

## Forecasting daily streamflow values: assessing heuristic models

Sepideh Karimi, Jalal Shiri, Ozgur Kisi and Tongren Xu

### ABSTRACT

Predicting streamflow values accurately is vitally important for hydrology studies. Two heuristic models, namely, gene expression programming (GEP) and support vector machine (SVM) are used and assessed utilizing data from four stations in China. The k-fold testing for local and external data management scenarios are tested extensively. Results indicate that models with inputs of current and one previous day's streamflow records provided the best accuracy. Both the GEP and SVM models can predict accurate streamflow values with respect to the observed records. GEP performed better than the SVM in all k-fold testing stages with lower skewness and standard deviation values for streamflow records. The test accuracy demonstrated high variations for the local and external k-fold case which proved the necessity of k-fold testing or data scanning procedure in daily streamflow prediction. Daily streamflow of downstream stations was also estimated using the data of upstream stations (external k-fold). The best results were obtained by the models trained using the data from the nearest upstream station. In some cases, the accuracy of the external models was found to be comparable to local models. This suggested the use of external models in streamflow prediction in the case of data scarcity.

**Key words** | gene expression programming, local and external k-fold testing, streamflow prediction, support vector machine

#### Sepideh Karimi

#### Jalal Shiri

Water Engineering Department, Faculty of Agriculture,  
University of Tabriz,  
Tabriz,  
Iran

#### Ozgur Kisi

Faculty of Natural Sciences and Engineering,  
Iliia State University,  
Tbilisi,  
Georgia

#### Tongren Xu (corresponding author)

State Key Laboratory of Earth Surface Processes  
and Resource Ecology, School of Natural  
Resources, Faculty of Geographical Science,  
Beijing Normal University,  
No. 19, Xijiekouwai Street, Beijing 100875,  
China  
E-mail: xutr@bnu.edu.cn

### INTRODUCTION

Predicting streamflow values accurately is one of the important issues in water resources system planning, design, operation, and management. It is also an important task for identifying hydrologic drought periods (Chemeda Edossa & Singh Babel 2011), flood control (Sarlak 2008), optimizing hydrologic system or planning for future expansions or reductions (Kisi 2008), determining instream environmental flow (Tennant 1976), and modeling river flow-groundwater flow interactions (Gunduz & Aral 2005) as well as mitigating the negative consequences of extreme flow events and economic use of rivers (Benninga 2015) and modeling suspended sediment load in rivers (Kisi *et al.* 2012).

Traditionally, autoregressive moving average (ARMA) models have been employed for water resources time-

series modeling including streamflow forecasting (Maier & Dandy 1996). The major drawback of univariate time-series approaches in streamflow forecasting is the incorporation of the past flow magnitudes as well as assuming a linear relation between the input-target variables (Kisi 2008).

Alternatively, the use of heuristic data-driven models for streamflow forecasting has been reported by numerous studies. Many studies have confirmed the superiority of data-driven models over traditional ARMA or autoregressive integrated moving average (ARIMA) techniques in forecasting streamflow magnitudes (e.g., Nayak *et al.* 2004; Wang *et al.* 2009; Yarar 2014). In this context, Wang *et al.* (2006), Kisi (2007, 2009), Kagoda *et al.* (2010), Humphrey *et al.* (2016), and Uysal *et al.* (2016) employed artificial neural network (ANN) models for forecasting streamflow. Guven

(2009) applied linear genetic programming (GP) for time-series modeling of daily flow rates. Wu & Chau (2010) applied different heuristic data-driven approaches for modeling monthly streamflow time series and found that a moving average ANN presented the most accurate results. Shiri & Kisi (2010) introduced a wavelet-neuro-fuzzy model to forecast streamflow values and found that when the periodicity component was introduced as an input parameter, the models gave the most accurate results. Guo *et al.* (2011) employed the support vector machine (SVM) in forecasting monthly streamflow values and found that SVM produced more accurate results than ANNs. Kisi *et al.* (2013) compared different heuristic models in simulating surface runoff magnitudes and found the gene expression programming (GEP) approach to be the most accurate method. Liu *et al.* (2014) applied a wavelet-SVM model for daily and monthly streamflow forecasting. Sharma *et al.* (2015) compared a neuro-fuzzy model with a physically based watershed model for streamflow forecasting and concluded that the neuro-fuzzy model was equally comparable to the physical model, especially when rain gauge stations were not adequate. Karimi *et al.* (2016) introduced a wavelet-GEP approach for forecasting streamflow. Badrzadeh *et al.* (2018) applied a wavelet-ANFIS method for predicting intermittent streamflow values.

However, all the mentioned studies have used a single data set assignment approach for introducing the input-output matrixes to the models, where a part of available data is used for training the models, and then the models are tested using the rest of available patterns. This is a common data management scenario in hydrologic time series predictions, where the models are trained and tested using data of the same station (local or at station scenario). Apart from not performing a whole accuracy assessment of the within-station (local) patterns, another important drawback of this procedure is the lack of generalizability of the obtained models, i.e., the models are not assessed outside the training station (Marti *et al.* 2011; Shiri *et al.* 2014a, 2014b, 2014c). Hence, it would be worth attempting to assess the external generalizability of the applied models. The present research aimed at evaluating local (within-station) and external (cross-station) data management scenarios for predicting streamflow rates using GEP and SVM techniques through k-fold testing. Despite the limited applications of k-fold testing approach in hydrological modeling,

this is the first application of this approach in local and external simulation of streamflow values in the literature. The developed external models can be used for predicting the streamflow records without the need of local patterns.

