Response to the Reviewers' Comments for the Manuscript Entitled "CFSv2-based sub-seasonal precipitation and temperature forecast skill over the contiguous United States"

By Di Tian, Eric F. Wood, and Xing Yuan

November 4, 2016

Response to Editor:

I find the work relatively straightforward and the results reasonable -- ie, the skill results and associations with teleconnections are in line with expectations. All reviewers touch on the need for greater clarity in various areas (including further work on diction and grammer), and the authors responses are appropriate. I recommend that the author proceed with reviews as outlined and that the revised paper undergo a further round of reviews.

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 and three reviewers. We have submitted the revised manuscript as requested.

In addition, I find it odd that the authors do not address more comprehensively the apparent noise in skill results shown in the spatial (pixel) plots -- e.g., in Figure 9, where adjacent pixels show signals of opposite sign. Although the authors note that the differences shown are 'significant', perhaps indicating a domain-aggregate sense, I suggest the authors go further to indicate where in each season (eg, again in Figure 9) they are significant. Surely where the cell-specific differences are near zero, they cannot also be significant. It is possible that I am misinterpreting the authors' discussion, but that alone suggests the need for greater clarity and thorough effort to convey where skills that are statistically different than zero are achieved. I would hope to see this point addressed in the revision.

RESPONSE: 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. The results are shown in Figures 9 and 10 (now Figures 12 and 13).

Response to Reviewer #1:

General comment: I think this paper has merit for publication but needs a little more work before acceptance. Most notable is the grammar and syntax which makes the paper difficult to read. Moreover, there are lots of different combinations of forecast times and forecast skill evaluations for different models. I found that following the different periods was confusing. Perhaps add a table with the appropriate information or structure the workflow differently. I think that the results and discussion are in line with what the paper aims to show. I did not find any methodological fault, although I must admit that the subject matter is not my main expertise.

RESPONSE: Thank you very much for your time and your insightful review. We have improved the grammar and syntax throughout the manuscript. We added tables with appropriate information to clarify the description of data and method. We have carefully revised the manuscript in order to include your comments. We believe that this manuscript is substantially improved as a result of the revision.

Other comments and suggestions:

Abstract: I think the first sentence is either too vague or too direct. Perhaps start with something like: "This paper explores the possibility of exploiting forecasts from global seasonal climate forecast models for sub-seasonal forecasts of precipitation and 2-m temperature". The current wording seems like a statement: "... forecast models can be...", but it is a vague statement because of the word "potentially".

RESPONSE: This sentence has been revised as suggested.

Page2, line 12: References are not ordered properly; please revise (all text).

RESPONSE: The orders of all the references have been adjusted throughout the text.

Page 2, line 12-15: This sentence is not clear, please revise.

RESPONSE: This sentence has been clarified as requested:

"Precipitation and 2-m temperature (hereafter temperature) are considered to be two of the most important climate variables that significantly influence irrigation scheduling, urban water supply, cooling water related to thermal power generation, and hydropower operations, etc."

Page 2, line 20: "... there have has been"

RESPONSE: It has been revised as requested.

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."

Page 2, line28: The link between GCMs and the CFSv2 is not clear. Is the CFSv2 a GCM? Please indicate that it is a reforecast product based on reanalysis (If i understood correctly).

RESPONSE: The CFSv2 is a GCM in that it is a fully coupled ocean-land-atmospheric model, developed by NOAA for dynamical seasonal forecasting, and has archived a retrospective forecast product. It has been clarified in the revised manuscript.

Page 3, line 2: "... demonstrated the high performance...": delete "the".

RESPONSE: It has been deleted as suggested.

Page 3 lines 18-19: This sentence is not clear and does not add much to the paper. I suggest modifying it by giving it more substance. "Leverage forecasting efforts" and "contribute to sectorial management decision making" are both very vague objectives.

RESPONSE: We agree that this sentence is redundant and does not add much to the paper. We have deleted it in the revised manuscript.

Page 3, line 22: "... forecast model components of the climate system... ".not clear what this means. Of all existing models?

RESPONSE: This sentence has been changed to: "The CFSv2 has the most updated data assimilation and modeling system. It became operational at NCEP since March 2011."

Page 3, lines 29-33: Check grammar here (everywhere, but particularly here). It is difficult to read.

RESPONSE: These sentences have been changed to: "The one season and 45-day reforecasts are initialized each day so that relatively new initial conditions can be incorporated into a large ensemble for making a potentially 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 the one season and 45-day reforecasts (1999-2010), which ensures a larger sample size for a more robust evaluation."

We have also proofed all the grammars throughout the text.

Page 4, line 14: Replace the sentence with something like: "Comparing those two forecasts will help understand..."

RESPONSE: We have changed the sentence to: "Comparing those two forecasts will help understand the usefulness of the CFSv2 daily precipitation or temperature forecasts for hydrological applications compared to the monthly disaggregated forecasts."

Page 4, line 30-31: This should refer to a figure or a table somehow. We cannot follow the given example because of the lack of a reference.

RESPONSE: We have added a table to explain this example.

Page 5, line 15: Merge to make a more fluid sentence? (e.g. All forecasts... and all observations...)

RESPONSE: This sentence has been changed to "All ensemble forecasts were converted into probabilistic forecasts in terciles with all observations converted into dichotomous values of 1 or 0. "

Page 6, lines 14-19: This section is suspiciously similar to the text in L'heureux and Higgins. Please reword or cite directly.

RESPONSE: We reduced this section and made a more direct reference to L'Heureux and Higgins (2008).

Figure 6-7: It is not clear to me why the score is higher for the 14-day (week 1-2) than for weeks 1 and 2 taken individually.

RESPONSE: We have clarified this in the revised manuscript and added these sentences: "It is worth noting that the skill 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 indicates that the temporal noise in predictions can be reduced through averaging, as noted by Roundy et al. (2015)."

Page 10, line 25: Reference to nonexistent figure 13.

RESPONSE: It should be Figure 9 (Figure 12 in the revised manuscript) instead of 13. We have revised that in the manuscript.

Response to Reviewer #2:

General Comment: This paper is interesting and provides information about the CFSv2 model that will be useful to the scientific community. The paper presents intriguing findings, but some are hard to decipher due to grammatical and syntax errors. I found the paper to be hard to follow in places due to the unclear description of different models and time steps. The "Data and Methodology" section which describes the models/time steps analyzed was particularly difficult to follow; some of the paragraphs could be split and the writing could be more concise. Overall, the scientific approach and methods seem adequate, though I am new to this subject. The paper also presents clear conclusions that are a substantial contribution to scientific progress.

Thank you very much for your time and your insightful review. We have carefully revised the manuscript in order to include your comments. We believe that this manuscript is substantially improved as a result of the revision.

Below are other specific comments and suggestions:

Pg 1, line 13: Missing a word - "and generally are more skillful"

RESPONSE: The missing word has been added as requested.

Pg 1, line 16: Remove "the" before "number of consecutive dry days..."

RESPONSE: It has been revised as requested.

Pg 1, line 26: Add a comma after the reference.

RESPONSE: It has been revised as requested.

Pg 2, lines 13-19: To make the aims of this study more clear, it would be good to list or number them. There are 3 in the results section. It would be nice to see the same differentiation here.

RESPONSE: Thanks for pointing this out. It has been revised as requested: "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 forecast skill of CFSv2 at different lead times, and 3) evaluate the effects of MJO and ENSO on CFSv2 sub-seasonal forecast skill."

Pg 3, lines 25-27: You refer to 0Z cycle (and different numbers of cycles). I am not familiar with this terminology, and am therefore confused when reading these sentences.

RESPONSE: Thanks for pointing this out. 0Z cycle means the 0 UTC (Coordinated Universal Time) assimilation and forecast cycle. We have deleted these sentences and used a table and a figure instead to explain and compare those three CFSv2 hindcast configurations.

Pg 3-4, lines 22-33, 1-11: This paragraph is confusing due to the large amount of information presented. The paragraph could be split into multiple paragraphs and reworded to clarify which time steps and models are being referred to.

RESPONSE: We have split this paragraph into multiple paragraphs for clarification. Some of the sentences have been replaced by a table and a figure to explain and compare the three CFSv2 hindcast configurations. The paragraph has been reworded and updated accordingly.

Pg 4, line 26: This could be the start of a new paragraph to separate the two different ideas.

RESPONSE: We agree. It has been separated into two different paragraphs.

Pg 5, line11: Replace comma with a semicolon or separate into two different sentences.

RESPONSE: It has been revised as requested.

Pg 8, line 23: Bold the title.

RESPONSE: It has been revised as suggested.

Pg 9, lines 26-28: This sentence is confusing and could be reworded to improve clarity.

RESPONSE: This sentence has been modified as:

"This finding is important since 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."

Pg 9, line 31: It is unclear which precipitation and temperature indices are being referred to with the word "these."

RESPONSE: This sentence has been modified as "... those indices describing frequency or duration of precipitation and temperature extremes ..."

