Summer precipitation prediction in the source region of the Yellow River using climate indices

Feifei Yuan, Ronny Berndtsson, Cintia Bertacchi Uvo, Linus Zhang and Peng Jiang

ABSTRACT

The summer precipitation from June to September in the source region of the Yellow River accounts for about 70% of the annual total, and its decrease would cause further water shortage problems. Consequently, the objectives of this study are to improve the understanding of the linkages between the precipitation in the source region of the Yellow River and global teleconnection patterns, and to predict the summer precipitation based on revealed teleconnections. Spatial variability of precipitation was investigated based on three homogeneous sub-regions. Principal component analysis and singular value decomposition were used to find significant relationships between the precipitation and global teleconnection patterns using climate indices. A back-propagation neural network was developed to predict the summer precipitation using significantly correlated climate indices. It was found that precipitation in the study area is positively related to North Atlantic Oscillation, West Pacific Pattern and El Niño Southern Oscillation, and inversely related to Polar Eurasian pattern. Summer precipitation was overall well predicted. The Pearson correlation coefficient between predicted and observed summer precipitation was, in general, larger than 0.6. The results can be used to predict the summer precipitation and to improve integrated water resources management in the Yellow River basin.

Key words | source region of the Yellow River, summer precipitation prediction, teleconnection pattern

INTRODUCTION

The Yellow River is extremely important to China since it supplies fresh water for 110 million people and 13% of China’s irrigated farming land (Wang et al. 2006). The source region of the Yellow River contributes about 35% of the basin’s total streamflow playing an important role in meeting downstream water resources requirements (Zheng et al. 2007). Consequently, it is an important area affecting agricultural productivity, municipal and industrial water supply for the entire basin. The last 50 years have witnessed a generally increasing trend in temperature and decreasing trend in annual precipitation and streamflow for the Yellow River source region. Decreasing precipitation during the monsoon period (June–September) is, to a great extent, causing the annual precipitation decrease (Yuan et al. 2015). Continued precipitation and streamflow decrease may cause further water shortage problems in the downstream of the Yellow River.

Climate variability is closely linked to patterns of flood and drought in different areas of the world and strongly affecting local and regional scale climate through teleconnections. Teleconnections are statistical associations among climate variables separated by large distances. They are a consequence of large-scale dynamics between the ocean and atmosphere linking disparate regional climates into one unified, global climatic system (Leathers et al. 1991; Jiang et al. 2013). Numerous studies have shown that
climate variability has a strong impact on basin water resources through changes in hydrologic variables (Uvo 2003; Lorenzo et al. 2008; Rana et al. 2012; Peng et al. 2013). The interannual variability in local hydroclimatic variables (temperature, precipitation and streamflow) could be a reflection of low-frequency climatic fluctuations. Understanding the linkage between local precipitation and global teleconnection patterns is essential for water resources management. Further, it could improve the ability to predict the local precipitation based on physical reasoning (Redmond & Koch 1991; Hartmann et al. 2008).

Much research has been devoted to the issue of precipitation change in the source region of the Yellow River (Tang et al. 2008; Cong et al. 2009; Hu et al. 2012; Zhang et al. 2015; Zhu et al. 2016). Research during recent decades has also tried to link global atmospheric circulation (e.g., ENSO (El Niño–Southern Oscillation) and global SST (sea surface temperature) with more localized hydrological response in the Yellow River basin (Lau & Weng 2001; Fu et al. 2007; Xu et al. 2007; Yasuda et al. 2009). Wang et al. (2000) found that ENSO events can affect the East Asian climate through the Pacific–East Asian teleconnection, with an anomalous anti-cyclone east of the Philippines during El Niño events. Zhang et al. (2013b) examined the influence of ENSO on precipitation in the East River basin, south China. Lü et al. (2011) explored the rainfall–ENSO relationship using Southern Oscillation Index for the source region of the Yellow River. Feng et al. (2014) investigated the influence of Indian Ocean subtropical dipole on spring rainfall over China, and found that it has a significant correlation with the boreal spring rainfall over the Yellow River valley. Lorenzo et al. (2008) examined the links between circulation weather types and teleconnection patterns and their influence on precipitation patterns in Galicia. Hartmann et al. (2008) predicted summer precipitation in the Yangtze River basin using neural networks and climate indices. Chan & Shi (1999) predicted the summer monsoon rainfall over south China using climate indices. The Indian summer monsoon rainfall was shown to be well predicted by neural networks using different climate indices (Venkatesan et al. 1997; Sahai et al. 2003). In view of the above, the summer precipitation in the source region of the Yellow River could possibly be predicted by exploring its relationship with global teleconnection patterns.

