



# Performance and uncertainty analysis of a short-term climate reconstruction based on multi-source data in the Tianshan Mountains region, China

LI Xuemei<sup>1,2,3\*</sup>, Slobodan P SIMONOVIC<sup>4</sup>, LI Lanhai<sup>5</sup>, ZHANG Xueting<sup>6,7</sup>, QIN Qirui<sup>1,3</sup>

<sup>1</sup> Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China;

<sup>2</sup> National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730070, China;

<sup>3</sup> Gansu Provincial Engineering Laboratory for National Geographic State Monitoring, Lanzhou 730070, China;

<sup>4</sup> Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, Western University, London N6A 3K7, Canada;

<sup>5</sup> State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China;

<sup>6</sup> Qilian Alpine Ecology & Hydrology Research Station, Key Laboratory of Ecohydrology of Inland River Basin, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China;

<sup>7</sup> University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract:** Short-term climate reconstruction, i.e., the reproduction of short-term (several decades) historical climatic time series based on the relationship between observed data and available longer-term reference data in a certain area, can extend the length of climatic time series and offset the shortage of observations. This can be used to assess regional climate change over a much longer time scale. Based on monthly grid climate data from a Coupled Model Inter-comparison Project phase 5 (CMIP5) dataset for the period of 1850–2000, the Climatic Research Unit (CRU) dataset for the period of 1901–2000 and the observed data from 53 meteorological stations located in the Tianshan Mountains region (TMR) of China during the period of 1961–2011, we calibrated and validated monthly average temperature (MAT) and monthly accumulated precipitation (MAP) in the TMR using the delta, physical scaling (SP) and artificial neural network (ANN) methods. Performance and uncertainty during the calibration (1971–1999) and verification (1961–1970) periods were assessed and compared using traditional performance indices and a revised set pair analysis (RSPA) method. The calibration and verification processes were subjected to various sources of uncertainty due to the influence of different reconstructed variables, different data sources, and/or different methods used. According to traditional performance indices, both the CRU and CMIP5 datasets resulted in satisfactory calibrated and verified MAT time series at 53 meteorological stations and MAP time series at 20 meteorological stations using the delta and SP methods for the period of 1961–1999. However, the results differed from those obtained by the RSPA method. This showed that the CRU dataset produced a low degree of uncertainty (positive connection degree) during the calibration and verification of MAT using the delta and SP methods compared to the CMIP5 dataset. Overall, the calibrated and verified MAP had a high degree of uncertainty (negative connection degree) regardless of the dataset or reconstruction method used. Therefore, the reconstructed time series of MAT for the period of 1850 (or 1901)–1960 based on the CRU and CMIP5 datasets using the delta and SP methods could be used for further study. The results of this study will be useful for short-term (several decades) regional climate reconstruction and longer-term (100 a or more) assessments of regional climate change.

**Keywords:** climate reconstruction; climate change; delta method; physical scaling method; artificial neural network

\*Corresponding author: LI Xuemei (E-mail: [lixuemei@mail.lzjtu.cn](mailto:lixuemei@mail.lzjtu.cn))

Received 2019-05-16; revised 2019-11-11; accepted 2019-12-19

© Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Science Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020

(ANN); CRU dataset; CMIP5 dataset

**Citation:** LI Xuemei, Slobodan P SIMONOVIC, LI Lanhai, ZHANG Xueting, QIN Qirui. 2020. Performance and uncertainty analysis of a short-term climate reconstruction based on multi-source data in the Tianshan Mountains region, China. *Journal of Arid Land*, 12(3): 374–396. <https://doi.org/10.1007/s40333-020-0065-y>

## 1 Introduction

Understanding past climates is critical for assessing the context of the natural climate and its consistency (Wu et al., 2010). The long-term continual observation of climate is a precondition for studying climate change (Misra et al., 2012; Ghosh, 2018). Both globally and in the Northern Hemisphere, the 20<sup>th</sup> century is considered the warmest century during the last 1000 a, but uncertainty still remains regarding the regional assessment of warming during the whole 20<sup>th</sup> century in western China, including northwestern China and the Tibetan Plateau, where recorded climatic data were rarely available prior to the 1950s (Fang et al., 2012; Li et al., 2013; Shen et al., 2019). The climate in northwestern China changed abruptly in the mid-1980s, with a shift from warm-dry to warm-wet conditions (Shi et al., 2007). However, it is not known whether climate change in northwestern China occurs at decadal or centennial time scales (Shi et al., 2007; Li et al., 2013). Furthermore, it is difficult to quantify and assess climate change over long historical periods and its impact on water resources, ecosystems, agriculture and other activities due to the shortage of instrumental records (Misra et al., 2012; Li et al., 2013; Yao et al., 2013; Shen et al., 2019).

In recent years, many studies have attempted to reconstruct local paleoclimate variability by applying climate proxy data, including fossil pollens, ice cores, lake sediment cores, tree rings and historical documents (Bradley and Jones, 1992; Shen et al., 2001; Esper et al., 2002; Yang et al., 2002; Fang et al., 2012; Wang et al., 2017; Klippel et al., 2018). This kind of reconstruction could provide long-term historical climatic time series. However, the process of paleoclimate reconstruction requires much time and effort. In addition, the uncertainty of such reconstructions is very high and the temporal resolution is low. Available comparatively long-term grid data, such as the Coupled Model Inter-comparison Project phase 5 (CMIP5) and Climate Research Unit (CRU) datasets, can provide historical climate data for historic periods extending over 100 a ago (Taylor et al., 2012; University of East Anglia Climatic Research Unit et al., 2017). These data offer an opportunity to back cast the historical local climate data in regions with a lack of long-term historical data (Li et al., 2013; Shen et al., 2019).

Single-value indices, such as the Nash-Sutcliffe coefficient (NSC) and percentage bias (PBIAS), are quantitatively used to assess the performance of hydrological models (e.g., the SDHydro model) by assessing the quantitative relationship between the observed and estimated runoff (Moriasi et al., 2007; Chen et al., 2012; Zhang et al., 2016a, b). They are also used to evaluate the capacity of wind erosion models, such as the Agricultural Policy Environmental eXtender (APEX) model (Wang and Li, 2012; Pi et al., 2017) and some environmental models (e.g., Hauduc et al., 2015). Chen et al. (2012) extended these single-value indices to the assessment of statistical downscaling methods. Li et al. (2013) analyzed the performance of reconstructions using these indices to test the quantitative relationship between the observed and reconstructed (calibrated and verified) hydro-meteorological time series. However, these indices cannot completely assess the qualitative and quantitative uncertainty associated with climate reconstruction (Li et al., 2013). There is a need to determine how best to qualitatively and quantitatively assess the reconstruction performance and uncertainty associated with different data sources and reconstruction methods.

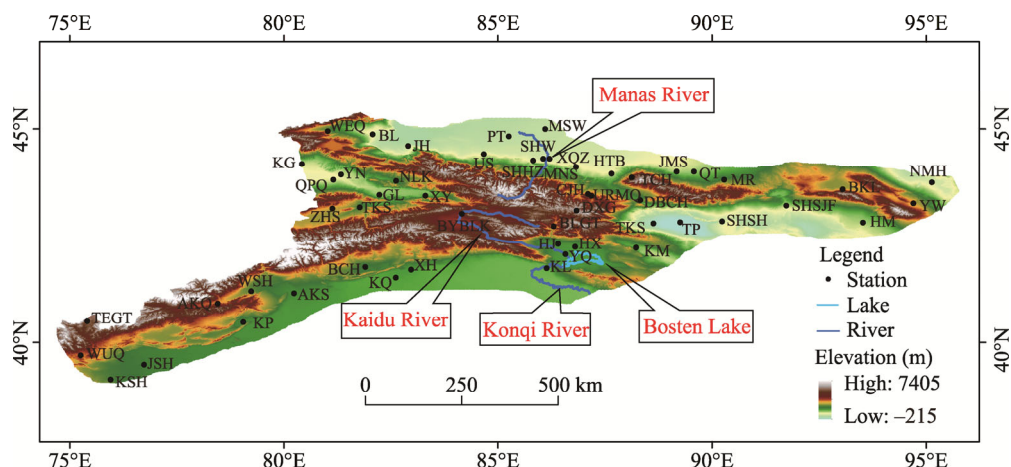
Based on the principles of set pair analysis (SPA) (Zhao, 1989; Zhao and Xuan, 1996), we developed a new method, namely, the revised SPA (RSPA), in this study. Then, we used this method to assess the performance and uncertainty of a short-term climate reconstruction in the Tianshan Mountains region (TMR) of China on the basis of two data sources and three reconstruction methods. This study will be useful and significant for short-term (several decades) regional climate reconstruction and longer-term (100 a or more) assessments of regional climate change.

## 2 Study area and data

### 2.1 Study area

Located in Central Asia and northwestern China, the Tianshan Mountains stretch approximately 2500 km in total, with about 1700 km in China. The area within China, about  $5.7 \times 10^5 \text{ km}^2$ , accounts for 34.5% of the total area of Xinjiang Uygur Autonomous Region (Hu, 2004) and is referred to as the TMR in this study (Fig. 1). The TMR starts from the Tarim Basin in the south and stretches to the Jungar Basin in the north. It forms the climate watershed between the northern and southern parts of Xinjiang (Wei et al., 2008). The average height of the mountain ridge is 4000 m a.s.l., with the highest peak being Tomor (7435 m a.s.l.). The TMR is a typical alpine mountainous area that is characterized by a continental climate, with large seasonal differences. Precipitation, which is mainly affected by the westerly circulation and topography, is generally regarded as the main source of water resources in Xinjiang (Liu et al., 2015; Zhang et al., 2019).

At the spatial scale, meteorological stations with long-term time series are unevenly distributed in the TMR (Fig. 1). At the temporal scale, meteorological records began in the late 1950s and no data are available prior to this time. The existing recorded data indicate that the annual average temperature is  $7.7^\circ\text{C}$  and is higher than  $0.0^\circ\text{C}$  for most meteorological stations, except those with elevations above 2000 m a.s.l. (e.g., Bayanbulak, Daxigou and Tuergate stations). The average annual precipitation is 189.58 mm and is larger at stations with higher elevations, such as Bayanbulak, Daxigou, Tianchi, Tuergate and Zhaosu stations (Li et al., 2016).



**Fig. 1** Distribution of the meteorological stations in the Tianshan Mountains region (TMR) of China. The full name and descriptive information of all meteorological stations can be seen in Table S1 in the Appendix.

### 2.2 Data

Observed monthly data and two sources of grid monthly data were used to reconstruct historical climatic time series in the TMR: (1) observed data from meteorological stations in the TMR; (2) grid data from the CRU dataset; and (3) a CMIP5 dataset of past and current climatic conditions. The monthly average temperature (MAT) and monthly accumulated precipitation (MAP) at 62 meteorological stations in the TMR were the target climate records for calibration and verification. The CRU TS v4.03 dataset covers the period of 1901–2018. Time series datasets for the 20<sup>th</sup> century indicate the month-by-month variation in climate at  $0.5^\circ$  spatial resolution. The variables include MAT, MAP, diurnal temperature range, monthly average daily maximum temperature and cloud cover (Harris et al., 2014; University of East Anglia Climatic Research Unit et al., 2020).

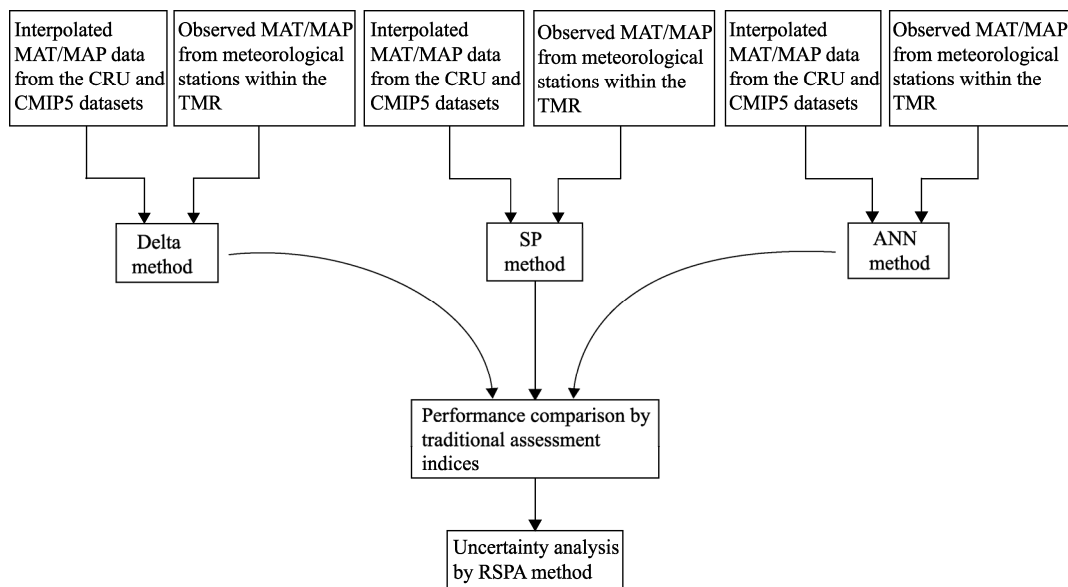
The CMIP5 multi-model dataset was compiled by the Working Group on Coupled Modelling (WGCM) and Program for Climate Model Diagnosis and Intercomparison (PCMDI) (Taylor et al., 2012). This multi-model dataset included the outputs from 42 global climate models (GCMs) (for details see Table S2 in the Appendix). Among these GCMs, 42 provided the MAP and 41 provided the MAT. Therefore, the MAP from 42 GCMs and MAT from 41 GCMs were used in this study.

