Research Article

An ensemble genetic programming model for seasonal precipitation forecasting

Ali Danandeh Mehr[1]

Abstract

Seasonal precipitation forecasting is one of the most challenging tasks in stochastic hydrology. This article proposes a new ensemble model, called EGP, to a season-ahead forecast of total seasonal precipitation. The EGP integrates evolutionary genetic programming (GP) and gene expression programming (GEP) techniques with multiple linear regression to increase forecasting accuracy of standalone GP and GEP models, while it secures their explicit structure. The EGP model was trained and validated using 88 years (1930–2017) of measured precipitation data from Muratpasa Station, Antalya, Turkey. The model performance was evaluated in terms of different statistical error measures and cross-validated with two other ensemble models as well as the state-of-the-art random forest developed in this study as the benchmark. The results showed that the proposed model can increase the forecasting accuracy of the best standalone GP and GEP models up to 30%. The EGP was also found to be superior to random forest, particularly in predicting low and high seasonal precipitation amount. This model is explicit, easy to evolve, and therefore, motivating to be used in practice.

Keywords Precipitation · Genetic programming · Seasonal forecast · Gene expression programming · Multiple regression · Random forest

1 Introduction

It is expected that the recent increase in the concentration of atmospheric carbon dioxide and global temperature will have a significant impact on precipitation patterns in both global and regional scales [1, 7, 40]. Like the other water cycle components, the precipitation process is highly complex owing to the stochastic attributes of its triggering factors, such as temperature, wind speed, and humidity. The level of uncertainty in precipitation forecasting is significantly higher than the other water cycle components such as temperature and streamflow [19, 38]. Consequently, the problem has been addressed in a variety of recent studies (e.g., [8, 22, 33, 39]). Many efforts have been made over the past decades to model precipitation patterns at specific locations. Using the theory of Markov chains and classic statistical models, observed precipitation series have been used to model and predict precipitating in different lead times [12, 43]. However, they are not well enough for long-lead-time forecasting owing to the highly nonlinear structure and presence of non-stationarity in the monthly or seasonal rainfall series [10, 15]. Nowadays, emerging artificial intelligent (AI) techniques such as artificial neural networks (ANNs), support vector machine (SVM), extreme learning machines, and genetic programming (GP) are from the AI techniques that were used for precipitation forecasting (e.g., [2, 11, 24, 34]). However, the techniques were criticized as a black box and the results were not precise enough, particularly in the predictions of drizzles or heavy rainstorms [25].

To increase predicting accuracy, the current pertinent literature showed an increasing trend in the implementation of hybrid AI models that combine different data pre- or post-processing approached with one or more AI

✉ Ali Danandeh Mehr, ali.danandeh@antalya.edu.tr | [1]Department of Civil Engineering, Antalya Bilim University, Antalya, Turkey.

techniques ([14, 18, 20, 32, 41, 44]. For example, Sivapragasam et al. [35] and Chau and Wu [4] coupled singular spectrum analysis with SVM and showed the hybrid model may perform better than ad hoc SVM and ANN for rainfall forecasting. In a recent study, Danandeh Mehr et al. [9] confirmed that the firefly optimization algorithm can be used to determination of SVM parameters and increase its predictive accuracy. More recently, Yaseen et al. [42] hybridized an adaptive neuro-fuzzy inference system model with the evolutionary algorithms including particle swarm, genetic algorithm, and differential evolution. The capability of the proposed hybrid models for precipitation forecasting was compared with the conventional neuro-fuzzy model. The results showed that all the hybrid models are suitable to improve the neuro-fuzzy inference system; however, the particle swarm optimization was superior.

Although the above-mentioned hybrid AI models may provide higher prediction accuracy than standalone ones, they are still black boxes possessing a complicated model structure in most cases. These are two main factors that caused demotivation for practitioners to apply such hybrid models in practice. Thus, additional efforts are needed to address the key issues of accuracy and complexity in precipitation forecasting models. The aim of this study is, therefore, to develop a new hybrid model that meets both precision and simplicity conditions. To this end, several linear and nonlinear integrations of GP and gene expression programming (hereafter GEP) are examined. The main contributions of this study are threefold. First, the present study, for the first time, investigates predictive capabilities of standalone GP and GEP for seasonal precipitation forecasting. These are emerging AI techniques with explicit tree structures typically known as grey-box techniques [16]. Thus, they overcome the ultimate black box feature of earlier models. Second, this study integrates linear and nonlinear regression techniques to develop an ensemble model that prioritizes both predictive accuracy and model simplicity. The proposed method can be used to forecast seasonal precipitation with an explicit tree structure. Finally, this is the first study that compares the ability of GP- and random forest (RF)-based models for long-term precipitation modeling. RF is a robust regression and classification technique that was considered as the benchmark in this study.

2 Study area and data preparation

The Antalya Province with a population of more than two million is located on the Mediterranean coast of southwest Turkey, between the Torus Mountains and the Mediterranean Sea (Fig. 1a). Modern precipitation measurement in Antalya has been commenced since 1929. The data used in this study are from Muratpasa station (36.9063 N, 30.7990 E) located in Antalya downtown (Fig. 1b, c). The daily precipitation data from January 1, 1930, till December 31, 2017, were collected from the Turkish State Meteorological Service and used in this study.

The mean monthly precipitation for the period of 1930–2017 is depicted in Fig. 1d. Because of a pronounced variation in the mean monthly precipitation, conventional autoregressive time series modeling approaches could not be a wise choice for long-term precipitation forecasting in the Muratpasa station. This implies the Mediterranean climate that prevails in Antalya with hot and dry summers and wet autumns, winters, and springs. From a meteorological perspective, summer precipitations in the region are generally convective with high intensity and short duration. Significant differences in the seasonal distribution of the number of rainy days are also observed. While the long-term average of rainy days in the summer is about 4.0 days, the corresponding numbers for winter, spring, and autumn are 34, 22, and 15 days, respectively [5]. Owing to such a significant difference in the precipitation pattern between summer and wet seasons, this study aimed at the modeling and forecasting seasonal precipitation at wet seasons (autumn, winter, and spring) in which the evolved models merely uses historical data from the wet seasons. From a modeling point of view, this strategy may yield in the elimination of abundant zero values from the input vectors (predictors) and consequently will lead to more precise predictions for wet seasons. There are 88-year observed data with a total of 352 (88×4) seasons. Hence, this study is based on 264 (88×3) seasons. Time series plot of the observed total seasonal precipitation (TSP) and their statistical characteristics are presented in Fig. 2 and Table 1, respectively. For information about the socioeconomic development of the Antalya Province, the interested reader is referred to the recent paper by Özkubat and Selim [30].

