Does An ENSO-Conditional Skill Mask Improve Seasonal Predictions?

KATHY PEGION

Cooperative Institute for Research in Environmental Sciences, University of Colorado, and Physical Sciences Division, NOAA/Earth System Research Laboratory, Boulder, Colorado

ARUN KUMAR

Climate Prediction Center, National Oceanic and Atmospheric Administration/National Centers for Environmental Prediction, College Park, Maryland

(Manuscript received 30 October 2012, in final form 25 July 2013)

ABSTRACT

The National Centers for Environmental Prediction Climate Prediction Center uses statistical tools together with the Climate Forecast System (CFS) to produce forecasts for seasonal outlooks of U.S. temperature and precipitation. They are combined using an optimal weighting procedure that depends on a skill mask consisting of the average historical forecast skill of each tool. However, it is likely that skill during El Niño–Southern Oscillation events is higher and the use of this information in developing forecasts could lead to improved seasonal predictions. This study explores the potential to improve the skill of seasonal predictions by developing an ENSO-conditional skill mask. The conditional masks are developed in a perfect-model framework using the CFS version 2 hindcasts and two indices of ENSO. The skill of the indices in forecasting variations in conditional skill is evaluated. The ENSO-conditional skill masks provide improvements in correlation skill over the unconditional mask when averaged over the globe. The masks are applied to tercile forecasts of seasonal temperature and precipitation during the spring and forecasts are verified in a perfect-model context. Application of the conditional masks to tercile forecasts results in modified Heidke skill scores of more than 10% less than using the average mask for temperature and little difference in skill for precipitation. This is attributed to the larger number of equal chances forecasts when using the conditional masks, particularly for temperature. For precipitation, the skill predicted by the average and conditional masks is frequently below 0.3, leading to low skill regardless of which mask is used.

1. Introduction

The National Centers for Environmental Prediction (NCEP) Climate Prediction Center (CPC) issues seasonal outlooks for temperature and precipitation indicating the probability of above, below, or normal conditions over the United States for lead times of 0.5–12.5 months. The official seasonal outlooks are based on an objective combination of four primary forecast tools along with forecasts for El Niño–Southern Oscillation (ENSO) and corresponding ENSO composites, discussions with partners via conference calls, and subjective input by the forecaster (O’Lenic et al. 2008). The primary objective forecast tools consist of optimal climate normals, canonical correlation analysis, screening multiple linear regression, and the Climate Forecast System (CFS). These tools are combined using an optimal weighting procedure and referred to as the consolidation (O’Lenic et al. 2008; Unger et al. 2009). When combining these tools, a skill mask is applied based on the average historical forecast skill of each tool (O’Lenic et al. 2008; van den Dool 2007). At locations where the average historical skill, in terms of anomaly correlation, for a given season and lead time is below 0.3, the forecast from that tool is not included in the consolidation and the forecaster is given information to indicate that the forecast is not likely to be skillful (O’Lenic et al. 2008).

It is well known that ENSO impacts temperature and precipitation over the United States (Ropelewski and Halpert 1986), which leads to higher seasonal prediction skill of U.S. temperature and precipitation during
ENSO events than during non-ENSO events, particularly during winter and early spring (Barnett and Preisendorfer 1987; Quan et al. 2006; Livezey and Smith 1999; Peng et al. 2012; Livezey and Timofeyeva 2008). Furthermore, since atmospheric predictability is a function of signal and noise and the atmospheric noise is relatively constant for ENSO versus non-ENSO events (Peng and Kumar 2005; Kumar et al. 2000; Kumar and Hoerling 1998), the year-to-year variations in forecast skill of atmospheric anomalies associated with ENSO should to first-order be a function of the strength of the signal, which in turn is likely to be a function of the amplitude of ENSO. This is demonstrated in Kumar and Hoerling (1997), who show that the amplitude of the signal in the extratropics (i.e., 500-hPa heights) is a linear function of the strength of ENSO-related SSTs, based on simulations by an atmospheric general circulation model. However, this assumption is not without controversy, e.g., Sardeshmukh et al. (2000). We will show, however, that this assumption is generally true for our study.

