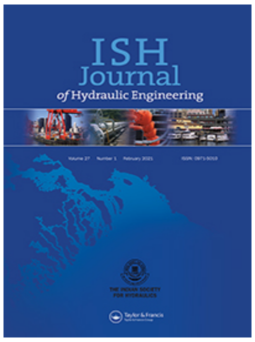
Forecasting monthly streamflows using heuristic models

Article *in* ISH Journal of Hydraulic Engineering · September 2018

DOI: 10.1080/09715010.2018.1516575

Sahar Mohsenzadeh Karimi
Simon Fraser University

Sepideh Karimi
University of Tabriz



ISH Journal of Hydraulic Engineering

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tish20

Forecasting monthly streamflows using heuristic models

Sahar Mohsenzadeh Karimi , Sepideh Karimi & Mohammad Poorrajabali

To cite this article: Sahar Mohsenzadeh Karimi , Sepideh Karimi & Mohammad Poorrajabali (2021) Forecasting monthly streamflows using heuristic models, ISH Journal of Hydraulic Engineering, 27:1, 73-78, DOI: 10.1080/09715010.2018.1516575

To link to this article: https://doi.org/10.1080/09715010.2018.1516575

Published online: 05 Sep 2018.

ISH JOURNAL OF HYDRAULIC ENGINEERING
2021, VOL. 27, NO. 1, 73–78
https://doi.org/10.1080/09715010.2018.1516575

Forecasting monthly streamflows using heuristic models

Sahar Mohsenzadeh Karimi, Sepideh Karimi and Mohammad PoorrajabaliDepartment of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

ABSTRACT
Forecasting streamflow values is of great importance in hydrology and water resources engineering as it affects the water related inflow-demand management, dam structure design and river engineering studies. Apart from using some physics-based models of this parameter forecast, using the previously recorded streamflow values for forecasting the future values would be very interesting, as only streamflow time series will be needed there. The present study aimed at assessing three heuristic data driven approaches, namely, gene expression programming (GEP), support vector machine (SVM) and interactive trees (IT) in forecasting monthly streamflow records. Monthly data from Soofi-Chai river in Iran covering a period of 13 years were used and a local k-fold testing cross validation process was adopted for training and testing the applied models. The obtained results revealed that all the applied models could predict riverflow time series with good accuracy. The results also showed the importance of defining a through train-test block mode (here, k-fold testing) to get a better insight about the applied models.

ARTICLE HISTORY
Received 25 March 2018
Accepted 23 August 2018

KEYWORDS
Streamflow; GEP; SVM; IT; cross validation

1. Introduction

Streamflow forecasting is a significant and important subject in water resources activities such as surface water systems (Hayes et al. 1996; Smith et al. 1998; Adamowski 2008; Kisi 2009; Shiri and Kisi 2010; Kisi et al. 2013a; Yarar 2014; Yavari et al. 2017; Karimi et al. 2018). Hence, anticipation of the streamflow data with reasonable precision would be effective in flood forecasting too, where, this type of predictions with sufficient lead time will decrease flood damages through the areas. Meanwhile, some hydrological parameters such as precipitation, evapotranspiration, groundwater level fluctuation and even the antecedent moisture content of the soil are sequentially contributed on the amount of the next day streamflow as discussed by Kisi (2008). In the process of forecasting, applications of heuristic data driven models such as gene expression programming (GEP), support vector machine (SVM) and interactive trees (IT) have been developed in many majors of hydrology area in the recent years. This methods have demonstrated admissible results in streamflow forecasting, rainfall-runoff modeling, reservoir inflow predictions, rainfall forecasting, river sediment modeling, etc (e.g. Smith and Eli 1995; Smith et al. 1998; Minns and Hall 1996; Tayfur 2002; Cancelliere et al. 2002; Supharatid 2003; Cigizoglu and Kisi 2006; Kisi 2007; Shiri and Kisi 2010; Kisi et al. 2012; Yarar 2014; Karimi et al. 2018).

