

## Bias Correction and Forecast Skill of NCEP GFS Ensemble Week-1 and Week-2 Precipitation, 2-m Surface Air Temperature, and Soil Moisture Forecasts

YUN FAN\* AND HUUG VAN DEN DOOL

*NOAA/NCEP/CPC, Camp Springs, Maryland*

(Manuscript received 17 September 2010, in final form 30 December 2010)

### ABSTRACT

A simple bias correction method was used to correct daily operational ensemble week-1 and week-2 precipitation and 2-m surface air temperature forecasts from the NCEP Global Forecast System (GFS). The study shows some unexpected and striking features of the forecast errors or biases of both precipitation and 2-m surface air temperature from the GFS. They are dominated by relatively large-scale spatial patterns and low-frequency variations that resemble the annual cycle. A large portion of these forecast errors is removable, but the effectiveness is time and space dependent. The bias-corrected week-1 and week-2 ensemble precipitation and 2-m surface air temperature forecasts indicate some improvements over their raw counterparts. However, the overall levels of week-1 and week-2 forecast skill in terms of spatial anomaly correlation and root-mean-square error are still only modest. The dynamical soil moisture forecasts (i.e., land surface hydrological model forced with bias-corrected precipitation and 2-m surface air temperature integrated forward for up to 2 weeks) have very high skill, but hardly beat persistence over the United States. The inability to outperform persistence mainly relates to the skill of the current GFS week-1 and week-2 precipitation forecasts not being above a threshold (i.e., anomaly correlation  $> 0.5$  is required).

### 1. Introduction

Precipitation ( $P$ ) and 2-m surface air temperature (T2m) are two meteorological variables that have most important impacts on human society. For soil moisture, the so-called sea surface temperature (SST) has also been considered important for weather and climate prediction, in particular during the warm season when the land and atmosphere are more tightly coupled (Dirmeyer 2000; Kanamitsu et al. 2003; Koster et al. 2003; Van den Dool et al. 2003; Zhang and Frederikson 2003; Van den Dool 2007). Soil moisture is also an important indicator for real-time drought and flood monitoring. Therefore, for obvious reasons, accurately predicting these three variables is of great practical importance. In 1997 the Climate Prediction Center (CPC) of the National Centers for Environmental Prediction

(NCEP) started a soil moisture “dynamical” week-1 and week-2 outlook program, over the United States only, on a daily basis, using CPC’s leaky bucket (LB) land surface hydrological model (Huang et al. 1996; Van den Dool et al. 2003) forced with week-1 and week-2  $P$  and T2m from a single-member forecast of the NCEP Medium-Range Forecast (MRF), now known as the Global Forecast System (GFS). From late 2001 onward, the GFS ensemble mean forecast was used to replace the single-member forecast and the procedure was further improved in late 2003 to include the bias-corrected GFS ensemble mean forecast.

In mid-2007, the CPC initiated its monitoring and prediction of the variability of global (African, Asian, Australian, and American) monsoon systems, in collaboration with the international community’s efforts on improving monsoon monitoring and providing timely and hopefully useful weather and climate information for different users and decision makers regionally and globally. With the release of the CPC gauge-based daily Global Unified Land Surface Precipitation Analysis in late 2007 (Chen et al. 2008), the daily bias-corrected GFS ensemble week-1 and week-2 precipitation forecasts have been expanded to the global land surface.

The reader should understand that the LB model is kept up to date every day by forcing it with the observed

---

\* Current affiliation: Meteorological Development Laboratory, Office of Science and Technology, National Weather Service, NOAA, Silver Spring, Maryland.

---

Corresponding author address: Dr. Yun Fan, Climate Prediction Center, Rm. 806, 5200 Auth Rd., Camp Springs, MD 20746.  
E-mail: yun.fan@noaa.gov

$P$  and  $T2m$ . One can look upon this as an integration of the LB from 1931 to 1200 UTC yesterday (to provide initial conditions), and then the GFS's  $T2m$  and  $P$  are appended to this ongoing LB integration to jump another 2 weeks ahead (to make soil moisture forecasts). We do not use the GFS's soil moisture directly, there by avoiding having to deal with the potentially very biased soil moisture states of the GFS. We note that the LB land surface hydrological model is integrated in, what is called, an offline fashion (i.e., not coupled to the atmosphere). More primitive approaches to avoiding the GFS soil moisture bias include considering the 2-week change in the GFS's soil moisture predicted by the GFS itself, a product launched by Center for Ocean–Land–Atmosphere Study (COLA) around 1995.

In this paper we report upon our research into operational products. When we talk below about research, we mean research “on the fly” applied to products that were generated in real time, that is, only a few years’ worth of data have been saved (sometimes with holes or a few days of missing data in the archive) and nothing was rerun since it is too costly to rerun the GFS. The work reported here is thus quite different from research in which hindcasts with a constant or frozen model were made with the express purpose of model calibration (Hamill et al. 2006; Saha et al. 2006).

The NCEP GFS is not a frozen system but has been upgraded frequently in terms of its dynamical core and physics package in recent years. In the early stages of CPC's soil moisture “dynamical” outlook, its performance was relatively poorer. In recent years, there is considerable anecdotal evidence that the NCEP GFS and GFS ensemble produce skillful forecasts, which motivates a formal quantitative verification of the modeling system. The first part of this work is to assess the GFS ensemble mean week-1 and week-2  $P$  forecasts over global land and  $T2m$  forecasts over the United States. The main attention is on the skill of the bias-corrected GFS ensemble mean  $P$  forecasts over the North American, South American, Asia–Australian, and African monsoon regions. Detailed analysis is conducted on the spatiotemporal distribution of the bias, in order to address questions like: What does the bias look like and is it removable? Does bias correction improve the GFS forecast skill? The second part of this research focuses on the predictability of the land surface, but over the United States only. Since the predictability of soil moisture critically depends on the quality of the GFS ensemble precipitation and the surface air temperature forecasts, further analysis is done on the spatiotemporal features of the GFS-driven soil moisture forecast skills, that is, when, where, and to what extent the soil moisture can be predicted on week-1 and week-2 time scales beyond the skill of a persistence forecast.

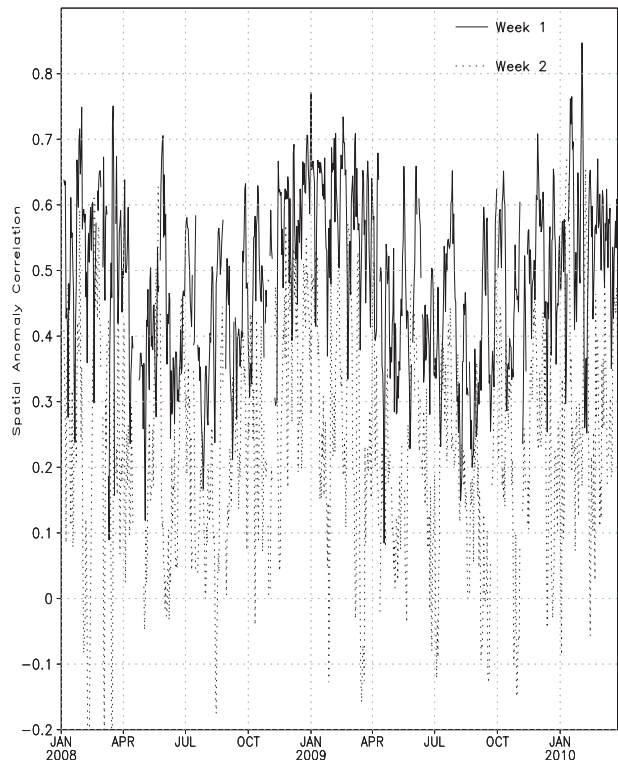


FIG. 1. Time series of the daily spatial correlation of week-1 (solid) and week-2 (dotted) observed and forecasted precipitation anomalies over North America (averaged over  $10^{\circ}$ – $55^{\circ}$ N,  $140^{\circ}$ – $60^{\circ}$ W) for the period 1 Jan 2008–31 Mar 2010. Bias correction is based on 30-day mean forecast errors on  $0.5^{\circ} \times 0.5^{\circ}$  grid. Units are dimensionless.

## 2. Methodology and data

When compared with other meteorological variables, precipitation is more difficult to forecast due to its episodic and short-lived nature. The forecasts for  $T2m$  are generally better than those for precipitation. The spatiotemporal distribution of  $T2m$  is also relatively more homogeneous, but this does not mean that  $T2m$  is easier to forecast everywhere because it is strongly impacted by the complexity of the lower boundary conditions (such as land and water surface, soil properties, and vegetation covers) and topography. The current level of forecast skill for  $P$  and  $T2m$  directly from numerical weather prediction (NWP) models in week-1 and week-2 time scales is still not good enough and a bias correction or postprocessing process is needed before the forecast is issued or applied elsewhere.