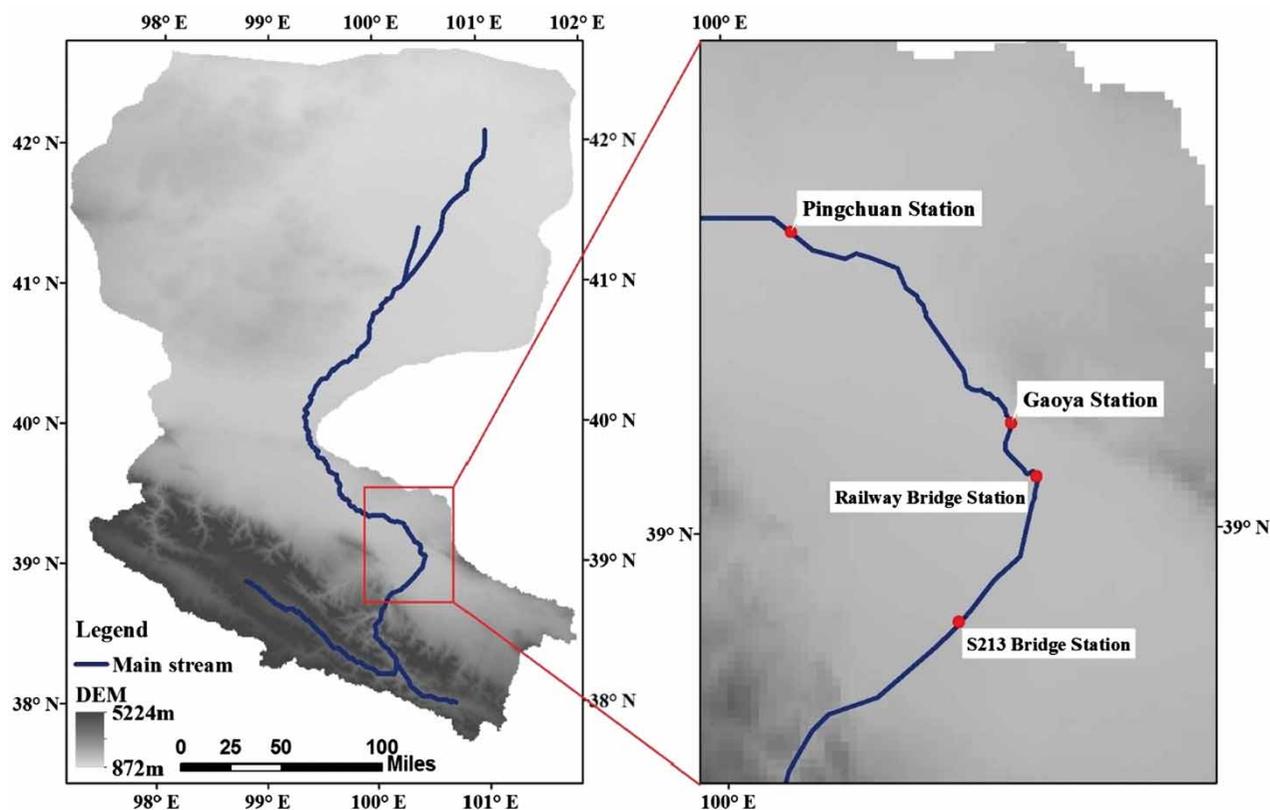
## MATERIALS AND METHODS

### Study area and used data

The Heihe River is the second largest inland river of the People's Republic of China (P.R.C.). It originates from southwest of Qilian Mountain, and runs through three provinces, namely, Qinghai, Gansu, and Inner Mongolia. The total length of the Heihe River is 821 km, and the catchment area is  $1.0 \times 10^4$  km<sup>2</sup> (Figure 1). Four stations, namely, S213 Bridge, Railway Bridge, Gaoya, and Pingchuan were established along the middle reach of the Heihe River to monitor the daily variations of streamflow (Liu *et al.* 2016). The site locations can be found in Figure 1. The daily streamflow data of the four stations collected from 1979 to 2014 (35 years) are used to construct and test the GEP and SVM models. Table 1 summarizes the statistical indices of the applied data set. The streamflow records of Gaoya station present the highest skewness while the others also have considerably high skewness coefficients indicating that the utilized streamflow data do not have a normal (Gaussian) distribution. Data of Railway Bridge and Pingchuan stations show the highest and lowest variations, respectively. The four stations located along the Heihe River, with distinct streamflow records, make it possible to test the local (within-station) and external (cross-station) performances of the two heuristic models.

### Gene expression programming

As a generalization of genetic algorithms (Goldberg 1989), GP (Koza 1992) is especially appropriate where inter-relationships among relevant variables are poorly understood; a theoretical examination is constrained by assumptions and there is a great deal of data in PC readable frames requiring tedious processing. GP-based models utilize a 'parse tree' structure in the search for their solutions



**Figure 1** | The study area: (left) the Heihe River basin and (right) the site locations.

**Table 1** | Statistical parameters of the applied river flow data

Station name	Min m <sup>3</sup> /s	Max m <sup>3</sup> /s	Mean m <sup>3</sup> /s	SD m <sup>3</sup> /s	CV	Skew
S213 Bridge	0.01	361.03	26.24	41.24	1.57	2.80
Railway Bridge	0.03	329.14	18.70	33.34	1.78	3.20
Gaoya	2.26	838.34	34.01	43.17	1.26	7.36
Pingchuan	5.03	274.17	27.84	16.83	0.61	2.62

Note: The min, max, mean, SD, CV, Skew stand for the minimum, maximum, mean, standard deviation, coefficient of variation, and skewness coefficient of the streamflow data.

(Banzhaf et al. 1998). GP is able to derive a set of explicit expressions which rule the subject phenomenon, to describe the inter-relationships between the input-target parameters by using different operators.

GEP is similar to GP, in the way that it selects the best governing formulations based on fitness values and presents genetic variety utilizing a unique or various genetic operators (Ferreira 2006). The most critical advantages of GEP are (Ferreira 2000): (i) the chromosomes are basic elements: linear, compact, relatively small, simple to manipulate

genetically (replicate, mutate, recombine, and so on); (ii) the expression trees are solely the statement of their particular chromosomes; they are elements upon which selection acts, and as indicated by fitness, they are chosen to recreate with adjustment.

