

Long-Lead Seasonal Prediction of China Summer Rainfall Using an EOF–PLS Regression-Based Methodology^{*,†}

WEN XING

College of Oceanic and Atmospheric Sciences, Physical Oceanography Laboratory/Qingdao Collaborative Innovation Center of Marine Science and Technology, Ocean University of China, Qingdao, China

BIN WANG

Department of Atmospheric Sciences, and International Pacific Research Center, University of Hawai'i at Mānoa, Honolulu, Hawaii, and Earth System Modeling Center, Nanjing University of Information Science and Technology, Nanjing, China

SO-YOUNG YIM

Korea Meteorological Administration, Seoul, South Korea

(Manuscript received 23 December 2014, in final form 29 December 2015)

ABSTRACT

Considerable year-to-year variability of summer rainfall exposes China to threats of frequent droughts and floods. Objective prediction of the summer rainfall anomaly pattern turns out to be very challenging. As shown in the present study, the contemporary state-of-the-art dynamical models' 1-month-lead prediction of China summer rainfall (CSR) anomalies has insignificant skills. Thus, there is an urgent need to explore other ways to improve CSR prediction. The present study proposes a combined empirical orthogonal function (EOF)–partial least squares (PLS) regression method to offer a potential long-lead objective prediction of spatial distribution of CSR anomalies. The essence of the methodology is to use PLS regression to predict the principal component (PC) of the first five leading EOF modes of CSR. The preceding December–January mean surface temperature field [ST; i.e., SST over ocean and 2-m air temperature (T2m) over land] is selected as the predictor field for all five PCs because SST and snow cover, which is reflected by 2-m air temperature, are the most important factors that affect CSR and because the correlation between each mode and ST during winter is higher than in spring. The 4-month-lead forecast models are established by using the data from 1979 to 2004. A 9-yr independent forward-rolling prediction is made for the latest 9 yr (2005–13) as a strict forecast validation. The pattern correlation coefficient skill (0.32) between the observed and the 4-month-lead predicted patterns during the independent forecast period of 2005–13 is significantly higher than the dynamic models' 1-month-lead hindcast skill (0.04), which indicates that the EOF–PLS regression is a useful tool for improving the current seasonal rainfall prediction. Issues related to the EOF–PLS method are also discussed.

* School of Ocean and Earth Science and Technology Publication Number 9563, International Pacific Research Center Publication Number 1165, and Earth System Modeling Center Publication Number 085.

† Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-15-0016.s1>.

Corresponding author address: Wen Xing, Physical Oceanography Laboratory/Qingdao Collaborative Innovation Center of Marine Science and Technology, Ocean University of China, Marine Meteorology, 238 Songling Road, Qing Dao 266100, China.
E-mail: xingwen1022@gmail.com

1. Introduction

The rainy season in China experiences a characteristic northward march starting from southern China in April–May and then retreats across China toward the end of August except for the southernmost part of the country. The climatological summer [June–August (JJA)] mean rainfall over China gradually decreases from southeastern China where the rainfall rate exceeds 8 mm day^{-1} to northwestern China where the rainfall rate is less than 1 mm day^{-1} (Fig. 1a).

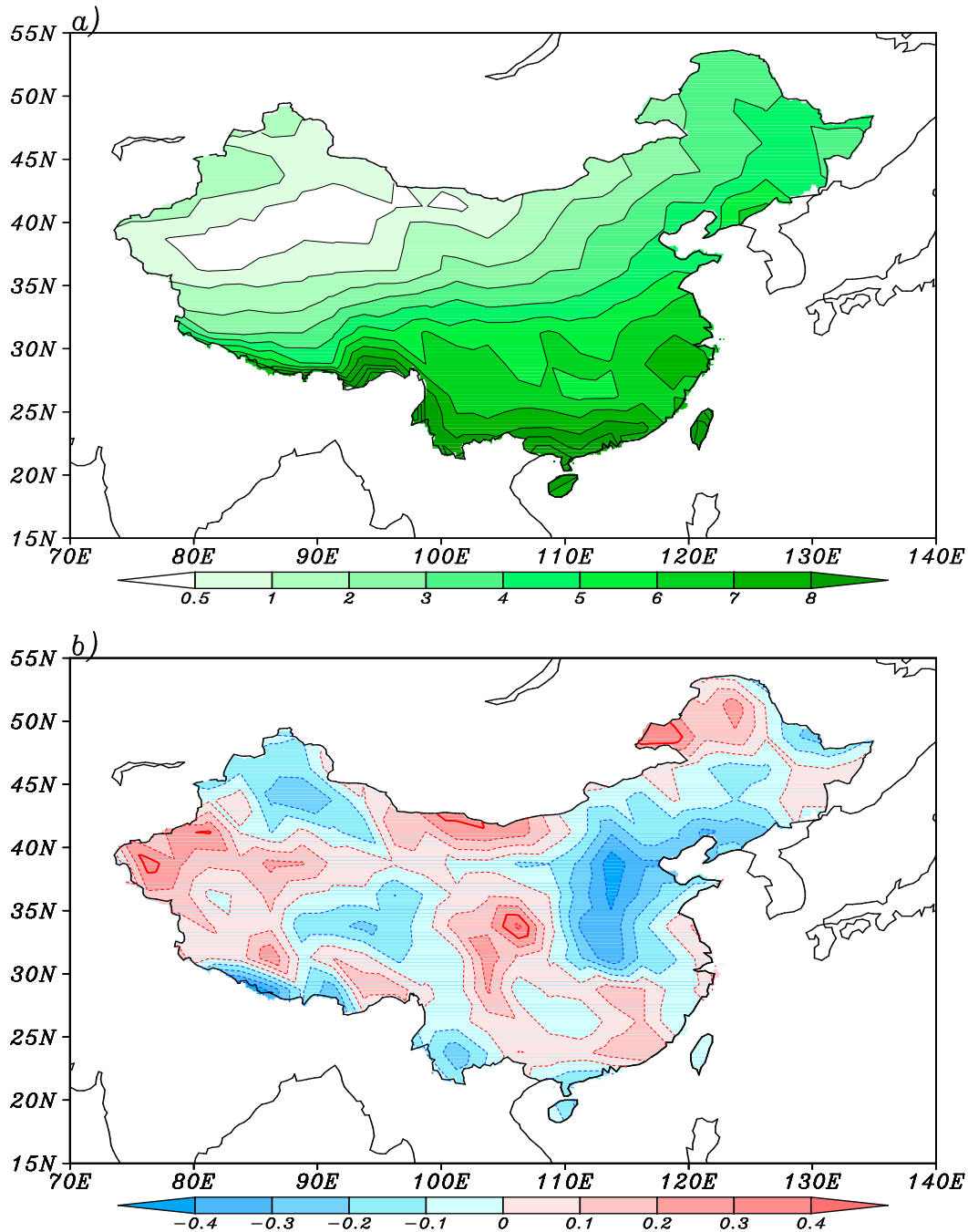


FIG. 1. (a) Climatological mean precipitation rate (mm day^{-1}) averaged for June–August (JJA) from 1979–2013. (b) The temporal correlation coefficient (TCC) skill for JJA precipitation prediction using the nine coupled models' multimodel ensemble initiated from the first day of May for the 32 yr of 1979–2010. The solid red contour is the TCC skill of 0.30 with statistical significance at the 0.1 confidence level. The averaged TCC over China is -0.02 .

Chinese JJA rainfall exhibits prominent interannual variability with complex spatial structures. Eastern and southern China are vulnerable to droughts and floods, which frequently cause devastating agricultural and economic losses (Tao and Chen 1987; Ding 1994; Yuan et al. 2008a,b; Gu et al. 2009a,b). The losses arising from

climate-related disasters have notably increased in the recent years due to the economic boom. For instance, the total economy loss in China due to extreme climate events in 2010 alone reached over RMB500 billion (about USD82 billion), which is the largest since 1990 (Sun et al. 2011).

Sea surface temperature anomalies (SSTA) over tropical Pacific associated with ENSO are a major cause of summer rainfall variability in East Asia (e.g., [Gong and Wang 1999](#); [Wang et al. 2000](#); [Wu and Wang 2002](#)). Besides, the interannual variability of the East and Southeast Asian monsoon is also tied to SSTA over the East China Sea, the Sea of Japan (East Sea), and the South China Sea (SCS) regions ([Lau et al. 2000](#)). The anomalous snow cover and snow depth over Eurasian land and the Tibetan Plateau as well as the Arctic ice concentration all exhibit, to various degrees, association with the eastern China summer rainfall (CSR) on an interannual time scale (e.g., [Chen and Wu 2000](#); [Qian et al. 2003](#); [Wu and Qian 2003](#); [Zhang et al. 2004](#); [Zhao et al. 2007](#); [B. Wu et al. 2009a,b](#); [Liu and Wang 2011](#)).

