

Does the NMME Capture a Recent Decadal Shift toward Increasing Drought Occurrence in the Southwestern United States?

ANTHONY G. BARNSTON AND BRADFIELD LYON

International Research Institute for Climate and Society, The Earth Institute, Columbia University, Palisades, New York

(Manuscript received 29 April 2015, in final form 30 September 2015)

ABSTRACT

A global-scale decadal climate shift, beginning in 1998/99 and enduring through 2013, has been documented in recent studies, with associated precipitation shifts in key regions throughout the world. These precipitation shifts are most easily detected during March–May when ENSO effects are weak. Analyses have linked this climate shift to a shift in the Pacific decadal variability (PDV) pattern to its negative phase. Here the authors evaluate the predictive skill of the North American Multimodel Ensemble (NMME), and the CFSv2 model alone, in maintaining the observed precipitation shifts in seasonal forecasts, emphasizing the southwestern United States where deficient precipitation has tended to prevail since the late 1990s.

The NMME hindcasts out to 6 months lead are found to maintain the observed decadal precipitation shifts in key locations qualitatively correctly, but with increasingly underestimated amplitude with increasing lead time. This finding holds in the separate CFSv2 model hindcasts. The decadal precipitation shift is relatively well reproduced in the southwestern United States. The general underestimation of the precipitation shift is suggested to be related to a muted reproduction of the observed shift in Pacific sea surface temperature (SST). This conclusion is supported by runs from a different (but overlapping) set of atmospheric models, which when forced with observed SST reproduce the decadal shifts quite well. Overall, the capability of the NMME model hindcasts to reflect the observed decadal rainfall pattern shift, but with weakened amplitude (especially at longer leads), underscores the broader challenge of retaining decadal signals in predictions of droughts and pluvials at seasonal-to-interannual time scales.

1. Introduction

In recent years there has been a substantial effort focused on evaluating the skill of initialized coupled climate model decadal climate predictions. Although evidence exists for some skill in model predictions of decadal variability in sea surface temperature (SST) (e.g., Meehl et al. 2014), preliminary assessments (e.g., Goddard et al. 2013) show that the skill of decadal precipitation forecasts is generally quite low over most land areas of the globe. Among other factors, a fundamental challenge to skillful decadal prediction is the ability of climate models to capture shifts in low-frequency behavior of the climate system, such as that associated with multidecadal variability of SSTs in the Pacific and Atlantic Oceans. In this paper the focus is on

Pacific decadal variability (PDV) and an observed shift from its warm to cool phase in 1998/99 (Dai 2013; Ding et al. 2013; Lyon et al. 2013). Here, PDV is defined in a generic sense, not using a specific definition such as the North Pacific index used in Trenberth and Hurrell (1994), the Pacific interdecadal oscillation presented in Mantua et al. (1997), the interdecadal Pacific oscillation used in Dai (2013), or the Pacific decadal time series developed in Lyon et al. (2013). The main goal of the study is to discern the extent to which the coupled climate models of the North American Multimodel Ensemble (NMME), in hindcasts of seasonal means extending up to 6 months into the future, are able to reproduce observed precipitation changes associated with this known decadal shift in Pacific SSTs and to investigate some of the factors (e.g., related to SSTs) if it does not.

Although the PDV patterns in SST and their teleconnected climate anomalies are real, some work has suggested that quasi-decadal shifts in the PDV may be the result of several dynamical processes superimposed, mainly coming from the tropical Pacific (i.e., ENSO;

Corresponding author address: Anthony G. Barnston, International Research Institute for Climate and Society, The Earth Institute, Columbia University, 61 Route 9W, Palisades, NY 10964.
E-mail: tonyb@iri.columbia.edu

e.g., Newman 2007). This view of PDV implies that it is not a climate mode in its own right and, because of the lack of inherent predictability of one or more of its components, may itself have little predictability and may vary on multiple time scales, interdecadal being just one of them.

Interest in PDV arises from associated observed variations in seasonal climate in various regions of the globe. For example, the cold phase of PDV has been associated with an increased occurrence of drought in the western United States (Dai 2013; Hidalgo and Dracup 2003), southern China (Chan and Zhou 2005), central and southwestern Asia (Hoell et al. 2015; Lyon et al. 2013), and eastern Africa (Lyon 2014; Yang et al. 2014). Such tendencies have substantial implications for water resource management (e.g., Allen et al. 2014) and the energy and agricultural sectors (e.g., Swetnam and Betancourt 2010). The recent hiatus in the upward global warming trend has also been linked in some studies to the PDV (e.g., Kosaka and Xie 2013).

A decadal-scale shift in Pacific SSTs and associated atmospheric conditions in 1998/99 have been documented in several recent studies (Dai 2013; Lyon et al. 2013; Trenberth and Fasullo 2013; Lin 2014, among others). For the March–May (MAM) season, an analysis of precipitation variability over the last several decades indicates that recurrent drought events since 1999 in East Africa, central-southwest Asia, parts of eastern Australia, and the southwestern United States are all regional climate variations associated with this global-scale multidecadal pattern (Lyon et al. 2013). Using simulations from atmospheric general circulation models forced with observed global SSTs, Lyon et al. (2013) and Dai (2013) further showed that many of the main precipitation and atmospheric circulation features associated with the observed shift are captured. In addition, model runs forced only with observed SSTs in the tropical Pacific were also able to capture many of the observed atmospheric changes, pointing to its central role in driving the regional and global atmospheric responses on the multidecadal time scale.

In the last two years, a North American Multimodel Ensemble has been developed for improved operational prediction of seasonal climate (Kirtman et al. 2014). By late 2014, the NMME included eight coupled general circulation models and had become a valued and routinely used input for the Climate Prediction Center's seasonal forecast production each month (Kirtman et al. 2014). Hindcasts from sets of ensembles from each of the NMME models are available over a 32-yr period (1982–2013). Here we examine these hindcasts to determine the extent to which the NMME detected the observed changes in seasonal precipitation following the decadal

shift in Pacific SSTs in 1998/99. In particular, we look for evidence of reproduction of the drought-like conditions in the southwestern United States, among other regions, considering the December–February (DJF) season as well as MAM.

The NMME hindcasts for the DJF or MAM seasons discussed here are made for up to only several months in advance, so they are not predictions of the decadal shift from the years preceding the shift. In this sense the examination is of seasonal prediction skill during two multidecadal periods having different climatological average SSTs and precipitation. Rather than looking for the hindcast skill for interannual variations such as those related to ENSO, we look for model skill in reproducing the observed mean differences between the two periods, perhaps due largely to retention of the mean differences in the initial conditions for the NMME hindcasts. Failure to reproduce these mean differences might manifest itself with hindcasts that drift toward the climatology of the entire 1982–2013 period even at short lead times. It should be noted that there is little or no indication in recent research that coupled models can reproduce observed decadal variability in the oceans, and thus in the atmospheric teleconnections. In fact, some studies have suggested a lack of decadal-scale predictability specifically in the North Pacific in today's coupled models (Branstator et al. 2012; Branstator and Teng 2012; Meehl et al. 2014).

The paper is outlined as follows. Section 2 first describes the model and observational data used in the study and introduces the analysis methods. Section 3 presents results evaluating the skill of the NMME seasonal precipitation forecasts for the full analysis period, 1982–2013. The observed precipitation shift for the pre- and post-1999 periods are then presented and compared with results from the NMME forecasts for these two periods at 0- and 3-month lead times. To help explain the model performance characteristics, errors in the SST field generated by the NMME are examined in section 4 along with the precipitation field from the AMIP runs. Finally, section 5 provides a summary and the main conclusions of the study.

2. Data and methods

a. Data

Here we use a set of six coupled climate models from the NMME project, which all have global 1982–2013 (32 yr) hindcast data available in a common format at the time of this writing. The six models include the 1) NCAR–University of Miami CCSM3, 2) NOAA/NCEP CFSv2, 3) Canadian Meteorological Centre (CMC)

TABLE 1. Basic information for the six models of the NMME used in the study. (Expansions of acronyms are available online at <http://www.ametsoc.org/PubsAcronymList>.)

Model	Expanded model name	No. of ensemble members	Max lead (months)
CMC1-CanCM3	Canadian coupled model 1	10	12
CMC2-CanCM4	Canadian coupled model 2	10	12
COLA-RSMAS CCSM3	COLA-University of Miami-NCAR coupled model	6	12
GFDL CM2pl-aer04	Modified version of GFDL coupled model	10	12
NASA GMAO-062012	Modified version of NASA coupled model	12	9
NCEP CFSv2	NOAA/NCEP coupled model	24	10

model 1 (CMC1), 4) CMC model 2 (CMC2), 5) NOAA/GFDL model, and 6) NASA model (see Table 1 for more information on models). All of the model data were placed onto a 1° latitude–longitude grid at the originating centers. Among the six models, the number of ensemble members varies from 6 to 24, and the maximum lead time varies from 9 to 12 months (Table 1). Most of the analyses here use the combined forecast [i.e., the multimodel ensemble (MME)]. The MME is formed by combining the individual ensemble members of all of the models. Because some models have more members than others, the number of members acts as a model weighting factor; for example, CFSv2 has 4 times as many ensemble members as CCSM3, so it exerts 4 times the weight of CCSM3 in forming the MME forecast. Given that the hindcast differences in average skill are not large among the models (Becker et al. 2014), assigning as much weight to a model with relatively few members as to a model with a large number of members will theoretically diminish the skill of the MME when the model forecasts have similar average skill. We do not attempt to weight models according to their hindcast skill, for reasons described in section 2b. In the analysis, we examine MME forecasts of 3-month mean precipitation for the first 3 months following initialization, and also for the following 3-month period.

For comparison with the NMME coupled model results, model simulations (runs forced with observed, global SSTs) from seven atmospheric models (hereafter referred to as AMIP runs) are utilized. The seven models used are ECHAM4.5; ECHAM5; NCAR Community Atmosphere Model, version 4 (CAM4), and Community

Climate Model, version 3 (CCM3); NASA GEOS-5; and two versions of CFSv2. These AMIP runs are made available by researchers participating in the NOAA Drought Task Force for an investigation of the recent California drought (Seager et al. 2015). Monthly precipitation data from these models for the period 1979–2013 is utilized following regridding to the common format of the NMME. Ideally, the atmospheric versions of the same coupled models comprising the NMME would be used for this purpose, but only an overlapping set of models is available (Table 2). The overlap consists of the NCEP CFSv2 and NASA GEOS-5 atmospheric models, whose atmospheric components are essentially the same as those of their NMME counterparts. The AMIP runs aim to provide a comparative measure of the success of climate models to capture the observed climate shift when forced with observed, versus predicted (i.e., the NMME), SSTs. For a better controlled comparison between the coupled and the AMIP simulations, the two sets of results for the CFSv2 model alone are also examined.

For verification of the precipitation forecasts two global precipitation analyses are utilized. The primary dataset is the monthly merged satellite–gauge precipitation product from the Global Precipitation Climatology Project, version 2.2 (GPCPv2.2; Adler et al. 2003; Huffman and Bolvin 2012). These are monthly data on a 1° latitude–longitude grid with the period 1979–2013 utilized. The GPCP data have the advantage of providing precipitation estimates over ocean areas, thus allowing for a more complete picture of the global shift pattern. The CPC Unified Gauge-Based Analysis of Global Daily Precipitation dataset (Chen et al. 2008)

TABLE 2. Basic information for the seven AMIP models used in the study. (GHG is greenhouse gas.)

Model	Source	Ensemble size	Resolution	SST forcing	Trace gases	Data available
CCM3	LDEO	16	T42L18	Hadley	Fixed	1856–2014
ECHAM4.5	IRI	24	T42L19	ERSST	Fixed	1950–2014
ECHAM5	NOAA/ESRL	20	T159L31	Hurrell	Varying GHGs	1979–2014
GEOS-5	NASA GSFC	12	1° × 1° L72	Hurrell	Varying	1871–2014
ESRL CFSv2	NOAA/ESRL	50	T126L64	Hurrell	Varying CO ₂	1979–2014
NCEP CFSv2	NOAA/CPC	18	T126L64	Hurrell	Varying CO ₂	1957–2014
CAM4	NOAA/ESRL	20	0.94° × 1.25° L26	Hurrell	Varying	1979–2014

is also used, primarily for assessing results over the western United States. These data are on a 1° latitude–longitude grid, which again matches the resolution of the NMME forecast data. Monthly and seasonal averages over the period 1982–2013 are computed from these daily grids. Monthly average SSTs from the Optimum Interpolation SST version 2 dataset (OISSTv2; Reynolds et al. 2002) for the same period are also employed. These data are on a 1° latitude–longitude grid. All data used in the study are obtained via the IRI Data Library (<http://iridl.ldeo.columbia.edu/>).

b. Methods

For assessing the quality of the NMME forecasts, we select the basic verification measure of temporal correlation, which measures discrimination by computing the extent to which the temporal phases of the variability in the observations are represented in the forecasts. We use a standard statistical test of significance of differences between two means (t test) to determine the robustness with which the models reproduce the shift in mean conditions as seen in the observations. The ensemble mean of the NMME forecasts is used for most analyses, in which case the probabilistic aspect of the NMME forecasts is ignored. However, we also examine the position of the observation within the distribution of model ensemble member hindcast to identify cases in which the observation is poorly accounted for in the model. In these analyses the ensemble member distribution is better defined using a Monte Carlo simulation method that greatly increases the effective number of ensemble realizations.

