

Journal of Environmental Informatics 37(1) 36-48 (2021)

Journal of Environmental Informatics

www.iseis.org/jei

Comparing the Performance of An Autoregressive State-Space Approach to the Linear Regression and Artificial Neural Network for Streamflow Estimation

Y. Yang^{1*}, T. T. Huang¹, Y. Z. Shi¹, O. Wendroth², and B. Y. Liu³

¹ State Key Laboratory of Earth Surface Processes and Resource Ecology, School of Geography, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

² Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY 40546, USA

³ School of Geography, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

Received 5 May 2019; revised 9 November 2019; accepted 27 January 2020; published online 06 August 2020

ABSTRACT. Accurate streamflow estimation remains a great challenge although diverse modeling techniques have been developed during recent decades. In contrast to the process-based models, the empirical data-driven methods are easy to operate, require low computing capacity and yield fairly accurate outcomes, among which the state-space (STATE) approach takes use of the temporal structures inherent in streamflow series and serves as a feasible solution for streamflow estimation. Yet this method has rarely been applied, neither its comparison with other methods. The objective was to compare the performance of an autoregressive STATE approach to the traditional multiple linear regression and artificial neural network in simulating annual streamflow series of 15 catchments located in the Loess Plateau of China. Annual data of streamflow (Q), precipitation (P) and potential evapotranspiration (PET) during 1961 ~ 2013 were collected. The results show that STATE was generally the most accurate method for Q estimation, explaining almost 90% of the total variance averaged over all the 15 catchments. The estimation of streamflow were almost equal. Missing data were estimated using the STATE method, which allowed inter-annual trend analysis of the streamflow. Significant downward trends were manifested at all the 15 catchments during the study period and the corresponding slopes ranged from -0.24 to -1.71 mm y⁻¹. These findings hold important implications for hydrological modelling and management in China's Loess Plateau and other arid and semi-arid regions.

Keywords: artificial neural network, climatic factors, Loess Plateau, state-space model, streamflow, time series analysis

1. Introduction

Water scarcity is one of the critical problems worldwide and receives increasing concern. Especially in the arid and semiarid regions, natural water resources are limited and experience severe stress exerted by expanding population and climate change (Wheater et al., 2008). Groundwater usually becomes one of the main water sources owing to its reliability (UNESCO, 2016). However, groundwater in most cases is hardly renewable and its over-extraction tends to cause a chain of environmental problems, such as land subsidence, seawater intrusion and salinization. Although highly variable in both time and space, streamflow usually dominates the allocation of renewable water resources and inextricably affects local ecosystems and catchment morphology (Lange and Leibundgut, 2003; Srivastava et al., 2006). It is therefore of utmost importance to accurately estimate or model streamflow time series at different spatial and temporal scales (e.g., Noori and Kalin, 2016; Zhu et al., 2016).

ISSN: 1726-2135 print/1684-8799 online © 2020 ISEIS All rights reserved. doi:10.3808/jei.202000440

The current methods for streamflow estimation are categorized into two groups in general (Besaw et al., 2010; Bourdin et al., 2012). One is characterized as process-based, aiming to simulate the functions and processes of the hydrological system, such as the Soil and Water Assessment Tool (SWAT) (Gassman et al., 2007) and Hydrologic Simulation Program-Fortran (HSPF) (Bicknell et al., 1996). However, they require massive input data covering diverse aspects of the system, as well as high computing power and capability for data processing, which are usually difficult to satisfy especially at the regional scale (Kokkonen and Jakeman, 2001; Srivastava et al., 2006). The other group consists of empirical data-driven models (Bourdin et al., 2012) that directly relates the impact factors such as climate and topography to streamflow without consideration of underlying physical and hydrological processes. Requiring no elevated computing power, these methods are easy to apply and able to derive fairly accurate estimates in a short time (Lauzon et al., 2006).

Within the group of data-driven methods, multiple linear regression (MLR), artificial neural network (ANN), and the Box-Jenkins time series model of autoregressive moving average (ARMA) are most commonly used (Wang et al., 2008; Besaw et al., 2010; Noori and Kalin, 2016). The first approach, i.e., MLR, assumes linear relationships between input

^{*} Corresponding author. Tel.: +(86) 10-58807473; fax: +(86) 58806955 *E-mail address*: yang.yang@bnu.edu.cn (Y. Yang).

and output parameters and usually attempts to establish the regression equation via ordinary or weighted least squares method (e.g., Schilling and Wolter, 2005; Mohamoud, 2008). Nevertheless, the relationships between streamflow and possible impact factors can be quite complex and may not be sufficiently described by a single linear function. As an alternative, the ANN modeling inspired by the neural networks of human brain has been introduced to simulate streamflow in the 1990s (Bourdin et al., 2012). Compared to MLR, it does not assume any particular model structure, i.e., linear or non-linear or a combination of both (Ghorbani et al., 2018). The input and output variables are actually connected via one or more hidden layers (Aleksander and Morton, 1990): the input variables are weighed and sent to the hidden layer(s) through, but not limited to non-linear transfer functions, which are weighed again and passed to the output variables in the same manner. In view of these merits, various forms of ANN models have been applied in streamflow estimation (e.g., Karunanithi et al., 1994; Chen et al., 2008; Makwana and Tiwari, 2014).

It is assumed in both traditional MLR and ANN that the streamflow and all the impact factors involved are independent over time, which is rarely the case (Yue et al., 2002). The ARMA approach, in contrast, rather takes advantage of the temporal dependence and derives estimation models based on the autocorrelation of the subject variable (Box and Jenkins, 1970). However, many previous studies have concluded that the ANN method generally performs better than the univariate ARMA modeling in streamflow estimation (e.g., Hsu et al., 1995; Jain and Kumar, 2007), even when the latter is enhanced by incorporating periodical fluctuations (ARIMA) or further by including the impact of possible factors (ARI-MAX) (Castellano-M éndez, 2004). The probable reason is that the prerequisite of dataset stationarity behind ARMA and related models is difficult to satisfy in reality.

To cope with these disadvantages of ARMA, a similar tool named the autoregressive state-space (STATE) model is recommended. This model, in the scenario of stationary data series and univariate analysis, is equivalent to the ARMA model (Shumway and Stoffer, 2011). In contrast to the latter considering model error only, the STATE model assumes that the measurement vector is a linear transform of the true state vector with a white noise and takes the measurement uncertainty into account in addition. An expectation-maximization (EM) algorithm (Shumway and Stoffer, 1982) with a filtering and smoothing procedure is therefore embedded in the STATE approach to deal with these uncertainties, which in the meantime allows to analyze nonstationary data series, to estimate missing values as well as to forecast values outside the measurement domain (Morkoc et al., 1985; Nielsen and Wendroth, 2003; Shumway and Stoffer, 2011). Originally developed by Kalman (1960) and Kalman and Bucy (1961), this method was initially used for filtering noise from signal in aerospace-related and economic research, yet has rarely been employed in hydrology to date. It remains unclear how it performs relative to the traditional datadriven methods such as MLR and ANN in estimating streamflow time series.

The semi-arid Chinese Loess Plateau (CLP) has long been

characterized by thick loess distribution, serious water scarcity and soil erosion (Jia et al., 2015). To manage this situation, numerous soil and water conservation practices have been implemented since the 1950s (Chen et al., 2007), among which the "Grain-for-Green" Project initiated around 2000 was the largest one and turned a great many of sloping croplands into forestlands and grasslands. These conservation measures directly affect water distribution and modify the water regime of CLP (Miao et al., 2010; Liang et al., 2015), which presents a great challenge for streamflow simulation using either process-based or data-driven techniques (Yaseen et al., 2019).

With such premise, the goal was to compare the performances of MLR, ANN, and STATE methods for annual streamflow estimation in CLP. Typical variables used for streamflow modeling include precipitation and temperature or evapotranspiration, depending on the availability of data (Wu and Chau, 2011). In the current study, annual data of streamflow (Q), precipitation (P), and potential evapotranspiration (PET) during 1961 ~ 2013 were collected or derived for 15 catchments across CLP. The specific objectives were to 1) compare the estimation results of annual Q series obtained with MLR, ANN, and STATE approaches, and 2) examine the inter-annual trends of streamflow Q in CLP.

