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

Yuejian Zhu*[1], Xiaqiong Zhou[2], Malaquias Pena[2], Wei Li[2],

Christopher Melhauser[2], and Dingchen Hou[1]

[1]EMC, NCEP, NWS, NOAA, College Park, Maryland

[2]IMSG at EMC, NCEP, NWS, NOAA, College Park, Maryland,

Submitted to Weather and Forecasting

(August 17 2017)

*Corresponding Author: Yuejian Zhu, Email: Yuejian.Zhu@noaa.gov, Environmental Modeling Center/NCEP/NOAA, 5830 University Research Court, College Park, MD 20740

ABSTRACT

The GEFS is being extended from 16-d to 35-d 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, CONUS accumulated precipitation, and MJO skill are examined using the GEFSv11 with various SST forcing configurations. The configurations consist of (1) the operational GEFS 90-d e-folding of the observed RTG-SST anomaly relaxed to climatology; (2) an optimal AMIP configuration using the observed RTG-SST analysis updated every 24-h; (3) a 2-tier approach using the CFSv2 predicted SST, updated every 24-h; and 4) a 2-tier approach using biased corrected CFSv2 predicted SST, updated every 24-h. The experiments are carried out over a six month period covering the fall and winter of 2013-14. This is the first study to examine an operational GEFS configuration in the extended-range.

The results indicate that there are minimal differences in RPSS between the various SST forcing experiments. The forecast skill of the Northern Hemisphere 2-m temperature and precipitation for weeks 3&4 are marginal, especially for North America during this period. The bias corrected CFSv2 predicted SST experiment generally had superior performance, but only had statistically significant improvement in spatially and temporally aggregated 2-m temperature RPSS 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 skill for the weeks 3&4 time range. For all configurations, the forecast MJO indices become stronger and are subject to larger error with an increase in lead-time.

1. Introduction

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 & 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. Recently, the need for numerical guidance covering the weeks 3 & 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.

Global efforts have been pursued to provide extended range forecast guidance to the public, helping to reduce the impact of high impact weather and extreme events. One such effort is the sub-season to season (S2S) project, a legacy project of The Observing System Research and Predictability Experiment (THORPEX). This project was endorsed in 2012 by the World Weather Research Program (WWRP) and World Meteorological Organization (WMO) World Climate Research Program (WCRP; Vitart et al. 2016). 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 & 4 forecast guidance (NOAA 5-year research and development plan: 2013-2017, http://nrc.noaa.gov/CouncilProducts/ResearchPlans/5YearRDPlan/NOAA5YRPHome.aspx). Since 2011, the NOAA 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 – NOAA's 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 & 4, subseasonal, and interaseasonal time periods. The seminal studies by Lorenz 1969a; 1969b; and 1982 set the foundation for understanding forecast predictability. Subsequent studies attempted to find and explain key phenomena that impact forecast predictability across temporal scales. In the tropics, the Madden-Julian Oscillation (MJO; Madden and Julian 1971, 1972) was found to be a key phenomena for extended-range forecast prediction due to its preferred 40- to 50-d 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 and 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 the summer and produce conditions conducive for severe storm in the winter (Rex, 1950). Several notable studies attempt to improve forecast skill and further understand phenomena to increase forecast skill on the subseasonal to seasonal timescales with emphasis placed on high-impact weather events (Kirtman et al., 2014).

The studies using numerical models focused on the scientific issues and relationships of key phenomena, including the MJO (Fu et al., 2013; Xiang et al., 2016), teleconnections (e.g., PNA, NAO; Dool et al., 2000; Chen and Dool, 2003), monsoons (Adams and Comrie, 1997; Chang et al., 2000; Lou et. al, 2016), extreme rainfall events (Lou et. al, 2016), sea-ice (Hunke and 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 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 for which 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 forecasts.

Operational global numerical guidance for weeks 3 & 4 and monthly prediction are available from several operational forecasting centers. NOAA’s National Center for Environmental Prediction (NCEP) Climate Forecasting System (CFS) Version 2 is a coupled (ocean, sea-ice, land, atmosphere) model (Saha et al., 2006; Saha et al., 2010; Saha et al., 2014) that combines 4 forecasts initialized 4 times daily into a daily 16 member time-lagged ensemble integrated out to 45 d with retrospective hindcasts for bias correction. The European Center for Medium-Range Weather Forecasting (ECMWF) runs a 51-member global coupled (ocean, sea-ice, land, and atmosphere) Ensemble Prediction System (EPS; Vitart et al., 2014) out to 46 d. The ECMWF EPS is initialized twice per week with a real-time hindcast for forecast calibration. Recently, Environmental Canada extended their 21 ensemble member uncoupled Global Ensemble Prediction System (GEPS; Côté et al. 1998; Buizza et al. 2005) to 32 d once-per-week with real-time forecasts for forecast calibration.

The NCEP Global Ensemble Forecast system (GEFS) has been designed to assimilate forecast uncertainty which results in improved forecast reliability (Buizza et al., 2005) in the medium-range. In recent years, GEFS has provided 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), pronounced at longer lead times. The NH

500hPa geopotential height anomaly correlation out to 16 d for the experimental period in this manuscript (fall 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 d) to cover the weeks 3 & 4 period provides additional benefits over the CFSv2, including a more frequent model upgrade cycle, higher model resolution, state-of-art flow-dependent initial perturbations from a hybrid 4DEnsVar data assimilation system, stochastic physics, and larger ensemble membership (84 members for every 24-hour cycle), all providing an improved sampling of forecast uncertainty.

In this study, the operational GEFS v11 configuration is extended to 35 d 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 contiguous United States (CONUS), and MJO indices. Section 2 describes the GEFS configuration; SST forcing experiments, experiment forecast period, and aspects of the verification methodology. Section 3 provides results and 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

2.1. Operational NCEP GEFS

The current operational configuration of GEFS uses the GFS Global Spectral Model V12.0.0 (GSM) for integration four times per day (0000, 0600, 1200 and 1800 UTC) out to 16 days (Sela 1988; Han and Pan 2011; Han et al. 2016). For days 0-8 the GEFS has a spectral resolution of T_L574 (semi-Lagrangian with 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; Wang et al., 2013; Kleist and Ide 2015; Zhou et al. 2016). If tropical cyclones are present in the initial conditions, TC perturbations are calculated after tropical cyclones (TCs) are separated from the environment (Kurihara et al. 1993, 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. 2006, 2008). The GEFS has the same GFS SST forcing which is initialized with the RTG analysis and damped to climatology (90-d e-folding) during model integration. The sea ice concentration is initialized from the daily 00 UTC sea ice and lace analysis from Interactive Multisensor Snow and Ice Mapping System (Ramsay 1998) and held constant throughout the model integration. Please see http://www.emc.ncep.noaa.gov/GFS/impl.php for additional information on GSM v12.0.0 settings used in the operational GEFS. For this study, the operational GEFS configuration is modified: (1) to extend the forecast to 35 d with the horizontal resolution reduced to T_L254 (approximately 78 km) for 16-35 d, (2) the SST is updated with various SST forcing schemes, and (3) the forecast is only initialized run once-per-day at 0000 UTC due to resource constraints.

2.2. SST Forcing Experiments

The SST configurations consist of the operational GEFS 90 d e-folding 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 2-tier approach using the CFSv2 predicted SST updated every 24-h during model integration (BC), and a 2-tier approach using biased 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 B.

2.3. Experiment Period

All experiments in this study span the fall and winter of 2013 and 2014 and initialized every 24 h starting 1 Sep 2013 and ending 28 Feb 2014. The 00 UTC initialization and corresponding 00 UTC forecast lead times (24 hour forecasts) for 2-m temperature and accumulated precipitation are verified to control for the diurnal variability found in the 2-m temperature (the 12 UTC verification lead times show similar, but slightly higher skill).

Over the experiment period, the MJO was weak or non-existent (Climate Prediction Center; 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 plain states and Colorado and New Mexico ranking within their ten 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. In Asia, Russia 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 2014, 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 meteorological winter since 1996 and coldest November to 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 southeastern states. Over the western US and Great Plains, drier-than-normal conditions persisted (National Climatic Data Center Climate Global Analysis: https://www.ncdc.noaa.gov/sotc/global).

2.4. Verification Procedure

2.4.1. Rank Probability Skill Score: 2-m Temperature and Accumulated Precipitation

The forecast skill for 2-m temperature and accumulated precipitation are evaluated using a tercile (below-normal, normal, or above-normal) Ranked Probability Skill Score (RPSS; e.g., Wilks, 2011); see Appendix A for additional details. The 2-m temperature is verified for land only against the 00 UTC GDAS analysis and the accumulated precipitation is verified for land only against the 00 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 & 4: days 15-28) and verified against the corresponding GDAS and CCPA data averaged or accumulated over the same lead times. Different methods and length of periods can be defined which can have a direct impact on forecast skill; generally longer averaging periods produces higher RPSS (not shown).

The week 2 and weeks 3 & 4 were chosen in this investigation to match the operational CPC week 2 and experimental weeks 3 & 4 forecasts.