Prediction of summer rainfall is one of the most challenging tasks in climate sciences. This challenge arises partly from its limited predictability and partly from the coupled climate models' deficiencies in capturing ENSO evolution and teleconnection, as well as atmosphere–ocean–land interaction ([Wang et al. 2008, 2009a](#)).

Although the dynamical models can capture, to a large degree, the leading modes of interannual variability of the Asian monsoon and the ENSO–monsoon relationship, useful skills for rainfall prediction are largely confined to tropical monsoon oceans ([Wang et al. 2008](#)). Over land monsoon areas, especially in the subtropics and mid-latitudes, the current climate models are still unable to adequately predict the mean intensity and seasonal variations ([Luo et al. 2005](#); [Shukla 2007](#); [Wang et al. 2009b](#)).

To see the current status of the dynamical seasonal prediction of the rainfall over China, we have examined hindcast skills of nine state-of-the-art atmosphere–ocean–land coupled models, including the 1) National Centers for Environmental Prediction (NCEP) CFS version 2 ([Saha et al. 2014](#)), 2) Geophysical Fluid Dynamics Laboratory (GFDL) Climate Model version 2.1 ([Delworth et al. 2006](#)), 3) Frontier Research Center for Global Change (FRCGC) SINTEX-F model ([Luo et al. 2005](#)), 4) European Centre for Medium-Range Weather Forecasts (ECMWF) model ([Gregory et al. 2000](#); [Molteni et al. 2011](#)), 5) Instituto Nazionale de Geofisica e Vulcanologia (INGV) model ([Madec et al. 1998](#)), 6) Leibniz Institute of Marine Sciences at Kiel University (IFM-GEOMAR) model ([Jungclauss et al. 2006](#)), 7) Predictive Ocean Atmosphere Model for Australia (POAMA) P24A model ([Zhong et al. 2005](#)), 8) Met Office (UKMO) model ([Roeckner et al. 1996](#); [Marsland et al. 2003](#)), and 9) Centre National de Recherches Météorologiques (CNRM) MF model ([Déqué 2001](#)). These models are collected in the Asian-Pacific Economic Cooperation (APEC) Climate Center (APCC) Climate Prediction and its Application to Society (CliPAS) project. A multimodel ensemble (MME)

hindcast from 1979 to 2010 with 1 May initial conditions was made by simply averaging the nine coupled models' ensemble mean anomalies after removing their own climatology.

[Figure 1b](#) shows the temporal correlation coefficient (TCC) for JJA precipitation prediction at each grid over China by using the nine climate models' MME. As can be seen, the TCC over China barely exceeds the 90% confidence level. The prediction skill is basically insignificant over all of China. The prediction is notoriously poor in central northern China.

Although the causes for CSR anomalies have been discussed, thus far only a limited number of studies have dealt with summer rainfall prediction. [Li and Zeng \(2008\)](#) predicted East Asia summer monsoon (EASM) rainfall based on SST and sea ice concentration. [Z. Wu et al. \(2009\)](#) established an empirical model to predict the EASM strength through a combination of ENSO and spring North Atlantic Oscillation (NAO) data. [Fan et al. \(2012\)](#) developed two statistical prediction schemes including the interannual increment approach to improve the seasonal prediction of the EASM's strength. Recently, a method called predictable mode analysis (PMA) was proposed by [Wang et al. \(2007\)](#). This method integrates empirical analysis, physical interpretation, and hindcast experiments. The hindcasts (retrospective predictions) are made by using physical–empirical models or coupled dynamical models. A series of works have been performed to predict East Asian rainfall anomalies during early (May–June) and peak summer (July–August) respectively by using this method ([Xing et al. 2014](#); [Yim et al. 2014](#)). All the results of these studies have shown that statistical models are more skillful than dynamical models in terms of predicting East Asian summer rainfall. In this study, we focus only on China rainfall during conventional summer season (JJA) in line with traditional seasonal prediction practice.

In the present study, we explore a new statistical forecast method called partial least squares (PLS) regression to predict each principal component (PC). PLS regression was first applied in the field of econometrics by [Wold \(1966\)](#). In recent decades, it has been improved and used in many fields such as computational biology ([Tan et al. 2004](#)) and neuro-imaging ([McIntosh and Lobaugh 2004](#)). For climate purposes, this method has been applied in paleoclimate reconstruction ([Kalela-Brundin 1999](#)), analysis of plant growth days ([McIntosh et al. 2005](#)), diagnosis of the factors that contribute to the variability of geophysical time series ([Smoliak et al. 2010, 2015](#)), exploration of the climate background of anomalous wet and cold winter in southern China ([Zhang et al. 2011](#)), and seasonal prediction of killing-frost frequency in south-central Canada ([Wu et al. 2013](#)).

Our study aims to identify the principal modes of CSR variations and then predict each mode with PLS regression, so it is called the combined EOF–PLS method (section 2). The characteristics of the major modes of CSR variability are discussed in section 3. Section 4 presents the prediction of CSR. Section 5 provides conclusions and discussion.

2. Methodology: The combined EOF–PLS regression method

The PLS is normally used to predict a single index (predictand) using a concurrent or precursory field. To predict the rainfall anomaly patterns, we propose an EOF–PLS method, that is, using observed EOF spatial patterns and PLS-predicted PCs to reconstruct the total anomaly rainfall distribution. EOF analysis of JJA mean precipitation anomalies over China from 1979 to 2013 is first performed to extract the first few principal modes of rainfall variability. The PC of each mode is then predicted by using PLS regression method. Finally, the anomaly prediction field is reconstructed through a linear combination of these orthogonal modes. We call it the combined EOF–PLS regression method.

The mathematical procedure of PLS regression is briefly introduced here. Assume we have a predictand time series of n years (the PC time series of $n = 35$ yr for each mode in the present study), which is denoted by a column vector \mathbf{y}_i ($i = 1, \dots, n$). A proper predictor field is then selected based on physical considerations. Assume that the selected predictor field has J grids with a time series of n years, which can be represented by an $n \times J$ matrix \mathbf{X}_{ij} ($i = 1, \dots, n$ and $j = 1, \dots, J$). Both the predictor matrix and predictand vector are standardized prior to the analysis.

First, a correlation map r_j can be derived by calculating correlation coefficients between the predictand \mathbf{y}_i and the predictor field \mathbf{X}_{ij} at each grid. The correlation map r_j shows how predictor anomalies in different regions are related to the predictand. In this way, the usefulness of predictor at each grid can be readily assessed and significant correlation patterns (regions) can be identified.

Next, the standardized predictor anomalies at each time are weighted by cosine latitude to account for area and projected onto the correlation map r_j to form a time series. This represents a weighted average of predictor anomalies, which is referred to as the first predictor z_1 (Smoliak et al. 2010). To be specific, the predictor z_1 is calculated by multiplying each grid value of the predictor anomalies by the correlation coefficient at the same grid point. The results of the

product at each grid are then summed, forming a weighted average of the predictor anomalies. Therefore, the predictor anomalies from the regions that are highly correlated (positive or negative) with the PCs get the most weight.

Finally, using conventional least squares fitting procedures, z_1 is regressed out of the predictand and the original predictor field, so that one obtains an approximation to \mathbf{y}_i (represented by \mathbf{y}_i') and \mathbf{X}_{ij} (represented by \mathbf{X}_{ij}'), respectively. A residual predictand–predictor field can be derived from the difference between the original predictand–predictor field and the corresponding approximation (i.e., $\mathbf{y}_i - \mathbf{y}_i'$ and $\mathbf{X}_{ij} - \mathbf{X}_{ij}'$). The aforementioned procedure is repeated on the residual matrices to obtain a second predictor time series z_2 , and so on. The optimal number of predictor component is determined by cross validation (Smoliak et al. 2010). To prevent overfitting, only two PLS predictors are retained, as determined through cross validation. A detailed explanation can be found in Smoliak et al. (2015). The predictors Z are mutually orthogonal and maximize the variance explained in \mathbf{y}_i and the correlation between \mathbf{X}_{ij} and \mathbf{y}_i .

Note that this is related to, but differs from, other statistical methods such as least squares, multiple linear regression, and principal component analysis (Martens and Naes 1989; De Jong and Phatak 1997; Phatak and De Jong 1997).

The data used in observational analyses comprise monthly mean precipitation from Global Precipitation Climatology Project (GPCP) version 2.2 (v2.2) datasets and the 160 stations provided by the Chinese Meteorological Data Center, monthly mean SST from NOAA Extended Reconstructed SST (ERSST.v3b; Smith et al. 2008), and monthly mean circulation data and 2-m air temperature from National Centers for Environmental Prediction–U.S. Department of Energy (NCEP–DOE) AMIP-II reanalysis products (Reanalysis-2; Kanamitsu et al. 2002). The data period chosen in this study is from 1979 to 2013.

