Response to the Editor:

Thank you for your time in editing our manuscript and for your insightful suggestions. We have responded to the reviews point by point in the document attached. The revised manuscript with tracked marks is also attached behind the responses to the editor. We have submitted the revised manuscript as requested.

1. The paper's treatment of the significance of the results is still less rigorous than I feel is appropriate for the forecasting topic. For instance, even statements such as (p5 l1) "The HSS above 0 indicates that the forecasts have skill" (positive, negative?) are simplistic. The skill estimate is certainly much stronger (ie significant) if the HSS has been calculated from a sample of 1000 obs-forecast pairs, compared to a sample of 10. I recommend adding to the discussion of the skill score a small general discussion regarding uncertainty (due to sampling error) in the Heidke skill score estimates, but related to the samples sizes used in this paper – ie, 28. It should be fairly straightforward to calculate a confidence interval given this sample size, which can then be referenced in the results discussion (Figures 3-10 excluding 5). For instance, given sampling uncertainty with N=28 the HSS is significantly (positively, $p<0.05$ or $p<0.10$) skillful when it is greater than X (where X is greater than zero).

RESPONSE: Thanks for pointing this out. We used a bootstrapping technique to estimate the confidence interval of the HSS, and the average of the HSS over the United States as the test statistic. We have added the following statement to the methodology section of the revised manuscript: "Since the number of forecast-observation pairs is 28 for each point, the HSS estimation has considerable uncertainty given this relatively small sampling size. To quantify this uncertainty, a bootstrapping technique (Wilks 2011; Hamilton et al. 2012) was applied to resample 28 samples (3000 times with replacement) from the 28-year reforecast averaged over the CONUS. Then a number of 3000 HSS was calculated for constructing a distribution; with the confidence interval and significance level of the HSS estimated from this distribution. With this treatment of the HSS estimation uncertainty, we can determine that the HSS is significantly skillful when it is greater than a given significance level." We have calculated confidence intervals for different indices using this method and referenced that in the result discussion: "It is worth noting that given the sample size (N=28) used for calculating the HSS, the confidence interval of the HSS for each index is relatively wide. Based on the bootstrapping technique described earlier, the HSS was found to be significantly skillful (significantly above 0 at the 0.05 level) when it was greater than the number between 20 and 24, depending on the indices."

For the difference between two HSS, the CI may be more difficult to calculate analytically thus the use of the bootstrap is a convenient empirical approach. Yet the authors use of the bootstrap on the sample of 6 mixed scores (3 months + 3 leads) is at least inadequate if not completely incorrect. One strategy would be to generate 1000 trials of two 28-member nonskilled samples (perhaps by shuffling the obs-forecast pairs in time) and calculate score difference thresholds that are exceeded with a desired probability, purely by chance. Another is to bootstrap similarly on the actual 28-member samples to assess the effect of their sample uncertainty on the difference in their scores. Because the HSS is a widely-used metric in weather/climate forecasting, and it is likely a common challenge to assess whether one forecast is better than another, I expect examples can be

found in the literature. It is still a striking feature of figures 12-13 that adjacent pixels with different signals (eg -15, +10) or signals of zero are all found significant, when the underlying climate maps (NLDAS and CFSv2) are much smoother. Despite an earlier request for analysis and discussion on this point, this issue remains unexplained.

RESPONSE: Thanks for your comments and suggestions. We used a bootstrap technique for significance tests for the difference between the HSS during active MJO and the whole period. We resampled 28 samples (3000 times with replacement) from the 28-year reforecast averaged over the CONUS. All the 28 samples were used to calculate the HSS during the whole period. The subset of the 28 samples under active MJO events was used to calculate the conditional HSS. The difference between the HSS during the MJO and the all period was then calculated. Since the resampling was conducted 3000 times, a number of 3000 HSS differences was obtained for constructing a distribution and used to estimate the confidence interval and significance level of the HSS. Similar to Peng et al. (2013), the significance level estimated based on the average over the CONUS was applied to test the local significance for each grid point over the region. The results have been updated in Figures 12 and 13 and relevant texts in the revised manuscript.

Finally, the use of the significance calculations in the figures can be improved. Rather than show the significance maps separately (eg in Figure 11), which makes it very difficult to match a pixel's significance and value, the example of skill masking at CPC could be followed – eg, http://www.cpc.ncep.noaa.gov/products/CFSv2/htmls/usPrece1SeaMask.html)

RESPONSE: Thanks for your suggestions. We have used mask to improve the use of the significance calculations in the Figures 12 and 13.

Before this paper can be accepted it will require a more rigorous and thoughtful treatment of uncertainties in the skill score estimates, referencing appropriate literature and clearly describing the approaches used to assign significance.

RESPONSE: Thanks for your comments. We have addressed those issues in the revised manuscript. Please refer to the responses to the comments above.

2. In general, I feel the authors have adequately addressed the reviewers' comments, but I ask that the authors go back through the first round of comments to reconsider and upgrade any perfunctory and limited responses such as the one highlighted below.

Reviewer: Page 2, line 23-25: Please explain how GCM outputs can be used for daily or short-term forecasts seeing as they are uncorrelated to current meteorological conditions.

RESPONSE: Thanks for pointing this out. This is an overstatement. We have changed the sentence to "Coupled Atmosphere-Ocean General Circulation Models (GCMs) are used to make forecasts at multiple timescales."

There are a number of interesting reasons why GCM outputs can and are being used for daily and short-term forecasts when they are initialized for weather and climate prediction, yet these are not discussed in the response or, more appropriately, the paper (as background, perhaps).

The reviewer may be confused by the use of GCMs in free-running climate projection mode (where there is no correlation) rather than in operational forecasting mode, and here it would have been appropriate to make the distinction and to describe briefly the sources of GCM predictability at short-to-medium ranges (eg the initializations, inertial dynamics at different scales and in different components, etc.).

RESPONSE: Thanks for your comments. We again went through the reviewers' comments and found that the explanation for the use of GCMs outputs is not sufficient. We have added the following to clarify this in the revised manuscript:

"Coupled Atmosphere–Ocean General Circulation Models (GCMs) are used to make forecasts at multiple timescales, from medium-range weather forecasting, seasonal climate predictions, and long-term climate projections. The reason GCMs can be used as operational models at these time scales is due to the predictability from different sources, such as initial conditions from the atmosphere, inertial dynamics from soil moisture and sea surface temperature, etc."

3. The writing still requires a careful proof-reading by an accomplished technical English language writer.

RESPONSE: Thanks for your comments. We have conducted a careful proof-reading to improve the writing of this manuscript.

CFSv2-based sub-seasonal precipitation and temperature forecast skill over the contiguous United States

Di Tian[1,2], Eric F. Wood[1], and Xing Yuan[3,2]

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey 08544, USA
[2]Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, Alabama 36849, USA
[3,2]CAS Key Laboratory of Regional Climate-Environment for Temperate East Asia (RCE-TEA)RCE-TEA, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

Correspondence to: Di Tian (dtiantiandi@auburnprinceton.edu)

Abstract. This paper explored the potential of a global seasonal climate forecast models for sub-seasonal forecasting of precipitation and 2-m air temperature. The probabilistic categorical sub-seasonal forecast skill of ten precipitation and temperature indices was investigated using the 28-years' sub-seasonal hindcasts of thefrom the Climate Forecast System version 2 (CFSv2) over the contiguous United States (CONUS). The forecast skill for mean precipitation and temperature as well as for the frequency and duration of extremes was highly dependent on the target forecasting indices, regions, seasons, and leads, and methods. Indices characterizing mean precipitation and temperature as well as measuring frequency or duration of precipitation and temperature extremes for 7-, 14-, and 30-day forecasts were skillful depending on seasons, regions, and forecast leads. Forecasts for 7- and 14-day temperature indices showed skill even at weeks 3 and 4, and generally were more skillful than precipitation indices. Overall, temperature indices showed higher skill than precipitation indices over the entire CONUS region at sub-seasonal scale. While the forecast skill related to mean precipitations indices were was low in summer over the CONUS, the number of rainy days, number of consecutive rainy days, and number of consecutive dry days showed considerable high skill for the west coastal region. The 30-day forecasts of precipitation and temperature indices calculated from the downscaled monthly CFSv2 forecasts were less skillful than those calculated from the daily CFSv2 forecasts, suggesting the potential usefulness of the CFSv2 daily forecasts for hydrological applications relative to the temporally disaggregated CFSv2 monthly forecasts. While Tthe presence of active Madden-Julian Oscillation (MJO) events improved CFSv2 weekly mean precipitation forecast skill over most partsmajor areas of CONUS, but it MJO and/or El Niño Southern Oscillation did not necessarily improve the weekly mean temperature forecasts. The 30-day forecasts of precipitation and temperature indices calculated from the downscaled monthly CFSv2 forecasts were less skillful than those calculated directly from CFSv2 daily forecasts, suggesting the usefulness of CFSv2 for sub-seasonal hydrological forecasting.

1. Introduction

Sub-seasonal (or intra-seasonal) timescale forecasts are typically between medium-range weather forecasts (1 or 2 weeks) and seasonal climate predictions (1 to 12 months). The medium-range weather forecast is strongly influenced by atmospheric initial

Formatted: Superscript

conditions (Vitart et al. 2008), while the seasonal climate forecast depends on slowly-evolving components of the climate system (e.g. sea surface temperature and soil moisture) (Troccoli 2010). However, since the sub-seasonal timescale is usually too long to be favoured by the atmospheric initial conditions (Vitart 2004) and too short to be strongly influenced by the variability of the ocean, making skillful sub-seasonal forecasts is particularly difficult and thus far has~~ve received much less attention~~less progress than the medium-range weather forecasts and seasonal climate forecasts.

Since many extreme events (e.g., flash drought, heat wave, and cold wave) and their corresponding management decisions fall into sub-seasonal timescales, accurate sub-seasonal forecast information will be central to the development of climate services and therefore has~~promise~~ great socio-economic value (Vitart et al. 2012). In fact, sub-seasonal forecast information can be useful~~d~~ for developing strategies for proactive natural disaster mitigation~~, which may be needed during those extreme events~~ (Brunet et al. 2010; Vitart et al. 2012). Previous studies have evaluated the potential of sub-seasonal to seasonal forecasts for heat wave forecasting (e.g. Hudson et al. 2011a; White et al. 2014), hydrological forecasting (e.g. Orth and Seneviratne 2013; Yuan et al. 2014), water resources management (e.g. Sankarasubramanian et al. 2009), hydropower production management (e.g. Garcia-Morales and Dubus, 2007), and crop yield prediction (e.g. Hansen et al. 2006; Zinyengere et al. 2011). ~~-~~Due to the improvement of numerical models, prediction techniques, and computing resources, there is an increasing focus on sub-seasonal forecasts (e.g. Toth et al. 2007; Vitart et al. 2008; Brunet et al. 2010; Hudson et al. 2011b; Hudson et al. 2013; Robertson et al. 2014).