2.4.2. MJO skill score

In this study, the MJO is evaluated using the traditional real-time multivariate MJO (i.e. RMM) 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 and winter 2013-2014 period 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 NCEP GDAS. The long term climatology is calculated from the NCEP/NCAR Reanalysis 1 (http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html) for the U200 and U850 and from the NCAR Interpolated Outgoing Longwave Radiation (http://www.esrl.noaa.gov/psd/data/gridded/data.interp_OLR.html) for the OLR, both for the period of 1981-2010. The long term mean and average of the previous 120 day 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 RMS error and spread for GEFS 500-hPa geopotential heights over the fall and winter of 2013-2014 for the lead day 18 (Fig. 2a,b) and lead day 25 (Fig. 2c,d) generally supports the notion that GEFS is performing well in the weeks 3 & 4 period, although deviations occur for other variables, lead times, and locations. For both lead day 18 and 25, the spread and error over the NH polar latitudes show similar spatial patterns and magnitudes, although slightly under dispersive over the NH storm tracks. In the SH, the GEFS appears to be slightly over dispersive over a large swath of the SH arctic circle.

Locating the source of uncertainty of the large-scale circulation is another necessary step towards a more accurate forecast for week 3 & 4 time frame. During the fall-winter of the NH, subtropical jet is one of the major large-scale circulations that modulate the NA weather. As such, demonstrating the uncertainty associated with the upper level circulation is helpful for model developer on evaluating the jet stream forecast.

The 6-month experiment period average of 200-hPa RMSE (Fig. 3) shows 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 mid-latitudes south of 30^{o}S and north of 30^{o}N. This suggests that for weeks 3 &4 forecasts, improving the skill of the large-scale circulation, especially over the subtropical jet region, shouldn’t be ignored.

The operational GEFS is an uncoupled system with the sea surface temperature (SST) prescribed using the NCEP real time SST analysis (RTG) 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. Figure 4 shows the area-average SST over the 15^{o}S-15^{o}N band for RTG (Fig. 4a) analysis and lead day 20 forecasts valid at the corresponding analysis verification date for the CTL RTG persisted SST damped to climatology (Fig. 4b), raw CFSv2 (Fig. 4c), and bias corrected CFSv2

(Fig. 4d). Comparing the coupled CFSv2 model output to CTL, the CFSv2 provides additional multi-scale information. Removing systematic biases in the CFSv2 model output (Fig. 4d) improves the correlation between the RTG analysis and the lead day 20 CFSv2 forecast data.

3.1. 2-m Temperature Forecast Skill

Over the experimental period, the global land only 2-m temperature RPSS is regionally and lead time dependent. The tropics have the highest RPSS for both week 2 (Fig. 5a and weeks 3 & 4 (Fig. 5b) with NA having the lowest. Comparing between week 2 and weeks 3 & 4, the RPSS remains similar for the tropics and SH with the NH and NA dropping ~0.1-0.3. Within each region, the forecast skill for the SST forcing experiments are generally statistically indifferent from CTL for both week 2 and weeks 3 & 4. RTG, CFS and CFS_BC show a statistically significant improvement during weeks 3 & 4 over NA with RTG showing statistically significant improvements over TR. It is interesting that RTG does not have more robust improvement compared to the other experiments, given this experiment is being forced with the observed SST forcing. During weeks 3 & 4 over NA, CFS and CFS_BC actually outperform the RTG experiment in terms of RPSS. This suggests that there may be deficiencies in the forecast model which are limiting the spread of information from the ocean boundary to atmospheric land areas. 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 & 4 spatial 2-m temperature RPSS score for CTL (Fig. 6a) indicates the highest skill over land extending from the western Sahara into the 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 grid-point 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. Minimal differences over land can also be found comparing CTL to CFS (Fig. 6c) and CFS_BC (Fig. 6d) experiments. In general, the experiments forced with the CFS (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 a degradation in RPSS over the northern high latitudes due to differences in modeling or representing sea ice. Also, along the western portion of South America and extending to the eastern equatorial region, the CFSv2 is known to overproduce low-level clouds and bias the SST. 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 US and Canada that were observed during the experiment period. This cold bias is reduced in RTG (Fig. 7b), CFS (Fig. 7c), and CFS_BC (Fig. 7d), and corresponds to improve RPSS in central NA. The dynamic sea ice in both CFS and CFS_BC indicate large regions of the northern high latitudes that were cooler than CTL with the sea ice/ocean boundary clearly evident.

Specifically comparing CFS (Fig. 7c) and CFS_BC (Fig. 8d), 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 the sea-ice and prescribing the SST and potential discrepancies between the model sea ice. However, the CFS_BC clearly reduces the bias over the western US extending into northern Mexico, improving the RPSS (Fig. 6d). Additionally, CFS_BC significantly reduces the warm bias in CFS over the western 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 this area.

The weeks 3 & 4 time frame falls within the gray zone between the weather and climate, thus one way to highlight the “sub-seasonal” time scale and increase the predictability is to remove the short-term noise associated with the synoptic weather using a 5-day running mean. The 5-day running mean RMSE 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 the Asia continent. 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 US. 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 RMSE provide the forecast error, signal-to-noise ratio (SNR) directly indicates the predictability for a certain forecast variable (Wang et al. 2013; Zhang et al. 2016). For 2-m temperature, the predictability mainly occurs over the tropical regions (Fig. 9). Over the western CONUS, there is more predictability compared with the central US, but overall the predictability is low. It should be noted the GEFS is under dispersive in 2-m temperature, especially in the tropics.

The minimal 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 cannot 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 & 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 and further investigation needs to be performed to determine if this trend holds over other forecast variables and verification metrics.

3.2. Accumulated Precipitation Forecast Skill - CONUS

Over the fall and winter of 2013-2014, the CONUS accumulated precipitation RPSS shows no statistically significant difference between CTL and RTG, CFS, or CFS_BC for week 1 (not shown), week 2 (Fig. 5b), or weeks 3 & 4 (Fig. 5d). The magnitude of the RPSS falls off drastically after week 1 - approx. 0.55 at lead day 1 and 0 at lead day 7 (die off curves not shown) - leveling off around approx.. 0 (no skill) for all experiments for the extended period. The aggregate accumulated week 2 RPSS is slightly higher than weeks 3 & 4, but overall, the results suggest minimal skill with the current model configurations, regardless of SST forcing.

The distribution of weeks 3 & 4 accumulated precipitation RPSS for CTL (Fig. 10) indicates the highest skill is over the northern plains with minimal or negative skill across the southwest, south central plains, and southeast. Comparing the RPSS differences from CTL, all CFS SST forcing experiments (Fig. 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 RPSS 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 are large enough in magnitude be statistically significant. The RTG and CFS_BC experiments have large coherent regions reduction of the dry bias in the central and portions of the eastern U.S. The minimal differences in bias between SST forcing experiments suggests the systematic model errors from model parameterizations dominate the biases at the extended period.

3.3. 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 version of 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 observations (RTG), the MJO forecast skill was improved up to ~2 days. For the weeks 3 & 4 range, the most skillful SST forcing is RTG with the CFS_BC being the most skillful scheme that could be practically used in operations.

The MJO skill averaged for weeks 3 & 4 was improved by ~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 works (Wang et al. 2015). Therefore, without changing the model, it is found that improving the SST results in an increase of the MJO skill.

The strength and variability of the MJO index are subject to forecast errors, increasing with lead time. Over the experiment period, the MJO is predicted to be weaker in September, late November-mid December of 2013, late January and February, 2014 but stronger over all other periods (Fig. 13). The bias in MJO strength was consistent across lead times. For longer lead time (e.g lead day=22), the forecast MJO indices tend to become stronger in most verification months except for December. Although the weeks 3 & 4 forecast MJO magnitude is generally too strong and slightly out of phase, there are some periods that GEFS performed well, for example, the late November - early December period for lead day 14. The investigation of this is left to future study since the purpose of the paper is to present the general skill of the GEFS for the weeks 3 & 4 time range.

4. Conclusions and Future Work

The NCEP GEFS is being extended from 16 d to 35 d to cover the subseasonal forecast period. The impact 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 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 2-tier approach using the CFSv2 predicted SST, updated every 24-h; and 4) a 2-tier approach using biased corrected CFSv2 predicted SST, updated every 24-h. The experiments are carried out over a six month period covering the fall and winter months of 2013-2014. This period was characterized by weak MJO events and a neutral ENSO conditions.

There was minimal to no improvement in land only 2-m temperature and accumulated precipitation found over the extended weeks 3 & 4 period. Forcing the GEFS with an optimal

SST setup did not show statistically significant improvements. This indicates there are deficiencies that need to be addressed. The GEFS in its current configuration cannot effectively propagate the information contained in the tropical SSTs to land regions around the globe. For accumulated precipitation over the CONUS, the minimal differences in RPSS between experiments and overall during the weeks 3 & 4 period along with the minimal differences in bias between SST forcing experiments also 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 skill during the experimental period for the operational GEFS is ~ 14.6 days. Using 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 time (e.g lead day=22), the forecast MJO indices tend to become stronger in most verification months except for December.