3. The major modes of CSR

EOF analysis of JJA mean precipitation anomalies over China from 1979 to 2013 is performed to extract the principal modes of rainfall variability. We focus on the first five EOF modes for two reasons. First, these modes that are derived from two different precipitation datasets (GPCP and China station data) show very similar spatial patterns and PCs (Fig. 2; see also Fig. S1 in the supplemental material), while the higher modes have different patterns and PCs, suggesting that those higher modes cannot be distinguished from noise due to the

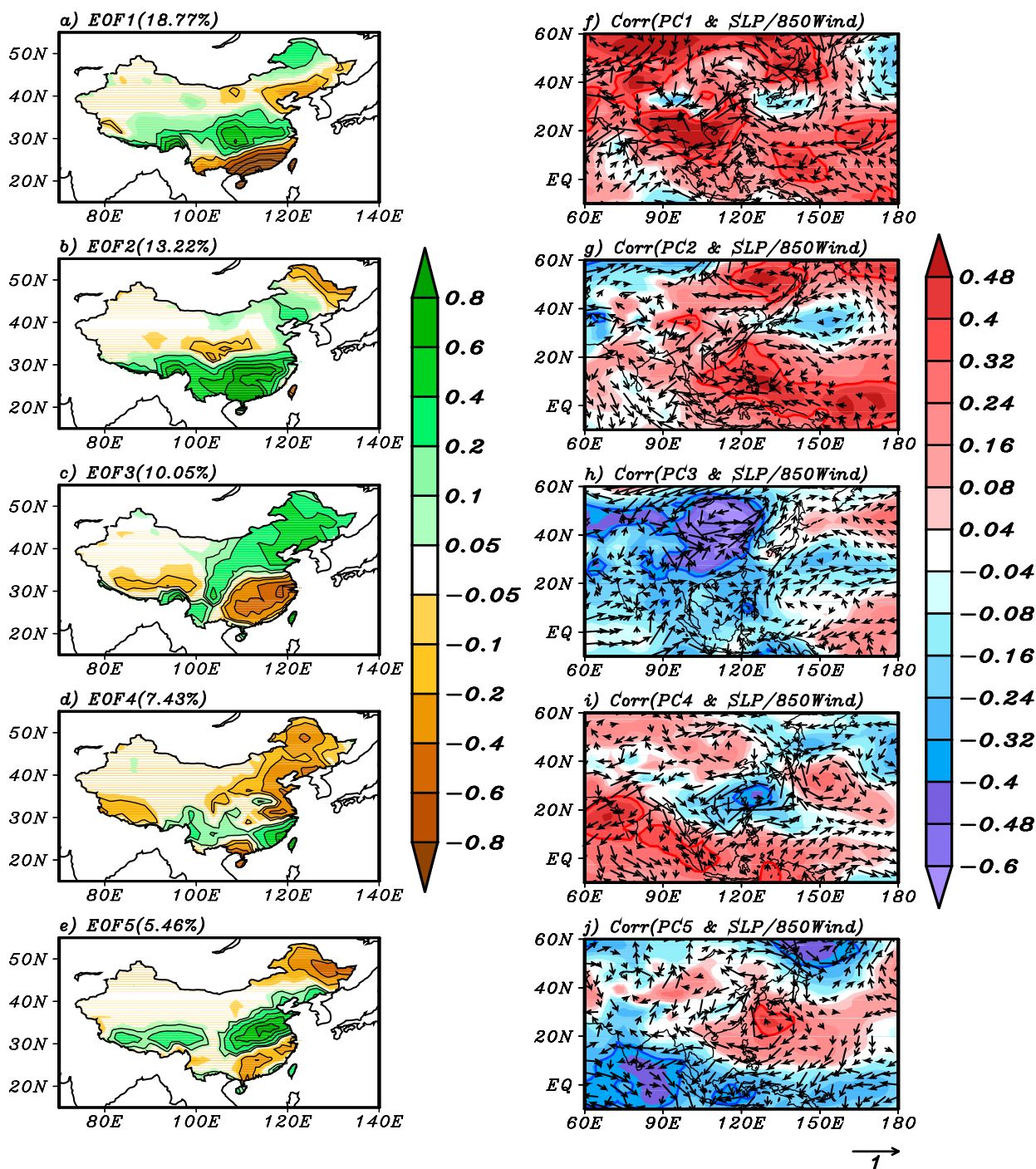


FIG. 2. (a)–(e) Spatial distribution of the first five EOF modes of JJA precipitation over China. The GPCP data from 1979 to 2013 were used for the EOF analysis. (f)–(j) The simultaneous correlation map (with reference to PC1 through PC5, respectively) of the anomalous sea surface pressure (color shading) and 850-hPa winds (vectors, m s^{-1}). The red (blue) contour is the correlation coefficient of +(-)0.33 with statistical significance at the 95% confidence level.

uncertainty of the observation. Second, the simultaneous correlation maps of the lower boundary anomalies with reference to the higher modes (starting from the sixth mode) show more “noisy” patterns and

irregular local signals compared with the first five modes. These facts suggest that the first five modes may represent the signals of China rainfall variability and are potentially predictable, but the higher modes,

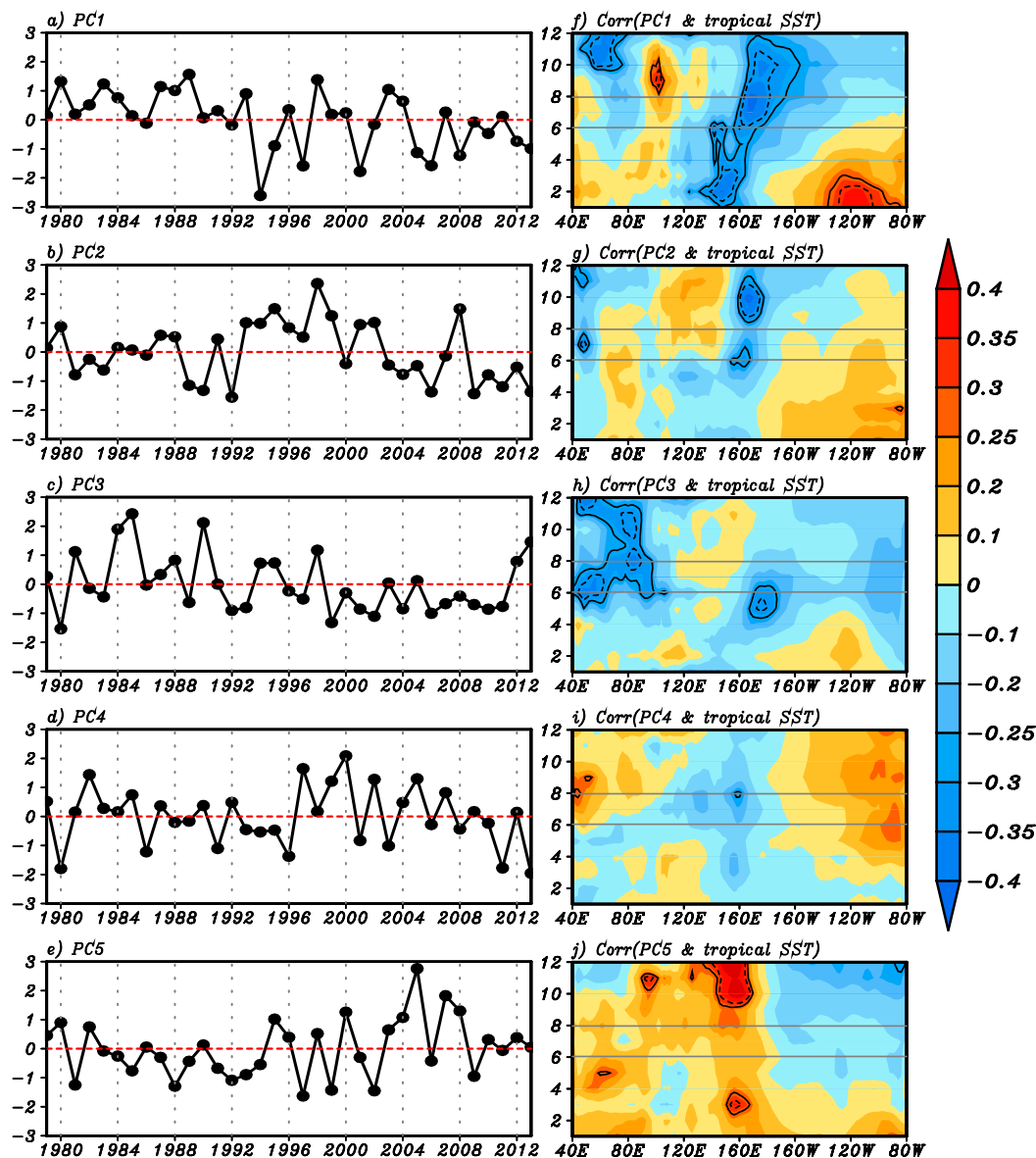


FIG. 3. (a)–(e) Principal components (PCs) of the first five EOF modes of JJA precipitation over China. (f)–(j) The evolution of equatorial Indo–Pacific (40°E–80°W) SST anomalies averaged between 10°S and 10°N that is associated with PC1–PC5, respectively, which is represented by the lead-lag correlation coefficients between the PCs and the monthly mean SSTA. The solid (dashed) contours are the correlation coefficient of ± 0.28 (± 0.33) with statistical significance at the 90% (95%) confidence level.

which are much more noisy and uncertain, may be unpredictable.