The GEP procedure starts with the random generation of the chromosomes of a certain number of individuals or programs (the initial population). The chromosomes are then expressed and the fitness of each program is assessed against a set of fitness cases (training set). The programs are then chosen based on their fitness to reproduce with modification, leaving progeny with new traits (Ferreira 2001). Here, absolute error-based root relative squared error (RRSE) fitness function was used, as advised by Kisi et al. (2013). The second step is selecting the terminal and function sets. Here, the terminal set includes the streamflow records (with different lag times) and the function set consists of basic arithmetic operators along with other functions:  $+$ ,  $-$ ,  $\times$ ,  $\div$ ,  $\sqrt[3]{\phantom{x}}$ ,  $\sqrt{\phantom{x}}$ ,  $\ln$ ,  $e^x$ ,  $x^2$ ,  $x^3$ ,  $\sin x$ ,  $\cos x$ ,  $\text{Arctg}x$ , linked with the addition linking function. The

**Table 2** | General description of the applied GEP models

Number of chromosomes	30	one point recombination rate	0.3
Head size	8	two point recombination rate	0.3
Number of genes	3	gene recombination rate	0.1
Linking function	addition	gene transposition rate	0.1
Fitness function error type	RRSE	insertion sequence transposition rate	0.1
Mutation rate	0.044	root insertion sequence transposition	0.1
Inversion rate	0.1	penalizing tool	parsimony pressure
Chromosomes	<i>the chromosomes of GEP are usually composed of more than one gene of equal length. Each gene codes for a sub-expression tree and the sub-expression trees interact with one another forming a more complex multi-subunit expression tree.</i>		
Head size	<i>determines the complexity of each term in the model</i>		
Mutation	<i>provides the evolution of good solutions for the studied models to virtually all problems</i>		
Inversion	<i>inversion is restricted to the heads of genes</i>		
One point recombination	<i>the parent chromosomes are paired and split up at exactly the same point</i>		
Two point recombination	<i>two parent chromosomes are paired and two points are randomly chosen as crossover points</i>		
Gene recombination	<i>entire genes are exchanged between two parent chromosomes, forming two daughter chromosomes containing genes from both parents</i>		
Gene transposition	<i>an entire gene works as a transposon and transposes itself to the beginning of the chromosome</i>		
IS transposition	<i>short fragments of the genome with a function or terminal in the first position that transpose to the heads of gene except the root</i>		
RIS transposition	<i>short fragments with a function in the first position that transpose to the start position of genes</i>		

genetic operators (Table 2) are used based on the results provided in previous studies (e.g., Kisi et al. 2012, 2013). By using GEP, the parsimony pressure tool was utilized to design the parsimonious solutions. Further details on GEP application for hydrological time series modeling can be found in, for example, Kisi et al. (2012).

### Support vector machine

SVMs are supervised learning models with related learning algorithms that analyze data and recognize patterns, and can be utilized for classification and regression analysis. While the original problem might be expressed in a finite dimensional space, the sets to discriminate are not linearly separable in that space. Consequently, it was suggested that the original finite-dimensional space be mapped into a much higher-dimensional space, probably making the partition less demanding in that space. There are four principal advantages of SVM. First, it has a regularization parameter, which makes the user consider about avoiding over-fitting. Second, it utilizes the

kernel trick, so one can construct an expert knowledge about the issue through designing the kernel. Third, an SVM is determined by a convex optimization problem (no local minima) for which there are accurate methods (e.g., SMO). Finally, it is estimated to a bound on the test error rate, and there is a substantial body of theory behind it which recommends it ought to be a good idea (Vapnik 1995).

The regression-SVM type 1 was applied here since its superiority to other types has been demonstrated in the literature (Shiri et al. 2014c). Using a trial and error process, SVM constants are selected as 8 (capacity) and 0.14 (epsilon). Linear, sigmoid, polynomial, and radial basis kernel functions were compared and it was found that the radial basis kernel (Gamma=0.125) gives the most accurate results. The SVM model is very sensitive to the Gamma parameter. This parameter indicates how far a single training example's influence reaches; low values mean 'far' and high values mean 'close'. The maximum number of iterations was found to be 1,500 (iteratively) and the models were stopped at error magnitudes of 0.005.

### Study flowchart and data splitting

The first step with defining the models' structures would be identifying the most relevant input parameters. In the

present study, auto correlation technique was employed to identify the suitable lags of streamflow time series at the studied stations. Figure 2 shows the partial auto correlation function (PACF) diagrams of the streamflow data at each