Overall, the one-way forcing of GEFS with more realistic SSTs does enhance MJO skill, but it does not significantly improve NA weather (2-m temperature and precipitation). This implies an (1) inherent predictability issue for NA weather over the weeks 3 & 4 period and that future work needs to be performed (2) to improve the GEFS model as well as (3) to improve boundary forcing such as sea ice, snowpack and soil moisture for potential gain in weeks 3 & 4 skill. Also, observations indicate that the fall and winter of 2013-2014 has a generally weak MJO. Future work will focus on a two year span that covers a stronger MJO period spanning 1 May 2014 to 31 May 2016 providing insight into the predictability from strong MJO and their relationship with 2-m temperature and CONUS accumulated precipitation from global teleconnections. Therefore, further experiments with higher resolution GEFS with improved model stochastic physics have been designed to improve MJO prediction for the period of 2014-2016.

ACKNOWLEDGEMENTS: The authors would like to thank Drs. Qin Zhang, Ping Liu for MJO discussion; Drs. Xingren Wu, Wanqiu Wang on the discussion of SST configuration; Dr. Shrinivas Moorthi's valuable comments through EMC internal review process. This work has been supported by NWS/NGGPS project and OAR/CPO/MMAP.

APPENDIX A

Rank Probability Skill Score

The Rank Probability Skill Score (RPSS) measures the improvement of a multi-category ensemble forecast relative to a reference forecast. It ranges from -inf 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 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 The RPSS is calculated as

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

where the forecast Ranked Probability Score ($RPSf$) is calculated as

$$RPS_f = \frac{1}{N} \cdot \sum_{k=0}^{N} \quad [(probB_n - obsB_n)^2 + (probN_n - obsN_n)^2 + (probA_n - obsA_n)^2]$$

with n corresponding to each forecast-observation pair, N are the total number of forecast-observation pairs, $probX_n$is the ranked cumulative forecast probability for each bin X, and $obsX_n$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 set to ⅓ since each forecast bin is defined as climatologically equal. See Wilks 2011 or the Climate Prediction Center (http://www.cpc.ncep.noaa.gov/products/verification/summary/index.php?page=tutorial) for more information.

APPENDIX B

SST Forcing Calculations

Operational GEFS SST Forcing (CTL):

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

$$SST_f^t = \left[SST_a^{t_0} - SST_c^{t_0}\right] e^{(t-t_0)/90} + SST_c^{t_0}$$

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

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-d period, calculated as

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

where f is the forecast, a 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

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

REFERENCES:

Adams, D. and A. Comrie, 1997: The North American Monsoon. *Bull. Amer. Meteor. Soc.*, **78**, 2197–2213, doi: 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, doi: 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, doi: 10.1175/MWR2905.1.

Chang, C., 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, doi: 10.1175/1520-0442(2000)013<4310:IAIVOT>2.0.CO;2.

Chang, C.-P. and et. al., 2004: East Asian Monsoon, World Scientific Series on Asia-Pacific Weather and Climate, Vol. 2. 1-572. ISBN: 978-981-270-141-1

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, doi: 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, doi: 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 Dynamics*, **41**, 1067, doi:10.1007/s00382-013-1859-9.

Gottschalck, 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, doi: 10.1175/2010BAMS2816.1.

Guan, H. and Y. Zhu, 0: Development of verification methodology for extreme weather forecasts. *Wea. Forecasting*, **0**, doi: 10.1175/WAF-D-16-0123.1.

Hou, D., Z. Toth, Y. Zhu, and W. Yang, 2008: Evaluation of the impact of the stochastic perturbation schemes on global ensemble forecast. Proc. 19th Conf. on Probability and Statistics, New Orleans, LA, Amer. Meteor. Soc. [Available online at https://ams.confex.com/ams/88Annual/webprogram/Paper134165.html.]

Han, J. and H. Pan, 2011: Revision of Convection and Vertical Diffusion Schemes in the NCEP Global Forecast System. *Wea. Forecasting*, **26**, 520–533, doi: 10.1175/WAF-D-10-05038.1.

Han, J., M. L. Witek, J. Teixeira, R. Sun, H-L Pan, J. K. Fletcher, 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, doi:10.1175/WAF-D-15-0053.1.

Hou, D., M. charles, Y. Luo, Z. Toth, Y. Zhu, R. Krzysztofowicz, Y. Lin, P. Xie, D. 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. Hydrometeor.*, **15**, 2542–2557, doi: 10.1175/JHM-D-11-0140.1.

Hunke, E. C., W. H. Lipscomb, A. K. Turner, 2-10: Sea-ice models for climate study: retrospective and new directions. *J. of Glaciology*, **56**, 1162-1172, doi: 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, doi:10.1175/JCLI-D-13-00480.1.

Kirtman B . P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, doi:10.1175/BAMS-D-12-00050.1.

Kleist, D. T., 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.

Kleist, D. 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, doi: 10.1175/MWR-D-13-00350.1.

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, doi:10.1175/MWR-D-13-00292.1.

Lindzen, R. S., 1987: On the development of theory of the QBO. *Bull. Amer. Meteor. Soc.*, **68**, 329-337, doi:10.1175/1520-0477(1987)068<0329:OTDOTT>2.0.CO;2.

Liu, Q., S. J. Lord, N. Surgi, Y. Zhu, R. Wobus, Z. Toth and T. Marchok, 2006: Hurricane Relocation in Global Ensemble Forecast System, Preprints, 27th Conf. on Hurricanes and Tropical Meteorology, Monterey, CA, Amer. Meteor. Soc., P5.13.

Lorenz, E., 1969a: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, doi:10.1111/j.2153-3490.1969.tb00444.x

Lorenz, E., 1969b: Three approaches to atmosphere predictability. *Bull. Amer. Meteor. Soc.*, **50**, 345–349.

Lorenz, E., 1982: Low-order models of atmospheric circulations. *J. Meteor. Soc. Japan*, **60**, 255-267.

Luo, Y., R. and Coauthors, 2016: The Southern China Monsoon Rainfall Experiment (SCMREX). *Bull. Amer. Meteor. Soc.*, **0**, doi: 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, doi: 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, doi: 10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2.

Madden, R. and P. Julian, 1972: Description of Global-Scale Circulation Cells in the Tropics with a 40–50 Day Period. *J. Atmos. Sci.*, **29**, 1109–1123, doi: 10.1175/1520-

0469(1972)029<1109:DOGSCC>2.0.CO;2.

National Academies of Sciences, Engineering, and Medicine, 2016: Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts. Washington, DC. *The National Academies Press*. doi: 10.17226/21873.

Melhauser. C. W. Li, Y. Zhu, X. Zhou, M. Pena and D. Hou, 2016: Exploring the Impact of SST on the Extended Range NCEP Global Ensemble Forecast System, STI Climate Bulletin: http://www.nws.noaa.gov/ost/climate/STIP/41cdpw_digest.htm pages 30-34

NOAA National Centers for Environmental Information, State of the Climate: National Overview for November 2013, published online December 2013, retrieved on November 30, 2016 from http://www.ncdc.noaa.gov/sotc/national/201311.

NOAA National Centers for Environmental Information, State of the Climate: National Overview for February 2014, published online March 2014, retrieved on November 30, 2016 from http://www.ncdc.noaa.gov/sotc/national/201402.

Pegion, K. and P. Sardeshmukh, 2011: Prospects for Improving Subseasonal Predictions. *Mon. Wea. Rev.*, **139**, 3648–3666, doi: 10.1175/MWR-D-11-00004.1.

Ramsay, B.H., 1998. The interactive multisensor snow and ice mapping system. *Hydrological Processes*, **12**, 1537-1546.

Rex, D., 1950: Blocking Action in the Middle Troposphere and its Effect upon Regional Climate, Tellus, **2**, 275-301, doi:0.1111/j.2153-3490.1950.tb00331.x.

Saha, S. and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517, doi: 10.1175/JCLI3812.1.

Saha, S. and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi: 10.1175/2010BAMS3001.1.

Saha, S. and Coauthors, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27**, 2185–2208, doi: 10.1175/JCLI-D-12-00823.1.

Sela, J., 1980: Spectral modeling at the National Meteorological Center, Mon. Wea. Rev., 108, 1279-1292.

Shelly, A., P. Xavier, D. Copsey, T. Johns, J. M. Rodr´ıguez, 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, doi:10.1002/2013GL059062.

Toth, Z. and E. Kalnay, 1993: Ensemble Forecasting at NMC: The Generation of Perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, doi: 10.1175/1520-

0477(1993)074<2317:EFANTG>2.0.CO;2.

Toth, Z. and E. Kalnay, 1997: Ensemble Forecasting at NCEP and the Breeding Method. *Mon. Wea. Rev.*, **125**, 3297–3319, doi: 10.1175/15200493(1997)125<3297:EFANAT>2.0.CO;2.

Toth, Z., Y. Zhu, and T. Marchok, 2001: The Use of Ensembles to Identify Forecasts with Small and Large Uncertainty. *Wea. Forecasting*, **16**, 463–477, doi: 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, doi: 10.1175/1520-0442(2000)013<1421:EOT>2.0.CO;2.

Vitart F, and Coauthors, 2014: Sub-Seasonal Predictions, ECMWF Research Department Technical Memorandum 734. ECMWF: Reading, UK.

Vitart, F., C. and Coauthors, 2017: The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, doi: 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, doi: 10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2.

Wang, W. and Coauthors, 2014: MJO prediction in the NCEP climate forecast system version 2. *Clim. Dyn.*, **42**, 2509–2520, doi: 10.1007/s00382-013-1806-9.

Wang, W., Kumar, A., Fu, J.X. and Hung, M.P., 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.