The five leading EOF modes account for 18.8%, 13.2%, 10.1%, 7.4%, and 5.5% of the total precipitation variance respectively. Together they account for 55% of the total variance, which may be viewed as the upper limit of the predictability estimated by the PC-based prediction method. These five modes are not statistically separable from each other by the North test (North et al. 1982) due to limited ensemble size (short length of

available record), but they are orthogonal and thus suitable for reconstruction of the total anomaly field.

The first five EOF patterns are shown in Figs. 2a–e and the corresponding PCs are shown in Figs. 3a–e. To see the circulation anomalies associated with each mode, the simultaneous correlation map between each PC and the sea level pressure (SLP) and 850-hPa wind anomalies is given in Figs. 2f–j. To understand the linkage between the first five modes and equatorial SST averaged between 10°S and 10°N, the lead-lag

correlation maps between monthly mean tropical SSTA and each PC are plotted in Figs. 3f–j. Significant rainfall anomalies of EOF spatial patterns are found over southern and eastern China where the climatological mean precipitation is relatively high (Fig. 1a).

The rainfall anomalies of the first EOF mode (EOF1) show a meridional distribution of zonal rainfall bands (Fig. 2a). Prominent suppressed rainfall shows up over southern China while the Yangtze River valley receives more rainfall. The anomaly signals over northern China are much weaker than over the southern part. The anomalous western Pacific subtropical high (WPSH) located in Southeast Asia (centered at the South China Sea) suppresses rainfall over the southern coast of China and induces above-normal rainfall to its north along the Yangtze River valley where SLP is relatively low (Fig. 2f). Except for the interannual fluctuation, the corresponding PC exhibits a downward trend or an interdecadal change from a positive value before 1993 to a negative value after 1994 (Fig. 3a). The results here are consistent with the decadal shift of summer rainfall in EASM previously documented by Kwon et al. (2007). The monthly lead–lag correlation map with the equatorial Indo–Pacific (40°E–80°W) SSTA averaged between 10°S and 10°N (Fig. 3f) shows that a prominent eastern Pacific (EP) warming appears during the previous winter and then rapidly decays from January to May followed by an onset of a subsequent cold event in the central Pacific (CP). So this mode occurs in the transition from the decaying phase of the EP type of El Niño to the development phase of the CP type of La Niña (Ashok et al. 2007).

The spatial pattern of EOF2 features increased precipitation over southern China, which represents an enhanced subtropical frontal rainfall associated with the anomalous WPSH over the SCS and Philippine Sea (Figs. 2b,g). The corresponding PC2 experiences both interannual and decadal change with a period of about 20 years (Fig. 3b). Similarly to EOF1, this mode is associated with slowly decaying of EP warming, but the development of CP cooling is not as significant as in the first mode (Fig. 3g).

EOF3 is characterized by a sharp contrast between eastern–northern China and southern China (Fig. 2c). The 850-hPa wind anomalies show anomalous anticyclones that are centered over western Japan and the Philippine Sea respectively. Abundant rainfall appears over northern China, which is associated with an anomalous cyclone centered over northern China and Mongolia (Fig. 2h). Over southeastern China, the anomalous high over the Philippine Sea that extends to the northern SCS induces deficient rainfall. The corresponding PC3 shows an irregular year-to-year fluctuation.

The SSTA over eastern Pacific during the preceding winter is insignificant, which means that this mode is a non-ENSO-related mode.

The rainfall anomalies of EOF4 are largely negative over almost the entire country except for Fujian province. This large-scale deficient monsoon is dominated by an anomalous SLP meridional dipole pattern over East Asia (Figs. 2d,i), with weakening of the western Pacific subtropical high and the northeast continental low. The northerly anomalies on the eastern edge of the anticyclone over northern Asia reduce the local rainfall in northern China. The corresponding PC4 has no significant correlation with tropical SSTA during the whole year (Fig. 3i).

The fifth EOF mode (EOF5) exhibits an abundant rainfall band over eastern-central China between Yellow River and Yangtze River (Fig. 2e). Affected by the two anomalous anticyclones over northwestern China and the East China Sea, there is a weak anomalous low SLP band over central China, which induces positive rainfall there (Fig. 2j). This mode also has little to do with equatorial SST anomalies from spring to summer (Fig. 3j).

These modes have also been identified in other studies using rotated EOF analysis (Zhang et al. 2009) or 160 stations' rainfall data (Nitta and Hu 1996). In summary, the five rainfall anomaly patterns are closely associated with different WPSH anomalies and Asian low anomalies. The first two modes are related to fast and slow decay of ENSO, respectively, whereas the other three are not related to ENSO.

4. Prediction of CSR anomaly pattern

The total rainfall anomalies can be predicted by using the sum of the observed five spatial EOF patterns multiplied by their corresponding predicted PCs as in the PMA analysis (Wang et al. 2015a). In our study, the PLS regression is applied to predict the five PCs.

A key element of PLS regression is to choose a predictor field. The CSR is strongly influenced by the slowly varying components of the climate system, such as SST and snow cover. The anomalous winter snowpack over Eurasia and the Tibetan Plateau may affect the springtime continental heating, which will influence the strength of Asian summer monsoon and rainfall (Yang and Xu 1994; Douville and Royer 1996; Chen and Wu 2000). Therefore, we take global surface temperature [ST; i.e., SST over oceans and 2-m air temperature (T2m) over continents] field as the predictor variable with a resolution of $2^\circ \times 2^\circ$ since snow cover variation, to a large extent, can be reflected by T2m (Namias 1962, 1985; Vavrus 2007).