Pg 10, lines 6-12: These sentences could be rewritten to be more concise.

RESPONSE: This paragraph has been revised to be more concise and clearer.

Pg 10, line 22: Figure 13 isn't used in the paper.

RESPONSE: Thanks for pointing this out. It should be Figure 9 (Figure 12 in the revised manuscript) instead of 13. We have revised that in the manuscript.

Pg 11, line 9: "Subseasonal" is spelled wrong and is also written differently than the other instances where the spelling was "sub-seasonal." This may be an issue throughout the paper and should be checked.

RESPONSE: All the instances of "subseasonal" have been changed to "sub-seasonal" throughout the paper.

Response to Reviewer #3:

General Comments and Suggestions:

The work presented here is interesting, and presents an opportunity for researchers to utilize sub-seasonal forecasts from the CFSv2 model. However, this manuscript would benefit from a thorough proofreading, as the authors' writing is difficult to follow throughout much of the paper. Overall, I believe the manuscript can be made more concise and direct, which I believe will improve the readability of the paper. I have noted some specific instances below, but I am sure it is not a comprehensive list of all the improvements that could be made to the manuscript.

Thank you very much for your time and your insightful review. We have carefully revised the manuscript in order to include your comments. We believe that this manuscript is substantially improved as a result of the revision.

On page 2, the authors state that, ": : :many extreme events and management decisions fall into sub-seasonal timescales: : :" I suppose I could use some examples. As the paper is currently, "sub-seasonal timescales" are on the order of 3-4 weeks. When I think of extreme events, I think of shorter term events such as flash flooding, tornadoes, extreme hail, wind, etc: : :, or I think of more persistent events such as drought or persistent floods. Many of the examples the authors provide in the second paragraph tend to be on the order of the seasonal timescale as defined by the authors on page 1. It would be beneficial if the authors could state what extreme events they specifically have in mind to be addressed by this research. Some of this is touched on near the end of the Discussion section by the authors, but could be more succinctly stated earlier.

RESPONSE: We added following sentences to this paragraph:

"… For example, flash drought, heat wave, and cold wave are extreme events at sub-seasonal timescale. Sub-seasonal forecast information can be used for developing strategies for proactive natural disaster mitigation, which may be needed during those extreme events. … In this study, we aim to evaluate forecasting for precipitation and temperature derivatives, or indices that are associated with those extreme events, and decision-making at sub-seasonal timescale."

At the end of the second paragraph on page 2, the authors state that the, ": : :derivitives or indices are directly associated with important events and decision making: : :". Similar to the previous comment, I do not believe the manuscript currently addresses this point.

RESPONSE: In this study, we aim to evaluate forecasting for precipitation and temperature derivatives or indices that are associated with important events and decision-making at sub-seasonal timescale – in particular 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. We have noted this in the revised manuscript.

Beginning on page 3 and continuing to page 4, the first paragraph of the Data and Methodology section is very confusing. A table comparing the differences between the 9-month, 45-day, and season runs would be beneficial. As it stands now, I believe this section assumes too much of the reader to interpret how NCEP runs the CFSv2 model.

RESPONSE: We added a table and a figure to explain those three hindcast configurations. We have also updated the text in this section to make it more concise and direct.

On page 4, the authors discuss an example and reference a "lead one" forecast. It appears they are referencing the 14-day forecast, but it's not clear.

RESPONSE: This sentence has been clarified: "Taking the 14-day forecast for WetSpell in January as an example, the first (second) forecast lead time is the number of consecutive rainy days from January 1 to 14 (from January 15 to 28)."

Equation 1 is not clear. The variable E is described as 1/3 the total number of forecasts, T. Why wouldn't the denominator simply be (T – T/3)? It sounds as if there is more to the variable E than is described.

RESPONSE: The variable E is the expected number of categories forecast correctly just by chance. Since there are three forecast categories in this study, it is described as one third of the total number of forecasts. This has been clarified in the revised manuscript.

With regards to the discussion of Figure 2 that starts near the end of page 6 and continues on to page 7, I am not sure why the authors do not include similar figures for MAM or SON, but do discuss the results. Wouldn't it be beneficial to include those figures? This comment applies to Figure 4 as well.

RESPONSE: We agree that it would be beneficial to include those figures. The figures for MAM and SON have been added in the revised manuscript as suggested.

Similarly, I'm not sure why the authors stress emphasizing the month of July in Figure 3. The authors state there is some difference in monthly spatial patterns, but without the other months, I do not have the proper context for the figure.

RESPONSE: In the new figure, we have added the month of January as a context for the figure.

With regards to section 3.3, it would be useful to have some sort of table or reference to see what periods of MJO and/or ENSO activity are being analyzed. The number of events considered could be limited enough that the HSS could be somewhat skewed. It would also be helpful if the authors discussed more clearly the impacts of active MJO and ENSO compared to the combined impacts of MJO and ENSO events. I think the authors begin to discuss this in the second paragraph on page 9, but do not offer enough insight on the particular point.

RESPONSE: We have included a table showing the periods of MJO and ENSO events and point out that the limited number of events could somewhat skew the skill score conditioned on ENSO. We also discuss the combined effects of MJO and ENSO compared to the individual effects from either MJO or ENSO, suggesting a potential benefit of using MJO and ENSO information for sub-seasonal forecasts. Please see the details in section 3.3 of the revised manuscript.

The first full paragraph on page 10, describing the role of the BM method is not clear. I would recommend the authors explain this conclusion more clearly, or simply remove the paragraph.

RESPONSE: Thanks for pointing this out. We have simplified this paragraph by removing the sentences for describing the role of the BM method.

On page 11, the authors state that forecast skill could be improved by simply having a larger ensemble. I am not convinced of that; a larger ensemble may not necessarily add useful information. It may be more appropriate to state that a sensitivity study on ensemble size could be performed to see if a larger ensemble does improve forecast skill.

RESPONSE: We agree that a larger ensemble may not necessarily add useful information. We changed the original sentence to:

"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."

Page 1, lines 27-28: I'm not sure what is meant by, ": : :sub-seasonal timescale is beyond the memory of the atmospheric initial conditions: : :"

RESPONSE: It means sub-seasonal timescale is sufficiently long that much of the memory of the atmospheric initial conditions is lost. We have clarified that in the revised manuscript.

Page 5, the NCDC has since been renamed the National Centers for Environmental Information (NCEI)

RESPONSE: It has been revised as suggested.

Page 6, line 26 should read "Figure 2 shows"

RESPONSE: It has been revised as suggested.

There is no legend for Figure 5, so I am unsure which regions match to each particular color.

RESPONSE: We have added a legend as suggested.

Page 8, line 10: "temperally" should be "temporally"

RESPONSE: It has been revised as suggested.

Page 8, line 17: "reasonaly" should be "reasonably"

RESPONSE: It has been revised as suggested.

Page 9, line 23: "depending" should be "dependent"

RESPONSE: It has been revised as suggested

Page 10, line 4, "spatial" and "temporal" should be "spatially" and "temporally"

RESPONSE: It has been revised as suggested.

Page 10, line 5: "to note" should be "noting"

RESPONSE: To address one of the comments raised above, this sentence has been deleted.

Page 10, line 23: There is no Figure 13 included in the manuscript. To this point, referencing a figure in another article (Jones et al. 2011) is a bit confusing. I think I would just note how the results of this study compare to Jones et al., rather than to a specific figure.

RESPONSE: Thanks for pointing this out. Figure 13 is a typo. It has been changed to Figure 12 in the revised manuscript. We agree it is a bit confusing to reference a figure in another article. We have changed that in the revised manuscript as suggested

Page 11, line 8, "highlighted" should be "highlight"

RESPONSE: It has been revised as suggested.

Page 11, line 9, "subeasonal" should be "sub-seasonal"

RESPONSE: It has been revised as suggested throughout the paper.

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

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

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey 08544, USA
[2]RCE-TEA, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

Correspondence to: Di Tian (dtian@princeton.edu)

Abstract. This paper explored~~s~~ the potential of global seasonal climate forecast models for sub-seasonal forecasting of precipitation and 2-m air temperature~~Forecasts from global seasonal climate forecast models can be potentially exploited for sub-seasonal forecasts of precipitation and 2-m temperature~~. The probabilistic sub-seasonal forecast skill of ten precipitation and temperature indices was~~is~~ investigated using the 28-years' hindcasts of the Climate Forecast System version 2 (CFSv2) over the contiguous United States (CONUS). The forecast skill was~~is~~ highly dependent on the target ~~forecast~~ indices, regions, seasons, 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 ~~are~~were more skillful than precipitation indices. Overall, temperature indices showed higher skill than precipitation indices over the entire CONUS region. While the forecast skill related to mean precipitation indices were low in summer over the CONUS, the number of rainy days, number of consecutive rainy days, and ~~the~~ number of consecutive dry days showed considerable high skill for the west coast region. The 30-day forecasts of precipitation and temperature indices calculated from the downscaled monthly CFSv2 forecasts ~~are~~ 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 the presence of active Madden-Julian Oscillation (MJO) events improved~~s~~ CFSv2 weekly mean precipitation forecast skill over major areas of CONUS, MJO and/or El Niño Southern Oscillation did not necessarily improve the ~~have same strong effects on~~ weekly mean temperature forecasts.