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 (doi: 10.1175/MWR-D-12-00141.1).

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 Ni~no/Southern Oscillation and Atlantic multidecadal oscillation. *Proc. Natl. Acad. Sci.* USA, 110, 5347–5352.

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, doi:10.1111/j.1600-0870.2007.00273.x.

Wei, M., Z. Toth, R. Wobus, Y. Zhu, C. H. Bishop, X. Wang, 2006: Ensemble Transform Kalman Filter-based ensemble perturbations in an operational global prediction system at NCEP. Tellus, **58A**, 28-44, doi:10.1111/j.1600-0870.2006.00159.x.

Wheeler, M. 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, doi: 10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

Wheeler, M. C., H. Zhu, A. H. Sobel, D. Hudson and F. Vitart, 2016: Seamless precipitation prediction skill comparison between two global models, *Q. J. R. Meteorol. Soc.*, **143**, 374-383, doi:10.1002/qj.2928.

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, doi: 10.1175/2007MWR2018.1.

Wilks, D.S., 2011. Statistical methods in the atmospheric sciences. *Academic press*. Vol. **100**.

Wu, W., R. J. Purser, and D. F. Parrish, 2002: Three dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, 130, 2905-2916.

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.

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, doi: 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.

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, doi: 10.1175/WAF-D-16-0109.1.

Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng and D. Wobus, 2016: The NCEP Global Ensemble Forecast System with the EnKF Initialization. *Submitted to Wea. and Forecasting*, In Process.

Zhu, Y., 2005: Ensemble Forecast: A New Approach to Uncertainty and Predictability. Advance in Atmospheric Sciences, **22**, 781-788, doi:10.1007/BF02918678.

Zhu, Y., G. Iyengar, Z. Toth, M. S. Tracton, and T. Marchok, 1996: Objective evaluation of the NCEP global ensemble forecasting system. Preprints, 15th Conf. on Weather Analysis and Forecasting, Norfolk, VA, Amer. Meteor. Soc., J79–J82.

Zhu, Y., 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, doi: 10.1175/1520-0477(2002)083<0073:TEVOEB>2.3.CO;2.

FIGURE CAPTION LIST

Figure 1. Average Anomaly Correlation by lead day for 500-hPa geopotential heights over the Northern Hemisphere covering the period of 1 September 2013 to 28 February 2014 for the deterministic GFS (blue) and the GEFS ensemble mean (red).

Figure 2. Spatial distribution of 5-day running mean RMS error (left column) and ensemble spread (right column) of 500-hPa geopotential heights for CTL over the 6-month experiment period for lead day 18 (top row) and 25 (bottom row).

Figure 3. Spatial distribution of RMS error of 200-hPa u-component of wind for CTL over the 6-month experiment for lead day 18 (top row) and lead day 25 (bottom row).

Figure 4. Hovmoller diagrams of SST area-average over the 15S-15N band for the (a) RTG analysis, (b) CTL initial conditions, and (c) CFSv2 and (d) bias corrected CFSv2 SST forecast at lead day 20. The three panels on the right verify with the dates of the RTG on the left. Time-longitude correlation is given for each SST forecast panels.

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

Figure 6. Land only 2.5^{o} global 2-m temperature Rank Probability Skill Score averaged over the 6-month experimental period for weeks 3 & 4 for (a) CTL and the difference from CTL of RTG (b), CFS (c), and CFS_BC (d). Hatching on (b,c,d) indicates the difference is statistically significant at the 95% confidence level.

Figure 7. Land only 2.5^{o} global 2-m temperature bias averaged over the 6-month experiment period for weeks 3 & 4 for CTL (a) and the difference from CTL of RTG (b), CFS (c), and CFS_BC (d).

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

Figure 9: The 2-m temperature signal-to-noise ratio for CTL averaged over weeks 3 & 4.

Figure 10. Spatial weeks 3 & 4 accumulated precipitation Rank Probability Skill Score over the CONUS averaged over the 6-month experimental period for CTL (a) and the difference from CTL of RTG (b), CFS (c), and CFS_BC (d).

Figure 11. Spatial weeks 3 & 4 accumulated precipitation bias over the CONUS averaged over the 6-month experimental period for CTL (a) and the difference from CTL of RTG (b), CFS (c), and CFS_BC (d).

Figure 12. MJO forecast skill (i.e. bivariate correlation between ensemble mean forecast and

analysis data) as a function of lead time for the period of September 1, 2013 - February 28, 2014. Climatology and previous 120-day mean are removed from the forecast and analysis data while calculating the RMMs.

Figure 13. a) MJO forecast skill (i.e. bivariate correlation between ensemble mean forecast and analysis data) as a function of lead time for the period of September 1, 2013 - February 28, 2014. Climatology and previous 120-day mean are removed from the forecast and analysis data while calculating the RMMs. b) Average of the MJO skill for weeks 3 & 4 (averaged over lead day 15-28)

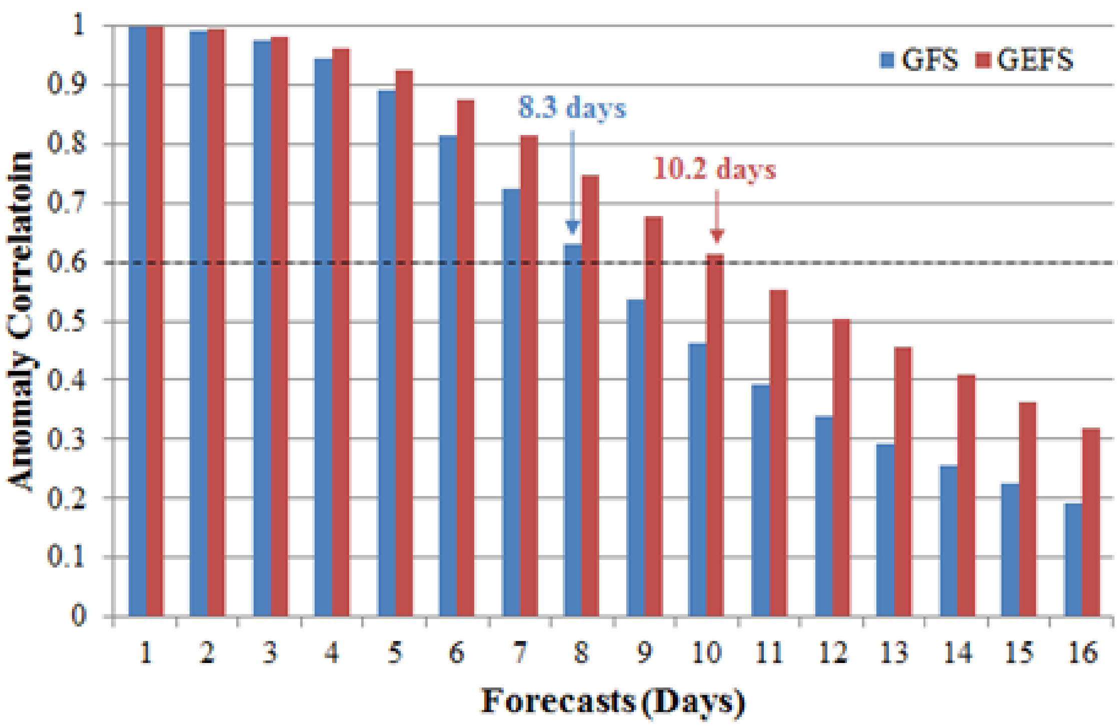

Figure 1. Average Anomaly Correlation by lead day for 500-hPa geopotential heights over the Northern Hemisphere covering the period of 1 September 2013 to 28 February 2014 for the deterministic GFS (blue) and the GEFS ensemble mean (red).

