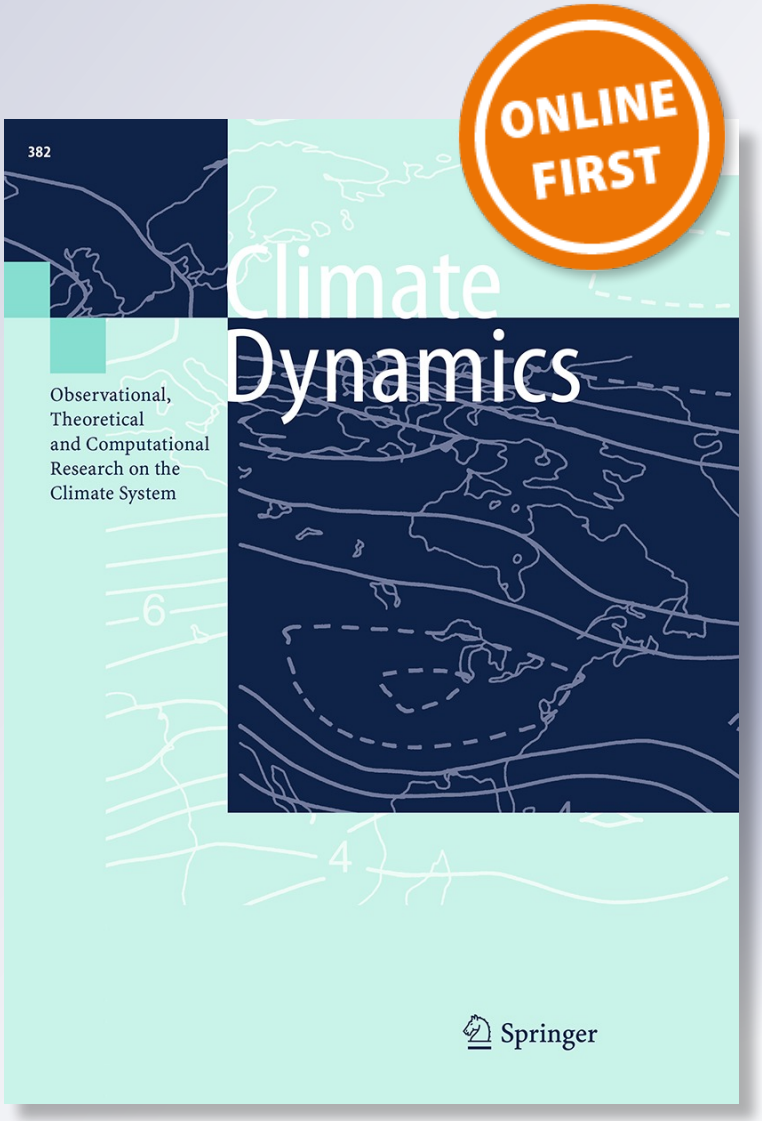
Predictability and prediction of summer rainfall in the arid and semi-arid regions of China

Wen Xing & Bin Wang

Climate Dynamics
Observational, Theoretical and Computational Research on the Climate System

ISSN 0930-7575

Clim Dyn
DOI 10.1007/s00382-016-3351-9



Springer

Clim Dyn
DOI 10.1007/s00382-016-3351-9

Predictability and prediction of summer rainfall in the arid and semi-arid regions of China

Wen Xing[1] · **Bin Wang**[2,3]

Received: 2 April 2016 / Accepted: 7 September 2016

Abstract Northwest China (NWC) is an arid and semi-arid region where climate variability and environmental changes are sensitive to precipitation. The present study explores sources and limits of predictability of summer precipitation over NWC using the predictable mode analysis (PMA) of percentage of rainfall anomaly data. Two major modes of NWC summer rainfall variability are identified which are tied to Eurasian continental scale precipitation variations. The first mode features wet northern China corresponding to dry central Siberia and wet Mongolia, which is mainly driven by tropical Pacific sea surface temperature anomalies (SSTA). The second mode features wet western China reflecting wet Central Asia and dry Ural–western Siberia, which strongly links to Indian Ocean SSTA. Anomalous land warming over Eurasia also provides important precursors for the two modes. The cross-validated hindcast results demonstrate these modes can be predicted with significant correlation skills, suggesting that they may be considered as predictable modes. The domain averaged temporal correlation coefficient (TCC) skill during 1979 to 2015 using 0-month (1-month) lead models is 0.39 (0.35), which is considerably higher than dynamical models' multi-model ensemble mean skill (−0.02). Maximum potential attainable prediction skills are also estimated and discussed. The result illustrates advantage of PMA in predicting rainfall over dry land areas and large room for dynamical model improvement. However, secular changes of predictors need to be detected continuously in order to make practical useful prediction.

Keywords Arid and semi-arid regions · Northwest China · Summer rainfall · Seasonal prediction · Predictable mode analysis

1 Introduction

Northwest China (NWC), which occupies 30 % of the total Chinese territory, is located in the innermost Eurasia continent. Because NWC is thousands of kilometers away from the ocean, this region includes an agro-pastoral transition zone (Gong et al. 2004) and vast desert areas. Affected by the arrival of summer monsoon, NWC receives most of its annual rainfall (more than 60 %) in boreal summer (June–July–August, JJA) (Samel et al. 1999). The climatological rainfall rate over NWC in JJA is normally less than 3 mm d^{-1}, the least among the entire China (Fig. 1). Hence, NWC is considered as an arid and semi-arid region compared with monsoonal southeast China.

The environment of arid and semi-arid regions is sensitive to the regional climate variability, particularly to the precipitation (Ye and Chen 1992). Even a small rainfall deficit would possibly exert heavy drought pressure on the plants and agriculture, especially during summer when vegetation growth reaches most active season. Besides, due to the increasing anthropogenic activities and significant rainfall variability, NWC is also vulnerable to desertification, which is one of the most serious problems in the current

✉ Wen Xing
xingwen1022@gmail.com

[1] Physical Oceanography Laboratory/CIMST, Ocean University of China and Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

[2] Department of Atmospheric Sciences and International Pacific Research Center, University of Hawaii at Manoa, Honolulu, HI 96822, USA

[3] Earth System Modeling Center, Nanjing University of Information Science and Technology, Nanjing 210044, China

Published online: 23 September 2016

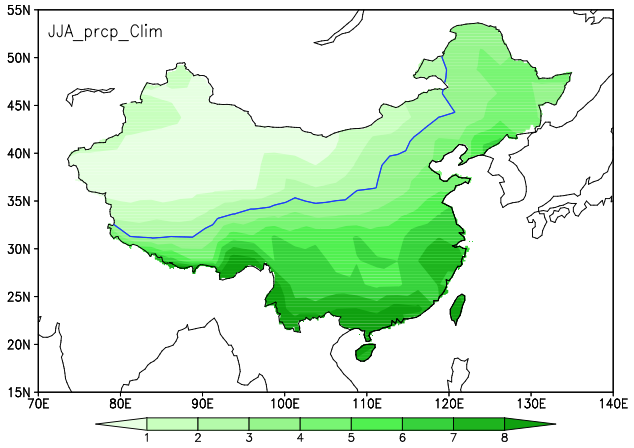


Fig. 1 Climatological mean precipitation rate (in unit of mm day^{-1}) averaged for June–July–August (JJA) from 1979 to 2015. The *blue contour* indicates the precipitation rate of 3 mm day^{-1}

global environment (FAO, Unesco, WMO 1977). Investigating and predicting the precipitation variability of NWC is helpful for developing agriculture and animal husbandry, and for assessing the rapidly growing environmental problems such as desertification.

Precipitation over NWC is considered to be affected by both monsoon circulation and mid-latitude westerly disturbance. The interannual variability of NWC summer rainfall exhibits pronounced regionality (Guo and Wang 1988; Chen et al. 1992; Kripalani and Singh 1993). Precipitation anomalies in northern China are well positively correlated with anomalous Indian summer rainfall (Yatagai and Yasunari 1995). The east-westward expansion of the Tibetan High combined with the westward shift of the Pacific Subtropical High is responsible for the wet years in northern China. Rainfall anomaly over northern China also seems to be closely related to El Niño/Southern Oscillation (ENSO) (Wang and Li 1990) and atmosphere–ocean interaction in the equatorial Pacific (Yasunari 1990; Shen and Lau 1995). Meanwhile, precipitation in Xinjiang Province locating over northwestern China shows a negative relationship with all-Indian summer rainfall, which is caused by a rather local circulation change over Central Asia. The interannual variation of summer precipitation in this region is also related to the windward mid-latitude circulation and eastward (westward) shift of the Tibetan High in a wet (dry) year (Yatagai and Yasunari 1995).