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 conditions (Vitart et al. 2008), while the seasonal climate forecast depends on slowly-evolving components of the ~~earth~~ climate system (e.g. sea surface temperature and soil moisture) (Troccoli 2010). However, since sub-seasonal timescale is usually ~~beyond~~ too long to be favoured by ~~the memory of~~ 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 have received much less attention 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 promise great socio-economic value (Vitart et al. 2012). ~~For example, flash drought, heat wave, and cold wave are extreme events at sub-seasonal timescale.~~ In fact, s~~S~~ub-seasonal forecast information can be used for ~~reducing risks in management decision making and~~ developing strategies for proactive natural disaster mitigation, which may be needed during ~~drought and heat waves~~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~~; Toth et al. 2007; Vitart et al. 2008~~). ~~Of many climatic variables, p~~Precipitation and 2-m temperature (hereafter temperature) are considered to be two of the most important ~~factors to~~climate variables that significantly influence ~~management decision making related to, for example:~~ irrigation scheduling, urban water supply, cooling water related to thermal power generation, ~~and~~ hydropower operations, etc. Many important sub-seasonal events including heat waves, cold waves, dry spells, and wet spells are directly derived from frequency, duration, and intensity of rainfall or hot (cold) temperatures. 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 ~~have~~ has been, thus far, no ~~compelete~~complete investigation of sub-seasonal forecasting capabilities for the other temperature and precipitation ~~derivatives or~~ indices that are directly associated with important events and ~~decision-~~ decision-making at sub-seasonal timescales. In this study, we aim~~ed~~ to evaluate forecasting for precipitation and temperature derivatives or indices that are associated with those important events and decision-making at sub-seasonal timescale~~. Those derivatives indices included not only~~ -- the mean, ~~but also the~~ frequency, duration, and intensity of ~~the rain~~precipitation and temperature at sub-seasonal timescale, such as the number of dry/~~rainy~~wet days, number of cold/hot days, etc.

Coupled Atmosphere–Ocean General Circulation Models ~~Global climate models~~ (GCMs) ~~have become useful tools~~are used to make forecasts at multiple timescales. While the GCMs have demonstrated advanced configurations and realistic representations of the climate~~earth~~ systems, the use of GCMs' predictions is still restricted by their coarse resolution and inherent systematic biases. To overcome these limitations, the GCMs' 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; ~~Wood et al., 2002;~~ Yuan et al., 2013; Tian et al., 2014~~; Wood et al., 2002; Yuan et al., 2013~~). The Climate Forecast System version 2 (CFSv2) is a recently developed GCM 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 from 1982 to 2009~~ at a daily or sub-daily time step~~. Despite the availability of those CFSv2 hindcasts ~~daily or sub-daily forecast archives from the newly developed Climate Forecast System version 2 (CFSv2) by National Centers for Environmental Prediction (NCEP) (Saha et al. 2014)~~ , 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 archived sub-daily hindcast datasets over almost 30 years, which provides an opportunity to make such assessments.~~ The CFSv2 has fully coupled atmospheric, oceanic, and land components of the ~~earth~~ climate systems and demonstrated ~~the~~ high 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 GCMs, 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 allows 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 ~~for~~of the precipitation and temperature hindcasts at sub-seasonal timescales. Specifically, t~~T~~he ~~aim of this study is~~aims of this study are to 1) ~~to understand the capacities of using~~assess ~~the~~ CFSv2 predictions for ~~sub-seasonal~~ precipitation and temperature ~~predictions~~~~derivatives or~~ indices at different locations and seasons within the first 30 days, 2) compare weekly and fortnight ~~and weekly~~ forecast skill of ~~the~~ CFSv2 at different lead times, and 3) evaluate the effects of MJO and ENSO on ~~the~~ CFSv2 sub-seasonal forecast skill~~ through a comprehensive evaluation of the precipitation and temperature hindcasts at sub-seasonal timescales~~. The assessment includes mean values of sub-seasonal predictions as well as related temperature and precipitation ~~derivatives or~~ 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 probabilistic temperature and precipitation forecast

Formatted: Do not check spelling or grammar

Formatted: Do not check spelling or grammar skill is also assessed. ~~This study will leverage sub-seasonal forecasting efforts and contribute to sectorial management decision making at sub-seasonal timescales.~~

2. Data and Methodology

The CFSv2 has the most updated data assimilation and forecast model components of the climate system and became operational at NCEP since March 2011 (Saha et al. 2014). The CFSv2 has archived three different types of hindcast (or reforecast): 6-hourly time series from 9-month runs, 45-day runs, and season runs. Table 1 shows those forecasts ~~have~~with different configurations. ~~The 9-month runs are forecasts from initial conditions every 5 days apart (for all 6Z cycles on that day) for each calendar year over the 28-year period (1982-2009). The season runs are forecasts from every 0Z cycle over a 12-year period (1999-2010). The 45-day runs are forecasts from every 6Z, 12Z and 18Z cycle over the same 12-year period (1999-2010).~~ Figure 1 gives an example of the three hindcast configurations. The CFSv2 hindcast has a T126 spatial resolution (roughly 100 km) and includes several near surface variables at a 6-hourly temporal resolution. ~~While the~~ The one season and 45-day reforecasts ~~have initializations~~ are initialized every ~~single~~ day~~,~~ so that ~~and the~~ relatively new initial conditions can be incorporated into a large ensemble size for making a potentially skillful forecast over this shorter forecast period.~~,~~ Nevertheless, we chose to use the 9-month reforecast. This is because ~~in that~~ the 9-month reforecast covered much longer period (1982-20~~10~~09) than one season and 45-day reforecasts (1999-2010), which ~~would~~ ensures a ~~. With a longer period available, we would have~~ larger sample size for~~to ensure~~ a more robust evaluation, ~~. In addition,~~especially for~~And also~~ the ~~consideration~~ evaluation of skill conditioned on~~of~~ MJO and/or ENSO ~~conditions~~ that require long hincasts for to provide sufficient ~~would result in even smaller~~ sample sizes f~~or evaluating the forecast skill conditioned on MJO and ENSO~~. ~~Therefore, this study used the 9-month hindcast.~~

[Insert Table 1 here]
[Insert Figure 1 here]

Formatted: Space After: 0 pt

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 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~~of~~ the ~~0000~~0, ~~0600~~6, ~~1200~~12, and ~~1800~~ 18 UTC (Coordinated Universal Time) ~~cycles for~~ every 5 days. For example, the 24 ensemble members for January are the four cycles for each of December 12^{th}, 17^{th}, 22^{nd}, and 27^{th} and January 1^{st} and 6^{th}.

The forecast validation dataset is from the North American Land Data Assimilation System version 2 (NLDAS-2; Xia et al., 2012). The forcing dataset of the NLDAS-2 merges a large observation-based and reanalysis data and is routinely used to drive land surface models over the CONUS. It has 0.125° (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~~s~~ 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 using daily forecasts from the ~~raw~~ CFSv2 daily and the BM downscaled CFSv2. Table 2 shows the forecast lead time for different periods and methods. For daily forecasts from the ~~raw~~ CFSv2 daily, all precipitation and temperature indices were calculated at 7-, 14-, and 30-day forecast timescales in month 1. For daily forecasts from the BM downscaled CFSv2, the precipitation and temperature indices were only calculated at 30-day forecast timescales in month 1, since these forecasts were temporally disaggregated from monthly forecasts and it would be useful to look at the performance of the ~~raw~~ CFSv2 daily forecast in comparison with the daily forecast disaggregated from the monthly forecast.

[Insert Table 2 here]

Table 3 shows the precipitation and temperature indices calculated in this study. ~~Five precipitation indices were calculated, including mean precipitation (Pmean), mean precipitation over wet days (RainWet), number of rain days (RainDay), maximum rain day (wet spell) length (WetSpell), and maximum dry spell length (DrySpell) during the 30-, 14-, or 7-day periods in forecast month 1.~~ Following Zhang (2011), a wet (dry) day is defined as days with precipitation above (below) 1 mm during the *n*-day period. The wet (dry) spell is defined as number of consecutive wet (dry) days. ~~Take~~ Taking ~~the~~ a 14-day forecast for ~~January 14-day~~ WetSpell as an example (as is shown in Table 2), ~~lead one (two)~~ forecast lead one is the number of consecutive rainy days from ~~January~~ day 1 ~~(15)~~ to day 14 ~~(28).~~ . ~~Similarly, five air temperature indices are calculated, including mean temperature (Tmean), number of high temperature days (HighDay), number of low temperature days (LowDay), maximum number of consecutive high temperature days (CosHighD), maximum number of consecutive low temperature days (CosLowD) during the 30-, 14-, or 7-day periods in forecast month 1.~~ As a way of defining heat (cold) wave (e.g. Spinoni et al. 2015), the threshold for high (low) temperature day is defined when the temperature is above (below) 90^{th} (10^{th}) percentile of climatological distribution of temperature during the *n*-day period for different months.