These monthly data were subjected to a quality control procedure, which involved a visual examination of individual stations or grid records to identify outliers. Abnormal values were identified and either removed or corrected. Only a small fraction of the data needed correction. Missing monthly data, accounting for less than 2% of the total data, were estimated for a given month by extrapolating the average value of the data from the one or two preceding and one or two following records. Any time series with over one year of missing data was omitted from the research. Finally, the observed data from 53 meteorological stations in the TMR were used, while data from nine stations were omitted. The details of the 53 stations are provided in Table S1 in the Appendix. The near gridded values of MAT and MAP data from the CRU and CMIP5 datasets were interpolated into the 53 selected meteorological stations in the TMR by the bi-linear interpolation method (Huang et al., 2018). As with the observed data, a similar quality control procedure was applied to the data from the CRU and CMIP5 datasets. Finally, the interpolated MAT data from 41 GCMs in the CMIP5 and CRU datasets and the interpolated MAP data from 42 GCMs in the CMIP5 and CRU datasets were used in this study.

Considering the length of the observed data in the CMIP5 and CRU datasets, the period of 1961–1999 could be used as the calibration and verification period. Hence, each time series was divided into two parts for both MAT and MAP in accordance with the two stages in the model-building process. The first stage (1971–1999: 348 months) was used for calibration and the second stage (1961–1970: 120 months) was used for verification using the delta, physical scaling (SP) and artificial neural network (ANN) methods, with the bi-linear interpolated CMIP5 and CRU data. During the process of calibration and verification, only the interpolated CMIP5 and CRU data obtained by the bi-linear interpolation method were used to represent local station data and this approach was referred to as "No method". Therefore, during each stage, 168 calibrated and verified MAT time series and 172 calibrated and verified MAP time series were obtained for each station.

### 3 Methods

Many methods and approaches were considered for use in this study. Figure 2 shows a flow chart of the methods and approaches that were considered.



**Fig. 2** Flow chart of methods and approaches used in this study. MAT, monthly average temperature; MAP, monthly accumulated precipitation; CRU, Climate Research Unit; CMIP5, Coupled Model Inter-comparison Project phase 5; TMR, Tianshan Mountains region; SP, physical scaling; ANN, artificial neural network; RSPA, revised set pair analysis.

### 3.1 Delta method

The delta method is a simple statistical downscaling approach that has been widely applied in previous assessments of climate change and water planning studies (Hay et al., 2000; Zhao and Xu, 2007; Hao et al., 2009; Li et al., 2013). The method has been described in previous literature (Li et al., 2013). In this study, the delta method was extended to use in climate reconstruction as follows:

$$T(x, t) = \overline{Tb}(x, t) + (Tr(y, t) - \overline{Tbr}(y, t)), \quad (1)$$

$$P(x, t) = Pr(y, t) \times \left( \frac{\overline{Pb}(x, t)}{\overline{Pbr}(y, t)} \right), \quad (2)$$

where  $t$  denotes the number of the month in the year,  $t=1, 2, 3, \dots, 12$ ;  $T(x, t)$  (°C) and  $P(x, t)$  (mm) are the reconstructed MAT and MAP for the target location  $x$  in the  $t^{\text{th}}$  month during the reconstruction period, respectively;  $Tr(y, t)$  and  $Pr(y, t)$  represent gridded values of MAT and MAP for the reference location  $y$  in the  $t^{\text{th}}$  month during the reconstruction period, respectively;  $\overline{Tb}(x, t)$  and  $\overline{Tbr}(y, t)$  (°C) are the mean values of MAT during the baseline time period for the target location  $x$  and the reference location  $y$  in the  $t^{\text{th}}$  month, respectively; and  $\overline{Pb}(x, t)$  and  $\overline{Pbr}(y, t)$  (mm) are the mean values of MAP during the baseline time period for the target location  $x$  and the reference location  $y$  in the  $t^{\text{th}}$  month, respectively.

### 3.2 Physical scaling (SP) method

To overcome the limitation of location-centric characteristics and model geophysical changes within a statistical downscaling framework, Gaur and Simonovic (2017a, b) proposed the SP model. The model was based on the hypothesis that the large-scale climate and elevation properties of a location can define the local-scale climate characteristics. The SP model builds a regression relationship between local-scale observed data and bi-linearly interpolated climate model and elevation data. A generalized additive model (GAM) (Schimek and Turlach, 2000) was chosen as the regression function to link predictant (local-scale observed data) and predictor variables (bi-linearly interpolated climate model and elevation data). In this study, the SP method was extended to climate reconstruction and was mathematically expressed as:

$$g(\text{MAT}_{\text{obs}}) = B_0 + f_1(\text{MAT}_{\text{mod}}) + f_2(E_p), \quad (3)$$

$$g(\text{MAP}_{\text{obs}}) = B_0 + f_1(\text{MAP}_{\text{mod}}) + f_2(E_p), \quad (4)$$

where variables  $g$  and  $f$  represent the link and smoothing functions, respectively;  $\text{MAT}_{\text{obs}}$  (°C) and  $\text{MAP}_{\text{obs}}$  (mm) represents the observed MAT and MAP data, respectively; subscript *obs* and *mod* denote if the climatic data are observed or model based, respectively;  $B_0$  denotes the regression parameter;  $E_p$  denotes the elevation (m) of the referred pixel, and the subscript  $p$  means that the data used are at a pixel scale; and  $\text{MAT}_{\text{mod}}$  and  $\text{MAP}_{\text{mod}}$  denote large-scale MAT and MAP data obtained by a bi-linear interpolation of GCM or CRU data at the referred pixel, respectively. In this study, smoothing functions were fitted using the penalized likelihood maximization algorithm, which was a variant of the maximum likelihood estimation algorithm that was obtained by incorporating a penalty function, with a tradeoff between model fit wiggleness and goodness of fit (Wood, 2000).

### 3.3 Artificial neural network (ANN) model

Artificial neural networks (ANNs) have been widely used in the estimation of climatic, hydrologic, economic and other variables in recent decades. In this study, a three-layered feed-forward neural network (FFNN) was used. Three layers containing an input layer, a hidden layer and an output layer were included within a general framework for representing nonlinear functional mapping between a set of input and output variables (Nourani et al., 2009). The weighted connections feed activations in an FFNN only in the forward direction from an input layer to an output layer. In addition, additional weighted connections were used to feed the previous activations back to the network in a feed-forward back propagation (BP) network (Srinivasulu and Jain, 2006). In a FANN

structure with a BP algorithm, it has been proven that the three-layer structure is satisfactory for forecasting and simulating water resources, hydrologic parameters and climate change (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000). The Levenberg-Marquardt (LM) fitting algorithm can be used to train a three-layered BP FFNN (TLBP-FFNN) as follows: (1) suppose a set of  $n$  empirical pairs of independent and dependent variables are defined as  $(x_i, y_i)$ ; and (2) in the model curve  $f(x_i, \beta)$ , the parameter  $\beta$  is optimized so that the sum of the squares of the deviations ( $S(\beta)$ ; Eq. 5) obtains a minimum value (Levenberg, 1944; Marquardt, 1963).

$$S(\beta) = \sum_{i=1}^n [y_i - f(x_i, \beta)]^2. \quad (5)$$

A heuristic procedure is used to determine the input dimension and the number of hidden nodes by testing different inputs, with an increasing number of hidden nodes (Cannas et al., 2006). The output value of the TLBP-FFNN could be specifically expressed as follows:

$$y_k = f_o \left[ \sum_{j=1}^n w_{kj} \times f_h \left( \sum_{i=1}^m w_{ji} x_i + w_{jo} \right) + w_{ko} \right], \quad (6)$$

where  $y_k$  represents the  $k^{\text{th}}$  computed output variable;  $m$  and  $n$  are the number of neurons in the input and hidden layers, respectively ( $i=1, 2, 3, \dots, m$  and  $j=1, 2, 3, \dots, n$ );  $f_o$  and  $f_h$  are the activation functions for the output and hidden neurons, respectively;  $w_{kj}$  is the  $k^{\text{th}}$  neuron in the output layer, with the  $j^{\text{th}}$  neuron in the hidden layer;  $w_{ji}$  is a weighting linking the  $j^{\text{th}}$  neuron in the hidden layer with the  $i^{\text{th}}$  neuron in the input layer;  $x_i$  stands for the  $i^{\text{th}}$  input variable for the input layer; and  $w_{jo}$  and  $w_{ko}$  are the biases for the  $j^{\text{th}}$  hidden neuron and the  $k^{\text{th}}$  neuron in the output layer, respectively.

In this study, MAT and MAP were modelled by the TLBP-FFNN method, with the linear activation function in the output neuron and Tan-sigmoid activation function in the hidden neuron trained using the LM fitting algorithm. For the ANN, which is a data-driven statistical model, the input variables were selected from the available data, and the model was subsequently developed. The MAT and MAP during 1961–1999 were available from the 53 meteorological stations.

The LM back propagation scheme was used as the training algorithm in the Matlab neural network toolbox (Beale et al., 2015) for the purpose of this study. The network architecture and training iteration number (epoch) are important in the TLBP-FFNN modeling. In this study, 10,000 epochs resulted in a satisfactory training network, with  $10^{-4}$  as the target performance.

### 3.4 Performance and uncertainty analysis

#### 3.4.1 Traditional performance assessment indices

The following performance indices were used to assess the reconstruction performance:

$$\text{NSC} = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (7)$$

$$\text{PBIAS} = \frac{\sum_{i=1}^n (y_i - y'_i) \times 100\%}{\sum_{i=1}^n y_i}, \quad (8)$$

where NSC and PBIAS are the Nash-Sutcliffe coefficient (Nash and Sutcliffe, 1970) and percentage bias, respectively; and  $n$ ,  $y_i$ ,  $y'_i$  and  $\bar{y}$  are the number of observations, the observed value, the predicted or reconstructed value and the mean value of the observed data, respectively. Both NSC and PBIAS were used to assess the performance of models. When the value of NSC ( $-\infty$  to 1.00) is close to 1.00 and the absolute value of PBIAS is close to 0, the estimated or reconstructed data series is close to the observed data series (Gupta et al., 1999; Moriasi et al., 2007). A positive PBIAS indicates that the model is underestimating the predicted values, while a negative value means that the model is overestimating the predicted values (Gupta et al., 1999). Values of NSC greater than 0.50 and absolute values of PBIAS less than 25% indicate a satisfactory model performance (Table 1).

**Table 1** Climate reconstruction performance criteria (revised from Moriasi et al. (2007))

Performance rating	Grade	NSC	PBLAS
Very good	A	$NSC \geq 0.75$	$PBIAS \leq \pm 15\%$
Good	B	$0.65 \leq NSC < 0.75$	$PBIAS \leq \pm 20\%$
Satisfactory	C	$0.50 \leq NSC < 0.65$	$PBIAS \leq \pm 25\%$
Unsatisfactory	D	$NSC < 0.50$	Any value

Note: NSC, Nash-Sutcliffe coefficient; PBIAS, percentage bias.

### 3.4.2 Set pair analysis (SPA) theory

The SPA theory was proposed by Zhao (1989), which provides a way to assess uncertainties. This theory overlaps with many other uncertainty theories, especially Boolean reasoning methods, evidence theory, fuzzy set theory and rough set theory (Yang et al., 2008). Emphasizing the relativity and fuzziness in information processing, the SPA theory can identify the relative certainty and relative uncertainty, and describe them quantitatively and systematically using identity, discrepancy and contrary (Zhao and Xuan, 1996; Ye and Wu, 2000; Wang and Li, 2012; Yan et al., 2017). The SPA theory has been widely applied in many fields, including artificial intelligence, system engineering, multi-attribute assessments, urban ecosystem health assessments, energy consumption forecasting, water resource evaluation, weather forecasting and hazard assessment (Wang et al., 2004; Zhou et al., 2007; Wang et al., 2008; Su et al., 2009; Wang and Li, 2012; Wang et al., 2014; Yan et al., 2017).