Considering the poor prediction skill of dynamical models, a series of studies have been performed on Asian summer rainfall prediction using physical-empirical combined models (Wu et al. 2009; Yim et al. 2014; Li and Wang 2015; Wu and Yu 2016; Xing et al. 2016a). The long-lead seasonal prediction of summer rainfall anomaly pattern over the whole China was also examined by Xing et al. (2016b). However, these studies focused on a broad section of East Asia or the entire China, and paid little attention to the arid and semi-arid China. This may be due to lack of drought and flood events in this region compared with southeastern China (Lv et al. 2015). Besides, the significant regionality of NWC rainfall variation make it difficult to predict the rainfall pattern (Yatagai and Yasunari 1995; Zhao and Li 2013). Lv et al. (2015) predicted the area-mean summer rainfall index averaged over eastern NWC. However, there is still little work that focuses on the prediction of rainfall pattern over the whole arid and semi-arid regions which remains a great challenge.

The present study is devoted to understanding the predictability and improving prediction of summer precipitation variations in the arid/semi-arid NWC. The analysis season is from June to August, the local wet season. These months are of great importance for agriculture and human activity. Section 2 briefly describes the datasets and methodology employed in this study. Section 3 explores physical processes governing the most important modes of summer precipitation variability. The physical–empirical (P–E) prediction models for each principal mode are also established in this section. In Sect. 4 we estimate the extent to which we can predict NWC summer rainfall using P–E models. Section 5 provides a summary.

2 Data and method

2.1 Data

The data used in this analysis comprise monthly mean precipitation from Global Precipitation Climatology Project (GPCP, v2.2) datasets (Huffman et al. 2011), monthly mean sea surface temperature (SST) from NOAA Extended Reconstructed SST (ERSST v4, Huang et al. 2014; Liu 2014), and the monthly mean circulation data from ERA-interim (Dee et al. 2011). The data period chosen in this study is from 1979 to 2015.

2.2 Method

In this study, the predictable mode analysis (PMA) method, which was first proposed by Wang et al. (2007), is applied to estimate the potential predictability and establish prediction model for NWC summer rainfall. This method has been applied to assess the predictability of Asia summer rainfall (Yim et al. 2014; Li and Wang 2015; Xing et al. 2016a) and other climate variability (e.g. Lee et al. 2011, 2013; Wang et al. 2013, 2014).

The PMA includes three elements: empirical analysis, physical interpretation and retrospective predictions. The empirical analysis detects frequently observed patterns of

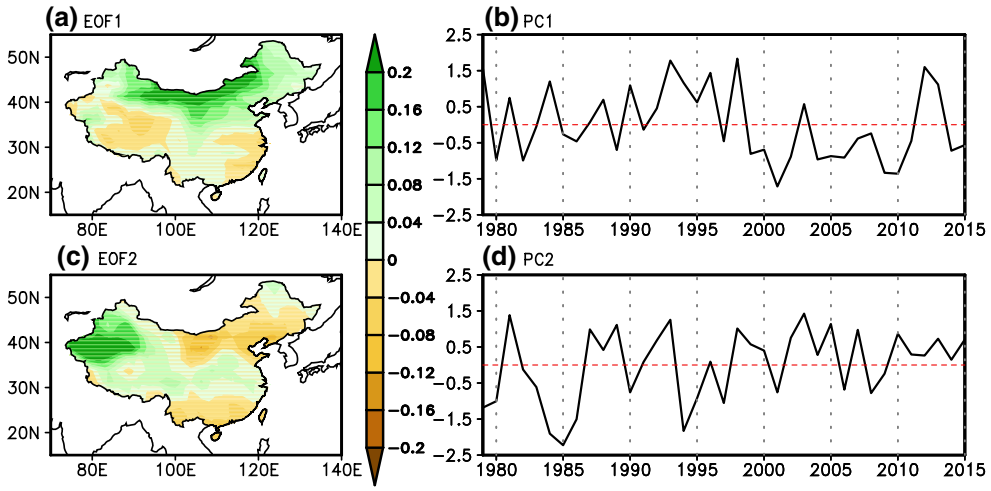


Fig. 2 **a**, **c** Spatial distribution of the first two leading EOF modes of percentage of rainfall anomaly during June to August over whole China from 1979 to 2015 and **b**, **d** the associated PC of each mode. The two modes account for 16.2 and 11.9 % of the total variance, respectively

variability which are often derived from empirical orthogonal function (EOF) analysis. The physical interpretation provides dynamical basis for prediction of each empirical pattern and thus establishes physical–empirical (P–E) prediction models for the principle component (PC) of each mode. The retrospective predictions (hindcast) with P–E models and/or dynamical models are used to identify the potential "predictable" modes. In the present study, since the dynamical models have limited capability in reproducing observed EOF modes, we use P–E prediction models to examine the reproducibility of these EOF modes.

To objectively select physically consequential predictors, we focus on surface temperature field, i.e., SST over ocean and 2 m air temperature (T2m) over land which reflect ocean and land surface thermal conditions. Two types of surface temperature anomalies are searched for the predictors: (a) *persistent* signals from late winter (February and March, FM) to spring (April and May, AM) (FMAM) and (b) *tendency* signals from FM to AM (AM-minus-FM). The persistent signals sometimes reflect positive feedback processes associated with the local atmosphere–ocean or atmosphere–land interaction, which may help maintain the lower boundary anomalies. The tendency predictors denote changes from late winter to spring and frequently tip-off the direction of subsequent evolution. Note that different from statistical method that fishes predictors from variety of fields and variables, the P–E approach seeks for large-scale statistically significant signals only in these two correlation maps (persistent and tendency). In the selection of predictors, we emphasize understanding the processes that explain the lead–lag relationships between the predictors and the NWC summer rainfall.

Stepwise multi-linear regression is used to establish the P–E model for each PC. The stepwise procedure identifies statistically important predictors at each step. The significance of each predictor selected is based on significance in increasing the regressed variance by a standard F-test (Panofsky and Brier 1968). A 95 % statistical significance level is used as a criterion to select new predictor at each step. Once selected into the model, a predictor can only be removed if its significance level falls below 90 % by the addition/removal of another variable. In this way those predictors that are relatively independent in their physical meanings can be selected and those that are well correlated predictors are avoided. To circumvent over-fitting, the number of predictors is required to be less or equal to three (i.e., less than 10 % of the sample size 37).

3 Major modes of NWC summer rainfall: origin and predictors

Since the climatological and anomalous rainfall amounts over NWC are much smaller than southern and eastern China (Fig. 1), the primary EOF modes may only reflect the rainfall variability over coastal regions if anomalous rainfall data over the whole China is used for EOF analysis (Xing et al. 2016b). In order to better depict the rainfall variability over NWC, the JJA mean percentage of precipitation anomaly (anomalous rainfall/climatology rainfall) from 1979 to 2015 is used for EOF analysis.

The first two modes, which are statistically separable from other modes based on the North's rule (North et al. 1982), represent the rainfall variability over northern and western China, respectively (Fig. 2a, c). The higher modes show irregularly weak signals and are not separable from each other. Because this study focuses on the rainfall prediction over arid/semi-arid NWC, another EOF analysis is