When combining the forecasts of individual models into a MME, we do not attempt to weight models by their hindcast skill. We make this decision because it has been found that when there are no more than moderate apparent skill differences among models, and only about 30 years of hindcast data are available, use of unequal model weighting fails to result in higher cross-validated MME skill than equal weighting (Kharin and Zwiers 2002; Tippett and Barnston 2008; Peña and van den Dool 2008; Barnston et al. 2012; DelSole et al. 2013). The reason given in these studies is that when the model skills vary by the amounts seen here for the NMME models (i.e., not drastically) the skill differences do not exceed differences explainable by sampling variability and hence may not reflect true model quality differences. When weights differ largely as a result of sampling variability, they often yield less skillful forecast results when applied to independent forecasts than when equal weighting is used (Tippett and Barnston 2008). When a particular model shows much lower skill than that of most of the other models, that model may be

removed entirely from the model set by subjective decision. Such action was not considered in our case, as the model showing lowest average skill over all seasons/leads is still found to contribute to the multimodel forecast skill during some seasons and lead times.¹

In this study, cross-validation is not used in assessing the hindcast skills of various methodological configurations. The main reason is that we are examining a decadal shift and have only two mean states (pre-1999 and 1999 onward; an earlier shift in the 1970s is before the beginning of the NMME archive). On this decadal time scale, and with only about three decades to use, cross-validation has little meaning and is thus ineffective in design.

As an objective method for identifying change points in precipitation time series, a variant of the standard normal homogeneity test (SNHT; Alexandersson 1986) as discussed in Haimberger (2007) is used. While this and related tests are frequently used to identify non-climate-related inhomogeneities in time series, here it is used to identify shifts associated with PDV. The SNHT is applied using a 21-yr moving window to the full time series for a particular season, generating a test statistic for each center value (year) of a given 21-yr period. This test statistic is evaluated for statistical significance by comparing its value to that obtained when applying the SNHT to a time series containing 30 000 years of synthetic data, having the same variance as the observed data.

3. Results

a. Skill of NMME precipitation forecasts during DJF and MAM seasons

Before focusing on the NMME's skill in representing the decadal shift, we briefly examine the ability of the NMME seasonal forecasts to reproduce the observed variability at all time scales (interannual, decadal, and slower trend) over the globe during 1982–2013. We focus on DJF and MAM because an observed decadal shift beginning in 1998/99 is exhibited most clearly during both the boreal winter (e.g., Dai 2013) and spring (Lyon et al. 2013) seasons.

Figure 1 (top) shows the temporal correlation skill for precipitation predictions at lead 0 for DJF (i.e., seasonal forecasts made at the beginning of December). Relatively high NMME forecast correlation skill is noted in eastern equatorial Africa, the Philippines region, northern Brazil, southeastern Australia, and the southern and western

¹ The Climate Prediction Center likewise does not weight the NMME models by their hindcast skill in forming the multimodel forecast average.

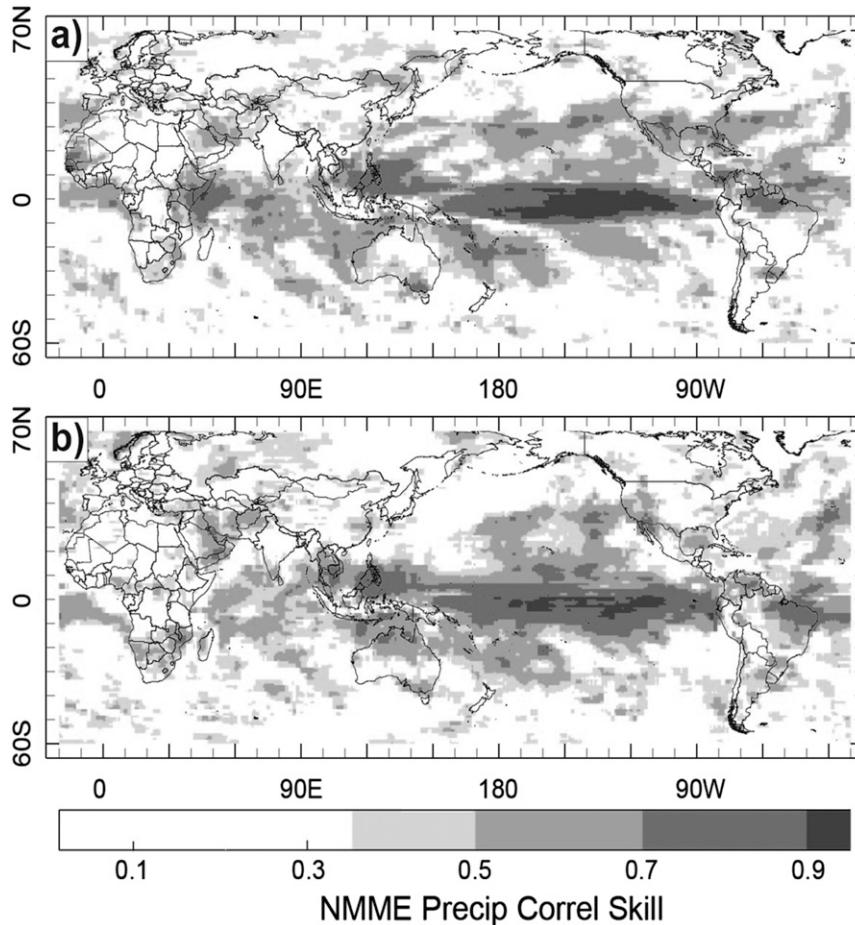


FIG. 1. Correlation (1982–2013) between NMME precipitation forecasts and observations for the (a) DJF and (b) MAM seasons at 0-month lead time (i.e., forecasts were initialized in early December and early March, respectively).

United States. The pattern of highest skill for MAM (Fig. 1, bottom) is somewhat similar, but the highest values are seen in northeastern Brazil, differing parts of Australia, part of southeast Asia, and again in part of the western United States.

In some regions, much of the skill shown in Fig. 1 may be coming from skillful predictions of high-frequency interannual variability (e.g., variability associated with ENSO, especially for DJF when ENSO teleconnections are strong) rather than from a correct reproduction of a decadal shift. To help determine whether the NMME shows a correct mean shift in its forecasts around 1999, we examine maps of the difference in the mean forecast during 1999–2013 compared with 1982–98, and compare the resulting patterns with those in the observed precipitation. Figures 2a and 2b show the difference in means in the observations for DJF (left) and MAM (right) seasons. Figures 2c,d and 2e,f show the difference in means in the NMME predictions

at 0- and 3-month lead times,² respectively, with predictions for DJF (MAM) again shown in the left (right) column. In the maps for observations and 0-month-lead forecasts, the red lines show statistical significance at the 90% and 99% confidence levels based on a *t* statistic, where dashed (solid) lines indicate a negative (positive) difference.

The difference map for DJF observed precipitation (Fig. 2a) shows a shift toward less precipitation in the eastern and central equatorial Pacific, extending westward to about 160°E. Lower precipitation is also noted in Central America and the southern tier of the United States, in eastern equatorial Africa, and in

² A 0-month lead time forecast for DJF, for example, is made at the beginning of December using observed data through the end of November. A 3-month-lead forecast for DJF is made at the beginning of September.

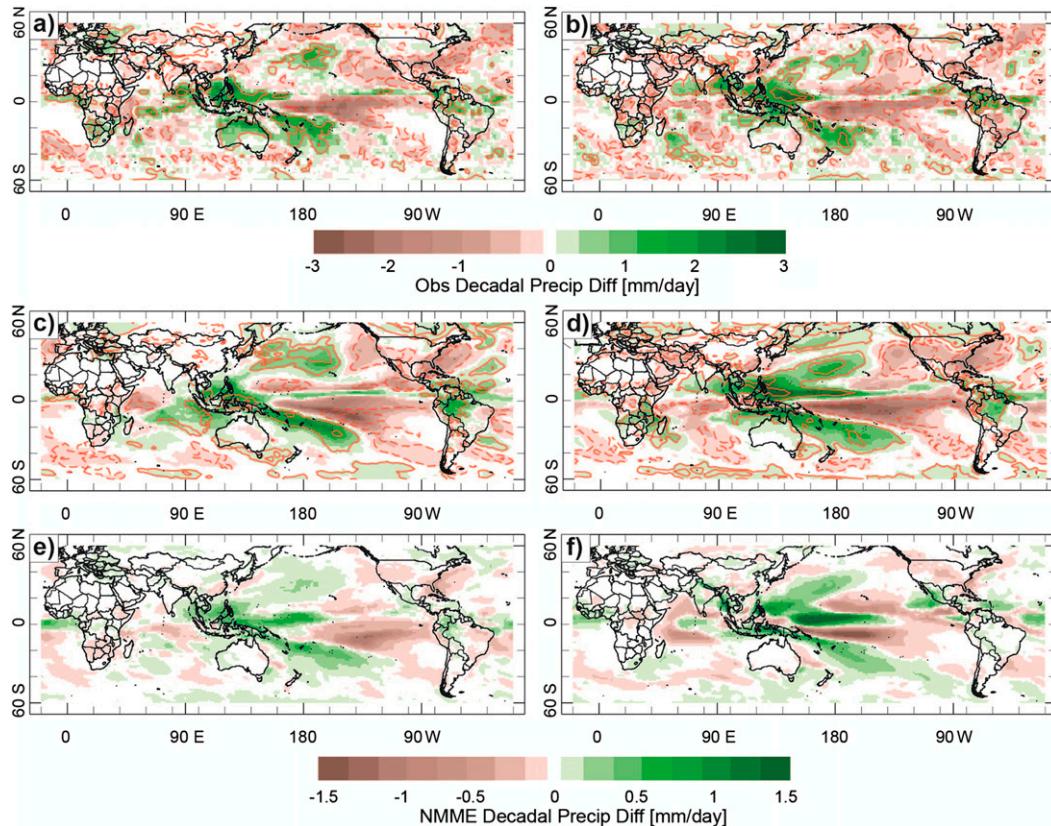


FIG. 2. (top) Difference in seasonal mean observed precipitation (mm day^{-1}) between averages for 1999–2013 and 1982–98 for (a) DJF and (b) MAM, using NASA’s GPCPv2.2. (middle) As at top, but for NMM forecast precipitation (mm day^{-1}) for (c) DJF and (d) MAM. Forecasts for both seasons made at 0-month lead time. Note the change in scale for the observed vs forecast difference. (bottom) As at top, but for forecasts made at 3-month lead time (3 months before the beginning of the first month of the predicted season) for (e) DJF and (f) MAM. In (a)–(d), the red lines show statistical significance at the 90% and 99% confidence levels based on a t statistic (dashed lines for negative t , solid for positive t).

parts of southern South America. A shift toward greater precipitation is seen in northern South America, most of Indonesia and the Philippines, northern Australia, southernmost Southeast Asia, and the southwest tropical Pacific islands. All of these differences are statistically significant at least at the 90% level, and some at the 99% level. The overall pattern somewhat resembles the teleconnection pattern associated with La Niña (Ropelewski and Halpert 1989).

For the MAM season, the difference map for observed precipitation (top right panel of Fig. 2) is similar to that for DJF in several respects and approximately replicates findings in Lyon et al. (2013, see their Fig. 1).³

³Differences from results of Lyon et al. (2013) may be due to slightly different study periods, and the fact that they use GPCP observed precipitation data (gauge data only), while here the NASA GPCP precipitation data (gauges and satellite) are used.