In this way, a full review of studies in water resources engineering is beyond the scope of our paper and we will review some examples especially regarding the streamflow modeling. In this context, Guven (2009) utilized linear genetic programming on daily flow rates values for time-series modeling. Guo et al. (2011) used the SVM for forecasting monthly streamflow values and found the SVM more accurate than artificial neural networks (ANNs). Shiri and Kisi (2011a) predicted the short term groundwater level fluctuations with neuro-fuzzy and GEP. The comparison of mentioned methods presented that GEP was better than the ANFIS. To predict daily lake level variations, Kisi et al. (2012) applied AI models such as GEP. In this regard, they found GEP as the most accurate model. Kisi et al. (2013b) found the GEP as the most acceptable method in modelling the rainfall-runoff in comparing with other different heuristic models. Karimi et al. (2016) studied short-term and long-term streamflow values and introduced a wavelet-GEP method as a capable model for forecasting them. Liu et al. (2016) utilized a combine model of wavelet-SVM for estimating daily and monthly streamflow values in Yingluoxia watershed. Hereupon, some of heuristic approaches like SVM and GEP models have considered among the most acceptable methods.The goal with the present paper was to compare the performance of GEP and SVM models with those of the IT technique (that has not been considered in this context yet) to forecast monthly streamflow records using a complete data assignment strategy.

2. Methodologies adopted

2.1. *Gene expression programming*

For the first time, Genetic Programming (GP) was offered as a generalization of Genetic Algorithms (GAs) (Goldberg 1989) by Koza (1992). This method is appreciated for poorly comprehend interrelationships among relevant variables. Meanwhile, assumptions are caused some constrains in theoretical analysis. Moreover, a tedious processing is essential for a large amount of data in computer readable forms. Therefore, the solutions of GP are of limited usage. Owing to these restrictions, GP is analogous to GEP which expand the computer programs in diverse sizes and encoded forms with fixed lengths of linear chromosomes. GEP as a genetic algorithm utilize sets of data which are populations of individuals. In this process, the chromosomes are formed from multiple genes that any of genes contains a smaller subprogram. On the other hand, structural organization of the linear chromosomes in GEP program authorizes it to operate

CONTACT Sahar Mohsenzadeh Karimi mohsenzadeh.sahar21@gmail.com

the notable genetic operators such as mutation, transposition and recombination without any limitation. In this way, GEP selects sets of data to fitness accordingly and makes models to introduce the relationships between variables.

Meanwhile, there are some advantages which make GEP approach stronger as a learning algorithm. Generally because of the nature of GEP, the creation of genetic diversity is greatly simple. On the other hand, this simple nature is mutagenic which allow the approach developing complex programs which are composed of several subprograms (Shiri and Kisi 2011b). Genetic diversity in GEP approach is extremely simplified. Furthermore, the expression trees are exclusively the expression of their respective chromosomes; they are entities upon which selection acts, and according to fitness, they are selected to reproduce with modification (Ferreira 2001a). Another advantage of GEP is being capable to express the relationship between the input-output variables. This unique feature made the GEP stronger than SVM and IT (Landeras et al. 2018; Shiri 2017).

2.2. *Support vector machine*

The original algorithm of SVM was devised by Vapnik (1995) which is a set of supervised learning algorithm that have been based on the concept of decisions in linear data categorization. In this process, SVM analyze data and recognize the patterns, which is used carry out the classification and regression challenges. The advantages of SVM technique which make it powerful can be summarized in following cases. By utilizing the kernel, SVMs can obtain more flexibility in the choice of the form of the threshold separating solvent from insolvent companies, which needs not be linear and even needs not have the same functional form for all data, since its function is non-parametric and operates locally. Furthermore, SVMs deliver is a unique solution, since the optimality problem is convex (Shiri et al. 2014; Kim et al. 2015). This case compared to Neural Networks, which have multiple solutions associated with local minima and for this reason may not be robust over different samples (Auria and Moro, 2008).

The regression-SVM type 1 was applied in this study which has demonstrated in the literature (Shiri et al. 2014a) as superiority to other types. This method estimates a functional relationship $f(\vec{x})$ between the sample points $X = \{\vec{x}_1, \vec{x}_2, \ldots\ldots, \vec{x}_l\}$ which are driven from Rn and target values $Y = \{y_1, y_2, \ldots\ldots, y_l\}$ with$y_i \in R$. Furthermore, SVM considers a regularized risk function ($R_{reg}(.)$) which makes the user consider about avoiding over-fitting. This parameter has been described as following (Vapnik et al. 1997; Gunn 1998):

$$R_{\mathrm{reg}}[f(\vec{x})] = C_C \sum_{x_i \in X} l_\varepsilon(y_i - f(\vec{x}_i)) + \frac{1}{2}\|\vec{w}\|^2 \tag{1}$$

Where the C_C demonstrate a positive constant in the calculation which can be introduced as a design parameter, $l_e(.)$ represents a regression vector which is estimated with using the data. This parameter is used for measuring the deviation between the target (y) and corresponding estimated ($f(\vec{x})$) values. Finally, $\|\vec{w}\|$ represents a regression vector which should be estimated using the data. In this context, SVM performs linear regression in the high-dimension feature space with using the ε-insensitive loss as well as tries to reduce model complexity by minimizing $\|\vec{w}\|$in parallel way. The loss function in this expression can be expressed as:

$$l_\varepsilon(y_i - f(\vec{x}_i)) = \begin{cases} 0 & \text{for} \quad |y_i - f(\vec{x}_i)| < \varepsilon \\ |y_i - f(\vec{x}_i)| & \text{otherwise} \end{cases} \tag{2}$$

