

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 125, No. 2, April – June, 2021, pp. 195–209

Comparison of the performances of GEP, ANFIS, and SVM artificial intelligence models in rainfall simulation

Seyed Mostafa Tabatabaei*, Mohammad Nazeri Tahroudi, and Bahareh Sadat Hamraz

Department of Water Engineering
University of Birjand, Birjand, Iran

**Corresponding author E-mail: tabatabaei1984@yahoo.com*

(Manuscript received in final form June 15, 2020)

Abstract—In this paper, evaluation the performances of GEP (gene expression programming), ANFIS (adaptive fuzzy interference system), and SVM (support vector machine) artificial intelligence models in two scales of daily and monthly rainfall data from Urmia meteorological station (Iran) and monthly rainfall data from Diata meteorological station (India) was used in rainfall simulation. The correlation coefficient of observed and simulated values was evaluated by the R^2 criterion, simulation error was evaluated by the root mean square error (RMSE), and MB criteria and model efficiency were evaluated by the Nash-Sutcliffe method. The results show that the correlation coefficients in the GEP model based on daily data from Urmia station and monthly data from Diata station are 23 and 58%, respectively, and R^2 in simulation with GEP is estimated to be 55% lower than with the other two models. The R^2 range in both ANFIS and SVM models varies from 91 to 93%. On average, the RMSE values in the GEP simulation are 50% and 55% higher than the ANFIS ratio for daily and monthly data at the two stations, respectively, and the RMSE values of ANFIS model are 1% and 3% higher than those of the SVM at Urmia and Diata stations, respectively. The bias values of the GEP model are 72% and 60% higher than those of ANFIS at Urmia and Diata stations, respectively. The GEP efficiency factors are 56% and 61% lower than those of ANFIS at Urmia and Diata stations, respectively. And the ANFIS efficiency ratio is 1 and 2% lower than SVM in Urmia and Diata stations, respectively. Therefore, rainfall simulation with the SVM model is associated with a lower error rate and better efficiency, the ANFIS model is close to the efficiency of SVM, and the GEP model is not suitable for rainfall simulation.

Key-words: artificial intelligence models, gene expression programming, rainfall simulation, support vector machine

1. Introduction

One way to study hydrological systems and water resources to predict the behavior of its components is to use a model, or in other words, to simulate its processes (*Refsgaard and Knudsen, 1996*). Rainfall is an important factor that is directly involved in the hydrological cycle. Simulation and forecasting of this factor play an important role in hydrological planning and water and agricultural resources. (*Refsgaard et al., 2005*) Water science experts pay special attention to modeling and forecasting effective strategies in rainfall analysis and its effect on agricultural activities (*Hoogenboom, 2000*). Moreover, the results of a simulation model can be used to verify or correct the data (*Sentelhas et al., 2001*). Intelligent models, meanwhile, have shown a relatively high ability to simulate and predict nonlinear hydrological time series (*Nourani and Komasi, 2013*). These models include the gene expression programming (GEP), the adaptive fuzzy inference system (ANFIS), and the support vector machine (SVM) (*Dibike, 2006; Jang, 1993; Ferreira, 2001*).

The GEP model, developed by *Ferreira (2001)*, is an evolutionary method based on Darwin's theory of evolution and on the ability to simulate completely nonlinear and dynamic processes (*Alvisi et al., 2005*). The ANFIS structure is equivalent to a post-diffusion network that uses the neural network learning algorithm in combination with fuzzy reasoning to create a mapping between the input and output space (*Jang, 1993*). The SVM method is also one of the supervised learning methods that can be used for both classification and regression. This method is based on the Vapnik's method of statistical learning theory (*Vapnik, 1998*), and it is a method for binary classification in the space of desired properties. The SVM is essentially a two-class binder that separates classes by a linear boundary. In this method, the closest samples to the decision boundary are called support vectors (*Hamel, 2011*). The mentioned models have been used in water engineering so far, the following are some of the researches conducted on them. *Ustoorikar and Deo (2008)* used GEP to estimate incomplete data on wave heights in the Gulf of Mexico, and they found that the structure was accurate in predicting time series data. *Aytek and Kisi (2008)* used ANN (artificial neural network) and GEP methods to rainfall-runoff modeling in the Juniata River Basin in Pennsylvania, USA. Comparing their results showed that GEP performed better than ANN. The results of a study by *Shiri and Kisi (2011)* showed that, in comparing the predictive performance of short-term static level fluctuations using two models, GEP and ANFIS, both models perform well in predicting static level fluctuations, but the GEP model has a simpler structure than ANFIS. *Kavehkar et al. (2013)* simulated the water level fluctuations of Urmia Lake using GEP and ANN. The results showed the optimal GEP accuracy in simulating water level fluctuations. Comparing the performance of the Bayesian network and GEP models in daily river flow forecasting, *Baba Ali and Dehghani (2016)* showed that GEP forecasts are associated with lower error estimation. *Chang and Chang (2001)* used