2. Materials and Methods

2.1. Study Area and Datasets

The study was conducted on the well-recognized CLP, which was dominated by the temperate semi-arid continental monsoon climate. The mean annual temperature ranges from 3.6 °C in the northwest to 14.3 °C in the southeast. The mean annual precipitation, similarly, gradually increases from ~ 150 mm in the northwest to 800 mm in the southeast, with major occurrence between July and September (Shi and Shao, 2000). The soil is mainly silt loam-textured loess.

Fifteen catchments were selected across CLP (Figure 1), which range from 1,121 to 106,498 km² in area and together cover 37.5% of the entire CLP. Annual streamflow data during 1961 ~ 2013 were collected from the Yellow River Conservancy Commission. However, the streamflow data were not available for one or two years at 5 catchments and for up to 6 years at the Qin catchment. Daily precipitation, minimum temperature and maximum temperature observed at 2,400 meteorological stations across China [China Meteorological Data Sharing Service System (http://cdc.nmic.cn/home.do)] were interpolated for a national gridded database during the study period, from which the corresponding precipitation and temperature data were derived for all the 15 catchments. Figure 1 displays the total 146 meteorological stations located in the study area. Except for the relatively small catchments of Gushan and Tuwei, such stations were distributed in each catchment, with a maximum number of 68 at the Wei catchment. The potential evapotranspiration was then calculated using the Hargreaves-Samani Equations requiring only temperature data and coordinates for each catchment (Hargreaves and Samani, 1985). Based on the data series above, annual streamflow (Q), precipitation (P), and potential evapotranspiration (PET) during 1961 ~ 2013 were acquired.



Figure 1. Map showing locations, hydrological gauging stations and meteorological stations of 15 catchments on the Loess Plateau. These catchments are numbered following the sequence from upstream to downstream and secondly from small stream order to large order.

2.2. Streamflow Estimation

Using P and PET, three approaches, i.e., MLR, ANN, and STATE, were adopted to estimate the annual series of Q. To avoid the numerical problem that might occur when the magnitudes of the variables differ by an order or more, a normalization procedure was performed for each time series x_i in advance (Nielsen and Wendroth, 2003):

$$x_{i}^{'} = \frac{x_{i} - (\overline{x} - 2\sigma_{x})}{4\sigma_{x}}$$
(1)

in which \bar{x} and σ_x refer to the mean and standard deviation of x_i , respectively.

2.2.1. Multiple Linear Regression (MLR)

The MLR models were generated for Q with a stepwise procedure, in which the significance levels for F-to-enter and F-to-remove were set at 0.05 and 0.01, respectively. The analysis was performed using SPSS 16.0.

2.2.2. Artificial Neural Network (ANN)

A typical ANN model consists of an input layer carrying independent variable(s), an output layer for dependent variable(s) and one or more hidden layers with varying nodes that connect the input and output layers. To estimate Q here, the feed-forward model with mainly nonlinear processing elements was selected. The initial weights in each layer were randomly assigned in each model run, resulting in varying estimations. The optimal ANN model with the smallest square error was selected from 100 models, i.e., performing ANN modeling for 100 times, which was believed to provide a reasonably reliable result (data not shown). Other than MLR, the ANN approach is not capable of selecting input variables. But it may not be appropriate to simply treat both P and PET as input variables, because the addition of PET to P tends to increase the complexity of network and the risk of overfitting (Anctil et al., 2004; Toth and Brath, 2007). Therefore, all the three combinations of input variables, i.e., P only, PET only, and P together with PET, were considered, and the optimal model providing the best estimation was selected as the final one. The ANN modeling was accomplished using the Neural Network Toolbox built in MatLab 2015a.

2.2.3. State-Space (STATE) Model

An autoregressive STATE model consists of a state equation and an observation equation (Shumway and Stoffer, 1982; Shumway, 1988). For a time series x_i , the commonly used firstorder state equation describes the state vector Z_i at time *i* with respect to the state Z_{i-1} at previous time *i* - 1:

$$Z_i = \Phi Z_{i-1} + \omega_i \tag{2}$$

in which Φ is the transition matrix and ω_i the uncorrelated model error. Considering the measurement uncertainty in addition to the model error, the observation model relates the observed vector X_i to the true state vector Z_i through a measurement matrix M_i , and an uncorrelated measurement error matrix v_i :

$$X_i = M_i Z_i + v_i \tag{3}$$

As the two matrices of Φ and M_i , and the distributions of the two error matrices are unknown, the commonly-used least square method is not applicable for model estimation. Instead, the likelihood based inference called Kalman filter (Kalman, 1960) allows the construction of the likelihood function associated with a STATE model and was employed to solve the models together with an iterative EM algorithm which terminated at the relative convergence limit of 0.005 (Shumway and Stoffer, 1982). Similar with the ANN modeling, the optimal STATE model was selected from the models considering three input variable combinations of P and PET.

All the variables were normalized to the same order of magnitude prior to the STATE modeling using Equation (1). In this case, the transition coefficient before each input vari- able describes its contribution to the output (estimate) in any STATE model. But it would not be sufficient when comparing the contributions of each variable in different models. The relative contribution of a particular variable was therefore calculated by dividing its transition coefficient by the sum of the coefficients in each model (Yang and Wendroth, 2014).

2.2.4. Data Validation

The extensive soil and water conservation practices, especially during the "Grain-for-Green" Project, tend to substantially modify the relationships between streamflow and climatic factors (Bi et al., 2009; Han et al., 2019). Under such circumstances, the traditional cross-validation scheme that adopts early half or more of the data series for calibration and later half or less for validation may lead to unsatisfactory results. Therefore, a leave-one-out procedure consisting of two scenarios was applied to cross validate the estimations of streamflow Q at the 15 catchments during 1961 ~ 2013. Scenario 1 eliminated the Q observations made in the even years and Scenario 2, on the contrary, removed the Q observations in the odd years. In each Scenario, the Q observations eliminated as planned were afterwards estimated with the corresponding remaining Q observations, together with P, PET or both.

2.2.5. Evaluation Criteria

The coefficient of determination (R^2) and the adjusted $R^2(R_{adj}^2)$ considering the number of regression variables (k) are usually employed to evaluate the performances of different estimation models, which are defined as:

$$R^{2} = \left[\sum_{i} (x_{i}^{*} - \overline{x_{i}})^{2}\right] / \left[\sum_{i} (x_{i} - \overline{x_{i}})^{2}\right]$$
(4)

$$R_{adj}^{2} = 1 - (1 - R^{2})(\frac{n - 1}{n - k - 1})$$
(5)

where x_i and x_i^* refer to the measurement and estimation at time *i*, respectively, x_i is the mean of measurements x_i , and *n* denotes the number of measurements.

Regarding hydrological models, a combination of absolute value error statistics, normalized goodness-of-fit statistics and graphical results are widely recommended as evaluation criteria (Ritter and Muñoz-Carpena, 2013). Therefore in addition to R_{adj}^2 , the root mean square error (*RMSE*), and Akaike information criterion (*AIC*) were also calculated, and the estimates were plotted against measurements to visually illustrate the model performance. The value of *RMSE* is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i}(x_{i} - x_{i}^{*})^{2}}$$
 (6)

The *AIC* balances between the fitting goodness and the simplicity of the regression model, and is reliable for model comparison and selection (Akaike, 1969):

$$AIC = \ln[\frac{1}{n}\sum_{i}(x_{i} - x_{i}^{*})^{2}] + \frac{2k}{n}$$
(7)

For a finite time series, e.g., k/n < 40, the corrected *AIC*, *AIC*_c, is suggested (Nielsen and Wendroth, 2003):

$$AIC_{c} = \ln[\frac{1}{n}\sum_{i}(x_{i} - x_{i}^{*})^{2}] + \frac{(n+k)}{(n-k-2)}$$
(8)

For either AIC or AIC_c , the smaller the value, the better the estimation quality.