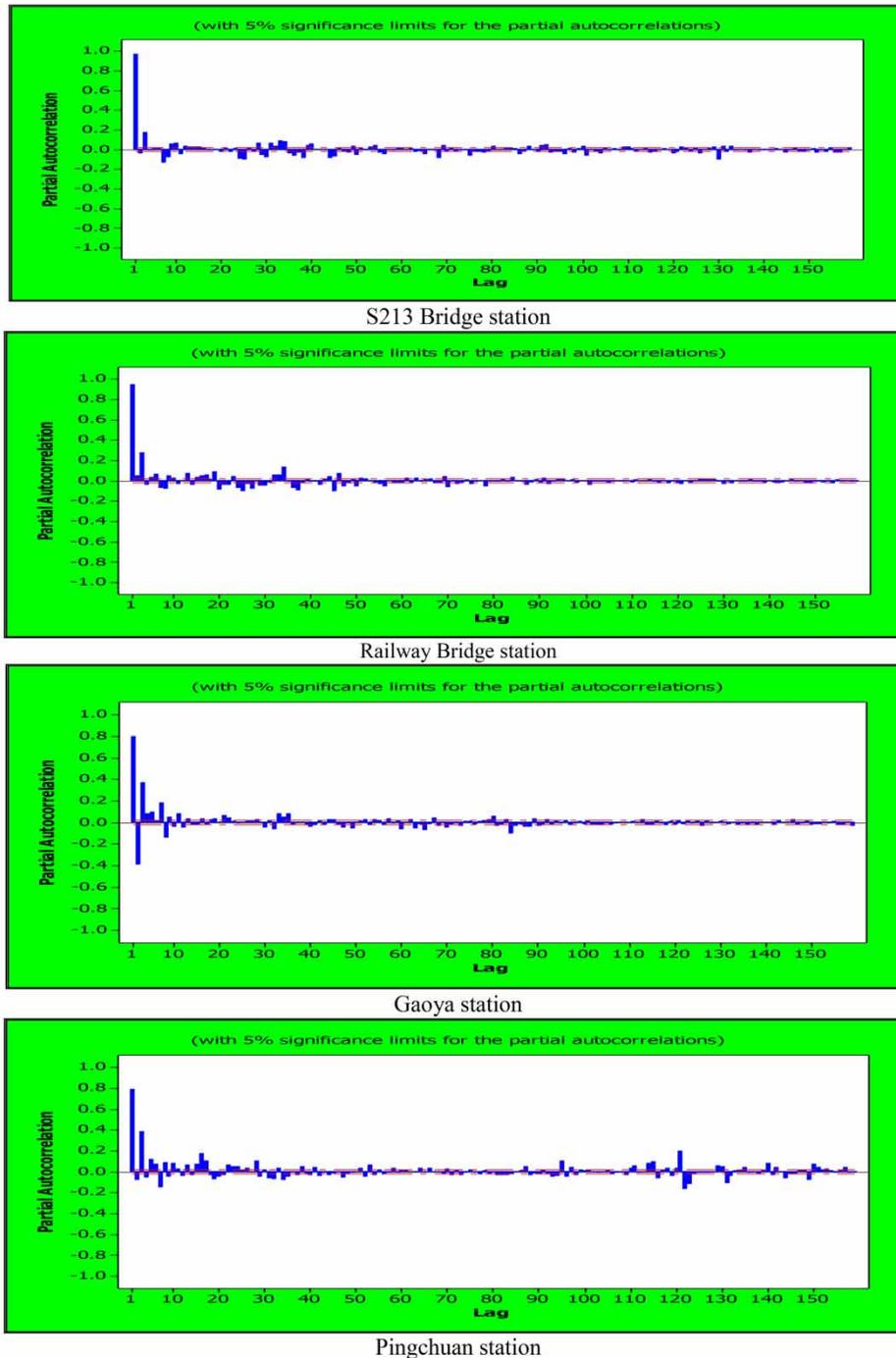


Figure 2 | PACF of streamflow data in the studied stations (from upstream to downstream).

station for the study period. As indicated, the first three time lags have significant correlations in all stations except for Gaoya station. Therefore, three-time lags were utilized for modeling the one-day ahead streamflow values in all four stations. Thus, given that the  $Q_{t+1}$  shows the streamflow values at the next immediate day (as models' target), the following input configurations were defined and evaluated using GEP- and SVM-based models:

Input configuration I:  $Q_t$  (GEP1, SVM1 models)

Input configuration II:  $Q_t, Q_{t-1}$  (GEP2, SVM2 models)

Input configuration III:  $Q_t, Q_{t-1}, Q_{t-2}$  (GEP3, SVM3 models)

Here, the assessment of the GEP and SVM models' performance has been carried out considering spatial as well as local k-fold testing approaches (Shiri et al. 2014a). Accordingly, a local k-fold testing was carried out over each station, using one-year patterns as minimum temporary test phase. Thus, a temporary k-fold testing was employed, leaving in every stage an alternate year for testing, until an entire scan of the series of each station was satisfied. This procedure was repeated for all stations, so a total 840 train-test processes (4 stations  $\times$  2 models  $\times$  3 input configurations  $\times$  35 years = 840) were carried out in local k-fold testing. Although a shorter time period (less than one year) might be assumable in local k-fold testing, this would involve a great deal of operational time and computational costs, as discussed by Marti et al. (2011) and Shiri et al. (2014b).

Additionally, a spatial (external) k-fold testing approach was also carried out using the best input configuration obtained through local modeling. In each step of this approach, the whole patterns (all available data) from upstream station(s) (see Figure 1) were utilized for training, and the models were tested using all data series of the remaining downstream stations. Therefore, in Pingchuan station, GEP and SVM models were at first trained using Gaoya station data, then trained by Railway Bridge station data, then trained by S213 Bridge station data, and finally trained using all patterns from the three mentioned train (upstream) stations. Similar processes were also repeated for Gaoya station (using Railway Bridge, S213 Bridge and pooled patterns). In the case of Railway Bridge, as it has only a unique upstream station, S213 Bridge patterns were used to train

the models. As this study considers four stations and only one input configuration for spatial analysis, the spatial approach consists of eight train-test processes per model.

### Performance criteria

Three statistical indices, namely, the Scatter Index ( $SI$ ), the Variance Accounted For ( $VAF$ ), and the Nash and Sutcliffe coefficient ( $NS$ ), were used for assessing the models performance:

$$SI = \frac{RMSE}{\overline{Q_o}} = \frac{\sqrt{1/n \sum_{i=1}^n (Q_{io} - Q_{ie})^2}}{\overline{Q_o}} \quad (1)$$

$$VAF = \left[ 1 - \frac{Var(Q_{io} - Q_{ie})}{Var(Q_{io})} \right] \times 100 \quad (2)$$

$$NS = 1 - \frac{\sum_{i=1}^n (Q_{io} - Q_{ie})^2}{\sum_{i=1}^n (Q_{io} - \overline{Q_o})^2} \quad (3)$$

where  $Q_{io}$  and  $Q_{ie}$  are the observed and simulated streamflow values, respectively, and  $\overline{Q_o}$  stands for the mean observed streamflow.  $n$  is the number of data patterns. The perfect values of  $SI$ ,  $VAF$ , and  $NS$  are 0, 1, and 1, respectively. The root mean square error ( $RMSE$ ) describes the average extent of errors by giving more weight to large errors. However, it is a site-specific index and would be affected by the magnitudes of the target variables at different stations. Thus, its dimensionless form (called  $SI$ ) can give a good insight for comparing the performances of various models in different sites as the effect of the observed values' magnitude has been ignored through incorporating the mean values of observations in  $SI$  equation. A  $NS$  (Nash & Sutcliffe 1970) coefficient value of 0 ( $NS = 0$ ) shows that the model simulations are as accurate as the mean of the observed data, while its negative values describe that the residual variance is larger than the original data variance. The  $VAF$  is generally used to verify the correctness of a model, by comparing the variances of the real output with the estimated output.