In this study, a simple running-mean error correction is implemented to remove biases from the NCEP GFS ensemble mean forecasts. That is, every day at 0000 UTC the week-1 and week-2 GFS ensemble mean  $P$  and  $T2m$  forecasts have been corrected with the past  $N$ -day

Time Series of Spatial Correlation of Weekly Observed & Forecasted Precip Anomalies

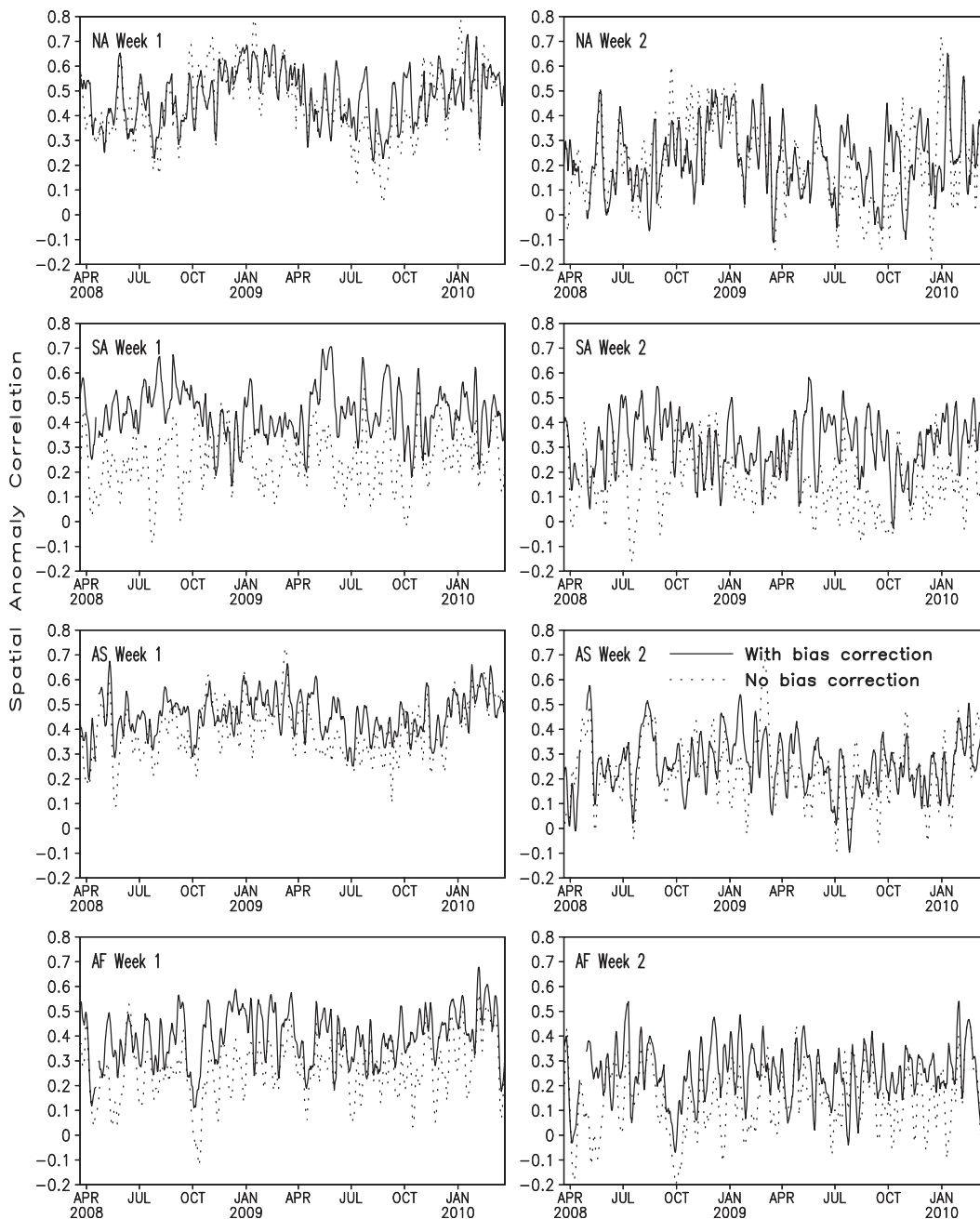


FIG. 2. Time series of 5-day running-mean daily spatial correlations of (left) week-1 and (right) week-2 observed and forecasted precipitation anomalies over (top to bottom) North America (NA), South America (SA), Asia-Australia (AS), and Africa (AF). Bias-corrected (raw) forecast scores are shown by the solid (dotted) line. The bias correction is based on 30-day mean forecast errors. Units are dimensionless.

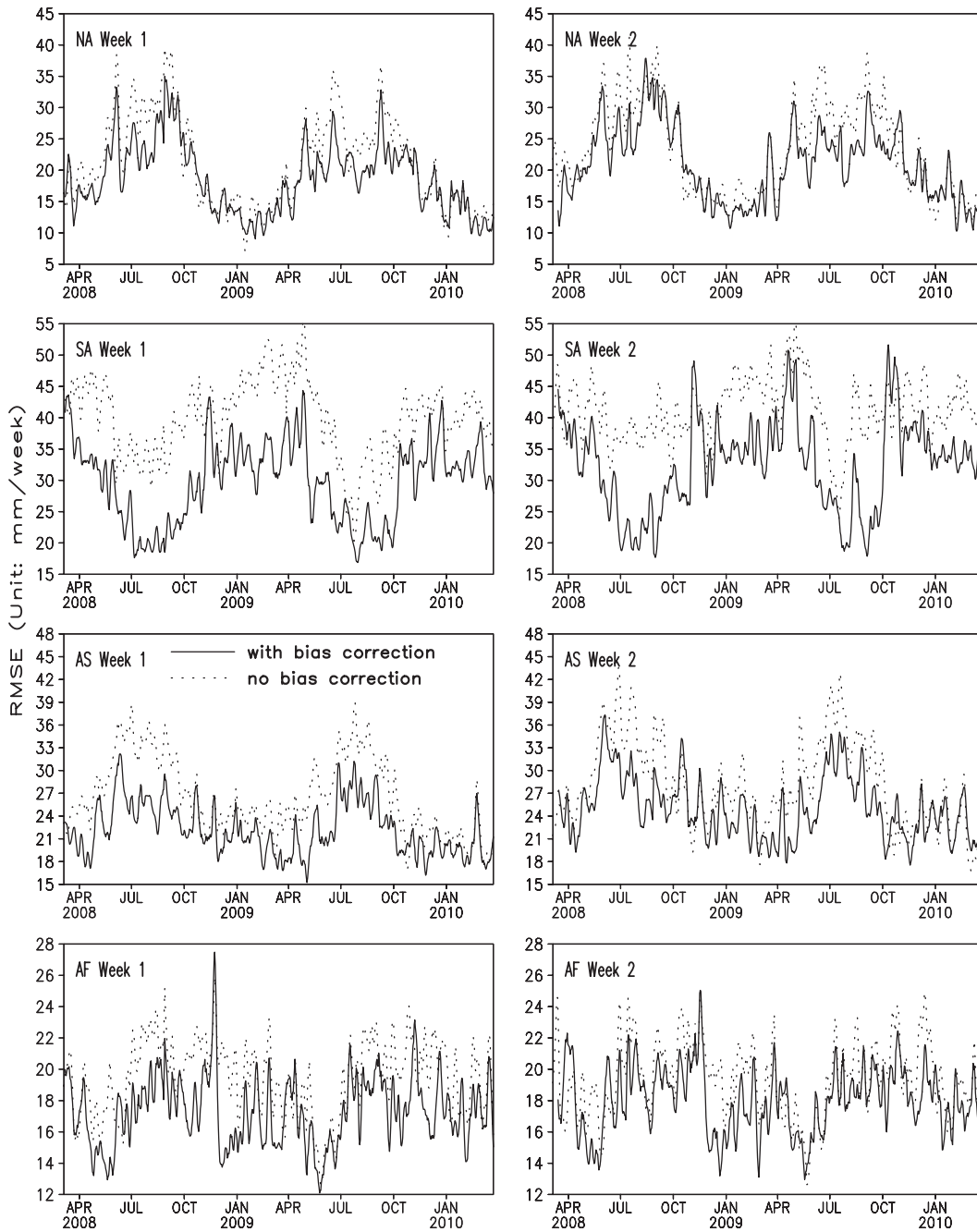
running-mean forecast errors (i.e., biases), defined as follows:

$$\text{Bias1} = \Sigma[\text{Pf}(\text{week1}) - \text{Po}(\text{week1})]/N \quad \text{and} \quad (1)$$

$$\text{Bias2} = \Sigma[\text{Pf}(\text{week2}) - \text{Po}(\text{week2})]/N, \quad (2)$$

where Pf is the weekly mean or accumulated model output and Po is the observed weekly mean or accumulated data. In addition, N is the number of days (e.g., 30 or 7 days, these being the only two choices being maintained in real time at CPC) and the choice of N is subjective. In general, the mean forecast errors calculated from larger N (e.g., 30

RMSE Time Series of GFS Forecasted Precip &amp; Observed Precip

FIG. 3. As in Fig. 2, but for RMSEs. Units are  $\text{mm week}^{-1}$ .

days) are more robust than those from the smaller  $N$  (e.g., 7 days or 1 day), but  $N$  cannot be too much larger than 30 because seasonality may interfere; keep in mind that the  $N$  day mean is backward looking. The  $N = 7$  estimate, while noisy, is also maintained because forecasters sometimes believe that a regime-dependent systematic bias can be applied.

The model data used here are from the NCEP Global Ensemble Forecast System (GEFS; Toth et al. 1997), which is a GFS (a global spectral data assimilation and forecast model system) based modeling system. It runs with 20 ensemble members per cycle plus one control at T126. The GEFS forecasts are produced for up to 28 levels every 6 h at 0000, 0600, 1200, and 1800 UTC.