2.3. Trend Analysis

The Mann-Kendall (MK) test is one of the most widelyused trend analysis tools for hydrological time series (Hamed, 2008). This non-parametric method is grounded on the ranks, rather than the actual values, of the measurements, and is therefore less sensitive to the outliers and does not require normal distribution (Yue and Pilon, 2004), which is perfectly suitable for the hydrological time series typically manifesting skewed distributions and data missing (Hirsch and Slack, 1984). The inter-annual trend of annual streamflow Q was examined by the MK test (Mann, 1945; Kendall, 1975; Hirsch et al., 1982) for each catchment. For a time series x_i with a total of n measurements, the statistic *S* of Kentall's tau is defined as:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}(x_j - x_i)$$
(9)

where the sign function $sgn(x_j - x_i)$ depends on the comparison between x_i and x_j :

$$sgn(x_{j} - x_{i}) = \begin{cases} 1 & x_{i} < x_{j} \\ 0 & x_{i} = x_{j} \\ -1 & x_{i} > x_{j} \end{cases}$$
(10)

The statistic Z determines whether the trend analyzed is upward or downward, and is defined as:

$$Z = \begin{cases} \frac{S-1}{\sqrt{\operatorname{var}(S)}} & S > 0\\ 0 & S = 0\\ \frac{S+1}{\sqrt{\operatorname{var}(S)}} & S < 0 \end{cases}$$
(11)

in which the variance of S is calculated as:

$$\operatorname{var}(S) = \frac{n(n-1)(2n+5) - \sum_{p=1}^{q} t_p(t_p-1)(2t_p+5)}{18}$$
(12)

here, q refers to the number of tied groups and t_p the pth group. The slope of the trend, if present, is calculated as the median slope of the adjacent observations:

$$slope = median(\frac{y_j - y_i}{j - i})$$
(13)

von Storch (1995) demonstrated that significant autocorrelations in the data series tend to increase the probability of significant trends detected in the MK test and thereby developed a pre-whitening procedure to remove the lag-one autocorrelation component from the time series. Hamed and Rao (1998) proposed another method for pre-whitening, in which the effective sample size (ESS) was computed via an empirical formula and the MK statistic *S* was modified accordingly to compensate the impact of autocorrelation on the variance. These two essential techniques, however, do not deal with the interaction between the trend and the autocorrelations, which tends to decrease the magnitude of existing trend. Therefore in the current study, the trend-free pre-whitening (TFPW) procedure developed by Yue et al. (2002), which removes the trend before ascertaining the magnitude of auto-correlation, was performed prior to the MK trend analysis for the Q series at each catchment.

3. Results

3.1. Descriptive Statistics

The mean and standard deviation (STD) of P, PET, and Q at each catchment during 1961 ~ 2013 are presented in Table 1. The mean P ranged from 379.0 to 716.4 mm, with the STD falling between 87.3 and 134.9 mm. The mean PET was much larger than P, ranging from 956.0 to 1163.3 mm, with a narrower STD range between 46.0 and 55.4 mm. There was no apparent spatial pattern in Q among the catchments studied. The maximum mean Q of 127.5 mm was observed at the Yiluo catchment with a STD of 82.2 mm and the minimum of 22.1 mm was obtained at the Fen catchment with a STD of 17.1 mm, both of which are located in the east of CLP (Figure 1).

According to the Pearson's correlation analysis, P was significantly correlated with Q at the level of 0.01, except for the correlation at the Tuwei catchment at the level of 0.05. The corresponding correlation coefficients were all positive and ranged from 0.308 at the Tuwei catchment to 0.832 at the Wei catchment. On the contrary, the correlations between PET and Q were all negative. Except for the Tuwei, Dali, and Qingjian catchments, they were significant (p < 0.05) at the rest 12 catchments. The corresponding correlation coefficients ranged from -0.301 at the Wuding catchment to -0.618 at the Xinshui catchment.

3.2. Streamflow Estimations Based on All Observations

3.2.1. Multiple Linear Regression (MLR)

Table 2 presents the MLR models for Q at each catchment. For 11 out of the 15 catchments, less than half of the total Q variance was explained with the MLR models, as the corresponding R_{adj}^2 were smaller than 0.5. Especially at the Tuwei

Table 1. Mean and Standard Deviation (STD) of Annual Precipitation (P), Potential Evapotranspiration (PET) and Streamflow (Q) at Each Catchment during 1961 ~ 2013

ID	Catchment	Hydrological Station	Area	P (mm)		PET (mm)		Q (mm)		
			(km ²)	Mean	STD	Mean	STD	Mean	STD	Year(s) of Missing Data
1	Gushan	Gaoshiya	1263	425.8	112.9	1030.8	52.5	48.4	39.7	
2	Kuye	Wenjiachuan	8645	399.7	103.4	998.8	50.1	57.8	32.6	
3	Tuwei	Gaojiachuan	3253	405.1	106.6	1006.6	50.5	95.9	27.8	
4	Jialu	Shenjiawan	1121	446.7	104.7	1055.5	52.7	48.7	30.7	1968
5	Qiushui	Linjiaping	1873	442.8	98.6	1163.3	55.4	35.1	26.7	2007, 2008
6	Sanchuan	Houdacheng	4102	553.1	108.2	956.0	50.5	51.3	23.3	2007, 2008
7	Dali	Suide	3893	448.9	97.4	1063.8	51.2	35.4	12.3	
8	Wuding	Baijiachuan	29662	379.0	87.3	1040.4	48.8	36.2	10.7	
9	Qingjian	Yanchuan	3468	506.8	104.1	1034.2	50.4	39.2	15.8	2007, 2008
10	Xinshui	Daning	3992	575.7	104.8	987.0	52.5	29.0	19.0	
11	Yan	Ganguyi	5891	541.9	109.2	1094.0	53.8	34.2	13.2	1990
12	Fen	Hejin	38728	512.0	92.4	1052.9	49.4	22.1	17.1	
13	Wei	Huaxian	106498	569.5	90.2	994.1	49.0	61.0	31.1	
14	Yiluo	Heishiguan	18563	716.4	134.9	1049.8	47.7	127.5	82.2	
15	Qin	Wulongkou	9245	602.8	110.2	1023.8	46.0	86.7	61.7	2001 ~ 2006

*All the catchments are numbered following the sequence from upstream to downstream and secondly from small stream order to large order

ID	Catchment	Model	RMSE	R^2_{adj}	AICc	
1	Gushan	Q = 0.603P + 0.198	0.199	0.352	-2.146	
2	Kuye	Q = 0.540P + 0.230	0.210	0.277	-2.037	
3	Tuwei	Q = 0.308P + 0.346	0.238	0.077	-1.792	
4	Jialu	Q = 0.503P + 0.249	0.216	0.238	-1.982	
5	Qiushui	Q = 0.522P + 0.239	0.213	0.257	-2.007	
6	Sanchuan	Q = 0.402P - 0.317PET + 0.458	0.193	0.381	-2.165	
7	Dali	Q = 0.590P + 0.205	0.202	0.335	-2.121	
8	Wuding	Q = 0.523P + 0.239	0.213	0.259	-2.012	
9	Qingjian	Q = 0.486P + 0.257	0.219	0.220	-1.958	
10	Xinshui	Q = 0.480P - 0.329PET + 0.424	0.172	0.510	-2.402	
11	Yan	Q = 0.593P + 0.203	0.201	0.339	-2.125	
12	Fen	Q = 0.636P + 0.182	0.193	0.393	-2.211	
13	Wei	Q = 0.832P + 0.084	0.139	0.687	-2.873	
14	Yiluo	Q = 0.699P - 0.209PET + 0.255	0.145	0.652	-2.744	
15	Qin	Q = 0.772P + 0.114	0.160	0.582	-2.576	

Table 2. Multiple Linear Regression (MLR) Models for Annual Streamflow (Q) at 15 Catchments during 1961 ~ 2013 based on Annual Precipitation (P) and Potential Evapotranspiration (PET), together with the Respective *RMSE*, R_{adi}^2 and *AICc*

catchment, R_{adj}^2 was as low as 0.077, indicating that the corresponding model accounted for only 7.7% of Q variance. The mean R_{adj}^2 averaging over the 15 catchments was 0.371, and the mean *RMSE* and *AIC_c* were 0.194 and -2.210, respectively. Using a stepwise procedure, P was included in all the 15 MLR models. But PET was only employed for the models at 3 catchments, i.e., Sanchuan, Xinshui and Yiluo, and the corresponding coefficients reflecting the relative contributions of PET to Q were all smaller than those of P. It is implied in these models that the variance of Q was mainly explained by the annual precipitation at most catchments. In other words, the magnitude of P, rather than PET, largely determined the size of streamflow.

