AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2014GL059586

Usefulness of ensemble forecasts from NCEP Climate Forecast System in sub-seasonal to intra-annual forecasting

Sanjiv Kumar[1,2], Paul A. Dirmeyer[1,3], and J. L. Kinter III[1,3]

[1]Center for Ocean-Land-Atmosphere Studies, George Mason University, Fairfax, Virginia, USA, [2]National Center for Atmospheric Research, Boulder, Colorado, USA, [3]Department of Atmospheric, Oceanic, and Earth Sciences, George Mason University, Fairfax, Virginia, USA

Key Points:

- Ensemble spread - a required component of the forecast
- CFSv2 provides useful ensemble climate forecasts from 0 to 9 months leads
- Forecast Biases: a function of forecasts initialization month and lead-time

Supporting Information:

- Readme
- Supplementary_text

Correspondence to:
S. Kumar,
kumar@uvic.ca

Citation:
Kumar, S., P. A. Dirmeyer, and J. L. Kinter III (2014), Usefulness of ensemble forecasts from NCEP Climate Forecast System in sub-seasonal to intra-annual forecasting, *Geophys. Res. Lett.*, *41*, 3586–3593, doi:10.1002/2014GL059586.

Received 10 FEB 2014
Accepted 25 APR 2014
Accepted article online 30 APR 2014
Published online 29 MAY 2014

Abstract Typically, sub-seasonal to intra-annual climate forecasts are based on ensemble mean (EM) predictions. The EM prediction provides only a part of the information available from the ensemble forecast. Here we test the null hypothesis that the observations are randomly distributed about the EM predictions using a new metric that quantifies the distance between the EM predictions from the National Centers for Environmental Prediction (NCEP) Climate Forecast System version 2 (CFSv2) and the observations represented by CFSv2 Reanalysis. The null hypothesis cannot be rejected in this study. Hence, we argue that the higher order statistics such as ensemble standard deviation are also needed to describe the forecast. We also show that removal of systematic errors that are a function of the forecast initialization month and lead time is a necessary pre-processing step. Finally, we show that CFSv2 provides useful ensemble climate forecasts from 0 to 9 month lead time in several regions.

1. Introduction

The National Centers for Environmental Prediction (NCEP) Climate Forecast System version 2 [CFSv2, *Saha et al.*, 2014] produces sub-seasonal to intra-annual ensemble climate forecasts. For most operational forecasts, only the ensemble mean (EM) is employed, discarding other potentially valuable information available in the ensemble. Hence, EM-only predictions can potentially leave society underprepared (or improperly prepared) for upcoming weather and climate events. For example, the forecast of El Niño and the Southern Oscillation (ENSO) issued on 9 September 2013 was as follows: "The CFSv2 ensemble mean predicts ENSO-neutral conditions into early 2014 [National Oceanic and Atmospheric Administration, Climate Prediction Center (NOAA CPC), 2013]". ENSO is a tropical Pacific climate phenomenon usually measured by the sea surface temperature (SST) anomaly in the NINO3.4 region (5°S–5°N, 120°W–170°W) that affects inter-annual hydroclimatic variability worldwide [*Ropelewski and Halpert,* 1987; *Trenberth,* 1997]. Following the NOAA CPC definition, we have determined ENSO states based on SST anomaly in the NINO3.4 region using the following thresholds: La Niña: SST anomaly $\leq -0.5^{\circ}\mathrm{C}$, ENSO neutral: $-0.5^{\circ}\mathrm{C} <$ SST anomaly $< 0.5^{\circ}\mathrm{C}$, and El Niño: SST anomaly $\geq 0.5^{\circ}\mathrm{C}$. Figure 1a shows an example of CFSv2 ENSO forecasts initialized in May 2006 from the retrospective forecasts set [CFSv2 reforecast; *Saha et al.,* 2014]. The terms "reforecast" and "forecast" are used interchangeably in this study since they both refer to predictions made using only antecedent information. While the EM predicted ENSO-neutral conditions for the ensuing 2006 fall and winter, the ensemble forecast also indicated the potential for El Niño conditions, which occurred (Figure 1a). The EM-only prediction left society underprepared in this example. CFSv2 reforecasts (1982 to 2009) provide an opportunity to investigate this issue in detail.

The EM predictions are also employed for forecast evaluations and hydrologic applications. *Yuan et al.* [2011], *Mo et al.* [2012] and *Dirmeyer* [2013] found that the CFSv2 forecasts lose considerable skill beyond the month in which forecasts are initialized (0 month lead time). While *Murphy* [1988] has shown that the EM prediction is the single best prediction in terms of minimum root mean square errors [also see *Whitaker and Loughe*, 1998], it has not been determined whether the EM prediction (anomaly prediction) is significantly better than a prediction made based solely on the assumption of white noise with zero mean and the standard deviation given by the ensemble predictions' standard deviation. This can be succinctly stated as a null hypothesis: the observations are randomly distributed about the EM. Here, we test the null hypothesis by quantifying the distance between the EM predictions and the observations. Results from this study are expected to inform operational forecasting.

