ORIGINAL PAPER

CrossMark

A hybrid support vector regression–firefly model for monthly rainfall forecasting

A. Danandeh Mehr[1,3] · V. Nourani[2,3] · V. Karimi Khosrowshahi[4] · M. A. Ghorbani[4]

Abstract

Long-term prediction of rainfalls is one of the most challenging tasks in stochastic hydrology owing to the highly random characteristics of rainfall events. In this paper, a novel approach is adopted to develop a hybrid regression model for 1-month-ahead rainfall forecasting at two rain gauge locations (namely: Tabriz and Urmia stations), in northwest Iran. The approach is based on the integration of support vector regression (SVR) and firefly algorithm (FFA) that results in truthful rainfall forecasts. The proposed hybrid model was trained and validated using weak stationary state of monthly rainfall data obtained from the gauges. The efficiency results of the model were also cross-validated with those of stand-alone SVR- and genetic programming-based forecasting models developed as the benchmarks in this study. For both rain gauge locations, the results showed that the hybrid model significantly outperforms the benchmarks. With respect to the average efficiency results at the gauge locations, the FFA-induced improvement in the SVR forecasts was matched by an approximately 30% decrease in root-mean-square error and around 100% increase in Nash–Sutcliffe efficiency. Such a promising accuracy in the proposed model may recommend its application at monthly rainfall forecasting in the present semiarid region.

Introduction

Accurate monthly rainfall forecast is required for a number of tasks in water resource management such as food production, water allocation, and flood risk mitigation. However, it is one of the most scientifically and technologically challenging problems in stochastic hydrology (Mekanik et al. 2013; Feng et al. 2014). Classic time series modeling such as autoregressive integrated moving average (ARIMA)

✉ A. Danandeh Mehr
  ali.danandeh@antalya.edu.tr

1 Department of Civil Engineering, Faculty of Engineering, Antalya Bilim University, Antalya, Turkey

2 Department of Water Resources Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

3 Department of Civil Engineering, Faculty of Engineering, Near East University, Mersin 10, Nicosia, Turkey

4 Department of Water Engineering, University of Tabriz, Tabriz, Iran

and seasonal ARIMA (SARIMA) have been performed for monthly rainfall forecasting in earlier studies (e.g., Delleur and Kavvas 1978). Although these models are applied for stationary state of rainfall series, they are basically linear models and have a limited ability to capture highly nonlinear characteristics of rainfall series. Moreover, they cannot be used for long lead time forecasts and have limited capability in monthly rainfall forecasting (Nourani et al. 2009; Delleur and Kavvas 1978).

Nowadays, artificial intelligence (AI) methods such as artificial neural network (ANN), fuzzy logic (FL), support vector machine (SVM), and genetic programming (GP) are widely used for rainfall forecasting (e.g., Aksoy and Dahamsheh 2009; Nourani et al. 2009; Wu et al. 2010; El-Shafie et al. 2011; Moustris et al. 2011; Abarghouei and Hosseini 2016). In spite of desired flexibility of AI methods for time series forecasting, recent studies have shown that they are not effective methods for long-term rainfall forecasting, particularly in arid and semiarid regions. Accordingly, application of hybrid AI models such as wavelet–ANN (Nourani et al. 2009), wavelet–SVM (Kisi and Cimen 2012), and adaptive network-based fuzzy inference system (Mekanik

et al. 2016) has been suggested. For example, Sivapragasam et al. (2001) improved SVM forecasts using singular spectrum analysis that decomposes original rainfall series into a set of high- and low-frequency components. Nasseri et al. (2008) and Saxena et al. (2014) optimized ANN model using genetic algorithm (GA) for short-term and long-term rainfall forecasting, respectively. Nourani et al. (2009) developed a wavelet–ANN coupled model for 1-month-ahead rainfall forecasting at Ligvanchai Catchment in Iran and reported that the hybrid wavelet–ANN model can accurately predict rainfalls in both short- and long-term basis in their study region. The reason behind such performance was mainly due to the use of wavelet-based multiscale signals that were utilized as the ANN input vectors. Kisi and Cimen (2012) investigated the performance of wavelet–SVM conjunction model for 1-day-ahead rainfall forecasting in Izmir and Afyon stations in Turkey and demonstrated that the conjunction method increases the forecast accuracy and performs better than the ad hoc SVM and ANN models. Solgi et al. (2014) combined wavelet transform with ANN to forecast daily rainfall with 1-day lead time in Verayneh station, Iran, and the hybrid model was compared with adaptive neuro-fuzzy interface system (ANFIS). The results showed that the WA-ANN is superior to the ANFIS. More recently, a new optimization approach called firefly algorithm (FFA) algorithm was used by Yaseen et al. (2017b) to improve efficiency of ANFIS networks for 1-month-ahead rainfall forecasts in Pahang River catchment, Malaysia. The authors compared efficiency of their hybrid model (ANFIS-FFA) with that of the stand-alone ANFIS using different statistical indices and explained that the ANFIS-FFA outperformed the ANFIS and could be adopted for the simulation of monthly rainfall in their study region.

Both short-term and long-term forecast of rainfalls are required to plan, operate, and optimize the activities associated with water resource systems. Each of them has their own benefits and applications. One-month-ahead forecast, as a long-term period, is beneficial for many watershed management applications such as food production, environmental protection, drought management, and optimal reservoir operation (Mekanik et al. 2013). However; short-term forecasts, with lead times of hours (up to 24 h) are necessary to drive hydrologic models, flood-warning systems, real-time reservoir operation, and others. The majority of the aforementioned studies showed that hybrid models can be effectively used for short-term forecast of rainfalls. However, only a few studies have investigated the possible application of hybrid models for forecast of rainfalls on a long-term basis (e.g., Mekanik et al. 2013; Yaseen et al. 2017b), and therefore, more studies are still required to identify unpredictable nature of rainfall events. The objective of this study is to investigate the possible application of a new hybrid model, namely SVR–FFA, for rainfall forecasting with

1-month lead time at heavily localized areas. To achieve this objective, power of FFA optimizer has been used to improve the support vector regression (SVR) estimations. Two rain gauge stations in northwest of Iran, each having 25 years observations, are chosen as the case studies. Performance of the proposed model has been compared with those of the stand-alone SVR and multigene GP (MGGP) models developed as the benchmarks in this study. Although the approach adopted in constructing the proposed hybrid model (i.e., SVR–FFA) has been recently applied for different hydrological forecasting problems such as solar radiation forecasting (Shamshirband et al. 2016), velocity estimation in sewer pipes (Ebtehaj and Bonakdari 2016), and field capacity prediction (Ghorbani et al. 2017), to the best of the authors' knowledge, the potential of SVR–FFA to forecast monthly rainfall has never been explored so far. This study has been carried out partly at university of Tabriz and partly at Near East University in 2017.

Materials and methods

Study area and observed data

The proposed SVR–FFA model is trained and tested using the total monthly rainfall (TMR) data from two rainfall gauges, Tabriz and Urmia stations, located in a semiarid region, northwest of Iran (Fig. 1 left). The data were obtained from the Iran Meteorological Organization (IRIMO; www.irimo.ir). Mean annual rainfall in the study region is about 250 mm with a maximum rate commonly in spring months (March to May). Time series of 25 years observed TMR data (January 1990 to December 2014) at the rain gauge stations are presented in Fig. 1 (right). The geographical information of the stations and their statistical characteristics are presented in Table 1.