3.2.2. Artificial Neural Network (ANN)

The optimal ANN model was selected according to AIC_c at each catchment (Table 3). Obviously, both P and PET were included in each optimal model, which performed much better than the corresponding MLR equation, as indicated in the consistently higher R_{adj}^2 , and smaller *RMSE* and AIC_c . Especially at the Tuwei catchment where the time series of Q was most poorly estimated with the MLR method, R_{adj}^2 was improved by 0.767 using the ANN modeling, suggesting an additional 76.7% of Q variance explained by the ANN model than by the MLR equation. Averaging over all the 15 catchments, the mean R_{adj}^2 , *RMSE*, and AIC_c were 0.877, 0.083 and -3.904, respectively.

3.2.3. State-Space (STATE) Approach

The optimal STATE models were selected with the same scheme used in the ANN modeling (Table 4). Both P and PET were adopted in the optimal models for 11 out of the total 15 catchments. Whereas for the rest 4 catchments, the Q estimations depended on P or PET only. In general, Q was better estimated with the STATE models than using MLR or ANN modeing. The mean R_{adj}^2 was as high as 0.891, suggesting that the STATE models were able to explain nearly 90% of the total Q variance on average. The three STATE models for the Fen, Wei,

and Yiluo catchments, furthermore, revealed R_{adj}^2 values even greater than 0.99, accounting for almost all the Q variance. The corresponding *RMSE* and *AIC_c* were lower than those of the MLR and ANN models, with their means of 0.065 and -4.979, respectively. However, at the five catchments of Gushan, Qiushui, Qingjian, Xinshui, and Yan, the ANN models seemed better than the STATE ones, as indicated by the relatively higher R_{adj}^2 , together with lower *RMSE* and *AIC_c* (Table 3, Table 4).

Table 3. Optimal Artificial Neural Network (ANN) Models for Annual Streamflow (Q) at 15 Catchments during 1961 ~ 2013 based on Annual Precipitation (P) and Potential Evapotranspiration (PET), together with the Respective *RMSE*, R_{adi}^2 and *AICc*

ID	Catchment	Factor(s)	RMSE	\mathbf{R}^2_{adj}	AICc
1	Gushan	P, PET	0.075	0.907	-4.060
2	Kuye	P, PET	0.097	0.845	-3.551
3	Tuwei	P, PET	0.097	0.844	-3.548
4	Jialu	P, PET	0.085	0.879	-3.801
5	Qiushui	P, PET	0.067	0.924	-4.265
6	Sanchuan	P, PET	0.075	0.907	-4.058
7	Dali	P, PET	0.122	0.752	-3.086
8	Wuding	P, PET	0.108	0.807	-3.336
9	Qingjian	P, PET	0.120	0.759	-3.110
10	Xinshui	P, PET	0.056	0.947	-4.629
11	Yan	P, PET	0.082	0.887	-3.867
12	Fen	P, PET	0.083	0.884	-3.847
13	Wei	P, PET	0.065	0.931	-4.357
14	Yiluo	P, PET	0.057	0.946	-4.613
15	Qin	P, PET	0.062	0.936	-4.432

3.2.4. Model Comparison

According to the evaluation criteria of *RMSE*, R_{adj}^2 and *AIC_c*, the Q series was generally better estimated by the STA-TE model than by the MLR or ANN model (Tables 2 ~ 4). To visually compare the performances of these three methods, the

resulting Q estimates were plotted against the observations at each catchment (Figure 2). Apparently, the scatters of the MLR model were most divergent from the 1:1 line. Especially for large Q observations located on the right end of each sub- plot, Q was generally underestimated. By comparison, the ANN modeling with the involvement of nonlinear processing elements (Aleksander and Morton, 1990) provided a better fit between the scatters and the 1:1 line, suggesting more accurate estimations. It is implied that the overall relationships of annual streamflow with the climatic factors of P and PET were prone to be nonlinear, rather than linear, during the study period. This deduction is to some extent consistent with Liang et al. (2015), which employed a Budyko framework and concluded highly nonlinear responses of streamflow discharge to P and PET in CLP during recent decades.

It is worth noting that the ANN method performed especially well for the large Q observations, but provided rather biased estimates for the small ones. As displayed in Figure 2, the corresponding red triangles were the closest to the 1:1 line on the right end of each subplot but rather more divergent from the 1:1 line on the left end. As a result, ANN surpassed both MLR and STATE approaches in estimating Q at 5 out of the total 15 catchments, i.e., Gushan, Qiushui, Qingjian, Xin-shui, and Yan catchments. For the rest 10 catchments, the STATE approach was the most accurate method for Q esti- mation, as the corresponding scatters were distributed gener- ally closest to the 1:1 line (Figure 2).

Table 4. Optimal State-space (STATE) Models for Annual Streamflow (Q) at 15 Catchments during 1961 ~ 2013 based on Annual Precipitation (P) and Potential Evapotranspiration (PET), together with the Respective *RMSE*, R_{adi}^2 and *AIC*_c

ID	Catchment	Model	RMSE	R^{2}_{adj}	AIC_c	
1	Gushan	$Q_i = 0.729 Q_{i\text{-}1} + 0.245 PET_{i\text{-}1} + w_i$	0.102	0.828	-3.449	
2	Kuye	$Q_i = 0.932 Q_{i\text{-}1} \text{-} 0.247 P_{i\text{-}1} \text{+} 0.256 PET_{i\text{-}1} \text{+} w_i$	0.046	0.964	-5.001	
3	Tuwei	$Q_i = 1.127 Q_{i\text{-}1} - 0.209 P_{i\text{-}1} + w_i$	0.029	0.986	-5.958	
4	Jialu	$Q_i = 1.027 Q_{i\text{-}1} \text{-} 0.294 P_{i\text{-}1} \text{+} 0.231 PET_{i\text{-}1} \text{+} w_i$	0.071	0.915	-4.124	
5	Qiushui	$Q_i = 1.035 Q_{i\text{-}1} - 0.306 P_{i\text{-}1} + 0.232 PET_{i\text{-}1} + w_i$	0.114	0.777	-3.163	
6	Sanchuan	$Q_i = 0.955 Q_{i\text{-}1} - 0.194 P_{i\text{-}1} + 0.216 PET_{i\text{-}1} + w_i$	0.035	0.979	-5.525	
7	Dali	$Q_i = 0.703 Q_{i\text{-}1} \text{-} 0.146 P_{i\text{-}1} \text{+} 0.393 PET_{i\text{-}1} \text{+} w_i$	0.108	0.802	-3.285	
8	Wuding	$Q_i = 1.162 Q_{i\text{-}1} - 0.259 P_{i\text{-}1} + w_i$	0.040	0.974	-5.334	
9	Qingjian	$Q_i = 0.790 Q_{i\text{-}1} + 0.191 PET_{i\text{-}1} + w_i$	0.179	0.467	-2.315	
10	Xinshui	$Q_i = 1.094 Q_{i\text{-}1} - 0.377 P_{i\text{-}1} + 0.262 PET_{i\text{-}1} + w_i$	0.066	0.926	-4.273	
11	Yan	$Q_i = 0.612 Q_{i\text{-}1} \text{-} 0.062 P_{i\text{-}1} \text{+} 0.434 PET_{i\text{-}1} \text{+} w_i$	0.118	0.762	-3.097	
12	Fen	$Q_i = 0.875 Q_{i\text{-}1} \text{-} 0.162 P_{i\text{-}1} \text{+} 0.275 PET_{i\text{-}1} \text{+} w_i$	0.011	0.998	-7.937	
13	Wei	$Q_i = 1.138 Q_{i\text{-}1} \text{-} 0.577 P_{i\text{-}1} \text{+} 0.413 PET_{i\text{-}1} \text{+} w_i$	0.020	0.994	-6.704	
14	Yiluo	$Q_i = 0.162 Q_{i\text{-}1} + 0.547 P_{i\text{-}1} + 0.273 PET_{i\text{-}1} + w_i$	0.008	0.999	-8.527	
15	Qin	$Q_i = 0.721 Q_{i1} - 0.056 P_{i1} + 0.326 PET_{i1} + w_i$	0.027	0.987	-5.997	



Figure 2. Annual streamflow (Q) observations versus estimates made by multiple linear regression (MLR), artificial neural networks (ANN) and state-space (STATE) models at all the 15 catchments studied. Note that Q and the climatic factors of P and PET, were normalized according to Nielsen and Wendroth (2003).