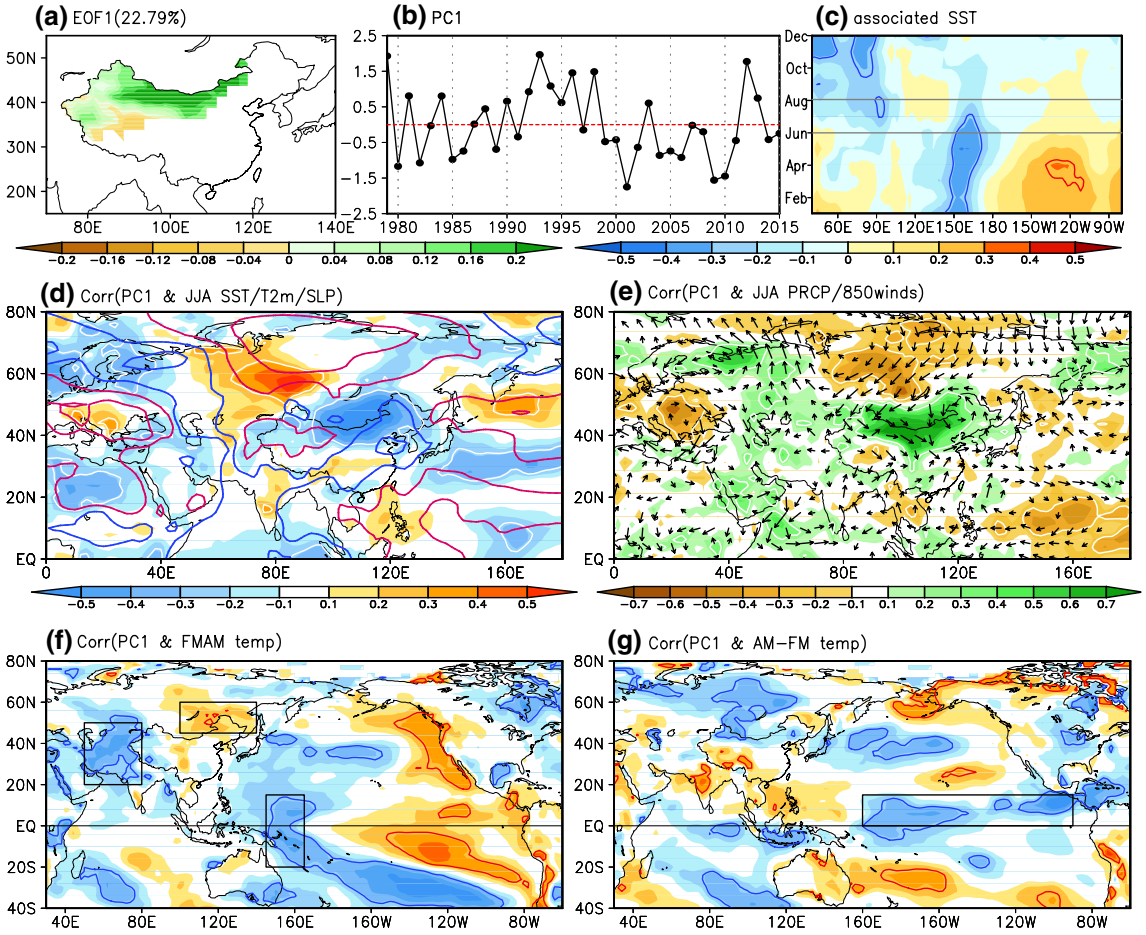


Fig. 3 **a** The spatial pattern and **b** corresponding PC (PC1) of the first EOF mode (EOF1) derived from percentage of JJA precipitation anomaly over northwestern China where the climatology rainfall rate is less than 3 mm day^{-1} for the period of 1979–2015. **c** The relationships between EOF1 and equatorial Indian-Pacific (40°E–80°W) SSTA averaged between 10°S and 10°N. The relationships are shown by the lead–lag correlation coefficients of monthly mean SSTA with reference of PC1. The *gray lines* show period of JJA. **d** The simultaneous correlation map (with reference to the PC1) of the anomalous JJA mean ST (SST over ocean and T2m over land) and SLP anomalies. The *red/blue* contours mean positive/negative correlation coefficient between SLP and PC1 starting from ±0.1 with an interval of 0.2. **e** The same as in Fig. 3d but for anomalous JJA mean precipitation (*color shading*) and 850 hPa wind anomalies (*vectors*). **f** The correlation map of FMAM mean ST with reference to the PC1. **g** The same as in Fig. 3f but for AM-minus-FM. The contour represents the correlation coefficient significant at the 90 % confidence level. The *color contours* in **c**, **f**, **g** and the *white* contours in (**d**–**e**) represent the correlation coefficient significant at the 90 % confidence level. The rectangular regions outline the area used to define the predictors

applied to the JJA mean percentage of precipitation anomaly over the region of NWC only, where the summer daily mean rainfall is less than 3 mm day^{-1} (Fig. 1). The spatial patterns and principal components of the first two modes are given in Figs. 3 and 5 respectively. These two modes are consistent with and resemble the first two modes derived from the whole China domain (Fig. 2). The correlation coefficients between the two PCs for the first and second mode are both 0.97. Thus the first two modes can be considered as the most important arid/semi-arid modes in NWC.

Out of the reasons above, we focus on the two leading modes over NWC, which account for 22.8 and 16.2 % of the total variance, respectively. In this section, we first discuss characteristics of the precipitation and low-level circulation patterns associated with each mode, then examine their leading relationship with surface temperature anomalies in February-to-May persistent and AM-minus-FM tendency fields, followed by selection of physically meaningful predictors for each mode. In order to understand the physical links between each predictor and rainfall for each mode, the leading and simultaneous correlation maps of lower boundary conditions or atmosphere circulations with reference to each predictor for PC1 and PC2 are shown in Figs. 4 and 6, respectively.

3.1 EOF1: Mongolia mode

The spatial pattern of the first leading mode (EOF1) is characterized by an elongated positive rainfall located

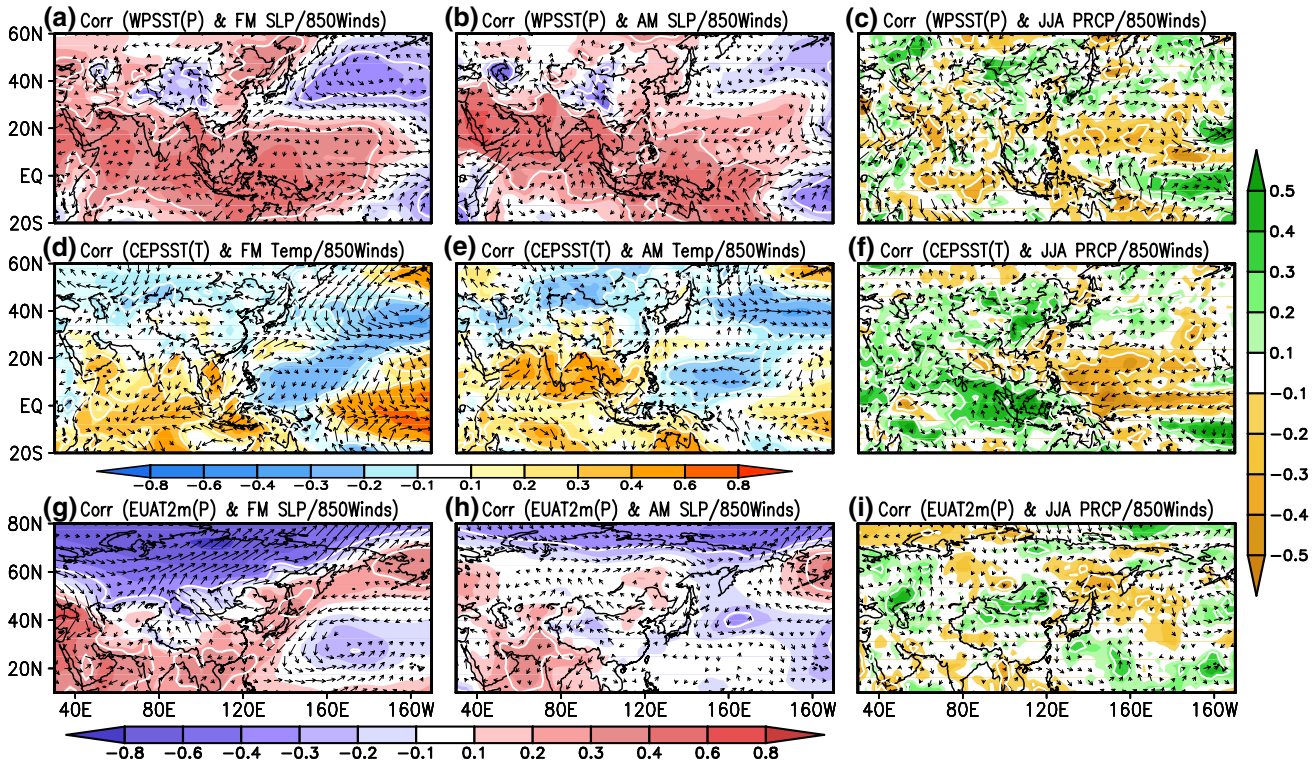


Fig. 4 **a**, **b** The correlation map of **a** FM mean and **b** AM mean SLP (*color shading*)/850 hPa wind anomalies (*vectors*) with reference to 0-month lead predictor WPSST (P) of PC1. **c** The correlation map of JJA mean precipitation (*color shading*)/850 hPa wind anomalies (*vectors*) with reference to 0-month lead predictor WPSST (P) of PC1. **d**, **e** The correlation map of **d** FM mean and **e** AM mean ST (*color shading*)/850 hPa wind anomalies (*vectors*) with reference to 0-month lead predictor CEPSST (T) of PC1. **f** The same as in (**c**) but for the 0-month lead predictor CEPSST (T). **g–i** The same as in **a–c** respectively but for the 0-month lead predictor EUAT2m (P). The *white* contours in each figure represent the correlation coefficient significant at the 90 % confidence level

along the northern border of China (Fig. 3a). This mode is part of a southeast-northwest oriented sandwich pattern of T2 m and precipitation anomalies, which is associated with a wave train pattern from Europe to East Asia (Fig. 3d, e). There is an anomalous anticyclone over Philippine Sea, expanding to South China Sea and eastern China, with higher SLP and suppressed rainfall in situ (Fig. 3e). The southwesterly anomalies associated with this anticyclone are also the southern branch of an anomalous cyclone located over northern China and Mongolia, inducing above-normal rainfall there. Thus, the anomalous southwest winds over southern and central China play a vital role in the enhanced rainfall in EOF1 mode. The T2m in northern China and Mongolia is also relatively low, consistent with the enhanced rainfall-induced greater cloudiness. To the north of Mongolia, a large-scale anticyclone and higher SLP appear, concurrent with suppressed rainfall in situ (Fig. 3d, e).