The large negative difference in the eastern and central tropical Pacific Ocean and the opposing positive difference over the Maritime Continent and far western tropical Pacific resembles the comparable features for DJF, except that the latter feature is located slightly farther north in MAM due to the seasonal migration of the intertropical convergence zone. Hence, a wetter MAM season is observed in Southeast Asia and southern India during the post-1999 period, which covers a smaller proportion of northern Australia as compared with DJF. Similar to DJF, shifts toward greater precipitation are noted in southern Africa and northern South America, while decreased precipitation is again found in eastern equatorial Africa, parts of southern South America, and the southern tier of the United States. However, in MAM, unlike DJF, a shift toward drier (wetter) conditions is seen in central and southwestern Asia (central and eastern Russia). The regions mentioned above all have statistically significant differences.

TABLE 3. Spatial correlation (COR) between observed field of mean precipitation difference between 1982–98 and 1999–2013 and the difference shown in the NMME forecasts, for the globe and for land areas only. Results are shown at 0-, 3-, and 6-month lead times for the DJF and MAM target seasons as well as the two in-between seasons. Correlations of 0.5 or more, significant at the 0.05 level (two sided) for 17 spatial degrees of freedom, are shown in boldface.

COR	DJF		JFM		FMA		MAM	
	Globe	Land	Globe	Land	Globe	Land	Globe	Land
0 month	0.65	0.50	0.71	0.51	0.60	0.60	0.67	0.60
3 month	0.46	0.34	0.55	0.29	0.44	0.19	0.35	0.16
6 month	0.29	0.17	0.29	0.00	0.26	0.07	0.17	0.24

To what extent does the NMME reproduce this decadal shift in the DJF season? For 0-month-lead forecasts (Fig. 2c), statistically significant results appear similar to their observational counterparts in many regions, but differ in a few. A decrease in precipitation in the eastern and central equatorial Pacific is common to both observations and forecasts, but a narrow break in the band of drying just north of the equator (5°–10°N) appears only in the forecasts. A drying over the southern tier of the United States and in eastern equatorial Africa is also replicated in the NMME forecasts, as is increased rainfall in northern South America, parts of Indonesia, and the Philippines. Noteworthy differences include a lack of increased rainfall in southern Africa and much of Australia. In regions showing good model reproduction, the decadal differences tend to be somewhat weaker than those in the observations (note the differing color scales in Fig. 2).

The mean shift in the NMME precipitation forecast for DJF at 3-month lead time (Fig. 2e) shows a roughly similar significant shift pattern, but with further degradation of the strength of the difference field, resulting in an inadequate reproduction of the precipitation decrease in eastern equatorial Africa, and a reversal of the observed increase over southern Africa.

In MAM, the NMME reproduces the observed decadal shifts for 0-month-lead forecasts to a greater extent than found in DJF (Fig. 2d). Most of the main regions of statistically significant shift in the observations are seen also in the forecasts, to first order, although the amplitude of the shift is less in the forecasts (again, note the differing color scales in Fig. 2). The narrow break in the west–east band of shift toward lower precipitation in the tropical Pacific is noted in the observations as well as the forecasts. However, the forecast drying responses over eastern equatorial Africa are weak, as are those in southern South America. The forecast for MAM season at 3-month lead time (Fig. 2f) shows a shift pattern with some resemblance to

TABLE 4. Standard deviation (SD; mm day⁻¹) over space, over the globe and for land areas only, of (top) the observed mean precipitation difference between 1982–98 and 1999–2013, and (remaining rows) the difference shown in the NMME forecasts at 0-, 3-, and 6-month lead times for the DJF and MAM target seasons as well as the two in-between seasons.

SD	DJF		JFM		FMA		MAM	
	Globe	Land	Globe	Land	Globe	Land	Globe	Land
Obs	0.49	0.16	0.50	0.18	0.48	0.18	0.49	0.19
0 month	0.25	0.05	0.39	0.10	0.37	0.10	0.31	0.11
3 month	0.15	0.05	0.23	0.10	0.24	0.11	0.24	0.09
6 month	0.18	0.08	0.16	0.08	0.16	0.07	0.16	0.06

that of 0-month-lead forecasts, but with further weakening of the difference field and degradation of the correspondence to the observed difference field. The basic pattern of shift in the tropical Pacific is itself poorly represented.

Table 3 shows the spatial correlation, for the entire globe and for land areas only, between observed and forecast differences at 0-, 3-, and 6-month lead times for the DJF and MAM seasons along with results for January–March (JFM) and February–April (FMA).⁴ The NMME forecasts for seasons at other times of the year (not shown) are found to have lower pattern correlation skill. Contributions to the correlation are weighted by the cosine of the latitude for an equal-area calculation. During DJF and MAM (and the two intermediate seasons) correlations for 0-month-lead forecasts exceed 0.5, which is statistically significant (at 0.05 level, two sided) if we assume at least 17 spatial degrees of freedom—a very conservative estimate for global precipitation, based on van den Dool and Chervin (1986). For 3-month-lead forecasts the correlations become weaker, and at 6-month lead time they are useless for practical purposes. Correlations for the globe tend to be stronger than those over land only, likely due to the well-forecast precipitation difference over the tropical Pacific associated with the SST shift underlying this decadal variation. Based on these correlations, we conclude that short-lead NMME forecasts capture the pattern of the observed decadal shift reasonably well. As forecast lead time increases, maintenance of the observed decadal signal decays.

Table 4 shows, for DJF through MAM, the amplitude of the difference fields as represented by the spatial standard deviation of the differences over the globe and for land areas only. The model forecast amplitude is

⁴ Note that correlations are affected not by the overall weakness of the forecast decadal differences, but only by the geographical phasing of the spatial patterns.

TABLE 5. Percentage of area over which the mean precipitation difference between 1982–98 and 1999–2013 precipitation is locally statistically significant at the two-sided 5% level, over the globe and for land areas only. Results are given for observations, and in the NMME forecasts at 0-, 3-, and 6-month lead times for the DJF and MAM target seasons as well as the two in-between seasons. The Student's t test is used to test local significance of the period difference at each grid point.

%	DJF		JFM		FMA		MAM	
	Globe	Land	Globe	Land	Globe	Land	Globe	Land
Obs	15	19	20	22	20	21	22	23
0 month	13	8	22	20	21	22	24	28
3 month	17	17	13	14	13	11	14	10
6 month	16	17	15	18	17	16	15	14

calculated without regard to the correctness of the spatial phasing of the difference fields. Because the primary source of the decadal signal is in the tropical Pacific Ocean, the amplitude of both observations and forecasts over land is generally less than half of that over the entire globe. For the globe, the amplitude of the forecasts at 0-month lead time is about half to three-quarters that of the observations, indicating that the decadal climate shift starting in 1999 is underrepresented in the NMME forecasts. At 3- and 6-month lead times, the forecast amplitude continues to decrease. Over land only, the ratio of the forecast to observed amplitude averages just slightly more than one-half for 0-month-lead forecasts.⁵

One reasonably might want to know the degree to which the decadal shift shown here and in recent studies is statistically significant. To address this significance issue in both observations and NMME forecasts, a t test is applied to the difference between precipitation for the periods 1982–98 and 1999–2012. Is the mean difference great enough to overcome the higher-frequency variability, such as that related to ENSO fluctuations, and unpredictable random variability? Table 5 shows the percentage area (both for global and for land only) attaining local significance at the 5% level (two sided) for the observations and for the NMME model forecasts for DJF through MAM at 0-, 3-, and 6-month lead times. In the absence of a decadal shift, the percentage of local significance is expected to be 5% due to random variability. The percentage is seen to be roughly 20% in the observations. In the NMME forecasts the percentage of

area with local significance varies between about 10% and 25%, often somewhat less than that of the observations. At 0-month lead time, with the exception of the DJF season for land areas, the percentage of area having local significance is comparable to that observed. The greatest percentages of local significance for both observations and for 0-month-lead NMME forecasts occurs for MAM, suggesting that this season may have a minimum of other sources of variability (e.g., ENSO related), allowing the decadal shift to be more easily distinguished without prefiltering the data for variability at other time scales such as ENSO and climate change.

With the percentages of local significance for observations and NMME forecasts well above the randomly expected 5% level, we conduct field significance tests to assess the probability of attaining these elevated percentages by chance. Seven hundred iterations of independent random shuffling of the year assignments are carried out, and the number of instances in which the percentage of local significance exceeds that found using the correct year assignments is counted. Resulting field significances for the DJF season for the observations for the globe (land only) are 0.006 (0.002), while for the NMME forecasts at 0- and 3-month lead times they are 0.07 (0.21) and 0.05 (0.05), respectively; for MAM, the field significances for the observations are <0.002 (<0.002)—that is, none of the 700 randomly shuffled cases outperformed the actual one—and for the forecasts at short and longer leads they are 0.006 (<0.002) and 0.07 (0.12), respectively. Strong field significance is achieved for the observations in both seasons. The 0-month-lead NMME forecasts are field significant at the 0.05 level for MAM season, but not for DJF. The 3-month-lead forecasts lack field significance at the 0.05 level for MAM, but are minimally significant for DJF. A tendency for a statistically more robust decadal signal over the globe than over land only is not seen in these significance results, suggesting that the statistical robustness of the decadal signal is equally strong over land as over the globe in both observations and predictions.

To evaluate the statistical significance of the *difference* between the decadal shift represented in the NMME hindcasts and that observed, the position of the observed shift within the distribution of shifts described by the individual ensemble members within the NMME is determined. The variation of shifts among ensemble members is thus representing the uncertainty distribution for the significance test. If the observation lies in the far tails of the ensemble distribution, or even completely outside of that distribution, we may conclude that the NMME did not reproduce what was observed (or predicted it with a very low probability). This analysis is carried out for each location for the 0-month-lead hindcasts.

⁵ It should be noted that because the observation is a single realization, while the NMME forecasts are ensemble averages, the latter are expected to have a *slightly* lower spatial standard deviation of difference due to their lesser contribution from noise, despite the averaging over at least 15 years in the pre- and postshift periods.

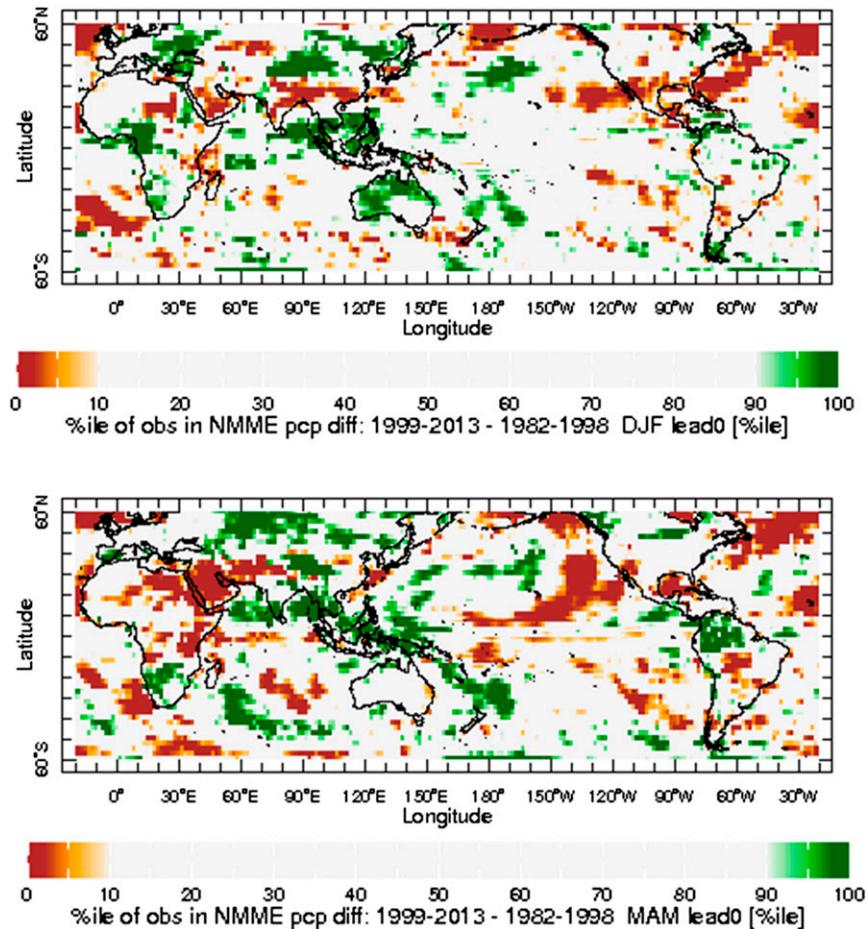


FIG. 3. Percentile of observed precipitation mean difference for 1999–2013 minus 1982–98 within Monte Carlo expanded ensemble distribution of precipitation mean difference in NMME hindcasts at 0-month lead time for (top) DJF and (bottom) MAM.