Formatted: Do not check spelling or grammar

[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 ~~raw~~ CFSv2 forecasts. 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 ~~raw~~ CFSv2 daily forecasts were evaluated at 30-day, 14-day, and 7-day time scales, the BM downscaled forecasts were only evaluated at 30-day time scale since they were disaggregated from monthly forecasts. Take ~~raw~~ native CFSv2 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 1 Jan-30 Jan, 1 Jan-14 Jan, and 1 Jan-7 Jan, respectively. All ensemble forecasts were converted into probabilistic forecasts in terciles ~~and~~with a~~. A~~ll observations ~~were~~ converted into dichotomous values of 1 or 0 ~~in terciles~~. 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 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 forecast 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 divided by the ensemble size. The HSS is expressed as:

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

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~~the reference forecast~~. In this study, since there are three forecast categories, E ~~, which~~ 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 0 representing the same skill as randomly generated forecast, which in this case is the climatological forecast. The HSS above 0 indicates that the forecasts have skill. The HSS was assessed for each method (~~raw~~ CFSv2 daily and BM), variable index, grid point, month, and forecast time. Since 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 predictability of ~~subseasonal~~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)

Formatted: English (United States)

~~National Climatic Data Center (NCDC)~~ climate regions as well as over the entire CONUS (Figure ~~1~~2). The HSS of the average forecasts over each of those regions were evaluated subsequently.

[insert Figure ~~1~~ 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 (http://cawcr.gov.au/staff/mwheeler/maproom/RMM/) for the period of 1982 to 2009. This index is 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), define eight MJO phases and an MJO amplitude. There are a few different ways to define active MJO events. While the simplest criterion was to define~~ing~~ MJO as RMM amplitude exceeds~~ing~~ a certain threshold (e.g. Johnson et al. 2014), this criteria did not consider minimum duration and eastward propagation of MJO. ~~Following L'Heureux and Higgins (2008), this~~ This study adopted a more rigorous definition of MJO: MJO days and events are 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)~~: (i) the aptitude of the index must be greater than one for consecutive pentads; (ii) phases of the index must be in numerical order (i.e., phases 6, 7, 8, 1, 2); and (iii) events must last for more than 30 days (six consecutive pentads) and cannot be in a particular phase for more than 20 days (four pentads)~~. ~~When the amplitude of one pentad is slightly below one, they are still included as part of a larger MJO event.~~ 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 are 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). ENSO periods of below and above normal SSTs are when the threshold is met for a minimum of 5 consecutive overlapping seasons.

3. Results

3.1 The 30-day forecast skill

Figure~~s~~ 3~~2~~ shows 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).

~~[insert Figure 2 here]~~

Formatted: English (United States)

Formatted: Do not check spelling or grammar

In DJF, the average skill of WetSpell over the CONUS is 34, which is much higher than the other indices; it showed high skill over major area of the CONUS including midwest and eastern US. The Pmean, RainDay, and DrySpell were skillful in the southeast and the southwest but also revealed skill in 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 coast regions with the WetSpell showing some skill in the midwest and northeast. For the other seasons, on average, the forecast skill for precipitation indices is between DJF and JJA for MAM but slightly lower than JJA for SON (~~not shown~~Figure 4).

[insert Figure 3 here]

[insert Figure 4 here]

Formatted: Space After: 0 pt

Spatial patterns in HSS are 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:

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

Where STD is the standard deviation of time series for each grid point, min(STD) is the minimum STD over all grid points, max(STD) is the maximum STD over all grid points, and STD’ is the normalized STD. ~~Figure 3~~Figure 5 shows normalized standard deviation of 30-day precipitation indices in January and July over 28-year period from 1982 to 2009 over the CONUS.

~~[insert Figure 3 here]~~

By comparing interannual variability (~~Figure 3~~Figure 5) with the forecast skill over the space (~~Figure 2~~Figure 3), we found that regions showing lower interannual variability usually have higher skill than the regions with higher interannual variability. Particularly in July, f~~F~~or 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 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]

Formatted: Line spacing: Double

Figure~~s~~ ~~4~~ 6 shows average HSS for 30-day forecasts of temperature indices calculated from the CFSv2 daily at different locations over DJF and JJA.

~~[insert Figure 4 here]~~

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 is skillful over major area of the CONUS, particularly in the northeast. HighDay and CosHighD showed notable high skill around south of the centeral area. For the other seasons, on average, the forecast skill for temperature indices is between DJF and JJA for MAM but slightly lower than JJA for SON (~~not shown~~Figure 7).

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

Formatted: Space After: 0 pt

~~Figure 5~~Figure 8 shows average HSS for 30-day forecasts of precipitation and temperature indices calculated from the CFSv2 daily or BM downscaled CFSv2 over 12 months for CONUS and its consistent ~~NCDC~~NCEI climate regions.

~~[insert Figure 5 here]~~

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 tempo~~e~~rally downscaled monthly forecasts for hydrological applications.

[insert Figure 8 here]

3.2 Weekly and f~~F~~ortnight ~~and weekly~~ forecast skill at different lead times

~~Figure 6~~Figure 9 (~~Figure 7~~Figure 10) shows average HSS of 14- and 7-day precipitation (temperature) indices forecasts from the ~~raw~~ CFSv2 daily over 12 months for CONUS and its consistent ~~NCDC~~NCEI climate regions.

~~[insert Figure 6 here]~~

~~[insert Figure 7 here]~~

In general, the skill scores for precipitation indices are reasonably ~~reasonaly~~ higher in the first two weeks than the second two weeks at both 14- and 7-day time scales, since first two weeks are within the range of weather forecast and are strongly influenced by the atmospheric initial conditions. While there 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 is worth noting that the skill 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 indicates that the temporal noise in predictions can be reduced through averaging, as ~~is~~ noted by Roundy et al. (2015).

[insert Figure 9 here]

[insert Figure 10 here]

3.3 Effects of MJO and ENSO

Formatted: Font: Bold

~~Figure 8~~Figure 11 shows skill differences between Pmean or Tmean weeks 2-4 forecasts during active ENSO, MJO, or combined active ENSO and MJO (MJO+ENSO) and the forecasts during all periods for CONUS and its consistent ~~NCDC~~NCEI climate regions. The Pmean and Tmean forecasts are calculated from the CFSv2 daily. In general, weeks 3 and 4 Pmean forecasts perform better during active ENSO or MJO states, while Tmean forecasts do not perform better.

[insert ~~Figure 8~~Figure 11 here]

For precipitation, forecast skill is inconsistent for the active ENSO, MJO, or combined ENSO and MJO relative to all 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 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 and for forecasts during all period at different locations over the CONUS for DJF, MAM, JJA, and SON (Figures 12 and 13). Since HSS evaluated forecast

Formatted: Space After: 0 pt

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. The results are shown in Figures 12 and 13. ~~(Figures 9 and 10).~~

[insert ~~Figure 9~~Figure 12 here]
[insert ~~Figure 10~~Figure 13 here]

In general, most of the skill is significantly different at different locations; MJO has strongly positive effects on CFSv2 ~~subseasonal~~sub-seasonal Pmean forecast skill over the CONUS; the effects on Tmean forecast skill is relatively weak and inconsistent among different regions. For precipitation, the influenced areas are greater during DJF and MAM than during JJA and SON, with the NE and NW regions being consiste~~a~~ntly influenced by MJO during four seasons. Aggregated over the spatial domain, ~~We~~ we further conducted statistical tests to compare whether precipitation forecast skills during active MJO, ENSO, or combined MJO and ENSO are greater than those during all 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) are statistically significant at a 5% level. The student's t-test showed the forecast skills during active MJO or combined MJO and ENSO were significantly greater than those during all period ($p<0.05$) over the CONUS for DJF, MAM, JJA, and SON; the forecast skills during active ENSO were significantly greater than those during all period over the CONUS for MAM. It is also worthwhile to note that the combined effects of MJO and ENSO are stronger than the individual effects of either MJO or ENSO, suggesting a potential benefit of using MJO and ENSO information for sub-seasonal forecasts. Table 4 shows there are 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 ~~subseasonal~~sub-seasonal forecast skill is highly depend~~ing~~ent on ~~forecasting~~ target 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 the ~~This finding is important since s~~sub-seasonal forecasting information~~, particularly frequency or duration of precipitation and temperature extremes,~~ is valuable to many decision ~~makings~~makers ~~and can thus be directly tailored to different application needs~~. 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 RainDay, WetSpell, and DrySpell weeks in advance will help farmers make irrigation scheduling decisions, save water costs and improve crop yields. Short-term planning of urban water supply could also benefit from this forecasting information since ~~these~~ those ~~precipitation and temperature~~ 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 hot/cold waves, forecasting information of these indices would also be useful for developing strategies for proactive disaster mitigation (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 resolutions. ~~It is worth to note that the skill scores only reflect the performance of the forecast anomalies since the terciles were defined individually for the forecasts and observations. The BM method, while also serving as a bias correction method in previous hydrological forecasting studies (e.g. Yuan et al. 2013), mainly played the role of spatial and temporal downscaling in this study.~~ For forecasting 30-day precipitation and temperature indices, since the precipitation and temperature indices calculated from CFSv2 daily have ~~slightly~~ mostly higher skill than the BM ~~in most cases~~, the comparison of these two methods implies that daily forecasts from the CFSv2 are potentially more useful than those disaggregated from the monthly forecasts. Thus, the CFSv2 daily forecast information should be used in application studies of sub-seasonal hydrological forecasts in contrast to temporal disaggregation of the monthly forecast.