ANFIS to predict reservoir inflow. According to their results, the predictions of the ANFIS model are more accurate than the classical models. *Choubey et al.* (2014) used the SVM model to predict and analyze the inflow of Narmada Reservoir Dam in the Indian state of Pragrah showing that this method has a very good ability to simulate and predict the average monthly flow. *He et al.* (2014) used the SVM model to predict river flow in mountainous and semi-arid regions in the northwestern part of China, and they found that SVM has better performance than ANN and ANFIS, to predict river flow in the semi-arid mountainous areas. *Tabari et al.* (2013) used SVM, ANFIS, regression, and meteorological models to simulate reference evapotranspiration using climate data. Their results showed the superiority of the SVM model over other methods. *Ahmadi et al.* (2014) used GP (genetic programming) and SVM to predict the monthly reference crop evapotranspiration. Their results showed that SVM was superior to GP. *Dehghani et al.* (2016) compared the performance of SVM, GEP, and Bayesian network models in predicting river flow. Their results showed that the SVM performed better with the least amount of error than the other models. *Mehdizadeh et al.* (2017) investigated the performance of empirical equations and soft computing approaches including GEP, SVM, as well as multivariate adaptive regression splines (MARS) in estimating monthly mean reference evapotranspiration. The performance of the SVM was better than the used empirical equations. *Hong et al.* (2018) coupled ANFIS with a genetic algorithm and differential evolution for flood spatial modeling. They combined two data mining techniques with the ANFIS model, including the ANFIS-Genetic Algorithm and the ANFIS-Differential Evolution. The result showed, the ANFIS-Differential Evolution hybrid model is more suitable for flood susceptibility mapping in their study area. *Kalantar et al.* (2018) assessed the training landslides random selection effects on support vector machine (SVM) accuracy, logistic regression (LR), and artificial neural network (ANN) models for landslide susceptibility mapping. Results showed that SVM and LR models performed better than ANN model. *Pashazadeh and Javan* (2020) compared the performance of GEP and ANN models with the equivalent Muskingum inflow model. Their results showed that the obtained outflow hydrograph by the GEP model had an excellent performance compared with ANN and equivalent Muskingum inflow models.

Due to the importance of rainfall simulation to analyze, verify, correct, and complete data and forecasts in future periods as well as the impact of long-term memory data on the performance of certain models, the aim of this study was to evaluate the efficacy and accuracy of the three models (GEP, ANFIS, and SVM) in rainfall simulation.

For this purpose, the monthly data of the 110-year-long time series of Diata station in India and the daily data of the 44-year-long time series of Urmia station have been used and the simulation performance has been evaluated using different criteria.

2. Materials and method

In this study, two series of monthly rainfall records (1900–2010) of Diata meteorological station (Shivapouri Region of Madhya State of India) and daily rainfall records (1961–2005) of Urmia synoptic station (Iran) were used. To perform the simulation, 90% of the data was used to train (learning phase) and the remaining 10% of the data was used to test (testing phase) the models, followed by running the SVM, ANFIS, and GEP algorithms and the evaluation criteria. *Fig. 1* shows the location of the Urmia and Diata meteorological stations in Iran and India, respectively.

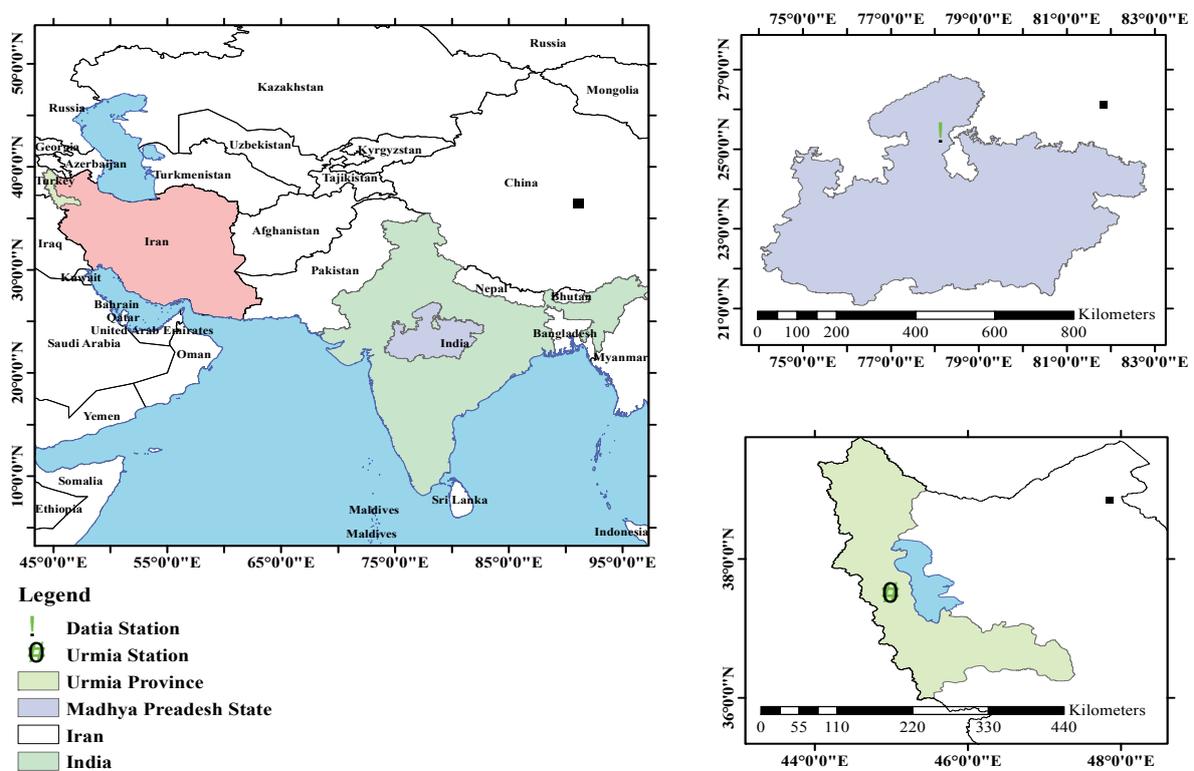


Fig. 1. Location map of the selected stations in Iran and India.