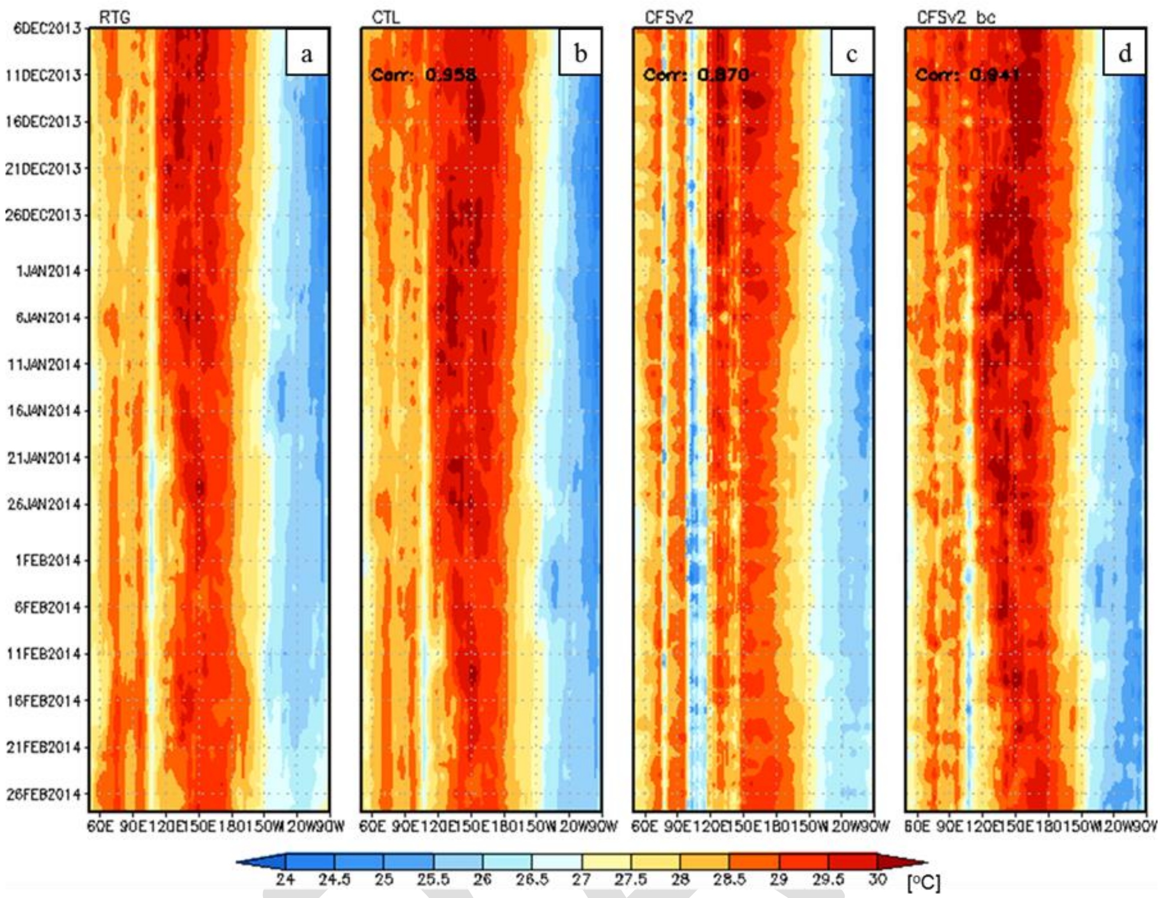

Figure 2. Hovmoller diagrams of area-average SST [K] over the 15°S-15°N band for the (a) RTG analysis, (b) CTL initial conditions, and (c) CFSv2 and (d) bias corrected CFSv2 SST forecast at lead day 20. The three panels on the right verify with the dates of the RTG on the left. Time-longitude correlation is given for each SST forecast panels.

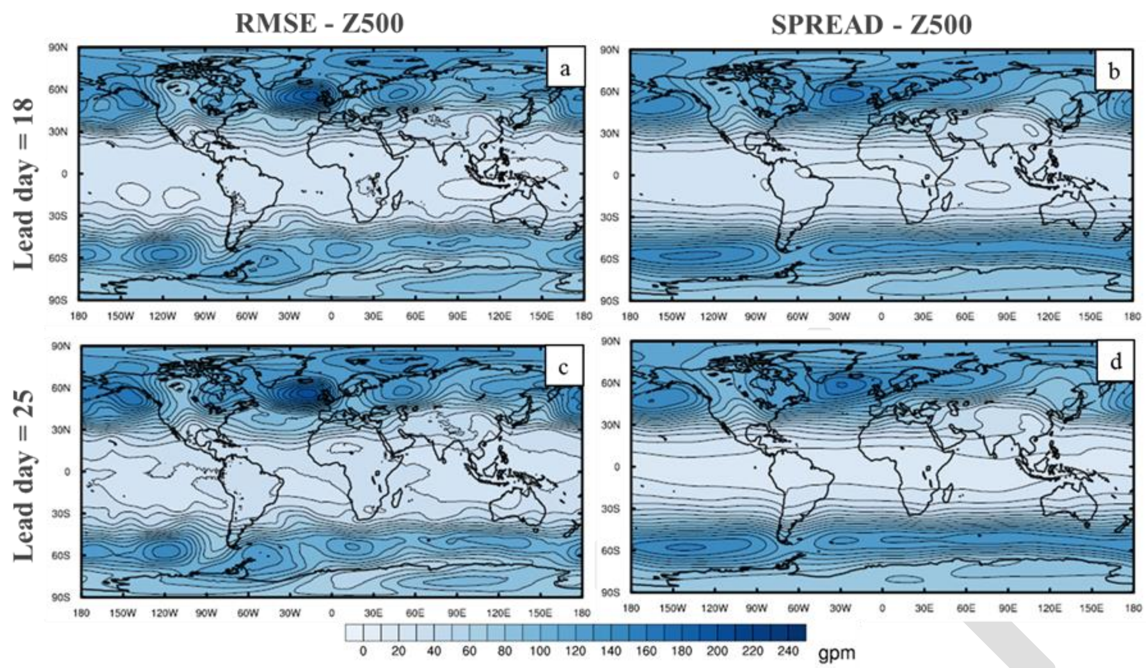

Figure 3. Spatial distribution of 5-day running mean RMS error (left column) and ensemble spread (right column) of 500-hPa geopotential heights [gpm] for CTL over the 6-month experiment period for lead day 18 (top row) and 25 (bottom row).

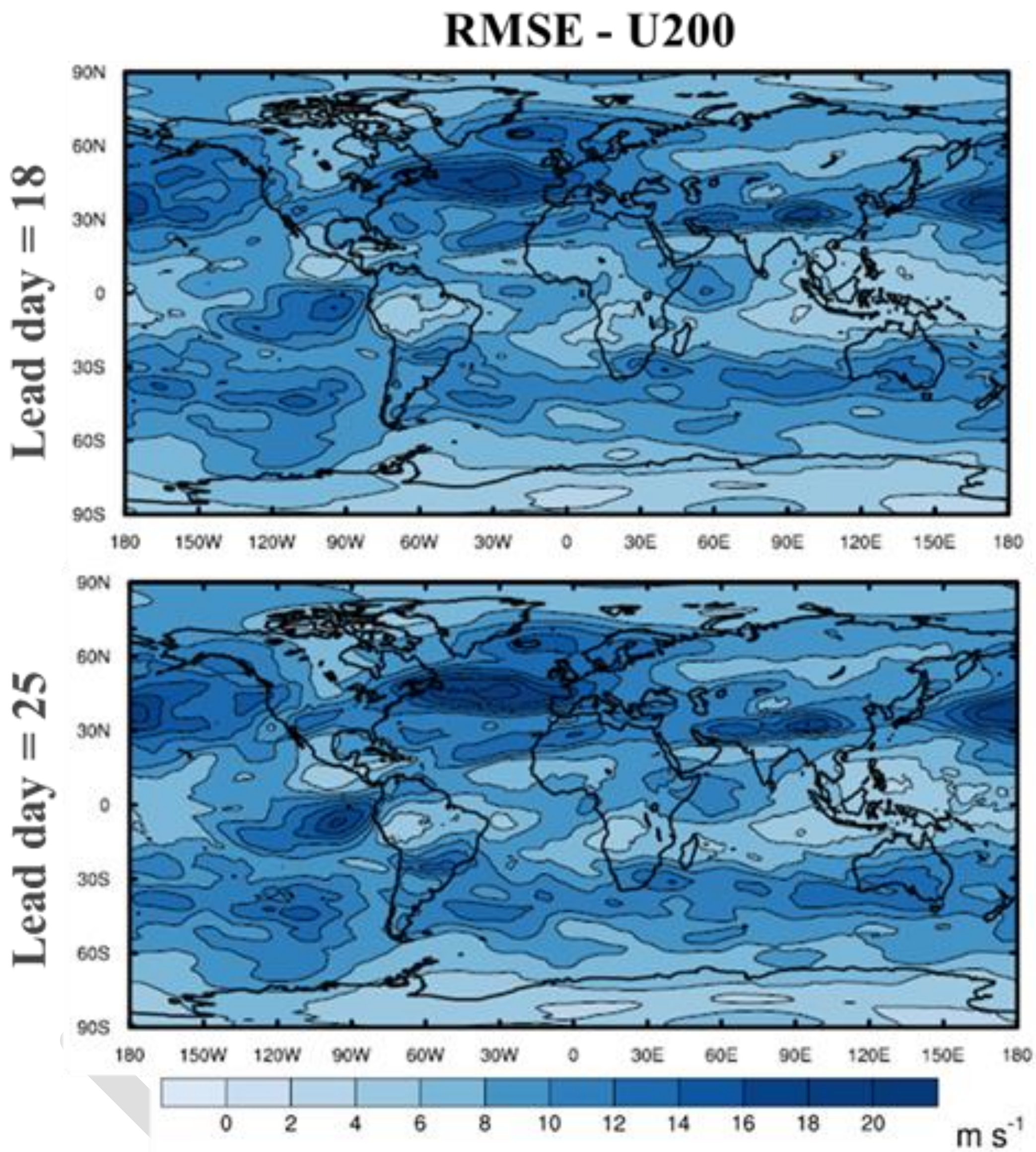

Figure 4. Spatial distribution of RMS error of 200-hPa u-component of wind [m s^{-1}] for CTL over the 6-month experiment for lead day 18 (top row) and lead day 25 (bottom row).