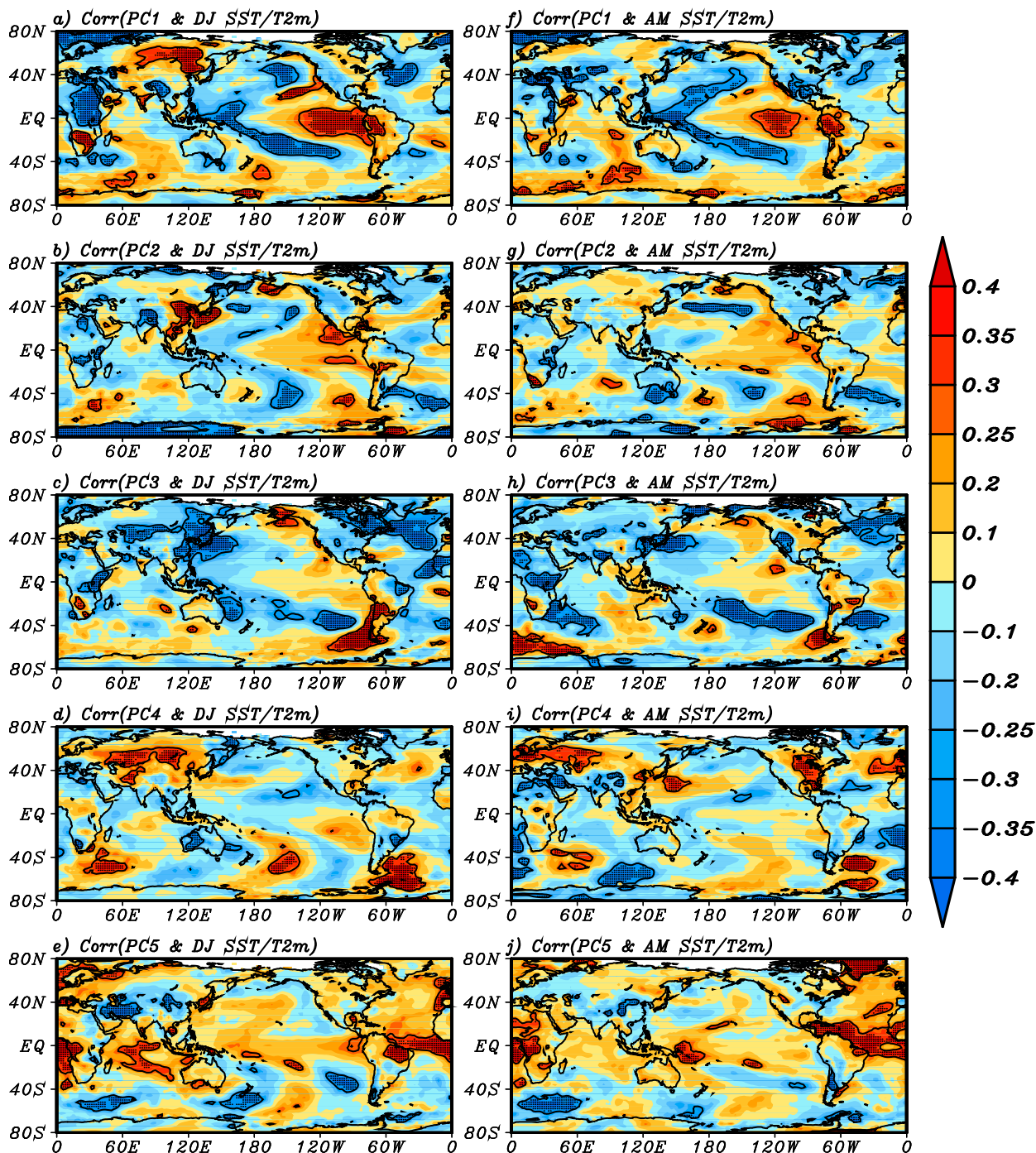


FIG. 4. (a)–(e) The correlation map (for PC1–PC5, respectively) of the anomalous December–January mean SSTA (color shading). (f)–(j) As in (a)–(e), but for April–May mean SSTA. The solid contours (dotted regions) are the correlation coefficient of ± 0.28 (± 0.33) with statistically significance at the 90% (95%) confidence level.

To select the lead period of the predictor field, a comparison is made between the correlation maps of ST with reference of each PC during two seasons respectively: Winter [from December in last year to January (DJ) mean; Figs. 4a–e] and spring (April–May

mean; Figs. 4f–j). Consistent with the discussion in section 3, the first two modes are mainly driven by the decay of EP-type ENSO (Figs. 4a,b,f,g). The significant positive correlation between T2m and PC1 over northeastern Eurasia indicates that anomalous heating over land

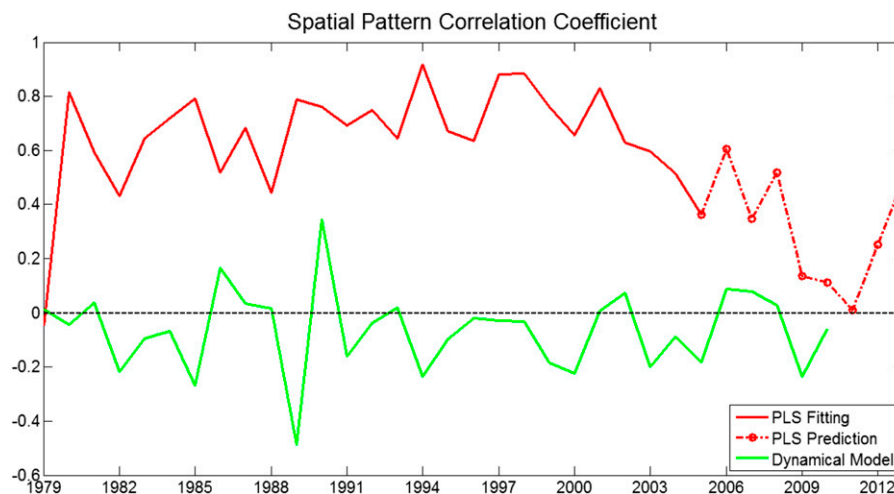


FIG. 5. The pattern correlation coefficient (PCC) skill for JJA precipitation prediction over China as a function of forecast year. The solid red line is the PCC between the observation and the fitted rainfall pattern using the first training period during 1979–2004. The mean value is 0.66. The dashed red line is the PCC between the observation and the forecast pattern during 2005–13. The mean value is 0.32. The green line shows the PCC between the observation and the hindcast result derived from a nine coupled models' ensemble mean with 1 May initial conditions. The mean value is only 0.04.

may play an important role for EOF1, EOF3, EOF4, and EOF5, which are non-ENSO-related modes, also have significant correlation with anomalous T2m over central and western Eurasia (Figs. 4c–e, h–j). In addition, the rainfall variability of EOF2 and EOF3 is tied to surface temperature anomalies over the Sea of Japan and the seaboard cities along East Asia (Lau et al. 2000).

Compared to the correlation maps during spring, we can see that all EOF modes have higher correlations with SST and T2m in the preceding winter. Besides, in order to make timely management decisions, a sufficient lead time is appreciated. For these reasons, we take the previous winter (DJ) mean for the predictor field. This long-lead-time forecast helps farmers to make decisions about cultivation in early spring. As mentioned in section 2, the first two PLS components (z_1 and z_2) are retained. For the average of the first five modes, the first PLS component z_1 explains about 51% of the variance in the PCs, and the second PLS component z_2 explains about 29% of the variance in the PCs.

In this study, we take the first 26 yr (1979–2004) as the first training period, which is reasonably long to get a sufficiently reliable predictor field. PLS regression is used to forecast the PC of each mode for the target year: the 27th year (2005). We predict rainfall anomalies of the target year by using the sum of the observed five spatial patterns derived from the training period multiplied by the forecasted PCs in the target year. It should be noted that the spatial patterns are derived using the data from 1979 to 2004 without information for

2005, so the forecast is independent of training data. Similarly, the anomalous rainfall pattern of the 28th year (2006) can be predicted by using the data from an extended 27-yr training period (1979–2005). The same procedures are repeated 9 times until the 35th year's rainfall is predicted with the data of the previous 34 yr. This independent test can rigorously reflect the ability of the statistical model to forecast CSR.

We use the correlation coefficient between the observed rainfall anomalies and the simulated rainfall pattern as a measure of fitting skill and that between the observation and the forecast patterns as a measure of forecast skill. The red solid line in Fig. 5 exhibits the pattern correlation coefficient (PCC) between the observation and the simulated rainfall pattern by using the first training period during 1979–2004. The average fitting skill is 0.66.

The PCC between the observation and the forecast pattern during 2005–13 is given by the red dashed line in Fig. 5. The average forecast skill is 0.32, which is much lower than the fitting skill. The reason is that the independent forecast skill is derived only from the past information while the fitting skill contains previous, present, and future information. Given the large number of the grid points (or independent sample size), the average forecast skills are statistically significant from zero skill. Among the nine years, the forecast skill of 2009–11 is relatively low. The low prediction skill of these three years may be due to the fact that the first three major modes during these three years project

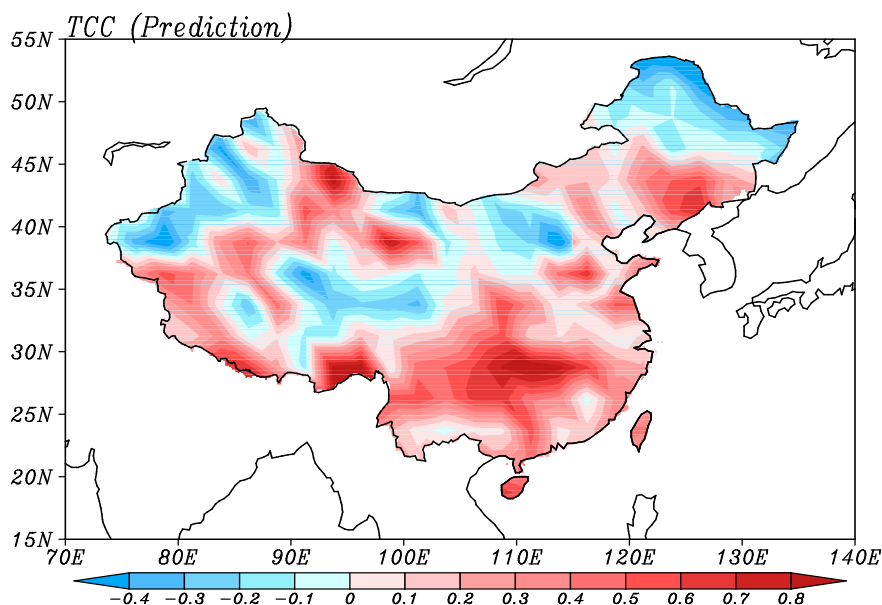


FIG. 6. The spatial variation of the independent forecast skill represented by the temporal correlation coefficient (TCC) between observed and forecast CSR anomalies during 2005–13.

weakly onto the observed precipitation pattern (Figs. 3a–c). The 9-yr averaged PCC between the rainfall anomalies reconstructed based on the first five EOF leading modes and the observational rainfall anomalies during 2005–13 is 0.47. This can be regarded as the potentially attainable prediction skill. The PCC skill predicted with the EOF–PLS combined model during this period accounts for about 66% of the potentially attainable prediction skill, suggesting the limitations of this predictive model.