3.3. Cross Validation for Streamflow Estimations

In both validation scenarios, the optimal MLR, ANN, and STATE models were derived for the calibration datasets applying the same strategy in Section 3.2 for the entire time series. With the significance level of 0.05 for F-to-enter in the stepwise procedure, no MLR model was generated for Q estimation at the Tuwei catchment in Scenario 2 eliminating all the odd-year Q observations or at the Qingjian catchment in Scenario 1 removing even-year observations. Regardless of these two cases, the resulting MLR models explained 15.5% ~ 73.6% of the total Q variance for calibration considering both scenarios. However, these models performed rather poorly in estimating Q of the validation years. Negative values of R^2_{adj} were even obtained for validation datasets at 3 catchments in Scenario 1 and up to 8 catchments in Scenario 2, suggesting strong deviations from the primary trends of Q distributions.

Such typical overfitting phenomenon, i.e., excellent estimations in calibration but poor ones in validation, was detected in the ANN modeling for all the catchments in both scenarios. The *RMSE* were all smaller than 0.03 and the R_{adj}^2 were all greater than 0.99 while calibrating, except for the ones at the Tuwei catchment in Scenario 1 with *RMSE* of 0.039 and R_{adj}^2 of 0.976, and at the Jialu catchment also in Scenario 1 with *RMSE* of 0.097 and R_{adj}^2 of 0.851. When applying the calibration models in validation, the resulting R_{adj}^2 were all negative except for a very small positive one of 0.003 at the Wei catchment in Scenario 1. It is then reasonable to draw a conclusion that relying on P and PET only, ANN modeling was not appropriate for annual streamflow estimation in the study region.

Overfitting was also spotted in the STATE modeling. As a typical example, negative R_{adj}^2 were obtained when applying the optimal calibration STATE models to the validation datasets in both scenarios at the Qingjian catchment and in Scenario 2 at the Yan catchment. For the rest 13 catchments, the optimal STATE models accounted for 75.9 ~ 99.9% of total Q variance in calibration and 22.6 ~ 91.1% of variance in validation.

In view of the negative R_{adj}^2 obtained in the validation datasets for almost all the combinations of catchment and scenario, the ANN method was excluded from the following model comparison, which simply categorized the 15 catchments into 3 groups. Group 1 consists of the Qingjian and Yan catchments, at which neither MLR nor STATE could properly simulate Q from P and PET. Both the optimal MLR and STATE calibration models resulted in negative R_{adj}^2 when applied in the validation datasets. Figure 3 plots the Q observations versus MLR and STATE estimates for both calibration and validation in Scenario 2 at the Qingjian catchment, as no significant MLR calibration model was generated in Scenario 1. Apparently, both the MLR and STATE models performed well in calibration as the corresponding scatters were close to the 1:1 line, resulting in R_{adi}^2 as high as 0.723 and 0.990, respectively. In contrast, the scatters in validation were quite biased from the 1:1 line and the corresponding R_{adi}^2 were both negative, i.e., -1.586 for the MLR model and -1.257 for the STATE one.

Group 2 comprises two catchments of Gushan and Wei, and illustrates a tie of model performance. The STATE approach

generated much better calibration models compared to the MLR method, but performed similarly in validation. It is clearly shown in Figures 4a and 4c that the scatters of the STATE approach were closer to the 1:1 lines in contrast to those of the MLR modeling. Nevertheless, such comparison became unobvious in the validation in Scenario 2 (Figures 4b, 4d). The corresponding R_{adj}^2 , *RMSE* and *AIC_c* were similar between these two types of models in validation (data not shown).



Figure 3. Annual streamflow (Q) observations versus estimates made by multiple linear regression (MLR) and state-space (STATE) models for (a) calibration and (b) validation in Scenario 2 at the Qingjian catchment. Note that Q and the climatic factors of P and PET, were normalized according to Nielsen and Wendroth (2003).

All the other 11 catchments fell into Group 3, where the STATE modeling was believed to surpass the MLR method in Q estimation. The corresponding STATE models performed much better not only for calibration but also for validation in both scenarios, accounting for averagely 93.3% of the total Q variance in calibration and 57.6% of variance in validation. In contrast, the mean percentages of the Q variance explained by the MLR models were only 36.8% and 28.0% in calibration and validation, respectively, disregarding the models possessing negative R_{adj}^2 in validation. A typical example at the Tuwei catchment is displayed in Figure 5. The scatters of the MLR models were closer to the 1:1 lines than those of the MLR models, no matter for calibration or validation, in Scenario 1 or Scenario 2.

3.4. Inter-Annual Trends of Streamflow

In view of its superiority in simulating annual streamflow Q, the STATE technique was applied to estimate the missing streamflow data at the Jialu, Qiushui, Sanchuan, Qingjian, Yan, and Qin catchments (Table 1). The resulting completed 6 time series as well as the series measured at the other 9 catchments were then subject to the MK test to explore the inter-annual trends of streamflow in CLP (Table 5).

There is no significant trend in P at any catchment except for a downward one at the Qin catchment (p < 0.05). The corresponding MK slope was -1.85 mm y⁻¹. Regarding PET, nevertheless, 9 out of the 15 catchments possessed significant trends during the study period, all of which were increasing with the slopes between 0.77 and 1.65 mm y⁻¹. Significant downward trends were detected in Q at all the catchments (p < 0.05) and



Figure 4. Annual streamflow (Q) observations versus estimates made by multiple linear regression (MLR) and state-space (STATE) models for calibration and validation in (a, b) Scenario 1 and (c, d) Scenario 2 at the Gushan catchment. Note that Q and the climatic factors of P and PET, were normalized according to Nielsen and Wendroth (2003).



Figure 5. Annual streamflow (Q) observations versus estimates made by multiple linear regression (MLR) and state-space (STATE) models for Q calibration and validation in (a, b) Scenario 1 and (c, d) Scenario 2 at the Tuwei catchments. Note that Q and the climatic factors of P and PET, were normalized according to Nielsen and Wendroth (2003).



Figure 6. Autocorrelograms for annual streamflow (Q) with 95% confidence limits (dash lines) at (a) Qingjian, (b) Gushan and (c) Tuwei catchments.

the corresponding MK slopes ranged from -0.24 mm y⁻¹ at the Qingjian catchment to -1.71 mm y⁻¹ at the Qin catchment. In view of the different catchment areas, the MK slopes acquired for all the catchments were averaged by area. The resulting MK slope for the entire study area was calculated at -0.49 mm y⁻¹. That is to say, the streamflow or the surface water resource has decreased by almost 0.5 mm per year for the 53 years from 1961 to 2013.

The value of MK slope derived is affected by the magnitude of the Q series analyzed. To compare different catchments, the MK slopes expressed in percentage per year (% y⁻¹) were calculated by dividing the MK slope over the corresponding mean. From 1961 to 2013, Q at all the catchments studied were decreased by an average of $1.74\% \text{ y}^{-1}$, ranging from 0.62% to 2.70% per year. The greatest decreasing rate was obtained at the Fen catchment, while the lowest one at the Qingjian catchment.