TABLE 1. Averaged (1 Apr 2008–31 Mar 2010) spatial correlations of observed and GFS forecast week-1 and week-2 precipitation anomalies over different monsoon regions.

	Week 1		Week 2	
	Bias correction	No bias correction	Bias correction	No bias correction
NA	0.49	0.48	0.24	0.26
SA	0.45	0.25	0.31	0.18
AS	0.47	0.40	0.29	0.26
AF	0.40	0.24	0.25	0.13

TABLE 2. Averaged (1 Apr 2008–31 Mar 2010) RMSEs of observed and GSF-forecasted week-1 and week-2 precipitation over different monsoon regions (mm week<sup>-1</sup>).

	Week 1		Week 2	
	Bias correction	No bias correction	Bias correction	No bias correction
NA	19.18	22.82	21.61	23.58
SA	29.55	41.06	32.27	41.72
AS	22.65	27.62	25.24	29.15
AF	17.06	19.47	17.66	19.33

All runs are up to 384 h at 6-h intervals. Data are interpolated to 1° × 1° resolution from 0 to 384 forecast hours. The NCEP GFS is an unfrozen system and major changes have been implemented frequently (information online at <http://www.emc.ncep.noaa.gov/GFS/doc.php>).

The above simple running-mean bias correction has been applied to the NCEP GEFS ensemble mean week-1 and week-2 *P* accumulation forecasts at 0000 UTC. The observed week-1 and week-2 precipitation is from the CPC daily U.S. (Higgins et al. 1996) and Global Unified Precipitation Analysis (Chen et al. 2008). The very same bias correction method is also applied every day to the week-1 and week-2 averaged GFS ensemble T2m forecasts, but over the United States only, because the current CPC daily T2m analysis (Janowiak et al. 1999) is available for the United States only.

Of course, one can calculate the mean forecast errors for the bias correction with more complicated methods, such as nonequal weighting (giving larger weights to more recent days) or the use of probability density function (PDF) adjustment (Wang and Xie 2007) based on the forecasted and observed precipitation in the past few days. Another more sophisticated model output statistics (MOS) technique (Glahn and Lowry 1972) is also used for operational weather forecasts at the National Weather Service, which relates observed meteorological elements (predictands) to appropriate variables (predictors, such as model outputs) through a statistical approach.

The daily bias-corrected GFS ensemble week-1 and week-2 *P* and T2m forecasts and other land surface variables (i.e., soil moisture, evaporation, and runoff) from the CPC leaky bucket hydrological model (Huang et al. 1996) forced with the above bias-corrected *P* and T2m over the United States are archived from late 2003. However, the daily bias-corrected GFS ensemble week-1 and week-2 *P* forecasts over the globe are archived from late 2007, with some missing data in the first few months. Therefore, when comparing or analyzing *P* and T2m forecast skills over the different regions, the common period from 1 April 2008 to 31 March 2010 was used. While for other variables (i.e., soil moisture) over the

U.S. region, the data were used for as long a period of record as possible (i.e., perhaps as far back as late 2003).

This study does not intend to compare bias correction methods, but focuses on the spatiotemporal features of model forecast errors or biases, and whether and to what extent they are removable. All results were based on the above simple running mean bias correction [Eqs. (1) and (2)] method.

### 3. Performance of NCEP GFS week-1 and week-2 bias-corrected ensemble mean *P* and T2m forecasts

Since the above simple running-mean bias corrections (with both 30- and 7-day mean forecast errors) are performed every day, the raw (no bias correction applied) datasets and bias-corrected datasets have been archived on a daily basis for verification and research. Figure 1 shows the time evolution of the daily spatial correlation between the week-1 and week-2 observed precipitation anomalies and the GEFS ensemble forecasted precipitation anomalies over North America (averaged over 10°–55°N, 140°–60°W), corrected with the 30-day running-mean forecast errors. The dominant features are a large day-to-day fluctuation in skill and a seasonal cycle in the GEFS precipitation forecast skill, with the relatively higher skill in the cold season and lower skill in the warm season. In general, the mean of the daily spatial correlation skill for the week-1 GEFS ensemble mean precipitation forecasts is around 0.49 and it is 0.24 for the week-2 GEFS ensemble mean precipitation forecasts over the period of 1 January 2008–31 April 2010. The same levels of daily forecast skill over other major monsoon regions (e.g., South America, Asia–Australia, and Africa) show similar large day-to-day fluctuation but with somewhat different levels of skill (not shown).

Here, one of our main questions is: Does bias correction actually improve the GEFS forecast skill? Figs. 2 and 3 display time series of the 5-day running-mean daily spatial correlation and root-mean-square error

## Spatial Correlation &amp; RMSE Time Series of Weekly Observed &amp; GFS Forecasted T2m

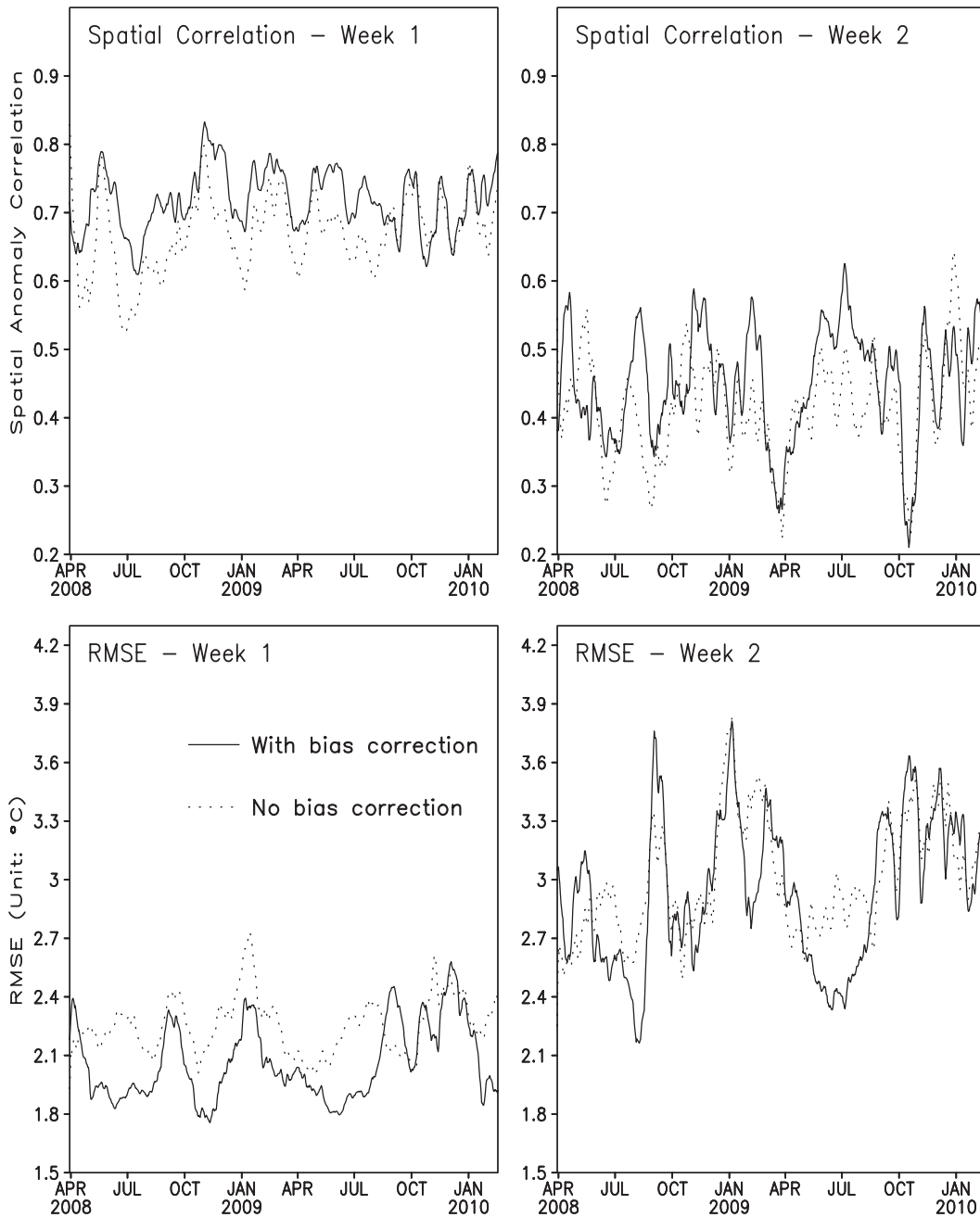


FIG. 4. Time series of 30-day running-mean (top) spatial anomaly correlation and (bottom) RMSE of (left) week-1 and (right) week-2 observed and forecasted T2m over the United States. Bias-corrected (raw) forecast scores are shown by the solid (dotted) line. Bias correction is based on 30-day mean forecast errors. Units are  $^{\circ}\text{C}$  for RMSE and dimensionless for the spatial correlation.