The lead–lag relationship of PC1 with equatorial SST anomalies (SSTA) over Indian Ocean to Pacific is illustrated in Fig. 3c. In tropical western Pacific, pronounced SST cooling lasts from previous winter to early summer, while positive SSTA appears over tropical eastern Pacific during previous winter and decays rapidly. These two features are also reflected in Fig. 3f, g. Thus, the SSTA from February to May averaged over tropical western Pacific (145°E–165°E, 20°S–15°N) is selected as one predictor for this mode, called WPSST (P). The cooling tendency from late winter (FM) to late spring (AM) averaged over central to eastern Pacific (160°E–90°W, 0–15°N) is selected as the second predictor which is called CEPSST (T). The physical processes that these two predictors make contributions to the formation of the southwesterly anomalies over central China and enhanced rainfall over Mongolia are discussed below.

The tropical WP cooling weakens local convective activities. The suppressed precipitation can generate low-level anomalous anticyclones to its west on both sides of the equator (Fig. 4a, b). The easterly components associated with the two anticyclones may enhance the total wind speed over tropical WP. Thus the cooling may maintain from the previous late winter to spring (Fig. 3f).

The second predictor represents the rapid decay of EP warming. The correlation map between this predictor and

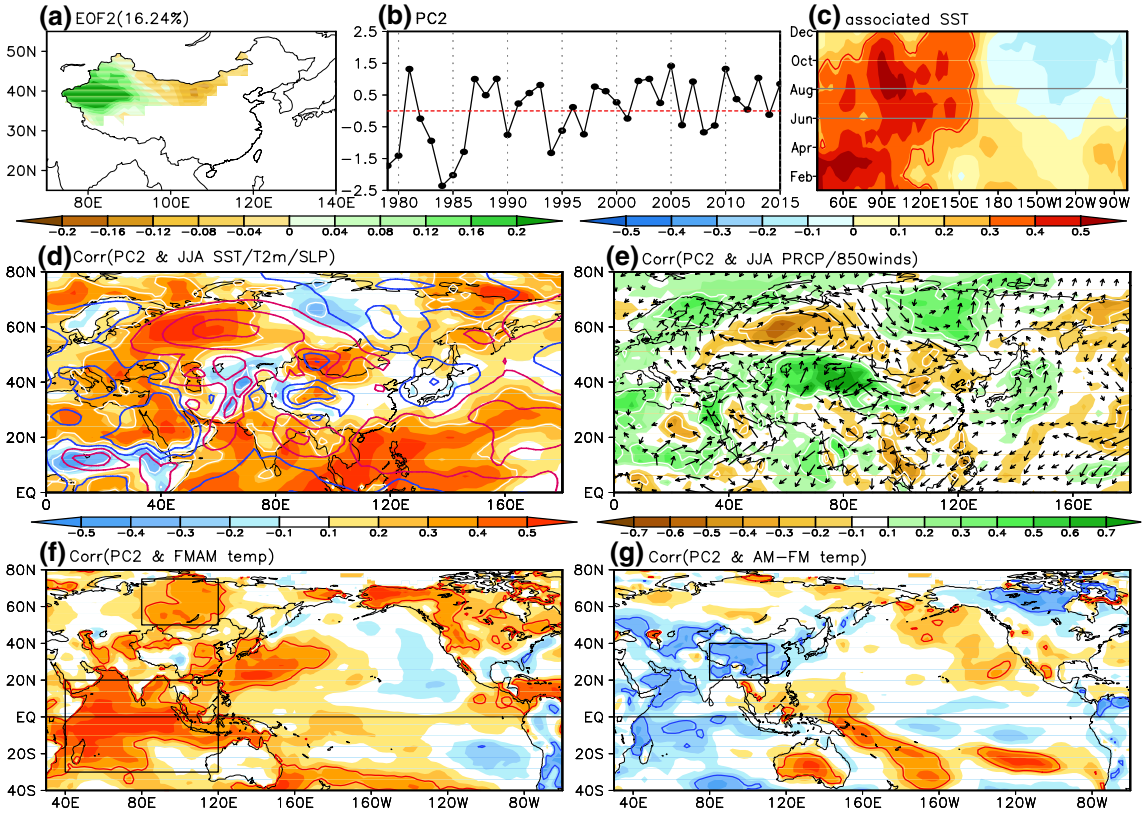


Fig. 5 The same as in Fig. 3 but for EOF2

850 hPa wind anomaly from February to August shows that an anomalous anticyclone with a ridge line along 20°N, extending from WP (Philippine Sea) to southeast China, persists from early spring through the ensuring summer (Fig. 4d–f). This is consistent with the finding of Wang and Zhang (2002). Wang et al. (2000) attributed the persistence of the Philippine Sea anticyclone from previous winter to summer to a positive thermodynamic feedback. As can be seen from Fig. 4d–f, the feedback occurs between atmospheric descending Rossby waves/suppressed convection and the underlying cold SST anomaly to the southeast of the anticyclone center. This air–sea interaction is also recognized as responsible for the origins of the second EOF mode of the Asian summer monsoon rainfall (Wang et al. 2014), and the first EOF mode of EA summer rainfall (Wang et al. 2009a; Xing et al. 2016a).

Therefore, both WPSST (P) and CEPSST (T) are important predictors for the maintenance of strengthened western Pacific subtropical high (WPSH), which induce suppressed precipitation over the Philippine Sea and enhanced rainfall over northern China and Mongolia.

The third predictor for PC1 is the persistent dipole differential T2m anomalies between southwestern (50°E–80°E, 20°N–50°N) and northeastern Eurasia (100°E–140°E, 45°N–60°N), which is called EUAT2m (P) (Fig. 3f). This predictor is associated with an anomalous low pressure trough over northwestern Eurasia (Fig. 4g). To the east of the low pressure trough, there is a northeast-southwest oriented high anomaly along the border of Eurasia continent. The anomalous high is centered over northeastern China. During summer, the anomalous anticyclone circulation over northeastern China may induce a local rainfall deficit. However, the southern flank of the anticyclone transports moisture from northern Pacific to northern China, inducing more rainfall there (Fig. 4h, i).

3.2 EOF2: the central Eurasia mode

EOF2 features a contrast between a prominently enhanced rainfall center over the far western China and a relatively weak rainfall suppression center to its east in the central northern China (Fig. 5a).

The large-scale low-level circulation and rainfall anomalies associated with PC2 are characterized by a dipole pattern: a suppressed rainfall center accompanied by an anomalous anticyclone and land warming is centered over Ural mountain (60°N, 60°E), whereas the enhanced rainfall anomalies and the associated anomalous cyclone are centered in central Asia (40°N, 80°E) (Fig. 5d, e). Thus, this mode mainly reflects the rainfall pattern over the central