Figure 6 shows the spatial variation of the mean independent forecast skill. Southeastern China rainfall anomalies are predicted better than those over the northwestern part of China where summer rainfall is scarce. High prediction skill is apparent over middle reach of the Yangtze River and the southern area of northeastern China. However, the prediction skill is essentially zero over the northern area of northeastern and northwestern China. More attention needs to be paid to these regions in future studies. To further check the spatial distribution of the forecast skill at the target years from 2005 to 2013, a comparison is made between the observed and forecast rainfall anomaly for each year. These results are shown in Figs. 7a–i, respectively. The authors also checked the prediction skill of 2014 after the main study (Fig. 7j). The PCC skill in this year is 0.28, which is close to the average. Figure 7 demonstrates that most regions of floods and droughts over eastern and southern China can be predicted with modest skill. Note that the 9-yr mean PCC value and spatial distribution of TCC skill derived from 160 stations are similar

to the results from GPCP data, showing the reliability of forecast skills of the combined EOF–PLS statistical model (Figs. S2 and S3 in the supplemental material).

The hindcast result represented by the green line in Fig. 5, which is derived from the nine coupled models' MME, is also shown for comparison. The nine coupled models that are available during 1979–2010 are all initialized on 1 May. The dynamic model exhibits deficiency in terms of predicting rainfall over mid- to high-latitude continental areas such as China because the skill during 2005–10 is significantly lower than the independent statistical forecast skill. The averaged skill of the dynamic MME forecast is not distinguishable from zero during 1979–2010. In addition, there is no seasonal variation in the MME's prediction skill in terms of East Asian summer rainfall. The prediction skill is very low regardless of forecast lead time (Lee et al. 2011). This indicates that the established statistical model can help to improve the current dynamic prediction.

5. Conclusions and discussion

The present study focuses on objective prediction of China summer rainfall (CSR) anomaly patterns. The potentially predictable part of the summer rainfall is presumably represented by the linear combination of the first five EOFs, which together explain about 55% of the total observed variability.

After obtaining the first five EOF modes, we predict the PC of each mode using partial least squares (PLS) regression. PLS regression is expected to yield a better

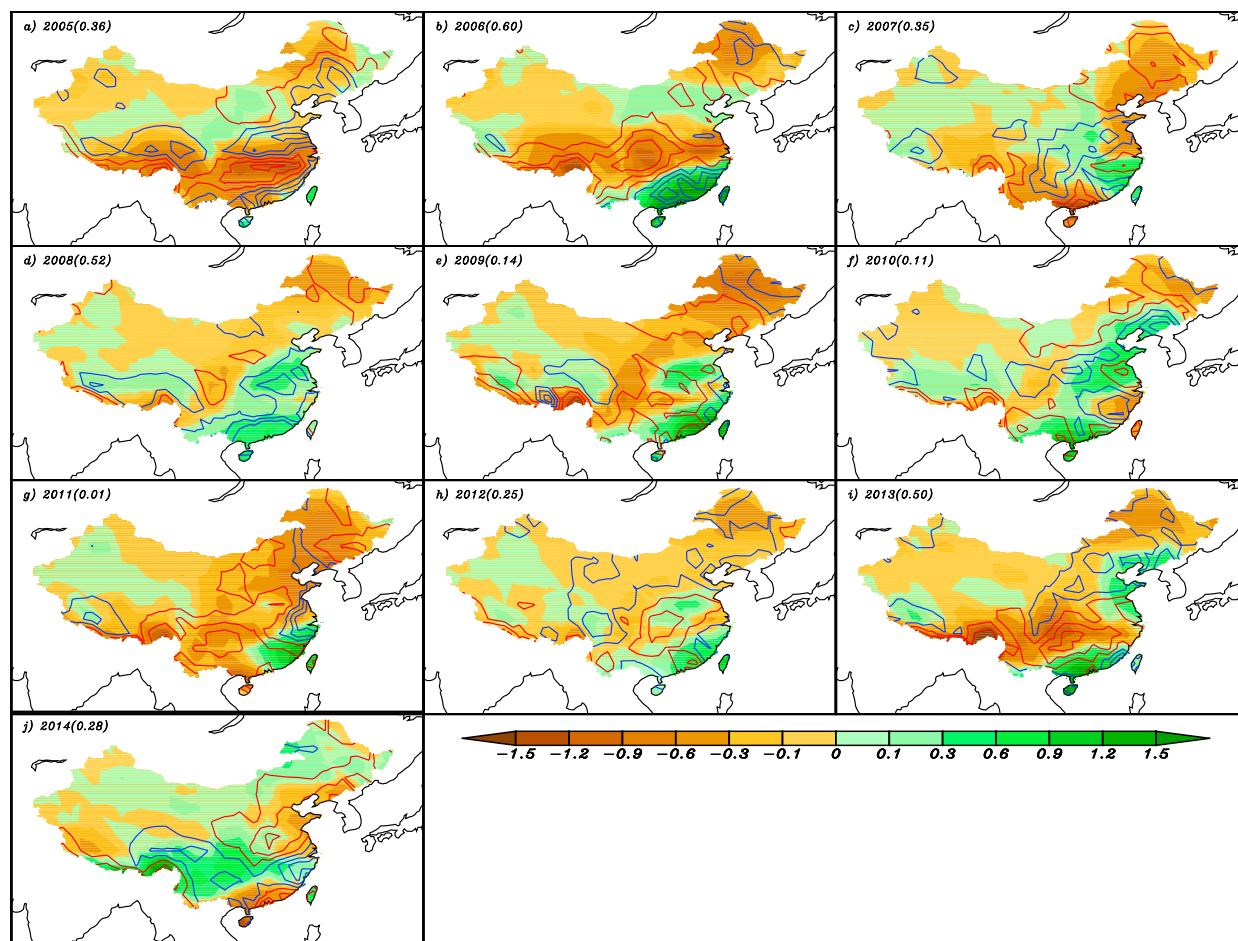


FIG. 7. The observed (contours) and forecast (color shading) rainfall anomalies during (a)–(j) 2005–14, respectively (mm day^{-1}). The blue (red) contour starts at $+(−)0.4$ with an interval of 0.8. The numbers within the parentheses in the figure legend indicate the PCC skill for each year.

prediction than those using regional predictors because the predictand is linked to an entire predictor field. surface temperature (ST; SST over ocean and 2-m air temperature over continental areas) during the previous December–January (DJ) mean is selected as the predictor field because snow cover (which is reflected by 2-m air temperature) and SSTA are the most important factors that affect CSR and the correlation between each mode, and ST during winter is higher than in spring.

An independent test is applied to check the forecast skill over the latest 9 years (2005–13) by using our combined EOF–PLS statistical model. The independent prediction test is carried out by utilizing only the predictor field before the target year. The pattern correlation coefficient (PCC) between the observation and forecast pattern during 2005–13 is used as a measure of independent forecast skill. The mean 4-month-lead PCC skill averaged across this 9-yr period is 0.32 which

is substantially higher than the dynamic models' 1-month-lead hindcast skill (0.04).

We also tested using the mean ST during other periods (e.g., April–May) as a predictor field. The forecast skill is lower than using the DJ ST. This is likely due to the fact that the CSR has, in general, a lower correlation with SST and 2-m air temperature in the preceding spring (as spring is often the SST transition period) than in the preceding winter with reference to each PC (Fig. 4). Therefore, the forecast skill using ST in winter yields better results.

Note that both the spatial patterns and the temporal evolutions of the EOF modes may also be subject to secular changes (Wang et al. 2015b). Therefore, the models derived here should be continuously tested and special attention should be given to detection of sudden changes in the predictor–predictand relationship. There are also some issues when applying PLS regression including whether the results are sensitive to

the predictor domain and the number of PLS components retained in predicting the predictand. These considerations need to be addressed carefully in conjunction with the implementation of PLS regression in climate prediction.