2.1. Gene expression programming (GEP)

This method is a combination of methods such as genetic programming (GP) and genetic algorithm (GA). In GEP, similarly to genetic algorithm, the individuals are encoded as linear strings of fixed length (the genome or chromosomes), and similarly to genetic programming, they are expressed as nonlinear entities of different sizes and shapes. In GEP, the genome or chromosome consists of a linear, symbolic string of fixed length composed of

one or more genes. It will be shown that despite their fixed length, GEP chromosomes can code expression trees (ETs) with different sizes and shapes. One of the strengths of GEP over GA and GP is that genetic operators' works at the chromosome level, which makes genetic diversity creation extremely simplified. The other strong point of GEP is its unique, multigeneric nature, which allows the evolution of more complex programs composed of several sub-programs (Ferreira, 2001). Further details on this method are provided by Shoaib et al. (2015). In this study, GEP simulation was performed using GeneXpro Tools (v5) software.

2.2. Adaptive-network-based fuzzy inference system (ANFIS)

The method of adaptive fuzzy-neural inference system is a combination method, in which the fuzzy part establishes the relationship between the input and output variables, and the characteristics related to the membership functions of the fuzzy part are determined by the neural network. This method has five layers according to Fig. 2, which are described in the following steps:

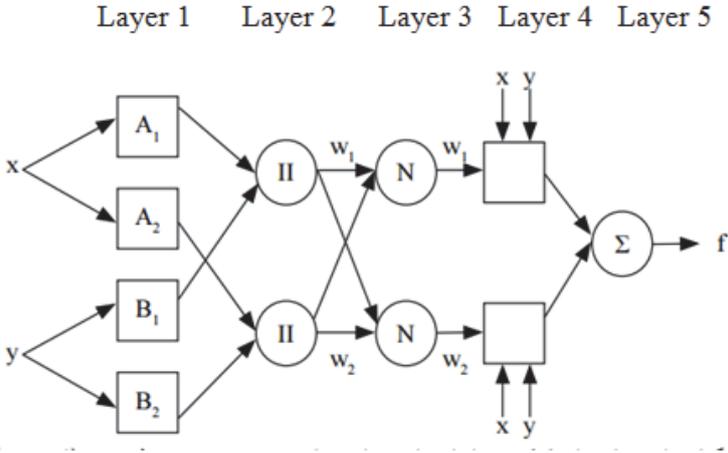


Fig 2. ANFIS network structure.

Layer 1, the input nodes: Every node in this layer acts as a member function that is assigned to each of the input variables of the model (x and y). Member amounts are determined based on the inputs belonging to each of the fuzzy sets A_i and B_i . In other words, the output of each node in this layer is the degree of members assigned to the input variables in the fuzzy sets, which are expressed as follows:

$$i = 1,2 \quad O_{1,i} = \mu_{A_i}(x), \quad (1)$$

$$i = 3,4 \quad O_{2,i} = \mu_{B_{i-2}}(x), \quad (2)$$

where x and y are the non-fuzzy inputs of node i , A_i , and B_i are fuzzy membership functions, and $O_{1,i}$ and $O_{2,i}$ represents the outputs of the first layer.

Layer 2, the rules nodes: Each node in this layer calculates the degree of activity of a rule. In this layer, the operator "and" is used to calculate the degree of participation of each rule. $O_{2,k}$ represents the output of the k -node in the second layer, and it is the product of the degrees of membership of each entry:

$$O_{2,K} = \mu_{A_i}(x) \cdot \mu_{B_j}(y). \quad (3)$$

Layer 3, the normalized nodes: This step calculates the ratio of the degree of participation of each rule to the total degree of participation of all the rules. As a result, this layer is defined as

$$O_{3,i} = \bar{w}_i = \frac{w_i}{\sum_{k=1}^4 w_k}, \quad (4)$$

where w_i is the i th output node of the previous layer.

Layer 4, the adaptive nodes: This step uses the result parameters to calculate the output of each node:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i), \quad (5)$$

where w_i is i th output node of the previous layer and $\{p_i, r_i, q_i\}$ are linear adaptive parameters.

Layer 5, the output node: This step expresses the final output value as the sum of the output nodes of the previous layer:

$$O_{5,i} = \bar{w}_i f_i = \frac{\sum_{i=1}^4 w_i f_i}{\sum_{i=1}^4 w_i}. \quad (6)$$

The ANFIS network learning algorithm is a hybrid learning algorithm consisting of a descending gradient algorithm and a minimum squared return method. The descending algorithm is used to update the network's nonlinear

parameters and the estimate of the minimum return square is used to adjust the network weights. Network training error is defined as:

$$E = \sum_{i=1}^N (f_i - \hat{f}_i)^2, \quad (7)$$

where f_i and \hat{f}_i are the optimal and estimated outputs of the network for the i th input, respectively, and N is the total number of input-output data pairs (training data) of the network (Jang, 1993).

2.3. Support vector machine (SVM)

SVM is a supervised learning model that is based on the constrained optimization theory and the method of structural risk minimization. In this model, the function related to the dependent variable Y , which is itself a function of several independent variables of x , is estimated. Similarly to other regression issues, it is assumed that the relationship between independent and dependent variables are specified with an algebraic function $f(x)$ plus an acceptable error ε .