3 Methodology

3.1 Determination of optimum predictors

Determination of optimum inputs is an important task in any system identification or hydrological modeling problem that dictates the accuracy/complexity of evolved models. In previous studies, different input combinations were examined via trial and error method to determine suitable predictors (e.g., [28, 42]). However, this is time-consuming, and the modeler might not achieve the best inputs if it is not considered as the potential predictor in advance. In this study, the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the TSP series were
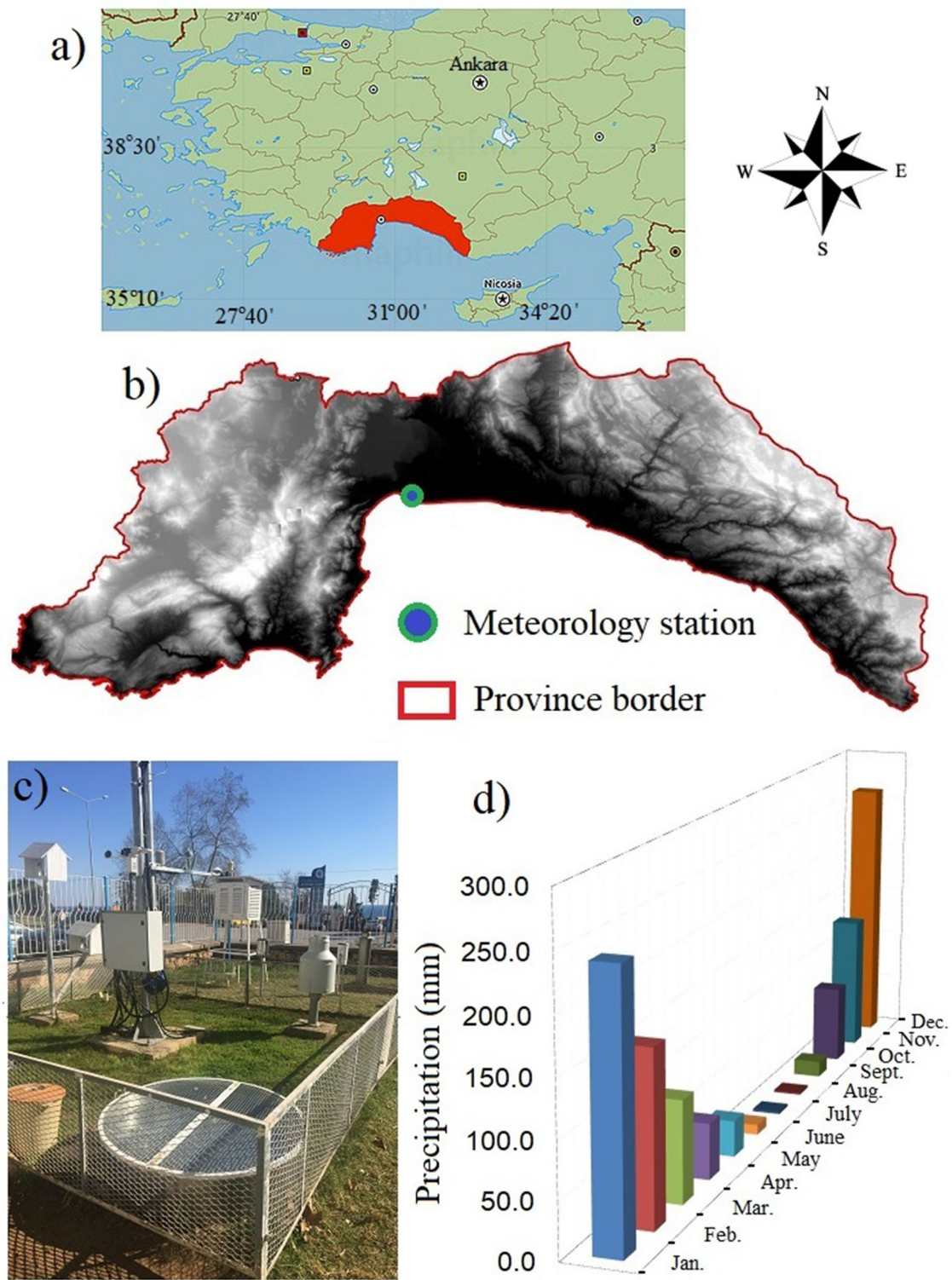
Fig. 1 Location of the Antalya Province (**a**), meteorology station (**b**, **c**), and mean monthly precipitation for the period 1930–2017 (**d**)

considered to determine optimum inputs (Fig. 3). The seasonal oscillation pattern is seen in the ACF diagram. As expected, the partial correlation markedly decreased after nine lags. These patterns point to the selection of nine lags as expressed in Eq. (1). The structure optimization features
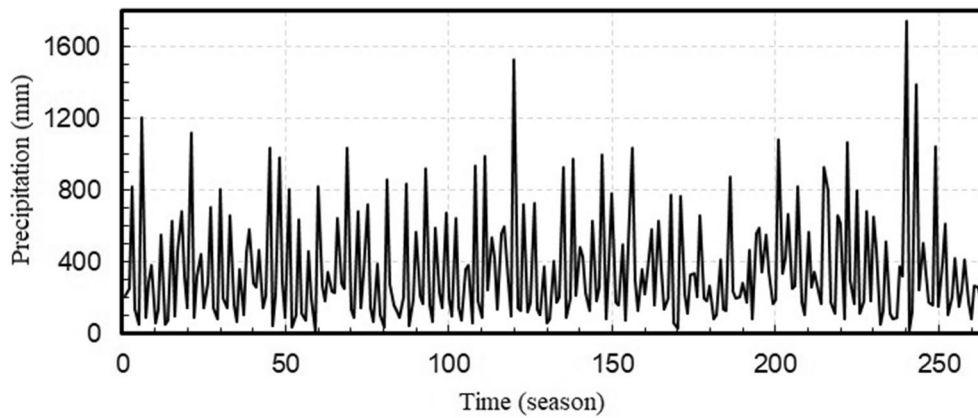
Fig. 2 Total seasonal precipitation observed in Antalya, Turkey

Table 1 Statistical characteristics of the total seasonal precipitation observed in rainy season at Antalya during the 1930–2017 period

Season	Mean (mm)	Minimum (mm)	Maximum (mm)	Standard deviation (mm)
Autumn	221.8	11.4	925.5	167.6
Winter	663.9	92.1	1744.6	301.2
Spring	178.7	21.8	611.4	95.6

of GP, GEP, and RF are used to select the most effective inputs in the present study.

$$P_t = f\left(P_{t-9}, P_{t-6}, P_{t-3}, P_{t-2}, P_{t-1}\right), \tag{1}$$

where P_t is TSP amount at the season t.

3.2 Multiple linear regression (MLR)

The MLR is the advancement for linear regression method in which the dependent variable (P_t) is linearly regressed to the selected five lagged TSP values as shown in Eq. (2). The m-values are coefficients corresponding to each parameter, and b is a constant value. Indeed, the attained equation is a straight line that best fits the TSP data and can be used to predict the target value based on unseen independent variables within their initial ranges. The unknown coefficients and the constant value can be calculated using the "least squares" method. The task can be easily done using the LINEST function in MS Excel.

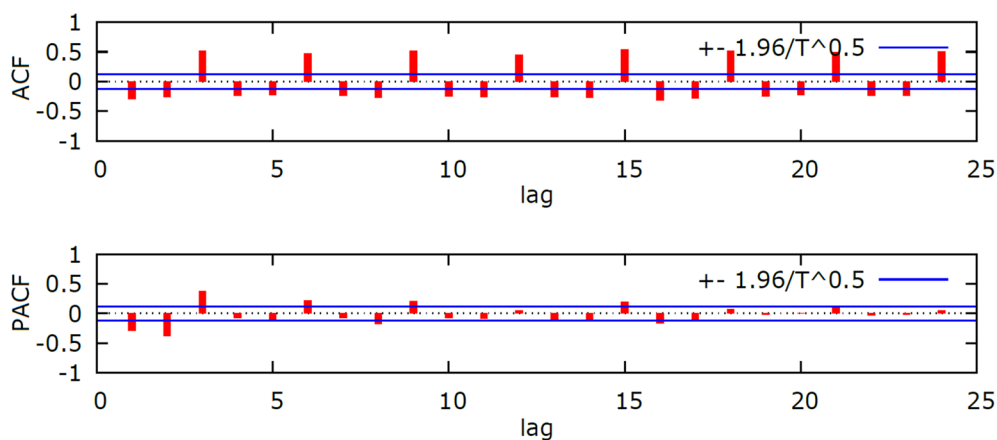$$P_t = m_1 P_{t-9} + m_2 P_{t-6} + m_3 P_{t-3} + m_4 P_{t-3} + m_5 P_{t-1} + b. \tag{2}$$