4. Discussion

In general, the STATE approach surpassed both the MLR and ANN methods and provided the most accurate estimates for annual streamflow Q in the current study. Using the "leave -one-out" validation procedure, the overfitting problem was

ID	Catchment	MK Slope (mm y ⁻¹)			MK Slope (% y ⁻¹)		
		Р	PET	Q	Р	PET	Q
1	Gushan	0.58	0.98*	-1.31*	0.14	0.10	-2.70
2	Kuye	0.60	0.74	-1.35*	0.15	0.07	-2.34
3	Tuwei	0.53	0.77*	-1.50*	0.13	0.08	-1.57
4	Jialu	0.66	0.99*	-1.27*	0.15	0.09	-2.61
5	Qiushui	-0.16	0.91*	-0.75*	-0.04	0.08	-2.14
6	Sanchuan	-0.45	1.24*	-0.85*	-0.08	0.13	-1.65
7	Dali	0.22	0.77*	-0.35*	0.05	0.07	-0.98
8	Wuding	0.48	0.54	-0.46*	0.13	0.05	-1.28
9	Qingjian	-0.28	0.81*	-0.24*	-0.06	0.08	-0.62
10	Xinshui	-1.23	1.65*	-0.67*	-0.21	0.17	-2.32
11	Yan	-0.75	0.89*	-0.33*	-0.14	0.08	-0.95
12	Fen	-1.05	0.76	-0.60*	-0.20	0.07	-2.70
13	Wei	-0.67	0.29	-0.72*	-0.12	0.03	-1.19
14	Yiluo	-1.41	-0.02	-1.29*	-0.20	0.00	-1.01
15	Qin	-1.85*	0.56	-1.71*	-0.31	0.05	-1.97

Table 5. Mann-Kendall (MK) Slopes of Annual Precipitation (P), PotentialEvapotranspiration (PET) and Streamflow (Q) at Each Catchment during 1961 ~ 2013

*: Significant trend at the level of 0.05

detected in the ANN modeling for each catchment. In contrast to MLR, the STATE method performed better for 11 out of the total 15 catchments according to the three evaluation criteria of R_{adj}^2 , *RMSE*, and *AIC_c* as well as the graphical results. The probable reason was that the STATE modeling, rather than the traditional MLR and ANN simulations, accounts for the temporal dependence in the measurements (Poulsen et al., 2003), which is commonly present in the hydrological time series (Beven, 2001). To be specific, a first-order autoregressive state-space modeling, i.e., as used here, is based on the auto-correlation of the output variable and the cross-correlations of the output with the input variables at the first lag (Shumway and Stoffer, 2011).

In accordance with the catchment categorization in the data validation section, Figure 6 depicts the auto-correlograms for Q at the Qingjian catchment from Group 1, Gushan catchment from Group 2, and Tuwei catchment from Group 3. On each auto-correlogram, the 95% confidence limits were also displayed, which were calculated according to Bartlett (1946) assuming approximately normal distribution of the auto-correlation coefficient r and considering the order of the stochastic process, or in other words, the lags for random distribution (Morkoc et al., 1985). It is obvious that for the Qingjian catchment where Q was unsatisfactorily estimated with the STATE method (Table 4, Figure 3), not a single auto-correlation coefficient r was significant at the lag h of 1 year or longer (Figure 6a), suggesting poor temporal structure inherent in the time series of Q. For the catchments within Group 2 where the STATE modeling performed equivalently well with the MLR method, r was found slightly above the 95% confidence limits at the first lag (Figure 6b). In addition, the "hole effect" describing significant r at longer lag h greater than those at shorter h, was observed in the autocorrelogram at the Gushan (Figure 6b) catchment, indicating that the temporal O distribution included one or more oscillating variation components (Nielsen and Wendroth, 2003; Webster and Oliver, 2007) and might not be sufficiently described by a simple MLR or first-order autoregressive STATE model. For the rest 11 catchments in Group 3, the auto-correlation coefficients r at the first lag were all significant and larger than those calculated for the catchments in Groups 1 and 2. As a typical example, r was as high as 0.789 at h of 1 year for the Q series at the Tuwei catchment and stayed significant until the lag h reached 5 years (Figure 6c).



Figure 7. Relative contribution of annual precipitation (P) and potential evapotranspiration (PET) in the optimal STATE model for annual streamflow (Q) estimation at each catchment.

In view of the generally superior performance of the STATE method in simulating streamflow, the transition coefficients of the corresponding optimal STATE model were used to evaluate the influence of the two climatic factors, i.e., P and PET, on the temporal distribution of streamflow Q. Figure 7 concludes the relative contributions of P and PET for the STATE modeling of Q at each catchment, which allowed the impact



Figure 8. Cross correlograms for annual streamflow (Q) versus (a) annual precipitation (P) and (b) annual potential evapotranspiration (PET), together with 95% confidence limits (dash lines) at the Yiluo catchment.

comparison among different catchments. For most catchments, i.e., 14 out of 15, the temporal distribution of Q was dominated by its temporal dependence at the first lag, as the relative contribution of Q in the previous year exceeded 50%. The only exception was observed at the Yiluo catchment, where the temporal behavior of Q was rather affected by P. In addition to a significant autocorrelation coefficient r obtained at the first lag (data not shown), the temporal correlations of Q with P and PET were also considerable (Figure 8). Especially for the one between Q and P, the corresponding cross-correlation coefficient Γ at lag h of 1 year was significant at the level of 0.05 (Figure 8a).

The relative contributions of P and PET to the temporal distribution of Q were almost equal (Figure 7). On one hand, there was no significant difference between the relative contributions of P and PET when taking all the catchments into account. The corresponding means of their contributions were 16.2% and 19.4%, respectively. On the other hand, the temporal pattern of Q mainly relied on P at 7 out of the 15 catchments and on PET at the remaining 8 catchments, regardless of the temporal dependence inherent in the Q series (Figure 7). This finding seems to disagree with the common knowledge that the streamflow in semi-arid regions such as CLP mainly depends on several heavy rainstorms during the rainy season (Love et al., 2010; Camacho Suarez et al., 2015), or in other words, the streamflow in this region is more sensitive to precipitation rather than evapotranspiration (Li et al., 2017). However, in view of the high temporal variability in rainfall events, the time span in between can be very long, and the evapotranspiration as well as its impact on streamflow are expected to be pronounced at a relatively long temporal scale (Lange and Leibundgut, 2003), such as the annual scale here. As an indirect evidence, PET has significantly increased at 9 out of the 15 catchments, whereas P has significantly decreased at only 1 catchment, along with the significant decreasing trends of Q at all the catchments.

During the study period of 1961 ~ 2013, the streamflow has decreased by 0.24 ~ 1.71 mm y⁻¹, resulting in an areaweighted mean of 0.49 mm y⁻¹. These decreasing rates are within the ranges detected by Zhang et al. (2008) and Liang et al. (2015) that were performed on different CLP catchments for slightly divergent periods. Considering domestic use, industry, fishing, and animal husbandry, Feng et al. (2016) estimated that at least 11.9 mm y⁻¹ of water should be acquired from streams to sustain the current socio-economic activities on CLP. The mean streamflow at the catchments investigated ranged from 12.8 mm at the Gushan catchment to 117.8 mm at the Yiluo catchment over the recent 5 years, i.e., 2009 ~ 2013. If the streamflow continues to decline at the current rates, it will deplete in 30 years for more than half of the catchments. In addition, the decrease of streamflow would also possess serious threat to groundwater recharge and discharge (Li et al., 2019), sediment transport (Wang et al., 2016), delta development (Kong et al., 2015), and fish assemblages (Xie et al., 2018).

5. Conclusions

Three data-driven methods, i.e., MLR, ANN, and STATE, were applied in the current study to estimate annual streamflow Q from the climatic factors of P and PET at 15 catchments across CLP. In general, the STATE approach surpassed the other two methods, especially at the 11 catchments categorized to Group 3 in data validation. This comparison is probably caused by the fact that in contrast to the ANN and MLR models assuming temporal independence, the STATE model accounts for the temporal dependence in the Q series as well as the temporal correlations of Q with P and PET. This reasoning is to some extent verified by the autocorrelograms of O manifesting significant autocorrelation coefficients at the first lag of 1 year at most catchments. Calculating the relative contributions of each variable based on the transition coefficients in the STATE models, the temporal distribution of Q mostly relied on its temporal dependence at the first lag of 1 year. Other than that, P and PET exerted almost equivalent impact on the temporal distribution of Q in CLP. The MK analysis was conducted to explore the inter-annual streamflow trends at the 15 catchments, after estimating the missing data at the 6 catchments with the optimal STATE models. The results manifested significant downward trends of Q at all the catchments during 1961 ~ 2013, at a rate ranging between -0.24 and -1.71 mm y^{-1} .

The Loess Plateau has undergone extensive land use changes as well as other soil and water conservation measures during recent decades. These anthropogenic activities are prone to modify the relationships between streamflow and climatic factors, thereby leading to unsatisfactory streamflow estimations made at several catchments, e.g., Gushan and Qingjian, no matter using which modeling technique. It is then of great importance to collect the detailed types and scales of relevant conservation practices and to systematically examine their impact on the water regime of different Loess Plateau catchments. At the long time scale, moreover, these activities may also interact with the local climate, which tends to further complicate the simulations of streamflow and other hydrological components. Future research is required to clarify these intricate interactions and to provide solid basis for sustainable catchment management.