According to the mentioned formula, when the difference between the target (y_i) and estimated $f(\vec{x}_i)$ values are less than ε, the loss value would be equal to zero. It is significant that choosing value for ε is easier than C_C, because it is commonly given as a desired percentage of the target values y_i. Therefore, nonlinear regression function will be expressed as (Vapnik et al. 1997; Gunn 1998):

$$f(x) = \sum_{i=1}^{N} (\alpha_i^* - \alpha_i) K(x, x_i) + B \tag{3}$$

Where $\alpha_i, \alpha_i^* \geq 0$ demonstrate the Lagrange multipliers, B is a bias term, and $K(x,\ x_i)$ represented the kernel function which makes the SVM to perform the operations in the input space. Detailed information for SVM can be found in Vapnik (1995).

2.3. *Interactive tress (IT)*

IT is a type of the learning algorithms which provides interactivity and complete control over the modelling process. It is a tree-based method and can be considered as a model with high accuracy, stability and ease of interpretation. This partitioning method, builds trees for predicting the dependent parameters which is supported by CART algorithm (Breiman et al. 1984). The applying method is the recursive partitioning to split the training records into segments with similar output field values. Finding the best split starts by examining the input data which is the duty of the tree node. In the following, the split defines two subgroups and each of them split to other subgroups until one of the stopping criteria is triggered.

One of the advantages of IT is related to its ability to optimize very large data sets. Furthermore, this model is flexible in handling of missing data. Performing 'what-if' analyses to obtain the best insights into the data by interactively deleting individual branches, growing other branches, and observing various results statistics for the different trees are another advantage of IT.

3. Applications

3.1. *Used data*

Monthly stream flow values of the Alaviam Dam on the Soofi-Cahi River in the Northwest of Iran which were recorded at the inlet (Latitude:37°25′N, Longitude:46°15′E), were used in the present study. The observed data includes 156 months (13 years) which is consisting on September 1997–September 2010. Meanwhile, training and testing process for developing of the applied models were applied using the k-fold test validation. So,at each time of modeling process, a block (one year) of data were considered as the test set and the models were trained using the remaining part of the available data. The process was repeated till all the data were particip[ated in modeling (training and testing) phase. Some monthly statistical parameters such as X_{mean}, X_{min}, X_{max}, S_d, C_v, C_{sx} and C_K which displayed the mean, minimum, maximum, standard deviation, coefficient of variation, skewness coefficient and kurtosis, respectively, have been presented in Table 1.

From the table, it is clear that the monthly streamflow data showed significantly high skewed distribution (C_{sx} = 2.039). In this area, the value of maximum streamflow as flood peak was also significant (95 m^3s^{-1}).

3.2. Study flowchart and data splitting

Identifying the most suitable mood for input data is one the first steps in modelling different processes. In the current study, auto correlation technique was applied to detect the number of effective lags in time series of steamflow at the Alaviam Dam station. Figure 1 illustrates the partial auto correlation function (PACF) diagrams of the streamflow data for the study period which shows that the first two lags have been effective based on correlation. Accordingly, two-times lags were used for modelling the one-month ahead streamflow values. So, Q_{t+1} as a model's target demonstrates the stream values at the next month. In this process, two Input configuration can considered for modelling with GEP, SVM and IT approaches:

Input configuration I: Q_t (GEP1, SVM1, IT1 models)
Input configuration II: Q_t, Q_{t-1} (GEP2, SVM2, IT2 models).

In the following steps, all the mentioned modelling methods were applied on input configurations with k fold testing (test size = 1 year) that repeated on total 78 train-testing processes (1 stations × 3 models × 2 input configurations × 13 years = 78). Finally, to evaluate models validation, four statistical parameters e.g. *R, RMSE, SI* and *NS* were applied. This process was once applied on whole data and once more on each year.

Table 1. Statistical parameters of the applied river flow data.