Fig. 3 Correlogram of the observed total seasonal precipitation

3.3 Overview of classic GP

GP [21] is an emerging evolutionary AI technique frequently used for symbolic regression problems. The best solution is an individual computer program which generated on the basis of the principle of the survival of the fittest [36]. In classic GP, each individual has a form of a tree, called the genome. For example, Fig. 4 represents a genome and corresponding mathematical equation encompassing a root node (addition), inner nodes of multiplication and addition, and terminal nodes of the independent variables (x_1 and x_2) and random numbers (R_1 and R_2). Each node in a genome accepts only a specific type(s) of elements. While the root node merely accepts a function (logical, mathematical, etc.), the inners can be fill by a function or independent variables. Therefore, creating a set of appropriate functions and variables is of the most important tasks of the GP modelers. In order to avoid overfitting problem and keep away from intricate solutions, use of low depth genomes has been suggested [17]. A wise decision on these factors not only supports the algorithm to achieve more accurate solutions but also reduce computation cost.

To solve a regression problem using GP, the algorithm starts with the formation of the initial population of potential solutions using the modeler defined functions/variables. The individuals that show better fit between the dependent and independent variables are considered as parents to produce offspring. Typically, three evolutionary operators including *Reproduction*, *Crossover*, and *mutation* are acted on the parents [6]. The reproduction is the relocating the best individual into the new set of offspring without any transform. The process commences with the goodness of fit assessment of initial programs and ceases after the identification of the best fitted one. The crossover is the exchange of branches/nodes between
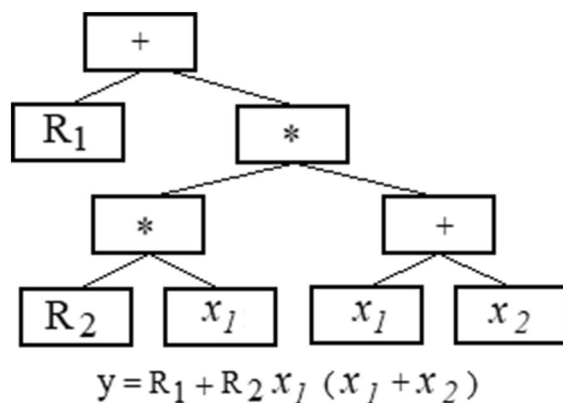
two high-performance individuals that results in two offspring [6]. Figure 5 illustrates the process in which a branch of *parent 1* (red nodes) is replaced with a branch of *parent 2* (blue nodes). The offspring are solutions that possess the genetic materials of their parents. Many studies have shown that offspring fit to the training set better than their parents [16]. As illustrated in Fig. 6, the mutation is the random change of branches/nodes (i.e., genetic materials) in a single parent at the mutation point. In the figure, the branch $x_1 \times \sin x_2$ in the parent was replaced with the randomly created Log x_2 in the offspring. All the above-mentioned operations are repeated till an individual displays the desired level of accuracy in both training and holdout testing data sets.

3.4 Overview of GEP

GEP is a GP variant that uses fixed-length linear genome, aka chromosomes, to represent computer programs in the form of expression trees [13]. It is trained with the same evolutionary process acting on a population of initial chromosomes of different sizes and shapes. The parents are selected according to fitness and genetic operators are implemented to improve their fitness. A linking function (e.g., addition, multiplication, etc.) is used to join the evolved offspring and the best model is selected among the joint expressions. An example of a linear GEP chromosome and associated mathematical function is presented in Fig. 7. The chromosome has four types of functions (Sin, $+$, $-$, $\times$, $\div$) and four terminals (X_1, X_2, 3, 2), where X_1 and X_2 are the input variables, and 3 and 2 are random numbers. the multiplication function between the parenthesis (i.e., sub-ETs) is called linking functions.

As illustrated in Fig. 7, the chromosome comprises two parts: the head (blue panel) which can take members from both functions and terminals, and the tail (green panel) which is formed merely with terminals. According to Ferreira [13], the length of the head h is chosen by the modeler, and then Eq. (3) is used to determine the length of the tail length t.

$$t = h \times (n - 1) + 1, \tag{3}$$

where n is the maximum arity of all predefined functions. For details about evolutionary operations in GP and GEP, the reader is referred to Danandeh Mehr et al. [6].

3.5 The proposed ensemble genetic programming (EGP) model

The proposed EGP is a hybrid model (Fig. 8) that aimed at increasing the accuracy of standalone classic multiple and symbolic evolutionary regression models. While
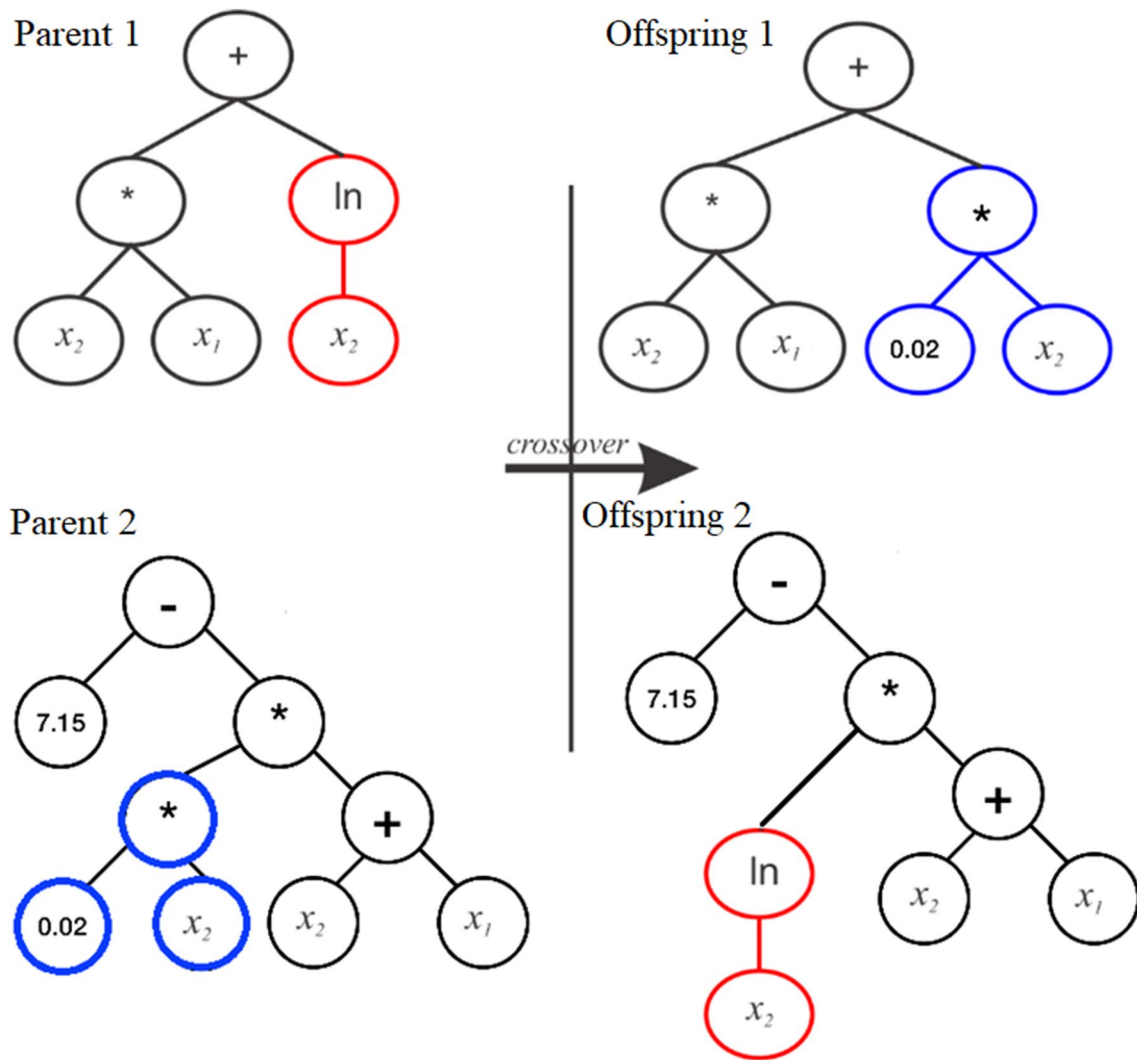
$$y = R_1 + R_2 x_1 (x_1 + x_2)$$