The certainty and uncertainty in a given system can be evaluated by the SPA theory. Assuming  $A$  and  $B$  are the sets given, the set pair  $H=(A, B)$ , made up of  $A$  and  $B$ , is determined by its characteristics. The value of its characteristics is represented by  $N, S, F$  and  $P$  represent the values of the identity, discrepancy and contrary characteristics, respectively. The values of  $S, F, P$  and  $N$  meet the condition  $N=S+F+P$ . Therefore, the values of  $S/N, F/N$  and  $P/N$  are termed the identity, discrepancy and contrary degrees, respectively. In this study, we set  $a=S/N, b=F/N$  and  $c=P/N$ . Obviously,  $a+b+c=1$  is undisputed (Wang and Li, 2012; Yan et al., 2017). The connection degree of  $H$  is then defined as follows:

$$\mu(A, B) = \frac{S}{N} + \frac{F}{N}i + \frac{P}{N}j = a + bi + cj, \quad (9)$$

where  $\mu$  is the connection degree of  $H$ ;  $i$  is the uncertainty coefficient of the discrepancy degree, with an uncertainty value between  $-1$  and  $1$ , which corresponds to various circumstances or can be considered only as a measure of discrepancy; and  $j$  is the coefficient of the contrary degree and is specified as  $-1$  (Su et al., 2009; Wang and Li, 2012; Yan et al., 2017).

Equation 9 is considered to be the three-element connection degree (Wang and Li, 2012; Yan et al., 2017). The multi-element connection degree is defined as follows:

$$\mu(A, B) = a + b_1i_1 + b_2i_2 + \dots + b_{k-2}i_{k-2} + cj, \quad (10)$$

where  $a+b_1+b_2+\dots+b_{k-2}+c=1$ ;  $b_1, b_2, \dots, b_{k-2}$  are components of the discrepancy degree, explaining different grades of discrepancy; and  $i_1, i_2, \dots, i_{k-2}$  are the uncertainty component coefficients of the discrepancy degree (Wang and Li, 2012). The connection degree  $\mu(A, B)$  may overcome the drawbacks of using a single index, such as the correlation coefficient, subordinate degree or grey correlation degree (Wang and Li, 2012; Yan et al., 2017).

### 3.4.3 Revised set pair analysis (RSPA) for climate reconstruction

Based on the principle and operating procedure of the SPA, a revised SPA, namely the RSPA, was proposed to assess the uncertainty of short-term climate reconstruction. We assumed that  $A=(y_1, y_2, \dots, y_m)$  was the reconstructed climatic time series, and  $B=(x_1, x_2, \dots, x_m)$  was the observed climatic time series ( $m$  is the number of observed values) in this study. For the performance assessment of  $k$  ( $k \geq 2$ ) grades of climate reconstruction, there are  $k-1$  thresholds ( $s_1, s_2, \dots, s_{k-1}$ ) whose intervals ( $0-s_1, s_1-s_2, \dots, s_{k-2}-s_{k-1}, s_{k-1}-+\infty$ ) correspond to different grades. For example,  $k=3$  means three grades, namely good, normal and bad performance (Wang and Li, 2012; Yan et al., 2017).

The  $k$ -element connection degree  $\mu(A, B)$  is described as follows:

$$\mu(A, B) = \begin{cases} 1 + 0i_1 + 0i_2 + \dots + 0i_{k-2} + 0j, & \text{when } \Delta_l \in [0, s_1] \\ \frac{S_1}{m} + \frac{S_2}{m}i_1 + 0i_2 + \dots + 0i_{k-2} + 0j, & \text{when } \Delta_l \in [0, s_2] \\ \frac{S_1}{m} + \frac{S_2}{m}i_1 + \frac{S_3}{m}i_2 + \dots + 0i_{k-2} + 0j, & \text{when } \Delta_l \in [0, s_3] \\ \dots & \dots \\ \frac{S_1}{m} + \frac{S_2}{m}i_1 + \frac{S_3}{m}i_2 + \dots + \frac{S_{k-1}}{m}i_{k-2} + 0j, & \text{when } \Delta_l \in [0, s_{k-1}] \\ \frac{S_1}{m} + \frac{S_2}{m}i_1 + \frac{S_3}{m}i_2 + \dots + \frac{S_{k-1}}{m}i_{k-2} + \frac{S_k}{m}j, & \text{when } \Delta_l \in [0, +\infty) \end{cases}, \quad (11)$$

where  $\Delta_l = |y_l - x_l|$ ,  $l=1, 2, 3, \dots, m$ , and  $s_1 \leq s_2 \leq \dots \leq s_{k-1}$ .

$$S_1 = \sum_{l=1}^m \text{sgn}(\Delta_l - s_1); \quad \text{sgn}(\Delta_l - s_1) = \begin{cases} 1, & \Delta_l - s_1 \leq 0 \\ 0, & \text{else} \end{cases}, \quad (12)$$

$$S_d = \sum_{l=1}^m \text{sgn}(\Delta_l - s_d); \quad d=2, 3, 4, \dots, k-1; \quad \text{sgn}(\Delta_l - s_d) = \begin{cases} 1, & s_{d-1} < \Delta_l \leq s_d \\ 0, & \text{else} \end{cases}, \quad (13)$$

$$S_k = m - S_1 - S_2 - \dots - S_{k-1}. \quad (14)$$

In this study, the absolute error ( $\Delta_l$ ) between the calibrated (or verified) and observed climatic time series was chosen as the assessment factor.  $S_1/m$  is the possibility of  $\Delta_l$  belonging to  $G_1$  grade;  $S_2/m$  is the possibility of  $\Delta_l$  belonging to  $G_2$  grade;  $S_3/m$  is the possibility of  $\Delta_l$  belonging to  $G_3$  grade;  $S_{k-1}/m$  is the possibility of  $\Delta_l$  belonging to the  $(k-1)^{\text{th}}$  grade; and  $S_k/m$  is the possibility of  $\Delta_l$  belonging to the  $G_k$  grade. For  $k=5$ , it means that there are five grades, i.e., very good performance ( $G_1$ ), good performance ( $G_2$ ), normal performance ( $G_3$ ), bad performance ( $G_4$ ) and very bad performance ( $G_5$ ). The grade standards for MAT and MAP reconstruction are shown in Table 2. According to the SPA theory,  $S_1/m$  corresponds to the identity degree;  $S_5/m$  corresponds to the contrary degree; and  $S_2/m$ ,  $S_3/m$  and  $S_4/m$  are three detailed discrepancy degrees.

**Table 2** Assessment of grade standards for the calibration and verification of monthly average temperature (MAT) and monthly accumulated precipitation (MAP) time series in the Tianshan Mountains region (TMR) of China

	Grade				
	$0-s_1$	$s_1-s_2$	$s_2-s_3$	$s_3-s_4$	$s_4-s_5$
Range of MAT ( $^{\circ}\text{C}$ )	0.0–0.5	0.5–1.0	1.0–1.5	1.5–2.0	2.0– $+\infty$
Range of MAP (mm)	0.0–2.0	2.0–4.0	4.0–6.0	6.0–8.0	8.0– $+\infty$
Performance	Very good	Good	Normal	Bad	Very bad
Degree	G1	G2	G3	G4	G5

In addition, we set the ratio of  $R:R_i^j$  representing different calibrated (or verified) climatic time series for one target station based on interpolated grid data from the CMIP5 and CRU datasets using four different reconstruction methods: delta, ANN, SP and No methods. Here,  $i$  means different data sources (GCMs from the CMIP5 and CRU) and  $j$  means different reconstruction methods (delta, ANN, SP and No methods). For example,  $R_{\text{CanCM4}}^{\text{ANN}}$  represents a calibrated (or verified) climatic time series for one target station based on interpolated grid data from the CanCM4 GCM model using the ANN method. In this study, we set  $T$  representing the observed MAT time series for one target station in the TMR, and  $P$  representing the observed MAP time series in the TMR. The connection degree ( $\mu(R, T)$  and  $\mu(R, P)$ ) between calibrated (verified) and observed data can be calculated using Equations 11–14.

## 4 Results

In this section, we assessed the performance of calibrated and verified MAT and MAP time series, taking Zhaosu station as an example.

### 4.1 Performance assessment of calibrated and verified MAT time series

Figure 3 shows the calibration and verification performance of MAT time series by the delta, SP, ANN and No methods based on the CMIP5 and CRU datasets at Zhaosu station. The results indicated that 128 satisfactory calibrated and verified MAT time series were produced from a total of 168 time series, with a satisfaction degree of 76%. All 128 satisfactory calibrated and verified MAT time series captured the intra-annual variation pattern (Fig. 3). Numerically, the calibration and verification based on two sources of data performed very well (most performance grades were A (very good)) for calibrated and verified MAT time series). Use of the three methods (delta, SP and ANN) greatly improved the calibration and verification performance. There was no large difference between calibrated (verified) and observed MAT values, but there was a small difference when the different sources of data were used (Fig. 3). Calibrated and verified results based on grid data from the CRU were better than those generated using the CMIP5. This may be because the CRU data were superior to the CMIP5 data. The CRU data were obtained based on actual pointwise observations, while the CMIP5 data were modelled based on greenhouse gases, anthropogenic aerosol solar variability and volcanic eruptions.

The same data sources also generated similar MAT time series using the different methods at Zhaosu station during the calibration and verification periods. There was almost no difference between calibrated (verified) and observed MAT time series based on the CRU data when using the four different methods (Fig. 3). Based on the CMIP5 dataset, we obtained a similar calibration and verification performance by the four methods according to the average NSC value. However, according to the PBIAS values, MAT values were underestimated or overestimated by the SP and ANN methods for both data sources during the calibration and verification periods.

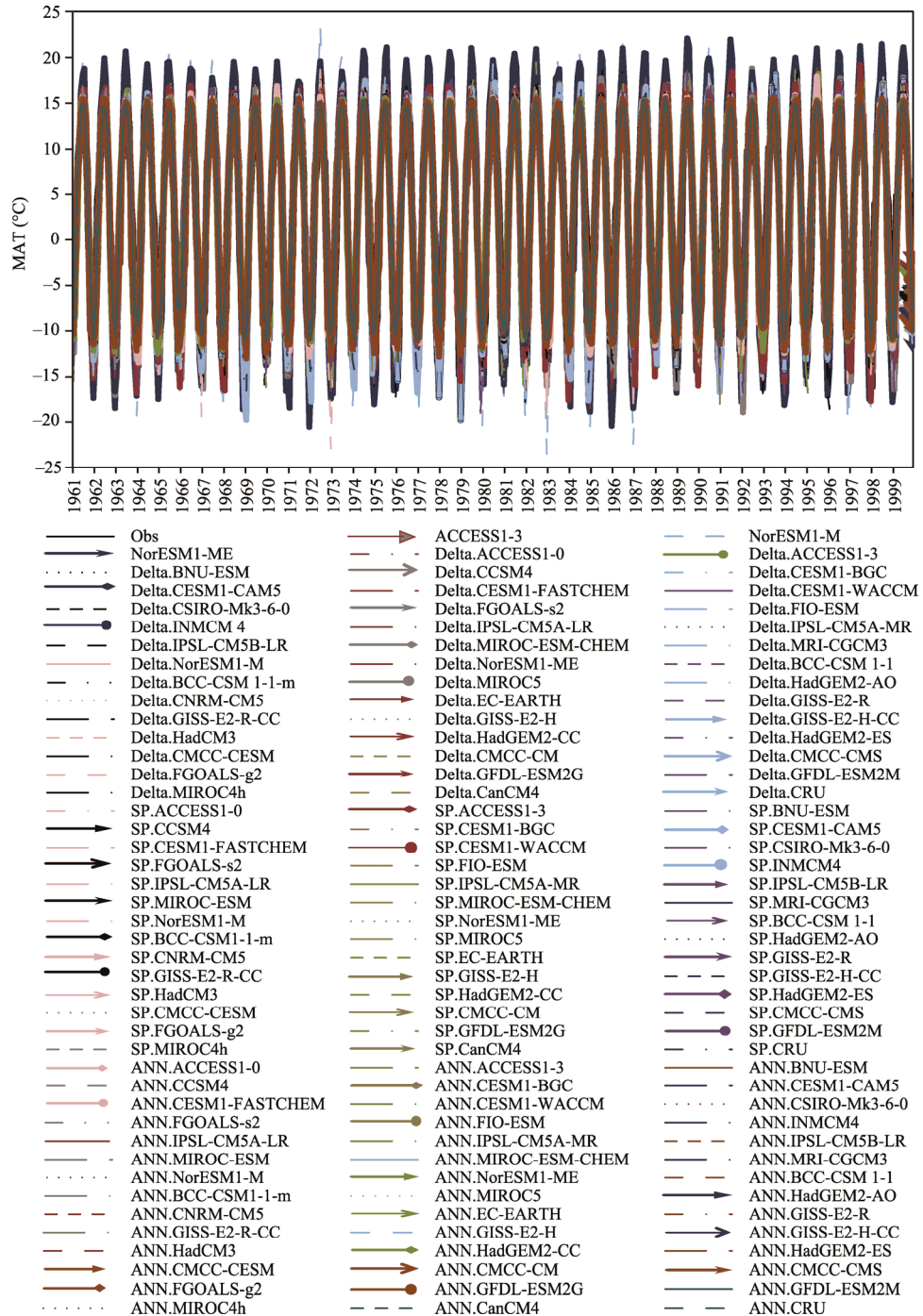
The numbers of satisfactory calibrated and verified MAT time series based on the CMIP5 and CRU datasets using the delta, SP and ANN methods in the TMR are presented in Figure 4 and Table S3 (see the Appendix). It can be seen that for all 53 meteorological stations selected in this study, satisfactory calibrated and verified MAT time series were obtained. The performance ratings were high, with most of them graded A (very good). Both the CMIP5 and CRU datasets performed very well during the MAT calibration and verification periods. The delta and SP methods produced satisfactory results based on both the CMIP5 and CRU datasets for all 53 meteorological stations. The ANN method produced a satisfactory calibrated and verified MAT time series for only 3–4 stations based on several GCMs from the CMIP5 and CRU datasets. Some bi-linear interpolated CMIP5 models and CRU data could generate satisfactory calibrated and verified MAT time series for most stations without the use of a reconstruction method.