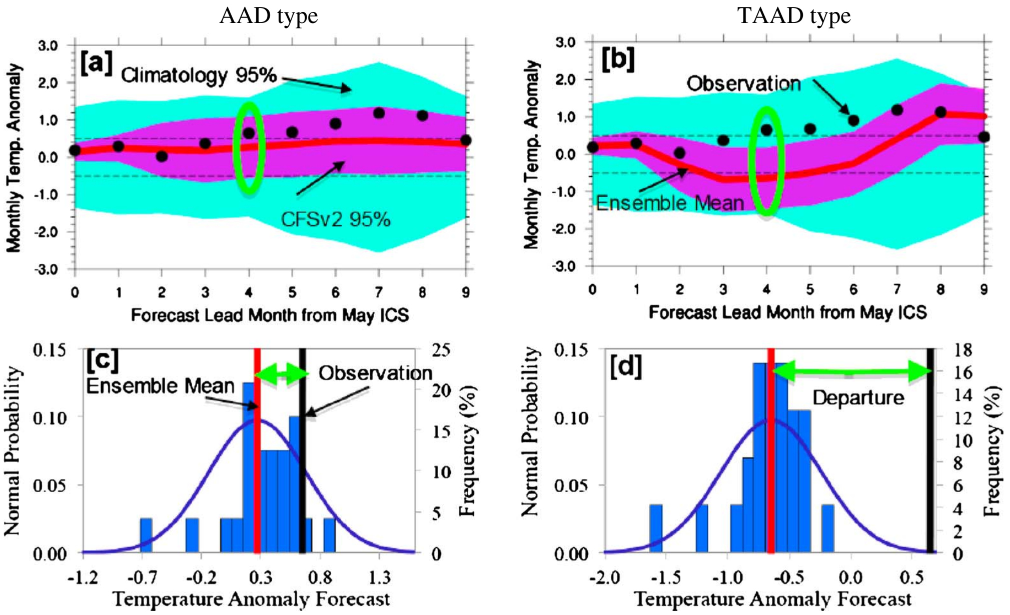

Figure 1. An example of an El Niño and the Southern Oscillation (ENSO) forecast initialized in May 2006. In the Absolute Anomaly Departure (AAD) type, reforecast anomalies are calculated relative to reforecast climatology (1999 to 2009) initialized in May at corresponding lead time and observation anomalies are calculated relative to the observation climatology (1999 to 2009). In Traditional Absolute Anomaly Departure (TAAD)-type forecasts, both anomalies are calculated relative to the observation climatology. See text for details. Figures 1c and 1d show the frequency distribution of 24 members ensemble forecast, as well as fitted normal distribution curves at 4 month lead time forecast.

2. Methodology

We have employed 24-member, 10 month near surface air temperature ensemble reforecasts for each month of the period 1982 to 2009. Forecasts for the month in which they have been initialized are referred to as 0 month lead forecasts, and forecasts for subsequent months are referred to as 1 to 9 month lead forecasts. The monthly means provided by the CFSv2 Reanalysis and Reforecast project are used [*Saha et al.,* 2014]. Each ensemble reforecast at a given lead time is evaluated using the corresponding observations from the CFSv2 Reanalysis. The CFSv2 Reanalysis provides continuous data coverage (spatially and temporally) by assimilating observations from sources including radiosonde data, satellite data, and surface observations using a coupled climate (atmosphere-ocean-land) modeling system, namely CFSv2, as the source of first-guess information in the assimilation cycle [*Saha et al.*, 2010]. The main difference between CFSv2 Reanalysis and CFSv2 Reforecasts is that observations are assimilated every 6 h in the CFSv2 Reanalysis, whereas the CFSv2 reforecasts are 10 month freely running coupled climate simulations using initial conditions from the CFSv2 Reanalysis [*Saha et al.,* 2010, 2014]. We have referred to reforecasts by their month of initialization.

The departure of the ensemble mean reforecasts from the observations is defined in two ways: (1) Absolute Anomaly Departure [AAD, equation (1)] and (2) Traditional Absolute Anomaly Departure [TAAD, equation (2)]. The difference between AAD and TAAD is that AAD calculates reforecast anomalies relative to the reforecast climatology from the corresponding initialization month and the corresponding lead time; whereas TAAD calculates reforecast anomaly relative to the observation climatology:

$$AAD_{m,l,y} = \frac{abs\left[\left(\frac{1}{n}\sum_{i=1}^{n} fA_{i,m,l,y}\right) - OA_{m,l,y}\right]}{S_{m,l,y}} \tag{1}$$

$$fA_{i,m,l,y} = f_{i,m,l,y} - \frac{1}{p}\sum_{y=1}^{p}\left(\frac{1}{n}\sum_{i=1}^{n} f_{i,m,l,y}\right) \tag{1.a}$$

$$OA_{m,l,y} = O_{m,l,y} - \frac{1}{p}\sum_{y=1}^{p}\left(O_{m,l,y}\right) \tag{1.b}$$

$$TAAD_{m,l,y} = \frac{abs\left[\left\{\left(\frac{1}{n}\sum_{i=1}^{n} f_{i,m,l,y}\right) - \left(\frac{1}{p}\sum_{y=1}^{p} O_{m,l,y}\right)\right\} - OA_{m,l,y}\right]}{s_{m,l,y}} \tag{2}$$