To expand the uncertainty distribution beyond that provided by the 72 NMME ensemble members, a Monte Carlo randomization scheme is used to shuffle the ensemble member selection separately for each year, providing a multiplicity of ensemble member combinations in each realization; 2000 such iterations are generated. Figure 3 shows the spatial distribution of the percentile of the observations within the expanded NMME distribution for DJF and MAM, highlighting percentiles of less than 10 or more than 90. When the observation shows a negative difference (later period drier than earlier period), and the NMME hindcasts systematically fail to reproduce the precipitation decrease, the percentile of the observation within the model ensemble distribution is on the low tail of the ensemble distribution, and vice versa. In some regions the pattern of high and low percentiles shown in the panels of Fig. 3 appears similar to the pattern of the observed decadal shift (Figs. 2a,b), indicating that the

observation falls on the tails of an ensemble distribution of NMME hindcasts, suggesting that the decadal shift is statistically not well represented. This is particularly the case during MAM, when the “noise” of ENSO variability is relatively weak and can be noted in Indonesia, southern Africa, and northern South America (systematically insufficient wetting starting in 1999), as well as in the Gulf of Mexico, eastern equatorial Africa, and parts of the northern tropical Pacific (insufficient drying). The resemblance of the patterns of percentile extremes in Fig. 3 with the pattern of the observed decadal shift, while sketchy, suggests a systematic underestimation of some key aspects of the shift signal in the NMME hindcasts. However, in some regions the observation is not statistically underestimated by the NMME. For example, for both DJF and MAM the percentile of the observation is not in the lower tail of the CFSv2 ensemble distribution in the southwestern United States, except in the southern portion, indicating that

the decadal difference in the NMME does not statistically differ from the observed drying tendency. This approximate equivalence of NMME with observed difference in the southwestern United States is consistent with the middle versus the top panels of Fig. 2 for both seasons.

In summary, the NMME forecast evaluations suggest that the NMME reproduces the observed pattern of decadal shift in precipitation fairly well, but at a reduced strength. The NMME representation is more strongly present in 0-month-lead than in 3-month-lead forecasts, and is somewhat better in MAM when the decadal shift stands out most clearly relative to higher-frequency variability, than in DJF. Although the models appear to reproduce reality more strongly when ocean areas are included, performance over land, including the United States, is also generally favorable despite some omissions and errors in details. Field significance tests for the robustness of the decadal signal are very strong in the observed data, and significant also in some of the model hindcasts, depending on season and lead time. Tests of the significance of the differences between shifts in the NMME hindcasts at 0-month lead time and those observed, based on the position of the observed shift within the distribution of the NMME ensemble members, suggest systematically weak reproduction of the shift in some of its key regions. This is seen more clearly in MAM than in DJF when the observed shift itself is less prominent. In section 4 we will show how the deficiencies in the model decadal signal may be related to errors in the NMME-generated SST forecasts.

b. Increased drought occurrence in the southwestern United States

The southwestern United States has experienced a protracted precipitation deficiency since the late 1990s; the timing suggests a relationship to PDV, which is consistent with previous investigations (see Dai 2013, and references therein). Here, Figs. 2a and 2b suggest that this decadal pattern is manifest by precipitation deficits during both DJF and MAM. Climatological precipitation in the southwestern United States during May is quite small, so changes in MAM seasonal totals are expected to be dominated by precipitation occurring in March and April. As an objective method to identify decadal precipitation shifts, we apply the SNHT to the area-average precipitation time series for the southwestern U.S. region (30°–40°N, 105°–125°W, land areas only), examining both the CPC unified and GPCP datasets (Fig. 4). Figures 4a,b show seasonal precipitation departures from a 1980–2013 mean for the 6-month December–May period. The horizontal lines represent mean values for periods before and after an identified

change point. The SNHT results indicate that a likely change point occurs in 1998/99 in both datasets (with significance level $p < 0.10$). For DJF (Figs. 4c,d) the SNHT did not identify a statistically significant change point, although the results did indicate an enhanced likelihood of a shift in 1998/99 for both datasets. In MAM (Figs. 4e,f), a statistically significant shift is identified as most likely occurring in 2000 for the CPC unified ($p < 0.05$) and GPCP ($p < 0.10$) data. These results provide evidence for the consistency of the negative phase of PDV being associated with (not necessarily “causing”) an increased occurrence of drought in the southwestern United States (e.g., Newman et al. 2015, manuscript submitted to *J. Climate*).

Figures 5a and 5c show similar time series generated from NMME hindcasts at 0- and 3-month lead times during the DJF seasons from 1982 to 2013; Figs. 5b,d show the same for MAM. Horizontal lines on the graphs again indicate mean values for the periods 1982–98 and 1999–2013. As shown in Fig. 1, the NMME shows skill in seasonal precipitation hindcasts (most at zero lead) in this region in both seasons. In DJF (Figs. 5a,c), a downward shift in precipitation starting in 1999 is reproduced quite well by the NMME at 0-month lead time, and the interannual variability of the NMME matches that of the observations fairly well (note peaks with the El Niño years of 1983 and 1998, and the brief “spring El Niño” of 1993). At 3-month lead time, the forecast decadal shift is weaker than that observed. The correlations between the NMME forecasts and observed time series at 0- and 3-month lead times are 0.56 and 0.52, respectively. In MAM (Figs. 5b,d), good model performance is seen both at the interannual time scale (e.g., good reproduction of precipitation peaks during El Niño endings in 1983, 1992, and 1998 and troughs during La Niña endings in 1989 and 2008) and the decadal time scale where the mean shift in the model approximates that observed. The correlation with the observed time series, reflecting variations at both time scales, is 0.66 for 0-month-lead forecasts—stronger than that of DJF despite the weaker contribution from ENSO. At 3-month lead time the NMME performance is still positive (correlation with observations is 0.54), but the decadal shift is underforecast.

c. Contribution of CFSv2 to NMME forecasts of a decadal shift

It is increasingly believed that forecasts from an ensemble of multiple models, each having its own set of ensemble members, usually results in a more skillful forecast than that from the most skillful individual model (e.g., Kharin and Zwiers 2002; Kirtman et al. 2014). Nonetheless, one might legitimately ask whether

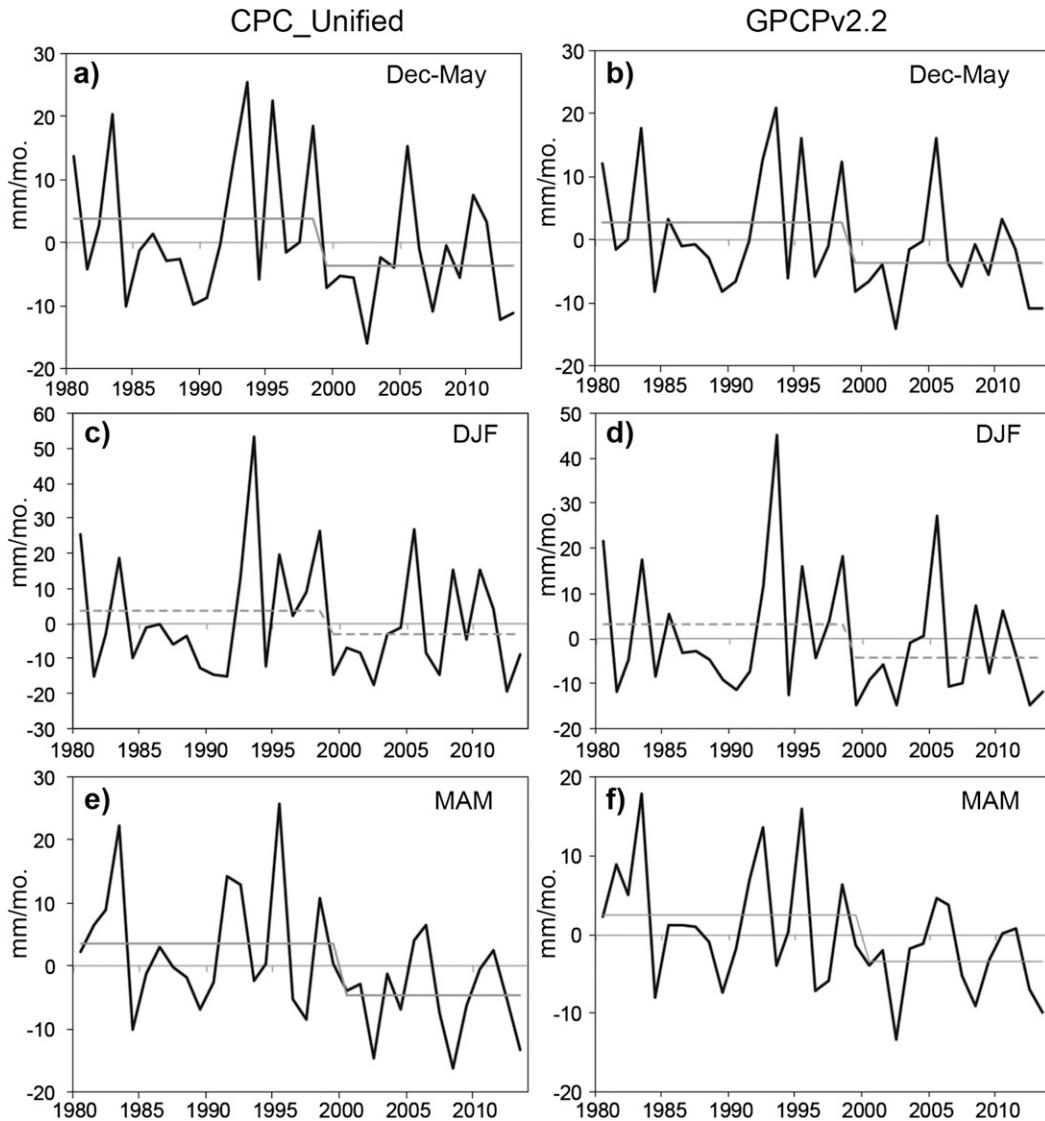


FIG. 4. (top) Observed precipitation anomalies during the December–May season for the southwestern United States (30° – 40° N, 105° – 125° W) during 1982–2013 (mm month^{-1}) for the (a) CPC and (b) GPCP datasets. The mean values for 1982–98 and 1999–2013 are also shown where solid (dashed) lines indicate a statistically significant (possible) changepoint observed based on the SNHT results. (middle) As at top, but for DJF in (c) CPC and (d) GPCP. (bottom) As at top, but for MAM in (e) CPC and (f) GPCP.

there are individual models in the NMME that better reproduce the observed decadal shift than others. Here we limit our investigation to NOAA’s official operational prediction model, CFSv2. Because of its relatively large ensemble size (24 members), CFSv2 exerts 33% of the weight in creating the NMME MME forecasts used here.