Precipitation and 2-m temperature (hereafter temperature) are considered ~~to be~~as two of the most important climate variables that significantly influence irrigation scheduling, urban water supply, cooling water related to thermal power generation, hydropower operations, etc. Many important sub-seasonal events including heat waves, cold waves, dry spells, and wet spells are directly derived from the frequency, duration, and intensity of rainfall or hot (cold) temperature~~s~~. ~~However, most of the studies for precipitation and temperature forecasts only focused on their mean or accumulated totals (e.g. Roundy et al. 2015).~~ While several studies have been conducted to forecast the duration of high temperature days (i.e. heat waves) (e.g. Hudson et al. 2011a; Luo and Zhang 2012; White et al. 2014), there has been, thus far, no complete investigation of sub-seasonal forecasting capabilities for the other temperature and precipitation indices that are directly associated with important events and decision-making ~~at sub-seasonal timescales~~. In this study, we aim to evaluate the skill of sub-seasonal forecasting for those precipitation and temperature ~~derivatives or~~ indices including ~~that are associated with those important events and decision-making at sub-seasonal timescale -- the~~ mean, frequency, duration~~,~~ and intensity of precipitation and temperature at sub-seasonal timescale, such as the number of dry/wet days, number of cold/hot days, etc.

Coupled Atmosphere–Ocean General Circulation Models ~~-~~(~~GCM~~GCMs) are used to make forecasts at multiple timescales, from medium-range weather forecasting, seasonal climate predictions,~~-~~ and long-term climate projections. ~~While GCMs are known for producing long-term climate projections to understand and predict climate change at centurial timescale, they have also been running in operational forecasting mode by major weather and climate centers to make relatively short-term weather~~

and seasonal climate forecasting. The reason that GCMs can be used as operational models at these time scales is due to the predictability from different sources, such as initial conditions from the atmosphere, inertial dynamics from soil moisture and sea surface temperature, etc. While the GCMGCMs have demonstrated advanced configurations and realistic representations of the climate systems, the use of GCMGCMs' predictions is still restricted by their coarse resolution and inherent systematic biases. To overcome these limitations, the GCMGCMs' predictions at seasonal timescales are usually downscaled and bias corrected before being used in hydrological applications (e.g. Wood et al., 2002; Luo and Wood, 2008; Yuan et al., 2013; Tian et al., 2014). The Climate Forecast System version 2 (CFSv2) is a recently developed GCMGCM by the National Centers for Environmental Prediction (NCEP) (Saha et al. 2014). The CFSv2 model has run retrospectively to produce forecasts (hereafter reforecasts or hindcasts) every month 5 days from 1982 to 2009. Despite the availability of those CFSv2 daily hindcasts, temporal downscaling of the seasonal predictions is still routinely done from monthly to daily without using any of the daily forecast information (e.g. Yuan et al. 2013), with the assumption that the accuracy of daily information is limited at seasonal time scale. At the sub-seasonal timescale, the usefulness of these daily or sub-daily precipitation or temperature forecasts compared to the monthly disaggregated forecasts has not been assessed. The CFSv2 has fully coupled atmospheric, oceanic, and land components of the climate systems and demonstrated high better performance for seasonal climate predictions when compared to other seasonal forecast models (Yuan et al. 2011). Since sub-seasonal precipitation or temperature forecasts are influenced jointly by the conditions of atmosphere, land, and ocean, the CFSv2 has great potential to make skillful precipitation or temperature forecasts at sub-seasonal timescales.

Besides GCMGCMs, teleconnections between large-scale climate patterns and local weather events have also been used to develop sub-seasonal precipitation or temperature forecasts. Recent examples included sub-seasonal winter temperature forecasts in North America using Madden-Julian Oscillation (MJO) or El Niño Southern Oscillation (ENSO) conditions (Yao et al., 2011; Rodney et al., 2013; Johnson et al., 2013). In addition, Jones et al. (2011) found that the deterministic forecast skill of the CFSv1 for extreme precipitation in the contiguous United States (CONUS) during winter is higher when the MJO is active. With the updated version of CFS, the CFSv2 hindcasts allow re-examining this issue by assessing the influence of MJO or ENSO on the probabilistic temperature and precipitation forecast skill over the CONUS.

This study will conduct a comprehensive evaluation of the precipitation and temperature hindcasts at sub-seasonal timescales. Specifically, the aims of this study are to 1) assess the CFSv2 predictions for precipitation and temperature indices at different locations and seasons within the first 30 days, 2) compare weekly and fortnight forecasting skill of the CFSv2 at different lead times, and 3) evaluate the effects of MJO and ENSO on the CFSv2 sub-seasonal forecast skill. The assessment includes mean values of sub-seasonal predictions as well as related temperature and precipitation indices at different forecast leads and scales. The downscaled CFSv2 monthly forecasts are compared with the native CFSv2 daily sub-seasonal forecasts. Furthermore, the influence of MJO or ENSO conditions on the CFSv2 probabilisticcategorical temperature and precipitation forecast skill is also assessed.

2. Data and Methodology

The CFSv2 hads the most updatedstate-of-the-art data assimilation and forecast model components of the climate system and became operational at NCEP since March 2011 (Saha et al. 2014). There CFSv2 has archivedwere three different types of hindcasts (or reforecasts): 6-hourly time series from 9-month runs, 45-day runs, and season runs. (Table 1) shows those forecasts with different configurations. Figure 1 gives an example of the three hindcast configurations. The CFSv2 hindcast data hads a T126 spatial resolution (roughly 100 km) and includeds several near surface variables at a 6-hourly temporal resolution. The one season and 45-day reforecasts wereare initialized every day so that relatively new initial conditions couldan be incorporated into a large ensemble size for making a potentially more skillful forecast over this shorter forecast period. Nevertheless, we chose to use the 9-month reforecast. This is because the 9-month reforecast covered much longer period (1982-2009) than one season and 45-day reforecasts (1999-2010), which ensures a larger sample size for a more robust evaluation, especially for the evaluation of skill conditioned on MJO and/or ENSO that require long hincasts for to provide sufficient sample sizes f.

[Insert Table 1 here]

[Insert Figure 1 here]

The daily precipitation total was aggregated from the 6-hourly precipitation data; the daily mean temperature was obtained by averaging daily maximum and minimum temperature, which were extracted from the 6-hourly maximum and minimum temperature. The ensemble members for each month were constructed in the same way as the monthly hindcasts from CFSv2 producing monthly means hindcasts. For each year, the daily hindcast had 28 members in November and 24 members in other months with initial conditions at the 0, 6, 12, and 18 UTC (Coordinated Universal Time) every 5 days. For example, the 24 ensemble members for January wereare initialized from the four cycles for each of December 12th, 17th, 22nd, and 27th and January 1st and 6th.

The forecast validation dataset was obtainedis from the North American Land Data Assimilation System version 2 (NLDAS-2; Xia et al., 2012). The forcing dataset of the NLDAS-2 mergeds a large observation-based and reanalysis data and wasis routinely widely used to drive land surface models over the CONUS. It hads 0.125^{o} (approximately 12 km) spatial resolution and hourly temporal resolution. The NLDAS-2 hourly precipitation (temperature) data were aggregated (averaged) into daily data.

Besides using CFSv2 daily hindcasts at its native spatial resolution (hereafter CFSv2 daily), the CFSv2 monthly hindcasts were also downscaled using the Bayesian merging (BM) method for hydrological applications (Luo et al., 2007). By comparing those two forecasts, it will help us understand the usefulness of the CFSv2 daily precipitation or temperature forecasts for hydrological applications compared to the monthly disaggregated forecasts. The BM method both spatially and temporally

downscaled the CFSv2 monthly hindcasts from its native spatial resolution into daily hindcasts at a 0.125° spatial resolution for hydrological applications. The BM method updated an observational climatology based on the hindcast skill using Bayesian theory and generated 20 daily ensemble members for each month using historical-analog criterion and random selection. For a more detailed description of the BM method, please see Luo et al. (2007) and Luo and Wood (2008).

Ensemble forecasts of precipitation and temperature indices at sub-seasonal timescale were calculated by using daily forecasts directly from the CFSv2 ~~daily~~ and the BM downscaled CFSv2 monthly forecasts. Table 2 shows the forecast lead time for different periods and methods. For daily forecasts directly from ~~the~~ CFSv2 ~~daily~~, all precipitation and temperature indices were calculated at 7-, 14-, and 30-day forecast timescales in the first month ~~1~~. For daily forecasts from the BM downscaled CFSv2 monthly data, the precipitation and temperature indices were only calculated at 30-day forecast timescales in the first month ~~1~~, since these forecasts were temporally disaggregated from monthly forecasts. ~~and It~~ would be useful to look at the performance of the CFSv2 daily forecast in comparison with the daily ~~forecast~~ data disaggregated from the monthly forecast.

[Insert Table 2 here]

Table 3 shows the precipitation and temperature indices calculated in this study. Following Zhang (2011), a wet (dry) day was~~is~~ defined as days with precipitation above (below) 1 mm during the *n*-day period. The wet (dry) spell was~~is~~ defined as number of consecutive wet (dry) days. Taking a 14-day forecast for WetSpell as an example (as is shown in Table 2), the first forecast lead ~~one~~ was~~is~~ the number of consecutive rainy days from day 1 to day 14 forecasts. As a way of defining heat (cold) wave (e.g. Spinoni et al. 2015), the threshold for high (low) temperature day was~~is~~ defined when the temperature was~~is~~ above (below) 90th (10th) percentile of climatological distribution of temperature during the *n*-day period for different months.