where the reforecast ($f_{i,m,l,y}$, $i = 1, 2, \ldots n$ ensemble members) initialized in month (m) and year ($y = 1, 2, \ldots. p$; p (=28)) is the total number of years for which reforecasts are available, i.e., 1982 to 2009) verifying at lead month (l) against the corresponding observations ($O_{m,l,y}$). Subscripts m, l, and y in $O_{m,l,y}$ are indicative of the observation's relationship with the forecasts ($f_{i,m,l,y}$,); as such observations do not have lead time dependency. For example, forecasts initialized in May 2006 are verified at 2 month lead ($l = 2$) using observations from July 2006; the same observations (July 2006) are also used to verify the forecast initialized in March 2006 at 4 month lead. $s_{m,l,y}$ is the standard deviation calculated across 24-member ensemble reforecasts; $s_{m,l,y}$ is also referred as Ensemble Spread [ES, *World Meteorological Organization,* 2012].

$AAD_{m,l,y}$ represents the absolute distance between the ensemble mean reforecast anomaly and the corresponding observation anomaly (Figure 1c). The reforecast anomaly ($fA_{i,m,y,l}$) is calculated relative to the reforecast climatology from the corresponding initialization month and the corresponding lead time (equation (1a)), and the observation anomaly (OA $_{m,y,l}$) is calculated relative to the corresponding observed climatology (equation (1b)). We employ two different climatologies (reforecast and observation) to calculate anomalies because we found significant biases in the reforecast climatology that are a function of reforecast initialization month and lead time (shown later). TAAD represents the traditional anomaly calculation methodology where a common climatology, i.e., the observed climatology, is used to calculate the forecast and observation anomalies (equation (2), Figures 1b and 1d). We used a threshold of $s_{m,l,y} \geq 0.25$℃, i.e., forecasts with $s_{m,l,y} < 0.25$℃ are not considered in the analysis. This threshold prevents either $AAD_{m,l,y}$ or $TAAD_{m,l,y}$ from becoming unrealistically large to reject the null hypothesis (equations (1) and (2)). We selected 0.25°C after preliminary investigations with lower values for which similar results (discussed later) are found for most of the world except for the eastern equatorial Pacific region where the lower ensemble spread (less than 0.25°C) at 0 and 1 month lead times results in higher AAD, and thereby rejection of the null hypothesis (see supporting information Figures S.1 and S.2). *Peng et al.* [2011] have found that the eastern equatorial Pacific region has high predictability. We have referred to our methodology as an "AAD-type forecast," and the traditional methodology is referred to as a "TAAD-type forecast."

Kumar et al. [2012] and *Zhang et al.* [2012] have found a discontinuity in CFSv2 reforecast and CFSv2 Reanalysis data in October 1998 due to the introduction of new satellite data in the assimilation system. To overcome this issue, we have used the 1982 to 1998 climatology (observations and reforecasts) in the 1982 to 1998 calculations, and the 1999 to 2009 climatology in the 1999 to 2009 calculations, as suggested by *Xue et al.* [2013].

The optimal outcome for AAD- or TAAD-type forecasts is AAD or TAAD = 0. A lesser degree of expectation is that errors in AAD- or TAAD-type forecasts are white noise (the null hypothesis), i.e., the observations are randomly distributed about the ensemble mean predictions. If x has a white noise Gaussian distribution (mean = 0, and standard deviation = 1) and u = abs(x), then u has a half-normal distribution. Using standard hypothesis testing procedures, it can be shown that the null hypothesis can be rejected at the 95% confidence interval, i.e., the significance level (α) = 0.05 if:

$$\overline{u} \leq \mu_0 - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{p}} \quad or \quad \overline{u} \geq \mu_0 + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{p}} \tag{3}$$

where $\overline{u}$ is the mean of u, μ_0 is the mean of the half-normal distribution (=0.80), σ is the standard deviation of the half-normal distribution (=0.60), p is the sample size used to calculate $\overline{u}$, and $Z_{\alpha/2}$ is the standard normal distribution corresponding to the significance level $\alpha/2 = 0.025$ for a two-tailed test. Further details about the half-normal distribution and the hypothesis testing design are provided in the Supplementary Materials. An advantage of using the absolute value function in equations (1) and (2) is if $\overline{u} \leq \mu_0 - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{p}}$, then the null hypothesis is rejected for the right reason (the optimal outcome), i.e., the EM prediction is statistically indistinguishable from the observation. Whereas $\overline{u} \geq \mu_0 + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{p}}$ indicates that the EM prediction is too far away from the observation that may fall outside the forecast ensemble 95% range (e.g., Figures 1b and 1d). We also used a standard "f-ratio test" to compare forecast ES against climatological inter-annual variability [*Miller and Miller*, 2004; see supporting information for details].