The global field of mean difference in DJF precipitation between 1982–98 and 1999–2013 predicted at 0-month lead time by the CFSv2 model is shown in Fig. 6a, to be compared with the same but from the NMME (Fig. 2c) and the observations (Fig. 2a). The

CFSv2-predicted difference pattern and statistical significance pattern (red lines) appear similar to those of the NMME, and in some regions the signal is stronger, including portions of the tropical Pacific and parts of Indonesia. During DJF the CFSv2 shows an “island” of positive difference at and just north of the equator near the date line, which does not appear as strongly for the full NMME (and hardly appears in the observations). Other differences include CFSv2 correctly forecasting a shift toward higher precipitation in southern Africa, and an incorrect forecast toward lower precipitation in northeast Brazil and the adjacent tropical Atlantic

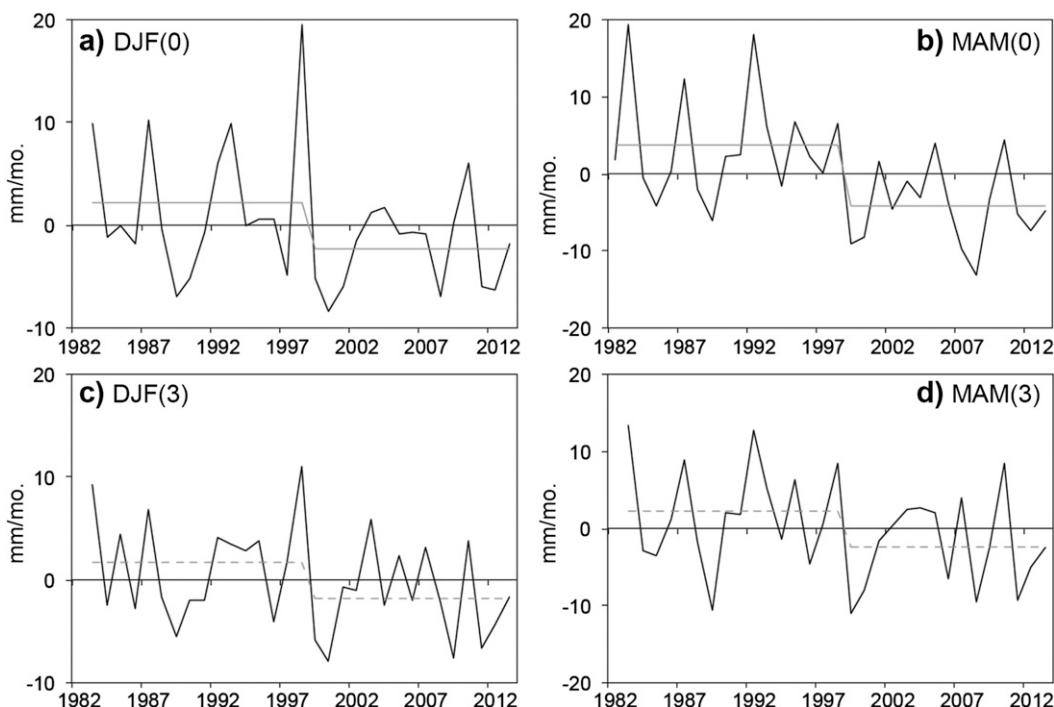


FIG. 5. Anomalies in NMME forecast precipitation for the southwestern United States (30° – 40° N, 105° – 125° W) during 1982–2013. The left (right) column is for the DJF (MAM) season with (a),(b) 0-month and (c),(d) 3-month lead time. The mean values for 1982–98 and 1999–2013 are also shown where solid (dashed) lines indicate a statistically significant (possible) forecast changepoint based on the SNHT results.

region. For the MAM season (Fig. 6b), differences in CFSv2 again mimic those of NMME quite closely, including their statistical significance. CFSv2 again shows slightly higher magnitudes of precipitation difference in the tropical Pacific, including an accentuated positive difference just north of the equator in the central Pacific. Differences from the NMME during MAM include a more realistic prediction by CFSv2 of wetter recent conditions in central Russia, and, similar to DJF, a faulty prediction for recent drying in northeast Brazil and adjacent Atlantic waters.

The overall similarity of CFSv2 with NMME results could reflect the fairly strong role of CFS in shaping the NMME forecast (33% weight); it is not known whether or not it is also a result of a strong commonality among features of the forecasts from most of the NMME models. Table 6 shows the percentage change in mean precipitation total during the DJF, JFM, FMA, and MAM seasons in the southwestern United States (30° – 40° N, 105° – 125° W) from 1982–1998 to 1999–2013 for the observations and for the forecasts of the NMME and the CFSv2 at 0- and 3-month lead times. The observations indicate an approximate 15%–20% precipitation reduction in the region during northern winter and spring seasons; generally, smaller reductions are found during

the remainder of the year (not shown). During the boreal winter and spring seasons, the NMME forecasts at 0-month lead time reproduce the decadal reduction of precipitation in the southwestern United States quite well during FMA and MAM but underestimate it by 25%–50% during DJF and JFM. At 3-month lead time the NMME underestimation of the reduction is more pronounced. The forecasts of the CFSv2 model behave somewhat similarly to those of the NMME but reproduce the precipitation reduction slightly less strongly than the NMME during the MAM. At 3-month lead time, CFSv2 fails to indicate a decadal precipitation reduction. It is possible that the CFS forecasts at 3-month lead time are affected by an issue involving the CFS reanalysis data used to initialize the model; this will be briefly discussed in section 4.

Although the mean differences in precipitation between the 1999–2013 and 1982–98 periods in the CFSv2 hindcasts are smaller than those observed, one might question the statistical significance of the difference between the model and observed decadal signal. As described above in section 3a for the NMME, we determine, for each location, the percentile of the observed decadal signal within the ensemble distribution of the CFSv2 hindcast decadal signal. The uncertainty

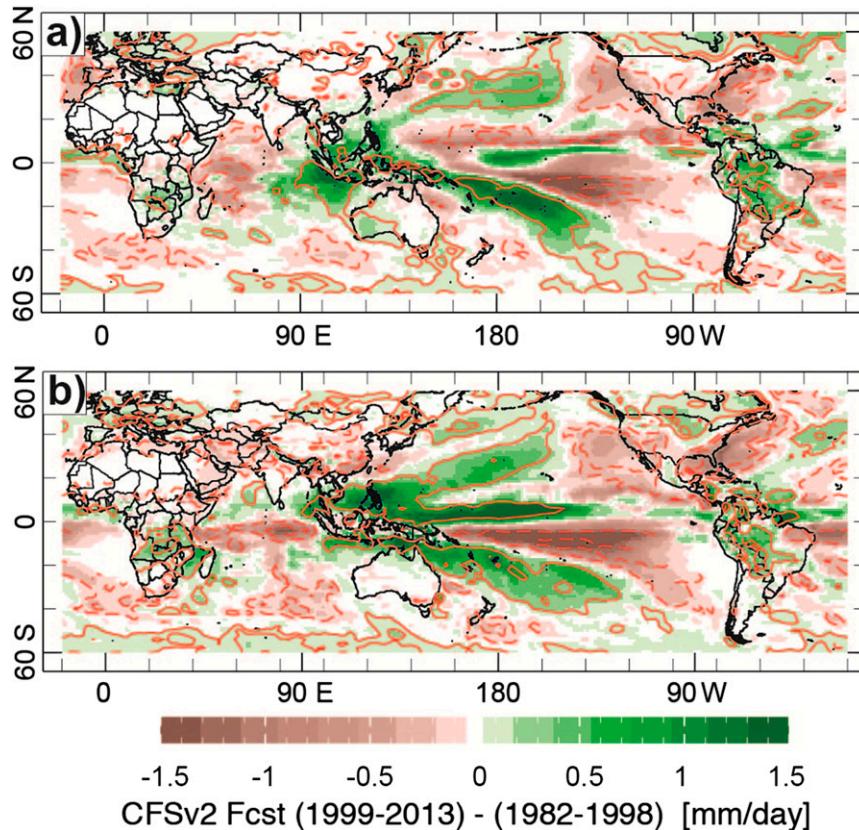


FIG. 6. Difference in seasonal mean CFSv2 forecast precipitation (mm day^{-1}) between averages for 1999–2013 and 1982–98 for (a) DJF and (b) MAM. Red lines show statistical significance at the 90% and 99% confidence levels based on a t statistic (dashed lines for negative t , solid for positive t). Forecasts are made at 0-month lead time.

distribution is again refined beyond what is provided by the 24 ensemble members, using Monte Carlo randomization of ensemble member separately for each year, providing 2000 realizations. Figure 7 shows the spatial distribution of the percentile of the observation within the expanded CFSv2 distribution, highlighting percentiles of less than 10 or more than 90, for a characterization of the CFSv2 model's reproduction of the observation as explained above for full NMME application.

In a number of regions, the pattern of high and low percentiles shown in Fig. 7 appears similar to the pattern of the observed decadal shift (Figs. 2a,b), indicating that the observation falls on the tails of the ensemble distribution of CFSv2 hindcasts. This resemblance to the observed pattern is stronger for MAM (when ENSO variability is weaker) than DJF and can be noted particularly in the central tropical Pacific in both seasons (drying), central South America in MAM (drying), parts of Indonesia, northern Australia, and the southwest subtropical Pacific in both seasons (wetting), and southern Africa and northern South America in

MAM (wetting). The result for CFSv2 alone is stronger than that found for the full NMME, despite the similarity between the CFSv2 and NMME shift signals, likely due to the somewhat narrower distribution of ensemble hindcasts for a single model than for a set of several models. The wider ensemble spread among the NMME comes about through the noticeable “differences of opinion” reflected by differing ensemble means among the NMME models (not shown; noted also in NMME ENSO forecasts;

TABLE 6. Percentage change in mean precipitation total in the southwestern United States (30° – 40° N, 105° – 125° W) from 1982–98 to 1999–2013, for the four running 3-month seasons from DJF to MAM, for the observations and for the forecasts of the NMME and the CFSv2 model. Model forecast results are shown at 0- and 3-month lead times.

	DJF	JFM	FMA	MAM
Observations	-14	-20	-14	-19
NMME, 0 month	-8	-14	-19	-20
NMME, 3 month	-5	-5	-5	-7
CFSv2, 0 month	-10	-13	-15	-16
CFSv2, 3 month	-1	3	0	-2

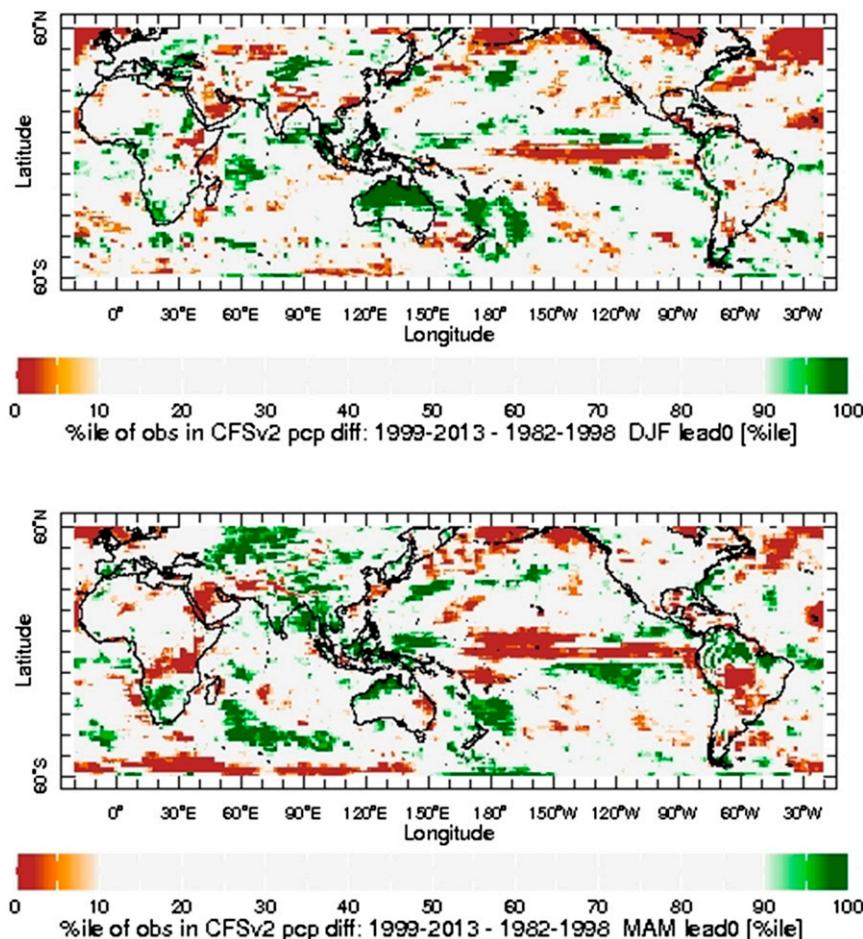


FIG. 7. Percentile of observed precipitation mean difference for 1999–2013 minus 1982–98 within Monte Carlo expanded ensemble distribution of precipitation mean difference in CFSv2 hindcasts at 0-month lead time for (top) DJF and (bottom) MAM.

Barnston et al. 2015). As found for the full NMME, in both DJF and MAM the percentile of the observation is not in the lower tail of the CFSv2 ensemble distribution in most of the southwestern United States, indicating that the decadal difference in CFSv2 is statistically representative of the observed drying tendency, as seen also in comparing Figs. 2c,d and the corresponding Figs. 2a,b.

4. Role of errors in the SST field generated by the NMME and CFSv2 alone

In coupled ocean–atmosphere models such as those of the NMME, SST predictions are developed simultaneously with atmospheric predictions, allowing for feedbacks between the two. While full coupling allows for potentially more realistic predictions, the models inevitably contain biases, including possible systematic errors in the SST predictions. Because of the strong role

SST plays in forcing the atmospheric circulation and surface climate anomalies, having realistic SST anomaly patterns is essential to the quality of the atmospheric predictions. Simulations of atmospheric general circulation models (AGCMs) using simultaneous observed SST anomalies as the lower boundary conditions (so-called AMIP simulations; Gates 1992) provide more skillful atmospheric predictions than the same simulations using predicted (imperfect) SSTs (Li et al. 2008, among others). In the case of the NMME, we probe whether using observed simultaneous SST as the lower boundary condition results in improved reproduction of the decadal precipitation shift compared with that shown here using fully coupled models. Such improvement would be particularly expected in the case of the decadal shift around 1998/99, given that a shift in the mean observed Pacific SST at that time was shown to have a fundamental linkage to the shift in the atmospheric anomaly pattern (e.g., Lyon et al. 2013).