Thus, in this paper we aim at improving the knowledge by investigating the relationship between precipitation in the source region of the Yellow River and global teleconnection patterns, and predicting the summer precipitation using significantly correlated climate indices by an artificial neural network (ANN). It is important to quantify the above relationships and to develop quantitative prediction techniques for the source region of the Yellow River, and establishing such links would improve the physical understanding of rainfall variability with important implications for water resources management. Also, the predicted precipitation is of utmost importance for food production and flood mitigation.

**STUDY AREA AND METHODS**

**Study area and data**

The source region of the Yellow River is located on the northeast Qinghai-Tibet Plateau between 32° 12’–35° 48’ N and 95° 50’–103° 28’ E and includes the area above the Tong-naihai runoff observation station. The area is $12.2 \times 10^4$ km$^2$ accounting for 16% of the Yellow River basin, and it has a great elevation change from 2,670 m in the east to 6,253 m in the west (Figure 1). Grassland covers 80% of the catchment and includes typical alpine swamp, steppe and shrub meadows. The area of lakes and swamps is about 2,000 km$^2$. There is a permanent snowpack and glaciers in the southern Animaqing, Bayankala and Northern Qilian mountains. The area has a comparably low population density with a total of about half a million inhabitants. The region is therefore regarded as relatively unaffected by human activities (Zheng et al. 2009; Hu et al. 2011). Neither large irrigation projects nor large dams exist in the area even though the population increase of humans and domestic livestock is increasingly affecting the grass cover and soil erosion.

Climatologically the area belongs to the semi-humid region of the Tibetan Plateau subfrigid zone and around 70% of the annual precipitation in this area fall during the wet summer season (June–September) due to the southwest monsoon from the Bay of Bengal (Hu et al. 2011). Thus, the climate of the source region of the Yellow River is greatly
influenced by the southwest monsoon and the East Asian summer monsoon (Ding & Chan 2005). The earliest onset of the East Asian summer monsoon occurs in the central and southern Indochina Peninsula. It displays a distinct stepwise northward and northeastward movement and then finally penetrates into the upper Yellow River from the south of China (Ding & Chan 2005). The effects of atmospheric circulation are, in general, different for the upper and lower Yellow River. The monsoon rain belt in the upper part is caused by southeasterly flow while the corresponding monsoon rain belt in the lower part is influenced by southwesterly flow (Qian et al. 2002). This causes differences in spatial distribution of summer precipitation between the two parts of the Yellow River. The upper part is characterized by low temperatures, sharp day–night temperature contrasts, long cold and short warm seasons, and intense sunlight (Liang et al. 2010). The precipitation in the region is generally of low intensity, long duration (10–30 days) and covers a large area (>100,000 km²) (Zheng et al. 2007; Hu et al. 2011). Snowfall is concentrated from November to March, when more than 78% of the total precipitation falls as snow. However, the total amount of annual snowfall accounts for less than 10% of the annual precipitation (Hu et al. 2011). The potential evaporation is 1,300–1,400 mm/year (Liang et al. 2010).