The statistical indices of the local k-fold testing refer to the complete period in each station, i.e., the simulations of each test stage (year) were pooled chronologically together

and the statistical indices were calculated for the complete study period. Regarding the spatial k-fold testing, the criteria correspond to the performance of the complete study period at each station.

## RESULTS AND DISCUSSION

Table 3 sums up the statistical indices of each input configuration averaging the results of the studied stations. Instead of averaging the statistical indicator calculating an average of 35 years, a global prediction vector of the complete period was made, and this vector was utilized for calculating the indicators of each station. From Table 3, it is seen that the predictions of those models relying on input configuration II (i.e., considering  $Q_t$ ,  $Q_{t-1}$  as input parameters) are more accurate than the rest. Hence, input configuration II was selected to perform the external k-fold testing from upstream to downstream stations. Table 3 shows that the local approach provides the best predictions, since it is trained using the streamflow time series of the same stations utilized for testing (using different years for training and testing). However, its generalization ability is limited to similar conditions of the training (and testing) location.

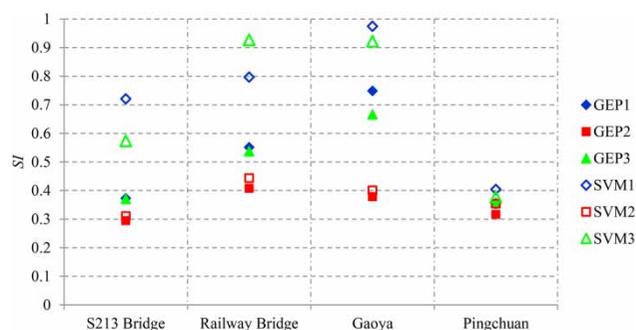
Analyzing the temporal GEP and SVM models relying on input configuration II (comprises the  $Q_t$ ,  $Q_{t-1}$  parameters as inputs) confirms their considerable superiority to the other input configurations. *SI* values have decreased by 19% and 7% for GEP and SVM models, respectively (compared to input configuration I), and by 14% and 4% for GEP and SVM models (compared to input configuration III), respectively. Predictions relying on input configuration I (GEP1 and SVM1) provide considerably

less accurate simulations than those of input configurations II and III. Additionally, considering the *NS* and *VAF* values, the superiority of the GEP2 and SVM2 models is confirmed. Figure 3 illustrates the *SI* values of the temporal GEP and SVM models split up per test station. These values were obtained through building a global prediction matrix of the complete period (35 years) at each station. Again, it is seen that the GEP2 and SVM2 models give the most accurate results at each station, so the external models were fed with input configuration I. The highest accuracy and the lowest variability among three models' results correspond to Pingchuan station followed by S213 Bridge station while Gaoya station has the worst accuracy and the highest variability. These results confirm the statistics given in Table 1, where Gaoya station has the highest skewness coefficient and standard deviation values followed by the Railway Bridge station. On the other hand, Pingchuan station presents the lowest skewness and standard deviation values followed by the S213 Bridge station, as can be seen from Table 1. Nevertheless, *SI* values of the models relying on input configuration II show the lowest variability among stations. Finally, GEP-based predictions also outperform the SVM-based models in all studied cases. Although presenting near *VAF* values, GEP2 increases the accuracy of SVM2 by 7.4% and 5% with respect to *SI* and *NS*, respectively. The reason of the superior accuracy of GEP compared to SVM may be the fact that the GEP method uses evolutionary genetic algorithm in the calibration process and this may provide an advantage in catching local minimum. The performance oscillations among studied sites dictate the need for assessing the applied models' performances through k-fold testing.

**Table 3** | Global average performance parameters of the local GEP and SVM models

Models	GEP			SVM		
	<i>SI</i>	<i>VAF</i>	<i>NS</i>	<i>SI</i>	<i>VAF</i>	<i>NS</i>
Input configuration I	0.507	0.788	0.788	0.724	0.751	0.630
Input configuration II	0.350	0.841	0.842	0.378	0.846	0.804
Input configuration III	0.479	0.815	0.816	0.701	0.796	0.723
External models	0.441	0.878	0.887	0.491	0.861	0.859