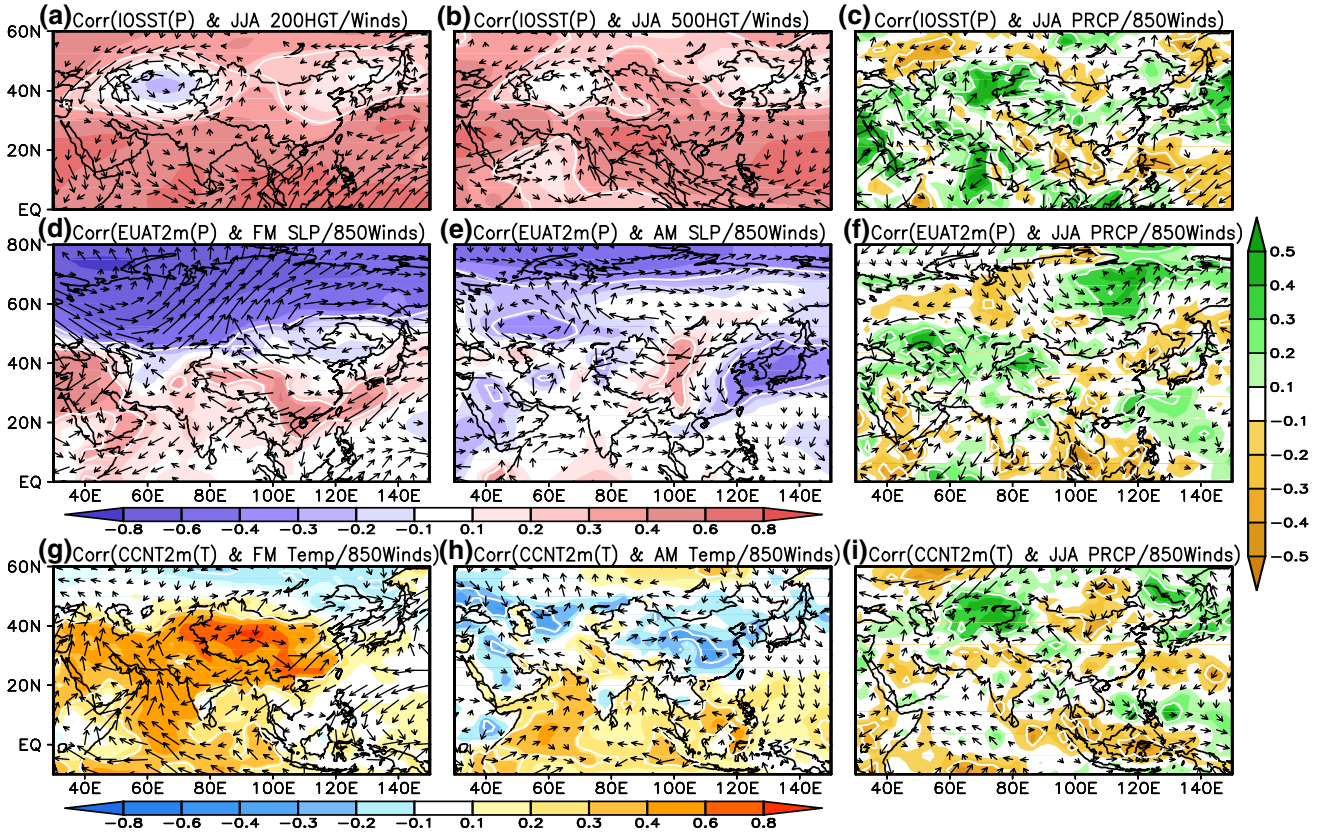


Fig. 6 **a**, **b** The correlation map of JJA mean (**a**) 200 hPa and **b** 500 hPa geopotential height (*color shading*)/wind anomalies (*vectors*) with reference to 0-month lead predictor IOSST (P) of PC2. **c** The same as in Fig. 4c but for the predictor IOSST (P) of PC2. **d–f** The same as in Fig. 4a–c but for the predictor EUAT2 m (P) of PC2. **g–i** The same as in Fig. 4d–f but for the predictor TPT2 m (T) of PC2. The *white* contours in each figure represent the correlation coefficient significant at the 90 % confidence level

Eurasia. For this reason, it is named as Central Eurasian mode. To the south of Tibetan Plateau, the extended positive SLP anomalies over northern Indian Ocean and western North Pacific weaken the tropical monsoon trough (Fig. 5d). The associated anomalous anticyclone in turn generates drying anomalies over the eastern India and Indo-China Peninsula (Fig. 5e).

The relationship of EOF 2 with equatorial SSTA is illustrated in Fig. 5c, f. This mode has a persistent positive relationship with tropical Indian Ocean and WNP SSTA (Fig. 5c, f). The warming SST over Indian Ocean may be regarded as a response to the weakened monsoon trough. This is because the southern branch of strengthened WPSH expands westward into tropical Indian Ocean, inducing warmed SST there since the total wind speed is reduced. In addition, eastern India and Indo-China Peninsula experience anomalous easterlies and suppressed rainfall associated with the weakened monsoon trough. As shown in Fig. 6a, an anomalous cyclone to the east of Caspian Sea (65°E, 40°N) is generated to the northwest of the suppressed rainfall area in the upper troposphere as a Rossby wave response (Ding and Wang 2005). This anomalous cyclone also exists in the middle and lower troposphere (Fig. 6b, c). Therefore, the anomalous cyclone is manifested as a quasi-barotropic structure and induces plenty of rainfall over western China and the vicinity (Fig. 6c). Thus, the western China rainfall has a negative relation with the strength of Indian summer monsoon, consistent with previous studies (Yatagai and Yasunari 1995). The spring persistent SSTA (FMAM mean) over Indian Ocean (40°E–120°E, 30°S–20°N), which can signal a weak South Asia summer monsoon, is selected as one predictor for this mode (Fig. 5f). This predictor is named as IOSST (P).

Another salient persistent precursor is the land warming over the central Siberia (Fig. 5f). The land warming occurs in the front of an anomalous Ural (mountain) trough during February and March (Fig. 6d). Warm and wet air is transported to northern Eurasia along the southwesterlies associated with the anomalous Ural trough, inducing anomalous land warming there. From FM to summer (JJA), it seems that the anomalous Ural low moves southward and arrives at central Asia in summer, inducing positive rainfall anomalies (Fig. 6e, f). For this reason, the persistent T2m anomaly over the central Siberia (80°E–120°E, 50°N–75°N),

which is called EUAT2 M (P), is selected as the second predictor (Fig. 5f).

A notable T2 m cooling tendency is noted over the Tibetan Plateau and vicinity. The cooling tendency is associated with more snow during AM, inducing weak south Asia summer monsoon (Fig. 6i). This signal can be regarded as an indirect precursor of more rainfall over central Asia. Thus, the third predictor named TPT2m (T) is defined as the AM-minus-FM 2 m air temperature anomalies averaged over (80°E–110°E, 20°N–40°N).

4 Predictability and prediction with P–E models

In the previous section, we have shown that the first two EOF modes have different origins and predictors. The potential maximum predictability and prediction of NWC summer rainfall are introduced in this section. Prediction of anomalous rainfall patterns over NWC includes two steps. First, each PC is predicted. Second, the rainfall prediction field is reconstructed using the sum of the two observed spatial EOF patterns multiplied by their corresponding predicted PCs. It should be pointed out that if the predictors used for prediction are 0-month ahead of June (i.e. the predictors include the information in May), the prediction is named as 0-month lead prediction. Similar definition applies to 1-month lead prediction.

4.1 Prediction of each PC with P–E models

In order to predict PC1, a P–E prediction model is established through stepwise multi-linear regression method using the three predictors shown in Fig. 3f, g. All predictors are selected by stepwise regression given the F-test at 95 % confidence level. The prediction (simulation) equation is

$$\mathrm{PC1}' = -0.28 \times \mathrm{WPSST(P)} - 0.35 \times \mathrm{CEPSST(T)} - 0.44 \times \mathrm{EUAT2m\ (P)} \quad (1)$$

The P–E prediction model can reproduce the rainfall variability of the first mode with a temporary correlation coefficient (TCC) skill of 0.75 for all 37 years.

In order to test the predictive capability of the empirical model, the cross-validation method (Michaelsen 1987) is performed to make a retrospective forecast (hindcast) for PC1. To lessen over-fitting problem, we leave 3 years of data out progressively centered on a forecast target year for the period 1979–2015, then train the model using the data in remaining years and finally apply the model to forecast the three target years. The hindcast PC1 using the P–E model in the cross-validated mode is shown by the red line in Fig. 7a compared with the corresponding observed PC1 (black line).The TCC skill of the 37-year cross-validated hindcast is 0.70. To further confirm whether the P–E model is actually useful, we use 1979–2003 as training period and derive prediction equation using the data during this training period. Then independent forecasts are made for the period 2004–2015. The independent forecast TCC skill is 0.71 (The green line in Fig. 7a).

Similarly, the P–E prediction model for PC2 is developed using the three predictors identified in Sect. 3.2. All predictors are relatively independent from each other. The prediction (simulation) equation is

$$\mathrm{PC2}' = 0.42 \times \mathrm{IOSST(P)} + 0.37 \times \mathrm{EUAT2m(P)} - 0.35 \times \mathrm{TPT2m(T)} \quad (2)$$

The simulation and cross validated hindcast skill during 1979–2015 are 0.72 and 0.67, respectively. The independent forecast TCC during the last 12 years is 0.60 (Fig. 7b).