Based on this knowledge and the episodic nature of ENSO, we expect that forecast skill is likely to be higher than the average skill during ENSO events and linearly related to the strength of the ENSO SST anomalies. As a consequence, using an average skill measure would underestimate the forecast skill and potentially exclude conveying valuable information to the users, particularly for strong ENSO events. On the other hand, during ENSO neutral conditions, and in the absence of signals from other predictors (e.g., snow cover, soil moisture, etc.), the forecast skill may be below average. In this case, the use of an average skill mask would overestimate forecast skill and likely include forecast information over regions where there is no skill.

This project investigates the potential for improving the skill of seasonal predictions and forecast information delivery by exploring ENSO as a predictor of forecast skill and developing skill masks conditional on two definitions of ENSO. This analysis is done in the context of a perfect-model framework (section 3). The ENSO-based skill masks are applied to seasonal tercile forecasts of U.S. temperature and precipitation. The skill of seasonal forecasts made using the average mask and ENSO-based masks are compared to determine if the ENSO-conditional mask provides added benefit (section 4).

2. Data and analysis methods

a. CFSv2 retrospective forecasts

The NCEP/CFS, version 2 (CFSv2), retrospective forecasts (CFSRR; S. Saha 2010, personal communication) provide a large suite of historical forecasts to explore the relationship between ENSO and seasonal forecast skill and the potential for creating an ENSO-conditional skill mask. These reforecasts were made by NCEP for lead times up to 9 months over the years 1982–2009, with four initializations (0000, 0600, 1200, and 1800 UTC) every fifth day. Ensembles are generated by using the 24 forecast initializations per month (N = 24), with the exception of the month of November, which has 28 initializations (N = 28). No explicit perturbations are added to the initial conditions.

We focus on the 0- and 3-month lead forecasts. Forecast leads are defined following CPC’s seasonal forecast operations. For example, a forecast made by CPC on the third Thursday in February, valid for February–April is initialized throughout the month of January and is considered the 0-month lead. A 3-month lead forecast is valid for May–July.

b. Perfect-model skill

The potential for creating a conditional skill mask is investigated in a perfect-model framework. Since sampling issues make it difficult to develop a robust skill mask using observations, we develop a conditional skill mask with the large ensemble available from the CFSv2 retrospective forecasts, in a perfect-model framework, and then apply it to actual predictions.

In the perfect-model framework, one ensemble member is withheld as the “truth” and the skill of the other ensemble members in “predicting” the withheld member is calculated. This is repeated over all ensemble members and the average skill over the ensemble members is taken as the perfect-model forecast skill in which variations with ENSO are analyzed.

c. Average unconditional skill

We calculate two types of perfect-model skill. The first is the unconditional or average skill over all, $M = 28$ yr. This is the same skill calculation used to produce the skill mask used by CPC, except that we are applying it here in a perfect-model context, meaning that we verify the model against the withheld ensemble member rather than actual observations.

Following (Kumar 2007, 2009), we calculate the unconditional skill, at each grid point, by first defining anomalies. Let $X_{i,a}$ be the seasonal mean for a given season, ensemble member $i$, and year $a$. Anomalies are computed by removing the model climatology calculated over the remaining years and ensemble members. Let the withheld ensemble member, representing the observed anomalies be $O_{i,a} = X'_{i,a}$, where the prime denotes anomalies. Then the forecast is the ensemble mean calculated over the remaining ensemble members, given by
The anomaly correlation for the unconditional skill averaged over all $M$ years can be calculated and the process can be repeated and averaged over all $N$ ensemble members. The average, perfect-model anomaly correlation is given by

$$AC = \frac{\langle F'_{t,a} O'_{t,a} \rangle_{MN}}{[\langle F'^2 \rangle_{MN} \langle O'^2 \rangle_{MN}]^{1/2}},$$

where brackets denote averaging.