### 4.2 Performance assessment of calibrated and verified MAP time series

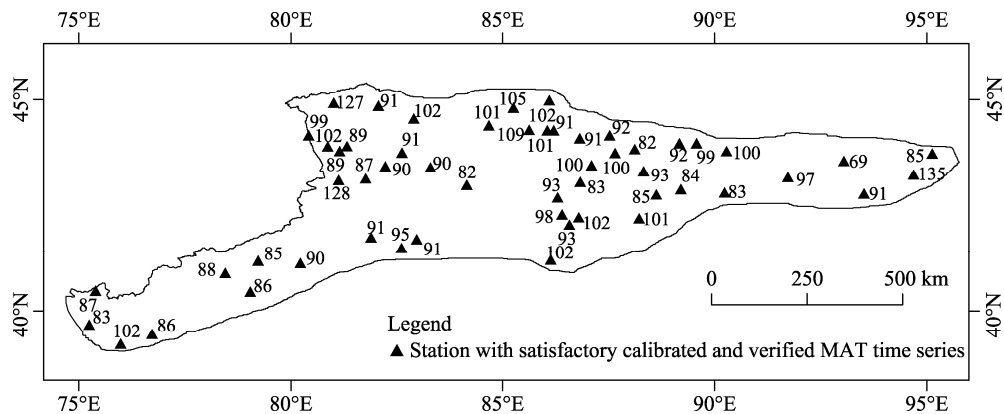
The calibration and verification performance of MAP time series produced by the delta, SP, ANN and No methods using the CMIP5 and CRU datasets at Zhaosu station is shown in Figure 5. In comparison with the calibration and verification results of MAT time series, the calibration and verification performance of MAP time series was inferior. Specifically, only 26 satisfactory calibrated and verified MAP time series among the total of 172 time series, with a satisfaction degree of 15%. Only 14 GCMs from the CMIP5 dataset produced a good or satisfactory calibrated and verified MAP time series (most performance grades were C (satisfactory)) at Zhaosu station, while the use of the CRU dataset did not result in any satisfactory series. The different GCMs from the CMIP5 dataset generated different performances, each with small differences (Fig. 5).

The same data source produced different results when the different methods were used at Zhaosu station during the calibration and verification periods. Among the 26 satisfactory calibrated and verified MAP time series, nine were generated by the delta method, nine by the SP method, six by the ANN method and two by the No method. Although all 26 satisfactory calibrated and verified

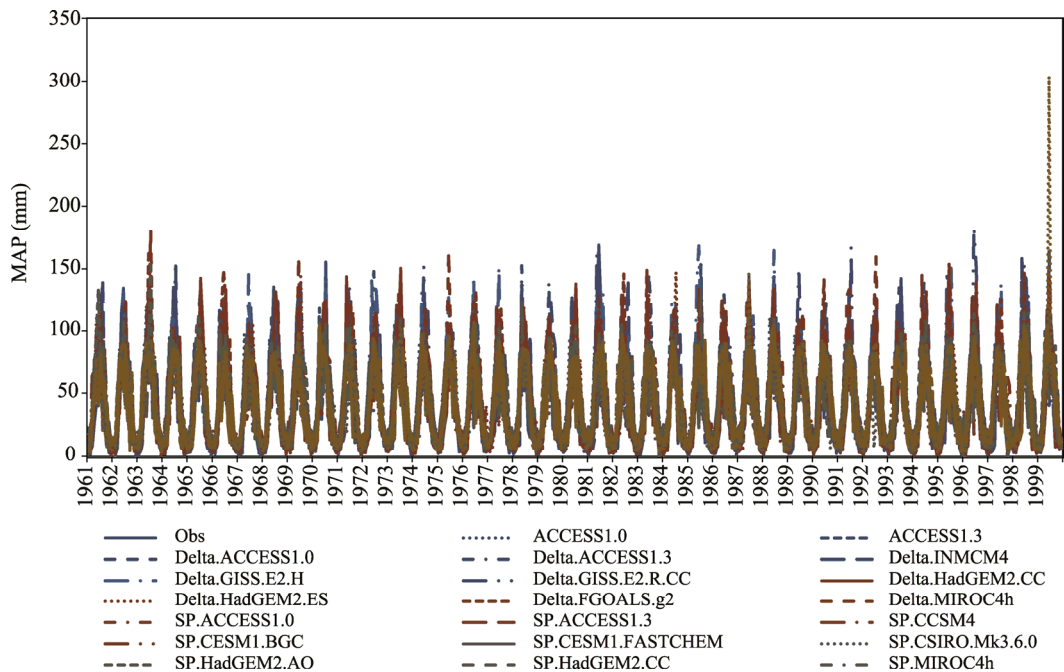
MAP time series could capture the intra-annual pattern, there were numerical differences between them and the observed time series (Fig. 5). The average NSC value for calibrated and verified MAP time series showed a decreasing trend with the order of delta, SP, ANN and No methods. Therefore, for the calibrated and verified MAP time series at Zhaosu station, the delta method performed better than the SP method, and the SP method performed better than the ANN method.



**Fig. 3** The 128 satisfactory calibrated, verified and observed monthly average temperature (MAT) time series at Zhaosu station. Obs, observed data; SP, physical scaling; ANN, artificial neural network. Delta.ACCESS1-0, the reconstructed data based on ACCESS1-0 data using the delta method; the others all follow this nomenclature scheme.



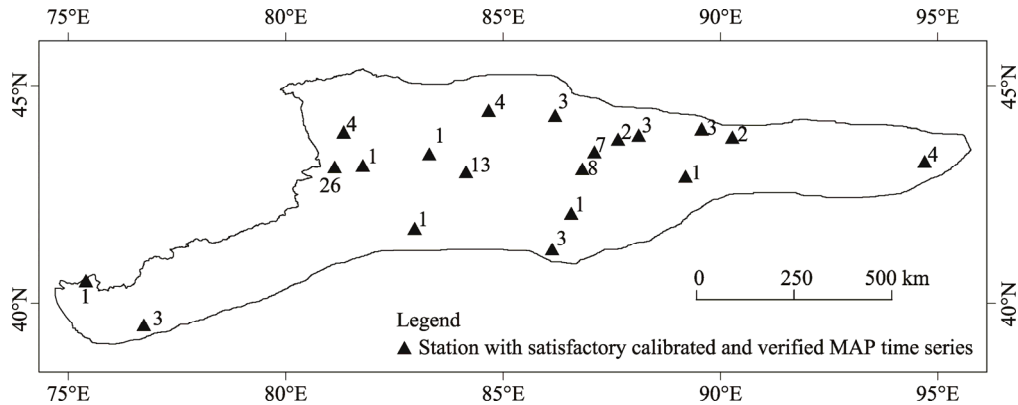
**Fig. 4** Distribution of numbers of satisfactory calibrated and verified MAT time series in the TMR. The value at each station represents the number of satisfactory calibrated and verified MAT time series based on the CMIP5 and CRU datasets using the delta, SP and ANN methods.



**Fig. 5** The 26 satisfactory calibrated, verified and observed monthly accumulated precipitation (MAP) time series at Zhaosu station. Obs, the observed data; SP, physical scaling. Delta.ACCESS1.0, the reconstructed data based on ACCESS1.0 data using the delta method; the others all follow this nomenclature scheme.

The number of satisfactory calibrated and verified MAP time series based on the CMIP5 and CRU datasets using the delta, SP, ANN and No methods in the TMR are shown in Figure 6 and Table S4 (see the Appendix). Satisfactory calibrated and verified MAP time series were obtained for only 20 stations in the TMR. The performance ratings were low and most of them were graded C (satisfactory). Although the number of satisfactory calibrated and verified MAP time series was very small (1, 2 or 3) at most of the 20 stations, it was slightly larger at Zhaosu station (26), followed by Bayanbulak (13), Daxigou (8) and Xiaoquzi (7) stations. With the exception of Zhaosu, Bayanbulak, Daxigou and Xiaoquzi stations, satisfactory MAP time series were only obtained with the CRU dataset. The delta and SP methods produced more satisfactory MAP time series than the ANN method. The ANN method produced satisfactory calibrated and verified MAP time series for 12 stations based on data from several GCMs in the CMIP5 and CRU datasets. The CRU data produced satisfactory calibrated and verified MAP time series for three stations, while

two GCMs from the CMIP5 dataset produced satisfactory calibrated and verified MAP time series for one station by the No method.



**Fig. 6** Distribution of numbers of satisfactory calibrated and verified MAP time series in the TMR. The value at each station represents the number of satisfactory calibrated and verified MAT time series based on the CMIP5 and CRU datasets using the delta, SP and ANN methods.

The above results indicate that the delta and SP methods performed very well during the calibration and verification periods of MAT time series for both the CMIP5 and CRU datasets at each station in the TMR. The ANN method provided a satisfactory MAT time series for several stations. For 40% of the selected stations in the TMR, good or satisfactory MAP time series were obtained by the delta and SP methods using several GCMs from the CMIP5 dataset. The ANN method generated satisfactory MAP time series at fewer stations than the delta and SP methods based on several GCMs from the CMIP5 and CRU datasets. One possible explanation for the difference was that the temperature was impacted mainly by the radiation balance and was easier to simulate. However, precipitation may be influenced by many factors, such as local climate, topography and vapor sources and was much more difficult to model.

## 5 Discussion

In the previous section, the best results were obtained for calibrated and verified MAT and MAP time series at Zhaosu station. Selecting Zhaosu station as an example again, this section presents the uncertainty determined during the calibration and verification of MAT and MAP time series using the RSPA method. Connection degree expressions between calibrated (verified) and observed climatic (MAT and MAP) time series by the delta, SP, ANN and No method based on the CMIP5 and CRU datasets at Zhaosu station in the TMR are provided in Tables S5 and S6 (see the Appendix). It should be noted that those two tables only show the first 20 entries by descending value of identity degree ( $a$ ).