The hindcast TCC skill in the cross-validation mode is significant at 99 % confidence level in both modes, which suggests that the two leading modes may be considered as predictable modes to a large extent. The high skills of independent forecast in the two modes also show that this set of P–E models can be used to make real time forecast.

4.2 Potential predictability

By far, we have shown that the first two EOF modes can be potentially predicted by the P–E models. In this subsection, we try to estimate the potentially maximum attainable prediction skill for the percentage of precipitation anomaly during summer over NWC. The observed predictable part of the predictand is reconstructed by the sum of the first two modes. Assuming that these modes can be predicted perfectly, the potentially maximum attainable prediction skill can be estimated from the correlation between the observed total field and the reconstructed predictable part of the rainfall using the two predictable modes.

The maximum attainable temporal correlation coefficient (TCC) skill at each grid point is shown in Fig. 8a. The domain-averaged correlation skill is 0.60. High predictability regions are found over the western Inner Mongolia and Xinjiang Province, which are consistent with the rainfall anomaly centers in the two modes.

The maximum attainable pattern correlation coefficient (PCC) skill for each year is shown by the black line in Fig. 9. The 37-year averaged PCC is 0.48. There is large year-to-year variation in the maximum attainable skill. There are 10 years during which the maximum attainable PCC skill is above 0.6 (1979, 1980, 1984, 1985, 1993, 1994, 2001, 2002, 2005 and 2010). 1988 and 1991 have relatively lower predictability. This may be due to the fact that the percentage rainfall fields of the 2 years have low projection onto the first two EOFs for the entire period of 1979–2015.

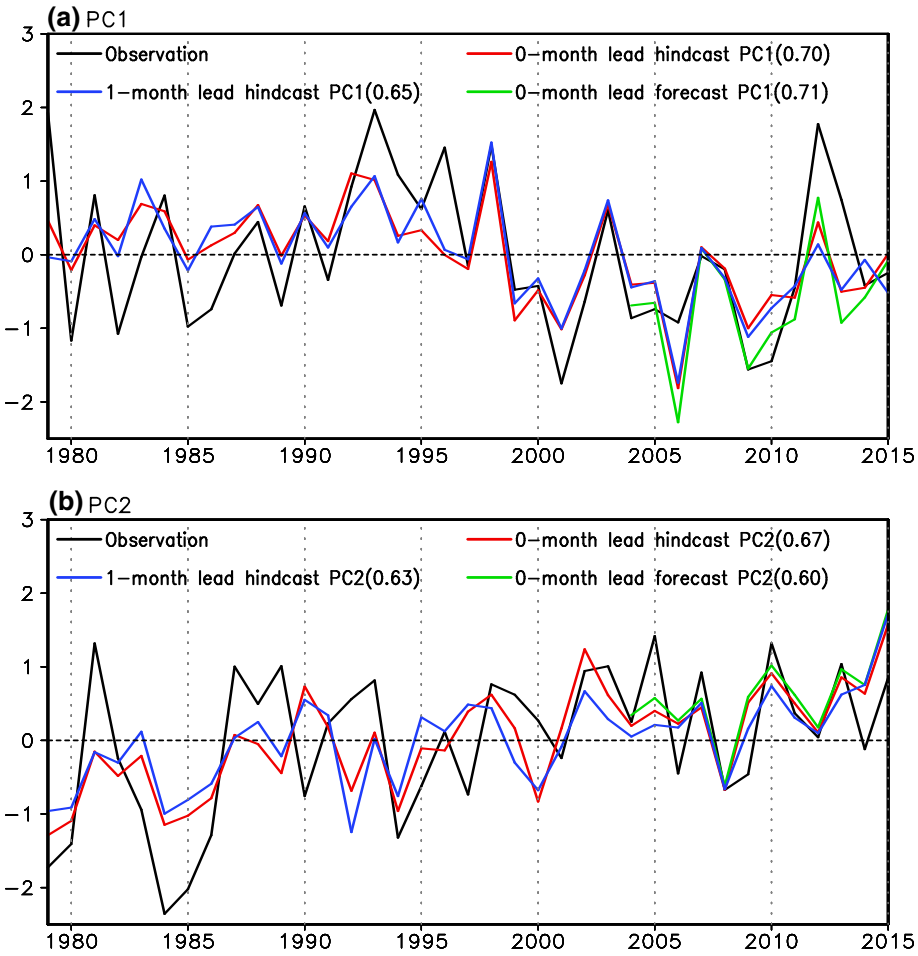


Fig. 7 **a** The corresponding PC of the first EOF mode in observation (OBS, *black line*), cross-validated 0-month lead prediction model (0 M, *red line*) and cross-validated 1-month lead prediction model (1 M, *blue line*) from 1979 to 2015. 0/1-month lead prediction model means the model is established using predictors which are 0/1-month ahead of June. The cross-validation was done by taking 3-year out around the predicted year. The 2004–2015 values shown by the *green line* are independent test predictions when the model is built using the data in the training period of 1979–2003. **b** The same as in Fig. 7a but for the second mode. The numbers within the parenthesis in the figure legend indicate the correlation coefficient between the observed and predicted PC

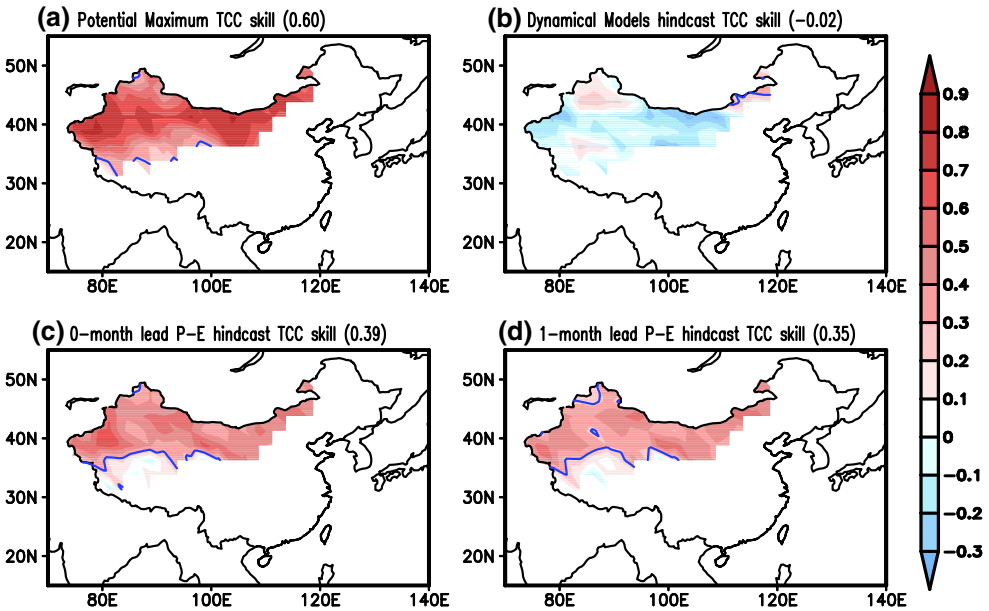


Fig. 8 The temporal correlation coefficient (TCC) skill for JJA precipitation prediction over NWC by using the **a** observed first two modes, **b** nine dynamical models' multi-model ensemble mean, **c** 0-month lead cross-validated P–E Models and **d** 1-month lead cross-validated P–E Models. The *dashed* contour is the TCC skill of 0.28 which is statistically significant at 90 % confidence level. The numbers within the parenthesis in the title of each figure indicate the domain-averaged TCC skill

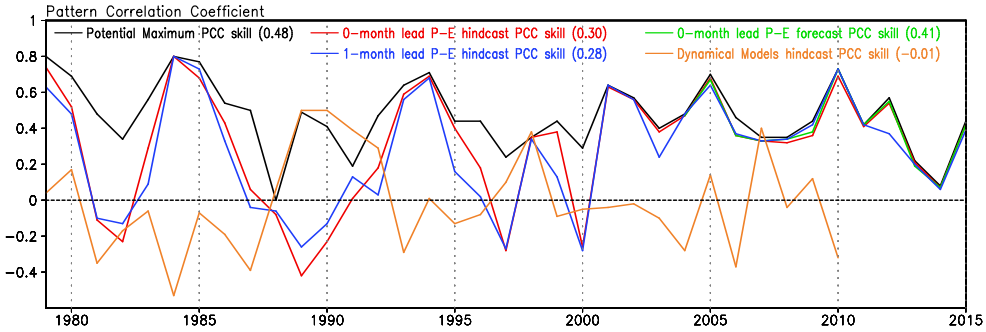


Fig. 9 The pattern correlation coefficient (PCC) skill for JJA precipitation prediction over NWC as a function of forecast year using the 3-year out cross-validated 0-month lead P–E model prediction (*red line*), and 1-month lead P–E model prediction (*blue line*). The potential attainable forecast skill obtained by using observed two PCs (OBS, *black line*) is also compared. The *orange line* shows the PCC skill by using the nine dynamical models' multi-model ensemble mean. The 2004–2015 values shown by the *green line* are independent forecast PCC skill when the model is built using the data in the training period of 1979–2003. The numbers within the parenthesis in the figure legend indicate the averaged PCC skill

4.3 The 0-month lead prediction of NWC summer rainfall

Given the fact that the first two EOF modes can be potentially predicted by the P–E models, we may reconstruct the prediction field of NWC summer rainfall by using linear combination of the first two observed spatial EOF patterns multiplied by their corresponding predicted PCs. Note that all predictors in Sect. 3 are 0-month ahead of June, so the prediction here is a 0-month lead prediction. But in practice, the forecast can be made 5 days before June 1st because the last 5 days' values for the predictors can be estimated by weather forecast.