(RMSE) of week-1 and week-2 observed and forecasted ensemble  $P$  over the different major monsoon regions [i.e., North America (NA), South America (SA), Asia-Australia (AS), and Africa (AF)] with bias correction based on 30-day running-mean forecast errors. The

results (also see the mean values in Table 1) show that in terms of spatial anomaly correlation the bias correction offers very little help to correct the spatial anomaly patterns of the GEFS-forecasted precipitation over North America (bias correction greatly reduces

TABLE 3. Averaged (1 Apr 2008–31 Mar 2010) spatial anomaly correlations and RMSEs (°C) of forecasted and observed week-1 and week-2 T2m over the United States.

	Week 1		Week 2	
	Bias correction	No bias correction	Bias correction	No bias correction
Spatial correlation	0.72	0.68	0.48	0.44
RMSE	2.13	2.30	3.04	3.15

the annual mean bias; see discussion concerning Fig. 5 later), some help in the Asia–Australia monsoon regions, and considerable help around South America and Africa, where the week-1 ensemble raw forecast skill and bias-corrected forecast skill changed from 0.25 to 0.45 (increased by 80%) and 0.24 to 0.4 (increased by 67%), respectively. In terms of RMSE, bias correction helps everywhere (see mean values in Table 2). Therefore, the effectiveness of the bias correction not only depends on the way the skill is measured, but it also depends on the location and time of year.

Two general conclusions suggest themselves about precipitation forecasts. The first is that bias correction is more effective over Africa and South America than over North America. Second, while bias correction decreases the root-mean-square error in all domains studied, the anomaly correlation improves considerably only over Africa and South America.

Because the resolution of the GEFS ensemble mean week-1 and week-2 forecasts used here is on a  $2.5^\circ \times 2.5^\circ$  grid and the observed CPC daily Unified Global Precipitation Analysis is on a  $0.5^\circ \times 0.5^\circ$  grid, one can do the verification on either grid. Therefore, one question raised here is whether the above levels of forecast skill are grid (or resolution) dependent. A test has been conducted on both grids and the results show that the skill assessment does not depend much on the grids, despite that some higher-resolution information may be lost when working on the  $2.5^\circ \times 2.5^\circ$  grid. Similar results were found in analyzing precipitation forecast skill (Higgins et al. 2008) from the NCEP Climate Forecast System (CFS; Saha et al. 2006).

Another question is whether the above levels of forecast skill are impacted by the number of days ( $N$ ) that is used to calculate the running-mean bias for bias correction on the GEFS ensemble raw forecast data. Some comparisons also have been done on the levels of week-1 and week-2 forecast skill from bias corrections based on 30- and 7-day running-mean forecast errors. The results indicate that the performance of the bias correction when using the 30-day running-mean forecast errors are better overall than those found when using the 7-day running mean forecast errors (not shown).

Since the daily global T2m gridded analysis is not available during the same period at the present time, the time evolution of the 30-day running mean of the daily spatial anomaly correlation and RMSE from the week-1 and week-2 observed T2m and bias-corrected ensemble T2m results over the United States only is displayed in Fig. 4 and Table 3. Compared to the precipitation, the spatial anomaly correlation between the bias-corrected GEFS ensemble week-1 and week-2 T2m forecast and observation results is relatively high (averaging around 0.72 for week 1 and around 0.48 for week 2). The week-1 and week-2 T2m RMSEs also show a clear seasonal cycle. Bias correction presents some improvements for both the week-1 and week-2 ensemble T2m forecasts, in terms of both spatial anomaly correlation and RMSE.

#### 4. Analysis of week-1 and week-2 forecast errors

For bias correction purposes, the best scenario is that the biases do not change with time; that is, they are constant, so that they can be easily removed by simply subtracting the operationally obtained estimate from model outputs. If they are not constant, it will be desirable that they have large-scale spatial structures and vary regularly and slowly with time. Thus, it is relatively easier to remove these parts of the biases. Obviously, if the biases have small-scale spatial structures and vary irregularly and quickly with time, it will be very difficult to remove them.

To understand why bias correction works while it varies in space and time, some detailed analysis on the spatiotemporal structure of the GEFS week-1 and week-2 ensemble mean forecast errors has been conducted. In general, the GEFS ensemble mean forecast errors can be separated into two parts, that is, the annual mean forecast errors and their variations around the annual means, which were further decomposed by using empirical orthogonal function (EOF) analysis as follows:

$$\text{Bias}_{1,2}(s,t) = \text{Mean} + \sum_{m=1}^M \text{PC}_m(t) \times \text{EOF}_m(s), \quad (3)$$

where 1, 2 refers to either the week-1 or week-2 ensemble means and  $s$  and  $t$  represent spatial and temporal points, respectively.

The annual means (averaged over 1 April 2008–31 March 2010) of the GEFS week-1 and week-2 ensemble precipitation forecast errors (Fig. 5) show that the GEFS ensemble forecasts tend to produce too much rainfall in most regions. The pattern and amplitude of the week-1 and week-2 forecast errors are very similar to each other, indicating this component of the GEFS forecast errors is nearly saturated in the week-1 period. The variation part

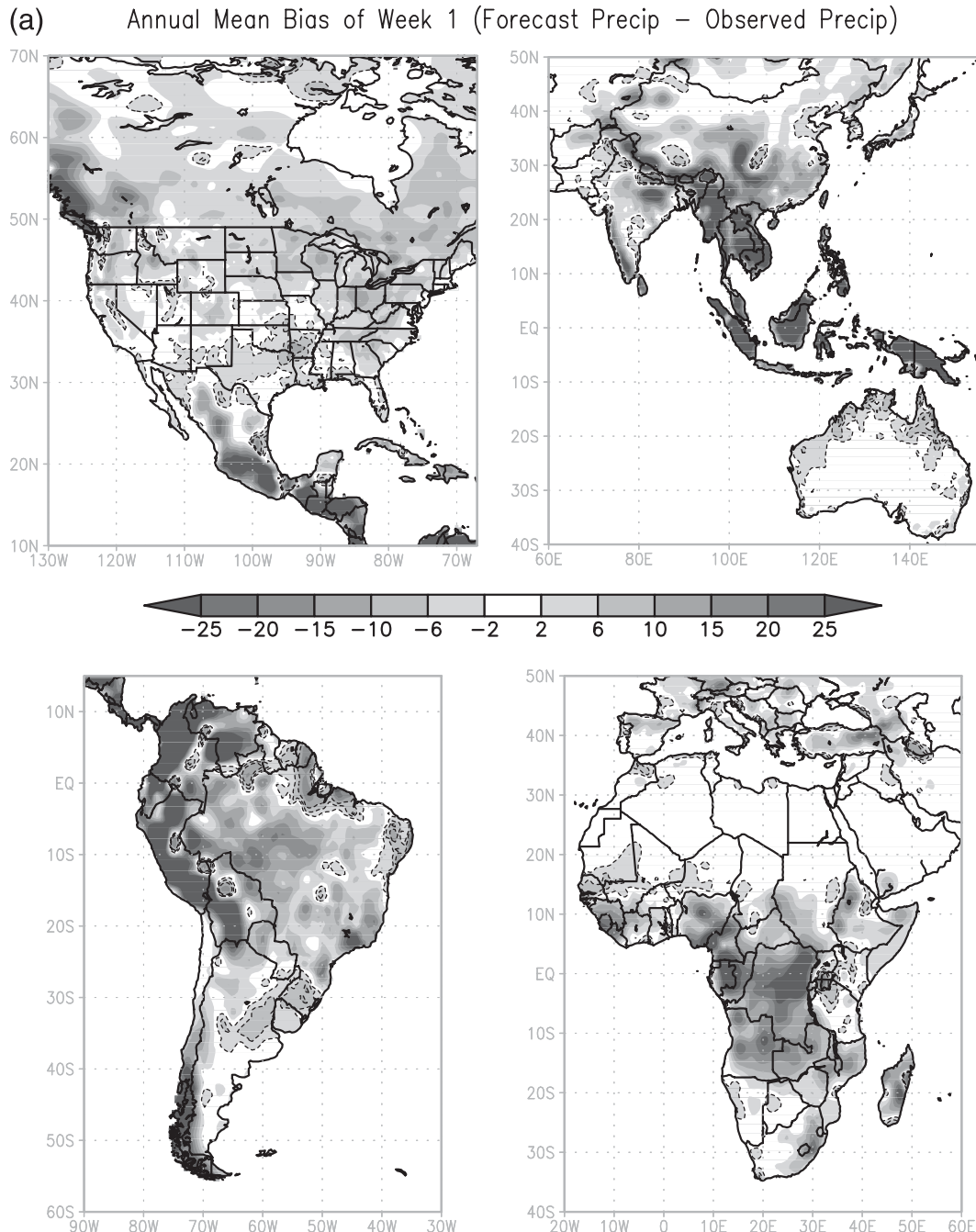


FIG. 5. (a) Annual mean bias of week-1 forecasted precipitation over NA, SA, AS, and AF for the period 1 Apr 2008–31 Mar 2010. Negative values are shown inside the dashed contour ( $\text{mm week}^{-1}$ ). (b) As in (a), but for week-2 forecasted precipitation.