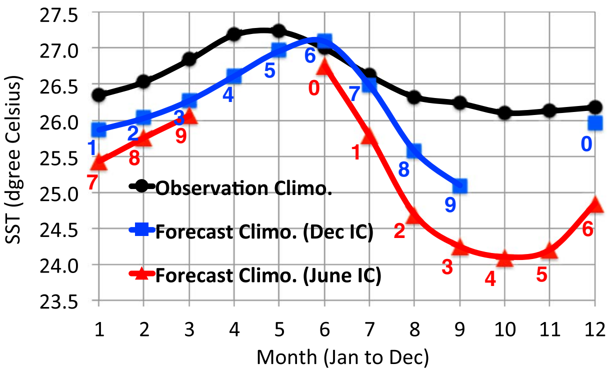

Figure 2. Observation (Climate Forecast System version 2 (CFSv2) Reanalysis) and CFSv2 reforecasts sea surface temperature climatologies (Climo.) initialized in June (red color) and December (blue color) in the NINO3.4 region. Numbers in red and blue colors represent lead time from forecast initialization in the respective cases. Climatologies are calculated for the 1982 to 1998 period for both observations and reforecasts. See text for details.

3. Results

Figure 1 shows an example of AAD- and TAAD-type ENSO forecasts (SST anomaly in the NINO 3.4 region) initialized in May 2006. Both AAD- and TAAD-type forecasts have smaller ES than the inter-annual climatological variability (or the climatological forecasts). For the AAD-type forecast, all observations are within the forecast ensemble's 95% range (~ two ES). For the TAAD-type forecast, 4 out of 10 observations are outside the forecast ensemble's 95% range. To investigate the differences between AAD- and TAAD-type forecasts, Figure 2 shows the SST climatology (1982 to 1998) in the NINO3.4 region from CFSv2 Reanalysis and two reforecasts initialized in June and December, respectively. Because of the discontinuity in data after 1998, we are using the larger sample size period (17 years). It is clear that forecast biases or the systematic errors are a function of forecast initialization month and forecast lead time. For example, at 4 month lead, the forecasts initialized in June have a systematic error of 2.01°C, and those initialized in December have a systematic error of 0.57°C. At 0 month lead, the forecast climatology is closer to the observations. Longer lead time forecasts do not necessarily have larger biases. For example, forecasts initialized in June have a bias of 0.78°C at 9 month lead, compared to 2.01°C at 4 month lead. This also indicates that seasonality plays an important role in CFSv2 forecasts. Overall, CFSv2 reforecasts have cold biases in the NINO3.4 region. Forecast biases are also a function of location (not shown).

Figure 3 shows results for the null hypothesis statistical significance test. We calculated AAD and TAAD for each month's initialized forecast at each lead time (0 to 9 months) and then averaged results for forecasts initialized in JJA (June–July–August, Figures 3a to 3h) and DJF (December–January–February, Figures 3c to 3h) at corresponding lead times. In AAD-type forecasts, the null hypothesis is not rejected at all lead times for most of the world (Figure 3a). For example, in forecasts initialized in JJA, the global land average AAD at 0 month lead is 0.83 ± 0.27, and the 1 to 9 month lead average is 0.86 ± 0.27. The error bars/shaded regions in Figures 3c to 3h show 2 standard error estimates of mean. The range for rejecting the null hypothesis is either less than 0.56 or greater than 1.04. Hence, the null hypothesis is not rejected for 0 to 9 month leads. Similar results are also found for DJF, and for the global oceans (Figures 3c and 3e). This result is different from many previous studies, which found that EM forecasts lose considerable skill after the 0 month lead time e.g., *Yuan et al.* [2011], and *Dirmeyer* [2013]; whereas our results show that if we consider full ensemble forecasts, the observations are contained within the 95% range of the ensemble forecasts for 0 to 9 month leads. The difference between this study and previous studies is also noteworthy: here we examine if the observations fall within the range of ensemble forecasts; whereas previous studies use anomaly correlations between the EM forecast and the observations or other similar skill scores. Hence, a direct comparison may not be justified. In the NINO3.4 region, there is a smaller sample size, particularly at 0 month lead, which results in a wider range for failing to reject the null hypothesis (equation (3), Figure 3g). The smaller sample size is due to smaller ES (less than 0.25°C) in most forecasts at 0 month lead in the NINO3.4 region [also see *Peng et al.*, 2011], which are excluded in this analysis.

In TAAD-type forecasts, the null hypothesis is rejected at most lead times for most of the world (Figure 3, right column). Since the null hypothesis is rejected at the higher end of the 95% confidence interval, i.e., $\overline{u} \geq \mu_0 + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{p}}$, this indicates that the observations depart significantly from the EM predictions. For example, in forecasts initialized in JJA, the global land average TAAD at 0 month lead is 1.71 ± 0.41, and the 1 to 9 month lead average is 1.62 ± 0.40. For DJF-initialized forecasts, these values are 1.39 ± 0.38 at 0 month lead and 1.88 ± 0.44 for the 1 to 9 month lead average. Differences between TAAD-type forecasts initialized in JJA and DJF are due to: (1) seasonality in the forecast biases (e.g., Figure 2), which get neutralized in AAD-type