Monthly precipitation data from 1961 to 2010, collected from 10 meteorological stations (Figure 1): Xinghai, Tongde, Zeku, Henan, Maduo, Dari, Jiuzhi, Maqu, Ruoergai and Hongyuan, were obtained from the China Meteorological Administration (CMA). The data quality has previously been checked by the CMA. The Shuttle Radar Topography Mission 90 m digital elevation data were downloaded from the Consortium for Spatial Information. Global monthly climate indices data representing teleconnection patterns, including North Atlantic Oscillation (NAO), East Atlantic (EA) Pattern, West Pacific (WP) Pattern, Pacific/North American (PNA) Pattern, East Atlantic/West Russia Pattern (EA/WR), India Ocean Dipole, El Niño-Southern Oscillation (NINO3.4), Scandinavia Pattern, Polar/Eurasia Pattern (POL) and Pacific Decadal Oscillation (PDO), were obtained from the National Weather Service, Climate Prediction Centre. Further explanation of each teleconnection pattern is described in Washington et al. (2000) and Barnston & Livezey (1987).

**METHODS**

To investigate precipitation trend and variability in the source region of the Yellow River, annual precipitation time series from 1961 to 2010 in different zones were examined by linear regression method and the non-parametric Mann–Kendall test. Mann–Kendall test is independent of the statistical distribution of the data. Statistical significance of the trend was evaluated at the 0.05 level of significance against the null hypothesis that there is no trend for the data series.
Principal component analysis (PCA) and singular value decomposition (SVD) were used to find relationships between precipitation in the source region of the Yellow River and teleconnection patterns using climate indices. PCA is a multivariate data analysis tool that offers a way to present complex data in a simplified way to identify relations between different parameters. It maximizes variance explained by weighted sum of elements in two or more fields and identifies linear transformations of the data set that concentrates as much of the variance as possible into a small number of variables (Uvo 2003; Rana et al. 2012). The PCA biplot is used to visualize the magnitude and sign of each variable’s contribution to the first two principal components, and how each observation is represented in terms of those components.

SVD is performed on the cross-covariance matrix of fields of two data sets and isolates the combinations of variables within the fields that tend to be linearly related to one another by maximizing the covariance between them (Wallace et al. 1992; Rana et al. 2012). Here, SVD was conducted to the cross-covariance matrix of the $s(t, x)$ matrix of $x$ climatic indices (called as predictand) at $t$ months and the $z(t, y)$ matrix composed by precipitation for three homogeneous precipitation zones at $t$ months. The latter one is called as predictor. All time series were standardized to zero mean and unit standard deviation prior to use in the SVD. The SVD of the cross-covariance matrix of two fields yields two matrices of singular vectors and one set of singular values. A singular vector describes spatial patterns for each field which has overall covariance given by the corresponding singular value (Uvo et al. 1998). In our case, the data time series $s(t)$ and $z(t)$ for each variable can be expanded in terms of a set of $N$ vectors, called patterns (Bretherton et al. 1992). Here, we try to estimate the predictor on basis of the predictand.

$$s(t) \leftarrow \hat{s}(t) = \sum_{k=1}^{N} a_k(t)p_k$$  \hspace{1cm} (1)

$$z(t) \leftarrow \hat{z}(t) = \sum_{k=1}^{N} b_k(t)q_k$$  \hspace{1cm} (2)

The time series $a_k(t)$ and $b_k(t)$ are called expansion coefficients; $p_k$ and $q_k$ are the patterns (Bretherton et al. 1992). The expansion coefficients are calculated as weighted linear combination of variables in data.

$$a_k(t) = \sum_{l=1}^{N_z} u_{lk}s_l(t) = u_k^Ts(t)$$  \hspace{1cm} (3)

$$b_k(t) = \sum_{l=1}^{N_z} v_{jk}z_l(t) = v_k^Tz(t)$$  \hspace{1cm} (4)