[Insert Table 3 here]

To validate the forecasts, the observed precipitation and temperature indices were ~~also~~ calculated using the NLDAS-2 daily precipitation and temperature data. The NLDAS-2 daily precipitation and temperature data were ~~also~~ upscaled using bin averaging in order to ~~validate the~~match CFSv2 ~~forecasts~~spatial resolution. The percentiles of defining high (low) temperature were obtained separately from distributions of forecasts and observations. All ensemble forecasts including raw and BM downscaled CFSv2 forecasts were verified against the NLDAS-2. ~~While the CFSv2 daily forecasts were evaluated at 30-day, 14-day, and 7-day time scales, the BM downscaled CFSv2 forecasts were only evaluated at 30-day time scale since they were disaggregated from monthly forecasts.~~ Take ~~native~~ CFSv2 raw forecasts for January as an example; there are 24 ensemble members for all 30-day, 14-day, and 7-day forecasts. The 24 member ensemble forecasts were considered as being initialized on the first day of the month regardless of which day the individual member of the forecasts was initialized~~. Those 24 member ensemble forecasts were verified for the common period of~~ (e.g., 1 Jan-30 Jan, 1 Jan-14 Jan, and 1 Jan-7 Jan~~), respectively~~. All ensemble forecasts were converted into ~~probabilistic~~categorical forecasts in terciles with all observations converted into

dichotomous values of 1 or 0. The terciles were defined separately based on the individual distributions of the observations and the forecasts (x), with x<1/3rd percentile for lower tercile, 1/3rd≤x≤2/3rd percentile for middle tercile, and x>2/3rd percentile for upper tercile.

The ~~probabilistic~~categorical forecasts were evaluated using the Heidke Skill Score (HSS), a common performance metric used by the Climate Prediction Center (CPC) (e.g. Johnson et al. 2013; Wilks 2011). The HSS assesses the proportion of correctly forecasted categories. The ~~probabilistic~~categorical forecast was~~is~~ assigned to three forecast categories (upper, middle, or lower tercile) based on the highest of the three forecast probabilities. The tercile category probabilities were obtained by counting the ensemble members in each of the three categories and then by divid~~ed~~inged by the ensemble size. The HSS is expressed as:

$$HSS = \frac{(H-E)}{(T-E)} \times 100 \quad (1)$$

The number of correctly forecasted categories is denoted as H. The random forecast, E, is the expected number of categories forecast correctly just by chance. In this study, since there are three forecast categories, E is defined as one-third of the total number of forecasts, T. The HSS ranges from -50 (no correct forecasts) to 100 (perfect forecasts), with a value of 0 representing the same skill as randomly generated forecast~~, which in this case is the~~or climatological forecast. The HSS above 0 indicates that the forecasts have skill. The HSS was ~~assessed~~ calculated for each method (CFSv2 daily and BM), variable, index, grid point, month, and forecast time. Since the number of forecast-observation pairs ~~is~~was 28 for each point, the HSS estimation had~~s~~ considerable uncertainty given this relatively small sampling size. To quantify this uncertainty, a bootstrapping technique (Wilks 2011; Hamilton et al. 2012) was applied to resample 28 samples (3000 times with replacement) from the 28-year reforecasts averaged over the CONUS. Then a number of 3000 HSS was calculated for constructing a distribution, with the confidence interval and significance level ~~(X)~~ of the HSS estimated from this distribution. With this treatment of the HSS estimation uncertainty, we can determine that the HSS is significantly skillful when it is greater than a given significance level~~X~~.

Since seasonal precipitation and temperature could be more predictable at larger scales (e.g. Luo and Wood 2006; Roundy et al. 2015), it is worthwhile to also look at the sub-seasonal predictive skill~~ability of sub-seasonal forecasts averaged~~ over a larger spatial domain. Therefore, each forecast was averaged over each of the nine National Centers for Environmental Information (NCEI, formerly known as National Climatic Data Center) climate regions as well as over the entire CONUS (Figure 2). The HSS of the average forecasts over each of those regions were evaluated subsequently.

[insert Figure 2 here]

The skill assessment of Pmean and Tmean was conducted not only for all forecasts but also for forecasts during active MJO, ENSO, or combination of the two. MJO is the dominant mode of the sub-seasonal variability in the tropical atmosphere. The MJO index used in this study was from the Australian Bureau of Meteorology

Formatted: Right: 0", Line spacing: 1.5 lines

(http://cawcr.gov.au/staff/mwheeler/maproom/RMM/) for the period of 1982 to 2009. This index wasis defined by the two leading principal components (PCs) from an empirical orthogonal function analysis of the combined near-equatorially averaged 850-hPa zonal wind, 200-hPa zonal wind, and satellite-observed outgoing longwave radiation data (Wheeler and Hendon 2004). The pair of these two leading PC time series at a daily time step, called the Real-time Multivariate MJO series 1 (RMM1) and 2 (RMM2), defined eight MJO phases and an MJO amplitude. There wereare a few different ways to define active MJO events. While tThe simplest criterion was to define MJO as RMM amplitude exceededs a certain threshold (e.g. Johnson et al. 2014), whichthis criteria did not consider minimum duration and eastward propagation of MJO. This study adopted a more rigorous definition of MJO: MJO days and events wereare identified using a pentad-averaged version of the Wheeler and Hendon RMM index subject to three major requirements as indicated by L'Heureux and Higgins (2008). Similar definition was also widely adopted by other researchers such as Jones (2009) and Jones and Carvalho (2011). In this work, ENSO was defined using the same criteria as CPC (http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml). ENSO periods wereare based on a threshold of +/- 0.5 ºC for the Oceanic Niño Index (3 month running means of SST anomalies in the Niño 3.4 region). Warm or cold ENSO periods of below and above normal SSTs wereare identified when the threshold wasis met for a minimum of 5 consecutive overlapping seasons.

3. Results

3.1 The 30-day forecast skill

Figure 3 shows the average HSS for 30-day forecasts of precipitation indices calculated from the CFSv2 daily at different locations over December-January-February (DJF) and June-July-August (JJA). In DJF, the average skill of WetSpell over the CONUS wasis 34 (with a confidence interval 34±22), which wasis much higher than the skill of the other indices; it showed high skill over major most areas of the CONUS including the midwest and eastern partsUS. The Pmean, RainDay, and DrySpell were skillful in the southeast and the southwest but also revealed skill in the other regions. RainWet showed minor skill over the entire region. The skill in JJA showed different spatial patterns with DJF. While the Pmean and RainWet showed modest forecast skill in JJA over the CONUS, the RainDay, WetSpell, and DrySpell all showed high skill in the west coastal regions with the WetSpell showing some skill in the midwest and northeast. For the other seasons, on average, tThe forecast skill for precipitation indices for MAM is was between DJF and JJA for MAM, but the skill for SON was slightly lower than JJA for SON (Figure 4).

[insert Figure 3 here]
[insert Figure 4 here]

Spatial patterns in HSS wereare very different among the indices, particularly in July. We calculated the standard deviation (STD) for observed precipitation indices in July to further examine the interannual variability of those indices at each grid

point over the space. To compare relative temporal variability in space, the STD was normalized spatially to a range of 0 to 1 using a feature scaling method:

$$STD' = \frac{STD - \min(STD)}{\max(STD) - \min(STD)}, \quad (2)$$

wWhere STD is the standard deviation of time series for each grid point, min(STD) and max(STD) areis the minimum and maximum STD over all grid points respectively, ~~max(STD) is the maximum STD over all grid points,~~ and STD' is the normalized STD. Figure 5 shows the normalized standard deviation of 30-day precipitation indices in January and July over 28-year period from 1982 to 2009 over the CONUS. By comparing interannual variability (Figure 5) with the forecast skill over the space (Figure 3), we found that regions showing lower interannual variability usually have higher skill than the regions with higher interannual variability. Particularly in ~~July~~JJA, for Pmean, the western CONUS showed relatively lower interannual variability and higher skill than the eastern CONUS; for RainDay, the western coastal areas showed much lower variability and higher skill than the other regions; for RainWet, all regions showed relatively equal variability and skills; for WetSpell, the southeastern CONUS showed higher interannual variability and lower skill than the other regions of the CONUS; for DrySpell, California and eastern CONUS showed relatively lower interannual variability and higher skill than the other areas.

[insert Figure 5 here]

Figure 6 shows the average HSS for 30-day forecasts of temperature indices calculated from the CFSv2 daily at different locations over DJF and JJA. Overall, the temperature indices showed reasonably higher skill than the precipitation indices in both DJF and JJA. For DJF, Tmean showed moderate high skill in the Great Lakes area and eastern US; ~~the~~ HighDay, LowDay, CosHighD, and CosLowD were skillful over most areas of CONUS and the skill was particularly high for LowDay and CosLowD in the center or north of the midwest region. The forecast skill of temperature indices in DJF showed different spatial patterns with JJA. Tmean and LowDay showed high skill over the west inland area. CosLowD wasis skillful over major area of the CONUS, particularly in the northeast. HighDay and CosHighD showed notable high skill around south of the central area. ~~For the other seasons, on average, t~~The forecast skill for temperature indices wasis between DJF and JJA for MAM but slightly lower than JJA for SON (Figure 7). It is worth noting that given the sample size (N=28) used for calculating the HSS, the confidence interval of the HSS for each index is relatively wide. Based on the bootstrapping approach described earlier, the HSS was found to be significantly skillful (significantly above 0 at the 0.05 level) when it was greater than the number between 20 and 24, depending on the indices.

[insert Figure 6 here]
[insert Figure 7 here]

Figure 8 shows the average HSS for 30-day forecasts of precipitation and temperature indices calculated from the CFSv2 daily or the BM downscaled CFSv2 over 12 months for CONUS and its consistent NCEI climate regions. The precipitation and temperature indices calculated from CFSv2 daily showed higher skill than BM for all regions. On average, the skill from the CFSv2 daily is approximately 20% higher than the skill from the BM, suggesting that the CFSv2 month-1 daily forecasts are potentially more useful than the temporally downscaled monthly forecasts for hydrological applications.

[insert Figure 8 here]

3.2 Weekly and fortnight forecast skill at different lead times

Figure 9 (Figure 10) shows the average HSS of 14- and 7-day precipitation (temperature) indices forecasts from the CFSv2 daily over 12 months for the CONUS and its consistent NCEI climate regions. In general, the skill scores for precipitation indices were~~are~~ reasonably higher in the first two weeks than the second two weeks at both 14- and 7-day time scales~~,~~ since the first two weeks were~~are~~ within the range of weather forecast and were~~are~~ strongly influenced by the atmospheric initial conditions. While there were~~are~~ differences among regions, the skill scores for indices measuring frequency or duration of precipitation (i.e. RainDay, WetSpell, and DrySpell) or temperature extremes (i.e. HighDay, LowDay, CosHighD, and CosLowD) were equally skillful as those measuring mean precipitation or temperature during the first two weeks. Temperature indices showed notably higher skill than any precipitation index, particularly in weeks 3 and 4. It was~~is~~ worth noting that the skill was~~is~~ higher for the 14-day forecast at the first lead than for 7-day forecast in weeks 1 and 2 taken individually. The improved forecast skill indicate~~d~~s that the temporal noise in predictions can be reduced through averaging, as noted by Roundy et al. (2015).