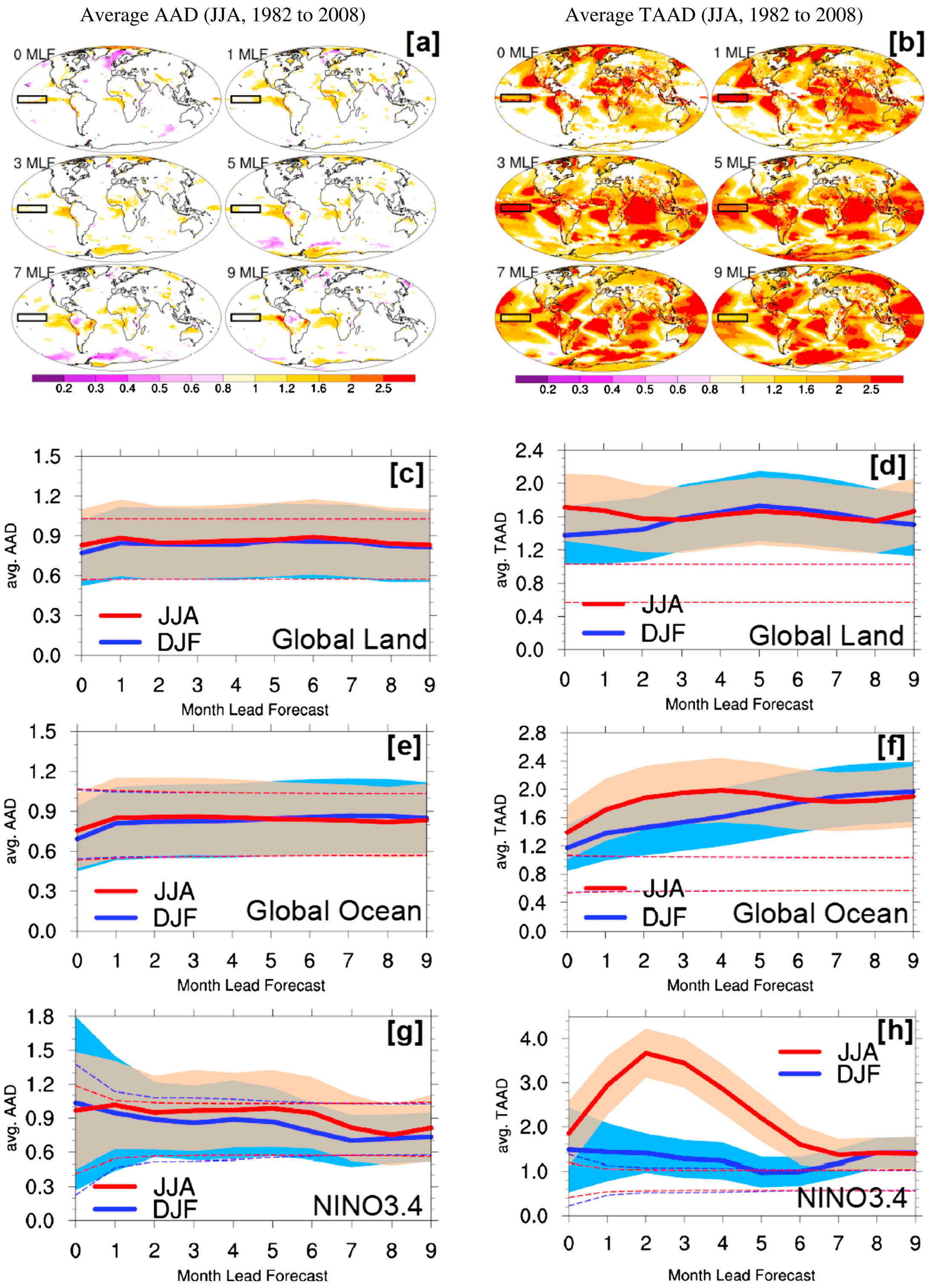

Figure 3. The Null Hypothesis tests results for (left column) AAD and (right column) TAAD forecasts from 1982 to 2008 (see text). In Figures 3a and 3b, JJA average values are shown, and the color shadings are shown only when the null hypothesis is rejected at the 95% confidence interval. Box regions represent the NINO3.4 region. In Figures 3c to 3h, color shadings represent the 95% uncertainty range of estimate in mean. Dashed lines in Figures 3c to 3h represent the 95% confidence interval range for rejecting the null hypothesis. MLF: month lead forecast.

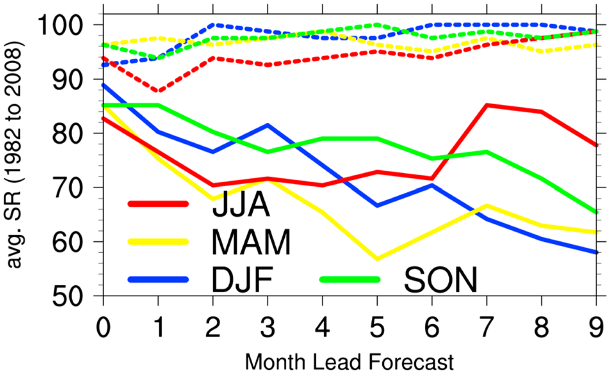

Figure 4. Average success rate (SR) in ensemble mean only ENSO forecasts (solid lines). Dashed lines show the rate at which observations are found within the forecast ensemble 95% range, DJF—December, January, February; MAM—March, April, May; JJA—June, July, August; and SON—September, October, November.

