(1971\)028<0702:DOADOI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2).
- , and —, 1972: Description of global-scale circulation cells in the tropics with a 40–50 day period. *J. Atmos. Sci.*, **29**, 1109–1123, [https://doi.org/10.1175/1520-0469\(1972\)029<1109:DOGSCC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1972)029<1109:DOGSCC>2.0.CO;2).
- Melhauser, C., W. Li, Y. Zhu, X. Zhou, M. Peña, and D. Hou, 2016: Exploring the impact of SST on the extended range NCEP Global Ensemble Forecast System. *STI Climate Bulletin*, NWS/Office of Science and Technology Integration, Silver Spring, MD, 30–34, http://www.nws.noaa.gov/ost/climate/STIP/41cdpw_digest.htm.
- Pegion, K., and P. Sardeshmukh, 2011: Prospects for improving subseasonal predictions. *Mon. Wea. Rev.*, **139**, 3648–3666, <https://doi.org/10.1175/MWR-D-11-00004.1>.
- Ramsay, B. H., 1998: The Interactive Multisensor Snow and Ice Mapping System. *Hydrol. Processes*, **12**, 1537–1546, [https://doi.org/10.1002/\(SICI\)1099-1085\(199808/09\)12:10<1537::AID-HYP679>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1099-1085(199808/09)12:10<1537::AID-HYP679>3.0.CO;2-A).
- Rex, D., 1950: Blocking action in the middle troposphere and its effect upon regional climate. *Tellus*, **2**, 275–301.
- Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517, <https://doi.org/10.1175/JCLI3812.1>.
- , and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, <https://doi.org/10.1175/2010BAMS3001.1>.
- , and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Sela, J., 1980: Spectral modeling at the National Meteorological Center. *Mon. Wea. Rev.*, **108**, 1279–1292, [https://doi.org/10.1175/1520-0493\(1980\)108<1279:SMATNM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<1279:SMATNM>2.0.CO;2).
- Shelly, A., P. Xavier, D. Copsey, T. Johns, J. M. Rodríguez, S. Milton, and N. Klingaman, 2014: Coupled versus uncoupled hindcast simulations of the Madden-Julian oscillation in the Year of Tropical Convection. *Geophys. Res. Lett.*, **41**, 5670–5677, <https://doi.org/10.1002/2013GL059062>.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, [https://doi.org/10.1175/1520-0477\(1993\)074<2317:EFANTG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2).
- , and —, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319, [https://doi.org/10.1175/1520-0493\(1997\)125<3297:EFANAT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2).
- , Y. Zhu, and T. Marchok, 2001: The use of ensembles to identify forecasts with small and large uncertainty. *Wea. Forecasting*, **16**, 463–477, [https://doi.org/10.1175/1520-0434\(2001\)016<0463:TUOETI>2.0.CO;2](https://doi.org/10.1175/1520-0434(2001)016<0463:TUOETI>2.0.CO;2).
- Van den Dool, H., S. Saha, and Å. Johansson, 2000: Empirical orthogonal teleconnections. *J. Climate*, **13**, 1421–1435, [https://doi.org/10.1175/1520-0442\(2000\)013<1421:EOT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1421:EOT>2.0.CO;2).
- Vitart, F., G. Balsamo, R. Buizza, L. Ferranti, S. Keeley, L. Magnusson, F. Molteni, and A. Weisheimer, 2014: Sub-seasonal predictions. ECMWF Tech. Memo. 734, 45 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2014/12943-sub-seasonal-predictions.pdf>.
- , and Coauthors, 2017: The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- Wallace, J., and D. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812, [https://doi.org/10.1175/1520-0493\(1981\)109<0784:TITGHF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2).
- Wang, B., J. Liu, H.-J. Kim, P. J. Webster, S.-Y. Yim, and B. Xiang, 2013: Northern Hemisphere summer monsoon intensified by mega-El Niño/Southern Oscillation and Atlantic multidecadal oscillation. *Proc. Natl. Acad. Sci. USA*, **110**, 5347–5352, <https://doi.org/10.1073/pnas.1219405110>.
- Wang, W., A. Kumar, J. X. Fu, and M. P. Hung, 2015: What is the role of the sea surface temperature uncertainty in the prediction of tropical convection associated with the MJO? *Mon. Wea. Rev.*, **143**, 3156–3175, <https://doi.org/10.1175/MWR-D-14-00385.1>.
- Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble-variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, <https://doi.org/10.1175/MWR-D-12-00141.1>.
- 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*, **59A**, 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](https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2).
- Whitaker, J., T. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482, <https://doi.org/10.1175/2007MWR2018.1>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.
- Wu, W., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916, [https://doi.org/10.1175/1520-0493\(2002\)130<2905:TDVAWS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2).
- Xiang, B., M. Zhao, X. Jiang, S. Lin, T. Li, X. Fu, and G. Vecchi, 2015: The 3–4-week MJO prediction skill in a GFDL coupled model. *J. Climate*, **28**, 5351–5364, <https://doi.org/10.1175/JCLI-D-15-0102.1>.
- Zhang, F., W. Li, and M. E. Mann, 2016: Scale-dependent regional climate predictability over North America inferred from CMIP3 and CMIP5 ensemble simulations. *Adv. Atmos. Sci.*, **33**, 905–918, <https://doi.org/10.1007/s00376-016-6013-2>.
- Zhou, X., Y. Zhu, D. Hou, and D. Kleist, 2016: A comparison of perturbations from an ensemble transform and an ensemble Kalman filter for the NCEP Global Ensemble Forecast System. *Wea. Forecasting*, **31**, 2057–2074, <https://doi.org/10.1175/WAF-D-16-0109.1>.
- , —, —, Y. Luo, J. Peng, and D. Wobus, 2017: The NCEP Global Ensemble Forecast System with the EnKF initialization. *Wea. Forecasting*, **32**, 1989–2004, <https://doi.org/10.1175/WAF-D-17-0023.1>.
- Zhu, Y., 2005: Ensemble forecast: A new approach to uncertainty and predictability. *Adv. Atmos. Sci.*, **22**, 781–788, <https://doi.org/10.1007/BF02918678>.
- , Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc.*, **83**, 73–83, [https://doi.org/10.1175/1520-0477\(2002\)083<0073:TEVOEB>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0073:TEVOEB>2.3.CO;2).