The unconditional skill for 0-lead, 2-m temperature and precipitation from the CFSv2 reforecasts is shown in Fig. 1. This figure shows that the average skill by the CFSv2 in predicting its own 2-m temperature forecast at 0-month lead is greater than 0.3 over most global land regions in all four seasons. In the winter and fall over the United States, the average skill bears much resemblance in pattern to the ENSO + trend composites of Higgins et al. (2000) and found on the CPC website (http://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/composites/), presumably because time-averaged skill is an integration of the response to ENSO and the trend, which are the largest signals. However, since we are
using a perfect-model calculation, correlations are much larger than for actual forecasts. This average skill will be used as the baseline skill mask that the ENSO-conditional skill masks, developed in this study, are expected to improve upon.

d. Conditional year-to-year skill

The second skill calculation is the conditional, or year-to-year skill. This is the forecast skill for individual seasons that we are attempting to capture with a skill mask. Conditional skill is calculated as the expected anomaly correlation based on the signal-to-noise ratio, where the signal is the average ensemble mean, the noise is the average ensemble spread, and \( N = 24 \) ensemble members (Kumar and Hoerling 2000; Kumar 2009, Sardeshmukh et al. 2000). Conditional skill is mathematically equivalent to the calculation of the unconditional skill, but calculated for each year separately (Kumar 2007, 2009).

Figure 2 provides an example of how the perfect-model skill can vary from year to year and how it can be above or below the average skill. The top panel shows the fraction of land area over the United States where the conditional, 2-m temperature skill is greater than 0.5. At times more than 80% of the United States has AC > 0.5. At other times, less than 10% of the United States has skill above 0.5. The middle panel shows the same calculation for the unconditional skill, which varies only by season. In the bottom panel, the ratio of the conditional to unconditional skill is shown. Where these values are greater (less) than 1, the fraction of land area with conditional skill above 0.5 is larger (smaller) than average. Our goal is to develop a conditional skill mask that can predict these variations in skill to improve the quality of seasonal predictions.

3. ENSO as a predictor of perfect-model forecast skill

We evaluate ENSO as a predictor of variations in forecast skill by exploring the linear relationship between two ENSO indices and perfect-model forecast skill of 2-m temperature and precipitation over the United States. The first index is Niño-3.4, a commonly used index of ENSO (Trenberth 1997). The index used by CPC in preparing operational seasonal forecasts, the oceanic Niño index, is a 3-month running mean of Niño-3.4 (Kousky and Higgins 2007). The second index is a seasonal EOF-based index. Both indices, and their relationship with perfect-model forecast skill, are described below.

a. Niño-3.4

A regression model to predict the anomaly correlation using the absolute value of the Niño-3.4 index as a predictor is created. Niño-3.4 is calculated as the simultaneous ensemble mean SST anomalies, predicted by the CFSv2 reforecasts, averaged over 5°S–5°N, 120°–170°W. The predictand is the anomaly correlation and the predictor is the absolute value of the simultaneous Niño-3.4 index, normalized by its standard deviation. The regression coefficient represents the relationship between the strength of ENSO and anomaly correlation. The regression coefficient is shown in Fig. 3 for 0-lead forecasts of 2-m temperature and precipitation, by season. Since the Niño-3.4 index has been normalized, regression coefficients are in units of correlation per factor of standard deviation. For example, a regression coefficient of 0.2 means the forecast anomaly correlation is increased (decreased) by 0.2, when the Niño-3.4 index is increased (decreased) by a factor of 1 standard deviation. Positive regression coefficients indicate that skill is higher during warm and cold ENSO events than
during ENSO neutral times. The regression coefficient indicates a strong relationship between ENSO strength and skill during the winter and spring months in both temperature and precipitation consistent with known ENSO teleconnections over the United States (e.g., Ropelewski and Halpert 1986) and globe (Ropelewski and Halpert 1987). In the model, the strongest relationship occurs during March–May.

There are also unexpected negative regression coefficients between ENSO and skill in some regions, indicating that the skill is higher during ENSO neutral cases in these regions and seasons. Particularly, for 2-m temperature there are large regions of negative regression coefficients in several seasons. There may be signals, such as land surface or snow cover, that at times are larger in amplitude and opposite the ENSO signal.

b. Seasonal EOF-based ENSO indices

There is evidence that the Niño-3.4 index may not properly represent the teleconnections associated with the decay of ENSO over the United States because the ENSO-related SST anomalies are not maximized in the Niño-3.4 region during this time of the year (Wang et al. 2012). Furthermore, the Niño-3.4 index may not be able to sufficiently capture ENSO teleconnections associated with events that are maximized in the central Pacific.
(Larkin and Harrison 2005). Although other ENSO indices that are fixed in space, such as Niño-3 or Niño-4 may seem like good alternatives, they may not be able to capture the times when the ENSO SST anomalies are maximized in other parts of the tropical Pacific. Therefore, we consider an alternate ENSO index based on seasonal EOFs. By developing the index this way, the seasonal variation in the location of SST anomalies can be captured, as well as the variability in ENSO-related SST anomalies across the tropical Pacific.