[insert Figure 9 here]
[insert Figure 10 here]

3.3 Effects of MJO and ENSO

Figure 11 shows skill differences between Pmean or Tmean weeks 2-4 forecasts during active events (ENSO, MJO, or ~~combined active ENSO and MJO (~~MJO+ENSO) and the forecasts during the whole~~all~~ period~~s~~ for the CONUS and its consistent NCEI climate regions. The Pmean and Tmean forecasts were~~are~~ calculated from the CFSv2 daily hindcasts. In general, weeks 3 and 4 ~~Pmean~~ forecasts perform~~ed~~ better during ~~active~~ anomalous ENSO or MJO states for Pmean but not for ~~, while~~ Tmean ~~forecasts do not perform better~~.

[insert Figure 11 here]

For precipitation, the forecast skill wasis inconsistent for the active anomalous ENSO, MJO, or combined ENSO and MJO phases relative to all the whole periods. There was a notable increase in skill when the forecasts were conditioned on active MJO for almost all regions, indicating the positive influence of MJO on the CFSv2 sub-seasonal precipitation forecasts. It is worthwhile to note that forecasts conditioned on combined MJO and ENSO, and forecasts conditioned on MJO, showed similar level of positive skill with a few differences, which may due to the modulation effects of ENSO on MJO. For temperature, while the MJO, ENSO, or combined MJO and ENSO mostly showed positive effects on the CFSv2 sub-seasonal temperature forecast skill for week 2 forecast, those influences became negative in most of the regions beyond week 2. We further examined differences between Pmean or Tmean average skill over weeks 2-4 for forecasts during active MJO phase and for forecasts during all the whole period at different locations over the CONUS for DJF, MAM, JJA, and SONand seasons (Figures 12 and 13). Since HSS evaluated forecast performance over a certain period, there is only one HSS for each month, location, and lead time. For this reason, the HSS sample size is only 6 for each season and location (3 months + 3 lead times). Therefore, we used a bootstrap method to test whether those differences are significantly ($p<0.05$) for each location. To do the bootstrap, we resampled 30 times from the sample of the HSS differences for each location and season, and conduct a t-test on each sample. We used a bootstrap technique for significance tests for the difference between the HSS during active MJO phase and the HSS during the wholeall period. We resampled 28 samples (3000 times with replacement) from the 28-year reforecast averaged over the CONUS. All the 28 samples were used to calculate the HSS during the wholeall period. The subset of the 28 samples under active MJO events was used to calculate the conditional HSS during the MJO. The difference between the HSS during the MJO and the HSS during allwhole period was then calculated. Since the resampling was conducted 3000 times, a number of 3000 HSS differences was obtained for constructing a distribution and used to estimate the confidence interval and significance level (X) of the HSS. Similar to Peng et al. (2013), the significance level estimated based on the average over the CONUS was applied to test the local significance for each grid point over the region. The results wereare shown in Figures 12 and 13.

[insert Figure 12 here]

[insert Figure 13 here]

In general, most of the skill wasis significantly different at different locations; MJO hads strongly positive effects on CFSv2 sub-seasonal Pmean forecast skill over the CONUS; the effects on Tmean forecast skill wasis relatively weak and inconsistent among different regions. For precipitation, the influenced areas wereare greater during DJF and MAM than during JJA and SON, with the NE and NW regions being consistently influenced by MJO during four seasons. Aggregated over the spatial domainCONUS, we further conducted statistical tests to compare whether precipitation forecast skills during active MJO, ENSO, or combined MJO and ENSO phases wereare greater than those during the wholeall period over the CONUS for DJF, MAM, JJA, and SON. We tested whether differences in mean HSS over the CONUS (averaged over 1024 grid points) wereare statistically significant at a 5% level. The student's t-test showed the forecast skills during active MJO or combined MJO and

ENSO phases wasere significantly greater than those during the wholeall period ($p<0.05$) over the CONUS for DJF, MAM, JJA, and SON; the forecast skills during active ENSO phases wasere significantly greater than those during all the whole period over the CONUS for MAM. It is also worthwhile to note that the combined effects of MJO and ENSO wereare stronger than their individual effects of either MJO or ENSO, suggesting a potential benefit of using combined information of MJO and ENSO information for sub-seasonal forecasts. Table 4 shows that there wereare much fewer ENSO events than MJO events during January 1982 to December 2009. The number of ENSO events could be limited enough to skew the skill score conditioned on ENSO.

[insert Table 4 here]

4. Discussion

The CFSv2 sub-seasonal forecast skill wasis highly dependent on target forecasting indices, regions, seasons, leads, and methods. The sub-seasonal forecasts for indices characterizing mean precipitation and temperature as well as frequency or duration of precipitation and temperature extremes showed skill in the first two weeks but no skill or modest skill for the second two weeks, since the first two weeks were within the range of medium-range weather forecasts. This finding is important since given the sub-seasonal forecasting information is valuable to many decision makers. In particular, sub-seasonal forecasts for frequency or duration of precipitation and temperature extremes can be directly tailored to different application needs. For example, knowing having the information of RainDay, WetSpell, and DrySpell weeks in advance will help farmers make decisions for irrigation scheduling decisions, to save water costs and improve crop yields. Short-term planning of urban water supply could also benefit from this sub-seasonal forecasting information, since those indices describing frequency or duration of precipitation and temperature extremes are known to be directly related the urban water demand forecasting (e.g. Donkor et al. 2012). As some temperature indices such as CosHighD and CosLowD were used to characterize heatot/cold waves, forecasting information of these indices would also be useful for developing strategies for proactive disaster mitigations (e.g. frost damage to crops).

The spatially and temporally downscaled CFSv2 monthly data using BM method was compared with the CFSv2 daily data for sub-seasonal forecasts at native its raw resolutions. For forecasting 30-day precipitation and temperature indices, sSince the 30-day precipitation and temperature indices calculated from CFSv2 daily hindcasts have mostly higher skill than the BM, the comparison of these two methods implies that daily forecasts from the CFSv2 are potentially more useful than those disaggregated from the monthly forecasts for. Thus, the CFSv2 daily forecast information should be used in application studies such asof sub-seasonal hydrological forecastings in contrast to temporal disaggregation of the monthly forecast.

This study demonstrated that the CFSv2 sub-seasonal forecast skill varies with space and time. These results identify seasons and regions where there is the potential for skillful sub-seasonal predictions for certain precipitation and temperature indices.

For example, water managers in California trying to predict WetSpell and DrySpell have confidence to use the forecasts from CFSv2 during summer seasons, while a decision maker in the southeast may benefit little by using such information.

Sub-seasonal forecast skill can be further improved by understanding the ~~attribution~~ sources of the skill. This study took a first look at the effects of MJO and ENSO on the CFSv2 sub-seasonal forecast skill. It was found that the presence of an active MJO improves weeks-2 to -4 ~~probabilistic~~categorical ~~CFSv2-based~~ forecast ~~performance~~ of precipitation over ~~major~~ most areas of CONUS. This finding corresponds to the study of Jones et al. (2011), ~~who found~~where improved deterministic CFSv1 forecast skill of extreme precipitation was also found during active MJO. We also compared the regions of improved skill ~~associated with the~~based on MJO in this study (Figure 9) with the results ~~in~~ from Jones et al. (2011). While there were spatial differences, the regions of improved skill associated with ~~the~~ MJO commonly occurred for the western coast of the ~~United States (US)~~CONUS. This result is consistent with current knowledge of the observed influence of the MJO on precipitation events along the CONUS west coast, which can be ~~viewed~~ found at the NOAA CPC website (http://www.cpc.ncep.noaa.gov)~~,~~ under the MJO section. Forecast skill of precipitation and temperature are inherently associated with the capacity of CFSv2 in forecasting MJO. The CFSv2 has shown useful MJO prediction skill out to 3 weeks (Wang et al. 2014). Improvements of the representation of the MJO in CFSv2 will likely further extend the forecast skill of precipitation and temperature. Furthermore, ~~related~~ recent studies have developed statistical forecasting models at sub-seasonal timescale using teleconnections of MJO and ENSO phases and local weather (e.g. Johnson et al. 2013). These statistical models could be potentially combined with CFSv2 forecasts to further improve the sub-seasonal forecast skill.

It is opportune to note some future directions of this work. Forecast skill could be potentially improved by having a larger ensemble size. A sensitivity study on ensemble size could be performed to assess whether a larger ensemble improves forecast skill. For future work, when one season or 45-day CFSv2 reforecasts (initialized everyday) are available over a longer period, we would choose to use those datasets instead of 9-month reforecasts (initialized every five days) in order to incorporate a large ensemble size for making a potentially more skillful forecast. Another approach to further improve the sub-seasonal forecast skill is through multi-model ensembles. The multi-model ensemble forecasts combine multiple seasonal forecast models and often have higher skill than ~~any~~ individual models, since it has an increased ensemble size and a wider spectrum of possible forecasts that takes into account model uncertainty due to differences in model configuration and physics (e.g. Hagedorn et al. 2005). Here we highlight two important endeavors: the North American Multi-Model Ensemble (NMME-2) system (Kirtman et al. 2013) is exploring sub-seasonal forecast in their next phase; the World Meteorological Organization (WMO) sub-seasonal to seasonal (S2S) prediction project (http://www.s2sprediction.net/) is archiving hindcast and real-time forecasts from a range of model systems. All of those efforts can facilitate sub-seasonal multi-model ensemble prediction and model inter-comparison studies. Furthermore, ~~this study focused on evaluation of the capacities of CFSv2 sub-seasonal precipitation and temperature forecasts. The~~ CFSv2 sub-seasonal precipitation and temperature forecasts can be used for subsequent application studies related to areas such as hydrology and agriculture. For example, flash drought refers to a sudden

onset of high temperatures and decreases of soil moisture and is a disastrous event at sub-seasonal timescale (e.g. Mo and Lettenmaier 2015; Wang et al., 2016). Sub-seasonal forecasting of flash drought will help decision makers develop mitigation strategies. The CFSv2 sub-seasonal precipitation and temperature forecasts can be used to drive land surface hydrological models to forecast soil moisture and evapotranspiration and consequently improve flash drought forecasts.