Support vector regression (SVR)

The state-of-the-art SVM (Cortes and Vapnik 1995) is a machine learning technique that learns through examples to find the best function of classifier/hyperplane to separate the two classes in the input space. SVMs are derived from the structural risk minimization principle in order to increase the generalization capability on the learning machine and decrease both the empirical risk and the confidence interval of the machine (Raghavendra and Deka 2014). The SVM has been proven to be a robust and efficient technique for both classification and regression problems in hydrology (e.g., Kisi and Cimen 2012; Nourani and Andalib 2015; Olyaie et al. 2017; Gizaw and Gan 2016). For the case of regression problem, using the sets of x and y as the input and output
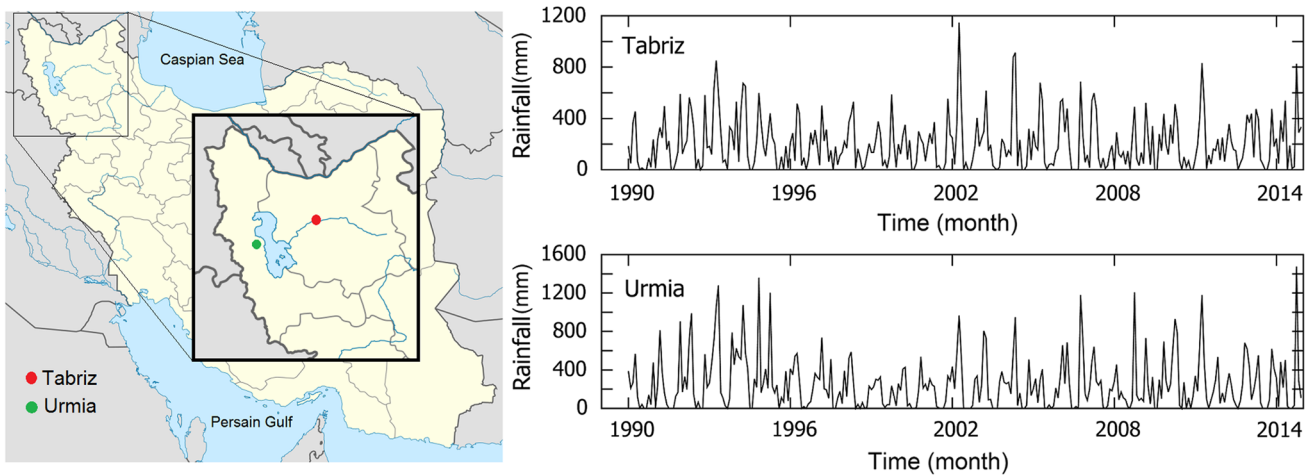
Fig. 1 Location of the rainfall gauges used in this study (Left) and observed total monthly rainfall data at Tabriz and Urmia stations for the period 1990–2014 (Right)

Table 1 Geographic and descriptive statistics using the observations 1990:01–2014:12

Stations	Latitude (°N)	Longitude (°E)	Elevation (MSL)	Min (mm)	Max (mm)	Mean (mm)	CV*
Tabriz	38.05	46.17	1345	0.00	1148.0	204.7	0.990
Urmia	37.40	45.03	1328	0.00	1475.0	257.2	1.110

*Coefficient of variation

space vectors, the SVM regression (hereafter SVR) function is expressed as (Raghavendra and Deka, 2014):

$$f(x) = w \times \phi(x) + b \qquad (1)$$

where w is a weight vector, b is a scalar called bias term, and ϕ represents a transfer function. This regression problem can be considered as the following convex optimization problem in order to get an appropriate SVR function $f(x)$.

$$
\begin{aligned}
\min \quad & \tfrac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \\
& y_i - (w \times \phi(x_i) + b) \le \varepsilon + \xi_i \\
\text{subject to} \quad & (w \times \phi(x_i) + b) - y_i \le \varepsilon + \xi_i^* \\
& \xi_i, \xi_i^* \ge 0, \quad i = 1, 2, \dots, n
\end{aligned}
\qquad (2)
$$

where ξ_i and ξ_i^* are slack variables introduced to evaluate the deviation of training samples outside ε-insensitive zone (deviation from the observed target). The parameter C is regularization factor that determines the trade-off between the flatness of $\phi(x)$ and the amount up to which the deviation ε can be tolerated (Sivapragasam et al. 2001).

In hydrological applications, the relationship between the input and output spaces is typically nonlinear. Therefore, data correlation is found through a nonlinear mapping methodology that SVR maps the input space on to some higher-dimensional space (feature space) using a kernel function. Sigmoid function, polynomial function, and radial

basis function (RBF) are from the basic kernel functions commonly used by SVR (Shamshirband et al. 2016). The RBF (Eq. 3) is an efficient and adaptable computation for the purpose of optimization (Sivapragasam et al. 2001; Kazem et al. 2013) Accordingly, it is adopted as the kernel function in the present study.

$$K(x, x_i) = \exp\left(-\|x - x_i\|^2 / 2\gamma^2\right) \qquad (3)$$

where x and x_i are input space vectors and the γ is the kernel function parameter. It is worth mentioning that the accuracy of the SVR function $f(x)$ is dependent on the selection the regularization factor C, the kernel parameter γ, and insensitive error term ε. Hence, a large number of trials must be carried out in order to find the optimum combination of these parameters. To cope with this problem, optimal values of these factors are determined using the optimization algorithm FFA in the present study. Principles of FFA are described in the following subsection. For details about SVR, the interested reader is referred to Cortes and Vapnik (1995).

SVR optimization using firefly algorithm (FFA)