Fig. 4 A tree-shaped genomes and corresponding mathematical representations

Fig. 5 An example of crossover operation
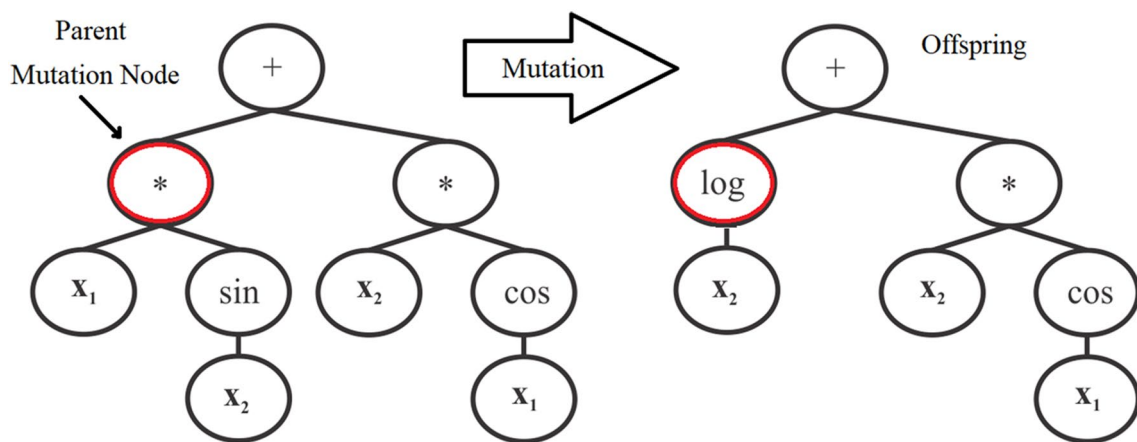
Fig. 6 Mutation operation acts on a GP chromosome
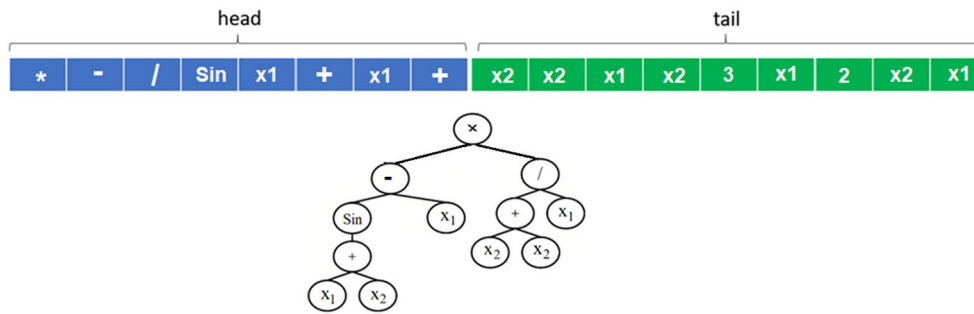
Fig. 7 An example of GEP chromosome representing $\left(\mathrm{Sin}\left(X_1 + X_2\right) - X_1\right)\frac{2X_2}{X_1}$
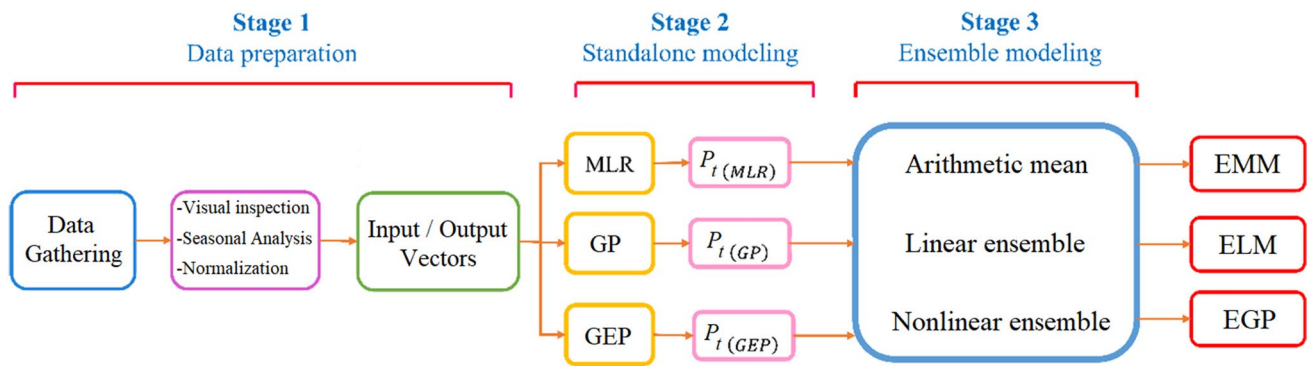

Fig. 8 Schematic of the proposed EGP model for seasonal precipitation forecasting

the standalone regression models map the dependent variables to the independent variable using either the method of least squares or evolutionary optimization technique, the new model improves the predictive accuracy through the combination of both tools. As illustrated in Fig. 8, the proposed model includes three stages of data preparation (Stage 1), standalone modeling (Stage 2), and ensemble modeling (Stage 3).

In the first stage, the historical data are gathered and preprocessed to determine input/output time series and to feed the standalone regression models. A classic linear regression model (i.e., MLR) together with two symbolic regression models (i.e., GP and GEP) was individually evolved to forecast P_t in the second stage. Finally, in the third stage (i.e., ensemble modeling unit), the results (outputs) obtained from the standalone modeling unit were integrated through three ensemble methods; (1) ordinary arithmetic mean, (2) linear ensemble, and (3) nonlinear ensemble techniques that, respectively, yields in ensemble mean model (Eq. 4; hereafter EMM), ensemble linear model (Eq. 5; hereafter ELM), and ensemble nonlinear model in which a classic GP is trained and verified to figure out optimum nonlinear structure (Eq. 6; hereafter EGP) for a season-ahead TSP forecasting.

$$\overline{P_t} = \frac{1}{3}\sum\left(P_{t\mathrm{MLR}}(i) + P_{t\mathrm{GP}}(i) + P_{t\mathrm{GEP}}(i)\right) \tag{4}$$

$$\overline{P_t} = m_1 P_{t\mathrm{MLR}}(i) + m_2 P_{t\mathrm{GP}}(i) + m_3 P_{t\mathrm{GEP}}(i) + c \tag{5}$$

$$\overline{P_t} = f\left(P_{t\mathrm{MLR}}(i), P_{t\mathrm{GP}}(i), P_{t\mathrm{GEP}}(i)\right), \tag{6}$$

where $\overline{P_t}$ is the ensemble model estimation, i represents the number of training data at each input/output vectors. The m_j (in the present study $j=3$) and c are, respectively, the weights of every single model and the constant component of the multivariable linear ensemble model. The optimum weights and constant are calculated by the least squares method via a multiple linear regression analysis in this study. Previous studies have proven that ensemble use of the outputs of several models improves the model accuracy [29, 36].