5. Conclusion

In this study, we have assessed the CFSv2 ~~probabilistic~~categorical sub-seasonal forecasts of precipitation and temperature indices over the CONUS. The ~~probabilistic~~categorical sub-seasonal forecast skill is highly dependent on forecasting indices, regions, seasons, and methods. Indices characterizing mean precipitation and temperature as well as measuring frequency or duration of precipitation and temperature extremes for 7-, 14-, and 30-day forecasts were skillful depending on seasons and regions. The f~~F~~orecasts for 7- and 14-day temperature indices even showed skill at weeks 3 and 4, and generally more skillful than precipitation indices. The f~~F~~orecasts ~~of~~ for 30-day temperature and precipitation indices calculated from the ~~daily forecasts BM~~statistically downscaled ~~from the monthly~~ forecasts mostly showed lower skill compared to those calculated directly from the CFSv2 daily forecasts, indicating the potential usefulness of the CFSv2 daily forecasts for hydrological applications relative to the temporally disaggregated CFSv2 monthly forecasts. The presence of an active MJO improves weeks 2 to 4 ~~probabilistic~~categorical forecast ~~performance~~ of precipitation over ~~major~~ most areas of CONUS in the CFSv2 system. The sub-seasonal forecast skill of precipitation and temperature could be further improved through combining with teleconnection-based statistical sub-seasonal forecasting models or multi-model ensemble.

Acknowledgments

This research was supported by the NOAA Climate Program through the grant NA12OAR4310090 entitled "A US National Multi-Model Ensemble ISI Prediction System", which is gratefully acknowledged. The third author also acknowledges National Natural Science Foundation of China (no. 91547103). The authors would like to thank Dr. Michelle L'Heureux of the NOAA Climate Prediction Center for sharing daily MJO events data. The authors acknowledge PICSciE/OIT at Princeton University for the supercomputing support.

References

Brunet, G., and Coauthors, 2010: Collaboration of the weather and climate communities to advance subseasonal-to-seasonal prediction. *Bulletin of the American Meteorological Society*, **91,** 1397-1406.

Donkor, E. A., T. A. Mazzuchi, R. Soyer, and J. Alan Roberson, 2012: Urban water demand forecasting: review of methods and models. *Journal of Water Resources Planning and Management*, **140,** 146-159.

Garcia-Morales, M. B., and Dubus, L., 2007. Forecasting precipitation for hydroelectric power management: how to exploit GCM's seasonal ensemble forecasts. *International Journal of Climatology*, 27(12), 1691.

Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. *Tellus A*, **57,** 219-233.

Hamilton, E., Eade, R., Graham, R.J., Scaife, A.A., Smith, D.M., Maidens, A. and MacLachlan, C., 2012. Forecasting the number of extreme daily events on seasonal timescales. Journal of Geophysical Research: Atmospheres, 117(D3).

Hansen, J. W., A. Challinor, A. V. M. Ines, T. Wheeler, and V. Moron, 2006: Translating climate forecasts into agricultural terms: advances and challenges. *Clim. Res.*, **33,** 27-41.

Hudson, D., A. Marshall, and O. Alves, 2011a: Intraseasonal forecasting of the 2009 summer and winter Australian heat waves using POAMA. *Weather and Forecasting*, **26,** 257-279.

Hudson, D., O. Alves, H. H. Hendon, and A. G. Marshall, 2011b: Bridging the gap between weather and seasonal forecasting: intraseasonal forecasting for Australia. *Quarterly Journal of the Royal Meteorological Society*, **137,** 673-689.

Hudson, D., A. G. Marshall, Y. Yin, O. Alves, and H. H. Hendon, 2013: Improving intraseasonal prediction with a new ensemble generation strategy. *Monthly Weather Review*, **141,** 4429-4449.

Johnson, N. C., D. C. Collins, S. B. Feldstein, M. L. L'Heureux, and E. E. Riddle, 2013: Skillful Wintertime North American Temperature Forecasts out to 4 Weeks Based on the State of ENSO and the MJO. *Weather and Forecasting*, **29,** 23-38.

Jones, C., 2009: A Homogeneous Stochastic Model of the Madden–Julian Oscillation. *J. Climate*, 22, 3270–3288.

Jones, C., and Carvalho, L. M., 2011. Stochastic Simulations of the Madden–Julian Oscillation Activity. *Climate Dynamics*, 36(1-2), 229-246.

Jones, C., L. M. V. Carvalho, J. Gottschalck, and W. Higgins, 2011: The Madden–Julian Oscillation and the Relative Value of Deterministic Forecasts of Extreme Precipitation in the Contiguous United States. *J. Clim.*, **24,** 2421-2428.

Kirtman, B. P., and Coauthors, 2013: The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bulletin of the American Meteorological Society*, **95,** 585-601.

L'Heureux M. L. and Higgins, R. W. 2008: Boreal Winter Links between the Madden–Julian Oscillation and the Arctic Oscillation. *J. Climate*, 21, 3040–3050.

Luo, L., and E. F. Wood, 2006: Assessing the idealized predictability of precipitation and temperature in the NCEP Climate Forecast System. *Geophys. Res. Lett.*, **33** (4). DOI: 10.1029/2005GL025292

Luo, L., and E. F. Wood, 2008: Use of Bayesian Merging Techniques in a Multimodel Seasonal Hydrologic Ensemble Prediction System for the Eastern United States. *Journal of Hydrometeorology*, **9,** 866-884.

Luo, L., and Y. Zhang, 2012: Did we see the 2011 summer heat wave coming? *Geophys. Res. Lett.*, **39,** L09708.

Luo, L., E. F. Wood, and M. Pan, 2007: Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **112**.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States. *J. Clim.*, **15,** 3237-3251.

Mo, K. C., and D. P. Lettenmaier, 2015: Heat wave flash droughts in decline. *Geophys. Res. Lett.*, 2015GL064018.

Orth, R., and S. I. Seneviratne, 2013: Predictability of soil moisture and streamflow on subseasonal timescales: A case study. *Journal of Geophysical Research: Atmospheres*, **118,** 10,963-910,979.

Peng, P., Barnston, A.G. and Kumar, A., 2013. A comparison of skill between two versions of the NCEP climate forecast system (CFS) and CPC's operational short-lead seasonal outlooks. *Weather and Forecasting*, 28(2), pp.445-462.

Robertson, A. W., A. Kumar, M. Peña, and F. Vitart, 2014: Improving and Promoting Sub-seasonal to Seasonal Prediction. *Bulletin of the American Meteorological Society*.

Rodney, M., H. Lin, and J. Derome, 2013: Subseasonal prediction of wintertime North American surface air temperature during strong MJO events. *Monthly Weather Review*, **141**(8), 2897-2909.

Roundy, J. K., X. Yuan, J. Schaake, and E. F. Wood, 2015: A framework for diagnosing seasonal prediction through canonical event analysis. *Monthly Weather Review*.

Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System Version 2. *J. Clim.*, **27,** 2185-2208.

Sankarasubramanian, A., U. Lall, N. Devineni, and S. Espinueva, 2009: The role of monthly updated climate forecasts in improving intraseasonal water allocation. *Journal of Applied Meteorology and Climatology*, **48,** 1464-1482.

Spinoni, J., M. Lakatos, T. Szentimrey, Z. Bihari, S. Szalai, J. Vogt, and T. Antofie, 2015: Heat and cold waves trends in the Carpathian Region from 1961 to 2010. *International Journal of Climatology* **, 35**(14), pp.4197-4209.

Tian, D., C. J. Martinez, and W. D. Graham, 2014: Seasonal Prediction of Regional Reference Evapotranspiration Based on Climate Forecast System Version 2. *Journal of Hydrometeorology*, **15,** 1166-1188.

Toth, Z., M. Peña, and A. Vintzileos, 2007: Bridging the gap between weather and climate forecasting: research priorities for intraseasonal prediction. *Bulletin of the American Meteorological Society*, **88,** 1427-1429.

Troccoli, A., 2010: Seasonal climate forecasting. *Meteorological Applications*, 17: 251–268. doi: 10.1002/met.184

Vitart, F., 2004: Monthly Forecasting at ECMWF. Mon. *Wea. Rev.*, 132, 2761–2779.

Vitart, F., A. W. Robertson, and D. L. Anderson, 2012: Subseasonal to Seasonal Prediction Project: bridging the gap between weather and climate. *Bulletin of the World Meteorological Organization*, **61,** 23.

Vitart, F., and Coauthors, 2008: The new VAREPS-monthly forecasting system: A first step towards seamless prediction. *Quarterly Journal of the Royal Meteorological Society*, **134,** 1789-1799.

Wang, L., Yuan, X., Xie, Z., Wu, P., and Li, Y., 2016: Increasing flash droughts over China during the recent global warming hiatus. Sci. Rep., 6, 30571, doi: 10.1038/srep30571.

Wang, W., M.-P. Hung, S. Weaver, A. Kumar, and X. Fu, 2014: MJO prediction in the NCEP Climate Forecast System version 2. *Clim Dyn*, **42,** 2509-2520.

Wheeler, M. C., and H. H. Hendon, 2004: An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Monthly Weather Review*, **132,** 1917-1932.

White, C. J., D. Hudson, and O. Alves, 2014: ENSO, the IOD and the intraseasonal prediction of heat extremes across Australia using POAMA-2. *Clim Dyn*, **43,** 1791-1810.

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

Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **107,** ACL 6-1-ACL 6-15.

Xia, Y., and Coauthors, 2012: Continental‐scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS‐2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **117**(D3).

Yao, W., H. Lin, and J. Derome, 2011: Submonthly forecasting of winter surface air temperature in North America based on organized tropical convection. Atmosphere-Ocean, **49**(1), 51-60.

Yuan, X., E. F. Wood, and M. Liang, 2014: Integrating weather and climate prediction: Toward seamless hydrologic forecasting. *Geophys. Res. Lett.*, **41,** 2014GL061076.

Yuan, X., E. F. Wood, L. Luo, and M. Pan, 2011: A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction. *Geophys. Res. Lett.*, **38,** L13402.

Yuan, X., E. F. Wood, J. K. Roundy, and M. Pan, 2013: CFSv2-Based Seasonal Hydroclimatic Forecasts over the Conterminous United States. *J. Clim.*, **26,** 4828-4847.

Zinyengere, N., T. Mhizha, E. Mashonjowa, B. Chipindu, S. Geerts, and D. Raes., 2011: Using seasonal climate forecasts to improve maize production decision support in Zimbabwe. *Agricultural and Forest Meteorology* 151 (12): 1792-1799.

Zhang, X., and Coauthors, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, **2,** 851-870.