The proposed SVR-FFA is a two-phase (i.e., simulation and optimization phases) hybrid model (see Fig. 2) that uses FFA optimizer (Yang et al. 2012) to determine optimum
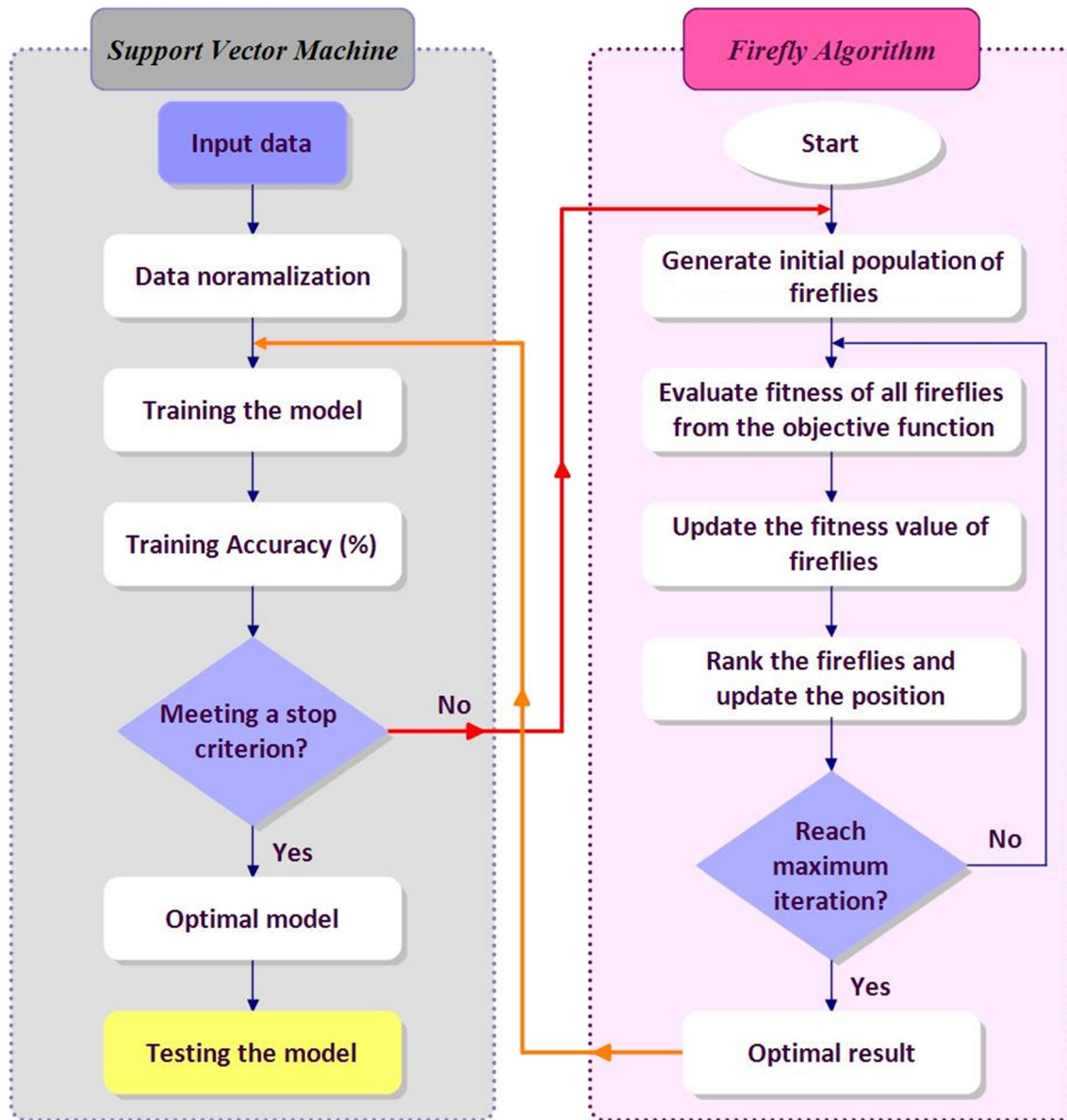
Fig. 2 Schematic structure of the proposed SVR–FFA model

values of SVR parameters (C, γ, ε) for monthly rainfall forecasts. The FFA is rather a new optimization technique from swarm intelligence optimization family that inspired by the movement of fireflies and their flashing characteristics. In hydrological studies, the technique has been successfully implemented to improve efficiency of ANFIS (e.g., Yaseen et a1. 2017a, b), ANN (e.g., Deo et al. 2018) and SVM (e.g., Ghorbani et al. 2017) methods. In FFA, the collective mind-set follows often a very simple rule tested over and over and maintained through the collective mind-set of the community. The solution of an optimization problem can be developed by treating each individual firefly as an agent, in which each firefly glows in proportion to its quality but is

attracted to the brightest firefly, regardless of its gender. This makes exploration within the search space more efficient than conventional search methods for most of the optimization problems under the presumption of swarm intelligence (Yaseen et a1. 2017a). Fireflies are attracted toward light so that the entire swarm moves toward the brightest (the most attractive) firefly. Thus, the attractiveness of a firefly is directly relative to its brightness, and therefore, fireflies come together more closely around the brightest firefly (i.e., optimal solution). The algorithm in FFA is formulated by developing the objective function and the variation of the light intensity. The Cartesian distance between any two fireflies x_i and x_j (i.e., r_{ij}), their light intensity at distance r (i.e.,

$I(r)$), and attractiveness β at distance r (i.e., $\beta(r)$) are mathematically expressed as (Yang et al. 2012):

$$r_{ij} = \left\| x_i + x_j \right\| = \sqrt{\sum_{k=1}^{d} x_{i,k} - x_{j,k}} \qquad (4)$$

$$I(r) = I_O \exp(-r^2) \qquad (5)$$

$$\beta(r) = \beta_O \exp(-\gamma^* r^2) \qquad (6)$$

where γ^* is the light absorption coefficient, I_0 is the initial light intensity of a firefly, and β_0 is the attractiveness at the distances $r = 0$. The next movement of firefly i can be represented as:

$$x_i^{i+1} = x_i + \Delta x_i \qquad (7)$$

$$\Delta x_i = \beta_0 e^{\gamma^* . r^2}(x_j - x_i) + \alpha \varepsilon_i \qquad (8)$$

where α is coefficient of randomization having a value is between 0 and 1 and ε_i is the random vector that is commonly derived from a Gaussian distribution (Yang et al. 2012).

Multigene genetic programming (MGGP)

GP (Koza 1992) is a generalization for genetic algorithm (GA). Since in GA one can manipulate chromosomes that represent binary, integer, or real numbers, it is possible to create chromosomes, which represent computer programs, that conduct evolutionary operations, and thus, they would get a program, which solves a particular problem. Among environmentalists, GP is frequently used as a self-structuring AI technique to generate alternative programs to identify the underlying system of a environmental process (e.g., Danandeh Mehr et al. 2014b; Shirani Faradonbeh et al. 2016). Three major evolutionary operations that conduct GP from initial population of random programs (GP trees) to a set of desired programs are: *reproduction*, *crossover*, and *mutation*. Reproduction is copying an existing population into the new population without alteration. Crossover is replacing chromosomes between desirable parents to produce two offspring and mutation is replacing a randomly selected node (function or terminal) or chromosomes with another node/chromosomes from the initial population (Danandeh Mehr and Nourani 2017). Figure 3 illustrates an example of crossover and mutation operators to generate two offspring. The dashed lines in the parents represent the crossover point selected randomly. The third offspring in the figure shows a mutated tree in which a terminal node in the first offspring (i.e., x_1) has been selected randomly and replaced with a new terminal node (i.e., x_2). The add3, mult3, and sqrt, in the figure, respectively, denote addition with tree argument, multiplication with three arguments, and root square functions.