The vectors $u_k$ and $v_k$ are called weight vectors. Together, each pair of patterns, the corresponding pair of weight vectors and the pair of expansion coefficients defines a mode, which combines the variability observed in different fields. The variables in the $s$ and $z$ fields are subscripted by $i$ and $j$, respectively, and individual modes by $k$. Homogeneous correlation maps of the left and right fields from SVD show correlation coefficients between the values of one field and the singular vector of the other field (Uvo et al. 1998). In our case, the patterns shown by the heterogeneous correlation maps for the $k^\text{th}$ SVD expansion mode indicate how well the pattern of the precipitation anomalies relate to the $k^\text{th}$ singular vector of climate indices. The correlation coefficients are a good indication of strength of the relationship between the two fields. This can be used to compare the relative importance of a particular mode in the expansion. A detailed procedure for the statistical procedure can be found in Bretherton et al. (1992) and Wallace et al. (1992).

A back propagation neural network was applied for predicting summer precipitation in the source region of the Yellow River using the significantly correlated global climate indices. ANN is a powerful tool for prediction of meteorological phenomena involving many complex and nonlinear physical processes (Uvo et al. 2000; Sahai et al. 2003; Hartmann et al. 2008; Yasuda et al. 2009; Li et al. 2013). The significantly correlated global climate indices were used as input variables, and the summer precipitation (June–September) in the source region of the Yellow River was used as output variable in the ANN. The architecture of the neural network was determined by a ‘trial and error’ approach. Two hidden layers were placed between the input and output layers, with three and two nodes, respectively. The activation function ‘tansig’ was used for
this purpose, which is given as:

\[ \text{tansig}(x) = \frac{2}{1 + e^{-2x}} - 1 \] (5)

The four-layer ANN was connected by weights. A training period (1961–1995) and a validation period (1996–2010) were selected for optimization of the weights. Pearson product-moment correlation coefficients \( r \) between observed and predicted summer precipitation were calculated for the assessment of model performance.

**RESULTS**

**Precipitation characteristics**

Due to the large and diverse topography of the area, the precipitation in the source region of the Yellow River shows great spatial and temporal variability. It has a strongly increasing gradient from the northwest part with approximately 309 to 755 mm/year in the southeast, and the topographical gradient of rainfall and the annual rainfall variability are greatly influenced by the southeasterly summer monsoon flow (Yuan et al. 2015). In order to identify the precipitation spatial variability in the source region of the Yellow River, cluster analysis was performed to separate the precipitation stations into homogeneous regions using monthly precipitation data from 1961 to 2010 (Yuan et al. 2016). Figure 2 shows the outcome of this analysis and that the region can be divided into three homogeneous areas. The precipitation stations were thus divided into three zones, namely, 1: Xinghai, Tongde and Maduo; 2: Zeku, Henan and Dari; and 3: Jiuzhi, Maqu, Ruoergai and Hongyuan.

The linear trends for annual precipitation were quantified using the Mann–Kendall test (Table 1, Figure 3). As seen from Table 1 and Figure 3, mean annual precipitations for the three different zones are 365.0, 517.9 and 692.4 mm/year for zone 1, 2, and 3, respectively.
zone 1, zone 2 and zone 3, respectively. Zone 1 had a non-significantly increasing trend with 0.24 mm/year during 1961 to 2010. In contrast, both zone 2 and zone 3 had significantly decreasing annual precipitation. The decrease was 1.49 and 1.79 mm/year for zone 2 and zone 3, respectively. The mean annual precipitation decrease for the whole area is approximately 0.6 mm/year. Previous studies have indicated a small but statistically non-significant trend for precipitation in the source region of the Yellow River with a mean annual precipitation decrease corresponding to approximately 0.6 mm/year since 1960 (Liu & Zheng 2004; Zheng et al. 2007; Tang et al. 2008; Yuan et al. 2013). However, dividing the area into homogeneous zones displays clear differences in precipitation characteristics and a clearer picture of the spatially dependent trend. This trend is the strongest for the wettest area of the source region and then gradually decreases with decreasing annual precipitation.