3.6 Overview of random forest (RF)

The RF [3] is an advancement of the conventional decision tree algorithm that refers to the creation of a set of random decision trees based on a random subset of variables

from the bootstrapped data set. In RF, a wide variety of random trees are created that makes the algorithm more effective than traditional decision trees. To predict a target variable, the corresponding features are taken and run them down all the created random trees. Then, voting is done among the outputs to see which alternative received more votes. This strategy has been proven that performs well in both classification and regression problems and is robust against overfitting [26]. Therefore, the approach is selected as a benchmark in this study. For the fundamentals of RF and its applications in water resources, the reader is referred to the review paper conducted by Tyralis et al. [37]. Details of RF created for seasonal precipitation forecasting are presented in the results section.

3.7 Performance evaluation

To validate the models' performance, Nash–Sutcliffe efficiency (NSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) measures were used in this study. These indicators are commonly used for the justification of hydrological models (e.g., [23, 26])

$$NSE = 1 - \frac{\sum_{i=1}^{n}(X_i^{obs} - X_i^{pre})^2}{\sum_{i=1}^{n}(X_i^{obs} - X_{mean}^{obs})^2} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i^{obs} - X_i^{pre})^2}{n}} \tag{8}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{X_i^{obs} - X_i^{pre}}{X_i^{obs}}\right|, \tag{9}$$

where X_i^{obs} is the observed TSP, X_i^{pre} is the TSP calculated by the models, and n is the number of samples.

4 Results

In this section, at first, the results of standalone modeling via MLR, GP, and GEP techniques are provided. Then, the outcomes from the ensemble modeling stage and discussion on different models are presented. The ensemble models are finally compared with an RF model developed

as the benchmark. It is worth mentioning that two software packages, namely GPdotNET [16] and GeneXproTools [13] were applied in this study to evolve the GP and GEP models, respectively. Both tools were trained and tested with the same input/target sample size, fitness function, and terminal sets. The best RF model was generated using RapidMiner Studio software.

In the evolution of AI models, particularly GP and GEP, dimensions and range of the inputs must be carefully adjusted before training the models [36]. As suggested by Tür [36], to achieve dimensionally truthful models, the TSP values were normalized into the range [0.1 and 0.9] using Eq. (10). Therefore, all the results presented in this paper are normal values.

$$P_{nt} = 0.8 \times \frac{P_t - P_{min}}{P_{max} - P_{min}} + 0.1. \tag{10}$$

4.1 Results of standalone linear and nonlinear models

Table 2 presents the efficiency results of the best evolved MLR, GP, and GEP models. The corresponding explicit expressions are also presented in Eqs. (11–13), respectively.

$$P_t = 0.271P_{t-9} + 0.206P_{t-6} + 0.205P_{t-3} \\ - 0.16P_{t-2} - 0.155P_{t-1} + 0.159 \tag{11}$$

$$P_t = ((((\cos((((\tan(P_{t-9})) \times (0.38016 - 0.38831)) \\ \times (\cos((P_{t-1} - 0.5559)))))) \times (sqrt((P_{t-9} \\ \times (sqrt((P_{t-6} \times P_{t-3})))))))) * (\cos((P_{t-9} \\ \times (sqrt((P_{t-9} \times (0.38016/0.38831))))))))) \tag{12}$$

$$P_t = [(0.0064((\exp((0.2418/P_{t-3}))^2) \times 0.2418))^2] \\ + [(0.7612^2) \times (\cos\left(\left(P_{t-6} + P_{t-3}\right)\right) \times ((0.4608P_{t-6} \\ + P_{t-3}))] + [(((\cos\left(P_{t-9}\right) \times \cos\left(P_{t-6}\right)) \times ((0.1566P_{t-9} \\ + P_{t-6})) \times P_{t-9}]. \tag{13}$$

According to the efficiency results, the GEP model provides the highest accuracy comparing to its counterparts both in the training and testing periods. The standalone GP model is also superior to the MLR. This is not a

Table 2 Performance indices calculated for the standalone models

Models	Training set			Testing set		
	RMSE	NSE	MAPE	RMSE	NSE	MAPE
MLR	0.091	0.529	0.284	0.127	0.289	0.341
GP	0.089	0.543	0.287	0.111	0.462	0.317
GEP	0.086	0.575	0.287	0.107	0.500	0.337

surprising result as the former uses a nonlinear structure having more complicated functions which indicate the complex relationship between the TSP and its antecedent values. From a model complexity perspective, the evolved GEP model is the most complex one as it uses more functional nodes. To further assess the models, the scatter plots of forecasts vs observations are presented in Fig. 9. According to the figure, MLR overestimates the lower TSP values in the testing period. By contrast, the GP underestimates the corresponding values. Comparing to the MLR and GP results, the observed experiments and corresponding GEP estimations are distributed more closely to the 1:1 line indicating its superiority.

4.2 Results of ensemble models

As previously mentioned in the ensemble modeling unit, the results of the MLR, GP, and GEP models are combined through two linear and one nonlinear schemes. Thus, the ensemble results are not sensitive to the limits of a standalone regression model. To attain optimum linear and nonlinear ensemble models in this stage, a new MLR and

GP models are developed using the results of standalone models as the inputs. As in the standalone GP, the GPdot-NET package was used in this stage where the algorithm was trained to minimize RMSE as the fitness function and the maximum depth of genes and best functions were obtained via a trial–error process.

The evolved ensemble GP models were also compared with the RF model developed as the benchmark in this study. Like ensemble GP models, RF is an ensemble of several tree structures. Here, several RF models were generated and the best one was selected. To this end, several random trees in the range of 2 to 1000 were generated with varying depth within the range 5 to 15. For each tree, a subset of examples was selected via bootstrapping and optimized with regard to the least square error criterion. The trials showed that the best model is attained with 50 trees, each of which has a depth of five. The results showed that a higher number of trees or depths increases the model performance in the training set insignificantly and no improvement occurred in the validation data. Table 3 compares the efficiency results of the best evolved EMM (Eq. 14), ELM (Eq. 15), and EGP (Fig. 10) models with the best evolved RF model.
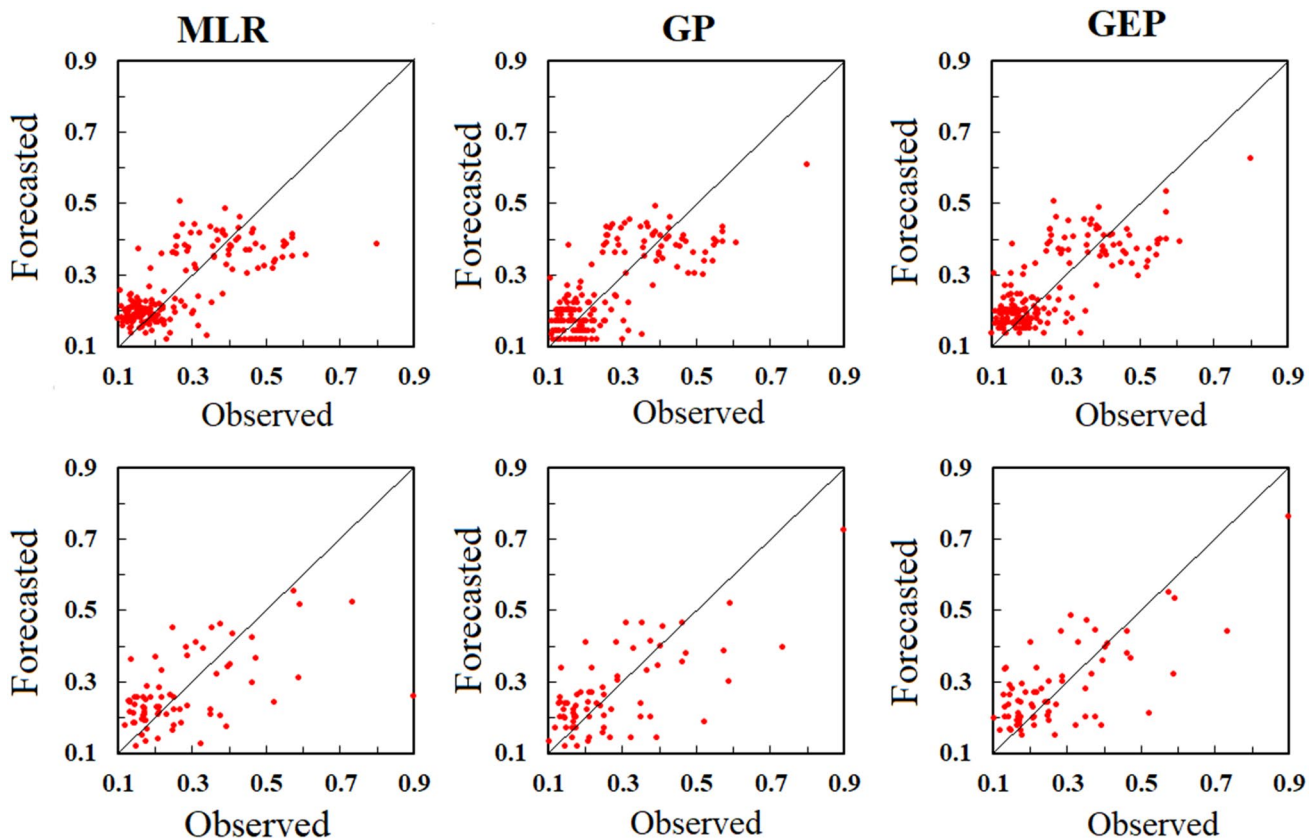