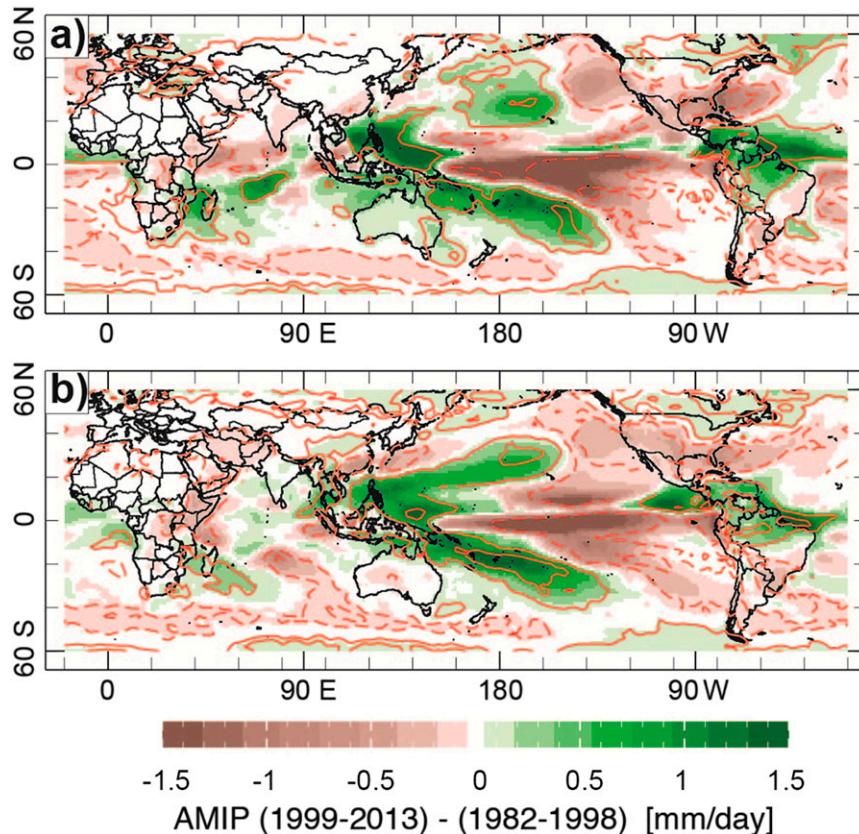


FIG. 8. Difference in seasonal mean observed precipitation (mm day^{-1}) between averages for 1999–2013 and 1982–98 for the multimodel mean AMIP runs for (a) DJF and (b) MAM.

Ideally, the atmospheric versions of the coupled models making up the NMME would be used for this “perfect SST” comparison. However, they are not available. As an alternative, we use simulations of a largely different set of seven atmospheric models (Table 2). However, two versions of the CFSv2 are used, one of which has essentially the same atmospheric model as the NCEP version used in the NMME, and the NASA GSFC atmospheric model (GEOS-5) is highly similar to that used in the NASA GMAO model of the NMME. The AMIP model simulations, made available courtesy of the NOAA Drought Task Force (Seager et al. 2015), provide output data during 1979–2013, converted to a common grid. These AMIP simulations are used here to provide a comparative measure of the ability of climate models to capture the observed decadal shift when forced with observed, in contrast to predicted, SST.

Figure 8 shows the difference between the AMIP multimodel simulated mean precipitation for the period 1999–2013 minus 1980–98 for the DJF and MAM seasons. These difference patterns, and their statistical significance, reasonably match those found in the observations (Fig. 2,

top), and more closely so than those for the NMME predictions at the shortest lead (Fig. 2, middle), including the primary shift region across the tropical and subtropical Pacific and Maritime Continent. An application of the SNHT to the simulated southwestern U.S. (30° – 40° N, 105° – 125° W) precipitation indices shows statistically significant change points for DJF ($p < 0.08$) and MAM ($p < 0.01$) occurring in 1998/99. This model finding helps confirm not only the reality of the 1998/99 decadal shift, but that the atmospheric component of the AMIP models is able to simulate the change in precipitation in the southwestern United States, given the correct change in the forcing from SSTs.

It was noted above that the decadal precipitation difference in CFSv2 alone (Fig. 6) is positive along (and just north of) the equator near the date line, particularly for DJF, while this feature does not appear in the multimodel AMIP result (nor in the observations). This feature is in a key location regarding atmospheric teleconnections and thus deserves further scrutiny. For a better controlled comparison between the coupled and AMIP precipitation hindcasts, the AMIP simulations are generated for the CFSv2 model alone. The resulting

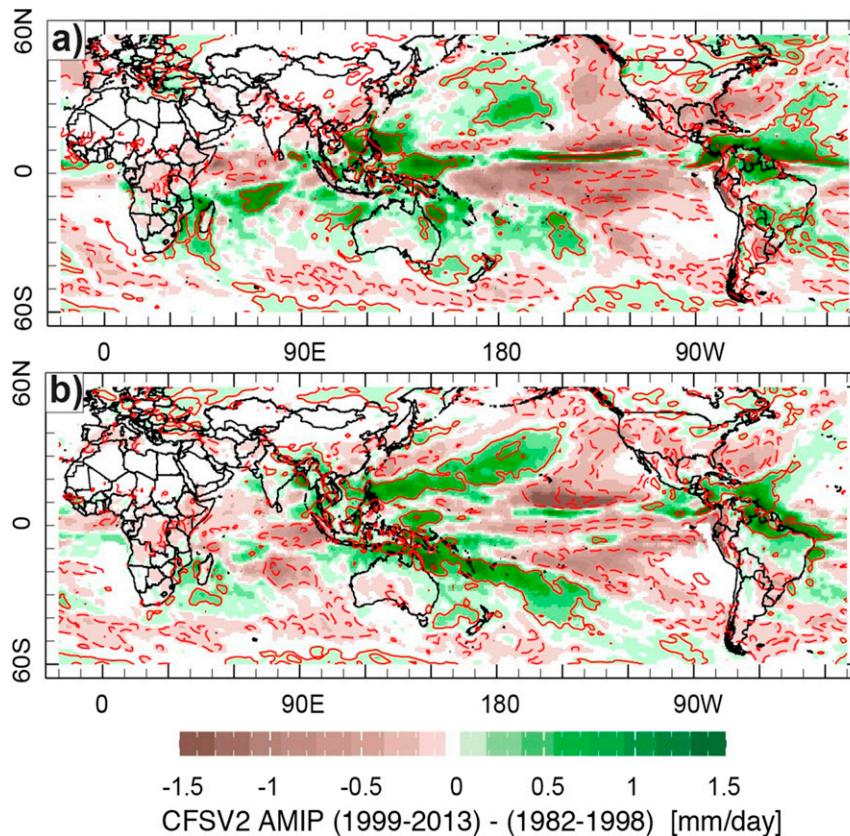


FIG. 9. As in Fig. 8, but for the CFSv2 AMIP runs for (a) DJF and (b) MAM.

geographical distribution of the decadal precipitation shift is shown in Fig. 9 for DJF and JFM, directly comparable with the result for the CFSv2 coupled hindcasts (Fig. 6). The decadal precipitation difference in the tropical Pacific in the CFSv2 AMIP result is not unlike that for the multimodel AMIP (Fig. 8) and lacks the positive difference near the date line in the tropics seen in the coupled CFSv2 result. Poor maintenance of the decadal mean difference in the SST forcing is therefore a likely reason for the imperfect reproduction of the decadal difference in precipitation in CFSv2. However, some role of the CFSv2 atmospheric model in the muted and slightly displaced coupled model decadal difference is also possible, as is a role of the CFSv2 reanalysis data used to initialize the CFSv2 coupled integrations (to be discussed further below). We suggest that these findings for CFSv2 likely approximately apply to the NMME coupled versus AMIP comparison also, despite the lack of equivalency between the sets of constituent models. Evidence for this generalization comes from the fact that while not identical, the “clean” comparison of coupled versus AMIP results for CFSv2 alone is similar to that for the NMME, helping to reinforce its outcome. Consequently, some additional AMIP-based analyses

are performed using the multimodel combination of models shown in Table 2, with an assumption that they would be approximately applicable to the set of NMME model (Table 1).

In view of the realistic decadal precipitation differentiation in the AMIP runs, it is reasonable to suspect that errors in the SST predictions of the NMME may be a cause of its somewhat weak reproduction of the decadal shift in mean precipitation.⁶ To investigate this possibility, we examine the errors in the hindcast NMME SST field at 0- and 3-month lead times using the OISSTv2 dataset for verification. An empirical orthogonal function (EOF) analysis is applied to the error fields of the SST predictions over the Pacific during 1982–2010 for DJF and MAM. Results are shown for the two leading error modes at 0- and 3-month lead times for DJF in Fig. 10, and for MAM in Fig. 11.

In DJF (Fig. 10), at 0-month lead time mode 2 represents the model error in the decadal shift in SST most closely, as indicated both by the spatial pattern and especially by

⁶ As will be discussed below, a problem in the ocean initializations may also contribute to the problems in the SST predictions.

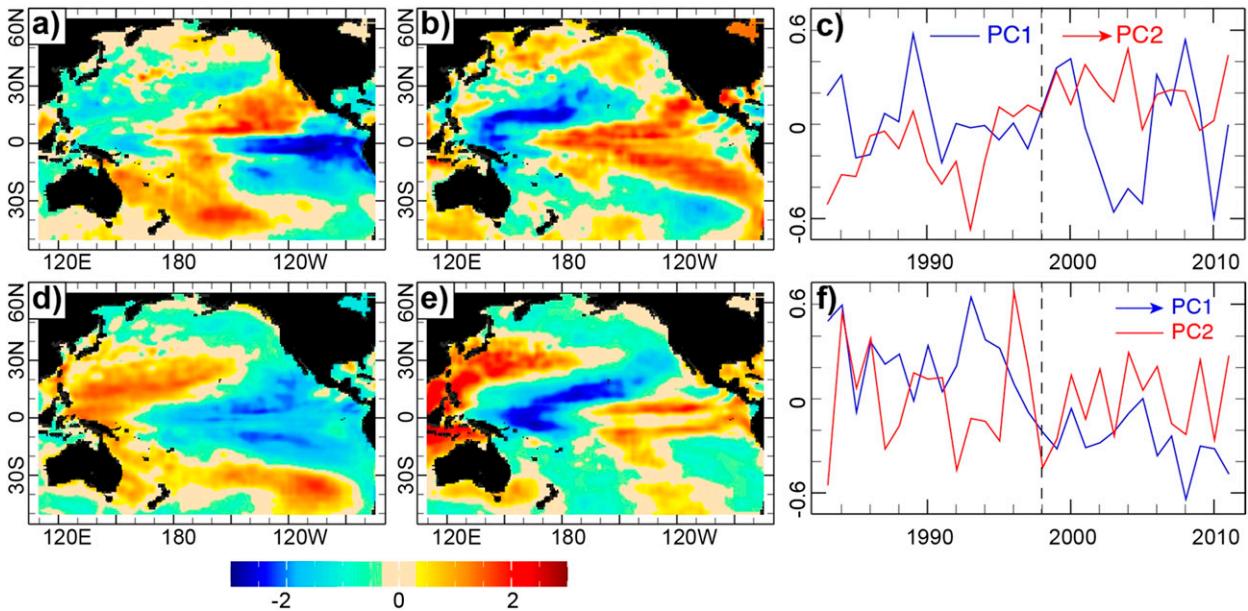


FIG. 10. EOF analysis of the error in seasonal NMME SST predictions for the Pacific during the DJF season (1982–2010). (top) The leading two EOFs for 0-month-lead forecasts for DJF with loadings for (a) EOF1 and (b) EOF2, and (c) the associated principal component time series for DJF, with the vertical dashed line separating the two epochs. (bottom) As at top, but for the 3-month-lead SST forecasts with loadings for (d) EOF1 and (e) EOF2, and (f) the PC time series of these EOFs.

the principal component time series. At 3-month lead time the shift is captured in mode 1, and the spatial pattern resembles that of the observed shift quite well. In the cases of both lead times, the direction of the decadal shift in the error indicates too weak a representation of the observed decadal shift by the ensemble of AMIP models.⁷ For the MAM season (Fig. 11), for 0-month lead time neither mode 1 nor mode 2 clearly captures the decadal shift, which appears mixed between the modes. At 3 months lead the shift in the model error appears most clearly in mode 2. At both leads the NMME SST forecasts underestimate the magnitude of the observed decadal shift.