Several advancements for the classic GP (i.e., monolithic GP) such as linear GP, gene expression programming, multigene GP, and fixed length gene GP have been suggested in recent studies. In parallel, successful implementations of these methods in hydrology have been reported (e.g., Hashmi et al. 2011; Uyumaz et al. 2014; Akbari-Alashti et al. 2015; Danandeh Mehr and Demirel 2016; Ravansalar et al. 2017). MGGP (Searson 2015) is one of the most recent advancements of GP that linearly combines low-depth GP trees in order to improve fitness of the classic GP. In MGGP, predictand variable is
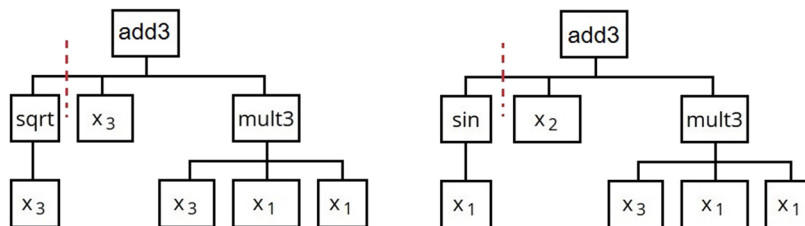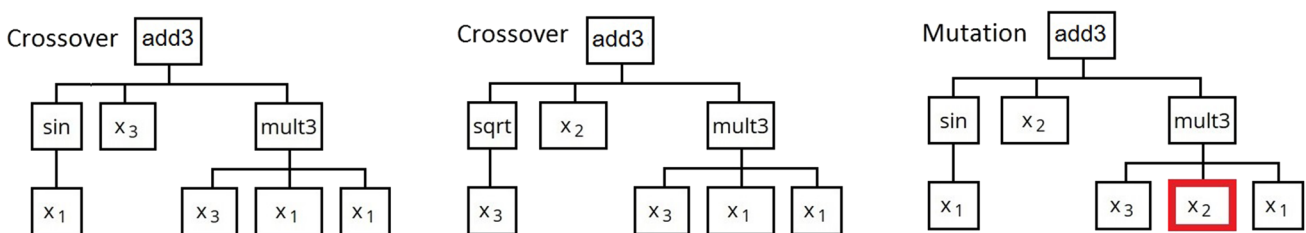
Fig. 3 An example of crossover and mutation operators acting on GP trees called parents. The dashed lines represent the crossover points in the parents

computed by the weighted outcome of each gene in the multigene chromosome plus a constant value called irregular term (bias). For example, a pseudo-linear MGGP chromosome shown in Fig. 4 predicts the predictand variable Y using three genes comprising three input variables X_1, X_2, and X_3. Mathematically, this chromosome can be expressed as:

$$Y = d_0 + d_1 \times 5.25(X_1 - X_2) + d_2 \times (X_1 \\ - X_2(X_3 - X_1)X_3) + d_3 \times ((X_1 - X_3) + X_2) \quad (9)$$

where d_0 is bias term and d_1, d_2, and d_3 are the gene weights (i.e., regression coefficients) that are typically determined using ordinary least-squares method. Therefore, it can be inferred that the MGGP employs the power of classical linear regression method to capture nonlinear behavior of the phenomenon of interest. For details on the theory and applications of MGGP, interested readers are referred to Searson (2015).

Efficiency criteria

Nash–Sutcliffe efficiency (NSE) and root-mean-square error (RMSE) have been used as the efficiency criteria in this study. These are statistical indices used in the majority of hydrological modeling studies (e.g., Danandeh Mehr et al. 2014a, 2015; Jalalkamali et al. 2015). The NSE (Eq. 10) is a normalized statistic that determines the relative magnitude of estimation error against the variance of observed data. It indicates how well the predicted and observed data fits the 1:1 line. The RMSE (Eq. 11) is a quadratic scoring rule which represents the average magnitude of the error. The error is the difference of predicted value from the corresponding predictand. Higher values for NSE (one for the perfect model) and lower values for RMSE (zero for the perfect model) stand for more efficient models

$$NSE = 1 - \frac{\sum_{i=1}^{n} \left(X_i^{obs} - X_i^{pre} \right)^2}{\sum_{i=1}^{n} \left(X_i^{obs} - X_{mean}^{obs} \right)^2} \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_i^{obs} - X_i^{pre})^2}{n}} \quad (11)$$

where X_i^{obs} = observed value of X (here monthly rainfall), X_i^{pre} = predicted value X_{mean}^{obs} = mean value of observed data, and n is the number of observed data.

In addition to these measures, Taylor diagram (Taylor 2001) was considered to examine concurrently the relative importance of multiple aspects such as the correlation coefficient between observed and predicted rainfall, root-mean-square centered difference and standard deviation. Importantly, Taylor diagram can highlight the goodness of different models compared to the observations within the same visual graph and, therefore, will complement the validity of the statistical metrics outlined in Eqs. (10) and (11).

Development of prediction scenarios

Identification of optimal lags (i.e., dominant input space vectors) is an important task in time series forecasting using any machine learning techniques. Optimal set of lags may lead a technique to create a parsimonious model. In contrast, insufficient or redundant lags will produce poorly performing or highly complex models (Danandeh Mehr and Nourani 2017). Autocorrelation function (ACF) and/or partial autocorrelation function (PACF) of time series is commonly used in identification of optimal lag for a time series forecasting model. Figure 5 shows the ACF, PACF, and the corresponding 95% confidence bands of TMR time series for the lag range of 0–50 months in Tabriz and Urmia rainfall gauge stations. The figure shows that the ACFs in both stations contain an oscillating pattern. They have the appearance of a sinusoidal function with a 12-month period (i.e., annual periodicity). This means that monthly rainfall at the stations is more correlated with its previous year amount than that of previous months. Such year-to-year serial dependent of
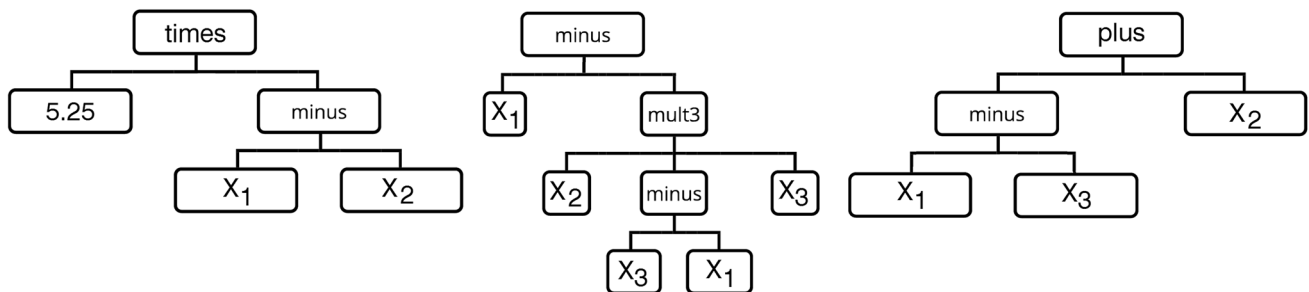
Fig. 4 An example of a multigene chromosome involving three genes with maximum depth of four. The mult3 function node denotes multiplication with three arguments