We perform a seasonal empirical orthogonal function (EOF) analysis of 0-lead, ensemble mean, tropical (30°S–30°N) SST anomalies from the CFSv2. This calculation is done by performing the EOF analysis, independently for each of the 12 sliding seasons. The three leading EOF patterns are consistent with those found in observations (Ashok et al. 2007; Lim et al. 2009; Wang and Hendon 2007) and other modeling studies (Hoerling and Kumar 2002; Kumar et al. 2005). The leading pattern represents 41%, 34%, 54%, and 63% of the variance for December–February (DJF), March–May (MAM), June–August (JJA), and September–November (SON), respectively and is consistent with well-known ENSO SST anomalies (e.g., Rasmussen and Carpenter 1982; Fig. 4). Although Fig. 4 shows only the DJF EOFs, the spatial pattern is similar for all seasons, but varies in amplitude and location of maximum SST anomalies, consistent with the seasonal cycle of ENSO. The second leading EOF represents 20% or less of the variance, depending on season, and appears to be related to a warming trend in tropical SSTs, over the 28-yr period, in the CFSv2. It is unknown whether this is related to a long-term warming trend associated with increasing CO₂, which is included in the CFSv2, or decadal variability, since we have only a 28-yr record. The corresponding EOF pattern and PC time series has very little variability between seasons. The third leading EOF represents less than 15% of the variance for all seasons and appears to be associated with warming and cooling near the coast of South America for some seasons and resembles the central Pacific or “El Nino Modoki” pattern (Ashok et al. 2007; Wang and Hendon 2007, Lim et al. 2009), although in the CFSv2, the percent variance explained by this pattern is much less than observed. The spatial pattern and time series for this EOF appear to vary from season to season.

Since the spatial pattern of the leading EOF appears to be related to ENSO for all seasons we define the seasonally dependent leading principal component time series (PC1) as an alternate ENSO index and consider the regression relationship between PC1 and anomaly correlation of temperature and precipitation (Fig. 5). Comparison of the correlation coefficient between ENSO index and skill for the Niño-3.4 index (Fig. 3) and PC1 index (Fig. 5) indicate that the two indices are most similar in global pattern in March–May and September–November for temperature as indicated by pattern correlations of 0.89 and 0.97, respectively. Overall, the root-mean-square (RMS) differences between the two indices are small, ranging from 0.02 in September–November to 0.09 in December–February. For precipitation, patterns are similar, except for December–February as indicated by pattern correlations in Fig. 5. Differences in amplitude are small, as indicated by the RMS differences with the largest difference occurring in December–February when the relationship between Niño-3.4 and skill is stronger than the relationship between PC1 and skill. Given that there are differences in the relationship between ENSO and skill for the two indices, depending on season, we will test the results using both indices when developing the regression models (section 3). Additionally, since each of the three leading PCs are by definition independent, we will explore the impact of including PCs 2 and 3 in the development of the regression models in order to capture the impact of the trend/decadal contribution and the inclusion of the central Pacific ENSO in predicting forecast skill (section 3).

The year-to-year perfect-model skill that we wish to model, using an ENSO index, is a function of the signal and noise. To understand under what signal-to-noise relationship the positive and negative regressions between ENSO and skill arise, we evaluate the relationship between signal and noise for ENSO and non-ENSO events over the United States (Fig. 6). ENSO events are defined to occur when the PC1 index is above or below 1σ for any season. All other times are designated as non-ENSO events. Both the signal and noise values have been normalized by the total variance at each grid point. The colors indicate the regression coefficients from Fig. 5 and the black lines indicate the average signal versus noise over all ENSO or non-ENSO cases shown in each panel.