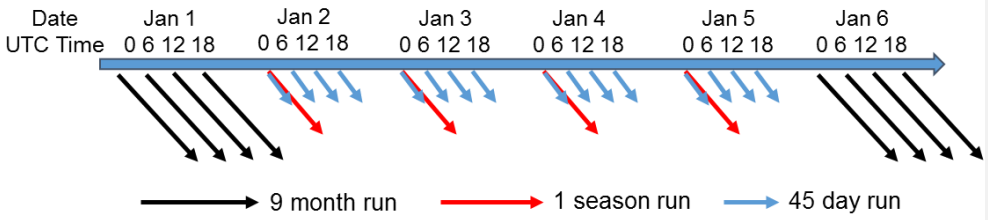


Figure 1. Three configurations of the CFSv2 hindcast: 9-month run, 1 season run, and 45-day run. UTC stands for Coordinated Universal Time.

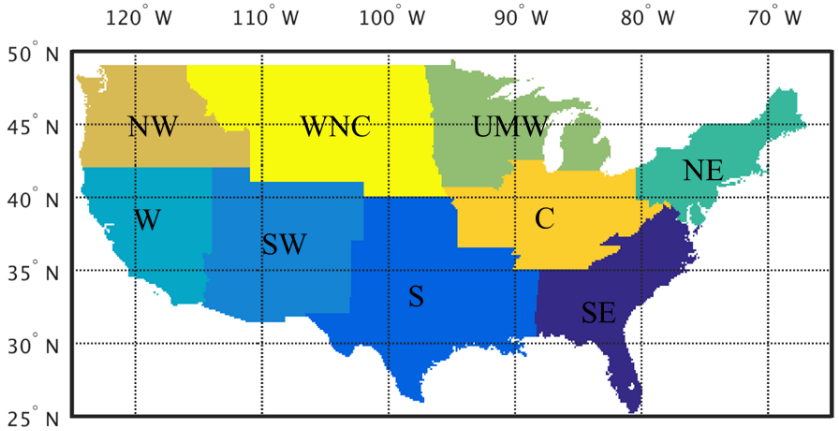


Figure 2. NCEI climate regions (described in Section 2) used as areal averaging domains for raw and BM downscaled CFSv2 forecasts. Regions are named as follows: Northwest (NW), West (W), Southwest (SW), West North Central (WNC), South (S), Upper Midwest (UMW), Central (C), Southeast (SE), and Northeast (NE).

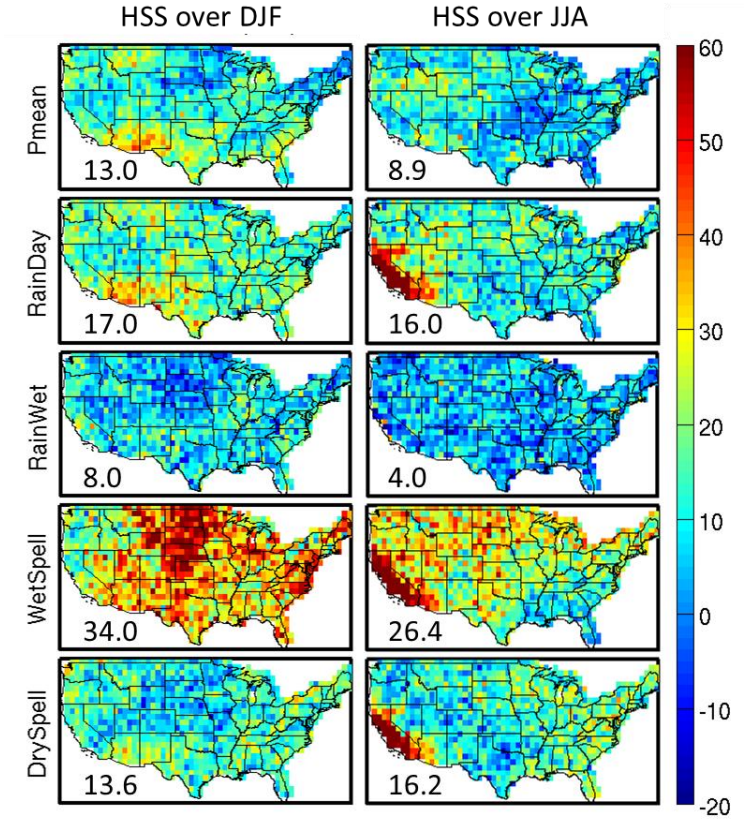


Figure 3. HSS of 30-day (from top to bottom columns) Pmean, WetRain, RainDay, WetSpell, and DrySpell from ~~(from left to right rows)~~ the CFSv2 daily hindcasts ~~and BM~~ over DJF (left) and JJA (right). The number in the bottom left is the overall average.

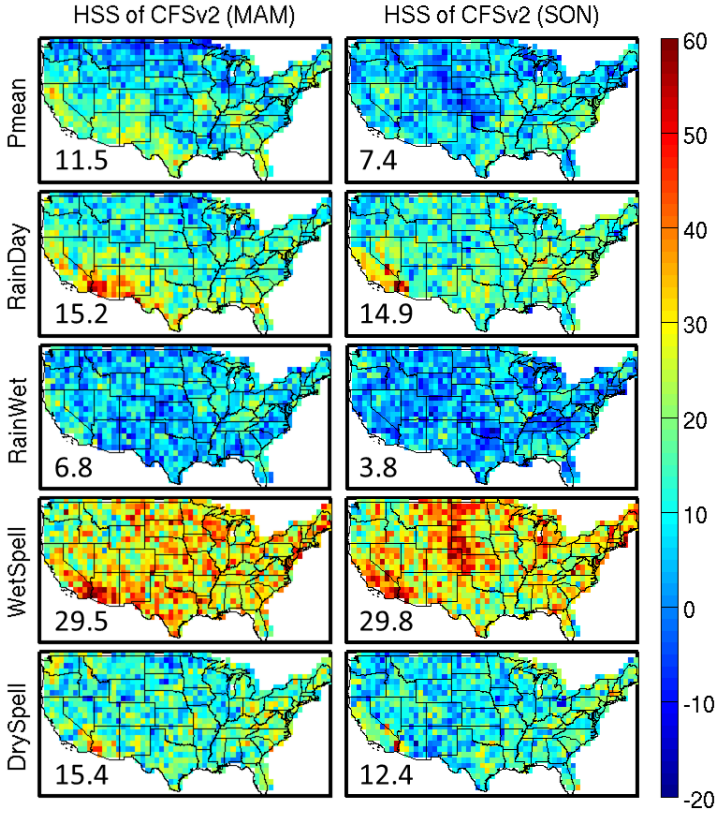


Figure 4. Same as in Figure 3, but for MAM and SON.

5

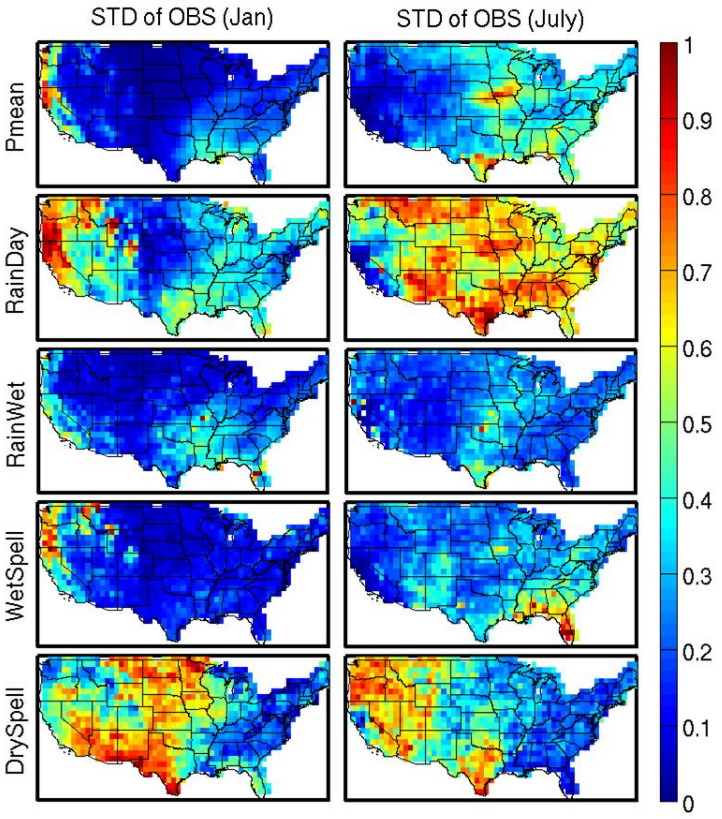


Figure 5. Spatially normalized standard deviations of observed 30-day precipitation indices in January and July over 28-year period from 1982 to 2009

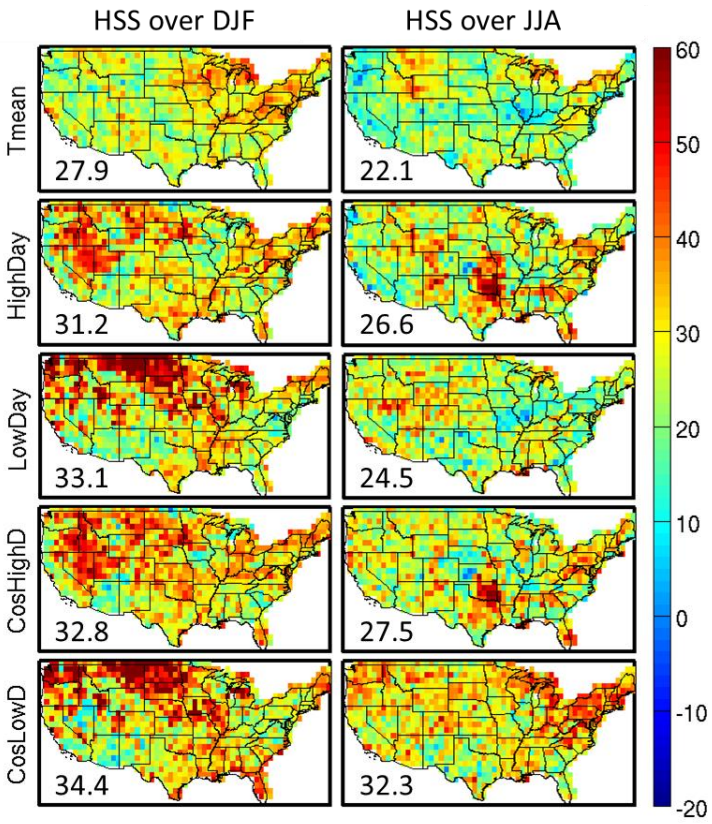


Figure 6. HSS of 30-day (from top to bottom columns) Tmean, HighDay, LowDay, CosHighD, and CosLowD from (from left to right rows) the CFSv2 ~~raw~~ daily hindcasts ~~and BM monthly diaggregated hincasts~~ during~~over~~ DJF (left) and JJA (right). The number in the bottom left is the ~~overall~~ average.