This study demonstrated 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 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 CFSv2-based forecast performance of precipitation over major areas of CONUS. This finding corresponds to the study of Jones et al. (2011), who found improved deterministic CFSv1 forecast skill of extreme precipitation during active MJO. We also compared the regions of improved skill associated with the MJO in this study (Figure ~~13~~9) with the ~~regions shown in Figure 5~~~~Figure 8~~results in 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). This result is consistent with current knowledge of the observed influence of the MJO on precipitation events along the US west coast, which can be viewed 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 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 ~~further~~ improved by having a larger ensemble size. A sensitivity study on ensemble size could be performed ~~to see if~~assess whether a larger ensemble ~~does~~ improves forecast skill. For future work, when one season or 45-day CFSv2 reforecasts are available over a longer period, we would choose to use those datasets instead of 9-month reforecasts 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 model, 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~~ed~~ 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 ~~subseasonal~~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). Sub-seasonal forecasting of flash drought will help decision makers develop mitigation strategies. 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 CFSv2 probabilistic sub-seasonal forecasts of precipitation and temperature indices over the CONUS. The probabilistic 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. Forecasts for 7- and 14-day temperature indices even showed skill at weeks 3 and 4, and generally more skillful than precipitation indices. Forecasts of 30-day temperature and precipitation indices calculated from the daily forecasts BM downscaled from the monthly forecasts mostly showed lower skill compared to those calculated 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 forecast performance of precipitation over major 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 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.

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.

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.~~, n/a n/a.~~

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.

Formatted: Font: (Asian) +Body Asian (SimSun), Not Bold, (Asian) Chinese (PRC), (Other) English (United States), Do not check spelling or grammar

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, 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.

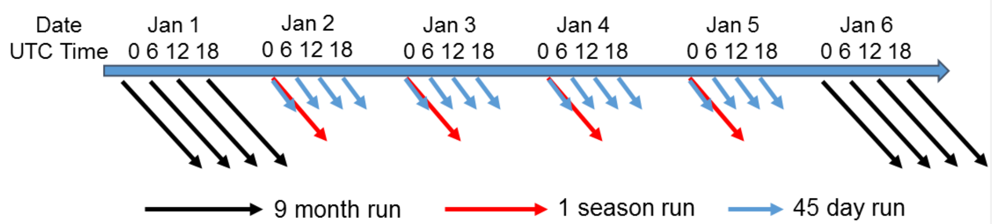


Figure 1. Three configurations of the CFSv2 hindcast: 9-month run, 1 season run, and 45-day run. UTC stands for Coordinated Universal Time.

Formatted: Font: 9 pt, Bold, Do not check spelling or grammar

Formatted: Line spacing: Double

Formatted: Font: 9 pt, Bold, Do not check spelling or grammar

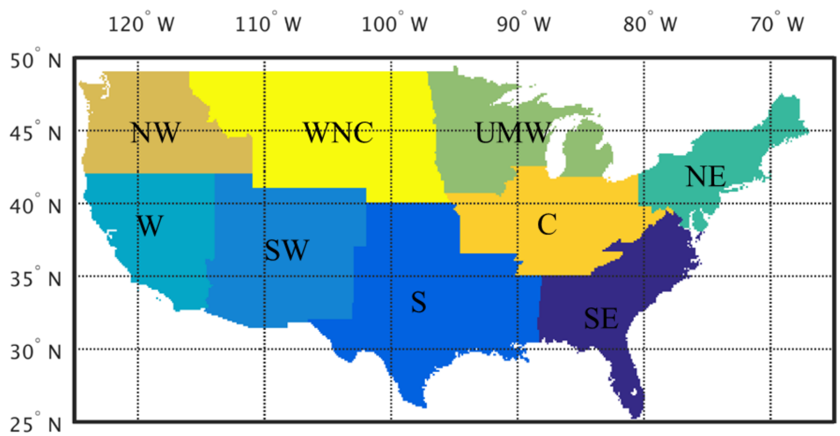
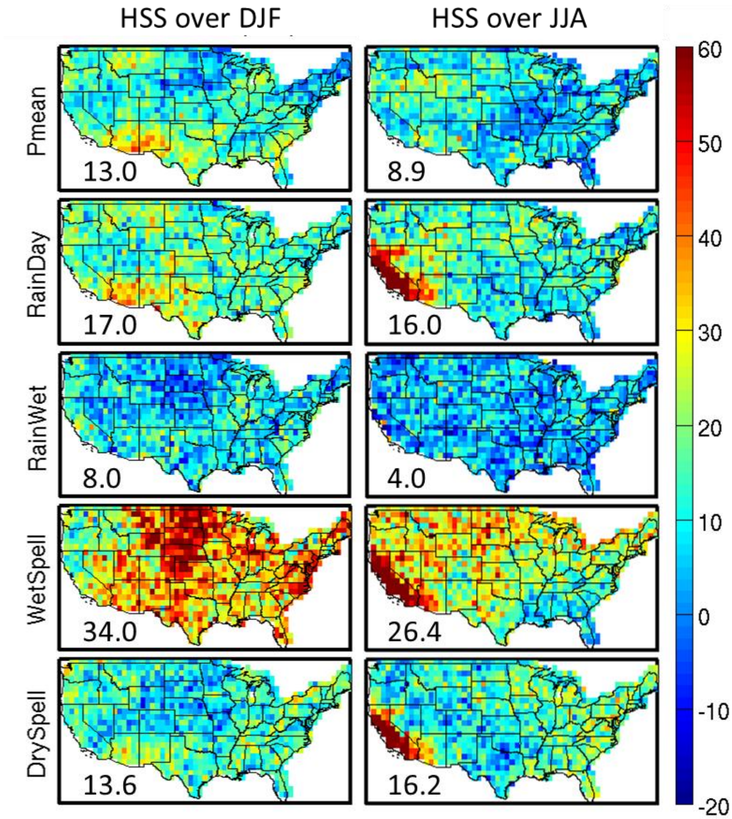


Figure ~~1~~2. ~~NCDC~~NCEI climate regions (described in Section 2) used as area averaging domains for raw and BM downscaled CFSv2 forecasts. Regions are named as follows: Northwest (NW), West (W), Southwest (SW), West North Central (WNC), South (S), Upper Midwest (UMW), Central (C), Southeast (SE), and Northeast (NE).



~~Figure 2~~Figure 3. HSS of 30-day (from top to bottom columns) Pmean, WetRain, RainDay, WetSpell, and DrySpell from (from left to right rows) the ~~Raw~~CFSv2 daily and BM over DJF (left) and JJA (right). The number in the bottom left is the overall average.