Fig. 5 ACF and PACF of TMR samples in Tabriz and Urmia stations
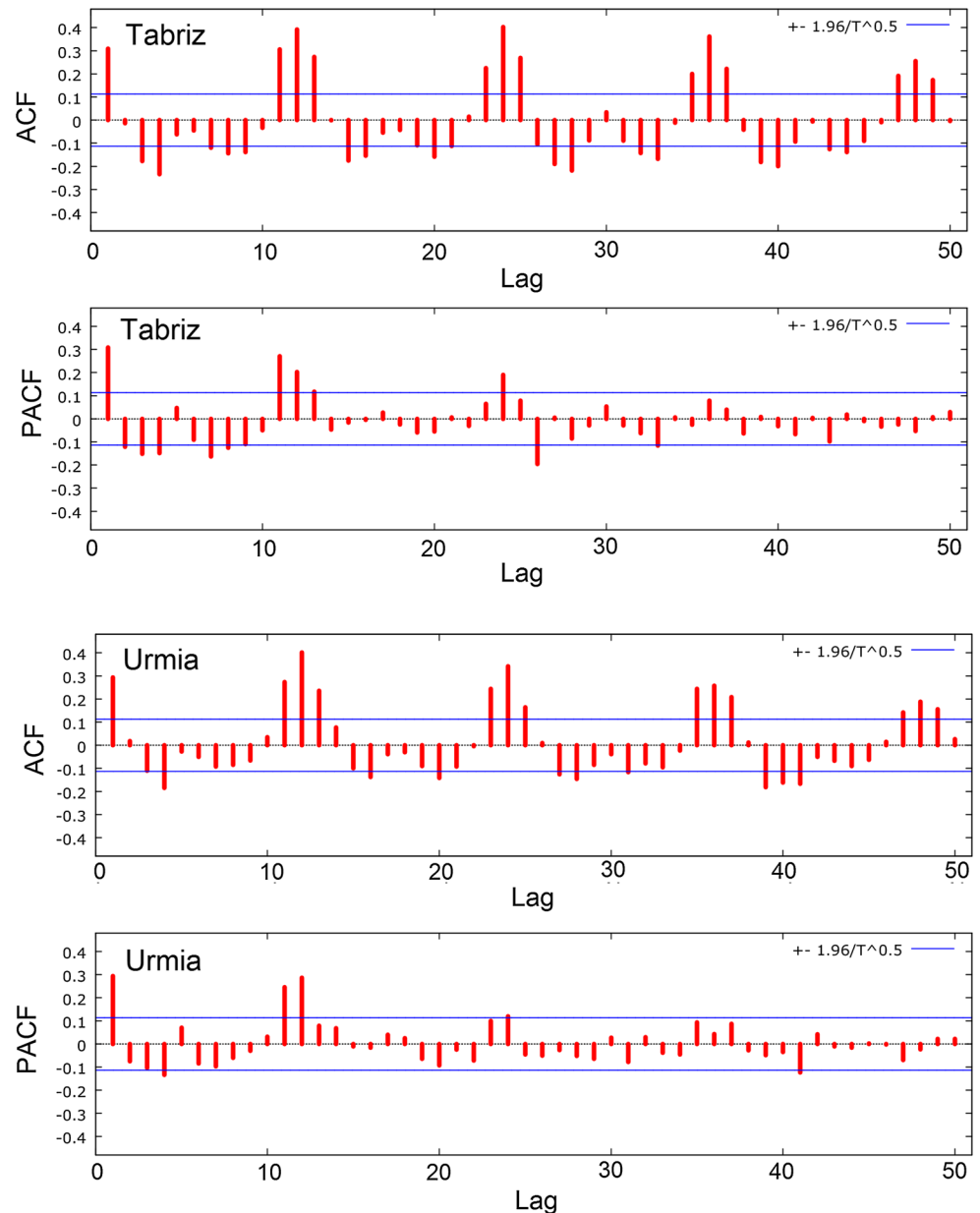
monthly rainfall values has been also shown in earlier studies (Delleur and Kavvas 1978; Aksoy and Dahamsheh 2009). In addition, PACFs show more correlation between the current month rainfall and its antecedent values at lag-1, lag-11, lag-12, and lag-24, respectively. Therefore, 1-, 11-, 12-, and 24-month lags have been selected as the extent of lag implemented for 1-month-ahead rainfall forecasting in the present study.

$$R_t = f\left(R_{t-1}, R_{t-11}, R_{t-12}, R_{t-24}, \varepsilon_t\right) \qquad (12)$$

where R_t represents monthly rainfall at the present time month t. The indices t-1 and t-11 are referred to as 1-month and 11-month lags and so on. The ε_t is bias (noise) term.

Results and discussion

In system identification process using AI methods, before data mining itself, a data preprocessing approach is commonly applied to make input/output variables dimensionless and put them within a certain range. Moreover, certain types of data preprocessing methods can remove nonstationary features of data. The type of data preprocessing approach suggested by Delleur and Kavvas (1978) was applied in this study. To this end, the benchmarks (i.e., SVR and MGGP) and SVR–FFA models were trained and verified using TMR series after the series were square-root-transformed and then

standardized so that they have zero mean and unit variance. It has been shown that subtracting the monthly means essentially removes the trend in the mean and variance and yields a time series having weak stationary states which is satisfied for practical purposes (Delleur and Kavvas 1978). After data preprocessing, the first step in developing SVR, MGGP, and SVR–FFA models is to determine the range of training and unseen testing data sets. The TMR series includes 25-year observations (see Fig. 1b). The first 2 years (i.e., 1990–1991) at each gauge have been considered as the years required to generate necessary lags for the time series (see Eq. 12). Therefore, the observations from January 1993 to December 2011 were selected for training the models. After creating the best models at each gauge, separately, the models were tested using the holdout TMR series at the testing period 2012–2014.

To obtain the best MGGP model and suitable values for SVR parameters (C, ε, γ) at each station, the RMSE is used as the fitness function. In addition to TMR series, to attain the best MMGP model, a set of random constants in the range of [−2, 2] were chosen as members of the terminal set in this study. Neglecting the potential complexity of evolved models, effect of various primitive functions is tested in order to find the best-fitted MMGP model. This step has been accomplished by trial-and-error method, and the results showed that beside the basic arithmetic operations (+, −, ×, and /), trigonometric (including sin, cos, and tan) and power functions play an important role in MGGP model. These functions were, therefore, considered as the other members of function set for MGGP runs. The other parameters/methods used for MGGP setup are given in Table 2. As shown in the table, the mutation transform has been applied with relatively high rate. The reason behind is the fact that the MGGP is dealing with very complex and nonlinear datasets, and the searching algorithm tends to converge very fast. To cope the problem, choosing such high mutation rate, new genetic materials are tried to bring in population set at each generation. Accordingly, MGGP could achieve a population diversity as much as possible.

As previously mentioned, in both SVR and SVR–FFA models, the RBF was used as the kernel function. The initial values of SVR and kernel parameters (i.e., C, ε and γ) for the best stand-alone SVR model were obtained by grid search approach in the training data set as suggested by Ghorbani et al. (2017). The approach tries to find an appropriate value for each parameter across a specific range in regard to RMSE at training period. This approach suggested that the SVR models with the parameters of (150.095, 0.091, and 50.270) and (0.775, 0.128, and 3.215) performed well for Tabriz and Urmia stations, respectively. As previously described, these parameters were subjected to be optimized using FFA in the SVR–FFA model. Accordingly, the optimal values of the parameters were estimated to be (76.47, 1.3, and 24.27) and (10.56, 0.8, and 6.49) in Tabriz and Urmia stations, respectively.

A comprehensive comparison among the efficiency results of the best evolved MGGP, SVR, and SVR–FFA models at each station is presented in Table 3. According to the results, the SVR–FFA provides better fit than both SVR and MGGP models in both Tabriz and Urmia stations. The FFA-induced improvement in SVR predictions was matched by an approximately 30% decrease in regard to RMSE and 100% increase in regard to NSE. This is due to the dependency of the optimal solution of SVR in finding the suitable parameters for kernel function. The performance of the MGGP model is comparable with SVR–FFA in Tabriz station, and it is superior to stand-alone SVR in both stations. But the proposed FFA–SVR model is still superior to the benchmark models in both stations.