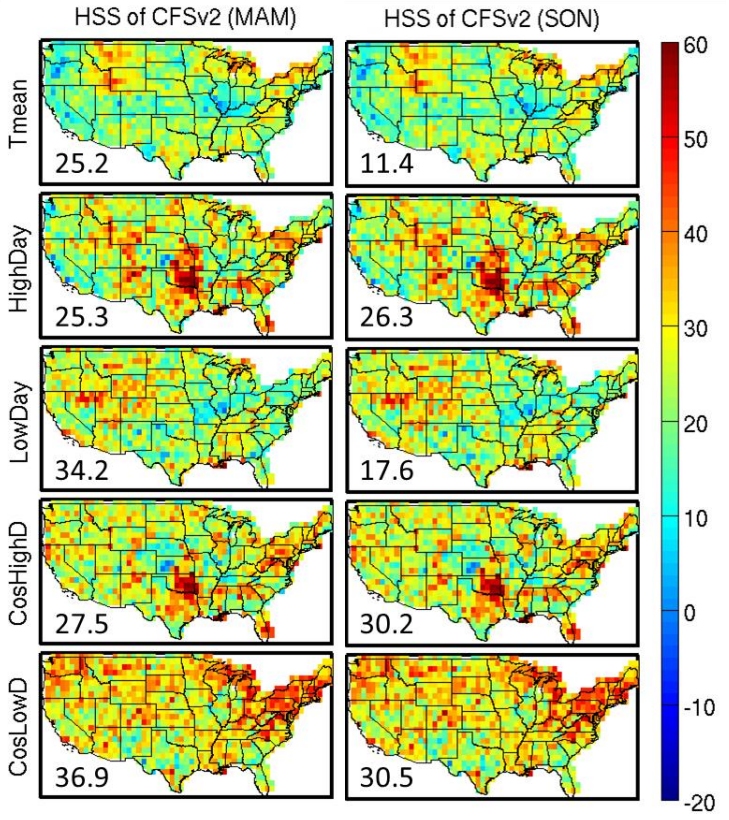


Figure 7. Same as in Figure 6, but for MAM and SON.

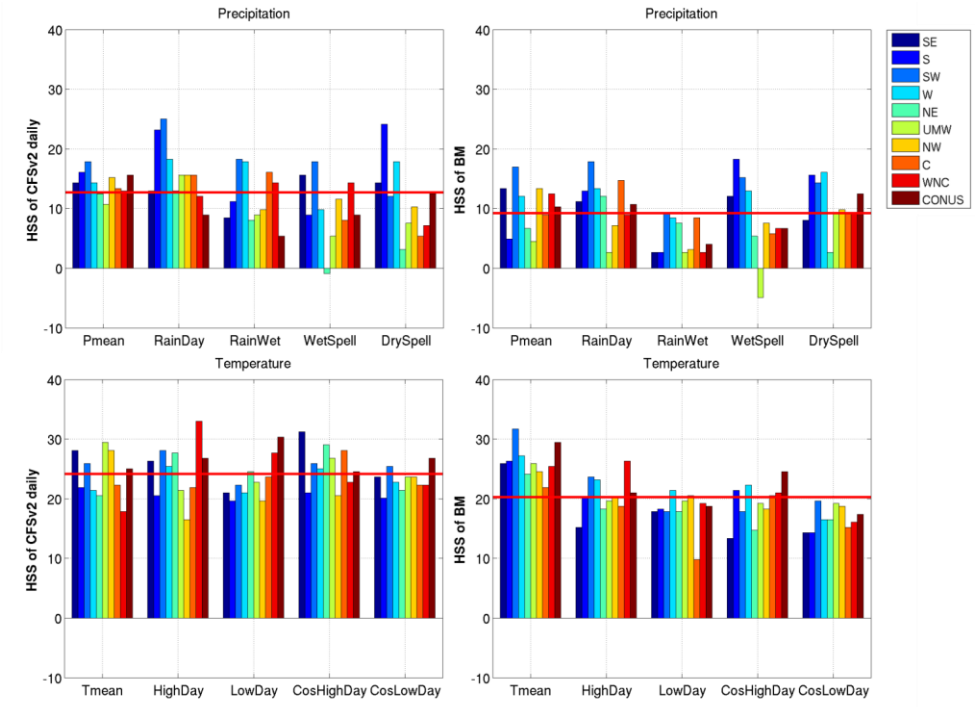


Figure 8. HSS of 30-day precipitation and temperature indices calculated from the CFSv2 and BM for CONUS and its consistent NCEI climate regions. The red line is the ~~overall~~ average.

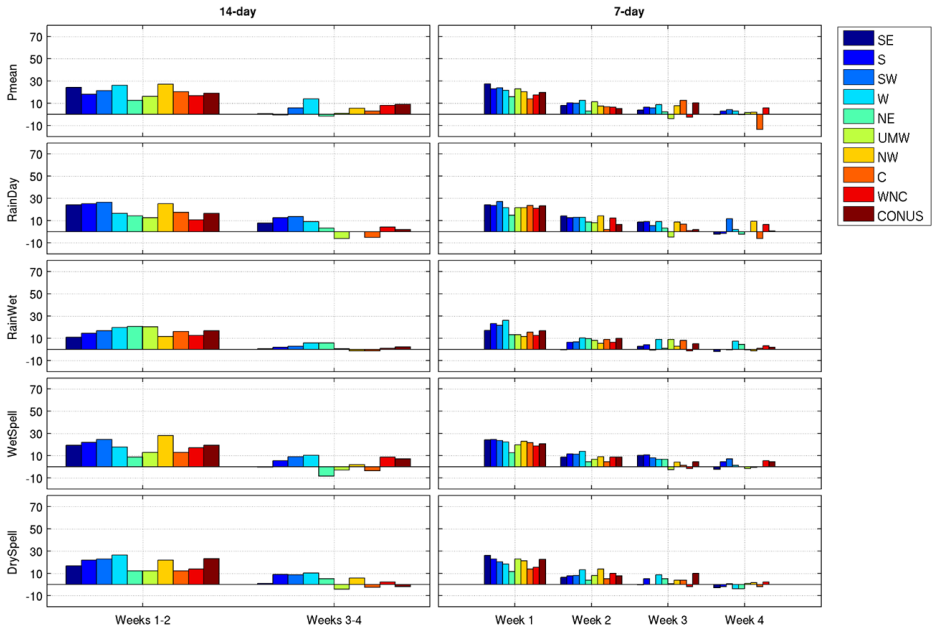


Figure 9. ~~Overall m~~Mean HSS of 14- and 7-day (from top to bottom rows) Pmean, WetRain, RainDay, WetSpell, and DrySpell from the CFSv2 daily hindcasts for CONUS and its consistent NCEI climate regions.

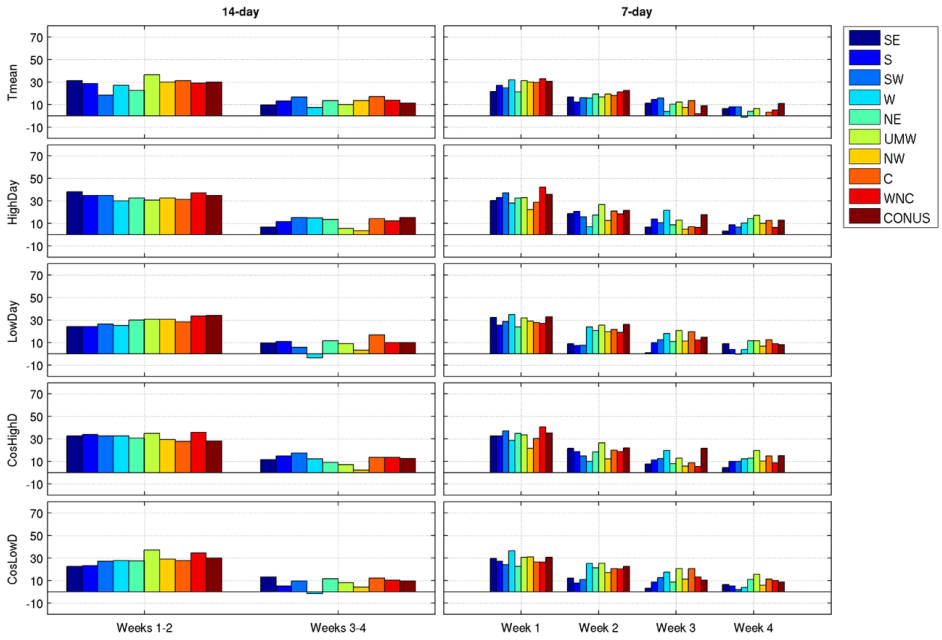


Figure 10. MOverall mean of HSS of 14- and 7-day (from top to bottom rows) Tmean, HighDay, LowDay, CosHighD, CosLowD from the CFSv2 daily hindcasts for CONUS and its consistent NCEI climate regions.