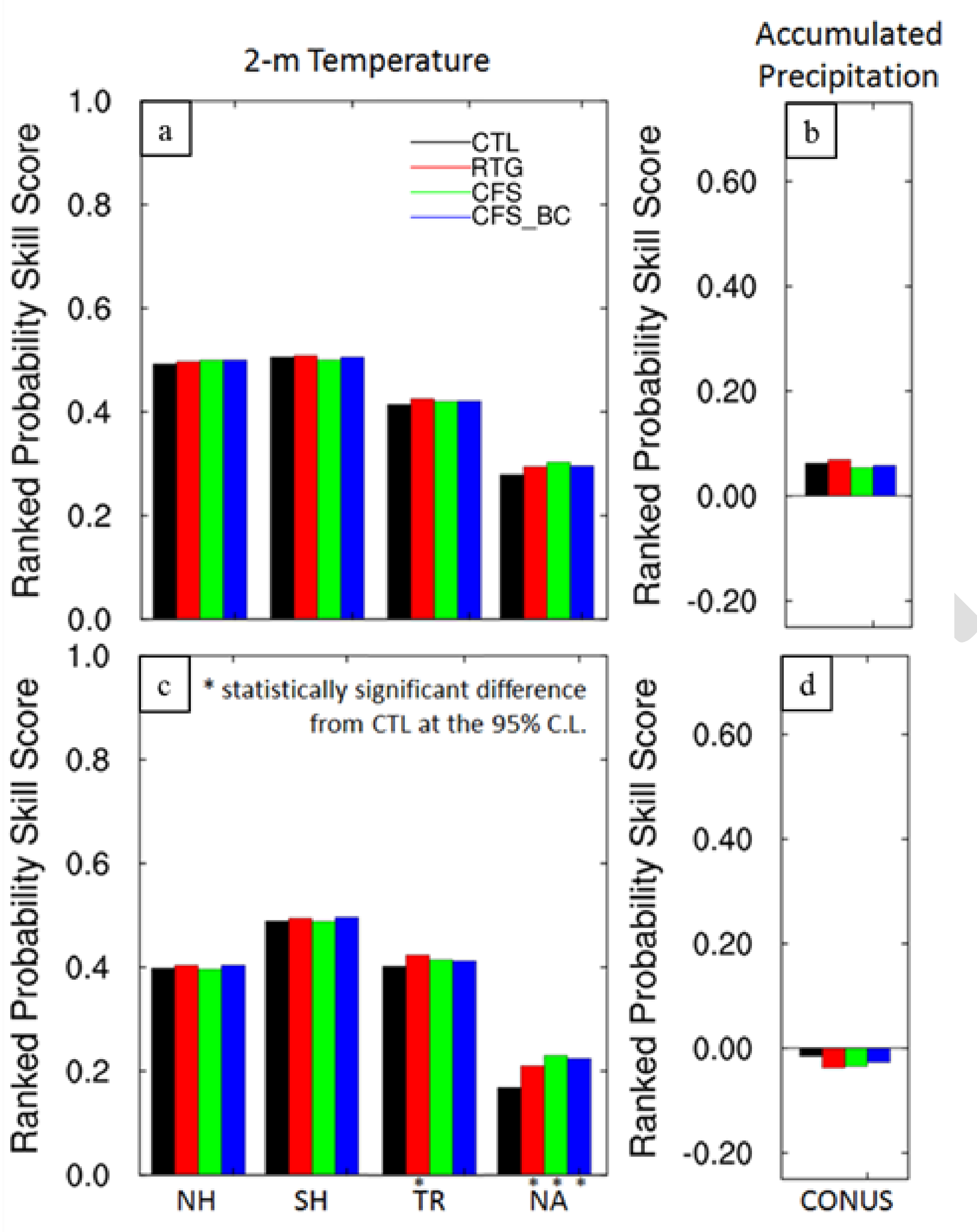

Figure 5. Rank Probability Skill Score for CTL (black), RTG (red), CFS (green), and CFS_BC (blue) calculated for week 2 (top row) and weeks 3 & 4 (bottom row) for 2-m temperature (a,c) and accumulated precipitation (b,d) averaged over the 6-month experiment period. Asterisks beneath the respective experiment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

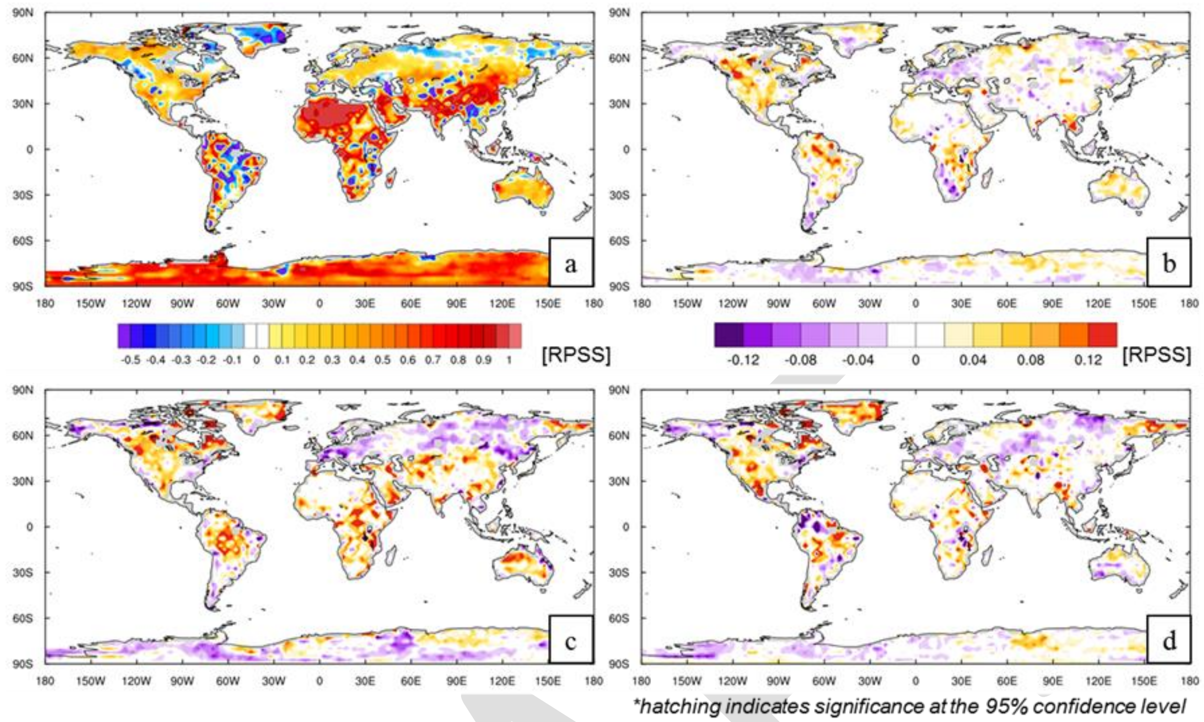

Figure 6. Land only 2.5o global 2-m temperature Rank Probability Skill Score averaged over the 6-month experimental period for weeks 3 & 4 for (a) CTL and the difference from CTL of RTG (b), CFS (c), and CFS_BC (d). Hatching on (b,c,d) indicates the difference is statistically significant at the 95% confidence level.

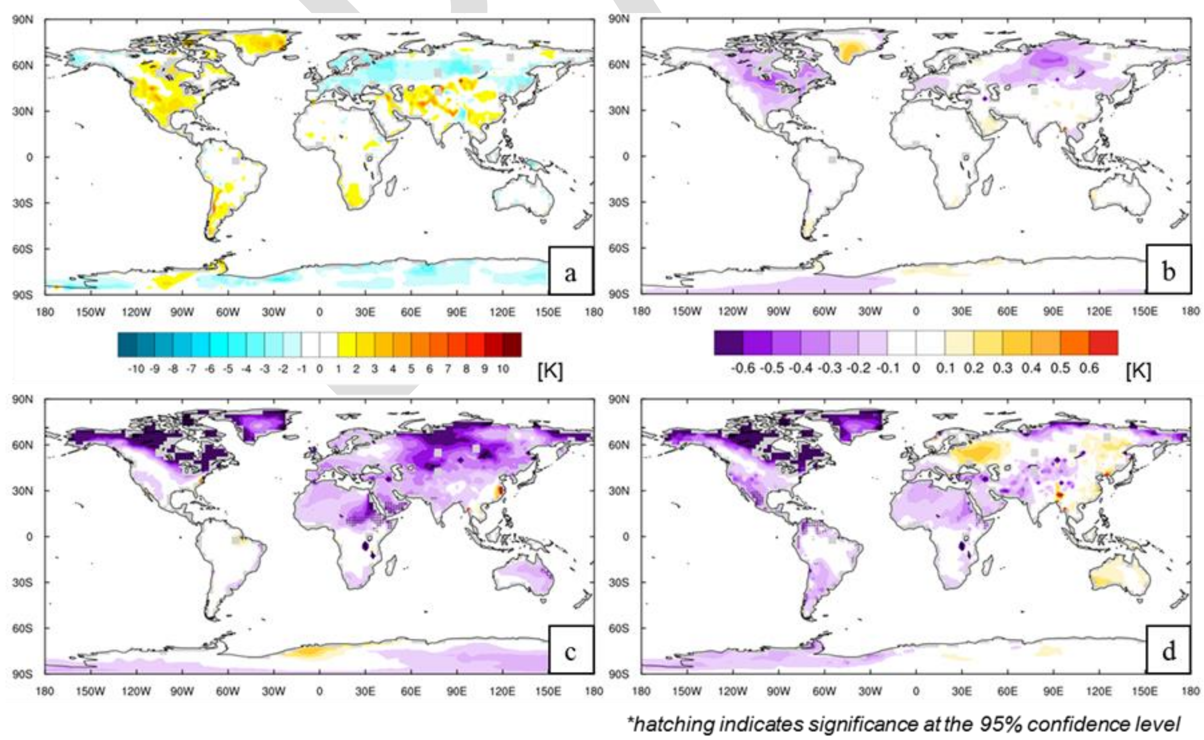

Figure 7. Land only 2.5^{o} global 2-m temperature bias [K] averaged over the 6-month experiment period for weeks 3 & 4 for CTL (a) and the difference from CTL of RTG (b), CFS (c), and CFS_BC (d).