(around the above annual mean) of the GEFS week-1 and week-2 ensemble mean precipitation (30-day running mean) forecast errors is displayed in Fig. 6. The unexpected and most prominent features are that the GEFS ensemble mean forecast errors are of relative large scale and low frequency (which appear to be annual and

semiannual cycles). The first two EOF modes [scaled by the RMS value of the associated principal components (PCs)] of the interannual GEFS week-1 and week-2 ensemble mean forecast errors explain about 60% of the total variance. The above features exist almost everywhere in four major monsoon regions (Asia–Australia



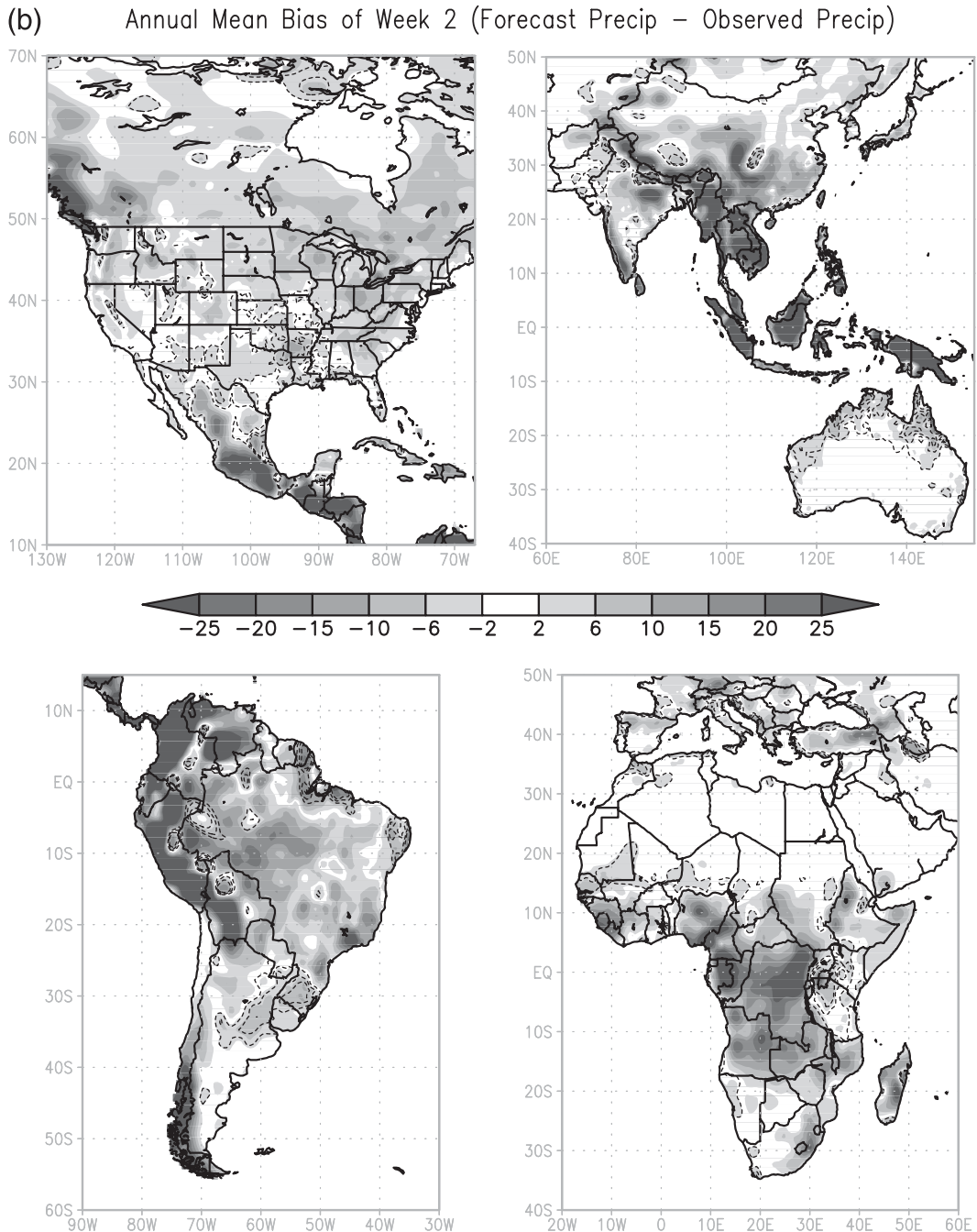


FIG. 5. (Continued)

and Africa are not shown here). These are the unexpected and most striking features found in this study. The bias correction shows that a very large portion of the annual mean raw forecast errors can be removed. The annual mean forecast errors after bias correction are about 5 times smaller than their annual mean raw forecast errors. The EOF analysis on the interannual variation portion of the bias-corrected forecast errors shows that

part of the variable forecast errors can also be effectively removed, especially in the cold season (not shown).

The annual means (also averaged over 1 April 2008–31 March 2010) of the GEFS week-1 and week-2 ensemble mean T2m forecast errors before and after bias corrections over the United States are displayed in Fig. 7. The results show that the GEFS ensemble mean week-1 and week-2 T2m forecast errors have relatively

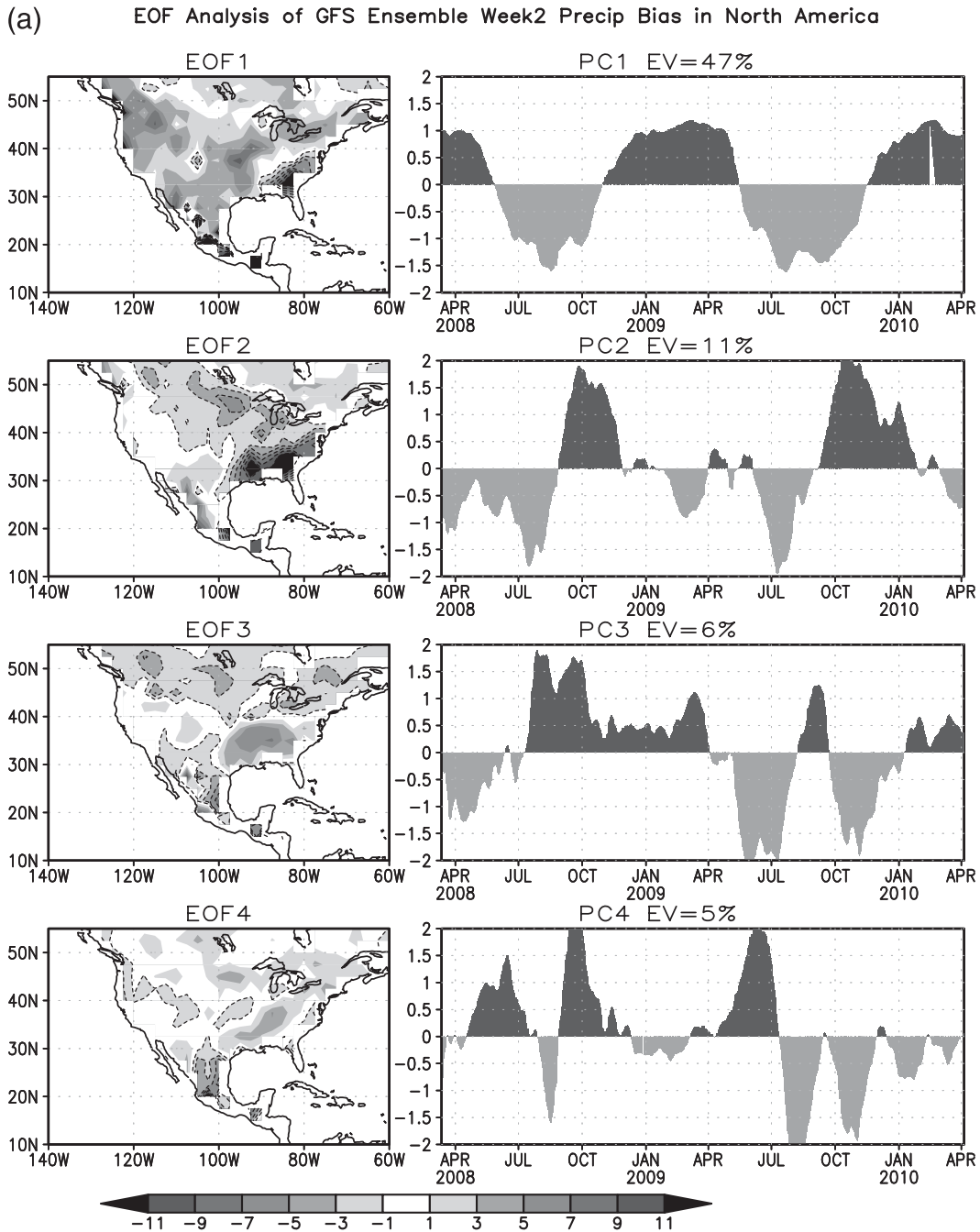


FIG. 6. (a) (left) EOF patterns (scaled by the RMS value of the associated PCs, with negative values inside the dashed contour) and (right) their PCs (normalized to unit variance) of week-2 forecasted precipitation biases over NA for the period 1 Apr 2008–31 Mar 2010. (b) As in (a), but for SA.

large-scale spatial patterns, with about  $1^{\circ}$ – $3^{\circ}\text{C}$  cold biases in the central United States and warm biases in the eastern and western United States, respectively. A very large portion of the annual mean bias can be easily removed by just simply subtracting the 30-day running-mean bias that was determined from previous data. However, the GEFS

ensemble mean T2m forecast errors seem more complicated in some U.S. western coastal and mountain regions, as the above simple running-mean bias correction method does not work well in these areas.