$$f(x) = W^T \cdot \phi(x) + b, \quad (8)$$

$$y = f(x) + \varepsilon, \quad (9)$$

where the W (vector of the equations) and b (constant) are the characteristics of the regression, and ϕ is Kernel function. The purpose is to find the function of $f(x)$. This is achieved by using the SVM training phase. Therefore, to calculate W and b , it is necessary to optimize the error function (Eq. (10)) in the ε -SVM model by considering the conditions in Eq. (11):

$$\frac{1}{2} W^T W + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^*, \quad (10)$$

$$\begin{aligned} W^T \cdot \phi(x_i) + b - y_i &\leq \varepsilon + \xi_i^* \\ y_i - W^T \cdot \phi(x_i) - b &\leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* &\geq 0, \quad i = 1, \dots, N, \end{aligned} \quad (11)$$

where C is a positive integer that determines the penalty when the model training error occurs, ϕ is the Kernel function, N is the number of samples, and

the two characters ξ_i and ξ^*_i are slack variables determining the upper and lower limits of the training error associated with the allowable error value ε . In the case of predictions, the data are placed within the ε -boundary range. But, if the data is out of range, then there is an equivalent error ξ_i and ξ^*_i (Eskandari et al., 2012). After defining the Lagrangian coefficients, the characteristics of W and b in the regression SVM model are calculated using Crash-Cohen-Tucker theory, in which W is consistent with Eq. (12), and as a result, for SVM model, we will calculate Eq. (13):

$$W = \sum_{j=1}^N (a_i + a_i^*) \phi(x_i) \quad (12)$$

$$W = \sum_{i=1}^N (a_i + a_i^*) \phi(x_i)^T \phi(x) + b \quad (13)$$

where a_i and a_i^* are Lagrangian terms and $(a_i + a_i^*)$ can be zero or non-zero. Therefore, only the data sets whose coefficients a_i^* are assumed to be non-zero, are entered into the final regression equation. These data sets are known as the support vectors. Finally, the regression SVM function can be rewritten as

$$f(x) = \sum_{i=1}^N a_i \phi(x_i)^T \phi(x_j) + b \quad (14)$$

In Eq.(14), the calculation of $\phi(x)$ in the specified space may be very complex. To solve this problem, a kernel function is selected in the regression SVM model. Different kernel functions can be used to construct different types of ε -SVM models (Hofmann, 2002):

$$K(x_i, x) = \phi(x_i)^T \phi \sqrt{b^2 - 4ac} \quad (15)$$

2.4. Performance criteria

The correlation coefficient R2 (Eq. (16)), root mean square error (RMSE) (Eq. (17)), Nash-Sutcliffe efficiency coefficient NSE (Eq. (18)), and the bias or model bias MB (Eq. (19)) were used as performance criteria to evaluate the accuracy and efficiency of the used models.

$$R^2 = \left(\frac{\text{cov}(P_i, P'_i)}{\sigma_{P_i} \sigma_{P'_i}} \right)^2, \quad (16)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - P'_i)^2}{S}}, \quad (17)$$

$$NSE = 1 - \frac{\sum_{i=1}^T (P_i - P'_i)^2}{\sum_{i=1}^T (P_i - \bar{P}_i)^2}, \quad (18)$$

$$MB = \frac{\sum_{i=1}^n (P'_i - P_i)}{P_i}, \quad (19)$$

where P_i are the observed values, P'_i are the simulated values, \bar{P}_i are the mean observed values, S is the number of data, σ_{P_i} is the variance of observed values, $\sigma_{P'_i}$ is the variance of simulated values (Nash and Sutcliffe, 1970; Swinscow and Campbell, 2013; Salas et al., 1980). The flowchart of the proposed methodology is presented in Fig. 2.

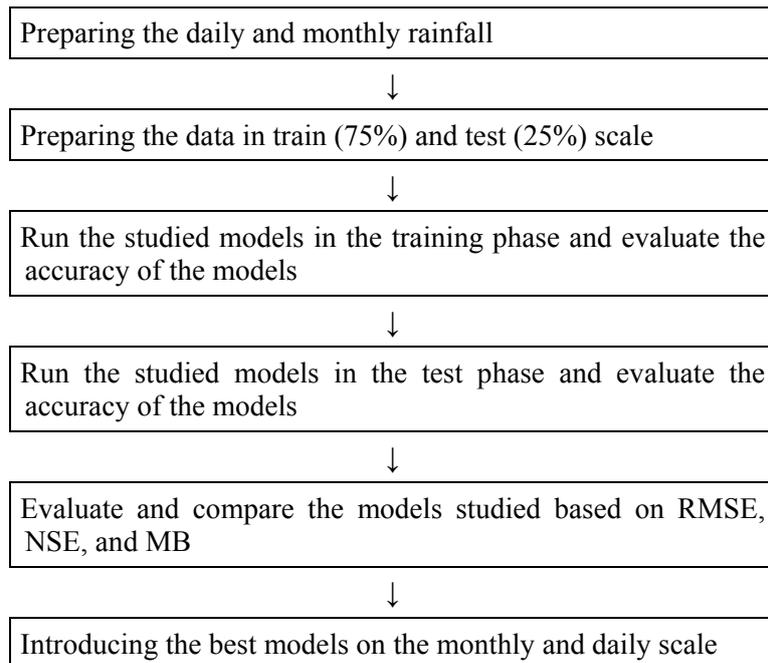


Fig 2. Flowchart of the proposed methodology

3. Results and discussion

Rainfall simulation has been demonstrated by GEP, ANFIS, and SVM models along with the correlation between observed and simulated value. The results are shown in *Figs. 4, 5, and 6*. According to these figures, in simulation with GEP, simulated rainfall is less than the observed values, and the correlation diagram of this simulation shows the relative scattering of simulated data, and the simulated values by ANFIS and SVM are more consistent with observational data. The correlation between the observed and simulated values of these two models is high and slightly different from each other. The data correlation in the ANFIS model is slightly higher than in the SVM. The correlation coefficient in simulation with GEP is estimated to be 55% lower than the other two models, and the R^2 range in both ANFIS and SVM models varies from 91 to 93%.