### 5.1 Uncertainty analysis of calibrated and verified MAT

For the set pair  $(R_{CRU}^{\text{delta}}, T)$ , the identity degree ( $a$ ) was 0.58 and 0.59 for the calibrated and verified MAT; the uncertainty component coefficients of the discrepancy degree ( $b_1$ ,  $b_2$  and  $b_3$ ) were 0.24, 0.11 and 0.04 for the calibrated MAT, and 0.26, 0.06 and 0.06 for the verified MAT, respectively; and the contrary degree ( $c$ ) was 0.03 and 0.03 for the calibrated and verified MAT, respectively (Table S5). Therefore,  $\mu(R_{CRU}^{\text{delta}}, T) = 0.58 + 0.24i_1 + 0.11i_2 + 0.04i_3 + 0.03j$  during the calibration period, which means that the connection degree between calibrated and observed MAT time series for Zhaosu station based on the CRU dataset using the delta method was expressed as:  $\mu(R, T) = 0.58 + 0.24i_1 + 0.11i_2 + 0.04i_3 + 0.03j$ , i.e., 58%, 24%, 11%, 4% and 3% of 348 values of  $\Delta_i$  during the calibration period fell into the ranges of  $0.0^\circ\text{C}$ – $0.5^\circ\text{C}$ ,  $0.5^\circ\text{C}$ – $1.0^\circ\text{C}$ ,  $1.0^\circ\text{C}$ – $1.5^\circ\text{C}$ ,  $1.5^\circ\text{C}$ – $2.0^\circ\text{C}$  and  $2.0^\circ\text{C}$ – $+\infty$ , respectively. The rest of the results can be explained in the same manner.

According to the criteria in Table 2, the calibrated and verified MAT time series based on the

CRU dataset obtained by the delta method produced a very good performance. There was a high possibility (51%–59%) of a very good performance during both the calibration and verification periods. For the time series based on the GCM data from the CMIP5 dataset, there was a lower possibility (20%–26%) of a very good performance compared with the possibility (30%–39%) of a very bad performance. The uncertainty component coefficients of the discrepancy degree based on both datasets decreased gradually from  $b_1$  to  $b_3$ . Almost all values of  $b_1$  were larger than values of  $b_2$  and  $b_3$ . This indicates that the possibility shifted from the discrepancy to the identity was larger than that from the discrepancy to the contradistinction. Values of  $\mu(R_{\text{CRU}}^{\text{delta}}, T)$ ,  $\mu(R_{\text{CRU}}^{\text{SP}}, T)$  and  $\mu(R_{\text{CRU}}^{\text{ANN}}, T)$  are positive if we let  $i_1=0.5$ ,  $i_2=0.0$ ,  $i_3=-0.5$  and  $j=-1.0$ . A larger positive value means that the calibrated (verified) and observed time series are more similar and the uncertainty is lower. Therefore, the lowest uncertainty for the reconstruction of MAT was based on the CRU data using the three methods (delta, SP and ANN).

Different methods based on the same data source can generate different results for the MAT. The CRU dataset showed that  $\mu(R_{\text{CRU}}^{\text{delta}}, T) > \mu(R_{\text{CRU}}^{\text{SP}}, T) > \mu(R_{\text{CRU}}^{\text{ANN}}, T)$  and the CMIP5 dataset showed that  $\mu(R_{\text{XXX}}^{\text{ANN}}, T) > \mu(R_{\text{XXX}}^{\text{SP}}, T) > \mu(R_{\text{XXX}}^{\text{delta}}, T) > \mu(R_{\text{XXX}}^{\text{No}}, T)$  (here XXX means any dataset from CMIP5) on average if  $i_1=0.5$ ,  $i_2=0.0$ ,  $i_3=-0.5$  and  $j=-1.0$ . In summary, the SP method produced 11 entries, and ANN and delta methods produced seven and two entries, respectively, among the first 20 entries.

The numbers of positive connection degrees for calibrated and verified MAT time series based on the CMIP5 and CRU datasets using the delta, SP and ANN methods in the TMR are shown in Table S7 (see the Appendix). It can be seen that 51 stations can obtain a positive connection degree if  $i_1=0.5$ ,  $i_2=0.0$ ,  $i_3=-0.5$  and  $j=-1.0$ . The CRU dataset produced low levels of uncertainty for many more stations than the CMIP5 dataset. The delta and SP methods produced low levels of uncertainty based on the CRU dataset and several GCMs from the CMIP5 dataset for all 51 stations. The ANN method could not produce good calibration and verification results for MAT.

Therefore, using the CRU dataset, we can obtain the best performance, with the lowest uncertainty for MAT; while we can acquire a good performance with higher levels of uncertainty for MAT, based on most GCMs from the CMIP5 dataset. The delta and SP methods were more robust in the calibration and verification of MAT time series.

## 5.2 Uncertainty analysis of calibrated and verified MAP

For the set pair  $(R_{\text{CRU}}^{\text{delta}}, P)$ , the identity degree ( $a$ ) was 0.22 and 0.23 for the calibrated and verified MAP; the uncertainty component coefficients of the discrepancy degree ( $b_1$ ,  $b_2$  and  $b_3$ ) were 0.16, 0.10 and 0.06 for the calibrated MAP, and 0.10, 0.07 and 0.08 for the verified MAP, respectively; and the contrary degree was 0.46 and 0.43 for the calibrated and verified MAP, respectively. Therefore,  $\mu(R_{\text{CRU}}^{\text{delta}}, P) = 0.22 + 0.16i_1 + 0.10i_2 + 0.06i_3 + 0.46j$  during the calibration period, which means that the connection degree between calibrated and observed MAP time series for Zhaosu station based on the CRU dataset using the delta method was expressed as:  $\mu(R, P) = 0.22 + 0.16i_1 + 0.10i_2 + 0.06i_3 + 0.46j$ , i.e., 22%, 16%, 10%, 6% and 46% of 348 values of  $\Delta_i$  during the calibration period fell into the ranges of 0–2, 2–4, 4–6, 6–8, and 8 mm– $+\infty$ , respectively (Table S6, see the Appendix). The rest of the results can be explained in the same manner.

According to the criteria in Table 2, calibrated and verified MAP time series based on the CRU dataset by the delta method produced a very good performance. There was a 22%–23% possibility of a very good performance during both the calibration and verification periods. For some GCMs from the CMIP5 dataset, the possibility of providing a very good performance (10%–18%) was much lower than that of making a very bad performance (49%–67%). As for MAP, the uncertainty component coefficients of the discrepancy degree for MAP based on both datasets decreased gradually from  $b_1$  to  $b_3$ . Almost all values of  $b_1$  were larger than the values of  $b_2$  and  $b_3$ . This indicates that the possibility shifted from the discrepancy to the identity was larger than that from the discrepancy to the contradistinction. All 20 connection degree values were negative, with  $\mu(R_{\text{CRU}}^{\text{delta}}, P)$  having the largest value, if we set  $i_1=0.5$ ,  $i_2=0.0$ ,  $i_3=-0.5$  and  $j=-1.0$ . Therefore, using the CRU dataset by the delta method, we can obtain the best performance, with the lowest uncertainty.

Different methods based on the same data source can also generate different results for the MAP.

The CRU dataset showed that  $\mu(R_{\text{CRU}}^{\text{delta}}, P) > \mu(R_{\text{CRU}}^{\text{No}}, P)$  if  $i_1=0.5$ ,  $i_2=0.0$ ,  $i_3=-0.5$  and  $j=-1.0$ , while the CMIP5 dataset showed that  $\mu(R_{\text{XXX}}^{\text{delta}}, P) > \mu(R_{\text{XXX}}^{\text{ANN}}, P) > \mu(R_{\text{XXX}}^{\text{SP}}, P) > \mu(R_{\text{XXX}}^{\text{No}}, P)$  (here XXX means any dataset from CMIP5) on average if  $i_1=0.5$ ,  $i_2=0.0$ ,  $i_3=-0.5$  and  $j=-1.0$ . In summary, the delta method produced 16 entries, and the ANN and No method produced one and three entries, respectively.

The number of positive connection degrees for calibrated and verified MAP time series based on the CMIP5 and CRU datasets using the delta, SP and ANN methods in the TMR are shown in Table S8 (see the Appendix). From the results it can be seen that 33 of the stations selected in this study obtained a positive connection degree if  $i_1=0.5$ ,  $i_2=0.0$ ,  $i_3=-0.5$  and  $j=-1.0$ . The CRU dataset produced low levels of uncertainty for more stations than the CMIP5 dataset. The delta, SP and ANN methods produced low levels of uncertainty based on both the CMIP5 and CRU datasets for all 33 stations. However, the delta method performed slightly better than the SP method while the SP method was better than the ANN method. For some stations, the use of the CMIP5 GCMs and CRU dataset could produce calibrated and verified MAP time series with low levels of uncertainty by the No method.

The best performance, with the lowest level of uncertainty, was obtained for MAP using the CRU dataset, while the performance based on model data from the CMIP5 dataset was not very good, with higher levels of uncertainty. The Zhaosu station is not shown in Table S8, which indicates that the satisfactory calibrated and verified MAT or MAP time series, as determined by traditional performance indices, did not always result in low levels of uncertainty for a given range of absolute error.

### 5.3 Uncertainty comparison between reconstructed MAT and MAP

Although this study used the latest grid data from the CRU and CMIP5 datasets and a range of reconstruction methods, the reconstruction of MAT was obviously superior to that of MAP. This was a similar outcome to that of a previous study by Li et al. (2013). The reconstruction of MAT based on the CRU and CMIP5 datasets and using the three reconstruction methods could capture not only the pattern of intra-annual changes but also the actual values, especially the extreme values, with low levels of uncertainty. For the reconstruction of MAP, the intra-annual pattern of variation could be captured, but the detailed description was poor in terms of actual values. It could not reconstruct the extreme monthly precipitation in one year. Large levels of uncertainty existed during the reconstruction of MAP based on the CRU and CMIP5 datasets using the three reconstruction methods.

The temperature was impacted mainly by the radiation balance, which was apparent from the CRU dataset and GCMs in the CMIP5 dataset. In contrast, precipitation was influenced by many factors such as local climate, topography and vapor sources and was difficult to model using the GCMs in the CMIP5 dataset compared to the CRU dataset. Therefore, the reconstructed time series of MAT could be used for the long-term (from 1850 or 1901 to now) assessment of regional climate change in the TMR.

## 6 Conclusions

By collecting available observed data from meteorological stations in the TMR and grid data from the CRU and the CMIP5 datasets, this study calibrated and verified MAT and MAP in the TMR mainly using the delta, SP and ANN methods. The reconstruction performance and uncertainty from different data sources and methods were assessed and compared using traditional performance assessment indices and the RSPA method. The delta and SP methods resulted in very good and similar performances for calibrated and verified MAT time series in the TMR based on the CRU and CMIP5 datasets, according to performance indices. The performance based on the grid data from the CRU dataset was better than that based on the data from the CMIP5 dataset. The calibration and verification of MAP did not indicate a very good performance compared to the results of MAT. The ANN method was more suitable for the reconstruction of MAP. However, for

a satisfactory calibrated and verified MAT or MAP time series, an assessment by performance indices did not always indicate a low level of uncertainty at a given range of absolute error according to the RSPA method. Most satisfactory calibrated and verified MAT time series and several MAP time series captured the characteristics of their annual periodicity, distribution or pattern well for the targeted stations. The accuracy of values could not be guaranteed for a given range of absolute error. The RSPA method described the numerical accuracy in more detail compared to traditional indices. It is recommended to combine the traditional performance indices with the RSPA method in qualitative and quantitative assessments of climate reconstruction.

This study was a new attempt to calibrate and verify climatic variables (MAT and MAP) using the SP method. The use of the SP method produced a comparable performance to the delta method. Therefore, it can be applied to simulate or reconstruct climatic variables due to its easy implementation. In future work, the data and methods used in this study that produced satisfactory time series, with low levels of uncertainty (for MAT), could be applied to the period of 1901–1960 based on the CRU dataset and the period of 1850–1960 based on the CMIP5 dataset for the assessment of regional climate change in the TMR at the century scale.

## Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (41401050, 41761014), the Foundation of A Hundred Youth Talents Training Program of Lanzhou Jiaotong University, and the Discovery Grant of Natural Sciences and Research Council of Canada. We acknowledge the Coupled Model Inter-comparison Project phase 5 (CMIP5) of the Program for Climate Model Diagnosis and Inter-comparison (PCMDI) and the World Climate Research Programme (WCRP), for collecting and archiving the model output and organizing the analysis of a multi-model dataset. We also acknowledge the help of Professor Slobodan P SIMONOVIC and the members of his research laboratory from the Facility for Intelligent Decision Support (FIDS) in Western University, Canada.