Like GEFS ensemble mean precipitation forecast errors, the EOF analysis (Fig. 8) reveals that the first EOF

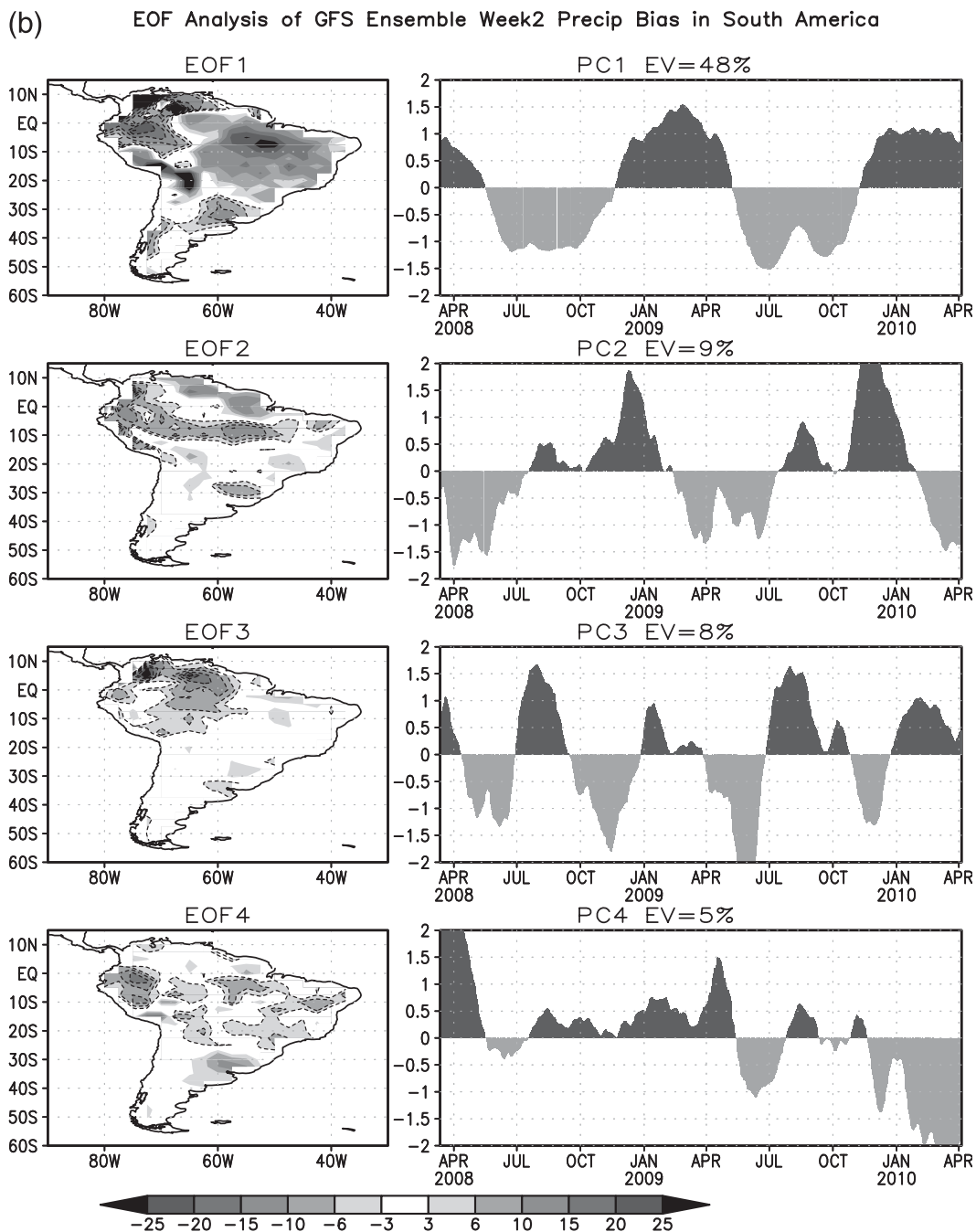


FIG. 6. (Continued)

mode (scaled by the RMS value of the associated PC1) of the interannual week-2 ensemble mean T2m forecast errors is dominated by relatively large-scale spatial structures and low-frequency (annual cycle) evolution. This mode represents that summers are too cold and winters are too warm over a large part of the United States for the GEFS ensemble mean T2m forecasts. The first two EOF modes of the GFS week-1 and week-2

ensemble mean T2m forecast errors can explain over 60% of the total variance. Parts of these varying forecast errors can be easily removed. The EOF analysis of the week-1 ensemble mean T2m forecast errors has features that are very similar to the above results (not shown).

One is tempted to conclude, based on Figs. 5–8, that the NCEP GEFS has some simple and correctable forecast errors with the annual cycle. In general, these features of

## Annual Mean Forecast Errors of Week 1 &amp; Week 2 GFS Ensemble Mean T2m (Unit: °C)

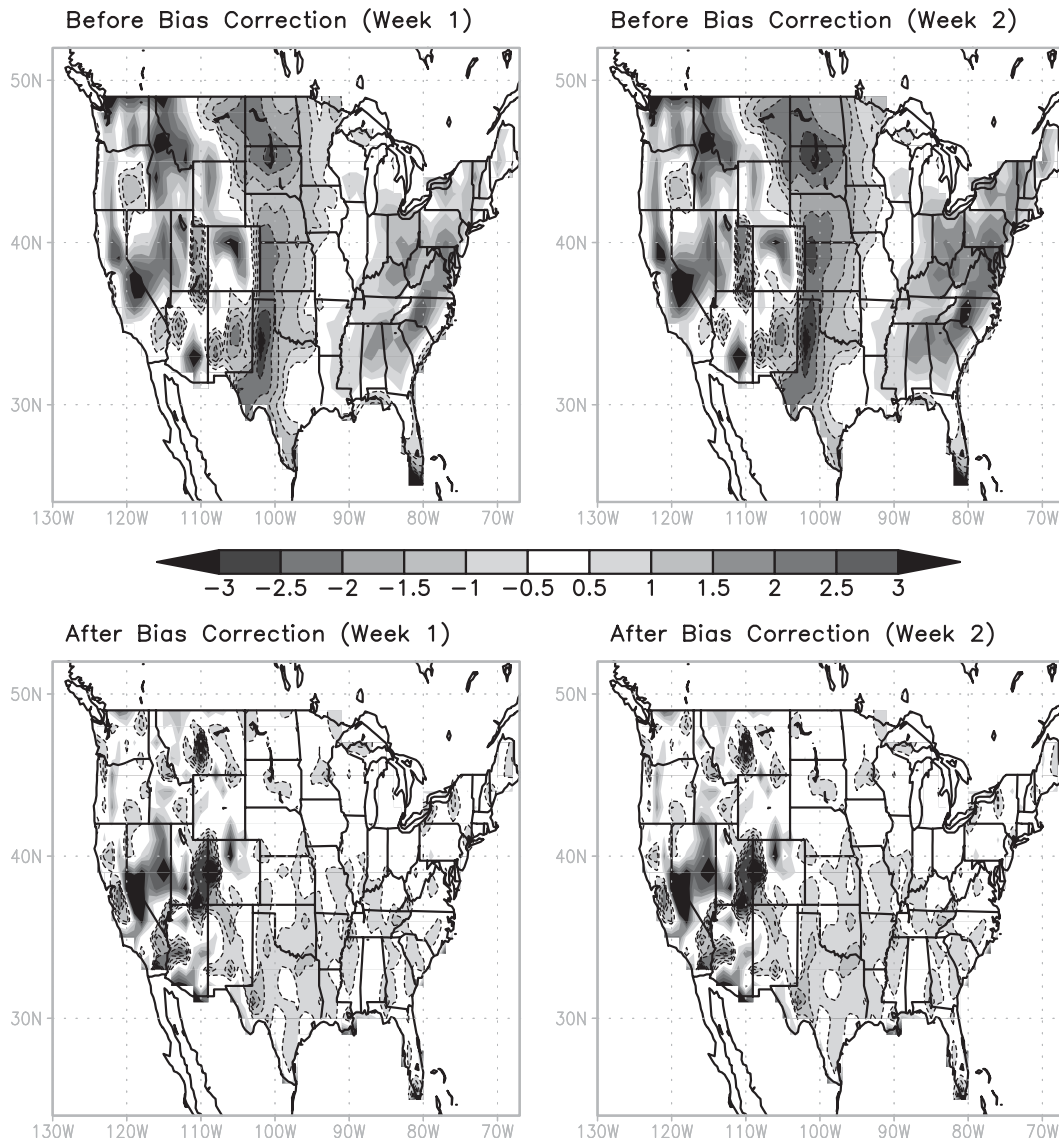


FIG. 7. Annual mean of (left) week-1 and (right) week-2 T2m forecast errors (top) before and (bottom) after bias correction over the United States for the period 1 Apr 2008–31 Mar 2010 (°C). Positive areas show forecasts warmer than the observations. Negative areas (inside the dashed contour) show forecasts colder than the observations.

$P$  and T2m annual mean forecast errors and their interannual variation components are also robust.

### 5. Application of the GEFS ensemble forecasts: Soil moisture outlook

The bias-corrected week-1 and week-2 GEFS ensemble mean  $P$  and T2m forecasts are used to drive the CPC leaky bucket land surface hydrological model forward up to 2 weeks over the United States. Obviously,

the prediction skill of the land surface soil moisture crucially depends on the quality of the forecasted  $P$  and T2m inputs. Because there is very little ground truth to be used, all land surface initial conditions and verification datasets are generated by running the CPC leaky bucket model forced with observed daily  $P$  and T2m, that is, after the fact.