Fig. 9 Scatter plots of the standalone model forecasts versus observed normalized TSP in the training (top panel) and testing (bottom panel) periods

Table 3 Performance indices calculated for the ensemble models

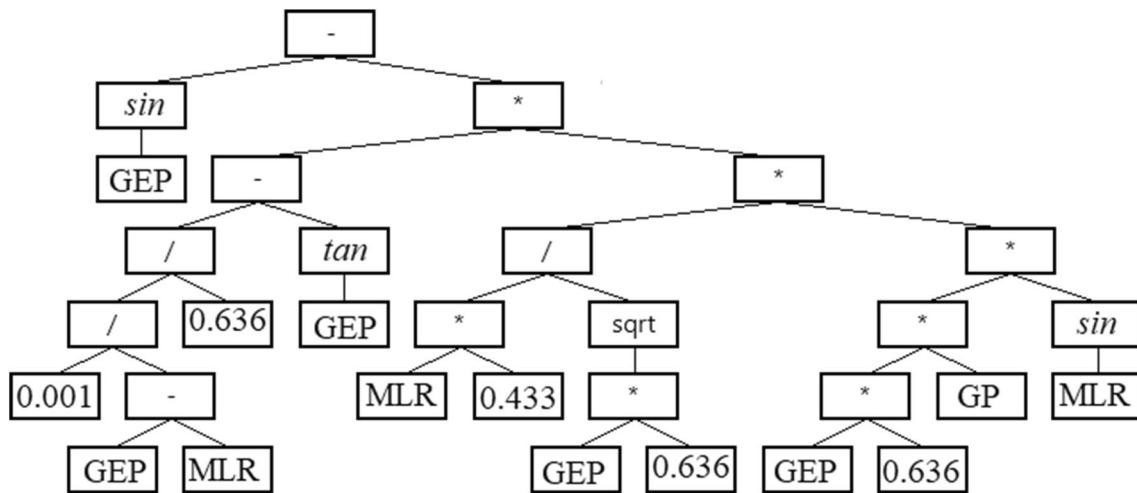| Models	Training set			Testing set		
	RMSE	NSE	MAPE	RMSE	NSE	MAPE
EMM	0.086	0.571	0.277	0.110	0.474	0.320
ELM	0.083	0.607	0.277	0.100	0.558	0.302
EGP	0.069	0.729	0.252	0.09	0.672	0.269
RF	0.067	0.746	0.207	0.116	0.410	0.337

Fig. 10 Optimum GP tree for the nonlinear ensemble model EGP

$$P_t = \left(P_{t\_MLR} + P_{t\_GP} + P_{t\_GEP}\right)/3 \tag{14}$$

$$P_t = 0.0729P_{t\_MLR} + 0.1036P_{t\_GP} + 0.7746P_{t\_GEP} + 0.01. \tag{15}$$

According to the results, the EGP performed better than the other ensemble models in terms of all the statistical efficiency measures. The improvement was more noticeable when the testing samples were considered. The EGP with the training (testing) NSE of 0.73 (0.67) is much more powerful in TSP forecasting than the ELM (EMM) model with NSE of 0.61 (0.56) and 0.57 (0.47), respectively. These values indicate that EGP is 20% and 28% superior to the ELM and EMM models, respectively. Comparing to the state-of-the-art RF model, the results indicated the superiority of the proposed ensemble EGP model. Lower generalizability of the RF was found as its main drawback in this study. As illustrated in Fig. 10, the EGP can be expressed in a single tree structure; however, the best RF, which is the combination of 50 trees, cannot be shown in a single tree.

Figure 11 exhibits the scatter plots of the ensemble models vs observed normalized TSP. As seen in the scatter plot, the EGP forecasts are closer to the 1:1 line that could predict the TSP better than the ELM and EMM models.

Therefore, the predicted TSP values improved with this method and the forecasts are closer to the observed data. It is noted the EGP is more powerful than MLR, GP, GEP, and RF models in the TSP forecast. In this model, the normalized TSP is predicted only by the knowledge of the historical precipitation events. The model does not depend on the large-scale climate indicators, and therefore, it is motivating to be used by practitioners.

Figure 12 illustrates the distribution of observed and ensemble models' forecasted TSP over the testing period (1990–2014) through displaying their quantiles and skewness. The figure demonstrates that all the models produced the same median which has an average seasonal precipitation of 242.6 mm. The EGP is superior to its counterparts in forecasting the precipitation amount in low and extreme rainy seasons. By contrast, the RF model shows better performance in predicting within the interquartile range.