For both DJF and MAM the model error in predicting the decadal shift appears more clearly in the 3-month than 0-month-lead SST predictions. This result could be related to the relatively greater loss of the decadal signal in the initial conditions as the lead time increases; that is, the models tend to revert to the overall base period climatology with increasing lead time.

While the EOF analyses of the model error fields reflect an underrepresentation of the decadal shift in Pacific SST, we wish to quantify more explicitly the extent

of the underrepresentation. Figure 12 shows the difference in mean SST between 1999–2010 and 1982–98 in the observations (Figs. 12e,f) and in the NMME model predictions for DJF and MAM at 0- and 3-month lead times. For both seasons, the models appear to capture the decadal shift pattern qualitatively correctly, but underpredict its magnitude by about one-third to one-half for 0-month-lead forecasts, and by about two-thirds at 3-month lead time. The decreasing retention of the decadal signal with increasing lead time appears to imply inadequate prediction capability. But why is there already a noticeable loss at 0-month lead time? One possibility is imperfect atmospheric and oceanic initialization.

An artificial discontinuity, unrelated to the shift discussed in this study, also beginning in 1999, has been noted in the CFS Reanalysis (CFSR) data (Saha et al. 2010). This shift induced a change in the pattern of the SST in the eastern tropical Pacific toward a weak El Niño (Kumar et al. 2012; Xue et al. 2013). Because this SST shift is largely in the opposite direction from that of the decadal SST shift discussed here, it may contribute to the models' underestimation of the observed shift. This nonnatural shift has been attributed to the introduction of the ATOVS radiance data in the atmospheric assimilation beginning in late 1998 (Zhang et al. 2012), due to forcing from the atmospheric to the oceanic aspects of the CFSR (Xue et al. 2011). The positive change in east-central tropical Pacific SST in 1999 opposes the downward shift in the observed central and

⁷ For example, for predictions of SST for DJF at 3-month lead time, the time series indicates a decrease in the amplitude of the mode 1 pattern starting in the late 1990s. Since the decadal pattern shown in mode 1 shows the negative phase, the model error is toward too little negative phase (i.e., an underestimation of the observed negative phase since the late 1990s).

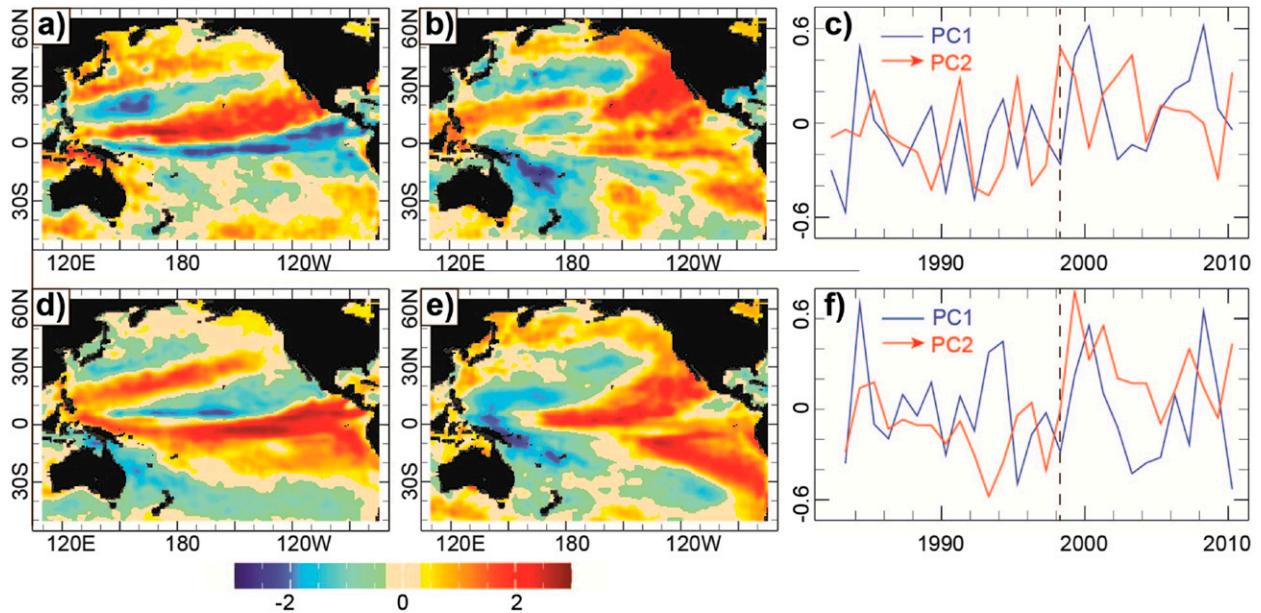


FIG. 11. As in Fig. 10, but for MAM.

east-central SST documented here and in other studies (e.g., Kumar et al. 2010, 2012; Deser et al. 2010; Lyon and DeWitt 2012; Lyon et al. 2013). The opposing directions of shift between this observationally induced

one and the natural one ensures that one of the shifts cannot be camouflaged as the other. However, because they both occurred around the same year, the apparent reproduction of the natural shift by models that use the

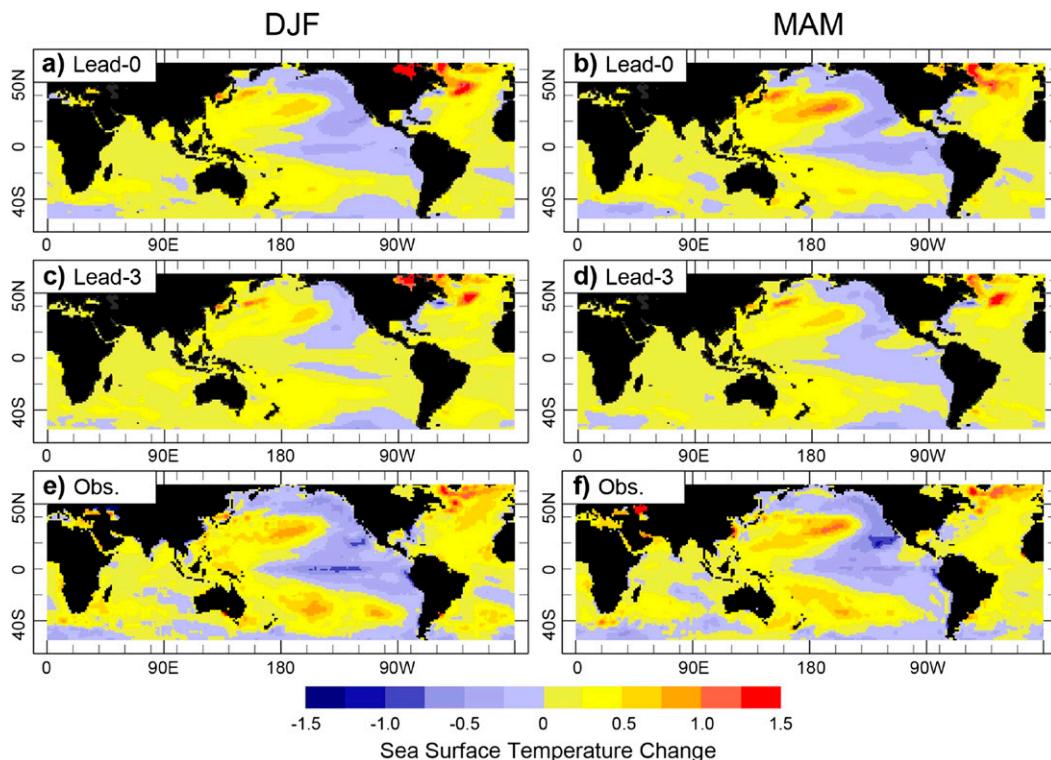


FIG. 12. (top) Difference in NMME 0-month-lead seasonal SST forecasts averaged over 1999–2010 and 1982–98 for (a) DJF and (b) MAM. (middle) As in (a) and (b), but for differences in the 3-month-lead forecasts for (c) DJF and (d) MAM. (bottom) Observed differences in seasonal averages for (e) DJF and (f) MAM.

changed atmospheric assimilation may be damped. The ENSO-related SST predictions of the CFSv2 model, for example, have been found to be affected (Barnston and Tippett 2013).

The relatively faithful reproduction of the difference in mean precipitation from before to after the late 1990s by the AMIP models (Fig. 9) and the inadequate reproduction of the SST anomalies associated with the decadal shift (Figs. 10 and 11) help provide evidence that the SST predictions (and possibly ocean initializations) play a critical role in the somewhat weak representation of the shift in precipitation shown by the NMME. Thus, at least for the NMME model hindcasts, the coupled models are able to reproduce seasonal precipitation changes associated with Pacific decadal SST variability most effectively only at relatively short lead times.

5. Discussion and conclusions

A global-scale decadal climate shift, beginning in 1998/99 and enduring through 2013,⁸ has been documented in recent studies in both observed and model-simulated rainfall. An associated shift in seasonal precipitation has appeared in a number of regions throughout the world, and is most easily detected during MAM when the effects of ENSO are relatively weak. Previous analyses have linked this climate shift to a shift in Pacific SSTs that occurred near 1999, and have characterized the climate pattern as being related to the Pacific decadal variability (PDV) pattern. Here we examine some of the details of the shift in the mean precipitation associated with this SST regime change during boreal winter and spring seasons, with special emphasis on the southwestern United States where deficient precipitation has generally persisted since the late 1990s. We focus the assessment on the predictive skill of the North American Multimodel Ensemble (NMME) in reproducing the precipitation shift in its seasonal hindcasts, and also examine the behavior of the CFSv2 model, which contributes to the NMME.

The NMME ensemble mean hindcasts are found to capture the decadal precipitation shifts qualitatively correctly, but with somewhat underestimated amplitude, particularly for longer lead forecasts. This weakened response applies to the decadal precipitation signal in several locations of the globe but is not severe in the southwestern United States at short lead time; it appears

also in the CFSv2 when considered in isolation from the NMME. The model shifts in precipitation are statistically significant in many of the regions for 0-month-lead forecasts, as they are in the observations. The NMME and the CFSv2 model alone reproduce the main region of negative precipitation anomalies over the east-central tropical Pacific Ocean and an opposing horseshoe pattern of rainfall increases over the Maritime Continent and off-equator western tropical Pacific Ocean, all associated with the decadal shift. Model performance is also favorable for the negative shifts over eastern equatorial Africa and the southwestern United States and the positive shifts over northern South America. Some details of the observed pattern of precipitation decadal shift are not reproduced well by the NMME, such as a shift toward more precipitation in southern Africa in DJF. The CFSv2 model erroneously predicts a dry shift over northeastern Brazil in both DJF and MAM, despite correct forecasts for wetter recent conditions over the rest of northern South America. CFSv2 also shows a spurious small region of wet shift near and north of the equator at the date line, especially for DJF. At 3-month lead time, both NMME and CFSv2 generally substantially underestimate the strength of the decadal shift in precipitation.

An examination of the precipitation field in AMIP style runs for a different (but overlapping) set of atmospheric models suggests that the insufficient amplitude of the precipitation shift is likely associated with an underestimation of the decadal shift in the Pacific SST predictions in the coupled model forecast integrations. This suggestion stems from the finding that when forced by simultaneous observed SSTs, the NMME atmospheric models reproduce the observed decadal precipitation shift quite well. The same analysis using AMIP simulations for the CFSv2 model alone leads to a similar result when compared with the coupled CFSv2 results. A specific problem in the oceanic and atmospheric initialization beginning in 1999 may be a secondary cause of the weak decadal SST signal in the SST forecasts, as the weakness appears even at short lead time. Errors in the atmospheric models of the NMME cannot be ruled out as additional contributors to the muted amplitude of the decadal shift, but they do not appear to be the main cause.