Station	Min (m^3s^{-1})	Max (m^3s^{-1})	Mean (m^3s^{-1})	SD (m^3s^{-1})	CV	C_{sx}	C_K
Alavian Dam	4.950	95.220	519.370	116.525	1.224	2.039	3.334

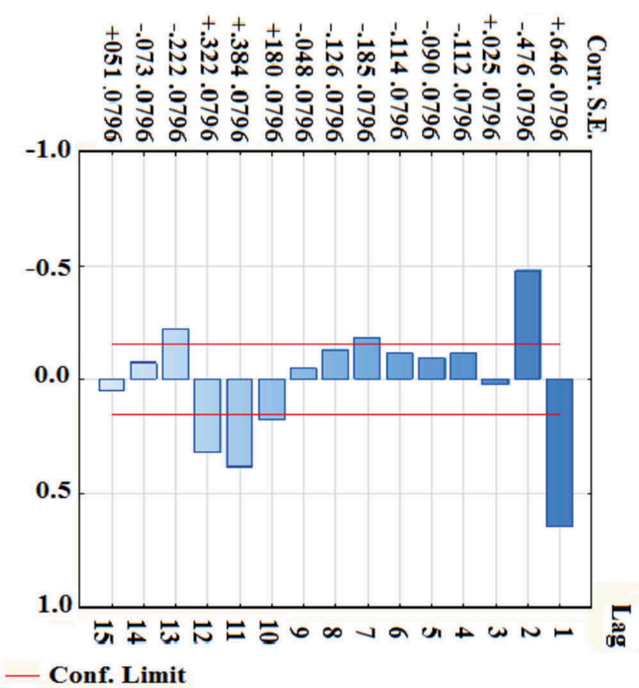


Figure 1. PACF of stream flow data in the studied river.

3.3. Statistical validation

To evaluate the performance of applied methods, four statistical indices were utilized as follows:

The coefficient of correlation (*R*); which is used to measure the strength of a linear relationship between variables. This Index considers a range of numbers between 0 and 1 which the higher values of *R* demonstrate the better performance of the model.

$$R = \frac{\sum_{i=1}^{n}(Q_{io} - \bar{Q}_{io})(Q_{ie} - \bar{Q}_{ie})}{\sqrt{\sum_{i=1}^{n}(Q_{io} - \bar{Q}_{io})^2 \sum_{i=1}^{n}(Q_{ie} - \bar{Q}_{ie})^2}} \quad (4)$$

Where, Q_{io} and Q_{ie} show the observed and forecasted streamflow values, respectively. *n* stands for the number of time steps. $\bar{Q}_{io}$ and $\bar{Q}_{io}$ represents the mean observed and forecasted streamflow values.

One the other hand, *R* is not capable to be used as a fitness measurement lonely. Thus, it is considerable to evaluate the amount of variable errors in the same unit. Hence, the root mean square error (*RMSE)*, the Scatter Index (*SI*) and the Nash-Sutcliffe (*NS*) are the other indicators which have been used in this way and defined as follows (Shiri et al. 2014b):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_{io} - Q_{ie})^2} \quad (5)$$

$$SI = \frac{RMSE}{\bar{Q}_{io}} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_{io} - Q_{ie})^2}}{\bar{Q}_{io}} \quad (6)$$

$$NS = 1 - \frac{\sum_{i}^{n}(Q_{io} - Q_{ie})^2}{\sum_{i}^{n}(Q_{io} - \bar{Q}_{io})^2} \quad (7)$$

If the *SI and NS* values are 0 and 1, respectively, this will represent the perfect fit. In addition, using this aforementioned statistical measures together make a better insight about the employed models.

4. Results and discussion

Statistical indices of each input configuration for the GEP, SVM and IT models are summarized in Table 2. The values presented in Table 2 were calculated by averaging the statistical indices of each year, instead of averaging the statistical indictors of 13 years. It can be detected from Table 2 that the prediction results of input configuration II were more precise than the first configuration. As mentioned, Q_t, Q_{t-1} have been applied as input parameters for input configuration II.

This process demonstrated that the *SI* values have decreased by 23%, 2.4% and 21.3% for GEP, SVM and IT models, respectively (compared to input configuration I). In addition, the superiority of the GEP2, SVM2 and IT2 is confirmed by the values of *R* and *NS* indices in Table 2. Moreover, from the table it can be seen that the minimum SI values of two configurations were corresponded to GEP (0.330 and 0.254, respectively), while, the maximum *SI* values are belonged to

Table 2. Global average performance parameters of the local GEP, SVM and IT models.