We use the TCC and PCC between the observed rainfall field and predicted patterns as a measure of prediction skill. The spatial distribution of the hindcast TCC skill in the cross-validated mode is given in Fig. 8c. The domain-averaged TCC skill is 0.39. The whole domain shows similar hindcast skill except the most western part of China. The red line in Fig. 9 shows the time series of PCC skill for each year between observed precipitation and hindcast field. The 37-year averaged PCC skill is 0.30. The green line in Fig. 9 indicates the independent forecast skill during 2004–2015. The forecast pattern is derived from the forecast PCs shown by the green line in Fig. 7 and the corresponding EOF patterns derived from 1979 to 2003. This independent test can rigorously reflect the ability of the P–E prediction model in NWC summer rainfall forecast. Actually, the forecast skill is similar to the hindcast skill during the latest 12 years. The 12-year mean PCC value is 0.41.

4.4 The 1-month lead prediction of NWC summer rainfall

In order to make timely management decisions, a longer lead time is appreciated. In this case we also try to make a 1-month lead prediction. The corresponding predictors of each PC are searched based on leading correlation maps with February-to-April persistent anomalies and April-minus-FM tendency anomalies. The selected predictors are almost the same as those in 0-month lead prediction (Table 1), which suggests that these predictors are basically stable with the lead time.

The blue lines in Fig. 7a, b give the cross-validated predicted PC in each mode. The hindcast TCC for the two modes are 0.65 and 0.63, respectively. The skills are systematically lower than the 0-monthlead hindcast skills. The corresponding domain-averaged TCC and 37-year mean PCC skill are 0.35 and 0.28, respectively (Figs. 8d and 9).

To compare the 1-month lead hindcast skills using P–E models with the current status of the dynamical seasonal prediction, we examine the hindcast skills of nine state-of-the-art atmosphere–ocean–land coupled models, including (1) NCEP (National Center for Environmental Prediction) CFS version 2 (Saha et al. 2014), (2) GFDL (Geophysical Fluid Dynamics Laboratory) CM version 2.1 (Delworth et al. 2006), (3) FRCGC (Frontier Research Center for Global Change) SINTEX-F model (Luo et al. 2005), (4) ECMWF (European Center for Medium Range Weather Forecasting) model (Gregory et al. 2000; Molteni et al. 2011), (5) INGV (Instituto Nazionale de Geofisica e Vulcanologia in Italy) model (Madec et al. 1998), (6) IFM-GEOMAR (Leibniz Institute of Marine Sciences at Kiel University) IFM model (Jungclaus el al. 2006), (7) POAMA (Predictive Ocean Atmosphere Model for Australia) P24A model (Zhong et al. 2005), (8) UKMO (United Kingdom met office) model (Roeckner 1996; Marsland et al. 2003), and (9) MF (Centre National de Recherches Meteorologiques in France) model (Madec et al. 1997; Deque 2004). These models are collected in the APCC/CliPAS project (Wang et al. 2009b). A multi-model ensemble

Table 1 Definition of each predictor selected for the *1-month lead prediction model* of each PC and the corresponding simulation equation

MODE	Name	Definition	Prediction equation
EOF1	WPSST(P)	FMA SST 145°E–165°E,20°S–15°N	−0.24 × WPSST(P) −0.41 × CEPSST (T) −0.39 × EUAT2 m (P)
	CEPSST(T)	April-minus-FM SST 160°E–90°W, 0–15°N	
	EUAT2 m(P)	FMA T2 m (50°E–80°E, 20°N–50°N) minus (90°E–120°E, 50°N–70°N)	
EOF2	IOSST(P)	FMA IO SST 40°E–120°E, 30°S–20°N	0.49 × IOSST (P) +0.40 × EUAT2 m(P) −0.37 × TPT2 m(T)
	EUAT2 m(P)	FMA T2 m 80°E–120°E, 50°N–75°N	
	TPT2 m(T)	April-minus-FM T2 m (80°E–110°E, 20°N–40°N)	

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

Figure 8b and the orange line in Fig. 9 show the TCC and PCC skill by using the nine climate models' MME. As can be seen, the TCC skill barely exceeds 90 % confidence level (Fig. 8b). The averaged skill of the dynamic MME hindcast is not distinguishable from zero during 1979–2010. The prediction is notoriously poor in the NWC. Therefore, compared with the 1-month lead MME hindcast, the P–E prediction model shows higher skills, suggesting the advantage of PMA in terms of predicting rainfall over mid-latitude land areas.

5 Summary

Northwest China (NWC) consists of an agro-pastoral transition zone and desert regions. It receives least rainfall in China and is regarded as the arid and semi-arid regions. The present study investigates the predictability and prediction of the percentage (fractional) rainfall anomaly (i.e., rainfall anomaly normalized by the local mean climatological rainfall) during June to August over NWC for the 37-year period of 1979–2015. These months are of great importance for plants and agriculture.

The first two EOF modes, which explain about 40 % of the total variance, are identified as the most important modes. The first mode reflects the rainfall variability over northeastern Eurasia including Inner Mongolia, Mongolia and central Siberia. This mode is mainly driven by spring persistent SST cooling over the tropical western Pacific and cooling tendency across spring over central and eastern Pacific. The second mode is associated with the meridional dipole distribution over the central Eurasia. It has a close relationship with tropical Indian Ocean SSTA which is a response to anomalous south Asian summer monsoon. Besides tropical SSTA, anomalous land warming over Eurasia continent also contributes to the two modes.

These findings provide dynamical insights into the physical processes that control the precipitation variability over NWC, and may have important implications for better prediction. To better understand the source of the potential predictability and to improve prediction skill, a set of 0-month lead physical–empirical (P–E) models is established for prediction of the first two leading PCs. The physically meaningful and statistically robust predictors are selected based on either persistent surface temperature anomalies (SSTA over ocean and 2 m air temperature anomalies over continent) during late winter to spring (February to May) or the tendency from late winter (February–March) to spring (April–May). Emphasis is put on discussion of the physical linkage between the predictors and predictand. Only three predictors are used for each PC. The cross-validated TCC skill for prediction of each PC is 0.70 and 0.67, respectively, which are both significant at 99 % level. The 10-year independent forecast skill is comparable to the cross-validated skills. The high TCC skills of independent forecast of the period 2004–2015 indicate the usefulness of the P–E models in the practical prediction. Thus, to a large extent, the first two modes can be regarded as predictable mode, at least in terms of the P–E model estimation. Besides, a set of 1-month lead P–E models is also developed accordingly for the purpose of making timely arrangement.

The predictable part of summer rainfall over NWC can be reconstructed by the two predictable modes. The area-averaged potentially maximum attainable skill over the entire domain is 0.60. The potential maximum attainable

pattern correlation coefficient (PCC) skill may fluctuate from year to year with an averaged PCC skill about 0.48.

Similarly, the 0-month and 1-month lead prediction field can be derived using the corresponding predicted PCs. We find that the 0-month lead models have slightly higher skills than 1-month lead models. That means that the signals during May are useful for summer rainfall prediction over NWC. Therefore, it is suggested that by late May, a more reliable forecast of JJA rainfall should be implemented in the operational centers. Even so, the 1-month prediction skill using P–E models shows higher skill than dynamical models' multi-model ensemble mean, suggesting that the P–E prediction model may be a useful approach for seasonal prediction compared with the current dynamical models over midlatitude continents and the current prediction models may have large room to improve.

In this study, we mainly focus on the impact of lower boundary forcing on the rainfall variability over NWC. The physically meaningful and statistically robust predictors are selected based on atmospheric lower boundary anomalies. A recent study demonstrated that the summer precipitation variability over most western part of China was also closely associated with the meridional teleconnection pattern around 50°E–80°E and the zonal teleconnection pattern along the Asian westerly jet in summer (Huang et al. 2015). In the future, the related predictors will be added in the pool of predictors to see whether this kind of predictors can improve the prediction skill of NWC summer rainfall.