Table 2 Parameter setting for the MGGP runs

Parameter	Value
Population size (chromosome)	1000
Mutation rate	% 25
Crossover rate	% 95
Reproduction rate	% 20
Maximum genes (trees)	5
Max tree depth	6
Max nodes per tree	Infinitive
Number of random constance	Infinitive
Initialization	Ramped half and half
Selection method	Fitness proportionate

Table 3 Goodness-of-fit results of the best MGGP, SVR, and SVR–FFA models for monthly rainfall forecasting at Tabriz and Urmia rain gauge stations

Rain gauge station	Model	Training		Testing	
		RMSE (mm)	NSE	RMSE (mm)	NSE
Tabriz	MGGP	156.4	0.428	145.0	0.410
	SVR	168.9	0.333	170.4	0.183
	SVR–FFA	120.6	0.661	122.8	0.593
Urmia	MGGP	221.0	0.392	247.9	0.272
	SVR	231.9	0.331	276.6	0.097
	SVR–FFA	165.5	0.659	206.9	0.493

To spatial assessment of the models' performance, the efficiency results of each model at different locations were compared. The results show that all the models provide more truthful predictions in Tabriz station. For example, the NSE values of the SVR–FFA models in Tabriz and Urmia stations are 0.593 and 0.493, respectively. This means that TMR prediction in Tabriz station is perhaps easier than Urmia station. The reason behind might be related to the more limited range of rainfall amounts in Tabriz station as shown by the relatively lower coefficient of variation in this station (see Table 1).

In order to further investigate performance of the models, the observed and predicted monthly rainfall time series at the testing period are shown in Fig. 6. The figure also shows comparative scatter plots between the results obtained by the SVR–FFA model and the measured values during the testing period. According to the time series plots, all the models more or less are capable to capture the oscillating pattern of the observed rainfall events. The SVR–FFA model is more efficient in finding both local and global maxima compared to the benchmarks. This result is consistent with previous studies (e.g., Shamshirband et al. 2016) that justify the

importance of FFA optimization in providing a better calibration for SVR model. Although the proposed SVR–FFA model significantly improved SVR forecasts and the results are better than those of the MGGP models, the scatter plots illustrates that the obtaining results are not at the desired level of accuracy. There is still room for more studies to develop more precise models for long-term rainfall forecasting in the study region.

Further to the common performance evaluation criteria, Taylor diagrams were depicted to summarize the overall performance of the models by identifying the pattern correlations, variability, and RMSE between the models and observed data. Figure 7 shows the Taylor diagrams of the TMR results of both the observed and three models for both training and testing sets. The modeling results in Tabriz station (Fig. 7a) indicates the pattern correlation of the forecasts with the observations lies in the range 0.55–0.85 and 0.35–0.65 at training and testing period, respectively. The models usually demonstrate lower performance in the training and higher performance in the testing period. The SVR and SVR–FFA exhibit the lowest and highest pattern correlation for training and testing periods, respectively. The
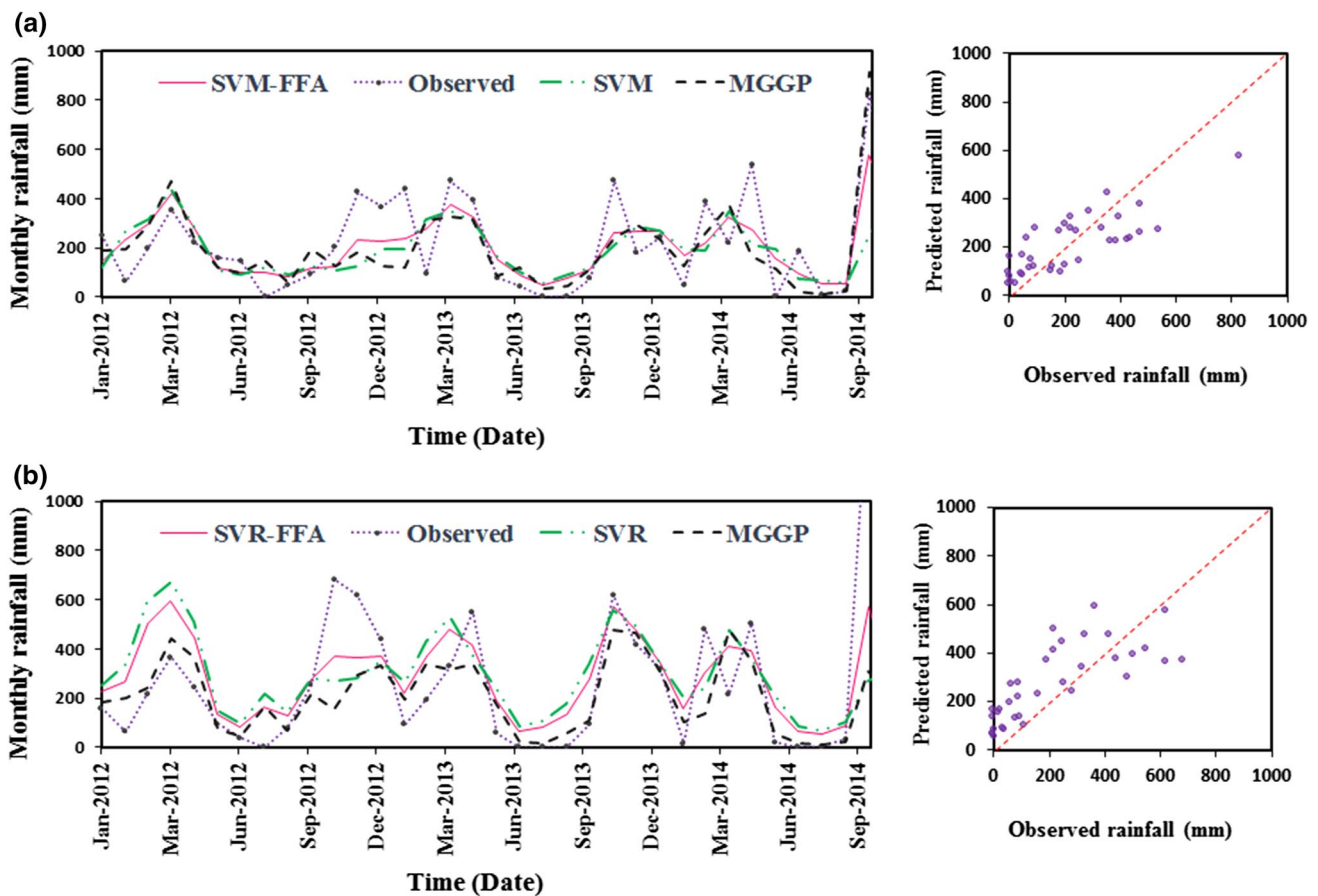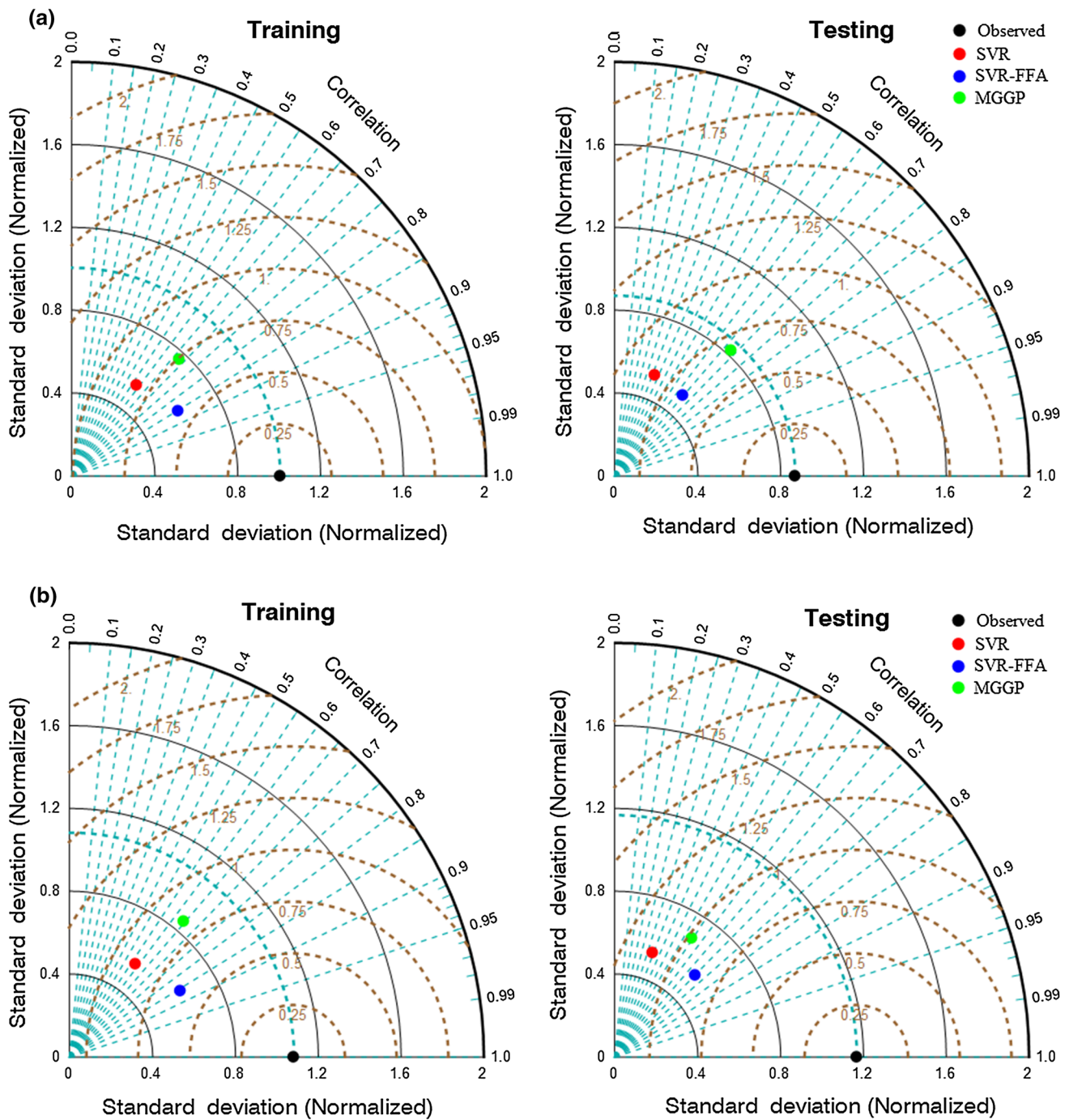