Although some dynamical predictability has been demonstrated on the decadal time scale in the North Pacific (e.g., Mochizuki et al. 2012; Chikamoto et al. 2012), the current findings suggest a gap in performance in the SST predictions within the coupled climate models of the NMME. This result is consistent with previous findings indicating that inherent predictability of North Pacific SST is lower than that in some other

⁸Recent observations suggest that the PDV mode may have shifted to a positive phase in 2014 and continued positive in 2015, but another year or two are needed to be more certain of this phase change.

extratropical ocean basins, due to both greater sensitivity to initial state uncertainty (Branstator et al. 2012; Branstator and Teng 2012) and lack of certainty regarding the mechanisms of internally based decadal variability in the North Pacific (Meehl et al. 2014). Although these studies apply to longer lead forecasts of decadal variability, such as predicting a phase change in the PDV pattern a year or two in advance, they also appear to include coupled models' ability to retain a decadal mean signal, present in the initial conditions, out to one or two seasons into the future. Based on the current work, today's coupled models appear to substantially regress such decadal mean conditions toward the overall average climate within the first 6 months of lead time.

In conclusion, the NMME seasonal hindcasts, and those of the constituent CFSv2, reproduce the pattern of precipitation impacts of a decadal climate shift beginning in 1999 fairly well, but with generally underestimated amplitude and minor errors in the details of the spatial pattern of change in various regions. While coupled models may not be able to forecast a decadal shift years before it occurs (Meehl et al. 2009; Ding et al. 2013), once such a shift has occurred the NMME reproduce its continuation, but with systematically somewhat weakened amplitude at the shortest lead times (up to 3 months), despite the presence of the decadal mean signal in the initial conditions, and with substantially weakened amplitude at longer lead times. This finding underscores a lesser challenge than successfully predicting decadal periods of drought or pluvials at 1 or more years lead time with coupled climate models—namely, that of just maintaining the decadal component of precipitation anomalies in short-lead and especially medium-lead seasonal forecasts.

Acknowledgments. This study was supported by NOAA's Climate Program Office's Modeling, Analysis, Predictions, and Projections (MAPP) program, Award NA12OAR4310088. Both of the authors participated in NOAA's Drought Task Force. The authors wish to thank Richard Seager, Arun Kumar, Martin Hoerling, and Siegfried Schubert for making their AMIP runs available to this study. They also appreciate the careful and constructive reviews by the three anonymous reviewers, as well as by the editor, Joseph Barsugli.

REFERENCES

- Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167, doi:10.1175/1525-7541(2003)004<1147:TVGP>2.0.CO;2.
- Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *J. Climatol.*, **6**, 661–675, doi:10.1002/joc.3370060607.
- Allen, D. M., K. Stahl, P. H. Whitfield, and R. D. Moore, 2014: Trends in groundwater levels in British Columbia. *Can. Water Resour. J.*, **39**, 15–31, doi:10.1080/070111784.2014.885677.
- Barnston, A. G., and M. K. Tippett, 2013: Predictions of Nino3.4 SST in CFSv1 and CFSv2: A diagnostic comparison. *Climate Dyn.*, **41**, 1615–1633, doi:10.1007/s00382-013-1845-2.
- , —, M. L. L'Heureux, S. Li, and D. G. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Amer. Meteor. Soc.*, **93**, 631–651, doi:10.1175/BAMS-D-11-00111.1.
- , —, H. M. van den Dool, and D. A. Unger, 2015: Toward an improved multimodel ENSO prediction. *J. Appl. Meteor. Climatol.*, **54**, 1579–1595, doi:10.1175/JAMC-D-14-0188.1.
- Becker, E., H. van den Dool, and Q. Zhang, 2014: Predictability and forecast skill in NMME. *J. Climate*, **27**, 5891–5906, doi:10.1175/JCLI-D-13-00597.1.
- Branstator, G., and H. Y. Teng, 2012: Potential impact of initialization on decadal predictions as assessed for CMP5 models. *Geophys. Res. Lett.*, **39**, L12703, doi:10.1029/2012GL051974.
- , —, G. A. Meehl, M. Kimoto, J. Knight, M. Latif, and A. Rosati, 2012: Systematic estimates of initial value decadal predictability for six AOGCMs. *J. Climate*, **25**, 1827–1846, doi:10.1175/JCLI-D-11-00227.1.
- Chan, J. C. L., and W. Zhou, 2005: PDO, ENSO and the early summer monsoon rainfall over south China. *Geophys. Res. Lett.*, **32**, L08810, doi:10.1029/2004GL022015.
- Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113**, D04110, doi:10.1029/2007JD009132.
- Chikamoto, Y., and Coauthors, 2012: Predictability of a stepwise shift in Pacific climate during the late 1990s in hindcast experiments by MIROC. *J. Meteor. Soc. Japan*, **90A**, 1–21, doi:10.2151/jmsj.2012-A01.
- Dai, A., 2013: The influence of the inter-decadal Pacific oscillation on US precipitation during 1923–2010. *Climate Dyn.*, **41**, 633–646, doi:10.1007/s00382-012-1446-5.
- DelSole, T., X. Yang, and M. K. Tippett, 2013: Is unequal weighting significantly better than equal weighting for multi-model forecasting? *Quart. J. Roy. Meteor. Soc.*, **139**, 176–183, doi:10.1002/qj.1961.
- Deser, C., A. S. Philips, and M. A. Alexander, 2010: Twentieth century tropical sea surface temperature trends revisited. *Geophys. Res. Lett.*, **37**, L10701, doi:10.1029/2010GL043321.
- Ding, H., R. J. Greatbatch, M. Latif, W. Park, and R. Gerdes, 2013: Hindcast of the 1976/77 and 1998/99 climate shifts in the Pacific. *J. Climate*, **26**, 7650–7661, doi:10.1175/JCLI-D-12-00626.1.
- Gates, W. L., 1992: AMIP: The Atmospheric Model Intercomparison Project. *Bull. Amer. Meteor. Soc.*, **73**, 1962–1970, doi:10.1175/1520-0477(1992)073<1962:ATAMIP>2.0.CO;2.
- Goddard, L., and Coauthors, 2013: A verification framework for interannual-to-decadal predictions experiments. *Climate Dyn.*, **40**, 245–272, doi:10.1007/s00382-012-1481-2.
- Haimberger, L., 2007: Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate*, **20**, 1377–1403, doi:10.1175/JCLI4050.1.
- Hidalgo, H. G., and J. A. Dracup, 2003: ENSO and PDO effects on hydroclimatic variations of the upper Colorado River basin. *J. Hydrometeorol.*, **4**, 5–23, doi:10.1175/1525-7541(2003)004<0005:EAPEOH>2.0.CO;2.
- Hoell, A., C. Funk, and M. Barlow, 2015: The forcing of southwestern Asia teleconnections by low-frequency sea surface

- temperature variability during boreal winter. *J. Climate*, **28**, 1511–1526, doi:10.1175/JCLI-D-14-00344.1.
- Huffman, G. J., and D. T. Bolvin, 2012: GPCP version 2.2 SG combined precipitation data set documentation. NASA Goddard Space Flight Center, 46 pp. [Available online at ftp://precip.gsfc.nasa.gov/pub/gpcp-v2.2/doc/V2.2_doc.pdf.]
- Kharin, V. V., and F. W. Zwiers, 2002: Climate predictions with multimodel ensembles. *J. Climate*, **15**, 793–799, doi:10.1175/1520-0442(2002)015<0793:CPWME>2.0.CO;2.
- Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal to interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, doi:10.1175/BAMS-D-12-00050.1.
- Kosaka, Y., and S.-P. Xie, 2013: Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature*, **501**, 403–407, doi:10.1038/nature12534.
- Kumar, A., J. Bhaskar, and M. L'Heureux, 2010: Are tropical SST trends changing the global teleconnection during La Niña? *Geophys. Res. Lett.*, **37**, L12702, doi:10.1029/2010GL043394.
- , 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. Wea. Rev.*, **140**, 3003–3016, doi:10.1175/MWR-D-11-00335.1.
- Li, S., L. Goddard, and D. G. DeWitt, 2008: Predictive skill of AGCM seasonal climate forecasts subject to different SST prediction methodologies. *J. Climate*, **21**, 2169–2186, doi:10.1175/2007JCLI1660.1.
- Lin, C.-Y., 2014: Sea surface salinity variability of the 1999 climate shift in the tropical Pacific. *J. Mar. Sci. Technol. Taiwan*, **2**, 264–268, doi:10.6119/JMST-013-0508-2.
- Lyon, B., 2014: Seasonal drought in the greater Horn of Africa and its recent increase during the March–May long rains. *J. Climate*, **27**, 7953–7975, doi:10.1175/JCLI-D-13-00459.1.
- , and D. G. DeWitt, 2012: A recent and abrupt decline in the East African long rains. *Geophys. Res. Lett.*, **39**, L02702, doi:10.1029/2011GL050337.
- , A. G. Barnston, and D. G. DeWitt, 2013: Tropical Pacific forcing of a 1998–1999 climate shift: Observational analysis and climate model results for the boreal spring season. *Climate Dyn.*, **43**, 893–909, doi:10.1007/s00382-013-1891-9.
- Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.*, **78**, 1069–1079, doi:10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.
- Meehl, G. A., and Coauthors, 2009: Decadal prediction: Can it be skillful? *Bull. Amer. Meteor. Soc.*, **90**, 1467–1485, doi:10.1175/2009BAMS2778.1.
- , and Coauthors, 2014: Decadal climate prediction: An update from the trenches. *Bull. Amer. Meteor. Soc.*, **95**, 243–267, doi:10.1175/BAMS-D-12-00241.1.
- Mochizuki, T., and Coauthors, 2012: Decadal prediction using a recent series of MIROC global climate models. *J. Meteor. Soc. Japan*, **90A**, 373–383, doi:10.2151/jmsj.2012-A22.
- Newman, M., 2007: Interannual to decadal predictability of tropical and North Pacific sea surface temperatures. *J. Climate*, **20**, 2333–2356, doi:10.1175/JCLI4165.1.
- Peña, M., and H. van den Dool, 2008: Consolidation of multimodel forecasts by ridge regression: Application to Pacific sea surface temperature. *J. Climate*, **21**, 6521–6538, doi:10.1175/2008JCLI2226.1.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625, doi:10.1175/1520-0442(2002)015<1609:AISAS>2.0.CO;2.
- Ropelewski, C. F., and M. S. Halpert, 1989: Precipitation patterns associated with the high index phase of the Southern Oscillation. *J. Climate*, **2**, 268–284, doi:10.1175/1520-0442(1989)002<0268:PPAWTH>2.0.CO;2.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:10.1175/2010BAMS3001.1.
- Seager, R., M. Hoerling, S. Schubert, H. Wang, B. Lyon, A. Kumar, J. Nakamura, and N. Henderson, 2015: Causes of the 2011–14 California drought. *J. Climate*, **28**, 6997–7024, doi:10.1175/JCLI-D-14-00860.1.
- Swetnam, T. W., and J. L. Betancourt, 2010: Mesoscale disturbance and ecological response to decadal climate variability in the American Southwest. *Tree Rings and Natural Hazards: Advances in Global Change Research*, D. R. Butler and B. H. Luckman, Eds., Springer, 329–359, doi:10.1007/978-90-481-8736-2_32.
- Tippett, M. K., and A. G. Barnston, 2008: Skill of multimodel ENSO probability forecasts. *Mon. Wea. Rev.*, **136**, 3933–3946, doi:10.1175/2008MWR2431.1.
- Trenberth, K. E., and J. W. Hurrell, 1994: Decadal atmosphere–ocean variations in the Pacific. *Climate Dyn.*, **9**, 303–319, doi:10.1007/BF00204745.
- , and J. T. Fasullo, 2013: An apparent hiatus in global warming? *Earth's Future*, **1**, 19–32, doi:10.1002/2013EF000165.
- van den Dool, H. M., and R. M. Chervin, 1986: A comparison of month-to-month persistence of anomalies in a general circulation model and in the earth's atmosphere. *J. Atmos. Sci.*, **43**, 1454–1466, doi:10.1175/1520-0469(1986)043<1454:ACOMTM>2.0.CO;2.
- Xue, Y., B. Huang, Z.-Z. Hu, A. Kumar, C. Wen, D. Behringer, and S. Nadiga, 2011: An assessment of oceanic variability in the NCEP Climate Forecast System Reanalysis. *Climate Dyn.*, **37**, 2511–2539, doi:10.1007/s00382-010-0954-4.
- , 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. Climate*, **26**, 5358–5378, doi:10.1175/JCLI-D-12-00600.1.
- Yang, W., R. Seager, M. A. Cane, and B. Lyon, 2014: The East African long rains in observations and models. *J. Climate*, **27**, 7185–7202, doi:10.1175/JCLI-D-13-00447.1.
- 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.