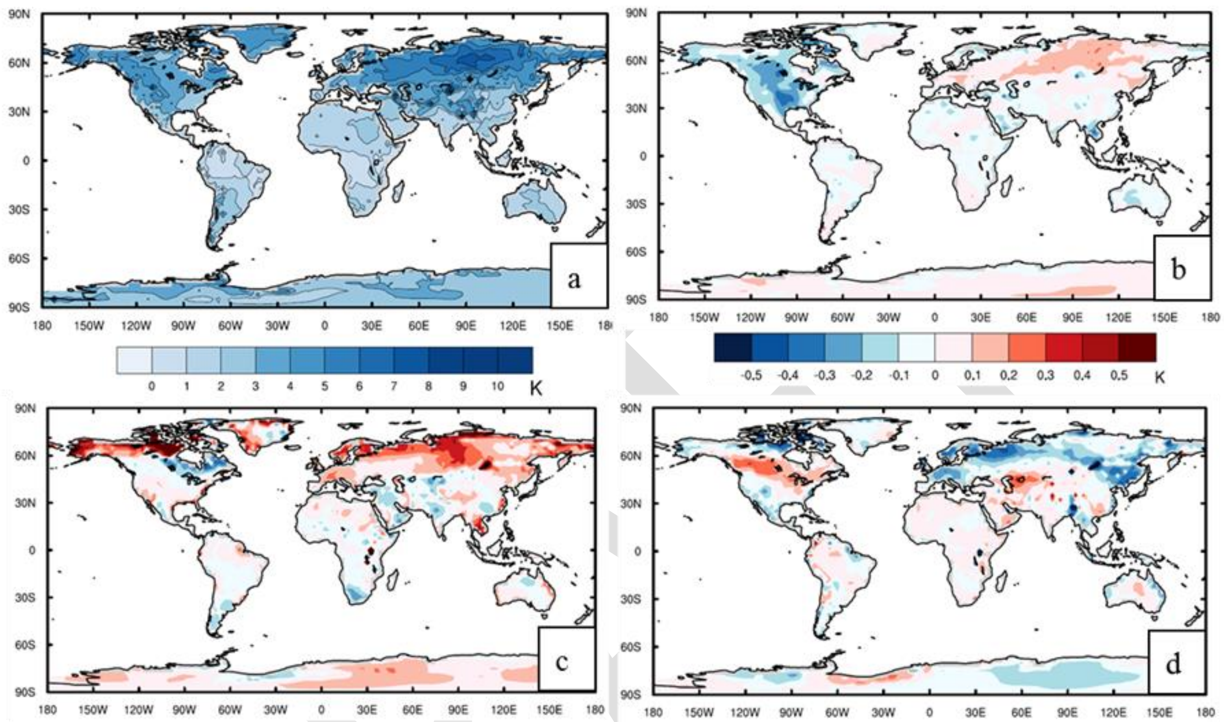

Figure 8. Land only 2.5^{o} 2-m temperature RMS error [K] for (a) CTL and the difference between CTL and RTG (b), CFS (c), and CFS_BC (d) averaged over the 6-month experiment period.

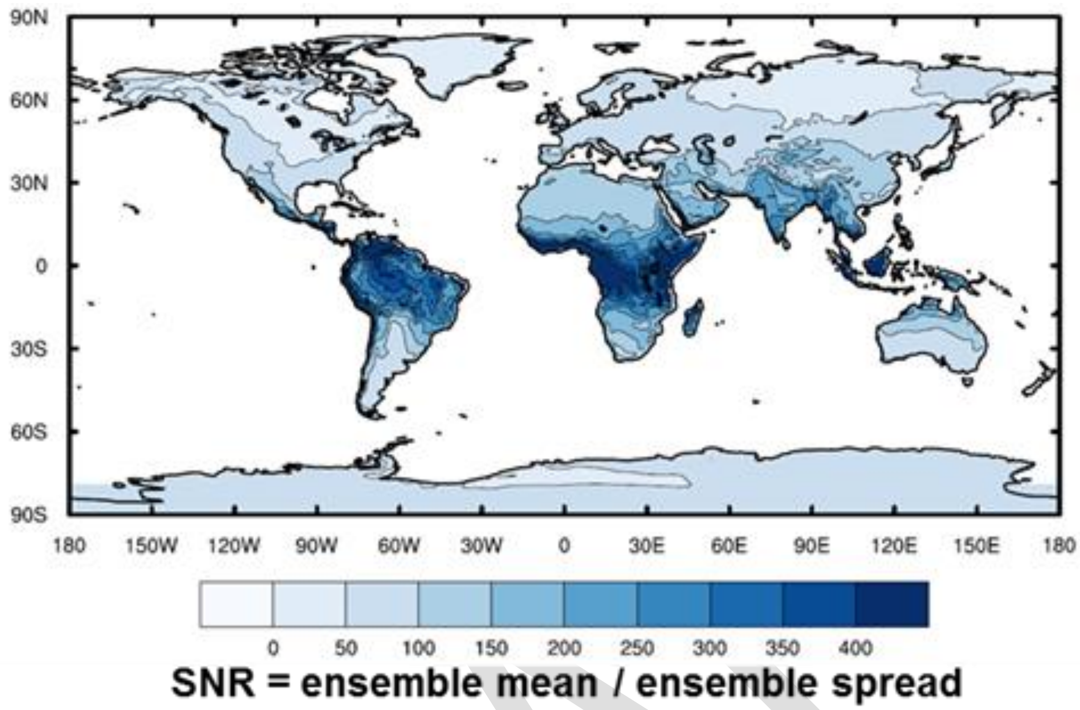

Figure 9: The 2-m temperature signal-to-noise ratio for CTL averaged over weeks 3 & 4.

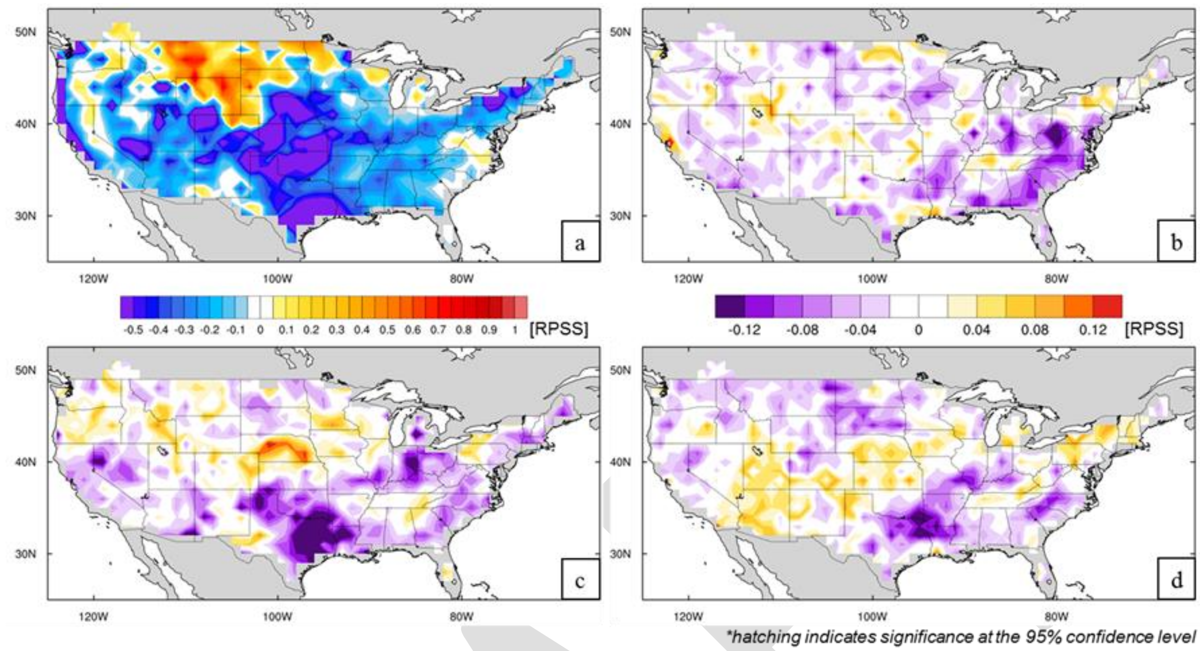

Figure 10. Spatial weeks 3 & 4 accumulated precipitation Rank Probability Skill Score over the CONUS averaged over the 6-month experimental period for CTL (a) and the difference from CTL of RTG (b), CFS (c), and CFS_BC (d).