5 Discussion

The results showed that combining the outcomes of classic and symbolic regression models may enhance the precision of the TSP forecasts. This is well in agreement
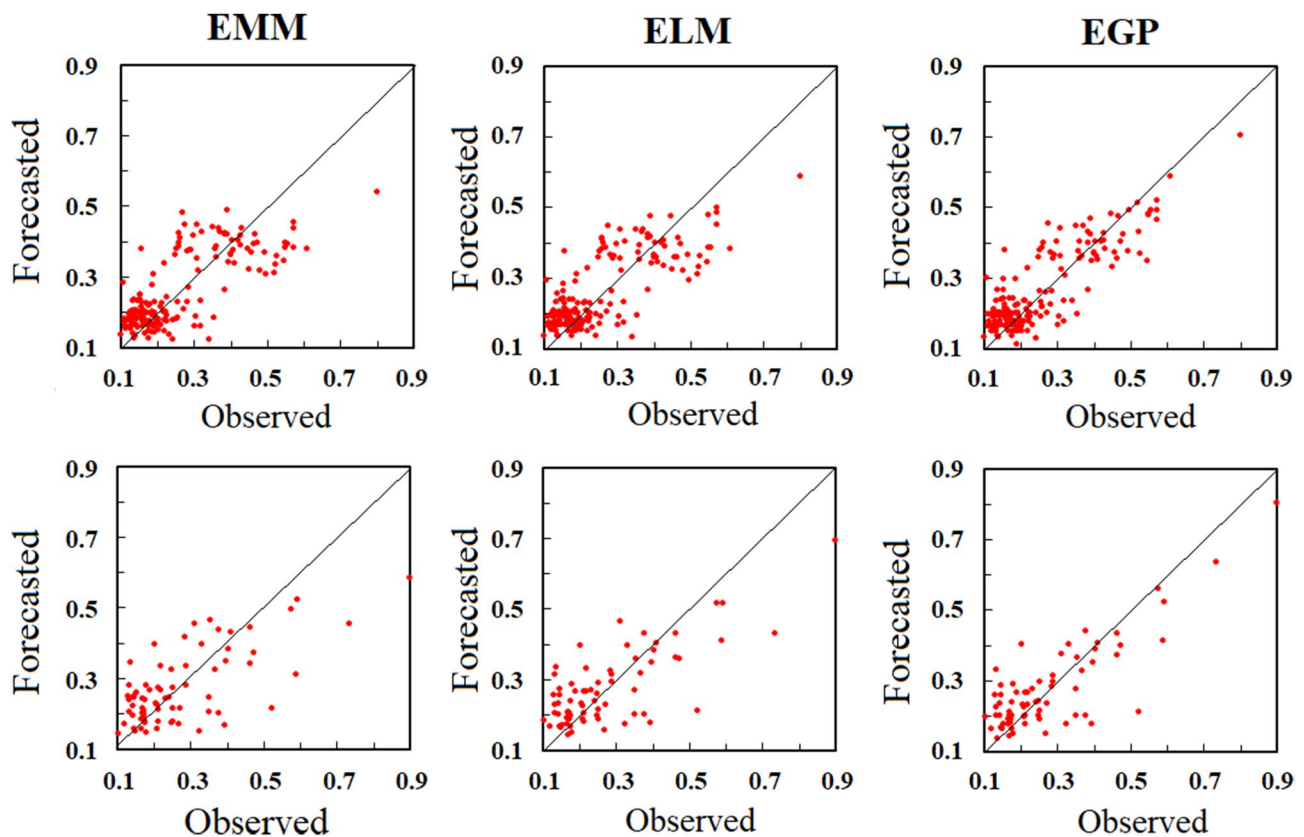
Fig. 11 Scatter plots of the estimations of ensemble models vs. observed normalized TSP (top panel) and testing (bottom panel) data sets

with the findings of the previous studies such as Rahmani-Rezaeieh et al. [31] and Danandeh Mehr and Safari [10] where the ensemble GP approach was introduced as the robust modeling technique in hydrological studies. In comparison with simple averaging and linear ensemble, the forecasted values of TSP using the new EGP model were closer to those of observed data set in both training and testing samples.

As noted previously, each of the evolved models has specific advantages and shortcomings. The MLR is so simple and easy to develop, use, and interpret but it underestimates heavy rainfall seasons. The GP and GEP models always produce better forecasts than MLR. But they are more complex. The ensemble models lead to higher performance by combining the results so that the underestimation problem is solved. As can be seen from Tables 2 and 3, the ability of nonlinear ensemble method, i.e., EGP, is higher than that of linear and arithmetic averaging ensembles. In terms of NSE, the best nonlinear ensemble technique (EGP) improved the modeling efficiency of the best standalone technique (GEP) by 25% and 35% in the training and testing periods, respectively. It is worth mentioning that the performance results tabulated in Tables 2 and 3 are the average

of the entire time series modeled and validated at the training and testing periods. Therefore, it is difficult to quantify the model's performance in each season. To address this problem, the residuals in at each season were calculated and the sum of the normalized absolute error of each model in each season is depicted in Fig. 13. It is seen that all the models show better performance in spring followed by autumn.

Returning to the relevant literature, it is seen that the implementation of AI-based models in precipitation forecasting has received more attention in recent years. However, the existing studies mostly focused on daily or monthly scenarios. The higher the modeling horizon, the lower the forecasting accuracy. For example, the standalone GEP model suggested by Mehr [25] for monthly rainfall forecasting in two rain gauge stations in Iran produces the maximum NSE values of 0.53 in the testing period. Such performance was considered as a good accuracy for monthly rainfall forecasting. For a seasonal forecasting scenario, such a level of accuracy can be considered as a desirable accuracy [28]. Thus, the proposed EGP model is an appropriate alternative for seasonal precipitation forecasting.
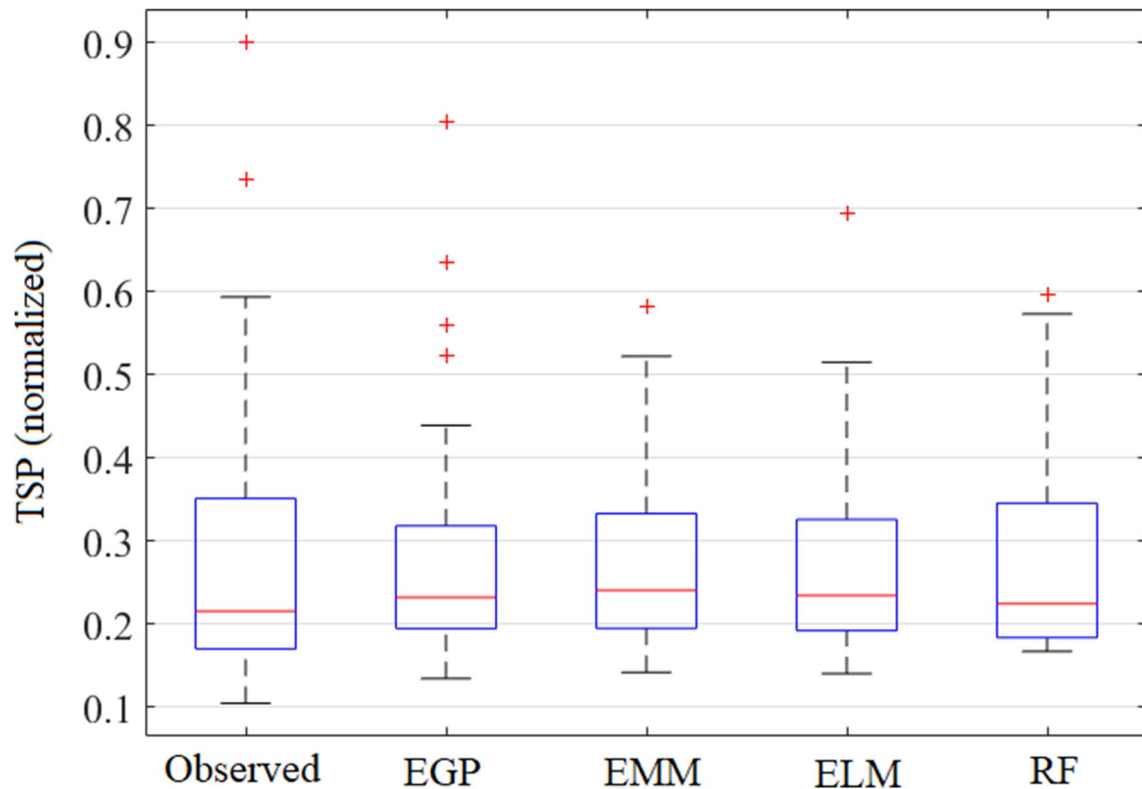
Fig. 12 Boxplot of normalized observed and forecasted seasonal precipitation for Antalya. The boxes indicate the 25th, 50th, and 75th percentiles. The ends of the whiskers indicate the lowest (highest) TSP within 1.5 times the interquartile range of the lower (upper) quartile. Cross-mark denotes extreme precipitation season considered as outlier
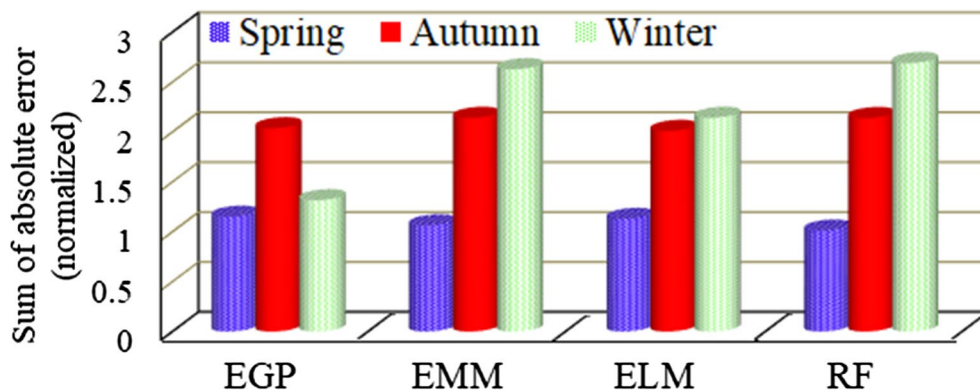