During ENSO events, the largely positive regression coefficients occur where the signal to noise ratio is much larger than average while negative regression coefficients generally occur where the signal to noise ratio is near or below average. However, the largest negative regression values occur for situations where the signal is very low (~0.2) relative to the noise (~0.8). During ENSO neutral months, negative regressions occur when the signal and noise are about average (black line) and positive regressions occur where the signal to noise ratio is below average. Similar behavior in signal to noise is seen for the relationship between the PC1 index and U.S. precipitation skill (Fig. 6, bottom panels). Like
temperature, precipitation regression values during ENSO events appear to be very closely related to the strength of the signal. These results support the earlier studies underlying atmospheric predictability associated with ENSO teleconnections—that the predictability during ENSO cases is associated mainly with the larger signal and that the noise is relatively unchanged. It also demonstrates that the negative regressions occur primarily when the signal is very weak.

c. Comparison of predictions of forecast skill

Now that we have identified two predictors of seasonal, perfect-model skill, based on ENSO, we demonstrate how well they fit to the year-to-year temperature and precipitation perfect-model forecast skill. Figure 7 (top panels) shows the fit, as a function of initial month, in terms of root-mean-square error (RMSE) over all global land points, for the average, or unconditional skill.
Despite the differences between the two ENSO indices described in the previous section, they are nearly indistinguishable in terms of their fit over the global land regions for 0- and 3-month lead forecasts of temperature and precipitation. However, they are both better fits to the skill than using the average. Including additional principal components as predictors (Fig. 7, red and blue dashed lines) improves the fit slightly, but since the additional PCs explain only a small percent of the variance relative to PC1, the differences are small.

Next we assess the ability of the average and regression models to predict perfect-model variations in conditional forecast skill in terms of the RMSE. This is done by withholding a year, constructing the regression model, and forecasting the conditional skill for the year withheld. This is repeated over all years. The same “jack-knifed” forecast calculation is performed using the average model (i.e., the unconditional skill). Figure 7 (bottom panels) shows the RMSE for 0- and 3-month lead perfect-model, seasonal forecasts of temperature (left) and precipitation (right) skill over the global land

![Figure 5: Regression coefficients between the simultaneous absolute value of normalized leading PC of 0-lead tropical SST anomalies and 2-m temperature conditional anomaly correlation. The contour interval is 0.05. The values are a correlation/factor of the standard deviation.](image-url)
regions. At all leads and for both temperature and precipitation, both the Niño-3.4 and PC1 masks are nearly indistinguishable in their ability to forecast the perfect-model skill. These ENSO-conditional forecasts of skill are better predictors than using average skill. Including PC2 and PC3 as predictors (red and blue dashed lines) produces slightly larger errors at both leads and for both variables than using PC1 or Niño-3.4 alone. Since the RMSE is calculated over all global land points, it is possible that there are regional differences in the ability to emulate forecast skill that are not shown here.
To demonstrate the ability of the regression models to predict the skill of U.S. temperature and precipitation we calculate the RMSE in predicting the forecast skill during ENSO and neutral cases at each grid point over the United States. Since the strongest relationship between the ENSO predictors and temperature and precipitation skill over the United States occurs in the spring (forecasts valid March–May), we will focus on comparing the skill during this season. The errors in predicting the forecast skill for the Niño-3.4 regression model and the average model for March–May are shown in Fig. 8 (top panel) for temperature and Fig. 9 (top panels) for precipitation, with their differences shown in the respective bottom panels. As indicated by these figures, the regression model based on Niño-3.4 better predicts the skill for both ENSO and non-ENSO cases.

**Fig. 7.** Root-mean-square error of (top) model fit and (bottom) model cross-validated skill to conditional anomaly correlation of (left) 2-m temperature and (right) precipitation for (top) 0- and (bottom) 3-month leads for average model (black), PC1 regression model (green dashed), Niño-3.4 regression model (green solid), PC1 + PC2 (blue dashed), and PC1 + PC2 + PC3 (red dashed) regression models as a function of initial month. Units are correlation.
than simply using the average skill to predict forecast skill. The average model has larger errors in predicting the skill during non-ENSO events with errors in correlation >0.5 in the Southwest and Florida for precipitation and the southern United States and Pacific Northwest for temperature. This occurs because the average model over predicts the skill during ENSO neutral times. The regression model using Niño-3.4 is able to mitigate this problem, indicated by the reduction of errors in correlation by more than 0.2. Since the average skill is presumably an integration of the response to ENSO (plus the trend), as previously discussed, the skill during ENSO events is much more similar than during ENSO neutral between the two masks, although the Niño-3.4 regression model has slightly smaller errors than the average mask over parts of the United States in both variables. These results indicate that the Niño-3.4-based regression model is a better predictor of forecast skill than using the average skill over all years primarily because it mitigates the overprediction of skill by the average mask during non-ENSO years.