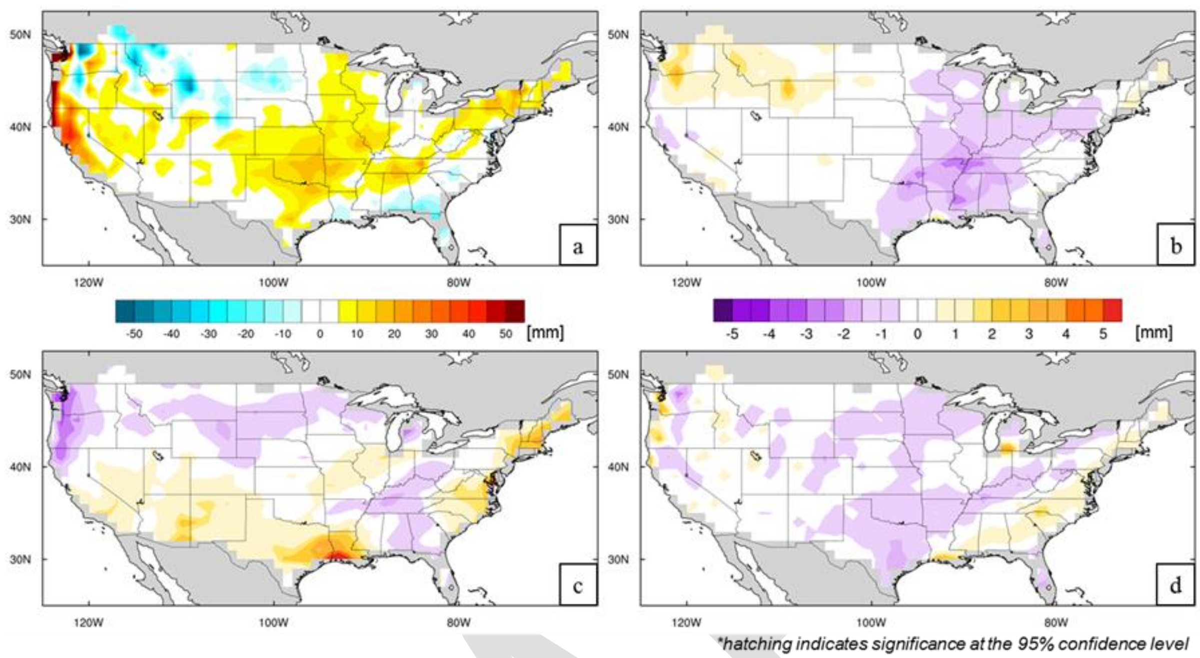

Figure 11. Spatial weeks 3 & 4 accumulated precipitation bias [mm] over the CONUS averaged over the 6-month experimental period for CTL (a) and the difference from CTL of RTG (b), CFS (c), and CFS_BC (d).

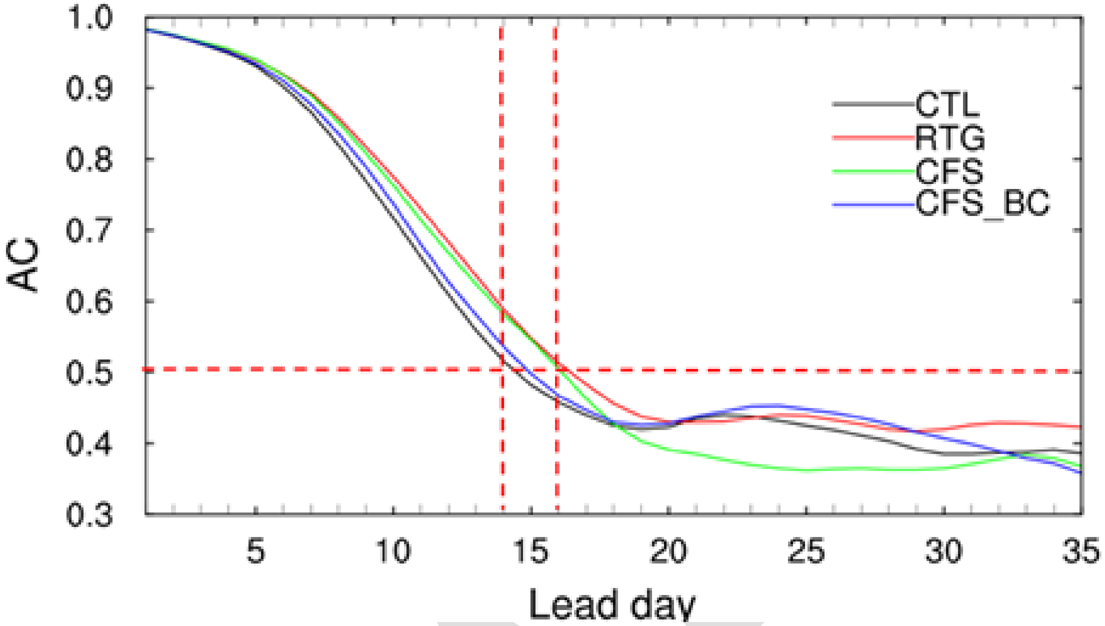

Figure 12. MJO forecast skill (i.e. bivariate correlation between ensemble mean forecast and analysis data) as a function of lead time for the period of September 1, 2013 - February 28, 2014. Climatology and previous 120-day mean are removed from the forecast and analysis data while calculating the RMMs.

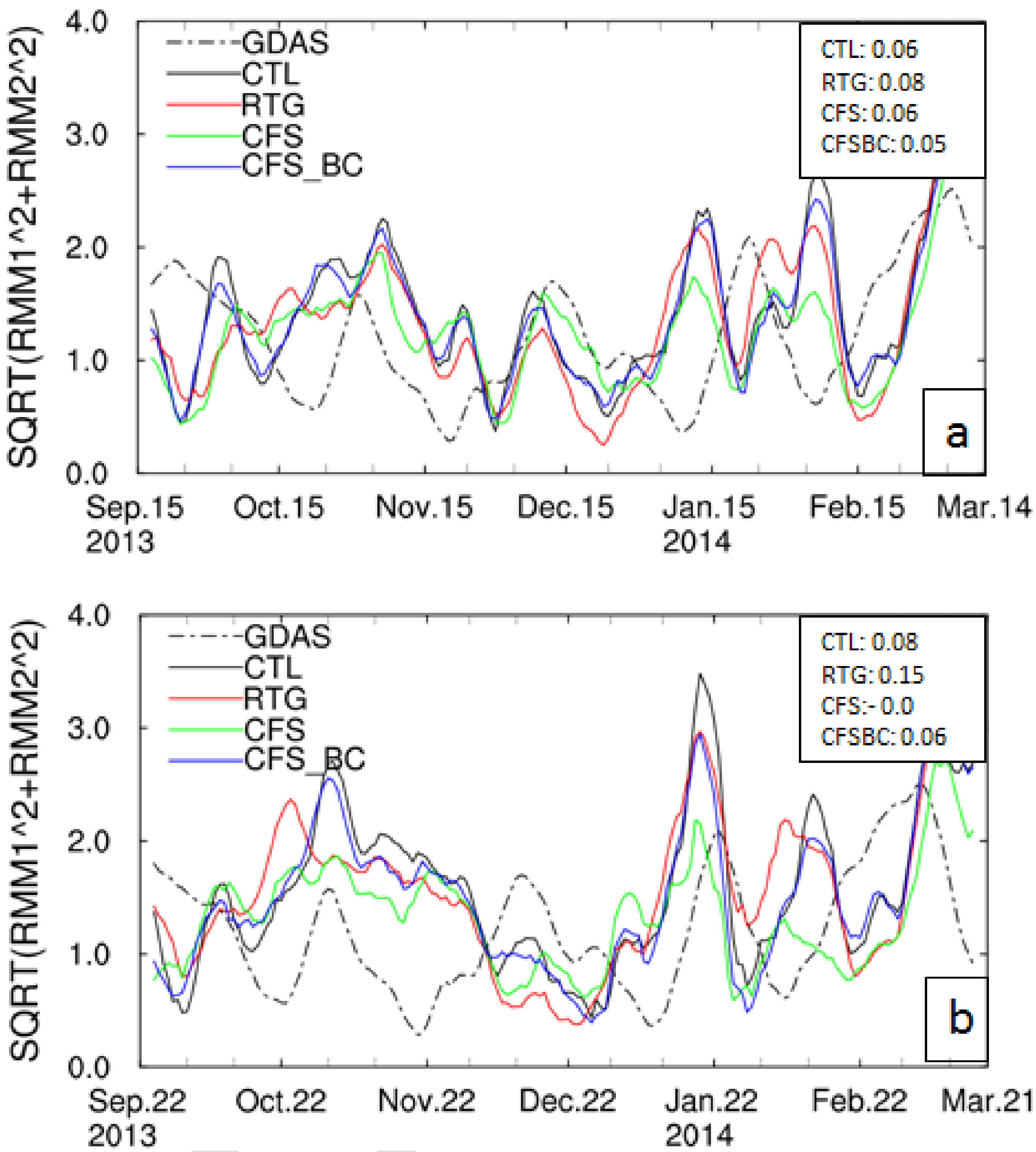

Figure 13. MJO index for different lead time. a). for lead day 14; b) for lead day 21. 7-point running mean is applied on the time series to smooth the data. Numbers in the text box are the variance of each experiment from the analysis for all initial times.