Fig. 6 Observed and predicted total monthly rainfall: **a** Tabriz station and **b** Urmia station. The scatter plots were given only for the SVR–FFA model

Fig. 7 Normalized Taylor diagrams displaying differences within pattern statistics of prediction models and observations at **a** Tabriz station and **b** Urmia station

variability in the MGGP is larger compared to the other model forecasts. On the other hand, the SVR results in the highest RMSE in both training at testing periods. As expected, the SVR–FFA forecasts in Urmia station reveal a fairly good agreement of the model estimations with the observations (Fig. 7b). The model has the highest pattern correlation (about 0.86 and 0.75 at training and testing

periods, respectively) as well as the lowest normalized RMSE (about 0.60 and 0.85 at training and testing periods, respectively). Similar to the MGGP results in Tabriz station, it is clear that the MGGP forecasts have the closest variation with the observations, most likely because of the high mutation rate that brings new genetic materials at each generation and the MGGP can achieve a population diversity

adequately. With respect to the results of Taylor diagrams at both stations, it can be said that SVR–FFA gives more consistent forecasts in terms of pattern correlation and RMSE. However, MGGP achieves generally better variability with regard to that of observed data.

Conclusion

In this study, the potential of SVR, MGGP, and newly developed SVR–FFA models were investigated to forecast TMR amounts in two rain gauges located in a semiarid region, Iran. The models were constructed for 1-month-ahead forecasting scenarios at each gauge so that 1-, 11-, 12-, and 24-month lags of observed monthly rainfall time series were used as input variables for modeling. To cope with the nonstationary feature of the observed TMR series, the original rainfall datasets were transformed using square root function and then standardized so that they have zero mean and unit variance.

Referring to the efficiency results of the SVR–FFA model, it can be inferred that the proposed model has higher ability than SVR and MGGP models to capture nonlinear feature of monthly rainfalls in the present region. Indeed, utilization of the FFA as an add-in optimizer in SVR model led to a significant improvement in the predictive accuracy, presumably owing to the optimal values of SVR parameters attained in the hybrid model. These results are in accord with earlier studies that demonstrate FFA may enhance feature extraction capability of other data-driven techniques (e.g., Yaseen et al. 2017a, b; Deo et al. 2018).

Returning to the literature, it should be reminded that neither stand-alone ANN (Aksoy and Dahamsheh 2009) nor SVR (Feng et al. 2014) was able to forecast monthly rainfalls with a desired level of accuracy in semiarid to arid regions. These results emphasize the complexity of monthly rainfall forecasting in arid regions. In spite of the higher accuracy of the SVR–FFA model, there is clearly room for more research on long-term forecasts of rainfalls, particularly in arid regions. The reason behind may rely on intermittent structure of the rainfall sequences as well as the high nonstationary feature of monthly rainfall series.

This study was limited to (1) 1-month-ahead forecasting scenario, (2) observations from semiarid region; and (3) weak stationary conditions of observed monthly rainfall series. To more efficiently use of rainfall predictions for many water resource applications, one way for future investigation could be examining the ability of SVR–FFA model to forecast monthly rainfall magnitudes in other climatic regions. Moreover, efficiency of the adopted method could be investigated for rainfall forecasts with higher lead times. From the standpoint of model improvement, future studies would also consider the effect of various data preprocessing approaches such as wavelet decomposition on the efficiency of SVR–FFA- or MGGP-based monthly rainfall prediction models.