## References

- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. 2000. Artificial neural networks in hydrology I: hydrology application. *Journal of Hydrologic Engineering*, 5(2): 124–137.
- Beale M H, Hagan M T, Demuth H B. 2015. *Neural Network Toolbox User's Guide*. Natick: The Mathworks Press, 1–906.
- Bradley R S, Jones P D. 1994. *Climate Since A.D. 1500*. London: Routledge Press, 511–537.
- Cannas B, Fanni A, See L, et al. 2006. Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. *Physics and Chemistry of the Earth*, 31(18): 1164–1171.
- Chen H, Xu C Y, Guo S. 2012. Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff. *Journal of Hydrology*, 434–435: 36–45.
- Esper J, Cook E R, Schweingruber F H. 2002. Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science*, 295(5563): 2250–2253.
- Fang K Y, Gou X H, Chen F H. 2012. Large-scale precipitation variability over Northwest China inferred from tree rings. *Journal of Climate*, 25: 1357–1357.
- Gaur A, Simonovic S P. 2017a. Accessing vulnerability of land-cover types to climate change using physical scaling downscaling model. *International Journal of Climatology*, 37(6): 2901–2912.
- Gaur A, Simonovic S P. 2017b. Extension of physical scaling method and its application towards downscaling climate model based near surface air temperature. *International Journal of Climatology*, 37(8): 3353–3366.
- Ghosh K G. 2018. Analysis of rainfall trends and its spatial patterns during the last century over the Gangetic West Bengal, Eastern India. *Journal of Geovisualization and Spatial Analysis*, 2: 15, doi: 10.1007/s41651-018-0022-x.
- Gupta H V, Sorooshian S, Yapo P O. 1999. Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, 4(2): 135–143.
- Hao Z H, Li L, Xu Y. 2009. Study on Delta-DCSI downscaling method of GCM output. *Journal of Sichuan University (Engineering Science Edition)*, 41(5): 1–7. (in Chinese)
- Harris I, Jones P D, Osborn T J, et al. 2014. Updated high-resolution grids of monthly climatic observations-the CRU TS3.10 dataset. *International Journal of Climatology*, 34(3): 623–642.

- Hauduc H, Neumann M B, Muschalla D, et al. 2015. Efficiency criteria for environmental model quality assessment: A review and its application to wastewater treatment. *Environmental Modelling & Software*, 68: 196–204.
- Hay L E, Wilby I L, Leavesley G H. 2000. A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States. *Journal of the American Water Resource Association*, 36(2): 387–397.
- Hu R J. 2004. *Physical Geography of the Tianshan Mountains in China*. Beijing: China Environmental Science Press, 139–142. (in Chinese)
- Huang Y, Yan Q, Zhang C. 2018. Spatial-temporal distribution characteristics of PM<sub>2.5</sub> in China in 2016. *Journal of Geovisualization and Spatial Analysis*, 2(2): 1–18.
- Klippel L, Krusic P J, Brandes R, et al. 2018. A 1286-year hydro-climate reconstruction for the Balkan Peninsula. *Boreas*, 47(4): 1218–1229.
- Levenberg K. 1944. A method for the solution of certain non-linear problems in least squares. *The Quarterly of Applied Mathematics*, 2: 164–168.
- Li X M, Li L H, Wang X X, et al. 2013. Reconstruction of hydrometeorological time series and its uncertainties for the Kaidu River Basin using multiple data sources. *Theoretical and applied climatology*, 113: 45–62.
- Li X M, Gao P, Li Q, Tang H. 2016. Multi-paths impact from climate change on snow cover in Tianshan Mountainous area of China. *Advance in Climate Change Research*, 12(4): 303–312. (in Chinese)
- Liu X K, Rao Z G, Zhang X J, et al. 2015. Variations in the oxygen isotopic composition of precipitation in the Tianshan Mountains region and their significance for the Westerly circulation. *Journal of Geographical Sciences*, 25(7): 801–816.
- Marquardt D. 1963. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11(2): 431–441.
- Misra V, DiNapoli S M, Bastola S. 2012. Dynamic downscaling of the twentieth-century reanalysis over the southeastern United States. *Regional Environment Change*, 13: 15–23.
- Moriasi D N, Arnold J G, van Liew M W, et al. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the American Society of Agricultural and Biological Engineers*, 50(3): 885–900.
- Nash J E, Sutcliffe J V. 1970. River flow forecasting through conceptual models Part 1: A discussion of principles. *Journal of Hydrology*, 10(3): 282–290.
- Nourani V, Komasi M, Mano A. 2009. A multivariate ANN-wavelet approach for rainfall-runoff modeling. *Water Resources Management*, 23: 2877–2894.
- Pi H, Sharratt B, Feng G, et al. 2017. Evaluation of two empirical wind erosion models in arid and semi-arid regions of China and the USA. *Environmental Modelling & Software*, 91: 28–46.
- Schimek M G, Turlach B A. 2000. Additive and generalized additive models. In: Schimek M G. *Smoothing and Regression: Approaches, Computation, and Application*. New York: Wiley Press, 277–327.
- Shen J, Zhang E L, Xia W L. 2001. Records from lake sediments of the Qinghai Lake to mirror climatic and environmental changes of the past about 1000 years. *Quaternary Sciences*, 21(6): 508–514.
- Shen S S P, Clarke G, Shen B W, et al. 2019. Spatiotemporal variations of the twentieth century Tibetan Plateau precipitation based on the monthly 2.5° reconstructed data. *Theoretical and Applied Climatology*, 135: 71–83.
- Shi Y, Shen Y, Kang E, et al. 2007. Recent and future climate change in northwest China. *Climate Change*, 80: 379–393.
- Srinivasulu S, Jain A. 2006. A comparative analysis of training methods for artificial neural network rainfall-runoff models. *Applied Soft Computing*, 6: 295–306.
- Su M R, Yang Z F, Chen B, et al. 2009. Urban ecosystem health assessment based on energy and set pair analysis—a comparative study of typical Chinese cities. *Ecological Modelling*, 220(18): 2341–2348.
- Taylor K E, Stouffer R J, Meehl G A. 2012. A summary of the CMIP5 experiment design. *Bulletin of the American Meteorological Society*, 93(4): 485–498.
- University of East Anglia Climatic Research Unit, Harris I C, Jones P D. 2020. CRU TS4.03: Climatic Research Unit (CRU) Time-Series (TS) version 4.03 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901–Dec. 2018). [2020-01-22]. Centre for Environmental Data Analysis, doi: 10.5285/10d3e3640f004c578403419aac167d82.
- Wang G Q, Zhao K Q, Zhen X J. 2004. Application of set pair analysis to fuzzy predictors of multiple regression weather forecast models. *Bulletin of Science and Technology*, 20(2): 151–155. (in Chinese)
- Wang H Q, Chen F, Ermenbaev B, et al. 2017. Comparison of drought-sensitive tree-ring records from the Tien Shan of Kyrgyzstan and Xinjiang (China) during the last six centuries. *Advances in Climate Change Research*, 8(1): 18–25.
- Wang M W, Wei D F, Li J, et al. 2014. A novel clustering model based on set pair analysis for the energy consumption forecast in China. *Mathematical Problems in Engineering*, 2014(1): 1–8, doi: 10.1155/2014/191242.

- Wang W S, Xiang H L, Li Y Q, et al. 2008. A new approach to annual runoff classification based on set pair analysis. *Journal of Sichuan University (Engineering Science Edition)*, 40(5): 1–6. (in Chinese)
- Wang W S, Jin J L, Ding J, et al. 2009. A new approach to water resources system assessment-set pair analysis method. *Science in China Series E: Technological Sciences*, 52(10): 3017–3023.
- Wang W S, Li Y Q. 2012. Hazard degree assessment of landslide using set pair analysis method. *Natural Hazards*, 60(2): 367–379.
- Wei W S, Yuan Y J, Yu S L, et al. 2008. Climate change in recent 235 years and trend prediction in Tianshan Mountainous area. *Journal of Desert Research*, 28(5): 803–808. (in Chinese)
- Wood S N. 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2): 413–428.
- Wu Z T, Zhang H J, Krause C M, et al. 2010. Climate change and human activities: a case study in Xinjiang, China. *Climatic Change*, 99: 457–472.
- Yan F, Xu K, Li D, et al. 2017. A novel hazard assessment method for biomass gasification stations based on extended set pair analysis. *PLoS ONE*, 12(9): e0185006.
- Yang B, Braeuning A, Johnson K R, et al. 2002. General characteristics of temperature variation in China during the last two millennia. *Geophysical Research Letters*, 29(9): 1029–1040.
- Yang J J, Zhou J Z, Liu L, et al. 2008. Similarity Measures between Connection Numbers of Set Pair Analysis. In: Sun F, Zhang J, Tan Y, et al. *Advances in Neural Networks - ISNN 2008*. ISNN 2008. *Lecture Notes in Computer Science*, Vol. 5263. Berlin, Heidelberg: Springer, 63–68.
- Yao T, Masson-Delmotte V, Gao J, et al. 2013. A review of climatic controls on delta  $^{18}\text{O}$  in precipitation over the Tibetan Plateau: observations and simulations. *Review of Geophysics*, 51(4): 525–548.
- Ye Y C, Wu Y L. 2000. The set pair analysis for the comprehensive evaluation of mine operation state. *Metal Mine*, 6: 23–33. (in Chinese)
- Zhang F Y, Ahmad S, Zhang H Q, et al. 2016a. Simulating low and high streamflow driven by snowmelt in an insufficiently gauged alpine basin. *Stochastic Environmental Research and Risk Assessment*, 30: 59–75.
- Zhang F Y, Bai L, Li L H, et al. 2016b. Sensitivity of runoff to climatic variability in the northern and southern slopes of the Middle Tianshan Mountains, China. *Journal of Arid Land*, 8(5): 681–693.
- Zhang X T, Li X M, Li L H, et al. 2019. Environmental factors influencing snowfall and snowfall prediction in the Tianshan Mountains, Northwest China. *Journal of Arid Land*, 11(1): 15–28.
- Zhao F F, Xu Z X. 2007. Comparative analysis on downscaled climate scenarios for headwater catchment of Yellow River using SDS and Delta methods. *Acta Meteorologica Sinica*, 65(4): 653–662. (in Chinese)
- Zhao K Q. 1989. Set pair and set pair analysis—a new concept and systematic analysis method. In: *Proceedings of the National Conference on System Theory and Regional Planning*. Baotou, China, 87–91. (in Chinese)
- Zhao K Q, Xuan A L. 1996. Set pair theory—a new theory method of non-define and its applications. *Systems Engineering*, 14(1): 18–23. (in Chinese)
- Zhou J X, Li X M, Chen L, et al. 2007. Updating methods for dynamic assessment of ecological footprint—a case study of Wuhan City. *Resources Science*, 29(3): 111–116. (in Chinese)

## Appendix

**Table S1** Descriptive information of the meteorological stations selected in the Tianshan Mountains region (TMR)

No.	Station name	Station code	Latitude	Longitude	Elevation (m)	Time period
1	Akqi	AKQ	78.45°N	40.93°E	1985.700	1961–2011
2	Aksu	AKS	80.23°N	41.17°E	1104.730	1961–2011
3	Baicheng	BCH	81.90°N	41.78°E	1230.000	1961–2011
4	Balguntay	BLGT	86.30°N	42.73°E	1739.470	1961–2011
5	Barkol	BKL	93.05°N	43.60°E	1675.370	1961–2011
6	Bayanbulak	BYBLK	84.15°N	43.03°E	2459.270	1961–2011
7	Bole	BL	82.07°N	44.90°E	532.567	1961–2011
8	Caijiahu	CJH	87.53°N	44.20°E	440.967	1961–2011
9	Dabancheng	DBCH	88.32°N	43.35°E	1104.770	1961–2011
10	Daxigou	DXG	86.83°N	43.10°E	3539.000	1961–2011
11	Gongliu	GL	82.23°N	43.47°E	775.970	1961–2011
12	Hami	HM	93.52°N	42.82°E	738.200	1961–2011
13	Hejing	HJ	86.40°N	42.32°E	1102.370	1961–2011
14	Hoxud	HX	86.80°N	42.25°E	1086.130	1961–2011
15	Hutubi	HTB	86.82°N	44.13°E	522.970	1961–2011
16	Jiashi	JSH	76.73°N	39.50°E	1208.600	1961–2011
17	Jimsar	JMS	89.17°N	44.02°E	735.570	1961–2011
18	Jinghe	JH	82.90°N	44.60°E	320.870	1961–2011
19	Kalpin	KP	79.05°N	40.50°E	1162.630	1961–2011
20	Kashi	KSH	75.99°N	39.27°E	1291.200	1961–2011
21	Khorgas	KG	80.42°N	44.20°E	772.970	1961–2011
22	Korla	KL	86.13°N	41.25°E	932.433	1961–2011
23	Kumux	KM	88.22°N	42.23°E	923.470	1961–2011
24	Kuqa	KQ	82.97°N	41.73°E	1082.930	1961–2011
25	Manas	MNS	86.20°N	44.32°E	471.400	1961–2011
26	Mori	MR	90.28°N	43.83°E	1271.970	1961–2011
27	Mosuowan	MSW	86.10°N	45.02°E	347.230	1961–2011
28	Naomaohu	NMH	95.13°N	43.77°E	479.030	1961–2011
29	Nilka	NLK	82.62°N	43.80°E	1106.130	1961–2011
30	Paotai	PT	85.25°N	44.85°E	337.100	1961–2011
31	Qapqal	QPQ	81.15°N	43.83°E	603.800	1961–2011
32	Qitai	QT	89.57°N	44.02°E	794.100	1961–2011
33	Shanshan	SHSH	90.23°N	42.85°E	398.930	1961–2011
34	Shawan	SHW	85.62°N	44.33°E	522.200	1961–2011
35	Shihezi	SHHZ	86.05°N	44.32°E	443.730	1961–2011
36	Shisanjianfang	SHSJF	91.73°N	43.22°E	722.900	1961–2011
37	Tekes	TKS	81.77°N	43.18°E	1210.530	1961–2011
38	Tianchi	TCH	88.12°N	43.88°E	1942.170	1961–2011
39	Toksun	TKS	88.63°N	42.80°E	1.670	1961–2011
40	Tuergate	TEGT	75.40°N	40.52°E	3506.400	1961–2011
41	Turpan	TP	89.20°N	42.93°E	34.970	1961–2011
42	Urumqi	URMQ	87.65°N	43.78°E	935.670	1961–2011
43	Usu	US	84.67°N	44.43°E	478.970	1961–2011
44	Wenquan	WEQ	81.02°N	44.97°E	1358.600	1961–2011
45	Wuqia	WUQ	75.25°N	39.72°E	2176.900	1961–2011
46	Wushi	WSH	79.23°N	41.22°E	1396.470	1961–2011
47	Xiaoquzi	XQZ	87.10°N	43.49°E	1873.800	1961–2011
48	Xinhe	XH	82.61°N	41.53°E	1014.200	1961–2011
49	Xinyuan	XY	83.30°N	43.45°E	929.200	1961–2011
50	Yanqi	YQ	86.57°N	42.08°E	1056.600	1961–2011
51	Yining	YN	81.33°N	43.95°E	663.200	1961–2011
52	Yiwu	YW	94.70°N	43.27°E	1728.600	1961–2011
53	Zhaosu	ZHS	81.13°N	43.15°E	1853.400	1961–2011