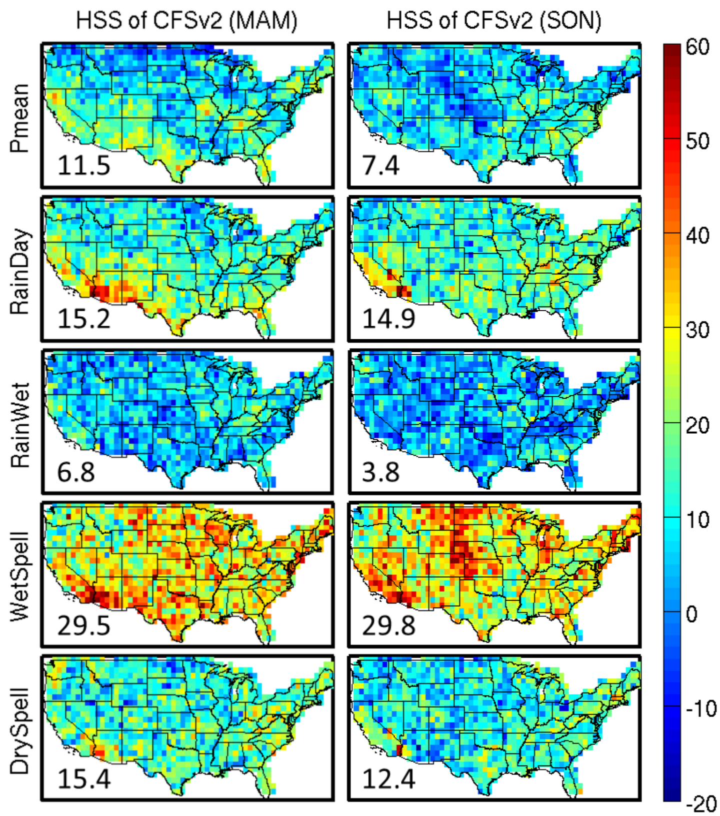
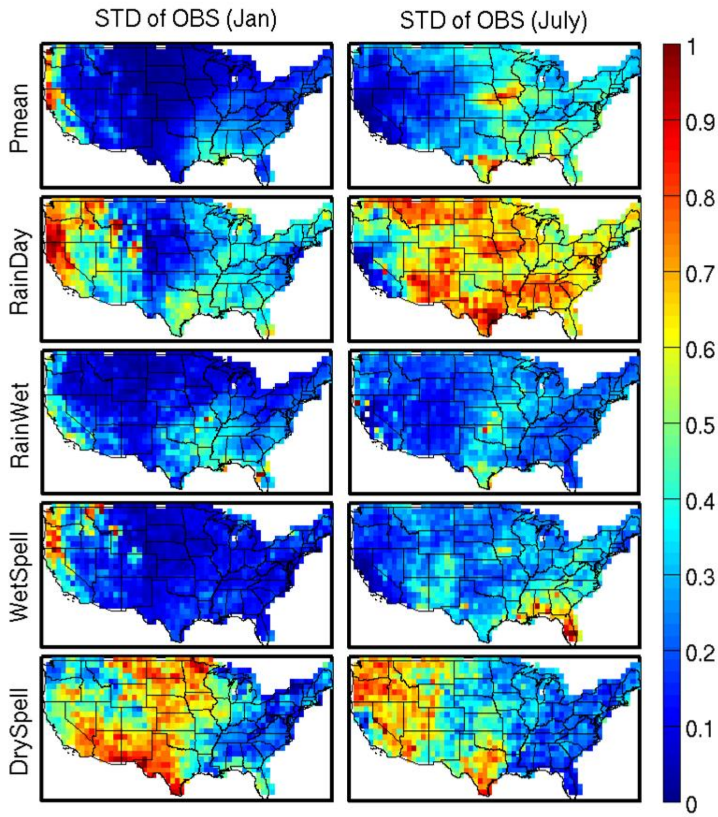
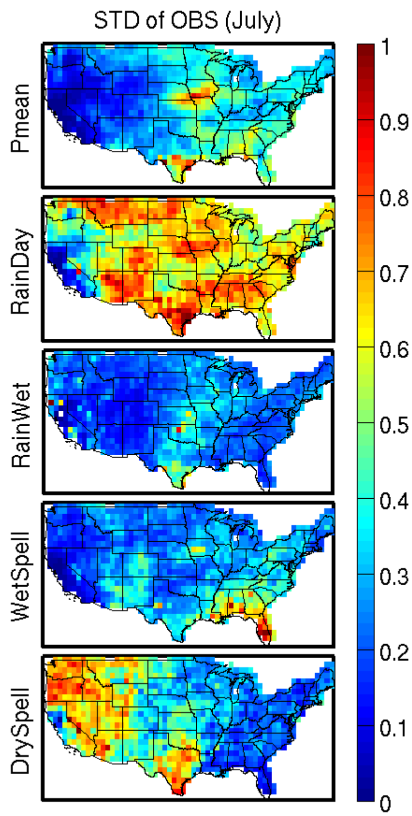


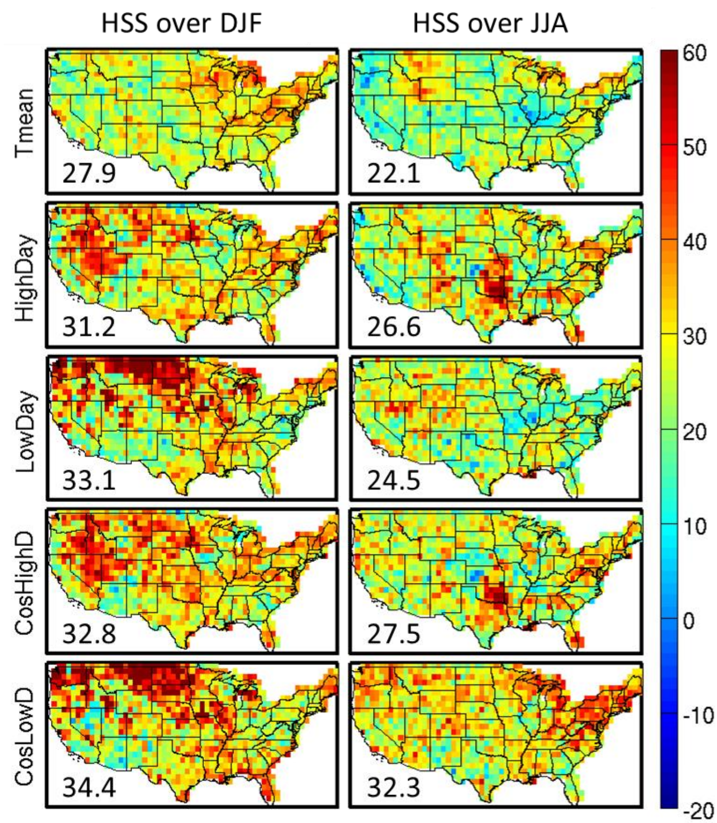
Figure ~~5~~4. Same as in Figure 3, but for MAM and SON.



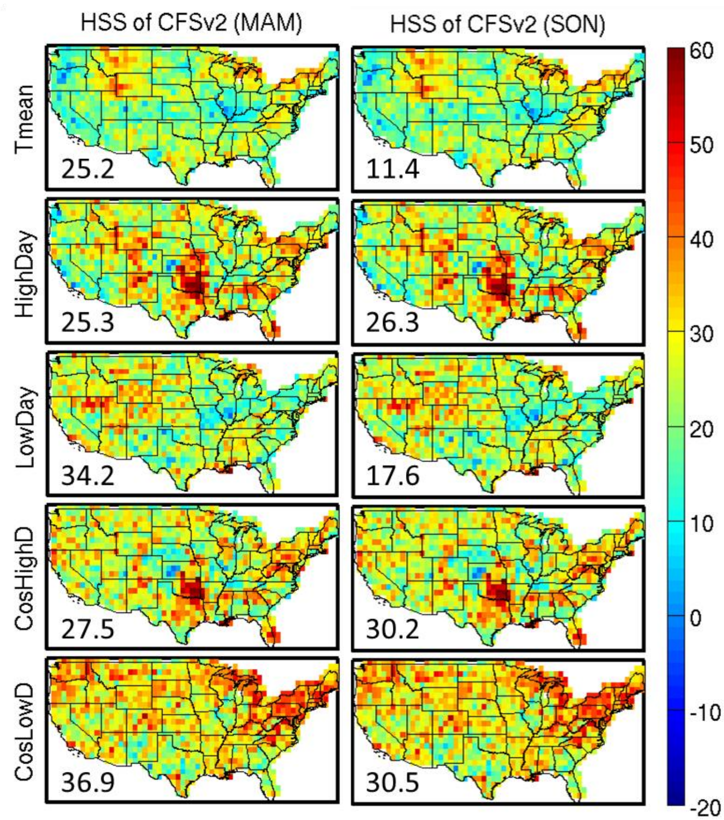
Formatted: Font: 9 pt, Bold



~~Figure 3~~Figure 45. Spatially normalized standard deviation of observed 30-day precipitation indices in ~~Janauary~~January and July over 28-year period from 1982 to 2009

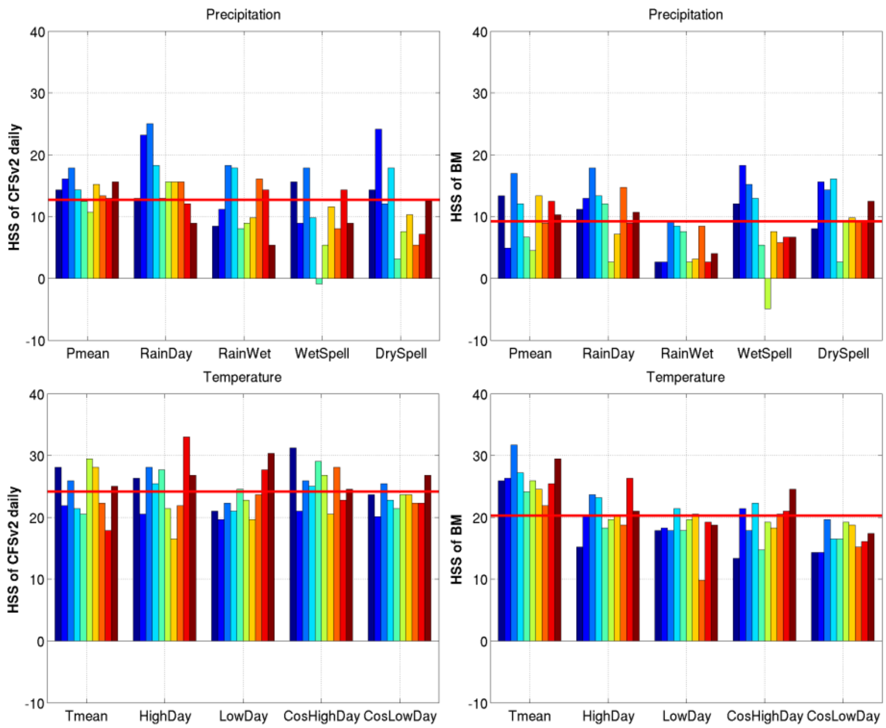


~~Figure 4~~Figure 6. HSS of 30-day (from top to bottom columns) Tmean, HighDay, LowDay, CosHighD, and CosLowD from (from left to right rows) the ~~Raw~~CFSv2 daily and BM over DJF (left) and JJA (right). The number in the bottom left is the overall average.