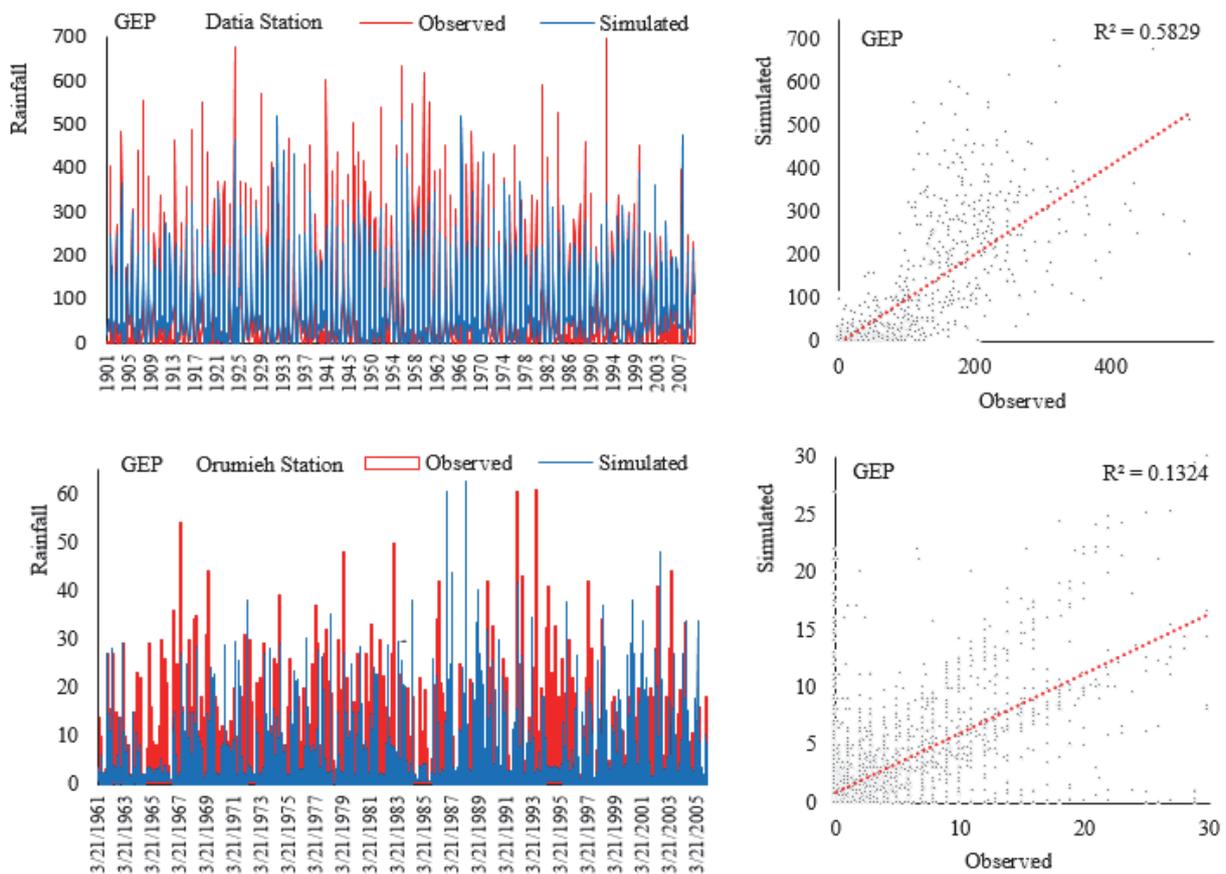


Fig 4. Rainfall simulation and the correlation between observational data and simulation in the GEP model (mm).

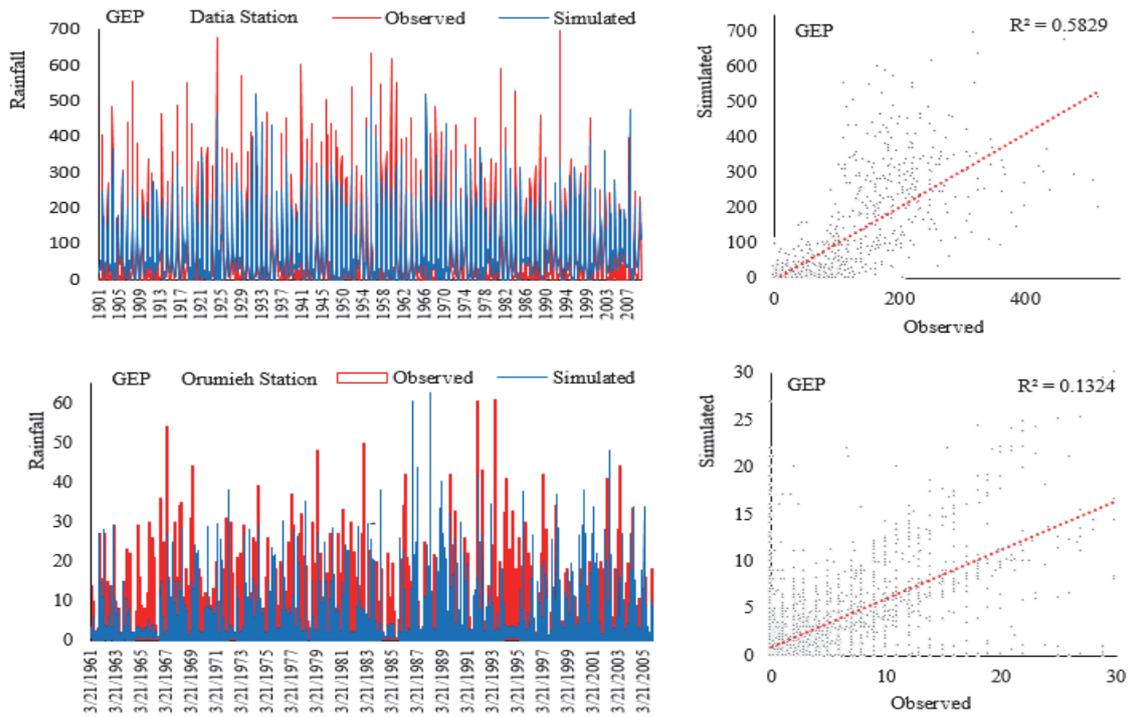


Fig 5. Rainfall simulation and the correlation between observational data and simulation in the ANFIS model (mm)