Acknowledgments. This work was jointly supported by the Atmosphere–Ocean Research Center (which is sponsored by Nanjing University of Information Science and Technology), the National Research Foundation of Korea through a Global Research Laboratory (GRL) grant of the Korean Ministry of Education, Science and Technology (MEST, 2011-0021927), and the NSFC-Shandong Joint Fund for Marine Science Research Centers (Grant U1406401). We also acknowledge support from China Scholarship Council (CSC) and the International Pacific Research Center (IPRC).

REFERENCES

- Ashok, K., S. Behera, S. Rao, H. Weng, and T. Yamagata, 2007: El Niño Modoki and its possible teleconnection. *J. Geophys. Res.*, **112**, C11007, doi:10.1029/2006JC003798.
- Chen, L., and R. Wu, 2000: Inter-annual and decadal variations of snow cover over Qinghai-Xizang Plateau and their relationship to summer monsoon rainfalls in China. *Adv. Atmos. Sci.*, **17**, 18–30, doi:10.1007/s00376-000-0040-7.
- De Jong, S., and A. Phatak, 1997: Partial least squares regression. *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling*, S. Van Huffel, Ed., SIAM, 25–36.
- Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674, doi:10.1175/JCLI3629.1.
- Déqué, M., 2001: Seasonal predictability of tropical rainfall: Probabilistic formulation and validation. *Tellus*, **53A**, 500–512, doi:10.1111/j.1600-0870.2001.00500.x.
- Ding, Y. H., 1994: *Monsoons over China*. Springer, 432 pp.
- Douville, H., and J.-F. Royer, 1996: Sensitivity of the Asian summer monsoon to an anomalous Eurasian snow cover within the Météo-France GCM. *Climate Dyn.*, **12**, 449–466, doi:10.1007/BF02346818.
- Fan, K., Y. Liu, and H. P. Chen, 2012: Improving the prediction of the East Asian summer monsoon: New approaches. *Wear Forecasting*, **27**, 1017–1030, doi:10.1175/WAF-D-11-00092.1.
- Gong, D., and S. Wang, 1999: Impacts of ENSO on rainfall of global land and China precipitation. *Chin. Sci. Bull.*, **44**, 852–856, doi:10.1007/BF02885036.
- Gregory, D., J. J. Morcrette, C. Jakob, A. C. M. Beljaars, and T. Stockdale, 2000: Revision of convection, radiation and cloud schemes in the ECMWF Integrated Forecasting System. *Quart. J. Roy. Meteor. Soc.*, **126**, 1685–1710, doi:10.1002/qj.49712656607.
- Gu, W., C. Y. Li, W. J. Li, W. Zhou, and J. C. L. Chan, 2009a: Interdecadal unstationary relationship between NAO and east China's summer precipitation patterns. *Geophys. Res. Lett.*, **36**, L13702, doi:10.1029/2009GL038843.
- , —, X. Wang, W. Zhou, and W. J. Li, 2009b: Linkage between mei-yu precipitation and North Atlantic SST on the decadal timescale. *Adv. Atmos. Sci.*, **26**, 101–108, doi:10.1007/s00376-009-0101-5.
- Jungclaus, J. H., and Coauthors, 2006: Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM. *J. Climate*, **19**, 3952–3972, doi:10.1175/JCLI3827.1.
- Kalela-Brundin, M., 1999: Climatic information from tree-rings of *Pinus sylvestris* L. and a reconstruction of summer temperatures back to AD 1500 in Femundsmarka, eastern Norway, using partial least squares regression (PLS) analysis. *Holocene*, **9**, 59–77, doi:10.1191/095968399678118795.
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1644, doi:10.1175/BAMS-83-11-1631.
- Kwon, M., J.-G. Jhun, and K.-J. Ha, 2007: Decadal change in East Asian summer monsoon circulation in the mid-1990s. *Geophys. Res. Lett.*, **34**, L21706, doi:10.1029/2007GL031977.
- Lau, K.-M., K.-M. Kim, and S. Yang, 2000: Dynamical and boundary forcing characteristics of regional components of the Asian summer monsoon. *J. Climate*, **13**, 2461–2482, doi:10.1175/1520-0442(2000)013<2461:DABFCO>2.0.CO;2.
- Lee, S. S., J.-Y. Lee, K.-J. Ha, B. Wang, and J. K. E. Schemm, 2011: Deficiencies and possibilities for long-lead coupled climate prediction of the western North Pacific–East Asian summer monsoon. *Climate Dyn.*, **36**, 1173–1188, doi:10.1007/s00382-010-0832-0.
- Li, F., and Q. Zeng, 2008: Statistical prediction of East Asian summer monsoon rainfall based on SST and sea ice concentration. *J. Meteor. Soc. Japan*, **86**, 237–243, doi:10.2151/jmsj.86.237.
- Liu, X., and Y. Wang, 2011: Contrasting impacts of spring thermal conditions over Tibetan Plateau on late-spring to early-summer precipitation in southeast China. *Atmos. Sci. Lett.*, **12**, 309–315, doi:10.1002/asl.343.
- Luo, J. J., S. Masson, S. Behera, S. Shingu, and T. Yamagata, 2005: Seasonal climate predictability in a coupled OAGCM using a different approach for ensemble forecast. *J. Climate*, **18**, 4474–4497, doi:10.1175/JCLI3526.1.
- Madec, G., P. Delecluse, M. Imbrad, and C. Levy, 1998: OPA version 8.1, Ocean general circulation model reference manual. Note du Pole de modélisation 11, IPSL, 91 pp.
- Marsland, S. J., H. Haak, J. H. Jungclaus, M. Latif, and F. Röske, 2003: The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates. *Ocean Modell.*, **5**, 91–127, doi:10.1016/S1463-5003(02)00015-X.
- Martens, H., and T. Naes, 1989: *Multivariate Calibration*. Wiley, 419 pp.
- McIntosh, A. R., and N. J. Lobaugh, 2004: Partial least squares analysis of neuroimaging data: Applications and advances. *Neuroimage*, **23**, S250–S263, doi:10.1016/j.neuroimage.2004.07.020.
- McIntosh, P. C., A. J. Ash, and M. Stafford Smith, 2005: From oceans to farms: The value of a novel statistical climate forecast for agricultural management. *J. Climate*, **18**, 4287–4302, doi:10.1175/JCLI3515.1.
- Molteni, F., and Coauthors, 2011: The new ECMWF Seasonal Forecast System (system 4). ECMWF Tech. Memo. 656, 49 pp.
- Namias, J., 1962: Influence of abnormal heat sources and sinks on atmospheric behavior. *Proc. Symp. on Numerical Weather Prediction*, Tokyo, Japan, Meteorological Society of Japan, 615–627.
- , 1985: Some empirical evidence for the influence of snow cover on temperature and precipitation. *Mon. Wea. Rev.*,