	GEP			SVM			IT		
Models	*R*	*SI*	*NS*	*R*	*SI*	*NS*	*R*	*SI*	*NS*
Input configuration I	0.944	0.330	0.877	0.941	0.370	0.851	0.945	0.368	0.863
Input configuration II	0.965	0.254	0.911	0.971	0.346	0.853	0.959	0.289	0.903

SVM models. Hence, predictions relying on SVM method present low precision simulation than GEP and IT methods. It should be noted that the high values of skewness and standard deviation which was given in Table 1 demonstrate a low variability and also these factors affect the accuracy of models. Nevertheless, values of validation indices (SI, R and NS) of GEP model relying on input configuration II display the lowest variability among other models. In conclusion, based on these reasons, the GEP method outperforms other models. This high accuracy perhaps is related to the nature of GEP which utilize the evolutionary genetic algorithm in the calibration process. Additionally, temporal analysis of GEP, SVM and IT models were performed and the results were presented in Figure 2.

According to Figure 2, the lowest values of SI during the test years of 1997, 2001, 2002, 2003, 2005, 2006 and 2008 assign to GEP method in two configurations. Generally, this notable point confirms the highest accuracy of GEP among the other models. Nevertheless, the mentioned result has also some exceptions. For example, in test years 1998 and 2000, GEP model has less accuracy in comparison to the IT model. Furthermore, there are some remarkable similarities and differences between model's accuracy in two configurations. For instance, during the test years 1998, 1999 and 2004 the GEP and IT models demonstrate the similar accuracy. On the other hand, the results of SVM model show the lowest accuracy during all of the test years except 1997, 2001 and 2003. The results of analyzing the *SI* differences of three applied models in Input figuration II ($\Delta SI = SI_{max} - SI_{min}$) introduced the SVM as a low precision method which was assigned the highest value of ΔSI. (0.370, 0.251 and 0.279 for SVM2, IT2 and GEP2 respectively).

By referring to the Figure 2, there are abnormal trends with are related to considerable differences between applied models in two configurations. For instance, during the test years of 2002, 2004, 2006 and 2009, the *SI* values of GEP1 and GEP2 have significant differences. Furthermore, these types of fluctuations are clearly can be also seen between other models in Figure 2. These fluctuations demonstrate the necessity of utilizing the temporal k-fold testing which applied in the present study. The traditional procedure (where the data is divided into training and test parts to run the model) due to utilizing a single data set assignment procedure may cause to produce the misleading results. Hence, using the k-fold testing cross- validation instead of the traditional procedure causes the superior outputs.

Finally, Figure 3 shows the observed vs. predicted riverflows for some sample test periods (years) for the applied models, that demonstrated the applicability of the all applied models (more or less) in the studied case.

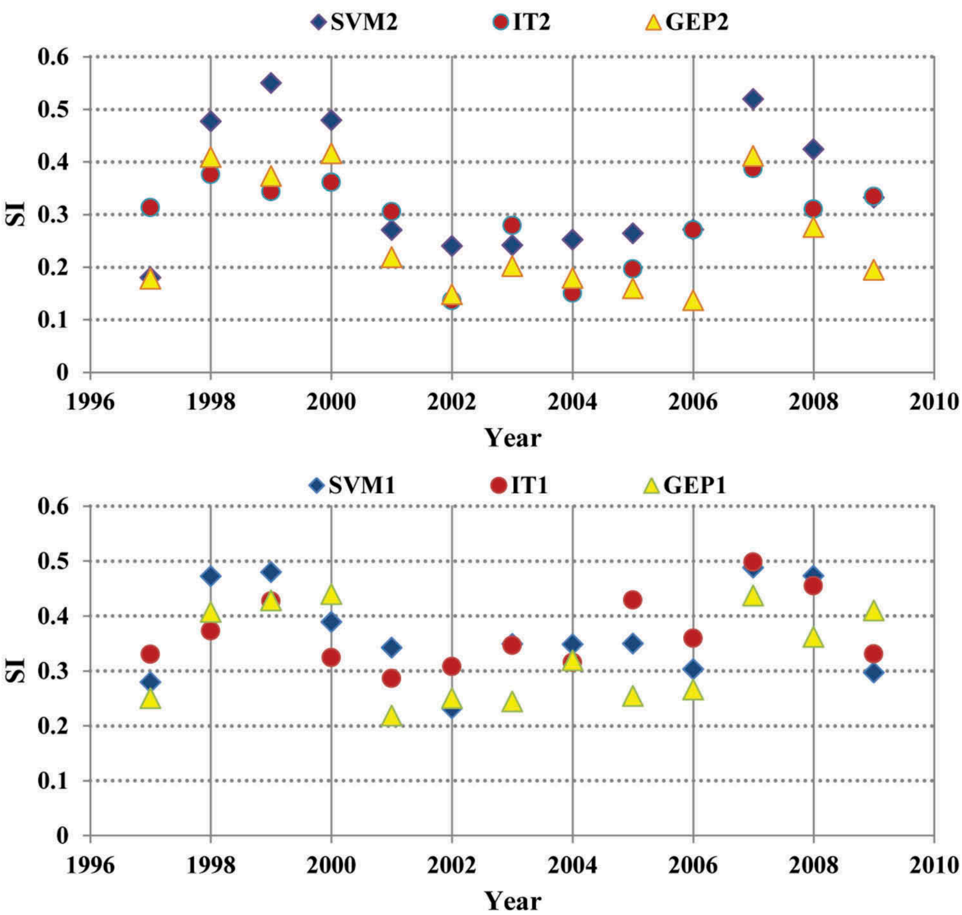


Figure 2. Temporal SI variation in Input configurations of GEP, SVM and IT models.