We now compare the skill of the PC-based regression models to that of the Niño-3.4 model (Fig. 10). The PC1 model is very similar to the Niño-3.4 for both temperature and precipitation. For precipitation, the other PC-based models have larger RMSEs than the Niño-3.4 model with the exception that there is improvement
in predicting skill in California when using the PC-based models. Based on this comparison, we will focus on the Niño-3.4 regression model when applying the conditional skill mask.

4. Application of the ENSO-based skill masks

Seasonal outlooks of temperature and precipitation at NCEP/CPC are issued as tercile probabilities, indicating above, below, or normal conditions (O’Lenic et al. 2008). If the probability of all three conditions is the same or cannot be distinguished, the forecast is for equal chances (EC). Here, we perform a similar forecast procedure to determine the impact of applying the average and ENSO-conditional skill masks to seasonal predictions from the CFSv2 reforecasts.

a. Application of the skill mask

We first apply the conditional skill masks to tercile forecasts in a perfect-model framework, using the model as truth. The CFSv2 retrospective forecasts are divided into observed (obs) and forecast (fcst) datasets. The obs dataset is determined by randomly choosing one ensemble for each initialization and the fcst dataset is composed of all other ensemble members. The anomalies and tercile boundaries for the two datasets are calculated separately. The forecast probabilities for above, below, and normal terciles are based on the fraction of
Fig. 10. Root-mean-square error differences in predicting anomaly correlation between (left) PC1 and Niño-3.4 regression models, (middle) PC1 + PC2 and Niño-3.4 regression models, and (right) PC1 + PC2 + PC3 and Niño-3.4 regression models for 0- and 3-month lead forecasts of (top) temperature and (bottom) precipitation for forecasts valid in March–May. Units are correlation.
ensemble members forecasting each category. In practice, using the model forecast, equal chances does not occur; however, we assign equal chances for the times when the skill masks forecast the skill to be below 0.3, meaning that either we cannot distinguish between tercile probabilities, or distinguishing them is not expected to lead to a useful level of skill above 0.3. We then quantify the skill of the tercile forecasts using the Heidke skill score (HSS), which is a percent improvement over random forecasts. We calculate the skill, including EC forecasts (sometimes referred to as the modified HSS), following O’Lenic et al. (2008). This skill metric is calculated as

\[ S_n = \frac{C - e + \frac{1}{3} ec}{n - e} \times 100, \]

where \( C \) is the number of correct forecasts, \( n \) is the total number of forecasts, \( e \) is the number of forecasts expected to be correct by chance, and \( ec \) is the number of forecasts for equal chances.

b. Does the ENSO-based mask improve perfect-model skill?

Next, we test the impact of applying the average versus Niño-3.4 conditional masks for forecasts valid during spring. Figure 11 shows the HSS when applied to tercile perfect-model forecasts using the average skill mask (left panels) and the forecasted skill from the Niño-3.4 regression model as a mask (middle panels). The difference between them is shown in the rightmost panels. For temperature (top), the average mask has higher skill scores with a difference of 6%–10% over much of the United States for both 0- and 3-month lead times. For precipitation (bottom panels), there is little difference in skill between the two masks.

Since the ENSO-based regression model is better at predicting the forecast skill than using the average skill (e.g., Figs. 7, 8, and 9), it is surprising that using the ENSO-conditional mask does not result in improved seasonal prediction skill for either temperature or precipitation. This can be understood by comparing the skill (Fig. 11) with the percentage of equal chances forecasts when the two different masks are applied (Fig. 12). At 0-lead times, the average mask has very few EC forecasts for temperature. However, the Niño-3.4 mask has a large percentage of EC forecasts. A similar situation occurs for precipitation at 0- and 3-month lead times. The skill of precipitation is low, with average perfect-model skill between 0.2 and 0.4 over much of the United States at 0-month leads (Fig. 1) and above 0.3 in the southwestern United States at 3-month leads (not shown). As a result, both masks have a large number of EC forecasts (Fig. 12), resulting in similar skill regardless of which mask is used.

c. Application of the skill masks to forecasts

We now apply the skill masks to the CFSv2 forecasts and verify against observations. This differs from the previous section, where we verified the forecasts in a perfect-model framework. By verifying against the observations, we are able to assess the potential for improving actual forecast skill by using the ENSO-conditional skill masks.