- 113**, 1542–1553, doi:[10.1175/1520-0493\(1985\)113<1542:SEEFTEI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1985)113<1542:SEEFTEI>2.0.CO;2).
- Nitta, T., and Z.-Z. Hu, 1996: Summer climate variability in China and its association with 500-hPa height and tropical convection. *J. Meteor. Soc. Japan*, **74**, 425–445.
- North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 699–706, doi:[10.1175/1520-0493\(1982\)110<0699:SEITEO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2).
- Phatak, A., and S. De Jong, 1997: The geometry of partial least squares. *J. Chemom.*, **11**, 311–338, doi:[10.1002/\(SICI\)1099-128X\(199707\)11:4<311::AID-CEM478>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1099-128X(199707)11:4<311::AID-CEM478>3.0.CO;2-4).
- Qian, Y.-F., Y.-Q. Zheng, Y. Zhang, and M.-Q. Miao, 2003: Response of China's summer monsoon climate to snow anomaly over the Tibetan Plateau. *Int. J. Climatol.*, **23**, 593–613, doi:[10.1002/joc.901](https://doi.org/10.1002/joc.901).
- Roeckner, E., and Coauthors, 1996: The atmospheric general circulation model ECHAM4: Model description and simulation of present-day climate. Max Planck Institut für Meteorologie Rep. 218, 90 pp.
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, doi:[10.1175/JCLI-D-12-00823.1](https://doi.org/10.1175/JCLI-D-12-00823.1).
- Shukla, J., 2007: Monsoon mysteries. *Science*, **318**, 204–205, doi:[10.1126/science.1150045](https://doi.org/10.1126/science.1150045).
- Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J. Climate*, **21**, 2283–2296, doi:[10.1175/2007JCLI2100.1](https://doi.org/10.1175/2007JCLI2100.1).
- Smoliak, B. V., J. M. Wallace, M. T. Stoelinga, and T. P. Mitchell, 2010: Application of partial least squares regression to the diagnosis of year-to-year variations in Pacific Northwest snowpack and Atlantic hurricanes. *Geophys. Res. Lett.*, **37**, L03801, doi:[10.1029/2009GL041478](https://doi.org/10.1029/2009GL041478).
- , —, P. Lin, and Q. Fu, 2015: Dynamical adjustment of the Northern Hemisphere surface air temperature field: Methodology and application to observations. *J. Climate*, **28**, 1613–1629, doi:[10.1175/JCLI-D-14-00111.1](https://doi.org/10.1175/JCLI-D-14-00111.1).
- Sun, J., X. B. Qu, and J. Z. Zhang, 2011: Analysis on the events and characters of meteorological disasters in China. *Wea. Forecasting Rev.*, **3** (1), 7–11.
- Tan, Y., L. Shi, W. Tong, G. T. G. Hwang, and C. Wang, 2004: Multi-class tumor classification by discriminant partial least squares using microarray gene expression data and assessment of classification models. *Comput. Biol. Chem.*, **28**, 235–243, doi:[10.1016/j.compbiolchem.2004.05.002](https://doi.org/10.1016/j.compbiolchem.2004.05.002).
- Tao, S., and L. Chen, 1987: A review of recent research on the East Asian summer monsoon in China. *Monsoon Meteorology*, C. P. Chang and T. N. Krishnamurti, Eds., Oxford University Press, 60–92.
- Vavrus, S., 2007: The role of terrestrial snow cover in the climate system. *Climate Dyn.*, **29**, 73–88, doi:[10.1007/s00382-007-0226-0](https://doi.org/10.1007/s00382-007-0226-0).
- Wang, B., R. Wu, and X. Fu, 2000: Pacific–East Asia teleconnection: How does ENSO affect East Asian climate? *J. Climate*, **13**, 1517–1536, doi:[10.1175/1520-0442\(2000\)013<1517:PEATHD>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1517:PEATHD>2.0.CO;2).
- , J.-Y. Lee, I.-S. Kang, J. Shukla, S. N. Hameed, and C.-K. Park, 2007: Coupled predictability of seasonal tropical precipitation. *CLIVAR Exchanges*, No. 12, International CLIVAR Project Office, Southampton, United Kingdom, 17–18.
- , and Coauthors, 2008: How accurately do coupled climate models predict the leading modes of Asian–Australian monsoon interannual variability? *Climate Dyn.*, **30**, 605–619, doi:[10.1007/s00382-007-0310-5](https://doi.org/10.1007/s00382-007-0310-5).
- , and Coauthors, 2009a: Advance and prospectus of seasonal prediction: Assessment of the APCC/CLIPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Climate Dyn.*, **33**, 93–117, doi:[10.1007/s00382-008-0460-0](https://doi.org/10.1007/s00382-008-0460-0).
- , J. Liu, J. Yang, T. Zhou, and Z. Wu, 2009b: Distinct principal modes of early and late summer rainfall anomalies in East Asia. *J. Climate*, **22**, 3864–3875, doi:[10.1175/2009JCLI2850.1](https://doi.org/10.1175/2009JCLI2850.1).
- , J.-Y. Lee, and B. Xiang, 2015a: Asian summer monsoon rainfall predictability: A predictable mode analysis. *Climate Dyn.*, **44**, 61–74, doi:[10.1007/s00382-014-2218-1](https://doi.org/10.1007/s00382-014-2218-1).
- , B. Xiang, J. Li, P. J. Webster, M. Rajeevan, J. Liu, and K.-J. Ha, 2015b: Rethinking Indian monsoon rainfall prediction in the context of the recent global warming. *Nature Commun.*, **6**, 7154, doi:[10.1038/ncomms8154](https://doi.org/10.1038/ncomms8154).
- Wold, H., 1966: Estimation of principal components and related rainfall predictability: A predictable mode analysis. *Multivariate Analysis*, P. R. Krishnaiah, Ed., Academic Press, 391–420.
- Wu, B., R. Zhang, and B. Wang, 2009a: On the association between spring Arctic sea ice concentration and Chinese summer rainfall: A further study. *Adv. Atmos. Sci.*, **26**, 666–678, doi:[10.1007/s00376-009-9009-3](https://doi.org/10.1007/s00376-009-9009-3).
- , —, —, and R. D'Arrigo, 2009b: On the association between spring Arctic sea ice concentration and Chinese summer rainfall. *Geophys. Res. Lett.*, **36**, L09501, doi:[10.1029/2009GL037299](https://doi.org/10.1029/2009GL037299).
- Wu, R., and B. Wang, 2002: A contrast of the East Asian summer monsoon–ENSO relationship between 1962–77 and 1978–93. *J. Climate*, **15**, 3266–3279, doi:[10.1175/1520-0442\(2002\)015<3266:ACOTEA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<3266:ACOTEA>2.0.CO;2).
- Wu, T.-W., and Z.-A. Qian, 2003: The relation between the Tibetan winter snow and the Asian summer monsoon and rainfall: An observational investigation. *J. Climate*, **16**, 2038–2051, doi:[10.1175/1520-0442\(2003\)016<2038:TRBTW>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<2038:TRBTW>2.0.CO;2).
- Wu, Z., B. Wang, J. Li, and F. F. Jin, 2009: An empirical seasonal prediction of the East Asian summer monsoon using ENSO and NAO. *J. Geophys. Res.*, **114**, D18120, doi:[10.1029/2009JD011733](https://doi.org/10.1029/2009JD011733).
- , H. Lin, Y. Li, and Y. Tang, 2013: Seasonal prediction of killing-frost frequency in south-central Canada during the cool/overwintering-crop growing season. *J. Appl. Meteor. Climatol.*, **52**, 102–110, doi:[10.1175/JAMC-D-12-059.1](https://doi.org/10.1175/JAMC-D-12-059.1).
- Xing, W., B. Wang, and S.-Y. Yim, 2014: Peak-summer East Asian rainfall predictability and prediction. Part I: Southeast Asia. *Climate Dyn.*, 1–13, doi:[10.1007/s00382-014-2385-0](https://doi.org/10.1007/s00382-014-2385-0).
- Yang, S., and L. Xu, 1994: Linkage between Eurasian winter snow cover and regional Chinese summer rainfall. *Int. J. Climatol.*, **14**, 739–750, doi:[10.1002/joc.3370140704](https://doi.org/10.1002/joc.3370140704).
- Yim, S.-Y., B. Wang, and W. Xing, 2014: Prediction of early summer rainfall over South China by a physical-empirical model. *Climate Dyn.*, **43**, 1883–1891, doi:[10.1007/s00382-013-2014-3](https://doi.org/10.1007/s00382-013-2014-3).
- Yuan, Y., H. Yang, W. Zhou, and C. Li, 2008a: Influences of the Indian Ocean dipole on the Asian summer monsoon in the following year. *Int. J. Climatol.*, **28**, 1849–1859, doi:[10.1002/joc.1678](https://doi.org/10.1002/joc.1678).
- , W. Zhou, J. C. L. Chan, and C. Li, 2008b: Impacts of the basin-wide Indian Ocean SSTA on the South China Sea summer monsoon onset. *Int. J. Climatol.*, **28**, 1579–1587, doi:[10.1002/joc.1671](https://doi.org/10.1002/joc.1671).
- Zhang, H., J. Qin, and Y. Li, 2011: Climate background of cold and wet winter in southern China: Part I. Observational analysis. *Climate Dyn.*, **37**, 2335–2354, doi:[10.1007/s00382-011-1022-4](https://doi.org/10.1007/s00382-011-1022-4).

- Zhang, Q., C.-Y. Xu, Z. Zhang, Y. D. Chen, and C.-L. Liu, 2009: Spatial and temporal variability of precipitation over China, 1951–2005. *Theor. Appl. Climatol.*, **95**, 53–68, doi:[10.1007/s00704-007-0375-4](https://doi.org/10.1007/s00704-007-0375-4).
- Zhang, Y., T. Li, and B. Wang, 2004: Decadal change of the spring snow depth over the Tibetan Plateau: The associated circulation and influence on the East Asian summer monsoon. *J. Climate*, **17**, 2780–2793, doi:[10.1175/1520-0442\(2004\)017<2780:DCOTSS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2780:DCOTSS>2.0.CO;2).
- Zhao, P., Z. Zhou, and J. Liu, 2007: Variability of Tibetan spring snow and its associations with the hemispheric extratropical circulation and East Asian summer monsoon rainfall: An observational investigation. *J. Climate*, **20**, 3942–3955, doi:[10.1175/JCLI4205.1](https://doi.org/10.1175/JCLI4205.1).
- Zhong, A., H. H. Hendon, and O. Alves, 2005: Indian Ocean variability and its association with ENSO in a global coupled model. *J. Climate*, **18**, 3634–3649, doi:[10.1175/JCLI3493.1](https://doi.org/10.1175/JCLI3493.1).