5. Conclusion

The present study has arranged for comparative purposes which compared the accuracies of three diverse artificial intelligence techniques. In this way, the GEP, SVM and IT models were applied to monthly streamflow data of Alavian Dam Station on the Soofi-Cahi River in the Northwestern Iran. The local k-fold testing cross validation process was employed on these 13 years period with two different input configurations that based on the partial auto-correlation function. The second input combination which includes the Q_t, Q_{t-1} values showed the best accuracy for all applications. By referring to the results of the local k-fold testing, all the applied models were able to predict the riverflow magnitudes with appropriate accuracy. Meanwhile the predictions relying on SVM-based model showed low precision simulation than the GEP and IT, which were based on statistical results. Moreover, the importance of defining a through train-test block mode to obtain a better insight about the applied models was observed.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Adamowski, J.F. (2008). "Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis." *J. Hydrol.*, 353, 247–266. doi: 10.1016/j.jhydrol.2008.02.013

Auria, L., and Moro, R.A., 2008. Support vector machines (SVM) as a technique for solvency analysis. DIW Berlin. (Published as a discussion paper at http://ideas.repec.org/s/diw/diwwpp.html).

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984). "*Classification and regression trees*." Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA.

Cancelliere, A., Giusiano, G., Ancarani, A., and Rossi, G. (2002). "A neural network for deriving irrigation reservoir operating rules." *Water Resour. Manag.*, 16, 71–88. doi: 10.1023/A:1015563820136

Cigizoglu, H.K., and Kisi, O. (2006). "Methods to improve the neural network performance in suspended sediment estimation." *J. Hydrol.*, 317, 221–238. doi: 10.1016/j.jhydrol.2005.05.019

Ferreira, C. (2001a). "Gene expression programming: A new adaptive algorithm for solving problems." *Complex Syst.*, 13(2), 87–129.

Goldberg, D.E. (1989). "*Genetic algorithms in search, optimization, and machine learning*." Addison-Wesley, Reading, MA.

Gunn, S.R., 1998. Support vector machines for classification and regression. Technical Report. University of Southampton, England.

Guo, J., Zhou, J., Qin, H., Zou, Q., and Li, Q. (2011). "Monthly streamflow forecasting based on improved support vector machine model." *Expert Syst. Appl.*, 38(10), 13073–13081. doi: 10.1016/j.eswa.2011.04.114

Guven, A. (2009). "Linear genetic programming for time-series modelling of daily flow rate." *J. Earth Syst. Sci.*, 118(2), 137–146. doi: 10.1007/s12040-009-0022-9

Hayes, M.J., Wilhite, D.A., Svoboda, M.D., and Smith, K.H., 1996. Drought management: Crisis vs management. Proc. 1996 Int. Conf. Exposition on Natural Disaster Reduction, ACSE, Washington, DC, 371–372.

Karimi, S., Shiri, J., Kisi, O., and Shiri, A. (2016). "Short-term and long-term streamflow prediction by using 'wavelet–Gene expression' programming approach." *ISH J. Hydraulic Eng.*, 22(2), 148–162. doi: 10.1080/09715010.2015.1103201

Karimi, S., Shiri, J., Kisi, P., and Xu, T. (2018). "Forecasting daily streamflow values: Assessing heuristic models." *Hydrol. Res.*, 49 (3), 658–669. doi: 10.2166/nh.2017.111