Acknowledgement: The study was funded by the National Natural Science Foundation of China [Grant No. 41571130082 and 41601277] and the State Key Laboratory of Earth Surface Processes and Resources Ecology [Grant No. 2017-ZY-09].

References

- Akaike, H. (1969). Fitting autoregressive models for prediction. Ann. Inst. Stat. Math., 21, 243-247. https://doi.org/10.1007/BF02532251
- Aleksander, I. and Morton., H. (1990). An introduction to neural computing. Chapman and Hall.
- Anctil, F., Perrin, C., and Andr éassian, V. (2004). Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models. *Environ. Model. Software*, 19, 357-368. https://doi.org/10.1016/S1364-8152(03)001 35-X
- Bartlett, M.S. (1946). On the theoretical specification of sampling properties of autocorrelated time series. J. Roy. Stat. Soc., 8, 27-41. https://doi.org/10.2307/2983611
- Besaw, L.E., Rizzo, D.M., Bierman, P.R., and Hackett., W.R. (2010). Advances in ungauged streamflow prediction using artificial neural networks. J. Hydrol., 386, 27-37. https://doi.org/10.1016/j.jhy drol.2010.02.037
- Beven, K.J. (2001). Rainfall-Runoff Modelling: The Primer. John Wiley & Sons.
- Bi, H., Liu, B., Wu, J., Yun, L., Chen, Z., and Cui, Z. (2009). Effects of precipitation and landuse on runoff during the past 50 years in a typical watershed in Loess Plateau, China. *Int. J. Sediment Res.*, 24, 352-364. https://doi.org/10.1016/S1001-6279(10)60009-1
- Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Donigian, A.S., and Johanson, R.C. (1996). Hydrological Simulation Program-Fortran User's Manual for Release 11.
- Bourdin, D.R., Fleming, S.W., and Stull, R.B. (2012). Streamflow modelling: A primer on applications, approaches and challenges. *Atmosphere-Ocean*, 50, 507-536. https://doi.org/10.1080/07055900. 2012.734276
- Box, G.E.P. and Jenkins, G.M. (1970). Time series analysis: forecasting and control. *Holden-Day*.
- Camacho Suarez, V.V., Saraiva Okello, A.M.L., Wenninger, J.W., and Uhlenbrook, S. (2015). Understanding runoff processes in a semiarid environment through isotope and hydrochemical hydrograph separations. *Hydrol. Earth Syst. Sci.*, 19, 4183-4199. https://doi.org/ 10.5194/hess-19-4183-2015
- Castellano-Méndez, M., Gonz ález-Manteiga, W., Febrero-Bande, M., Prada-Sánchez, J.M., and R. Lozano-Calderón. (2004). Modelling of the monthly and daily behavior of the runoff of the Xallas river using Box-Jenkins and neural networks methods. J. Hydrol., 296, 38-58. https://doi.org/10.1016/j.jhydrol.2004.03.011
- Chen, L., Wei, W., Fu, B., and L ü, Y. (2007). Soil and water conservation on the Loess Plateau in China: Review and perspective. Prog.

Phys. Geogr., 31, 389-403. https://doi.org/10.1177/0309133307081 290

- Chen, X., Wu, J., and Hu, Q. (2008). Simulation of climate change impacts on streamflow in the Bosten Lake Basin using an artificial neural network model. J. Hydrol. Eng., 13, 180-183. https://doi. org/10.1061/(ASCE)1084-0699(2008)13:3(180)
- Feng, X., Fu, B., Piao, S., Wang, S., Ciais, P., Zeng, Z., Lü, Y., Zeng, Y., Li, Y., Jiang, X., and B. Wu. (2016). Revegetation in China's Loess Plateau is approaching sustainable water resource limit. *Na. Clim. Change*, 6, 1019-1022. https://doi.org/10.1038/nclimate3092
- Gassman, P.W., Reyes, M., Green, C.H., and Arnold, J.G. (2007). The Soil and Water Assessment Tool: Historical development, applications, and future directions. *Trans. ASABE*, 50, 1211-1250. https: //doi.org/10.13031/2013.23637
- Ghorbani, M.A., Kazempour, R., Chau, K., Shamshirband, S., and Ghazvinei, P.T. (2018). Forecasting pan evaporation with an integrated artificial neural network quantum-behaved particle swarm optimization model: a case study in Talesh, Northern Iran. *Eng. Appl. Comput. Fluid Mech.*, 12, 724-737. https://doi.org/10.1080/ 1994 2060.2018.1517052
- Hamed, K.H. (2008). Trend detection in hydrologic data: the Mann-Kendall trend test under the scaling hypothesis. J. Hydrol., 349, 350-363. https://doi.org/10.1016/j.jhydrol.2007.11.009
- Hamed, K.H. and Rao, A.R. (1998). A modified Mann-Kendall trend test for autocorrelated data. Journal of Hydrology. 204:182-196. https://doi.org/10.1016/S0022-1694(97)00125-X
- Han, J., Gao, J., and Luo, H. (2019). Changes and implications of the relationship between rainfall, runoff and sediment load in the Wuding River basin on the Chinese Loess Plateau. *Catena*, 175, 228-235. https://doi.org/10.1016/j.catena.2018.12.024
- Hargreaves, G.H.A., and Samani, Z. (1985). Reference crop evapotranspiration from temperature. *Appl. Eng. Agric.*, 1, 96-99. https:// doi.org/10.13031/2013.26773
- Hirsch, R.M., and Slack, J.R. (1984). A nonparametric trend test for seasonal data with serial dependence. *Water Resour. Res.*, 20, 727-732. https://doi.org/10.1029/WR020i006p00727
- Hirsch, R.M., Slack, J.R., and Smith, R.A. (1982). Techniques of trend analysis for monthly water quality data. *Water Resour. Res.*, 18, 107-121. https://doi.org/10.1029/WR018i001p00107
- Hsu, K., Gupta, H.V., and Sorroshian, S. (1995). Artificial neural networks modelling of the rainfall-runoff process. *Water Resour. Res.*, 31, 2517-2530. https://doi.org/10.1029/95WR01955
- Jain, A., and Kumar, A.M. (2007). Hybrid neural network models for hydrologic time series forecasting. *Appl. Soft Comput.*, 7, 585-592. https://doi.org/10.1016/j.asoc.2006.03.002
- Jia, X., Shao, M., Zhang, C., and Zhao, C. (2015). Regional temporal persistence of dried soil layer along south-north transect of the Loess Plateau, China. J. Hydrol., 528, 152-160. https://doi.org/10. 1016/j.jhydrol.2015.06.025
- Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. J. Basic Eng., 82D, 35-45. https://doi.org/10.11 15/1.3662552
- Kalman, R.E., and Bucy, R.S. (1961). New results in linear filtering and prediction theory. J. Basic Eng., 83, 95-108. https://doi.org/ 10.1115/1.3658902
- Karunanithi, N., Grenney, W.J., Whitley, D., and Bovee, K. (1994). Neural networks for river flow prediction. J. Comput. Civil Eng., 8, 201-220. https://doi.org/10.1061/(ASCE)0887-3801(1994)8:2(20 1) Kendall, M.G. (1975). Time Series, second edition. Hefner.
- Kokkonen, T.S., and Jakeman, A.J. (2001). A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications. *Water Resour. Res.*, 37, 2345-2352. https://doi.org/10. 1029/2001WR000299
- Kong, D., Miao, C., Borthwick, A.G.L., Duan, Q., Liu, H., Sun, Q., Ye, A., Di, Z., and Gong, W. (2015). Evolution of the Yellow River Delta and its relationship with runoff and sediment load from 1983

to 2011. J. Hydrol., 520, 157-167. https://doi.org/10.1016/j.jhyd rol.2014.09.038