forecasts, and (2) seasonality in the ensemble spread; e.g., the snow-dominated land region has higher ensemble spread during winter compared to the summer ensemble spread irrespective of the lead time (not shown). The effect of seasonality in forecast biases is evident in the NINO3.4 region TAAD-type forecasts (Figure 3h). As shown in Figure 2, forecasts initialized in June have significantly higher biases during 2 to 6 month lead; hence, at these lead times, the null hypothesis is strongly rejected (Figure 3h). On the other hand, forecasts initialized in December have minimal biases at 5 to 7 month lead (Figure 2); consequently, at these leads, the null hypothesis is only weakly rejected or not rejected (Figure 3h). Overall, we argue that removing such known systematic errors, as in equation (1a), is a necessary pre-processing step in using CFSv2 forecasts. Further results are presented only for AAD-type forecasts, i.e., after removing the systematic errors as a function of forecast initialization month and lead time.

Here, we evaluate ENSO forecasts during 1982 to 2008 using EM-only predictions in AAD-type forecasts. Figure 4 shows the average success rate (%) for each season-initialized forecasts, calculated for each month-initialized forecasts and then averaged for 3 month seasons (for example, results from analysis of forecasts initialized in December, January, and February were averaged to produce the results for the DJF mean). A success rate is defined as 1 if the predicted categorical ENSO state (El Niño, La Niña, or neutral) is the same as the observed categorical state in the CFSv2 Reanalysis data; otherwise, it is zero. The forecast success rate decreases with lead time except for forecasts initialized in JJA (Figure 4). The average success rate is 63% for forecasts initialized in winter and spring at 5 to 9 month lead time (longer lead time). Also, *Luo et al.* [2008] have identified long-lead time ENSO predictability. Even in this relatively higher predictability region, the EM-only prediction at longer lead times fails frequently (up to ~40%).

Dashed lines in Figure 4 show the success rate for the ENSO forecasts that consider forecasts within the ensemble's 95% range. A success rate is defined as 1 if the observed SST anomaly falls within the forecast ensemble 95% range; otherwise, it is zero. The success rate remains considerably higher, greater than 95% in most forecasts. The slightly higher success rate at longer lead times (average SR 98%) than the shorter lead times (0 to 4 month lead, average SR 96%) is due to higher ES at longer lead times (shown next). Thus, ENSO observations are encompassed in CFSv2 24-member ensemble forecasts for 0 to 9 month lead times.

Ensemble spread forecasts are useful if the ensemble spread is smaller than the climatological variability (inter-annual variability) and the observations fall within the range of the ensemble forecast (shown earlier for AAD-type forecasts, also see Figure 1a); otherwise, climatology is a better forecast (the null expectation of climate in any region). Here, we test the forecast spread against the climatological variability using an "f-ratio test" Figure 5 shows the average variance ratio of the forecast to the climatological variance from 1982 to 1998. We calculated the variance ratio and applied the statistical test for forecasts initialized in each month and then averaged the results. Although the ensemble forecast variance is not significantly lower than the climatological variance for most of the world except for the tropical oceans at shorter lead times, the variance ratio remains lower than 1 for a considerable fraction of the world. Oceans have relatively higher predictability than the land (variance ratio less than 1, Figures 5c and 5d). In the NINO3.4 region, the ensemble spread remains significantly lower than the climatological variability up to 7 month lead in forecasts initialized in JJA and up to 5 month lead in forecasts initialized in DJF (Figure 5e). The variance ratio remains less than 1 in the NINO3.4 region for all leads. At longer lead times, the fractional areas of land and ocean having a variance ratio less than 1 decrease, i.e., lesser areas have useful ensemble forecasts. At 0 month lead time, 74% of the land area and 78% of the ocean area (average of JJA and DJF forecasts) have a variance ratio less than one. At 9 month lead, the areas of land and ocean having a