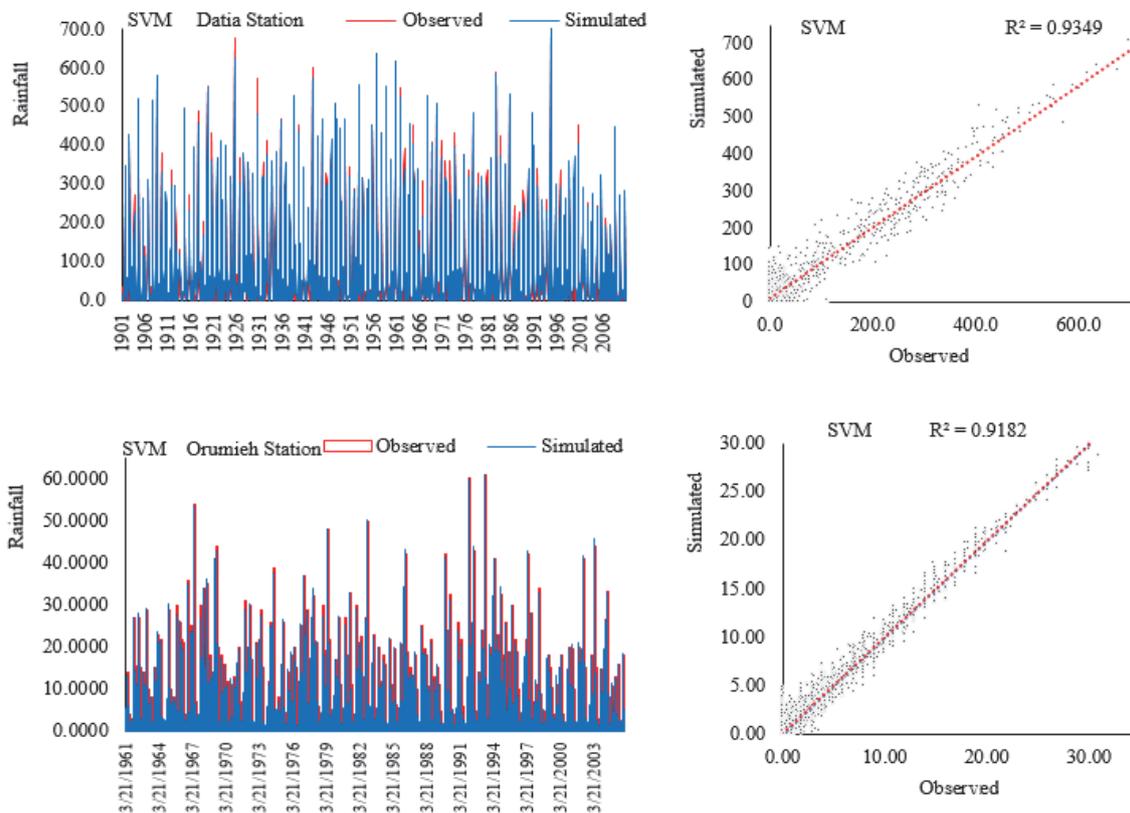


Fig 6. Rainfall simulation and the correlation between observational data and simulation in the SVM model (mm)

3.1. Evaluation of the performance of simulation models

The evaluation of GEP, ANFIS, and SVM models using RMSE, MB, and NSE criteria are presented in *Table 1*. The comparison of simulations based on daily and monthly rainfall data shows that the root mean square error (RMSE) in simulation with daily data is less than with monthly data and the model bias (MB) is more in daily data compared to monthly. The Nash-Sutcliffe efficiency (NSE) criteria of models with daily data are higher than monthly, the details of which can be examined in the table. On average, the root mean square error (RMSE) in the GEP simulation is 50 and 55% higher than that of the ANFIS for Urmia and Diata stations, respectively. However, ANFIS is superior to SVM in the training phase and performs worse in the test phase. But on average, the RMSE of the ANFIS model is 1 and 3% higher than that of the SVM at Urmia and Diata stations, respectively. The bias model of GEP is 72 and 60% higher than that of ANFIS at Urmia and Diata stations, respectively. The ANFIS bias is 62 and 15% higher than the SVM bias at Urmia and Diata stations, respectively. The GEP's Nash-Sutcliffe efficiency (NSE) criteria are 56 percent and 61% lower than the NSE of ANFIS at Urmia and Diata stations, respectively. The NSE of ANFIS is 1 and 2% lower than the NSE of SVM in Urmia and Diata stations, respectively.