Like sea surface temperature, land surface soil moisture is one of the most important lower boundary conditions for the atmosphere and it also has very high persistence (or memory). So one interesting question

EOF analysis of Week1 GFS Ensemble mean T2m Bias

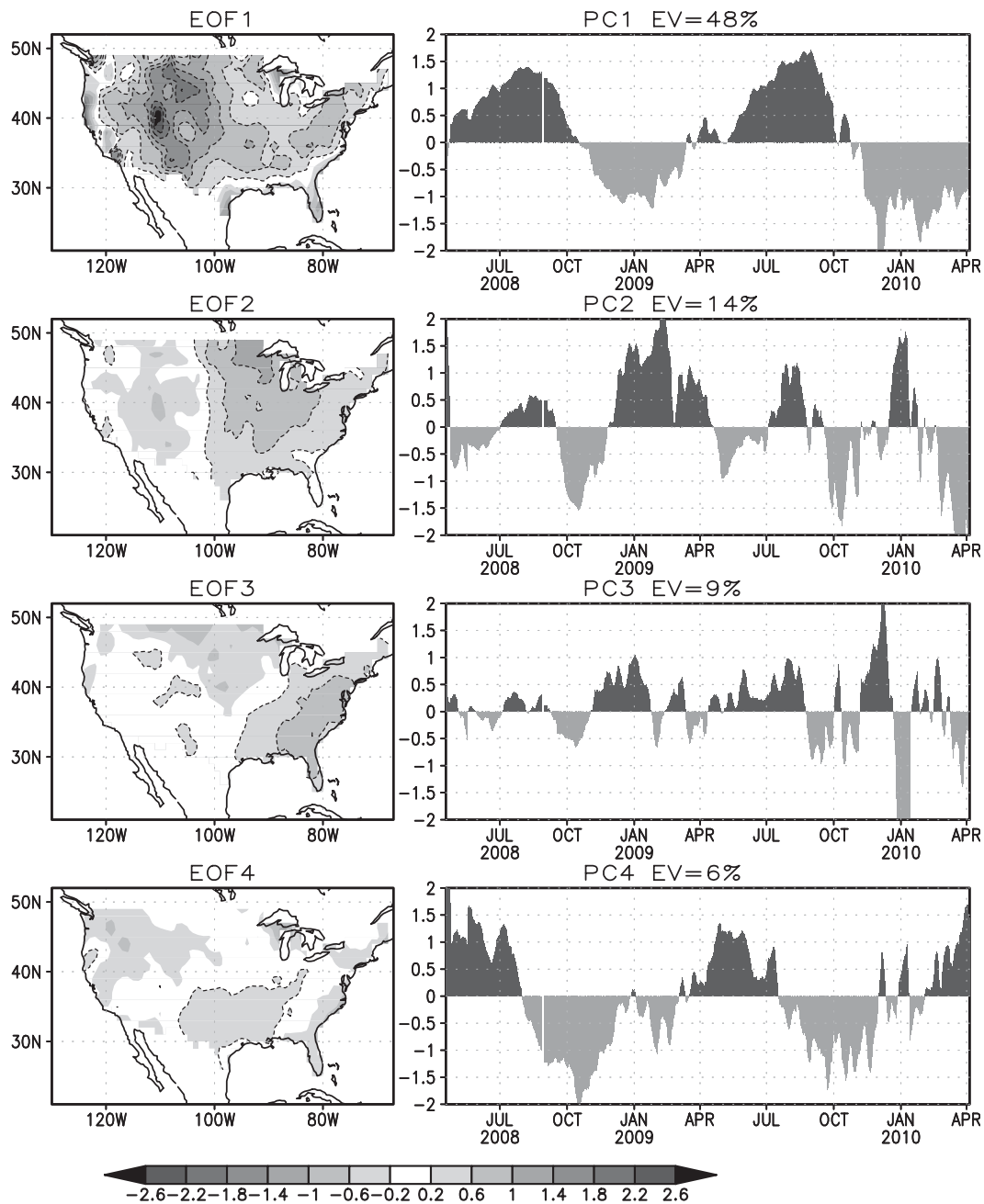


FIG. 8. (left) EOF patterns (scaled by the RMS values of the associated PCs, with negative values inside the dashed contour) and (right) their PCs (normalized to unit variance) of the week-2 T2m forecast errors over the United States for the period 1 Apr 2008–31 Mar 2010.

(and an old “standard” in meteorology) is: Can the soil moisture “dynamical” outlooks (forced with the bias-corrected GEFS week-1 and week-2 ensemble mean  $P$  and T2m forecasts) beat persistence of the initial states? For most land surface models, the land surface hydrological budget can be represented as

$$dW/dt = P - E - R = F \quad \text{or} \quad (4)$$

$$W(t + 1) = W(t) + F, \quad (5)$$

where  $W$  is soil moisture and  $P$ ,  $E$ , and  $R$  are precipitation, evaporation, and total runoff, respectively.

Week-2 Forecasted Soil Moisture Anomaly Correlation – Persistence (x100) (2004–2009)

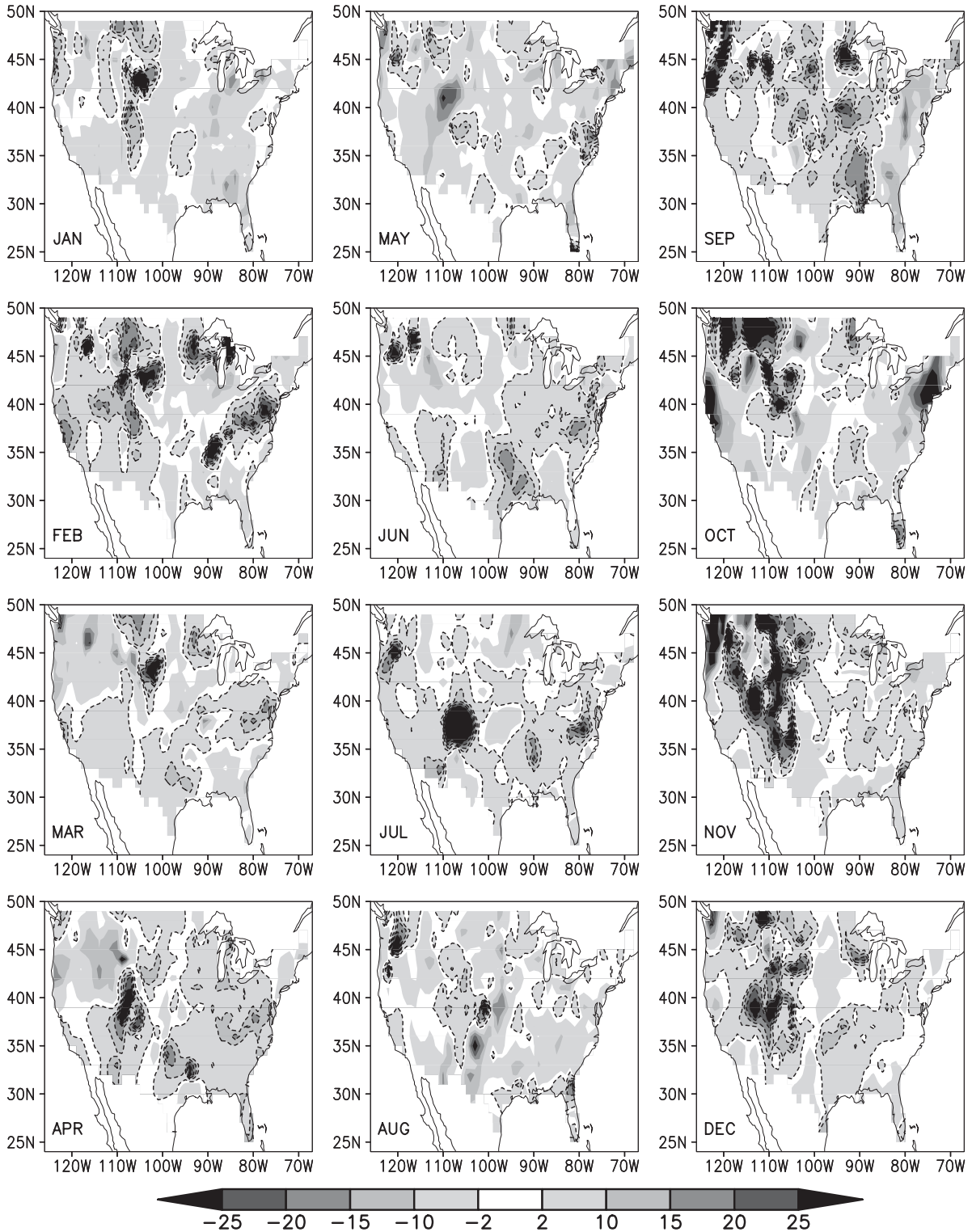


FIG. 9. The temporal anomaly correlation of the GFS week-2 soil moisture forecast minus its persistence, for all 12 months during the period 2004–09. Each pixel shows a difference based on about 180 cases (6 yr, ~30 forecasts per month). Positive regions denote that the forecast is better than persistence. Negative regions (dashed contours) mean the forecast cannot beat persistence.