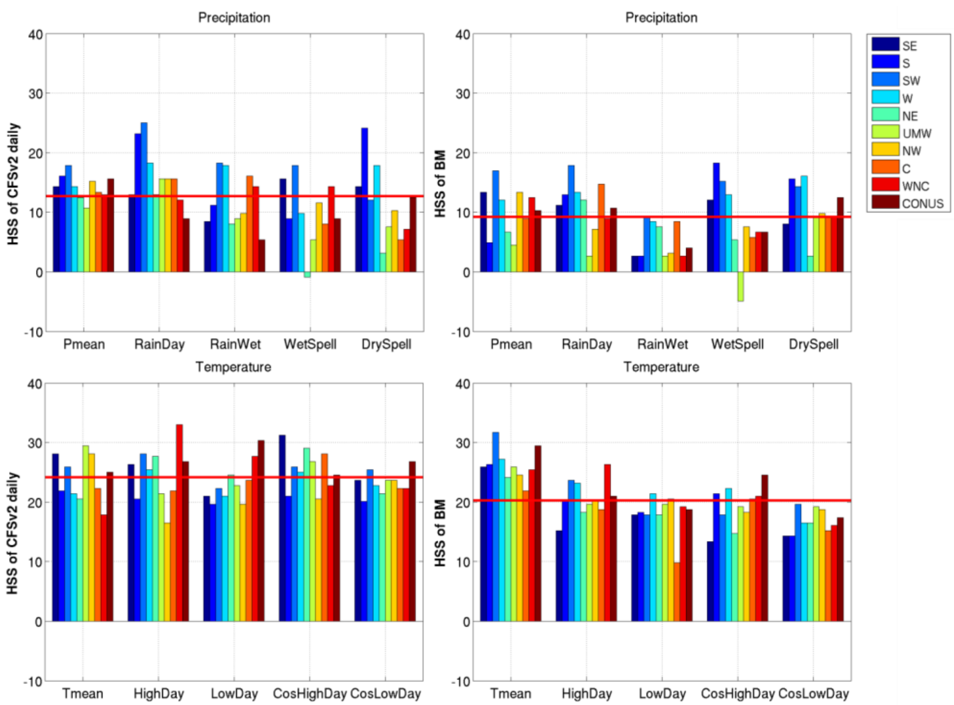


Formatted: Font: 9 pt, Bold

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

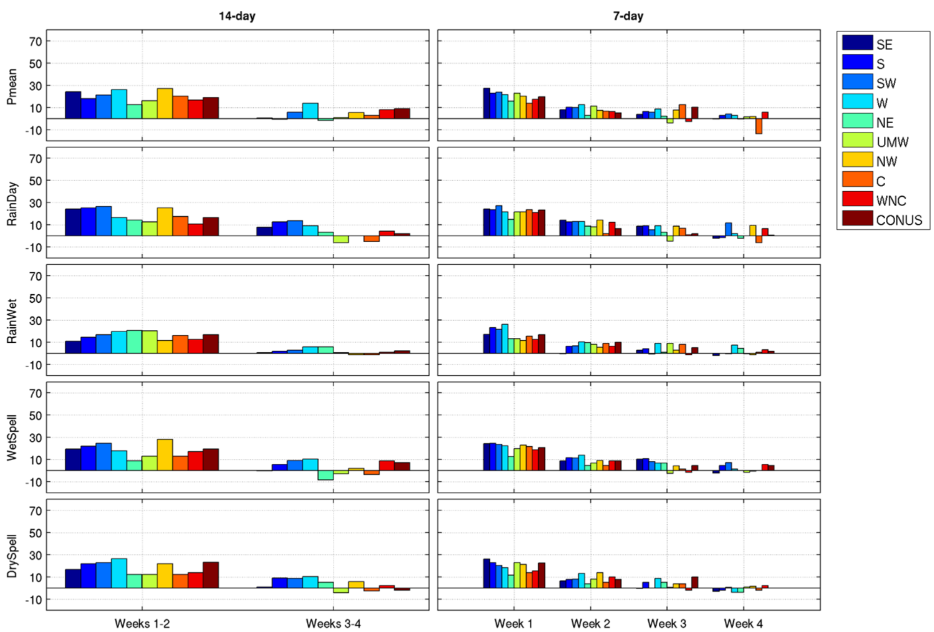


Formatted: Font: 9 pt, Bold

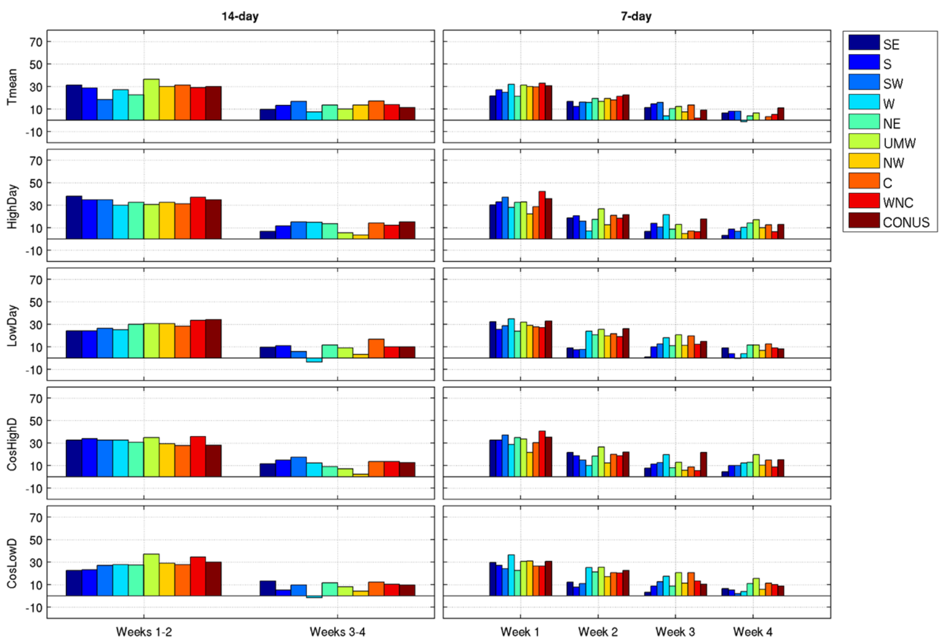


Formatted: Font: 9 pt, Bold

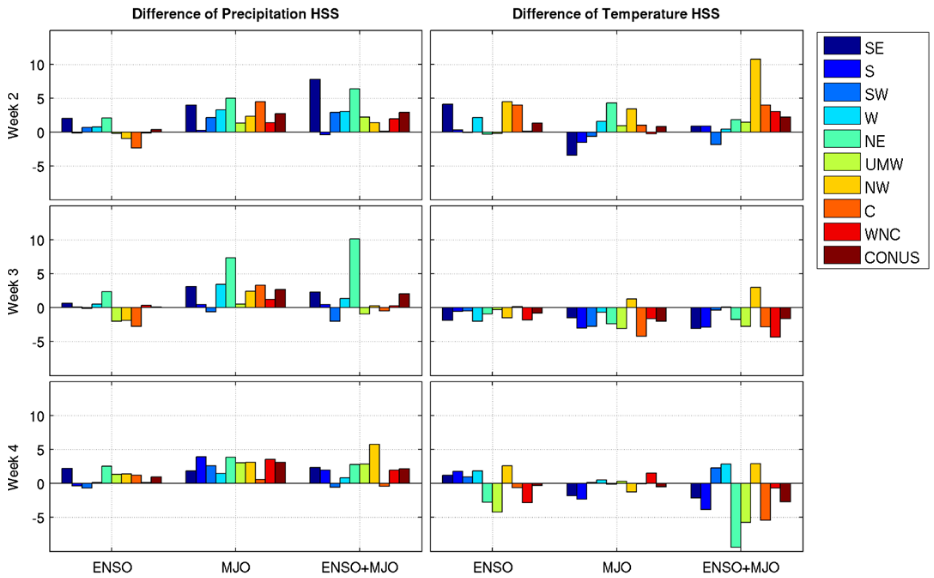
~~Figure 5~~Figure 8. HSS of 30-day precipitation and temperature indices calculated from the CFSv2 ~~daily~~ and BM for CONUS and its consistent ~~NCDC~~NCEI climate regions. The red line is the overall average.