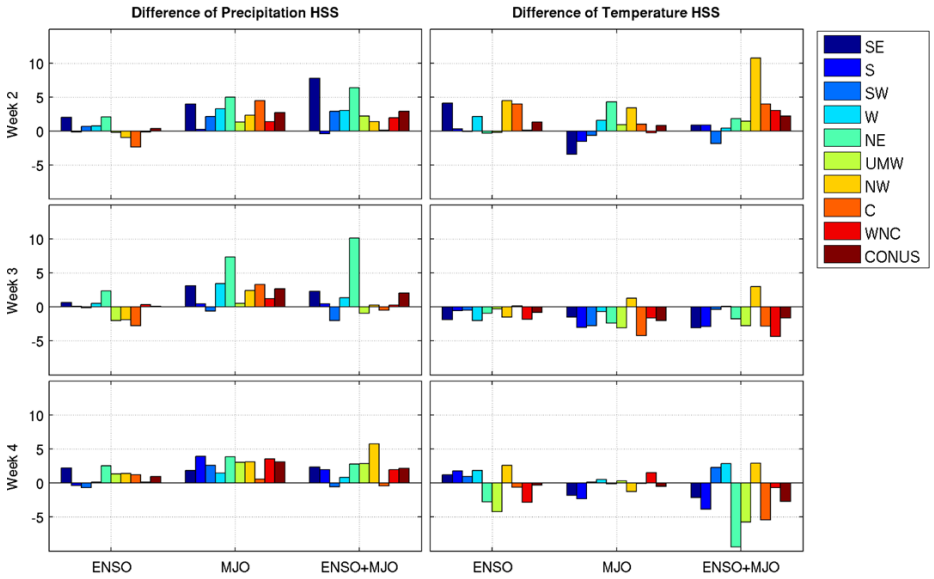
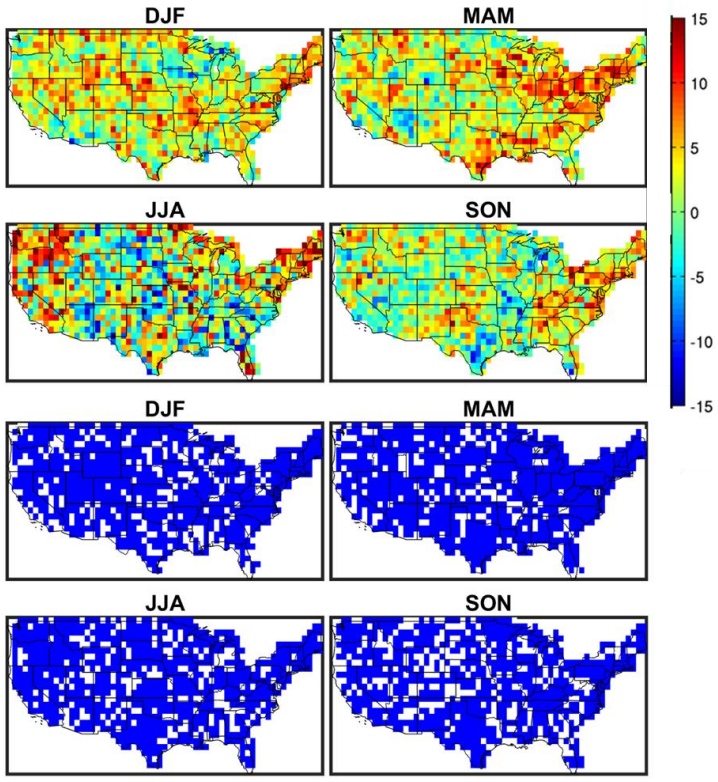
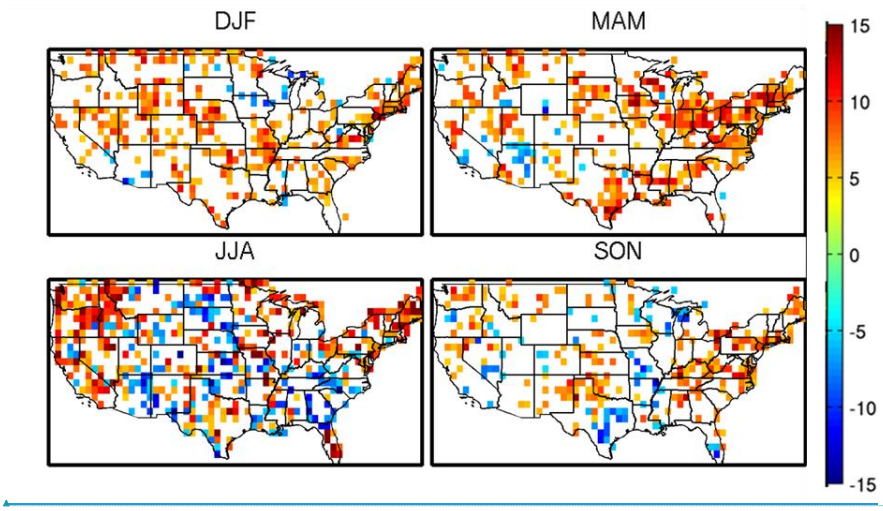


Figure 11. HSS differences between Pmean (left column) or Tmean (right column) weeks 2-4 forecasts during active ENSO, MJO, or combined active ENSO and MJO (MJO+ENSO) phases and those ~~forecasts~~ during the whole~~all~~ period~~s~~ for CONUS and its consistent NCEI climate regions. Positive values indicate more skillful forecasts ~~for the~~during active MJO, ENSO, or ENSO+MJO phases.

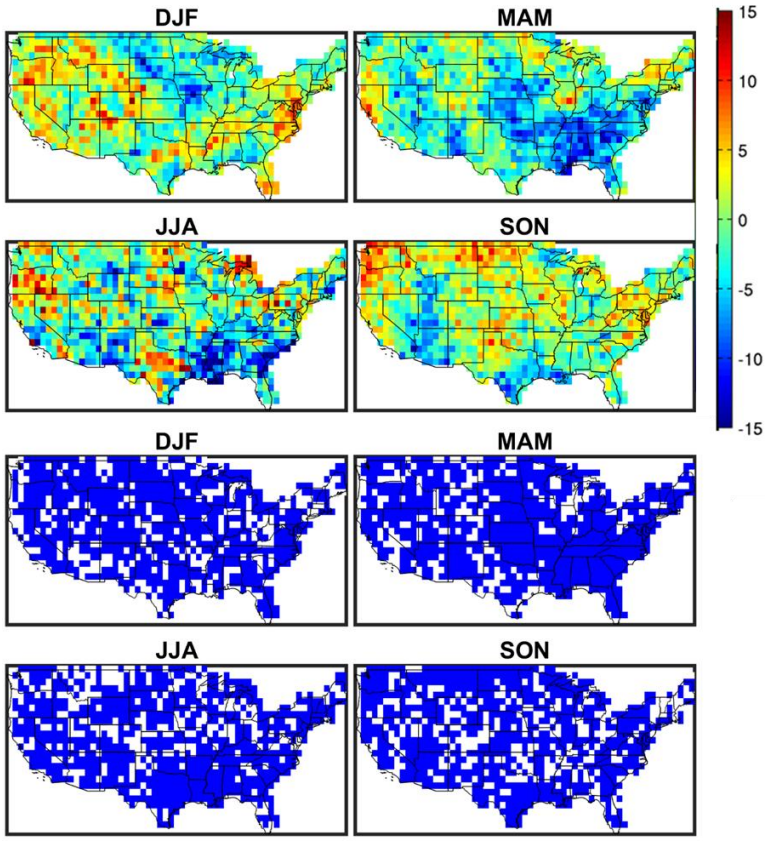


Formatted: Font: 9 pt, Bold



Formatted: Font: 9 pt, Bold

Figure 12. ~~The upper panel shows d~~**D**ifferences between HSS of weeks 2-4 **P**mean ~~average HSS over weeks 2-4 for~~ **forecasts during active MJO phases and** ~~for forecasts during all~~**the whole period at different locations over the CONUS for DJF, MAM, JJA, and SON. The** ~~blue pixel in the lower panel shows where the difference is significant~~areas in white indicate the differences are not significant (at the 0.05 level) ~~(p<0.05)~~**.**



Formatted: Font: 9 pt, Bold

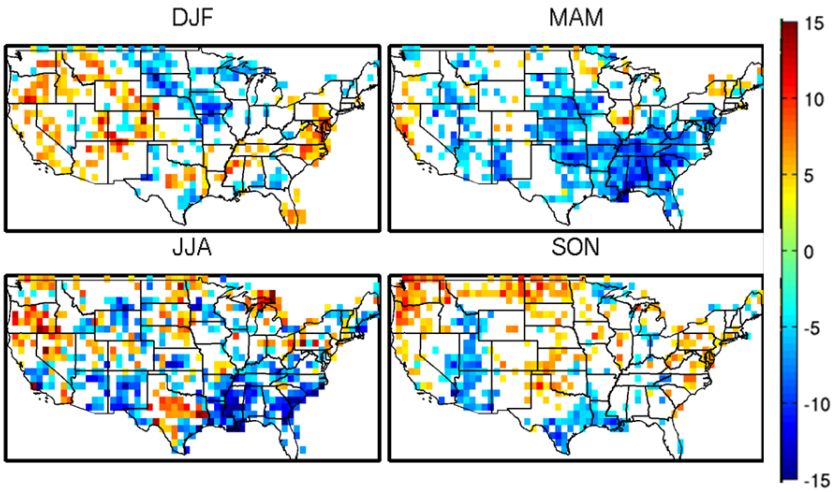


Figure 13. Same as in Figure 12, but for Tmean.

Formatted: Font: 9 pt, Bold

Table 1. Configurations of the CFSv2 hindcast. UTC stands for Coordinated Universal Time.

Configurations	9-month runs	one season	45-day runs
Initiated day	Every 5 days beginning from Jan 1 of each year	Every day	Every day
Initiated UTC time	0, 6, 12, 18	0	0, 6, 12, 18
Covered Period	1982-2010	1999-2010	1999-2010

Table 2. Forecast lead times for different periods and methods.

Period	CFSv2 daily				BM
	Lead 1	Lead 2	Lead 3	Lead 4	Lead 1
30-day	Day 1 to Day 30	-	-	-	Day 1 to Day 30
14-day	Day 1 to Day 14	Day 15 to Day 28	-	-	-
7-day	Day 1 to Day 7	Day 8 to Day 14	Day 15 to Day 21	Day 22 to Day 28	-

Table 3. Precipitation and temperature indices ~~calculated~~ used in this study.

Index	Description	Period
Pmean	mean precipitation	30-day, 14-day, and 7-day
RainWet	mean precipitation over wet days	30-day, 14-day, and 7-day
RainDay	number of rainy days	30-day, 14-day, and 7-day
WetSpell	maximum wet spell length	30-day, 14-day, and 7-day
DrySpell	maximum dry spell length	30-day, 14-day, and 7-day
Tmean	mean temperature	30-day, 14-day, and 7-day
HighDay	number of high temperature days	30-day, 14-day, and 7-day
LowDay	number of low temperature days	30-day, 14-day, and 7-day
CosHighD	maximum number of consecutive high temperature days	30-day, 14-day, and 7-day
CosLowD	maximum number of consecutive low temperature days	30-day, 14-day, and 7-day

Table 4. Active ENSO, MJO, and ENSO+MJO during January 1982 to December 2009. The red areas indicate active ENSO periods. The green areas indicate the periods with active MJO happening. The yellow areas indicate combined active ENSO and MJO events. The last three lines show the total number of ENSO, MJO, and ENSO+MJO events for each month.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1982												
1983												
1984												
1985												
1986												
1987												
1988												
1989												
1990												
1991												
1992												
1993												
1994												
1995												
1996												
1997												
1998												
1999												
2000												
2001												
2002												
2003												
2004												
2005												
2006												
2007												
2008												
2009												
No. ENSO	16	15	12	11	11	12	12	13	15	17	17	17
No. MJO	22	22	24	23	24	15	15	15	16	24	23	25
No. ENSO+MJO	12	12	10	7	8	4	6	6	9	13	13	14