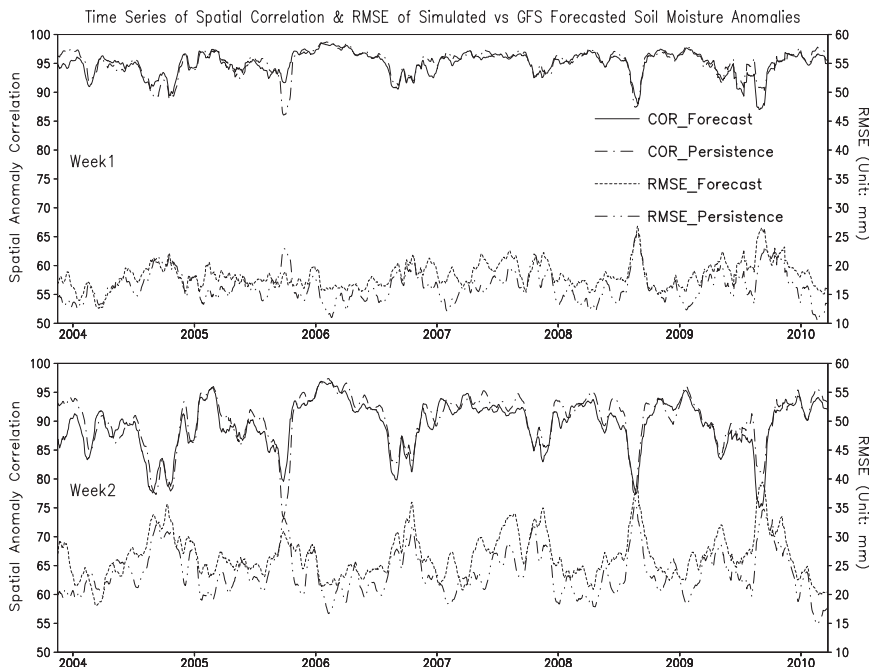


FIG. 10. The time series of 30-day running-mean spatial correlation ( $\times 100$ ) and RMSEs from (top) GFS week-1 and (bottom) week-2 forecasted soil moisture anomalies (solid and dashed) and persistence of soil moisture anomalies (dotted-dashed and dotted-dotted-dashed) over the United States for the period 1 Nov 2003–20 Jun 2009. Units are dimensionless for the spatial correlation and in mm for RMSE.

From Eq. (5) it is clear that if the  $F$  term does not have sufficient skill, the GEFS dynamical forecasts will lose against persistence (i.e., when  $F = 0$ ). So the question about persistence corresponds to the question of  $F = P - E - R$  having (sufficient or useful) skill. Figure 9 displays the spatiotemporal distribution of the correlation of the daily week-2 forecasted soil moisture anomalies minus the persistence of the soil moisture anomalies (i.e., anomalies from the GEFS ensemble mean forecasts forced run against anomalies from the observation forced run) for 12 months for the period of 1 January 2004–31 December 2009 (total 6 yr of daily forecasts). So for each month it has about total  $6 \times 30 = 180$  day daily records. Regions with positive values mean that the forecast beats persistence. In general, the GEFS shows some useful forecast skill over the west coast region, the southeastern United States, and Texas, but constantly (except May) loses against persistence over the Rocky Mountain regions, which seriously degenerates the U.S. overall performance of the GEFS. It indicates indirectly that the GEFS week-1 and week-2 ensemble mean  $P$  (or  $P - E - R$  rather, but this is dominated by  $P$ ) forecast skill is much poorer in these areas.

Figure 10 depicts the time evolution of the forecast skill and its persistence of week-1 and week-2 soil moisture anomalies averaged over the United States from late

2003 to middle April of 2010. If the GEFS was perfect (i.e., if it can reproduce the observed  $P$  and T2m on a 2-week time scale), the spatial correlation of the soil moisture forecasts will be one and its RMSE will be zero. In general, both forecast skill and persistence reach their lowest values (least predictable time) around September, when soil moisture is in its driest season climatologically in the year. Since the persistence of soil moisture is very high, the overall GEFS dynamical soil moisture forecasts beat persistence only by a very small margin in week 1 and lose to persistence in week 2 in terms of their spatial correlation over the United States. In terms of RMSE, the GEFS dynamical forecast loses to persistence in both weeks 1 and 2.

### 6. Summary

In this paper, a simple operationally feasible bias correction method was used to correct the NCEP GEFS ensemble mean week-1 and week-2  $P$  and T2m forecasts on a daily basis. The results show the bias-corrected forecast skill (i.e., anomaly correlation) of the GEFS week-1 and week-2 ensemble mean  $P$  and T2m has large day-to-day fluctuations but a clear seasonal cycle with better scores in winter. The bias-corrected forecasts, in general, are better than the raw forecasts and the degree

of improvement is regional and time-of-year dependent. Overall, the levels of bias-corrected week-1 and week-2 ensemble mean  $P$  and T2m forecast skill are still only modest for the current GEFS.

Ideally, if the forecast errors were constant, they could be easily removed by subtracting an operationally available estimate. If they are not constant, it would be desirable that their spatio-temporal structures be simple so that it would not be difficult to partly remove them. The annual mean of the forecast errors from both week-1 and week-2 GEFS ensemble mean  $P$  and T2m forecasts presents relatively large-scale spatial structures. This portion of the forecast errors was found to be the easiest part to remove. Surprisingly, the variation parts of the GEFS ensemble mean  $P$  and T2m forecast errors are dominated by low frequency (annual and semiannual cycles) and also have relatively large-scale spatial patterns. Part of these forecast errors is removable as well. The effectiveness of the bias correction is time and space dependent. Although the reasons for the above unexpected large-scale and slowly varying forecast errors are not clear at the present time; these bias features may provide some indications or hints to model developers looking to improve the GFS.

The dynamical soil moisture forecasts (i.e., land model forced with the bias-corrected GEFS week-1 and week-2 ensemble mean  $P$  and T2m) have very high skill because of the high persistence of soil moisture, but our results indicate that in general the current GEFS is barely good enough to beat soil moisture persistence (which is very high) over the United States. The inability to outperform the persistence by a noteworthy margin relates mainly to the skill of the forecasted week-1 and week-2  $P$  not being above the threshold (i.e., anomaly correlation  $>0.5$  is required).

*Acknowledgments.* The authors thank Jae Schemm for providing NCEP GFS week-1 and week-2 ensemble mean raw forecast data and John Janowiak for helping with bias correction. The authors are grateful to Doug Lecomte for continuously monitoring the forecasts in real time and invaluable comments. We also would like to thank Jin Huang who initiated the soil moisture forecasts and two anonymous reviewers for their constructive comments. This project was supported by CPPA Grant GC08-292.

## REFERENCES

- Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. Wayne 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.
- Dirmeyer, P., 2000: Using a global soil wetness data set to improve seasonal climate simulation. *J. Climate*, **13**, 2900–2922.
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.
- Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: Reforecasts: An important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33–46.
- Higgins, R. W., J. E. Janowiak, and Y.-P. Yao, 1996: A gridded hourly precipitation data base for the United States (1963–1993). *NCEP/Climate Prediction Center Atlas 1*, NOAA, 46 pp.
- , V. Silva, V. Kousky, and W. Shi, 2008: Comparison of daily precipitation statistics for the United States in observations and in the NCEP Climate Forecast System. *J. Climate*, **21**, 5993–6014.
- Huang, J., H. M. Van den Dool, and K. P. Georgakakos, 1996: Analysis of model-calculated soil moisture over the United States (1931–1993) and applications to long-range temperature forecasts. *J. Climate*, **9**, 1350–1362.
- Janowiak, J. E., G. D. Bell, and M. Chelliah, 1999: A gridded data base of daily temperature maxima and minima for the conterminous United States: 1948–1993. *NCEP/Climate Prediction Center Atlas 6*, NOAA, 50 pp.
- Kanamitsu, M., C. Lu, J. Schemm, and W. Ebisuzaki, 2003: The predictability of soil moisture and near-surface temperature in hindcasts of the NCEP Seasonal Forecast Model. *J. Climate*, **16**, 510–521.
- Koster, R. D., M. J. Suarez, R. W. Higgins, and H. M. Van den Dool, 2003: Observational evidence that soil moisture variations affect precipitation. *Geophys. Res. Lett.*, **30**, 1241, doi:10.1029/2002GL016571.
- Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517.
- Toth, Z., E. Kalnay, S. M. Tracton, R. Wobus, and J. Irwin, 1997: A synoptic evaluation of the NCEP ensemble. *Wea. Forecasting*, **12**, 140–153.
- Van den Dool, H., 2007: *Empirical Methods in Short-Term Climate Prediction*. Oxford University Press, 215 pp.
- , J. Huang, and Y. Fan, 2003: Performance and analysis of the constructed analogue method applied to U.S. soil moisture over 1981–2001. *J. Geophys. Res.*, **108**, 8617, doi:10.1029/2002JD003114.
- Wang, W., and P. Xie, 2007: A multiplatform-merged (MPM) SST analysis. *J. Climate*, **20**, 1662–1679.
- Zhang, H., and C. S. Frederikson, 2003: Local and nonlocal impacts of soil moisture initialization on AGCM seasonal forecasts: A model sensitivity study. *J. Climate*, **16**, 2117–2137.