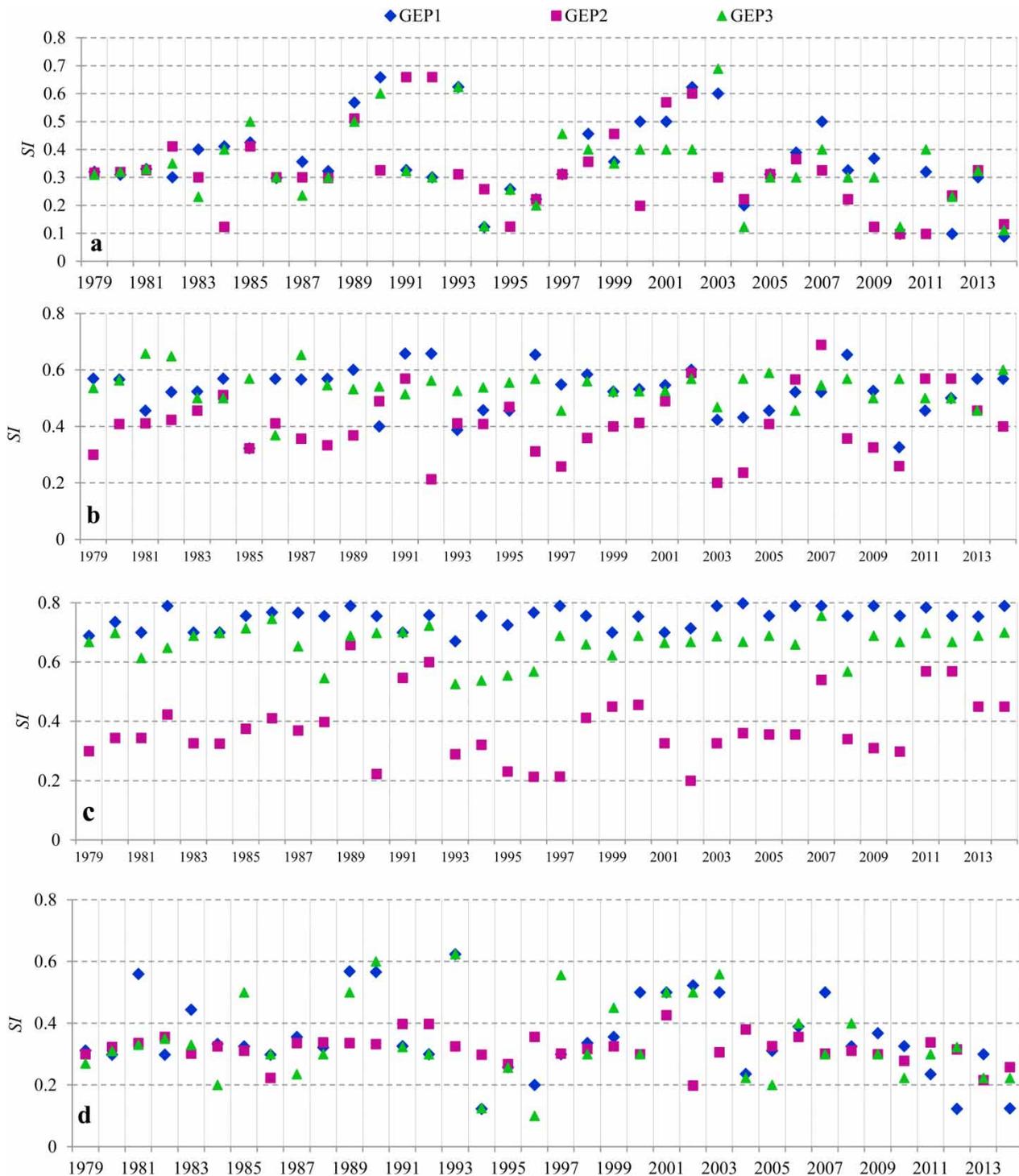
Note: The indices belong to the averaged statistics of the applied input configurations in the studied locations. External models were fed with input configuration II.



**Figure 3** | *SI* values of temporal GEP and SVM-based predictions split up per test station.

Figure 4 represents the temporal (split per test year)  $SI$  variations of the GEP models in the studied sites. As indicated in Figure 4, GEP2 outperforms GEP1 and GEP3 in

all sites and throughout the studied period with some minor exceptions. This difference is seen more clearly for Gaoyia station compared to the other stations. However,



**Figure 4** | Temporal GEP models'  $SI$  variations per test year: (a) S213 Bridge station, (b) Railway Bridge station, (c) Gaoyia station, (d) Pingchuan station.

there are considerable differences between models' accuracy. For instance, considering the test years 1979 and 1980, all three models present the same accuracy in term of *SI* values in S213 and Pingchuan station, while notable differences in *SI* values are seen in the models' performance for Railway Bridge station and Gaoya station at the same test years. However, in the test year 2000, GEP2 gives the highest accuracy in all stations except Pingchuan station, where GEP2 had less accuracy in comparison to the other applied models. Analyzing the *SI* differences of three input configurations ( $\Delta SI = SI_{\max} - SI_{\min}$ ) shows considerable variations in all stations. Considering GEP2,  $\Delta SI$  is 0.561, 0.489, 0.458, and 0.228 for S213 Bridge, Railway Bridge, Gaoya, and Pingchuan stations, respectively. Lower variations in Pingchuan station might be linked to the lowest skewness, standard deviation, and variation coefficient of streamflow time series in this station, among others. Analyzing the SVM performance variations (not presented here) showed similar statements.

Such variations in the performance of the applied models during the study period at each site might be linked to the variations of streamflow records considered in train–test stages. Thus, if the test year shows an abnormal trend with respect to the training patterns, e.g., comprising outliers, the derived GEP/SVM model might provide lower accuracy; for example, the test years of 1989, 1990, 1993, 2002, and 2003 for S213 Bridge station, the test years of 1991, 1992, 1996, 2002, 2007, and 2008 for Railway Bridge station, the test years of 1989, 1992, 1996, 2007, 2011, and 2012 for Gaoya station, and the test years of 1981, 1989, 1990, 1993, and 2003 for Pingchuan station. These fluctuations in the models' performance indicators clearly show the necessity of using a temporal k-fold testing for assessing the models' performance accuracy instead of using traditional data management scenarios (where a part of data is used for training, then the models are tested using the remaining part of the data), which may produce misleading results as they use a single data set assignment procedure.

The cross-correlation matrix of streamflow data among the studied stations is given in Table 4. The highest correlation is between the S213 Bridge and Railway Bridge stations. The reason for this may be the fact that these two stations are near to each other and there is no tributary between them. Gaoya station has the same correlation

**Table 4** | Correlation matrix of streamflow data

	S213 Bridge	Railway Bridge	Gaoya	Pingchuan
S213 Bridge	1.000	0.972	0.589	0.219
Railway Bridge	0.972	1.000	0.589	0.223
Gaoya	0.589	0.589	1.000	0.424
Pingchuan	0.219	0.223	0.424	1.000

with the S213 Bridge and Railway Bridge stations which show similar characteristics. Pingchuan station has low correlations with the other stations because it is located downstream of the basin and it has a higher drainage area than those of the others.