**Relationships between precipitation and teleconnection patterns**

The precipitation for the three identified homogeneous zones in the source region of the Yellow River was tested for relation against different climate indices using PCA and SVD. Monthly precipitation and climate indices for the same period from 1961 to 2010 were used for analysis. PCA revealed a close relationship that is direct or inverse between the precipitation for the three zones and climate indices. The first two modes of PCA were analysed as they represented a major variance, with explained variance of 22.8% and 18.0%, respectively. Figure 4 gives the PCA biplot, and it is seen that the precipitation for the three zones is, in general, well represented in the first PCA mode. It is positively related to NAO, WP and NINO3.4, and it is negatively related to POL. The PCA analysis shows strong evidence that the precipitation in the different zones is closely related to the different teleconnection patterns.

SVD was applied to the cross-variance matrix between monthly precipitation for the three zones and monthly climate indices for the same period from 1961 to 2010. Figure 5 presents the time series of precipitation for the different zones and climate indices in the first mode of SVD with explained variance of 84.5%. A similar monthly variation shows that precipitation for the different zones is closely related to the teleconnection patterns. Table 2 shows the heterogeneous correlation for the different indices. It is noteworthy that the precipitation for the three zones is positively related in first mode to NAO, WP and NINO3.4, and negatively related to POL at 0.05 significance level. The result of SVD confirms the relationship from the
PCA results between precipitation for the three zones and climate indices.

**Summer precipitation prediction**

Based on the revealed relationships between precipitation in the source region of the Yellow River and global teleconnection patterns, a back propagation neural network was developed to predict the summer precipitation (June–September) using significantly correlated climate indices as input layer, including NAO, WP, POL and NINO3.4. One limitation of ANN analysis is that the model results vary depending on the random setting of the initial weights. Therefore, the neural network model run for one experiment was repeated 10 times. The predicted summer precipitation refers to the most accurate results of 10 model runs. Here, a ‘trial and error’ process was used to choose different nodes for the two hidden layers. Table 3 shows the Pearson correlation coefficients between observed and predicted summer precipitations for training and validation periods. The correlation coefficients for the validation period are generally larger than 0.6. Figure 6 shows the observed and predicted summer precipitation for the validation period depending on the different zones and the whole area.

**DISCUSSION**

The results revealed in this study showed that ENSO, NAO, WP and POL events have an influence on precipitation in the source region of the Yellow River. Xu et al. (2007) found that the La Nina phase corresponds to a relatively rainier season in the Yellow River basin. Fu et al. (2013) examined the trend and variability of extreme rainfall events in China and found that it is mainly influenced by ENSO and the magnitude of the East Asian monsoon. Yuan et al. (2016) examined the summer precipitation in the source region of the Yellow River teleconnections with global sea surface temperatures, and found that higher sea surface temperature in equatorial Pacific areas corresponding to El Niño coincides with less summer precipitation. Cuo et al. (2012) showed that precipitation change in winter at the northern Tibetan Plateau could be attributed to changes in the East Asian westerly jet, NAO and ENSO. Zhang et al. (2013a) also observed that the warm phase of the North Atlantic SST is related to NAO that leads to less precipitation or more frequent droughts in the semi-arid subarea in the upper reaches of the Yellow River. Liu et al. (2015) found that NAO greatly controls the variability of summer precipitation between the northeastern and the southeastern