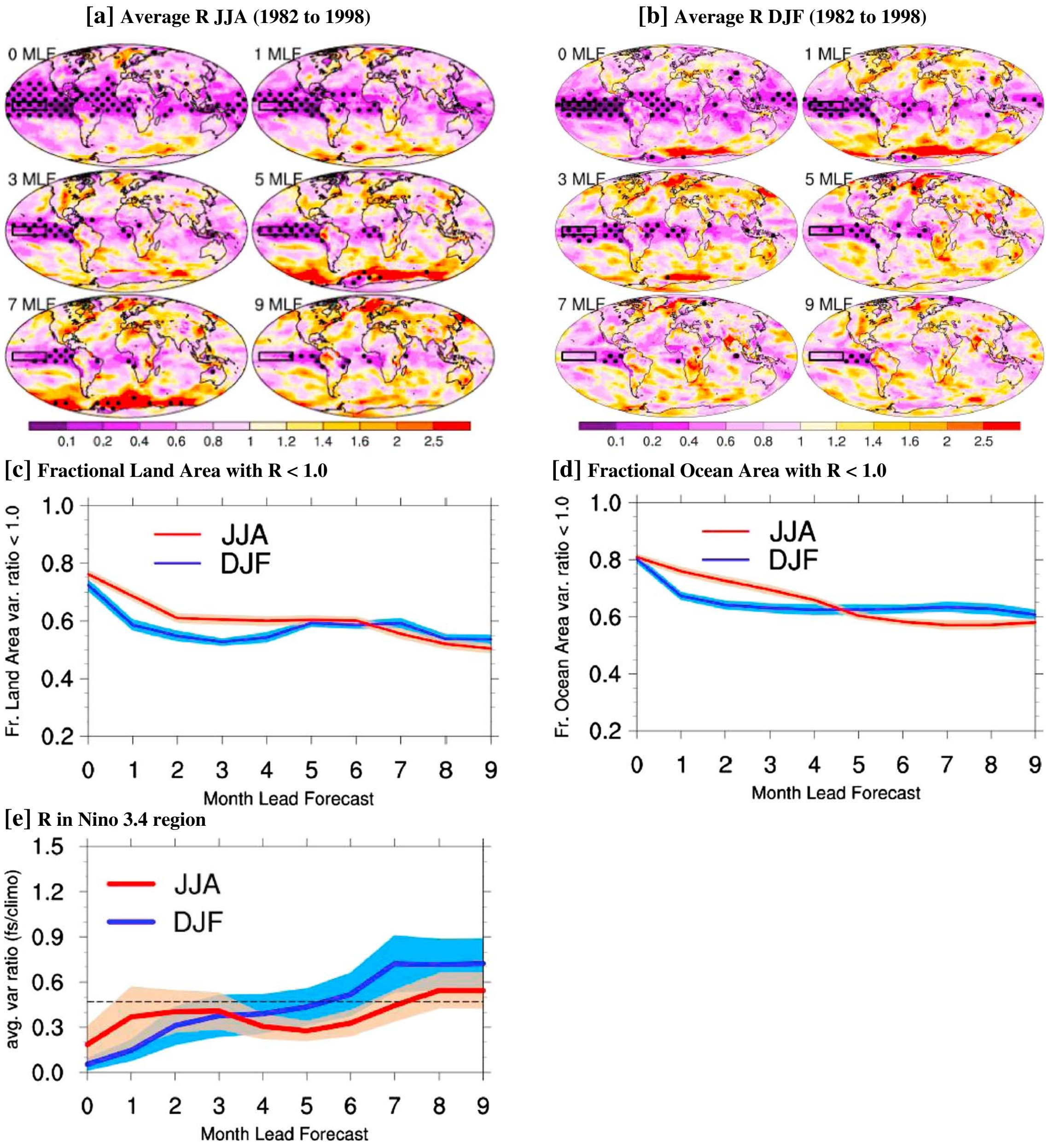

Figure 5. Average variance ratio (forecast/climatology, R) for the 1982 to 1998 period. Variance ratio is calculated for each month's initialized forecast then averaged across different years, and JJA and DJF months. Stippling in Figures 5a and 5b show statistically significant results if in 80% or greater times out of 17 years, results are found statistically significant using "f-ratio test" at the 95% confidence interval. Box regions represent the NINO3.4 region. In Figures 5c to 5e, color shadings represent the 95% uncertainty range of estimate in mean. In Figure 5e, dashed line represents the 95% confidence interval range (upper limit not shown) for the forecast variance to be significantly lower than the climatological variance. MLF: month lead forecast.

variance ratio less than one are 52% and 56%, respectively. Inter-annual variability is minimal in fractional areas having a variance ratio less than 1 (Figures 5c and 5d). Overall, these results show that CFSv2 provides useful ensemble forecasts at all lead times for a considerable part of the world.

4. Conclusion and Discussion

This study focused on the application of seasonal climate forecasts for regional climate predictions. Using ENSO as an example, we have shown that CFSv2 can provide useful climate forecasts from 0 to 9 month lead. We have also shown that the ensemble mean only forecasts provide limited information, and ensemble spread is a required component of the forecast. In 27 years of CFSv2 reforecasts data (1982 to 2008), we found that approximately 40% of ensemble mean ENSO forecasts fail at longer lead times, i.e., observed categorical ENSO states are different from the forecasted categorical ENSO states; whereas all ENSO observations were found within the forecast ensemble 95% range, which is narrower than the inter-annual climatological variability at all lead times.

It is important to note that CFSv2 forecasts have systematic errors that are a function of forecast initialization month and forecast lead times. To remove the systematic errors, we used a simple methodology where anomalies are calculated relative to the reforecast climatology from the corresponding initialization month and the corresponding lead time. While this issue may seem trivial, there is a great deal of diversity in the literature. For example, *Dirmeyer* [2013] examined several anomaly calculation methods, e.g., anomalies relative to reanalysis data, anomalies relative to reforecast climatologies (same as this study), and anomalies relative to zero month lead time forecasts. *Peng et al.* [2011] used reforecast climatologies to calculate reforecasts anomalies. Several other studies, e.g., *Yuan et al.* [2011] and *Luo et al.* [2013], are not clear about removing the systematic errors before forecast evaluations. Based on results presented in this study, we would argue that removing the systematic error is a necessary pre-processing step, and debate over anomaly calculation methodologies should be settled (at least for CFSv2), i.e., anomalies should be calculated relative to the reforecast climatology from the corresponding initialization month and the corresponding lead time.