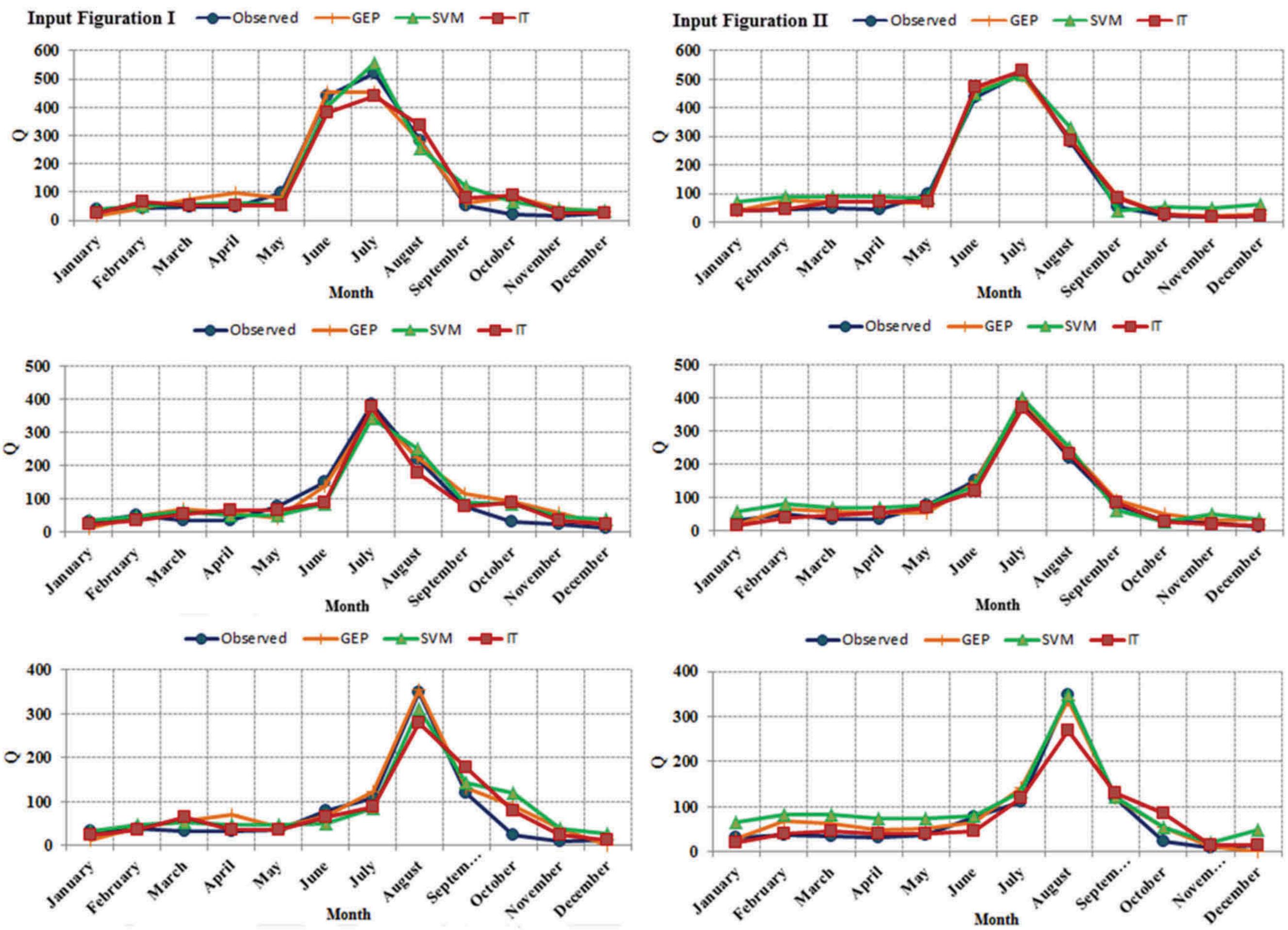


Figure 3. Some plots for sample test periods (years) for the modeled case.

Kim, S., Shiri, J., Singh, V.P., Kisi, O., and Landeras, G. (2015). "Predicting daily pan evaporation by soft computing models with limited climatic data." *Hydrol. Sci. J.*, 60(6), 1120–1136. doi: 10.1080/02626667.2014.945937

Kisi, O. (2007). "Evapotranspiration modeling from climate data using a neural computing technique." *Hydrol. Process.*, 21(6), 1925–1934. doi: 10.1002/hyp.6403

Kisi, O. (2008). "River flow forecasting and estimation using different artificial neural network techniques." *Hydrol. Res.*, 39(1), 27–40. doi: 10.2166/nh.2008.026

Kisi, O. (2009). "Neural networks and wavelet conjunction model for intermittent streamflow forecasting." *J. Hydrol. Eng.*, 14(8), 773–782. doi: 10.1061/(ASCE)HE.1943-5584.0000053

Kisi, O., Shiri, J., and Nikoofar, B. (2012). "Forecasting daily lake levels using artificial intelligence approaches." *Comput. Geosci.*, 41, 169–180. doi: 10.1016/j.cageo.2011.08.027

Kisi, O., Akbari, N., Sanatipour, M., Hashemi, A., Teimourzadeh, K., and Shiri, J. (2013a). "Modeling of dissolved oxygen in river water using artificial intelligence techniques." *J. Environ. Inform.*, 22(2), 92–101. doi: 10.3808/jei.201300248

Kisi, O., Shiri, J., and Tombul, M. (2013b). "Modeling rainfall-runoff process using soft computing techniques." *Comput. Geosci.*, 51, 108–117. doi: 10.1016/j.cageo.2012.07.001

Koza, J.R. (1992). "*Genetic programming: On the programming of computers by means of natural selection.*" MIT Press, Cambridge, MA.