---

**Table 2** | Heterogeneous correlation between precipitation and climate indices

<table>
<thead>
<tr>
<th>Zone 1</th>
<th>Zone 2</th>
<th>Zone 3</th>
<th>NAO</th>
<th>EA</th>
<th>WP</th>
<th>PNA</th>
<th>EA/WR</th>
<th>SCA</th>
<th>POL</th>
<th>PDO</th>
<th>NINO3.4</th>
<th>IOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode 1</td>
<td>-0.25</td>
<td>-0.25</td>
<td>-0.23</td>
<td>-0.16</td>
<td>0.01</td>
<td>-0.17</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.02</td>
<td>0.10</td>
<td>-0.04</td>
<td>-0.11</td>
</tr>
<tr>
<td>Mode 2</td>
<td>0.04</td>
<td>-0.01</td>
<td>-0.05</td>
<td>0.04</td>
<td>0.05</td>
<td>-0.04</td>
<td>0.00</td>
<td>0.05</td>
<td>0.10</td>
<td>0.05</td>
<td>0.01</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Values in bold are statistically significant at the 0.05 level.
Tibetan Plateau by modifying the atmospheric circulation over and around the Tibetan Plateau. During the positive phase of the NAO, warm moist air from the oceans around Asia is transported by the southeastern flank of the anticyclone anomaly over East Asian to the northeastern Tibetan Plateau, and this northward-moving warm moist air encounters cold air masses transported by the northwestern flank of the cyclonic anomaly over the northeastern Tibetan Plateau (Liu et al. 2013). This confluence of the cold and warm air masses subsequently strengthens cumulus convective activities and ultimately results in excessive precipitation over the northeastern Tibetan Plateau. Research showed that the strong positive and negative WP patterns are related to the east–west and north–south movements of the East Asian jet stream, indicating that the change from cold to warm season results from the northward movement of the East Asian jet stream and thus affects aspects of the East Asian climate such as precipitation and temperature (Barnston & Livezey 1987; Choi & Moon 2012). Yan (2002) found that POL was positively associated with winter precipitation in China, indicating the significance of the winter monsoon in producing rainfall pattern. Lin (2014) showed that the POL has negative correlation with precipitation in North China.

**CONCLUSION**

In this study, the spatial variability of precipitation was investigated based on the results from cluster analysis using monthly precipitation data from 1961 to 2010. Subdivision of the source region of the Yellow River into three homogeneous zones was made to investigate spatial variability of trends. PCA and SVD were used to find relationships between the precipitation in the

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Pearson correlation coefficients between observed and predicted summer precipitations for different areas for training and validation periods from ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period</td>
<td>Zone 1</td>
</tr>
<tr>
<td>Training</td>
<td>0.90</td>
</tr>
<tr>
<td>Validation</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Figure 6 | Observed and predicted summer precipitation for the validation period for the different zones and the whole area.
source region of the Yellow River and global teleconnection patterns using climate indices. The summer precipitation was predicted based on the revealed relationships using an ANN. The precipitation trend varies at different stations due to the temporal and spatial variation. The PCA analysis revealed relationships between some of the climate indices and precipitation. The first two modes of PCA were analysed since they can readily be associated with teleconnection patterns. The results showed that precipitation is positively related to the NAO, WP Pattern and El Niño Southern Oscillation and negatively related to the Polar Eurasian teleconnection. SVD was applied to the cross-covariance matrix between precipitation and climate indices. The results of SVD confirmed the relationship from the PCA. An ANN model was used to predict the summer precipitation in the source region of the Yellow River. The Pearson correlation coefficients between the predicted summer precipitation and observed summer precipitation are generally larger than 0.6. Thus, it is shown that significantly correlated climate indices can be used to predict the summer precipitation of the source region of the Yellow River.

Improving the knowledge regarding the relationship between precipitation in the source region of the Yellow River and global teleconnection patterns has important implications for water management. We also conclude that there is sufficient evidence to support the suitability of significantly correlated climate indices as predictors of summer precipitation in the source region of the Yellow River. The results are useful for integrated water resources management in the Yellow River basin.

ACKNOWLEDGEMENTS

This work was supported by the Key Program of Natural Science Foundation of China under Grant No. 40830639 and the MECW project from Swedish Science Research Council under Grant No. 2009-1056.

REFERENCES


Tang, Q. H., Oki, T., Kanae, S. & Hu, H. P. 2008 Hydrological cycles change in the Yellow River basin during the last half of the twentieth century. J. Climate 21 (8), 1790–1806.


Zhang, J., Li, D. L., Li, L. & Deng, W. T. 2015a Decadal variability of droughts and floods in the Yellow River basin during the last five centuries and relations with the North Atlantic SST. Int. J. Climatol. 35 (15), 3217–3228.