It should be noted that the predictable modes identified in this study are confined for the recent 37 years. We expect that these modes are still valid for the next few years, so useful prediction can be made. However, both the rainfall variability and the source of the predictability (predictors) may be generally involved with decadal change. Thus, continuous detection of long term changes and modifications of the predictors/prediction equations are required.

Acknowledgments The work was supported by the Global Change Research Program of China (2015CB953904, 2012CB955604) and National Natural Science Foundation of China (No. 41575067). This work was also supported by Atmosphere–Ocean Research Center that is sponsored by Nanjing University of Information Science and Technology. We also acknowledge support from the International Pacific Research Center (IPRC). This is publication No 9830 of the SOEST, publication No 1213 of IPRC and publication No 128 of Earth System Modeling Center (ESMC).

References

Madec GP et al (1998) OPA version 8.1, Ocean general circulation model reference manual. LODYC Technical Report No. 11, Paris

Chen LX, Dong M, Shao YN (1992) The characteristics of interannual variations of the East Asian Monsoon. J Meteor Soc Jpn 70:397–421

Dee DP et al (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597

Delworth TL et al (2006) GFDL's CM2 global coupled climate models. Part I: formulation and simulation characteristics. J Clim 19:643–674

Deque M (2004) Seasonal predictability of tropical rainfall: probabilistic formulation and validation. Tellus 53(4):500–512

Ding QH, Wang B (2005) Circumglobal teleconnection in the Northern Hemisphere Summer. J Clim 18:3483–3505

FAO, Unesco, WMO (1977) World map of desertification. Food and Agricultural Organization, Rome

Gong DY, Shi PJ, Wang JA (2004) Daily precipitation changes in the semi-arid region over northern China. J Arid Environ 59(4):771–784

Gregory D et al (2000) Revision of convection, radiation and cloud schemes in the ECMWF Integrated Forecasting System. Q J R Meteorol Soc 126:1685–1710

Guo QY, Wang J (1988) A comparative study on summer monsoon in China and India. J Tropical Meteorol 4:53–60 **(in Chinese with English abstract)**

Huang B et al (2014) Extended reconstructed sea surface temperature version 4 (ERSST.v4): Part I. Upgrades and intercomparisons. J Clim 28(3):911–930

Huang W, Feng S, Chen J et al (2015) Physical mechanisms of summer precipitation variations in the Tarim basin in Northwestern China. J Clim 28(9):3579–3591

Huffman GJ, Bolvin DT, Adler RF (2011) Last updated GPCP version 2.2 combined precipitation data set. WDC-A, NCDC, Asheville, NC. http://www.ncdc.noaa.gov/oa/wmo/wdcamet-ncdc.html

Jungclaus JH et al (2006) Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM. J Clim 19:3952–3972

Kripalani RH, Singh SV (1993) Large scale aspects of India–China summer monsoon rainfall. Adv Atmos Sci 10:71–84

Lee JY (2013) Seasonal prediction and predictability of the Asian winter temperature variability. Clim Dyn 41:573–587

Lee JY et al (2011) How predictable is the Northern Hemisphere summer upper-tropospheric circulation? Clim Dyn 37:1189–1203

Li J, Wang B (2015) How predictable is the anomaly pattern of the Indian summer rainfall? Clim Dyn. doi:10.1007/s00382-015-2735-6

Liu W (2014) Extended reconstructed sea surface temperature version 4 (ERSST.v4): Part II. Parametric and structural uncertainty estimations. J Clim 28(3):931–951

Luo JJ et al (2005) Seasonal climate predictability in a coupled OAGCM using a different approach for ensemble forecast. J Clim 18:4474–4497

Lv TZ, Deng SG, Hu YJ, Zhang K (2015) Quantitative prediction of precipitation in the east of northwest china during the flood period by using the year-to-year increment approach. J Arid Meteorol 33(3):386–394 **(in Chinese with English abstract)**

Madec GP et al (1997) OPA release 8, Ocean general circulation model reference manual. LODYC Internal Report, Paris

Marsland SJ et al (2003) The Max Planck Institute global ocean/sea ice model with orthogonal curvilinear coordinates. Ocean Model 5:91–127

Michaelsen J (1987) Cross-validation in statistical climate forecast models. J Clim Appl Meteor 26(11):1589–1600

Molteni F et al (2011) The new ECMWF seasonal forecast system (System 4). ECMWF Tech Memorandum 656:1–49

North GR, Bell TL, Cahalan RF, Moeng FJ (1982) Sampling errors in the estimation of empirical orthogonal functions. Mon Wea Rev 110(7):699–706

Panofsky HA, Brier GW (1968) Some application of statistics to meteorology. Pennsylvania State University Press, University Park, PA

Roeckner E (1996) The atmospheric general circulation model ECHAM4: model description and simulation of presentday climate. Max Planck Institut fur Meteorologie, Report No. 218, Hamburg, Germany

Saha S et al (2014) The NCEP climate forecast system version 2. J Clim 27:2185–2208

Samel AN, Wang WC, Liang XZ (1999) The monsoon rainband over China and relationships with the Eurasian circulation. J Clim 12:115–131

Shen S, Lau KM (1995) Biennial oscillation associated with the East Asian summer monsoon and tropical sea surface temperature. J Meteor Soc Jpn 73:105–123

Wang WC, Li K (1990) Precipitation fluctuation over semi-arid region in Northern China and the relationship with El Nino/Southern Oscillation. J Clim 3:769–783

Wang B, Zhang Q (2002) Pacific-East Asian teleconnection, Part II: how the Philippine Sea anticyclone established during development of El Nino. J Clim 15:3252–3265

Wang B, Wu R, Fu X (2000) Pacific-East Asia teleconnection: how does ENSO affect East Asian climate? J Clim 13:1517–1536

Wang B, Lee JY, Kang IS, Shukla J, Hameed SN, Park CK (2007) Coupled predictability of seasonal tropical precipitation. CLIVAR Exchanges 12:17–18

Wang B, Liu J, Yang J, Zhou T, Wu Z (2009a) Distinct principal modes of early and late summer rainfall anomalies in East Asia. J Clim 22:3864–3875

Wang B et al (2009b) Advance and Prospect of Seasonal Prediction: assessment of the APCC/CliPAS 14-model ensemble retroperspective seasonal prediction (1980-2004). Clim Dyn 33:93–117

Wang B, Xiang B, Lee JY (2013) Subtropical high predictability establishes a promising way for monsoon and tropical storm predictions. PNAS 10:2718–2722

Wang B, Lee JY, Xiang BQ (2014) Asian summer monsoon rainfall predictability: a predictable mode analysis. Clim Dyn 44:1–14

Wu ZW, Yu LL (2016) Seasonal prediction of the East Asian summer monsoon with a partial-least square model. Clim Dyn 46:3067–3078

Wu ZW, Wang B, Li JP, Jin FF (2009) An empirical seasonal prediction model of the East Asian summer monsoon using ENSO and NAO. J Geophys Res 114:D18120. doi:10.1029/2009JD011733

Xing W, Wang B, Yim S-Y (2016a) Peak-summer East Asian rainfall predictability and prediction part I: Southeast Asia. Clim Dyn 47:1–13

Xing W, Wang B, Yim S-Y (2016b) Long-lead seasonal prediction of china summer rainfall using an EOF–PLS regression-based methodology. J Clim 29(5):1783–1796. doi:10.1175/jcli-d-15-0016.1

Yasunari T (1990) Impact of Indian monsoon on the coupled atmosphere/ocean system in the tropical Pacific. Meteor Atmos Phys 44:29–41

Yatagai A, Yasunari T (1995) Interannual variations of summer precipitation in the arid/semi-arid regions in China and Mongolia: their regionality and relation to the Asian summer monsoon. J Meteor Soc Jpn 73:909–923

Ye DZ, Chen PQ (1992) Global change in China: a preliminary study. Meteorological Press, Beijing **In Chinese**

Yim SY, Wang B, Xing W (2014) Peak-summer east asian rainfall predictability and prediction Part II: extratropical East Asia. Clim Dyn 89–98

Zhao CG, Li ZC (2013) Regional modeling forecast of precipitation in summer in Northwest China. J Desert Res V 33(5):1544–1551 **(in Chinese with English abstract)**

Zhong AH, Hendon H, Alves O (2005) Indian Ocean variability and its association with ENSO in a global coupled model. J Clim 18:3634–3649