The differences in skill of tercile seasonal forecasts of 2-m temperature and precipitation are compared to assess the impact of applying the ENSO-conditional skill masks as opposed to the average skill mask in current operational use at CPC. Figures 13 shows the HSS when using the skill masks for forecasts valid during spring, when verified against NCEP/CPC Global Historical Climate Network/Climate Anomalies Monitoring System (GHCN/CAMS; Fan and van den Dool 2008) for temperature and NCEP/CPC precipitation reconstruction using gauge observations over land (PRECL; Chen et al. 2002) for precipitation. The results are similar to those using perfect-model skill, indicating that when applied to forecasts the average mask is more skillful than the ENSO-conditional mask for temperature and similarly skillful for precipitation.

5. Summary and discussion

We investigate the potential for improving seasonal predictions by developing skill masks conditional upon the amplitude of ENSO. We explore two ENSO indices (Niño-3.4 and a PC-based index) as predictors of perfect-model U.S. seasonal temperature and prediction skill, at 0- and 3-month leads, in a simple linear regression model. We test and compare the RMSE in forecasting anomaly correlation by the regression models and by the average model and found that when the entire globe is considered, the ENSO-conditional skill masks indicate the potential for small improvements in perfect-model forecast skill of temperature and precipitation, depending on season. We apply the Niño-3.4 regression model as a skill mask to tercile predictions of temperature and precipitation and compare the skill of forecasts made using them. Our results indicate that the application of the ENSO-conditional mask does not improve seasonal predictions over the average mask. Therefore, the
FIG. 11. Modified Heidke skill score for 0- and 3-month lead perfect-model forecasts of (top) temperature and (bottom) precipitation when the (left) average mask and (middle) Niño-3.4 mask are applied, and (right) differences between them. Units are %.
Fig. 12. Percentage of equal chances forecasts when the (left) average and (right) Niño-3.4 masks are applied to tercile forecasts of (top) temperature and (bottom) precipitation for 0- and 3-month lead times.
FIG. 13. Modified Heidke skill score for 0- and 3-month lead forecasts of (top) temperature, verified against GHCN/CAMS, and (bottom) precipitation, verified against CAMS/OPI when the (left) average mask and (middle) Niño-3.4 mask are applied, and (right) differences between them. Units are %.
average mask should remain the mask to be used for operational seasonal predictions at CPC. Since the conditional masks are better able to predict forecast skill in terms of RMSE, it is possible that perhaps a different application of the conditional masks could be identified to utilize this information for improving forecasts. This will be the subject of future research.

Previous studies have demonstrated that the global teleconnections associated with ENSO are changing because of changes in the global temperature trends (Kumar et al. 2010; Higgins et al. 2004; Peng et al. 2012). If this is true, then the regression relationships between ENSO and skill, determined from reforecasts made for 1982–2010, may not be realistic for the forecast skill in the future climate. In the current study, we did not directly consider the potential impact of the trend. However, the trend is included indirectly in the SST forecasts from CFSv2. Additionally, the second PC identified from seasonal EOFs of tropical SST appear to contain a trend component, although our 28-yr record is short enough that this trend cannot be separated from decadal variability. Including this second PC in our regression model did not improve the ability to forecast perfect-model forecast skill over PC1, because it only represents a small amount of the variance in the model. This indicates that for a perfect-model forecast from the CFSv2, ENSO appears to be the dominant predictor of forecast skill from tropical SST anomalies.

There is also indication that central Pacific ENSO events may have different global teleconnection patterns than events centered in the eastern Pacific (Larkin and Harrison 2005). We explored this possibility by including the third PC in our PC-based ENSO index. In the CFSv2, this PC represents a very small amount of the variance, much smaller than in observations. As a result, we did not find that including this central Pacific mode in our skill mask lead to improvement in prediction skill over using the PC1 or Niño-3.4 masks.

REFERENCES