Table 5 represents the statistical indices of the external models at each test station. In the case of Pingchuan station, using the streamflow patterns of Gaoya station, which is its nearest station with the highest correlation between two sites ( $r = 0.424$ ), provides the most accurate results. Comparing the GEP and local SVM models, the external GEP model (fed with Gaoya data) is comparable to local SVM model (Table 3) with a *SI* difference of around 0.098, which confirms the ability of the external GEP model in predicting streamflow patterns of the target site using data from upstream stations (exogenous data). From Table 5, it is also

**Table 5** | Statistical indices of the external modeling

Train stations	GEP			SVM		
	<i>SI</i>	<i>VAF</i>	<i>NS</i>	<i>SI</i>	<i>VAF</i>	<i>NS</i>
Test station: Pingchuan						
Gaoya	0.458	0.896	0.896	0.490	0.858	0.858
Railway Bridge	0.342	0.952	0.952	0.409	0.946	0.932
S213 Bridge	0.521	0.810	0.881	0.568	0.787	0.787
All upstream stations <sup>a</sup>	0.422	0.929	0.927	0.456	0.900	0.900
Test station: Gaoya						
Railway Bridge	0.523	0.898	0.890	0.600	0.801	0.801
S213 Bridge	0.418	0.914	0.914	0.468	0.885	0.885
All upstream stations	0.423	0.901		0.478	0.880	0.875
Test station: Railway Bridge						
S213 Bridge	0.385	0.798	0.788	0.413	0.888	0.888

<sup>a</sup>All upstream stations' indicate the models fed with all data from upstream stations of the test site, pooled together.

seen that using data from Railway Bridge ( $r = 0.223$ ) and S213 Bridge ( $r = 0.219$ ) stations is less accurate than the first case, while pooling all patterns of the upstream stations (i.e., Gaoya, Railway Bridge, and S213 Bridge) improves their performance accuracy. The accuracy improvements of GEP models using pooled data were 7.8% and 19% (*SI* reduction) for the models trained by Railway Bridge and S213 Bridge data, respectively. For the same cases, these improvements in SVM models were 6.9% and 28%, respectively. This performance improvement is probably due to the inclusion of Gaoya station data in the pooled training patterns. It can be said that only data of Gaoya station (the nearest upstream station) should be used in estimating Pingchuan station because adding data of other upstream stations worsens the models' accuracy.

Regarding the Gaoya test station, similar to the previous case, the GEP and SVM models trained using the nearest station (Railway Bridge station;  $r = 0.589$ ) provide the best results. Utilizing pooled patterns of the upstream stations (Railway Bridge and S213 Bridge stations) improves the predictions when compared to the models fed with S213 Bridge ( $r = 0.589$ ) data. Comparing Tables 3–5 for the models fed with Railway Bridge station, it is seen that the external models give comparable results to local models with *SI* differences of about 0.010 and 0.024 for the GEP and SVM models, respectively. The performance accuracy order of the external models follows the magnitudes of the skewness coefficient (measure of the probability distribution asymmetry) of streamflow records. Among three stations, Pingchuan station (lowest skewness) provides the most accurate external results, followed by Railway Bridge and Gaoya (highest skewness). Higher skewness values make the extrapolation of the external models difficult.

Summarizing, it can be stated that external streamflow models can be a good surrogate for local ones when local data are absent or unreliable. Nonetheless, when using external data for building streamflow prediction models, it is necessary to use the data from the nearest station (or the station with the highest correlation with the target station), when there is no sink/source between two successive stations. In the case of missing any low-distance or high-correlation station, using pooled data from all upstream stations would be a good solution for developing prediction models of streamflow records, when local data at test station are scarce.

The results obtained in the present paper confirm the necessity of using a complete data scanning procedure in building heuristic-based prediction models of streamflow and using a single data assignment would be misleading as the model accuracy would fluctuate throughout the study period and among the studied sites. This confirms previous statements given in the literature regarding other hydrological parameters (e.g., Shiri *et al.* 2014b).

## CONCLUSIONS

The study has investigated the accuracy of GEP and SVM methods in forecasting streamflow by employing local and external data management scenarios and k-fold testing. Daily data collected from four stations, S213 Bridge, Railway Bridge, Gaoya, and Pingchuan, Heihe River, China were used in the applications. Three different input configurations determined based on correlation analysis were used to build the GEP and SVM models for prediction of one-day ahead streamflow. Local k-fold testing was applied at each station using both methods and the models comprising inputs of current and one previous day's streamflow data provided the best accuracy. According to the local k-fold testing results, the highest accuracy was obtained for Pingchuan station while the Gaoya station has the worst accuracy. GEP-based model performed better than the SVM-based models in all k-fold testing cases. The high variation of test results for each case indicated the necessity of k-fold testing in daily streamflow forecasting. Spatial (external) k-fold testing was also employed by setting upstream stations as inputs to the applied models and the downstream station as target. The models trained using the nearest upstream station provided the most accurate results. Comparing the external and local models showed that the external GEP model might be comparable to the local SVM model. External models can be successfully used in streamflow prediction when local data are unavailable or unreliable.

## ACKNOWLEDGEMENTS

This work was funded by the National Basic Research Program of China (2015CB953702) and National Natural

Science Foundation of China (41671335 and 91647104). All the data used in this study were obtained from Cold and Arid Regions Science Data Center (<http://westdc.westgis.ac.cn>).