- Lange, J. and Leibundgut, C. (2003). Surface runoff and sediment dynamics in arid and semi-arid regions. In: Simmers, I. (editor), Understanding water in a dry environment: Hydrological processes in arid and semi-arid zones. A A Balkema Publisher, 115-150. https://doi.org/10.1201/9780203971307.ch4
- Lauzon, N., Anctil, F., and Baxter, C.W. (2006). Classification of heterogeneous precipitation fields for the assessment and possible improvement of lumped neural network models for streamflow forecasts. *Hydrol. Earth Syst. Sci. Discuss.*, 3, 201-227. https:// doi.org/10.5194/hessd-3-201-2006
- Li, Z., Coles, A., and Jun, X. (2019). Groundwater and streamflow sources in China's Loess Plateau on catchment scale. *Catena*, 181, 104075. https://doi.org/10.1016/j.catena.2019.104075
- Li, Z., Ning, T., Li, J., and Yang, D., (2017). Spatiotemporal variation in the attribution of streamflow changes in a catchment on China's Loess Plateau. *Catena*, 158, 1-8. https://doi.org/10.1016/j.catena. 2017.06.008
- Liang, W., Bai, D., Wang, F., Fu, B., Yan, J., Wang, S., Yang, Y., Long, D., and Feng, M. (2015). Quantifying the impacts of climate change and ecological restoration on streamflow changes based on a Budyko hydrological model in China's Loess Plateau. *Water Resour. Res.*, 51, 6500-6519. https://doi.org/10.1002/2014WR016 589
- Love, D., Uhlenbrook, S., Corzo-Perez, G., Twomlow, S., and van der Zaag, P. (2010). Rainfall-interception-evaporation-runoff relationships in a semi-arid catchment, northern Limpopo basin, Zimbabwe. *Hydrol. Sci. J.*, 55, 687-703. https://doi.org/10.1080/02626667. 2010.494010
- Makwana, J.J., and Tiwari, M.K. (2014). Intermittent streamflow forecasting and extreme event modelling using wavelet based artificial neural networks. *Water Resour. Manage.*, 28, 4857-4873. https://doi.org/10.1007/s11269-014-0781-1
- Mann, H.B. (1945). Non-parametric test against trend. *Econometrica*, 13, 245-259. https://doi.org/10.2307/1907187
- Miao, C., Ni, J., and Borthwick, A.G.L. (2010). Recent changes of water discharge and sediment load in the Yellow River Basin, China. *Prog. Phys. Geogr.*, 34, 541-561. https://doi.org/10.1177/03091333 10369434
- Mohamoud, Y.M. (2008). Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves. *Hydrol. Sci. J.*, 53, 706-724. https://doi.org/10.1623/hysj. 53.4.706
- Morkoc, F., Biggar, J.W., Miller, R.J. and Nielsen, D.R. (1985). Statistical analysis of sorghum yield: A stochastic approach. *Soil Sci. Soc. Am. J.*, 49, 1342-1348. https://doi.org/10.2136/sssaj1985.0361 5995004900060003x
- Noori, N. and Kalin, L. (2016). Coupling SWAT and ANN models for enhanced daily streamflow prediction. J. Hydrol., 533, 141-151. https://doi.org/10.1016/j.jhydrol.2015.11.050
- Nielsen, D.R., and Wendroth, O. (2003). Spatial and temporal statistics: Sampling field soils and their vegetation. *Catena*.
- Poulsen, T.G., Moldrup, P., Wendroth, O., and D.R. Nielsen. (2003). Estimating saturated hydraulic conductivity and air permeability from soil physical properties using state-space analysis. *Soil Sci.*, 168, 311-320. https://doi.org/10.1097/01.ss.0000070906.55992.75
- Ritter, A., and Muñoz-Carpena, R. (2013). Performance evaluation of hydrological models: Statistical significance for reducing subjecttivity in goodness-of-fit assessments. J. Hydrol., 480, 33-45. https:// doi.org/10.1016/j.jhydrol.2012.12.004
- Schilling, K.E. and Wolter, C.F. (2005). Estimation of streamflow, base flow, and nitrate-nitrogen loads in Iowa using multiple linear regression models. J. Am. Water Resour. Assoc., 41, 1333-1346. https://doi.org/10.1111/j.1752-1688.2005.tb03803.x
- Shi, H., and Shao, M. (2000). Soil and water loss from the Loess

Plateau in China. J. Arid Environ., 45, 9-20. https://doi.org/10. 1006/jare.1999.0618

- Shumway, R.H. (1988). Applied Statistical Time Series Analysis. Prentice Hall.
- Shumway, R.H. and Stoffer, D.S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. J. Time Ser. Anal., 3, 253-264. https://doi.org/10.1111/j.1467-9892.1982.tb0 0349.x
- Shumway, R.H., and Stoffer, D.S. (2011). Time Series Analysis and Its Applications, third edition. Springer. https://doi.org/10.1007/978-1-4419-7865-3
- Srivastava, P., McNair, J.N., and Johnson, T.E. (2006). Comparison of process-based and artificial neural network approaches for streamflow modeling in an agricultural watershed. J. Am. Water Resour. Assoc., 42, 545-563. https://doi.org/10.1111/j.1752-1688.2006. tb04475.x
- Toth, E. and Brath, A. (2007). Multistep ahead streamflow forecasting: Role of calibration data in conceptual and neural network modelling. *Water Resour. Res.*, 43, W11405. https://doi.org/10.10 29/2006WR005383
- United Nations Educational, Scientific and Cultural Organization (UNES-CO). (2016). The United Nations World Water Development Report 2016. http://unesdoc.unesco.org/images/0024/002439/ 243938e.pdf
- von Storch, H. (1995). Misuses of statistical analysis in climate research. In: von Storch, H., and Navarra, A. (Editors.): Analysis of Climate Variability-Applications of Statistical Techniques. Springer-Verlag. 11-26. https://doi.org/10.1007/978-3-662-03167 -4_2
- Wang, S., Fu, B., Piao, S., Lü, Y., Ciais, P., Feng, X., and Wang, Y. (2016). Reduced sediment transport in the Yellow River due to anthropogenic changes. *Nat. Geoscience*, 9, 38-10. https://doi.org/ 10.1038/ngeo2602
- Wang, Y., Chen, S., Yu, P., and Yang, T. (2008). Storm-event rainfallrunoff modelling approach for ungauged sites in Taiwan. *Hydrol. Process.*, 22, 4322-4330. https://doi.org/10.1002/hyp.7019
- Webster, R., and M.A. Oliver. (2007). Geostatistics for Environmental Scientists, second edition. John Wiley & Sons. https://doi. org/10.1002/9780470517277
- Wheater, H., Sorroshian, S., and Sharma, K.D. (2008). Hydrological Modelling in Arid and Semi-Arid Areas. Cambridge University Press. https://doi.org/10.1017/CBO9780511535734
- Wu, C.L. and Chau, K.W. (2011). Rainfall-runoff modeling using artificial neural network coupled with singular spectrum analysis. J. Hydrol., 399, 394-409. https://doi.org/10.1016/j.jhydrol.2011.01. 017
- Xie, J.Y., Tang, W.J., and Yang, Y.H. (2018). Fish assemblage changes over half a century in the Yellow River, China. *Ecology* and Evolution, 8, 4173-4182. https://doi.org/10.1002/ece3.3890
- Yang, Y. and Wendroth, O. (2014). State-space approach to analyze field-scale bromide leaching. *Geoderma*, 217-218, 161-172. https:// doi.org/10.1016/j.geoderma.2013.11.013
- Yaseen, Z.M., Sulaiman, S.O., Deo, R.C., and Chau, K. (2019). An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. J. Hydrol., 569, 387-408. https://doi.org/10.1016/j.jhydrol.2018.11.069
- Yue, S., and Pilon, P. (2004). A comparison of the power of the t test, Mann-Kendall and bootstrap tests for trend detection. *Hydrol., Sci.* J., 49, 21-37. https://doi.org/10.1623/hysj.49.1.21.53996
- Yue, S., Pilon, P., Phinney, B., and Cavadias, G. (2002). The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrol. Process.*, 16, 1807-1829. https://doi.org/10. 1002/ hyp.1095
- Zhu, S., Zhou, J., Ye, L., and Meng, C. (2016). Streamflow estimation by support vector machine coupled with different methods of time series decomposition in the upper reaches of Yangtze River, China. *Environ. Earth Sci.*, 75. https://doi.org/10.1007/s12665-016-5337-7