Landeras, G., Bekoe, E., Ampofo, J., Logah, F., Diop, M., Cisse, M., and Shiri, J. (2018). "New alternatives for reference evapotranspiration estimation in West Africa using limited weather data and ancillary data supply strategies." *Theor. Appl. Climatol.*, 132, 701–716. doi: 10.1007/s00704-017-2120-y

Liu, S., Xie, Z., and 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." *J. Beijing Normal Univ.*, 52(3), 393–401.

Minns, A.W., and Hall, M.J. (1996). "Artificial neural networks as rainfall-runoff models." *Hydrol. Sci. J.*, 41(3), 399–418. doi: 10.1080/02626669609491511

Shiri, J. (2017). "Evaluation of FAO56-PM, empirical, semi-empirical and gene expression programming approaches for estimating daily reference evapotranspiration in hyper-arid regions of Iran." *Agric. Water Manag.*, 188, 101–114. doi: 10.1016/j.agwat.2017.04.009

Shiri, J., and Kisi, O. (2010). "Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model." *J. Hydrol.*, 394, 486–493. doi: 10.1016/j.jhydrol.2010.10.008

Shiri, J., and Kisi, O. (2011a). "Comparison of genetic programming with neuro-nuzzy systems for predicting short-term water table depth fluctuations." *Comput. Geosci.*, 37(10), 1692–1701. doi: 10.1016/j.cageo.2010.11.010

Shiri, J., and Kisi, O. (2011b). "Application of artificial intelligence to estimate daily pan evaporation using available and estimated climatic data in the Khozestan Province (South-Western Iran)." *J. Irrigation Drainage Eng.*, 137(7), 412–425. doi: 10.1061/(ASCE)IR.1943-4774.0000315

Shiri, J., Marti, P., and Singh, V.P. (2014b). "Evaluation of gene expression programming approaches for estimating daily evaporation through spatial and temporal data scanning." *Hydrol. Proces.*, 28(3), 1215–1225. doi: 10.1002/hyp.v28.3

Shiri, J., Nazemi, A.H., Sadraddini, A.A., Landeras, G., Kisi, O., Fakheri Fard, A., and Marti, P. (2014a). "Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran." *Comput. Electron. Agric.*, 108, 230–241. doi: 10.1016/j.compag.2014.08.007

Smith, J., and Eli, R.N. (1995). "Neural networks models of rainfall-runoff process." *J. Water Resour. Plann. Manag.*, 121(6), 499–508. doi: 10.1061/(ASCE)0733-9496(1995)121:6(499)

Smith, L.C., Turcotte, D.L., and Isacks, B. (1998). "Stream flow characterization and feature detection using a discrete wavelet transform." *Hydrol. Process.*, 12, 233–249. doi: 10.1002/(ISSN)1099-1085

Supharatid, S. (2003). "Application of a neural network model in establishing a stage-discharge relationship for a tidal river." *Hydrol. Process.*, 17, 3085–3099. doi: 10.1002/(ISSN)1099-1085

Tayfur, G. (2002). "Artificial neural networks for sheet sediment transport." *Hydrol. Process.*, 4(6), 879–892. doi: 10.1080/02626660209492997

Vapnik, V. (1995). "*The nature of statistical learning theory*". Springer Verlag, New York, 1–47

Vapnik, V., Golwich, S., and Smola, A.J. (1997). "Support vector method for function approximation, regression estimation, and signal processing." *Advances in neural information processing systems 9*, M. Mozer, M. Jordan, and T. Petsche, eds., MIT Press, Cambridge, Massachusetts, USA, 281–287

Yarar, A. (2014). "A hybrid wavelet and neuro-fuzzy model for forecasting the monthly streamflow data." *Water Resour. Manag.*, 28, 553–565. doi: 10.1007/s11269-013-0502-1

Yavari, S., Maroufpoor, S., and Shiri, J. (2017). "Modeling soil erosion by data-driven methods using limited input variables." *Hydrol. Res.*, doi: 10.2166/nh.2017.041