Table 1. Evaluating the performance of simulation models GEP, ANFIS, and SVM

Model	Station	Index					
		NSE (%)		MB (mm)		RMSE (mm)	
		Test	Train	Test	Train	Test	Train
GEP	Urmia	27	31	0.672	2.13	2.647	3.389
	Diata	21	24	0.605	0.0276	70.558	81.327
ANFIS	Urmia	88	92	0.662	0.111	0.998	0.997
	Diata	83	93	0.145	0.106	35.592	32.453
SVM	Urmia	90	92	0.291	0.358	0.952	1.007
	Diata	88	93	0.099	0.189	32.702	33.002

By comparing the results of this study with the result of other researchers, it can be shown that the result is consistent with the results of *Ahmadi et al.* (2014) in the superiority of the SVM model over GEP to estimate reference evapotranspiration over a 37-year-long period. Results are consistent also with the results of *Tabari et al.* (2013), that they simulated the reference crop transpiration by using SVM and ANFIS, and showed the relative superiority of SVM. On the other hand, results of *Shiri et al.* (2013), which used ANN, ANFIS, SVM, and GEP

methods to predict rainfall and evaporation parameters to predict groundwater level fluctuations, are inconsistent with the research results of the present study showing better GEP efficiency. The reason for the inefficiency of ANFIS and SVM in this study is related to the limited groundwater data due to their nature. Also, they used eight-year time series data in their research. Therefore, if the data used is limited, other models may be as efficient as ANFIS and SVM, which require an evaluation through different indicators.

4. Conclusion

Based on the performance criteria, evaluating efficiency, and accuracy of three models, GEP, ANFIS, and SVM in daily and monthly rainfall simulation, the correlation coefficient of rainfall simulation with daily data is higher than that of the simulations with monthly data in all three models. The correlation in GEP simulation is significantly lower than in the other two models, and ANFIS performs 1% better than SVM with daily and monthly data. According to the RMSE criterion, the use of daily data performs better than the use of monthly data in all three models, GEP simulation performs worse than the other two models, and ANFIS has a slight superiority over SVM, which requires further research. In all three simulations, the model bias (MB) with daily data is higher than monthly, and the simulation bias in GEP is 76% and 83% higher than the bias in ANFIS and SVM, respectively. The efficiency of all three models performs better in simulation with daily data than with monthly data, and the efficiency of the GEP model is 27% better on average according to this criterion, and SVM has about 4% better performance than that of ANFIS. Therefore, the SVM rain simulation is associated with a lower error rate and better efficiency, the ANFIS model has the same capability as the SVM, and the GEP model is not suitable for simulating rainfall. Therefore, the higher MB and RMSE, as well as less R^2 and NSE of GEP model than the other two models show poor GEP performance compared to ANFIS and SVM. Due to the slight advantage of ANFIS over SVM based on R^2 and RMSE criteria and the relative superiority of SVM over ANFIS based on MB and NSE criteria, SVM can be introduced as the superior model.

References

- Ahmadi, F., Radmanesh, F., MirAbbasi, R., and Aishm, S., 2014: Prediction of Monthly Evaporation and Transpiration of Reference Plant in the Northwest of the Country Using Genetic Programming and Support Vector Machine. J. Iranian Irrigation Drainage 1, 65–54.*
- Alvisi, S., Mascellani, G., Franchini, M., and Bardossy, A., 2005: Water Level Forecasting Through Fuzzy Logic and Artificial Neural Network Approaches. J. Hydrol. Earth Syst. Sci. 2, 1107–1145. <https://doi.org/10.5194/hessd-2-1107-2005>*