References

Abarghouei HB, Hosseini SZ (2016) Using exogenous variables to improve precipitation predictions of ANNs in arid and hyper-arid climates. Arab J Geosci 9(15):663

Akbari-Alashti H, Bozorg Haddad O, Mariño MA (2015) Application of fixed length gene genetic programming (FLGGP) in hydro-power reservoir operation. Water Resour Manag 29(9):3357–3370

Aksoy H, Dahamsheh A (2009) Artificial neural network models for forecasting monthly precipitation in Jordan. Stoch Environ Res Risk Assess 23(7):917–931

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

Danandeh Mehr A, Demirel MC (2016) On the calibration of multigene genetic programming to simulate low flows in the Moselle River. Uludağ Univ J Fac Eng 21(2):365–365

Danandeh Mehr A, Nourani V (2017) A Pareto-optimal moving average-multigene genetic programming model for rainfall-runoff modelling. Environ Model Softw 92:239–251

Danandeh Mehr A, Kahya E, Bagheri F, Deliktas E (2014a) Successive-station monthly streamflow prediction using neuro-wavelet technique. Earth Sci Inform 7(4):217–229

Danandeh Mehr A, Kahya E, Yerdelen C (2014b) Linear genetic programming application for successive-station monthly streamflow prediction. Comput Geosci 70:63–72

Danandeh Mehr A, Kahya E, Şahin A, Nazemosadat MJ (2015) Successive-station monthly streamflow prediction using different artificial neural network algorithms. Int J Environ SciTechnol 12(7):2191–2200

Delleur JW, Kavvas ML (1978) Stochastic models for monthly rainfall forecasting and synthetic generation. J Appl Meteorol 17(10):1528–1536

Deo RC, Ghorbani MA, Samadianfard S, Maraseni T, Bilgili M, Biazar M (2018) Multi-layer perceptron hybrid model integrated with the firefly optimizer algorithm for windspeed prediction of target site using a limited set of neighboring reference station data. Renew Energy 116:309–323

Ebtehaj I, Bonakdari H (2016) A support vector regression-firefly algorithm-based model for limiting velocity prediction in sewer pipes. Water Sci Technol 73(9):2244–2250

El-Shafie AH et al (2011) Artificial neural network technique for rainfall forecasting applied to Alexandria, Egypt. Int J Phys Sci 6(6):1306–1316

Feng Q, Wen X, Li J (2014) Wavelet analysis-support vector machine coupled models for monthly rainfall forecasting in arid regions. Water Resour Manag 29(4):1049–1065

Ghorbani MA, Shamshirband S, Haghi DZ, Azani A, Bonakdari H, Ebtehaj I (2017) Application of firefly algorithm-based support vector machines for prediction of field capacity and permanent wilting point. Soil Tillage Res 172:32–38

Gizaw MS, Gan TY (2016) Regional flood frequency analysis using support vector regression under historical and future climate. J Hydrol 538:387–398

Hashmi MZ, Shamseldin AY, Melville BW (2011) Statistical downscaling of watershed precipitation using Gene Expression Programming (GEP). Environ Model Softw 26(12):1639–1646

Jalalkamali A, Moradi M, Moradi N (2015) Application of several artificial intelligence models and ARIMAX model for forecasting drought using the Standardized Precipitation Index. Int J Environ Sci Technol 12(4):1201–1210

Kazem A, Sharifi E, Hussain FK, Saberi M, Hussain OK (2013) Support vector regression with chaos-based firefly algorithm for stock market price forecasting. Appl Soft Comput 13(2):947–958

Kisi O, Cimen M (2012) Precipitation forecasting by using wavelet-support vector machine conjunction model. Eng Appl Artif Intell 25(4):783–792

Koza JR (1992) The genetic programming paradigm: Genetically breeding populations of computer programs to solve problems. In: Soucek B, The IRIS Group (eds) Dynamic, genetic, and chaotic programming. Wiley, New York, pp 203–321

Mekanik F, Imteaz MA, Gato-Trinidad S, Elmahdi A (2013) Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. J Hydrol 503:11–21

Mekanik F, Imteaz MA, Talei A (2016) Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals. Clim Dyn 46(9–10):3097–3111

Moustris KP, Larissi IK, Nastos PT, Paliatsos AG (2011) Precipitation forecast using artificial neural networks in specific regions of Greece. Water Resour Manag 25(8):1979–1993

Nasseri M, Asghari K, Abedini MJ (2008) Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. Expert Syst Appl 35(3):1415–1421

Nourani V, Andalib G (2015) Daily and monthly suspended sediment load predictions using wavelet based artificial intelligence apprroaches. J Mount Sci 12(2):85–100

Nourani V, Alami MT, Aminfar MH (2009) A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Eng Appl Artif Intell 22(3):466–472

Olyaie E, Abyaneh HZ, Danandeh Mehr A (2017) A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River. Geosci Front 8(3):517–527

Raghavendra S, Deka PC (2014) Support vector machine applications in the field of hydrology: a review. Appl Soft Comput J 19:372–386

Ravansalar M, Rajaee T, Kisi O (2017) Wavelet-linear genetic programming: a new approach for modeling monthly streamflow. J Hydrol 549:461–475

Saxena A, Verma N, Tripathi KC (2014) Neuro-genetic hybrid approach for rainfall forecasting. Int J Comput Sci Inf Technol 5(2):1291–1295

Searson D (2015) GPTIPS 2: an open-source software platform for symbolic data mining. In: A. H. G. et al. (ed) Chapter 22 in Handbook of genetic programming applications. Springer, New York

Shamshirband S, Mohammadi K, Tong CW, Zamani M, Motamedi S, Ch S (2016) A hybrid SVM-FFA method for prediction of monthly mean global solar radiation. Theor Appl Climatol 125(1–2):53–65

Shirani Faradonbeh R, Armaghani DJ, Majid MA, Tahir MM, Murlidhar BR, Monjezi M, Wong HM (2016) Prediction of ground vibration due to quarry blasting based on gene expression programming: a new model for peak particle velocity prediction. Int J Environ Sci Technol 13(6):1453–1464

Sivapragasam C, Liong S, Pasha M (2001) Rainfall and runoff forecasting with SSA-SVM approach. J Hydroinformatics 3(3):141–152

Solgi A, Nourani V, Pourhaghi A (2014) Forecasting daily precipitation using hybrid model of wavelet-artificial neural network and comparison with adaptive neurofuzzy inference system (case study: Verayneh station, Nahavand). Adv Civ Eng 2014. ID 279368, 12 p

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 106(D7):7183–7192

Uyumaz A, Danandeh Mehr A, Kahya E, Erdem H (2014) Rectangular side weirs discharge coefficient estimation in circular channels using linear genetic programming approach. J Hydroinformatics 16(6):1318–1330

Wu CL, Chau KW, Fan C (2010) Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. J Hydrol 389(1–2):146–167

Yang XS, Hosseini SSS, Gandomi AH (2012) Firefly Algorithm for solving non-convex economic dispatch problems with valve loading effect. Appl Soft Comput J 12(3):1180–1186

Yaseen ZM, Ebtehaj I, Bonakdari H, Deo RC, Danandeh Mehr A, Mohtar WHMW, Diop L, El-Shafie A, Singh VP (2017a) Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model. J Hydrol 554:263–276

Yaseen ZM, Ghareb MI, Ebtehaj I, Bonakdari H, Siddique R, Heddam S, Yusif AA, Deo R (2017b) Rainfall pattern forecasting using novel hybrid intelligent model based ANFIS-FFA. Water Resour Manag 32(1):105–122