## REFERENCES

- Badrzadeh, H., Sarukkalige, R. & Jayawardena, A. W. 2018 Intermittent stream flow forecasting and modelling with hybrid wavelet neuro-fuzzy model. *Hydrology Research* **49** (1), 27–40. doi: 10.2166/nh.2017.163.
- Banzhaf, W., Nordin, P., Keller, P. E. & Francone, F. D. 1998 *Genetic Programming*. Morgan Kaufmann, San Francisco, CA, 512 pp.
- Benninga, H. F. 2015 *Performance and Limitations of Ensemble River Flow Forecasts*. MSc Thesis, Civil Engineering and Management, University of Twente, Enschede, The Netherlands.
- Chemed Edossa, D. & Singh Babel, M. 2011 Application of ANN-based streamflow forecasting model for agricultural water management in the Awash River Basin, Ethiopia. *Water Resources Management* **25**, 1759–1773.
- Ferreira, C. 2001 Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems* **13** (2), 87–129.
- Ferreira, C. 2006 *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*. Springer, Berlin, Heidelberg, New York, 478 pp.
- Goldberg, D. E. 1989 *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA, 432 pp.
- Gunduz, O. & Aral, M. M. 2005 River networks and groundwater flow: a simultaneous solution of a coupled system. *Journal of Hydrology* **301** (1–4), 216–234.
- Guo, J., Zhou, J., Qin, H., Zou, Q. & Li, Q. 2011 Monthly streamflow forecasting based on improved support vector machine model. *Expert Systems with Applications* **38** (10), 13073–13081.
- Güven, A. 2009 Linear genetic programming for time-series modelling of daily flow rate. *Journal of Earth System Sciences* **118** (2), 137–146.
- Humphrey, G. B., Gibbs, M. S., Dandy, G. C. & Maier, H. R. 2016 A hybrid approach to monthly streamflow forecasting: integrating hydrological model outputs into a Bayesian artificial neural network. *Journal of Hydrology* **540**, 623–640.
- Kagoda, P. A., Ndiritu, J., Ntuli, C. & Mwaka, B. 2010 Application of radial basis function neural networks to short-term streamflow forecasting. *Physics and Chemistry of the Earth* **35** (13–14), 571–581.
- Karimi, S., Shiri, J., Kisi, O. & Shiri, A. A. 2016 Short-term and long-term streamflow prediction by using ‘wavelet-gene expression’ programming approach. *ISH Journal of Hydraulic Engineering* **22** (2), 148–162.
- Kisi, O. 2007 Streamflow forecasting using different artificial neural network algorithms. *ASCE Journal of Hydrologic Engineering* **12** (5), 532–539.
- Kisi, O. 2008 River flow forecasting and estimation using different artificial neural network techniques. *Hydrology Research* **39** (1), 27–40.
- Kisi, O. 2009 Neural networks and wavelet conjunction model for intermittent streamflow forecasting. *ASCE Journal of Hydrologic Engineering* **14** (8), 773–782.
- Kisi, O., Hossein zadeh Dalir, A., Cimen, M. & Shiri, J. 2012 Suspended sediment modeling using genetic programming and soft computing techniques. *Journal of Hydrology* **450–451**, 48–58.
- Kisi, O., Shiri, J. & Tombul, M. 2013 Modeling rainfall-runoff process using soft computing techniques. *Computers & Geosciences* **51**, 108–117.
- Koza, J. R. 1992 *Genetic Programming: On The Programming of Computers by Means of Natural Selection*. The MIT Press, Cambridge, MA, 840 pp.
- Liu, Z., Zhou, P., Chen, G. & Guo, L. 2014 Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting. *Journal of Hydrology* **519**, 2822–2831.
- Liu, S., Xie, Z. & Zeng, Y. 2016 Estimation of streamflow in ungauged basins using a combined model of black-box model and semi-distributed model: case study in the Yingluoxia watershed. *Journal of Beijing Normal University* **52** (3), 393–401. (Published in Chinese with English abstract.)
- Maier, H. R. & Dandy, G. C. 1996 Use of artificial neural networks for prediction of water quality parameters. *Water Resources Research* **32** (4), 1013–1022.
- Marti, P., Manzano, J. & Royuela, A. 2011 Assessment of 4-input artificial neural network for ETO estimation through data set scanning procedures. *Irrigation Science* **29**, 181–195.
- Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models. Part I: A discussion of principles. *Journal of Hydrology* **10**, 282–290.
- Nayak, P. C., Sudheer, K. P., Rangan, D. M. & Ramasastri, K. S. 2004 A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology* **291**, 52–66.
- Sarlak, N. 2008 Annual streamflow modelling with asymmetric distribution function. *Hydrological Processes* **22**, 3403–3409.
- Sharma, S., Srivastava, P., Fang, X. & Kalin, L. 2015 Performance comparison of Adoptive Neuro Fuzzy Inference System (ANFIS) with Loading Simulation Program C++ (LSPC) model for streamflow simulation in El Niño Southern Oscillation (ENSO)-affected watershed. *Expert Systems with Applications* **42** (4), 2213–2223.
- Shiri, J. & Kisi, O. 2010 Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model. *Journal of Hydrology* **394**, 486–493.
- Shiri, J., Sadraddini, A. A., Nazemi, A. H., Kisi, O., Landaras, G., Fakheri Fard, A. & Marti, P. 2014a Generalizability of gene

- expression programming-based approaches for estimating daily reference evapotranspiration in coastal stations of Iran. *Journal of Hydrology* **508**, 1–11.
- Shiri, J., Marti, P. & Singh, V. P. 2014b Evaluation of gene expression programming approaches for estimating daily evaporation through spatial and temporal data scanning. *Hydrological Processes* **28** (3), 1215–1225.
- Shiri, J., Nazemi, A. H., Sadraddini, A. A., Landaras, G., Kisi, O., Fakhri Fard, A. & Marti, P. 2014c Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran. *Computers and Electronics in Agriculture* **108**, 230–241.
- Tennant, D. L. 1976 Instream flow regimes for fish, wildlife, recreation and related environmental resources. *Fisheries* **1**, 6–10.
- Uysal, G., Şorman, A. A. & Şensoy, A. 2016 Streamflow forecasting using different neural network models with satellite data for a snow dominated region in Turkey. *Procedia Engineering* **154**, 1185–1192.
- Vapnik, V. 1995 *The Nature of Statistical Learning Theory*. Springer Verlag, New York, pp. 1–47.
- Wang, W., Van Gelder, P., Vrijling, J. K. & Ma, J. 2006 Forecasting daily streamflow using hybrid ANN models. *Journal of Hydrology* **324**, 383–399.
- Wang, W., Chau, K. W., Cheng, C. T. & Qiu, L. 2009 A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology* **374**, 294–306.
- Wu, C. L. & Chau, K. W. 2010 Data-driven models for monthly streamflow time series prediction. *Engineering Applications of Artificial Intelligence* **23**, 1350–1367.
- Yarar, A. 2014 A hybrid wavelet and neuro-fuzzy model for forecasting the monthly streamflow data. *Water Resources Management* **28**, 553–565.

First received 15 June 2017; accepted in revised form 5 October 2017. Available online 24 November 2017