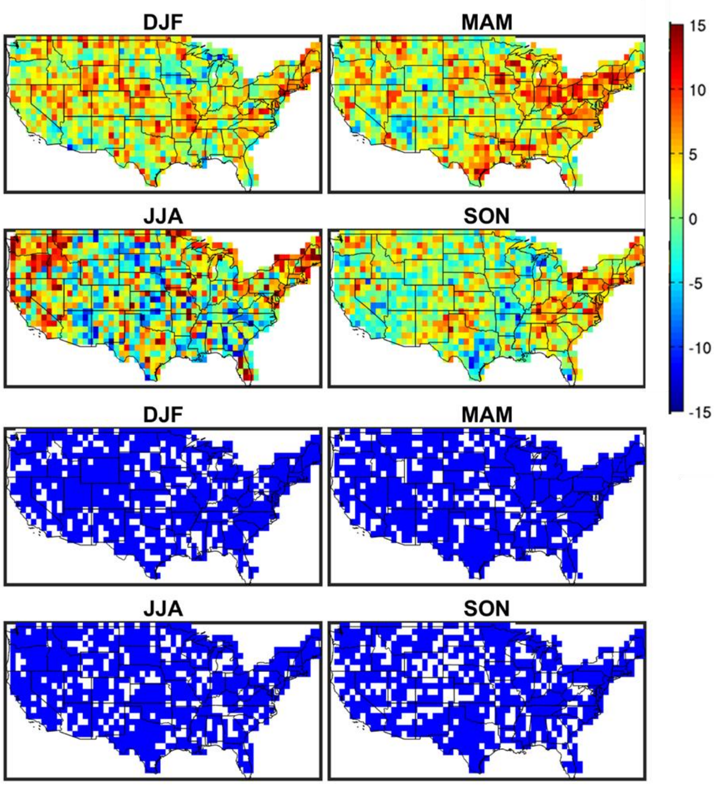
~~Figure 6~~Figure 9. Overall mean HSS of 14- and 7-day (from top to bottom rows) Pmean, WetRain, RainDay, WetSpell, and DrySpell from the ~~Raw~~CFSv2 daily for CONUS and its consistent ~~NCDC~~NCEI climate regions

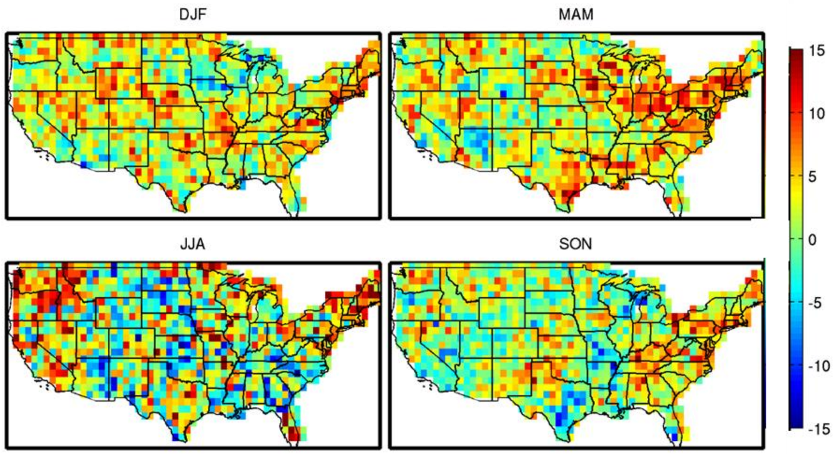


~~Figure 7~~Figure 10. Overall mean of HSS of 14- and 7-day (from top to bottom rows) Tmean, HighDay, LowDay, CosHighD, CosLowD from the ~~Raw~~CFSv2 daily for CONUS and its consistent ~~NCDC~~NCEI climate regions

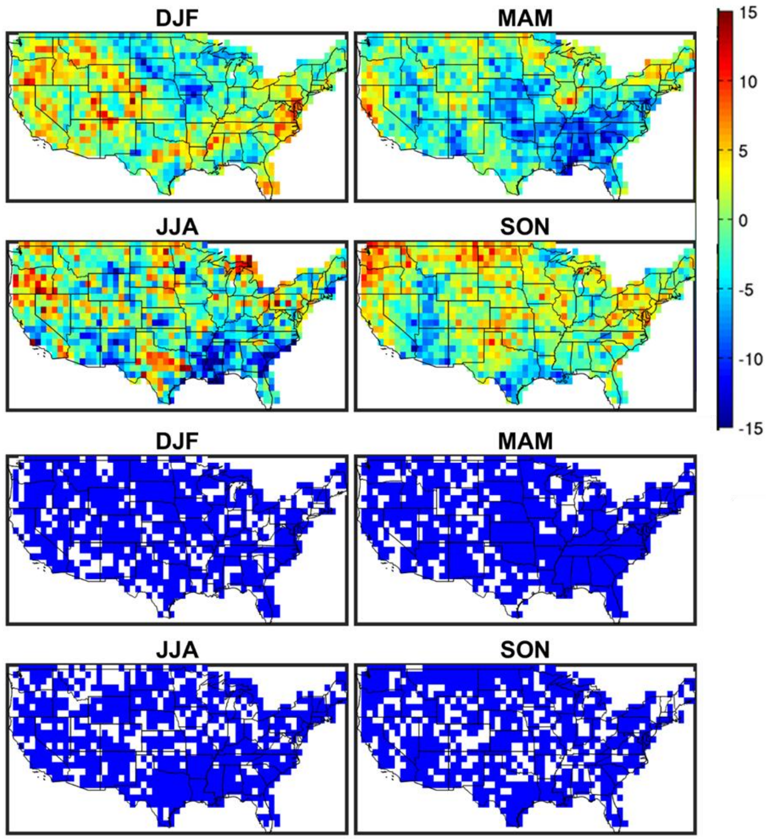


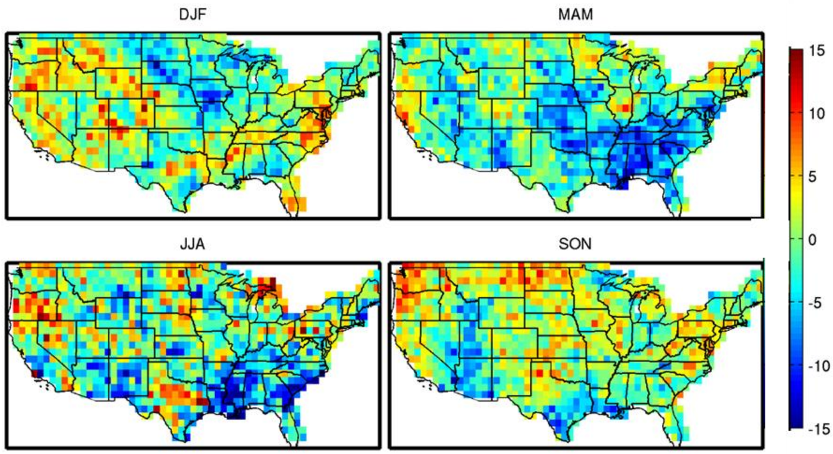
~~Figure 8~~Figure 11. HSS differences between Pmean (left column) or Tmean (right column) weeks 2-4 forecasts during active ENSO, MJO, or combined active ENSO and MJO (MJO+ENSO) and the forecasts during all periods for CONUS and its consistent ~~NCDC~~NCEI climate regions. Positive values indicate more skillful forecasts for the active MJO, ENSO, or ENSO+MJO.


DJF
MAM
JJA
SON
15
10
5
0
-5
-10
-15
DJF
MAM
JJA
SON



~~Figure 9~~Figure 12. The upper panel shows ~~d~~Differences between Pmean average HSS over weeks 2-4 for forecasts during active MJO and for forecasts during all period at different locations over the CONUS for DJF, MAM, JJA, and SON. The blue pixel in the lower panel shows whether the difference is significant ($p<0.05$).


DJF
MAM
JJA
SON
15
10
5
0
-5
-10
-15
DJF
MAM
JJA
SON



~~Figure 10~~Figure 13. Same as in ~~Figure 9~~Figure 12, but for Tmean.

Formatted: Normal

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 time 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 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

Formatted: Font: 9 pt, Bold