**Table S2** Basic information of the CMIP5 (Coupled Model Inter-comparison Project phase 5) models

No.	Model name	Source and country	Resolution	Historical period
1	ACCESS1-0	CSIRO-BOM, Australia	192×145	Jan 1850–Dec 2005
2	ACCESS1-3	CSIRO-BOM, Australia	192×145	Jan 1850–Dec 2005
3	BCC-CSM1-1	BCC, China	128×64	Jan 1850–Dec 2012
4	BCC-CSM1-1-M	BCC, China	320×160	Jan 1850–Dec 2012
5	BNU-ESM	GCESS, China	128×64	Jan 1850–Dec 2005
6	CanCM4	CCCMA, Canada	128×64	Jan 1850–Dec 2005
7	CanESM2	CCCMA, Canada	128×64	Jan 1850–Dec 2005
8	CCSM4	NCAR, U.S.	288×192	Jan 1850–Dec 2005
9	CESM1-BGC	NSF-DOE-NCAR, U.S.	288×192	Jan 1850–Dec 2005
10	CESM1-CAM5	NSF-DOE-NCAR, U.S.	288×192	Jan 1850–Dec 2005
11	CESM1-FASTCHEM	NSF-DOE-NCAR, U.S.	288×192	Jan 1850–Dec 2005
12	CESM1-WACCM	NSF-DOE-NCAR, U.S.	144×96	Jan 1850–Dec 2005
13	CMCC-CESM	CMCC, Italy	96×48	Jan 1850–Dec 2005
14	CMCC-CM	CMCC, Italy	480×240	Jan 1850–Dec 2005
15	CMCC-CMS	CMCC, Italy	192×96	Jan 1850–Dec 2005
16	CNRM-CM5	CNRM-CERFACS, French	256×128	Jan 1850–Dec 2005
17	CSIRO-Mk3-6-0	CSIRO-QCCCE, Australia	192×96	Jan 1850–Dec 2005
18	EC-EARTH	EC-EARTH, 10-European nations	320×160	Jan 1950–Dec 2012
19	FGOALS-g2	FGOALS, China	128×60	Jan 1960–Jan 1999
20	FGOALS-s2	FGOALS, China	128×108	Jan 1850–Dec 2005
21	FIO-ESM	FIO, China	128×64	Jan 1850–Dec 2005
22	GFDL-ESM2G	NOAA GFDL, U.S.	144×90	Jan 1961–Dec 2000
23	GFDL-ESM2M	NOAA GFDL, U.S.	144×90	Jan 1961–Dec 2000
24	GISS-E2-H	NASA GISS, U.S.	144×90	Jan 1951–Dec 2005
25	GISS-E2-H-CC	NASA GISS, U.S.	144×90	Jan 1951–Dec 2010
26	GISS-E2-R	NASA GISS, U.S.	144×90	Jan 1951–Dec 2005
27	GISS-E2-R-CC	NASA GISS, U.S.	144×90	Jan 1951–Dec 2005
28	HadCM3	MOHC, England	96×73	Jan 1959–Dec 2005
29	HadGEM2-AO	NIMR/KMA, SK/England	192×145	Jan 1860–Dec 2005
30	HadGEM2-CC	MOHC, England	192×145	Dec 1959–Nov 2005
31	HadGEM2-ES	MOHC, England	192×145	Dec 1959–Nov 2005
32	INMCM4	INM, Russia	180×120	Jan 1850–Dec 2005
33	IPSL-CM5A-LR	IPSL, French	96×96	Jan 1850–Dec 2005
34	IPSL-CM5A-MR	IPSL, French	144×143	Jan 1850–Dec 2005
35	IPSL-CM5B-LR	IPSL, French	96×96	Jan 1850–Dec 2005
36	MIROC4h	MIROC, Japan	640×320	Jan 1961–Dec 2000
37	MIROC5	MIROC, Japan	256×128	Jan 1850–Dec 2012
38	MIROC-ESM	MIROC, Japan	128×64	Jan 1950–Dec 2005
39	MIROC-ESM-CHEM	MIROC, Japan	128×64	Jan 1850–Dec 2005
40	MRI-CGCM3	MRI, Japan	320×160	Jan 1850–Dec 2005
41	NorESM1-M	NCC, Norway	144×96	Jan 1850–Dec 2005
42	NorESM1-ME	NCC, Norway	144×96	Jan 1850–Dec 2005

**Table S3** Numbers of satisfactory calibrated and verified monthly average temperature (MAT) time series based on the CMIP5 and Climatic Research Unit (CRU) datasets using the delta, physical scaling (SP), artificial neural network (ANN) and No methods

No.	Station name	Number of satisfactory calibrated and verified MAT time series						CRU	CMIP5
		Total	A (very good)	Delta method	SP method	ANN method	No method		
1	Akqi	88	87	41	42	0	5	2	86
2	Aksu	90	85	41	42	0	7	3	87
3	Baicheng	91	86	41	42	0	8	2	89
4	Balguntay	102	89	41	42	0	19	2	100
5	Barkol	69	55	41	27	0	1	2	67
6	Bayanbulak	82	82	41	41	0	0	2	80
7	Bole	91	84	41	42	0	8	2	89
8	Caijiahu	92	84	41	41	0	10	2	90
9	Dabancheng	93	87	41	42	0	10	3	90
10	Daxigou	83	83	41	42	0	0	2	81
11	Gongliu	90	88	41	42	0	7	2	88
12	Hami	91	86	41	42	0	8	3	88
13	Hejing	93	91	41	42	0	10	3	90
14	Hoxud	93	90	41	42	0	10	3	90
15	Hutubi	91	85	41	42	0	8	3	88
16	Jiashi	86	83	41	42	0	3	2	84
17	Jimsar	92	86	41	42	0	9	2	90
18	Jinghe	102	93	41	42	0	19	2	100
19	Kalpin	86	83	41	42	0	3	2	84
20	Kashi	102	93	41	42	0	19	2	100
21	Khorgas	99	89	41	42	0	16	3	96
22	Korla	102	93	41	42	0	19	3	99
23	Kumux	101	90	41	42	0	18	3	98
24	Kuqa	91	85	41	42	0	8	3	88
25	Manas	101	90	41	42	0	18	2	99
26	Mori	100	90	41	42	0	17	3	97
27	Mosuowan	102	90	41	42	0	19	2	100
28	Naomaohu	85	84	41	42	0	2	3	82
29	Nilka	91	89	41	42	0	8	2	89
30	Paotai	101	91	41	42	0	18	2	99
31	Qapqal	102	98	41	42	0	19	3	99
32	Qitai	99	90	41	42	0	16	3	96
33	Shanshan	83	83	41	42	0	0	2	81
34	Shawan	105	99	41	42	0	22	2	103
35	Shihezi	109	91	41	42	0	26	2	107
36	Shisanjianfang	97	87	41	42	0	14	3	94
37	Tekes	87	87	41	42	0	4	2	85
38	Tianchi	82	75	41	40	0	1	1	81
39	Toksun	85	83	41	42	0	2	3	82
40	Tuergate	87	85	41	42	0	4	2	85
41	Turpan	84	83	41	42	0	1	2	82
42	Urumqi	100	89	41	42	0	17	2	98
43	Usu	101	90	41	42	0	18	3	98
44	Wenquan	127	94	41	41	35	10	2	125
45	Wuqia	83	83	41	42	0	0	2	81
46	Wushi	85	83	41	41	0	3	1	84
47	Xiaoquzi	91	79	41	41	1	8	2	89
48	Xinhe	95	91	41	42	0	12	3	92
49	Xinyuan	90	88	41	42	0	7	2	88
50	Yanqi	98	90	41	42	0	15	3	95
51	Yining	89	86	41	42	0	6	2	87
52	Yiwu	135	106	41	42	37	15	3	132
53	Zhaosu	128	119	41	42	42	3	3	125

**Table S4** Numbers of satisfactory calibrated and verified monthly accumulated precipitation (MAP) time series based on the CRU and CMIP5 datasets using the delta, SP, ANN and No methods

No.	Station name	Number of satisfactory calibrated and verified MAP time series							CRU	CMIP5
		Total	C/B	C/C	Delta method	SP method	ANN method	No method		
1	Bayanbulak	13	0	12	12	1	0	0	0	13
2	Daxigou	8	5	2	8	0	0	0	0	8
3	Kashi	3	0	0	1	1	1	0	3	0
4	Korla	3	0	0	1	1	1	0	3	0
5	Kuqa	1	0	0	1	0	0	0	1	0
6	Manas	3	1	2	1	1	1	0	3	0
7	Mori	2	2	0	0	1	1	0	2	0
8	Qitai	3	0	0	1	1	1	0	3	0
9	Tekes	1	0	1	0	1	0	0	0	1
10	Tianchi	3	0	0	1	1	1	0	3	0
11	Tuergate	1	0	0	1	0	0	0	1	0
12	Turpan	1	0	0	1	0	0	0	1	0
13	Urumqi	2	0	0	0	1	1	0	2	0
14	Usu	4	0	0	1	1	1	1	4	0
15	Xiaoquzi	7	0	5	2	3	2	0	3	4
16	Xinyuan	1	0	1	0	1	0	0	1	0
17	Yanqi	1	0	1	1	0	0	0	1	0
18	Yining	4	0	0	1	1	1	1	4	0
19	Yiwu	4	0	0	1	1	1	1	4	0
20	Zhaosu	26	6	20	9	9	6	2	0	26

Note: B, good; C, satisfactory.