- Aytek, A. and Kisi, O., 2008: A Genetic Programming Approach to Suspended Sediment Modeling. *J. Hydrology* 351, 288–298. <https://doi.org/10.1016/j.jhydrol.2007.12.005>
- Babaali, H.R, and Dehghani, R., 2016: Comparison of the Performance of Business Network Models and Gene Expression Programming in Predicting Daily River Flow (Case Study: Marbara River). *J. Environ. Sci. Technol.* 2(2), 96–108.
- Chang, L.C. and Chang, F.J., 2001: Intelligent Control for Modelling of Real-Time Reservoir Operation. *J. Hydrolog. Proc.* 15, 1621–1634. <https://doi.org/10.1002/hyp.226>
- Choubey, V., Mishra, S., and Pandey, S.K., 2014: Time Series Data Mining in Real Time Surface Runoff Forecasting Through SVM. *Int. J. Comput. Appl.* 98(3), 23–30. <https://doi.org/10.5120/17163-7223>
- Dehghani, R., Younesi, H., and Torabi Poodeh, H., 2016: Comparison of the Performance of Support Vector Machine Models, Gene Expression Programming and Business Network in Predicting River Flow (Case Study: Kashkan River). *J. Water Soil Conserv. Res.* 24(4), 161–177.
- Dibike, B.Y. and Coulibaly, P., 2006: Temporal Neural Networks for Downscaling Climate Variability and Extremes. *J. Neural Networks* 19, 135–144. <https://doi.org/10.1016/j.neunet.2006.01.003>
- Eskandari, A.S., Ruhollah, N., Mearaji, H., and Kiaqadari, A., 2012: Development of an Appropriate Model Based on Artificial Neural Network and SVM for Predicting 5-day Biochemical Biochemistry. *J. Environ.* 38, 71–82.
- Ferreira, C., 2001: Gene Expression Program-Ming a New Adaptive Algorithm for Solving Problems. *J. Complex Syst.* 13(2), 87–129.
- Hamel, L.H., 2009: Knowledge Discovery with Support Vector Machines. Wiley and Sons. <https://doi.org/10.1002/9780470503065>
- He, Z., Wen, X., Liu, H., and Du, J., 2014: A Comparative Study of Artificial Neural Network, Adaptive Neuro Fuzzy Inference System and SVM for Forecasting River Flow in the Semiarid Mountain Region. *J. Hydrology* 509, 379–386. <https://doi.org/10.1016/j.jhydrol.2013.11.054>
- Hofmann, T., Tsochantaridis, I., and Altun, Y., 2002: Learning Over Structured Output Spaces Via Joint Kernel Functions. Sixth Kernel Workshop.
- Hong, H., Panahi, M., Shirzadi, A., Ma, T., Liu, J., Zhu, A.X., Chen, W., Kougiyas, L., and Kazakis, N., 2018: Flood Susceptibility Assessment in Hengfeng Area Coupling Adaptive Neuro-Fuzzy Inference System With Genetic Algorithm and Differential Evolution. *Sci. Total Environ.* 621, 1124–1141. <https://doi.org/10.1016/j.scitotenv.2017.10.114>
- Hoogenboom, G. 2000: Contribution of Agrometeorology to the Simulation of Crop Production and Its Applications. *J. Agricult. Forest Meteorol.* 103, 137–157. [https://doi.org/10.1016/S0168-1923\(00\)00108-8](https://doi.org/10.1016/S0168-1923(00)00108-8)
- Jang, J.S.R., 1993: ANFIS: Adaptive- Network- Based Fuzzy Inference Systems. IEEE Transactions on Systems. *Man Cybernetics* 23, 665–685. <https://doi.org/10.1109/21.256541>
- Kalantar, B., Pradhan, B., Naghibi, S.A., Motevalli, A., and Mansor, S., 2018: Assessment of the Effects of Training Data Selection on the Landslide Susceptibility Mapping: A Comparison Between Support Vector Machine (SVM), Logistic Regression (LR) and Artificial Neural Networks (ANN). *J. Geomat. Nat. Hazards Risk* 9, 49–69. <https://doi.org/10.1080/19475705.2017.1407368>
- Kavehkar sh, Ghorbani, M., Ashrafzadeh, A., and Darbandi, P., 2013: Simulation of Water Level Fluctuations Using Gene Expression Programming. *J. Civil Engin. Environ.* 43(3), 72–75.
- Mehdizadeh, S., Behmanesh, J., and Khalili, K., 2017: Using MARS, SVM, GEP and Empirical Equations for Estimation of Monthly Mean Reference Evapotranspiration. *Comput. Electronics Agricult.* 139, 103–114. <https://doi.org/10.1016/j.compag.2017.05.002>
- Nash, J.E. and Sutcliffe, J.V., 1970: River flow forecasting through conceptual models part I-A discussion of principles. *J. Hydrology* 10, 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nourani, V. and Komasi, M., 2013: A geomorphology-based ANFIS model for multistation modeling of rainfall–runoff process. *J. Hydrology* 490, 41–55. <https://doi.org/10.1016/j.jhydrol.2013.03.024>
- Pashazadeh, A., and Javan, M., 2020: Comparison of the gene expression programming, artificial neural network (ANN), and equivalent Muskingum inflow models in the flood routing of multiple branched rivers. *Theor. Appl. Climatol.* 139, 1349–1362. <https://doi.org/10.1007/s00704-019-03032-2>

- Refsgaard, J.C., Henriksen, H.J., Harrar, W.G., and Scholten Kassahun, A., 2005: Quality assurance in model based water management - review of existing practice and outline of new approaches. *J. Environ. Modell. Software*. 20, 1201–1215. <https://doi.org/10.1016/j.envsoft.2004.07.006>
- Refsgaard, J.C. and Knudsen, J., 1996: Operational validation and intercomparison of different types of hydrological models. *J. Water Resour. Res.* 32, 2189–2202. <https://doi.org/10.1029/96WR00896>
- Salas, J.D., Delleur, J.W., Yevjevich, V., and Lane, W.L., 1980: Applied Modeling of Hydrologic Time Series. Water resource Publications. Littleton, Colorado . U.S.A.
- Sentelhas, P.C., deFaria, R.T., Chaves, M.O., and Hoogenboom, G., 2001: Evaluation of the WGEN and SIMMETEO weather generators for the Brazilian tropics and subtropics, using crop simulation models. *Revista Brasileira de Agrometeorologia. Santa Maria.*; 9, 357–376.
- Shiri, J., Kisi, O., Yoon, H., Lee, K.K., and Hossein Nazemi, A., 2013: Predicting groundwater level fluctuations with meteorological effect implications . A comparative study among soft computing techniques. *J. Comput. Geosci.* 56, 32–44. <https://doi.org/10.1016/j.cageo.2013.01.007>
- Shiri, J. and Kisi, O., 2011: Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. *J. Comput. Geosci.* 37, 1692–1701. <https://doi.org/10.1016/j.cageo.2010.11.010>
- Shoaib, M., Shamseldin, A.Y., Melville, B.W., and Khan, M.M., 2015: Runoff Forecasting using Hybrid Wavelet Gene Expression Programming(WGEP) Approach. *J. Hydrology.* 527, 326–344. <https://doi.org/10.1016/j.jhydrol.2015.04.072>
- Swinscow, T.D.V. and Campbell, M.J., 2013: Statistics at Square One. London: BMJ Publication.
- Tabari, H., Martinez, C., Ezani, A., and Hosseinzadeh Talaei, P., 2013: Applicability of SVMs and adaptive neuro- fuzzy inference system for modeling potato crop evapotranspiration. *J. Irrigat. Sci.* 31, 575–588. <https://doi.org/10.1007/s00271-012-0332-6>
- Ustoorikar, K. and Deo, M.C., 2008: Filling up Gaps in Wave Data with Genetic Programming. *J. Marine Struct.* 21, 177–195. <https://doi.org/10.1016/j.marstruc.2007.12.001>
- Vapnik, V.N., 1998: Statistical Learning Theory. Wiley, New York.