Finally, we would like to comment on the differences between CFSv2 reforecast and operational forecast configurations. The operational forecast configuration is based on 28-member forecast ensembles initialized 4 times daily in the last 7 days. The reforecast configuration, in contrast, is based on 24-member forecast ensembles initialized 4 times daily every 5th day over the last 30 days [*Saha et al.,* 2014]. Whether the operational configuration that samples only the last 7 days of initial states provides enough ensemble spread to encompass the observations remains an open research question as only a few samples are available (since 2011). We would also like to make an argument for having the same CFSv2 reforecast and operational forecasts configurations so that the systematic errors can be appropriately removed from the operational forecasts.

Acknowledgments
This study received support from the following grants: NSF 0947837, NSF 0830068, NOAA NA09OAR4310058, and NASA NNX09AN50G. We thank Jennifer Adams (Center for Ocean-Land-Atmosphere Studies) for archiving and elegantly documenting the CFSv2 reforecasts and CFSv2 Reanalysis data. We also thank the editor and two anonymous reviewers whose comments led to improvements in the manuscript.

The Editor thanks Lifen Luo and an anonymous reviewer for assistance in evaluating this manuscript.

References

Dirmeyer, P. A. (2013), Characteristics of the water cycle and land-atmosphere interactions from a comprehensive reforecast and reanalysis data set: CFSv2, *Clim. Dyn.*, doi:10.1007/s00382-013-1866-x.

Kumar, A., M. Chen, L. Zhang, W. Wang, Y. Xue, C. Wen, L. Marx, and B. Huang (2012), An analysis of the nonstationarity in the bias of sea surface temperature forecasts for the NCEP Climate Forecast System (CFS) version 2, *Mon. Weather Rev.*, *140*, 3003–3016.

Luo, J.-J., S. Masson, S. K. Behera, and T. Yamagata (2008), Extended ENSO predictions using a fully coupled ocean–atmosphere model, *J. Clim.*, *21*, 84–93, doi:10.1175/2007JCLI1412.1.

Luo, L., W. Tang, Z. Lin, and E. F. Wood (2013), Evaluation of summer temperature and precipitation predictions from NCEP CFSv2 retrospective forecast over China, *Clim. Dyn.*, *41*, 2213–2230, doi:10.1007/s00382-013-1927-1.

Miller, I., and M. Miller (2004), *John E. Freund's Mathematical Statistics With Applications*, 7th ed., Prentice-Hall, Englewood Cliffs, N. J.

Mo, K. C., S. Shukla, D. P. Lettenmaier, and L.-C. Chen (2012), Do Climate Forecast System (CFSv2) forecasts improve seasonal soil moisture prediction?, *Geophys. Res. Lett.*, *39*, L23703, doi:10.1029/2012GL053598.

Murphy, J. M. (1988), The impact of ensemble forecasts on predictability, *Q. J. R. Meteorol. Soc.*, *114*, 463–493.

NOAA Climate Prediction Center (2013), ENSO cycle: Recent evolution, current status and predictions, issued on 9 September.

Peng, P., A. Kumar, and W. Wang (2011), An analysis of seasonal predictability in coupled model forecasts, *Clim. Dyn.*, *36*, 637–648, doi:10.1007/s00382-009-0711-8.

Ropelewski, C. F., and M. S. Halpert (1987), Global and regional scale precipitation patterns associated with the El Nino/Southern Oscillation, *Mon. Weather Rev.*, *115*, 1606–1626.

Saha, S., et al. (2010), The NCEP Climate Forecast System Reanalysis, *Bull. Am. Meteorol. Soc.*, *91*, 1015–1057, doi:10.1175/2010BAMS3001.1.

Saha, S., et al. (2014), The NCEP Climate Forecast System Version 2, *J. Clim.*, *27*, 2185–2208, doi:10.1175/JCLI-D-12-00823.1.

Trenberth, K. E. (1997), The definition of EL Nino, *Bull. Am. Meteorol. Soc.*, *78*(12), 2771–2777.

Whitaker, J. S., and A. F. Loughe (1998), The relationship between ensemble spread and ensemble mean skill, *Mon. Weather Rev.*, *126*, 3292–3302.

World Meteorological Organization (2012), *Guidelines on Ensemble Prediction Systems and Forecasting*, No. 1091, 32 pp., World Meteorological Organization, Geneva, Switzerland.

Xue, Y., M. Chen, A. Kumar, Z.-Z. Hu, and W. Wang (2013), Prediction Skill and Bias of Tropical Pacific Sea Surface Temperatures in the NCEP Climate Forecast System Version 2, *J. Clim.*, *26*, 5358–5378.

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, doi:10.1029/2011GL047792.

Zhang, L., A. Kumar, and W. Wang (2012), Influence of changes in observations on precipitation: A case study for the Climate Forecast System Reanalysis (CFSR), *J. Geophys. Res.*, *117*, D08105, doi:10.1029/2011JD017347.