**Table S5** Connection degree expressions between calibrated (verified) and observed MAT time series (first 20 entries) at Zhaosu station

No.	Connection degree	$a$		$b_1$		$b_2$		$b_3$		$c$	
		C	V	C	V	C	V	C	V	C	V
1	$\mu(R_{\text{CRU}}^{\text{delta}}, T)$	0.58	0.59	0.24	0.26	0.11	0.06	0.04	0.06	0.03	0.03
2	$\mu(R_{\text{CRU}}^{\text{SP}}, T)$	0.53	0.56	0.28	0.28	0.13	0.08	0.04	0.06	0.03	0.03
3	$\mu(R_{\text{CRU}}^{\text{ANN}}, T)$	0.51	0.56	0.31	0.27	0.13	0.10	0.03	0.05	0.02	0.03
4	$\mu(R_{\text{CanCM4}}^{\text{ANN}}, T)$	0.26	0.25	0.15	0.13	0.17	0.10	0.10	0.13	0.32	0.39
5	$\mu(R_{\text{FGOALS-g2}}^{\text{SP}}, T)$	0.26	0.19	0.16	0.24	0.13	0.14	0.14	0.13	0.32	0.30
6	$\mu(R_{\text{BNU-ESM}}^{\text{SP}}, T)$	0.25	0.14	0.15	0.27	0.16	0.13	0.11	0.13	0.33	0.33
7	$\mu(R_{\text{CESM1-BGC}}^{\text{SP}}, T)$	0.25	0.26	0.19	0.16	0.14	0.15	0.10	0.15	0.32	0.28
8	$\mu(R_{\text{CESM1-WACCM}}^{\text{SP}}, T)$	0.25	0.23	0.16	0.14	0.14	0.16	0.11	0.14	0.34	0.33
9	$\mu(R_{\text{CSIRO-Mk3-6-0}}^{\text{delta}}, T)$	0.24	0.20	0.16	0.16	0.14	0.19	0.11	0.15	0.35	0.30
10	$\mu(R_{\text{MIROC-ESM-CHEM}}^{\text{SP}}, T)$	0.24	0.15	0.16	0.25	0.14	0.17	0.11	0.09	0.36	0.34
11	$\mu(R_{\text{GISS-E2-R}}^{\text{SP}}, T)$	0.24	0.19	0.20	0.19	0.13	0.11	0.11	0.18	0.32	0.33
12	$\mu(R_{\text{CESM1-BGC}}^{\text{ANN}}, T)$	0.24	0.23	0.22	0.20	0.14	0.15	0.11	0.13	0.30	0.28
13	$\mu(R_{\text{CESM1-FASTCHEM}}^{\text{ANN}}, T)$	0.24	0.22	0.17	0.24	0.14	0.11	0.10	0.11	0.35	0.33
14	$\mu(R_{\text{GISS-E2-R}}^{\text{ANN}}, T)$	0.24	0.23	0.19	0.18	0.14	0.15	0.12	0.14	0.32	0.30
15	$\mu(R_{\text{CESM1-FASTCHEM}}^{\text{SP}}, T)$	0.23	0.23	0.21	0.18	0.11	0.10	0.11	0.11	0.34	0.35
16	$\mu(R_{\text{FGOALS-g2}}^{\text{SP}}, T)$	0.23	0.23	0.15	0.19	0.17	0.17	0.10	0.12	0.35	0.30
17	$\mu(R_{\text{GISS-E2-H}}^{\text{SP}}, T)$	0.23	0.17	0.18	0.24	0.13	0.18	0.12	0.08	0.34	0.33
18	$\mu(R_{\text{CMCC-CM3}}^{\text{SP}}, T)$	0.23	0.21	0.18	0.15	0.14	0.18	0.12	0.12	0.34	0.35
19	$\mu(R_{\text{MIROC5}}^{\text{ANN}}, T)$	0.23	0.21	0.21	0.19	0.15	0.18	0.11	0.09	0.29	0.33
20	$\mu(R_{\text{HadGEM2-ES}}^{\text{ANN}}, T)$	0.23	0.18	0.18	0.21	0.15	0.14	0.12	0.13	0.32	0.34

Note:  $a$ , identity degree;  $b_1$ ,  $b_2$  and  $b_3$ , uncertainty component coefficients of the discrepancy degree;  $c$ , contrary degree; C, calibrated; V, verified.

**Table S6** Connection degree expressions between calibrated (verified) and observed MAP (first 20 entries) at Zhaosu station

No.	Connection degree	$a$		$b_1$		$b_2$		$b_3$		$c$	
		C	V	C	V	C	V	C	V	C	V
1	$\mu(R_{CRU}^{\text{delta}}, P)$	0.22	0.23	0.16	0.10	0.10	0.07	0.06	0.08	0.46	0.53
2	$\mu(R_{MIROC4h}^{\text{delta}}, P)$	0.18	0.13	0.11	0.09	0.11	0.13	0.08	0.06	0.53	0.58
3	$\mu(R_{bcc-csm1-1-m5}^{\text{delta}}, P)$	0.16	0.13	0.12	0.09	0.10	0.13	0.06	0.08	0.57	0.56
4	$\mu(R_{inmcm4}^{\text{delta}}, P)$	0.16	0.12	0.14	0.13	0.09	0.11	0.07	0.05	0.54	0.60
5	$\mu(R_{MIROC4h}^{\text{ANN}}, P)$	0.16	0.10	0.08	0.08	0.11	0.10	0.07	0.11	0.57	0.62
6	$\mu(R_{MIROC5}^{\text{delta}}, P)$	0.15	0.12	0.13	0.13	0.10	0.08	0.08	0.08	0.55	0.59
7	$\mu(R_{CCSM4}^{\text{delta}}, P)$	0.15	0.17	0.14	0.13	0.11	0.06	0.07	0.07	0.53	0.58
8	$\mu(R_{CESM1-FASTCHEM}^{\text{delta}}, P)$	0.15	0.13	0.11	0.13	0.10	0.11	0.11	0.04	0.53	0.60
9	$\mu(R_{SRO-Mk3-6-0}^{\text{delta}}, P)$	0.15	0.13	0.13	0.13	0.07	0.08	0.09	0.10	0.56	0.57
10	$\mu(R_{MRI-CGCM3}^{\text{delta}}, P)$	0.15	0.18	0.12	0.13	0.11	0.10	0.07	0.07	0.55	0.53
11	$\mu(R_{NorESM1-M}^{\text{delta}}, P)$	0.15	0.13	0.11	0.10	0.10	0.09	0.08	0.11	0.56	0.57
12	$\mu(R_{CMCC-CM}^{\text{delta}}, P)$	0.15	0.14	0.11	0.08	0.11	0.10	0.09	0.10	0.54	0.58
13	$\mu(R_{ACCESS1-0}^{\text{No}}, P)$	0.14	0.13	0.11	0.10	0.08	0.11	0.08	0.08	0.59	0.58
14	$\mu(R_{HadGEM2-CC}^{\text{No}}, P)$	0.14	0.12	0.09	0.06	0.06	0.08	0.06	0.10	0.66	0.64
15	$\mu(R_{HadGEM2-ES}^{\text{No}}, P)$	0.14	0.11	0.09	0.08	0.08	0.09	0.07	0.07	0.61	0.65
16	$\mu(R_{ACCESS1-0}^{\text{delta}}, P)$	0.14	0.08	0.13	0.14	0.11	0.12	0.08	0.09	0.54	0.58
17	$\mu(R_{CESM1-BGC}^{\text{delta}}, P)$	0.14	0.13	0.14	0.15	0.12	0.12	0.08	0.07	0.53	0.53
18	$\mu(R_{IPSL-CM5A-MR}^{\text{delta}}, P)$	0.14	0.14	0.13	0.11	0.09	0.13	0.08	0.07	0.56	0.55
19	$\mu(R_{MIROC-ESM}^{\text{delta}}, P)$	0.14	0.1	0.13	0.16	0.09	0.08	0.09	0.06	0.56	0.60
20	$\mu(R_{CNRM-CM5}^{\text{delta}}, P)$	0.14	0.15	0.14	0.08	0.09	0.12	0.07	0.05	0.56	0.60

Note:  $a$ , identity degree;  $b_1$ ,  $b_2$  and  $b_3$ , uncertainty component coefficients of the discrepancy degree;  $c$ , contrary degree; C, calibrated; V, verified.

**Table S7** Number of positive connection degrees for calibrated and verified MAT time series based on the CMIP5 and CRU datasets using the delta, SP, ANN and No methods

No.	Station name	Number of positive connection degrees for calibrated and verified MAT time series						
		Total	Delta method	SP method	ANN method	No method	CRU	CMIP5
1	Akqi	4	2	2	0	0	2	2
2	Aksu	19	8	10	0	1	3	16
3	Baicheng	2	1	1	0	0	2	0
4	Barkol	2	1	1	0	0	2	0
5	Balguntay	19	1	18	0	0	2	17
6	Bayanbulak	2	1	1	0	0	2	0
7	Bole	2	1	1	0	0	2	0
8	Caijiahu	2	1	1	0	0	2	0
9	Dabancheng	3	1	1	0	1	3	0
10	Daxigou	14	1	13	0	0	2	12
11	Gongliu	2	1	1	0	0	2	0
12	Hami	3	1	1	0	1	3	0
13	Hejing	3	1	2	0	0	2	1
14	Hoxud	2	1	1	0	0	2	0
15	Hutubi	2	1	1	0	0	2	0
16	Jiashi	4	1	3	0	0	2	2
17	Jinghe	2	1	1	0	0	2	0
18	Kalpin	3	2	1	0	0	2	1
19	Khorgas	2	1	1	0	0	2	0
20	Korla	10	4	6	0	0	2	8
21	Kuqa	5	3	1	0	1	3	2
22	Kumux	4	2	2	0	0	2	2
23	Manas	2	1	1	0	0	2	0
24	Mori	3	1	1	0	1	3	0
25	Mosuowan	2	1	1	0	0	2	0
26	Naomaohu	2	1	1	0	0	2	0
27	Nilka	2	1	1	0	0	2	0
28	Paotai	2	1	1	0	0	2	0
29	Qapqal	2	1	1	0	0	2	0
30	Qitai	3	1	1	0	1	3	0
31	Shanshan	1	1	0	0	0	1	0

To be continued

Continued

No.	Station name	Number of positive connection degrees for calibrated and verified MAT time series						CMIP5
		Total	Delta method	SP method	ANN method	No method	CRU	
32	Shawan	2	1	1	0	0	2	0
33	Shihezi	2	1	1	0	0	2	0
34	Shisanjianfang	3	1	1	0	1	3	0
35	Tekes	2	1	1	0	0	2	0
36	Tianchi	3	1	1	1	0	3	0
37	Toksun	2	1	1	0	0	2	0
38	Tuergate	5	2	3	0	0	2	3
39	Turpan	2	1	1	0	0	2	0
40	Urumqi	2	1	1	0	0	2	0
41	Usu	2	1	1	0	0	2	0
42	Wenquan	1	1	0	0	0	1	0
43	Wuqia	2	1	1	0	0	2	0
44	Wushi	3	2	1	0	0	3	0
45	Xiaoquzi	4	1	1	1	1	4	0
46	Xinhe	22	10	11	0	1	3	19
47	Xinyuan	2	1	1	0	0	2	0
48	Yanqi	15	6	8	0	1	3	12
49	Yining	2	1	1	0	0	2	0
50	Yiwu	3	1	2	0	0	2	1
51	Zhaosu	3	1	1	1	0	3	0

**Table S8** Number of positive connection degrees for calibrated and verified MAP time series based on the CMIP5 and CRU datasets using the delta, SP, ANN and No methods

No.	Station name	Number of positive connection degrees for calibrated and verified MAP time series						CMIP5
		Total	Delta method	SP method	ANN method	No method	CRU	
1	Aksu	101	43	30	28	0	3	98
2	Balguntay	48	43	0	0	5	2	46
3	Bayanbulak	7	7	0	0	0	0	7
4	Caijiahu	3	1	1	1	0	3	0
5	Dabancheng	97	43	30	24	0	3	94
6	Hami	137	43	41	36	17	3	134
7	Hejing	121	43	38	34	6	4	117
8	Hoxud	61	43	4	10	4	3	58
9	Jiashi	110	43	34	28	5	3	107
10	Jinghe	54	34	3	8	9	3	51
11	Kalpin	53	43	3	4	3	4	49
12	Kashi	127	43	43	40	1	4	123
13	Korla	135	43	41	34	17	4	131
14	Kuqa	105	43	29	29	4	4	101
15	Kumux	129	43	39	37	10	4	125
16	Manas	4	1	1	1	1	4	0
17	Mosuwon	4	1	1	1	1	4	0
18	Naomaohu	137	43	43	42	9	4	133
19	Paotai	3	1	1	0	1	3	0
20	Qitai	4	1	1	1	1	4	0
21	Shanshan	132	43	43	43	3	3	129
22	Shisanjianfang	131	43	43	43	2	3	128
23	Toksun	63	0	42	21	0	1	62
24	Tuergate	1	1	0	0	0	1	0
25	Turpan	63	1	41	21	0	1	62
26	Urumqi	3	1	1	1	0	3	0
27	Usu	4	1	1	1	1	4	0
28	Wuqia	1	1	0	0	0	1	0
29	Wushi	10	9	1	0	0	2	8
30	Xinhe	61	43	3	10	5	4	57
31	Yanqi	59	43	4	7	5	4	55
32	Yining	4	1	1	1	1	4	0
33	Yiwu	58	43	1	3	11	4	54