Fig. 13 Sum of the normalized absolute error of the ensemble models in different seasons

6 Summary and conclusion

A three-stage hybrid modeling approach was suggested for TSP prediction in this study. The model was evolved and verified using long-term observed precipitation from Antalya, Turkey. To this end, first, the MLR, GP, and GEP-based models were developed and compared with each other. Afterward, three ensemble techniques were employed to enhance forecasting accuracy. In the ensemble modeling stage, the outputs of the best evolved MLR, GP, and GEP methods were used as the TSP predictors. The results were three hybrid ensemble models, namely EMM, ELM, and EGP in which both linear and nonlinear combination schemes were considered. The GP and GEP models used in the present study are from emerging AI techniques that are known as grey-box models with explicit model structures like a tree.

The standalone modeling stage showed that the ad hoc GP and GEP are not well enough for TSP forecasting. This agrees with findings in the relevant literature in which GEP was reported as an imprecise engine for monthly precipitation forecasting [25]. Comparing the efficiency results of the ensemble models with those of RF, it was shown that the EGP was the best ensemble model. It outperformed all the standalone and ensemble models that implies the combination of different modeling outputs could reduce uncertainties of individual models. The models evolved in the present study were limited to the historical TSP as the predictors. The evolution of similar models using large-scale climate indicators is suggested as a topic for future studies.

Compliance with ethical standards

Conflict of interest The author states that there are no conflicts of interest.

References

1. Adefisan E (2018) Climate change impact on rainfall and temperature distributions over west Africa from three IPCC scenarios. J Earth Sci Clim Change 9:476
2. Aksoy H, Dahamsheh A (2009) Artificial neural network models for forecasting monthly precipitation in Jordan. Stoch Environ Res Risk Assess 23(7):917–931
3. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
4. Chau KW, Wu CL (2010) A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. J Hydroinform 12(4):458–473
5. Çiçek İ, Türkoğlu N, Ceylan A, Korkmaz N (2006) Seasonal rainfall intensity and frequency in Turkey. In: Proceedings book of conference on water observation and information system for decision support, Ohrid, Republic of Macedonia, pp 23–26
6. Danandeh Mehr A, Nourani V, Kahya E, Hrnjica B, Sattar AM, Yaseen ZM (2018) Genetic programming in water resources engineering: a state-of-the-art review. J Hydrol 566:643–667
7. Danandeh Mehr A, Kahya E (2017a) Climate change impacts on catchment-scale extreme rainfall variability: case study of Rize Province, Turkey. J Hydrol Eng 22(3):05016037
8. Danandeh Mehr A, Kahya E (2017b) Grid-based performance evaluation of GCM–RCM combinations for rainfall reproduction. Theor Appl Climatol 129(1–2):47–57
9. Danandeh Mehr A, Nourani V, Karimi Khosrowshahi V, Ghorbani MA (2018) A hybrid support vector regression—firefly model for monthly rainfall forecasting. Int J Environ Sci Technol 1–12:643–667
10. Danandeh Mehr A, Safari MJS (2020) Multiple genetic programming: a new approach to improve genetic-based month ahead rainfall forecasts. Environ Monit Assess 192(1):25
11. Danandeh Mehr A, Jabarnejad M, Nourani V (2019) Pareto-optimal MPSA-MGGP: a new gene-annealing model for monthly rainfall forecasting. J Hydrol 571:406–415
12. Delleur JW, Kavvas ML (1978) Stochastic models for monthly rainfall forecasting and synthetic generation. J Appl Meteorol 17(10):1528–1536
13. Ferreira C (2006) Gene expression programming: mathematical modeling by an artificial intelligence, vol 21. Springer, Berlin
14. He X, Guan H, Qin J (2015) A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall. J Hydrol 527:88–100
15. Hossain I, Esha R, Alam Imteaz M (2018) An attempt to use non-linear regression modelling technique in long-term seasonal rainfall forecasting for australian capital territory. Geosciences 8(8):282
16. Hrnjica B, Danandeh MA (2019) Optimized genetic programming applications: emerging research and opportunities: emerging research and opportunities. IGI Global, Hershey, pp 1–310
17. Hrnjica B, Mehr AD, Behrem Š, Ağıralioğlu N (2019) Genetic programming for turbidity prediction: hourly and monthly scenarios. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi 25(8):992–997
18. Johny K, Pa, ML, Adarsh S (2020) Adaptive EEMD-ANN hybrid model for Indian summer monsoon rainfall forecasting. Theor Appl Climatol 141:1–17
19. Kirsta YB, Lovtskaya O'lga V (2020) Spatial year-ahead forecast of air temperature and precipitation in large mountain areas. SN Appl Sci 2:1044
20. Kisi O, Shiri, J (2011) Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. Water Resour Manag 25(13):3135–3152
21. Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection, vol. 1. MIT press
22. Kuwajima JI, Fan FM, Schwanenberg D, Dos Reis AA, Niemann A, Mauad FF (2019) Climate change, water-related disasters, flood control and rainfall forecasting: a case study of the São Francisco River, Brazil. Geol Soc Lond Spec Publ 488(1):259–276
23. Liu S, Tai H, Ding Q, Li D, Xu L, Wei Y (2013) A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. Math Comput Model 58(3–4):458–465
24. Luk KC, Ball JE, Sharma A (2000) A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. J Hydrol 227(1–4):56–65
25. Mehr AD (2018) Month ahead rainfall forecasting using gene expression programming. Am J Earth Environ Sci 1(2):63–70
26. Mehr AD, Safari MJS, Nourani V (2021) Wavelet packet-genetic programming: a new model for meteorological drought hindcasting. Teknik Dergi 32(4)
27. 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
28. 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
29. Nourani V, Behfar N, Uzelaltinbulat S, Sadikoglu F (2020) Spatiotemporal precipitation modeling by artificial intelligence-based ensemble approach. Environ Earth Sci 79(1):6
30. Özkubat G, Selim S (2019) Socio-economic development of provinces in turkey: a spatial econometric analysis. Alphanumer J 7(2):449–470
31. Rahmani-Rezaeieh A, Mohammadi M, Mehr AD (2020) Ensemble gene expression programming: a new approach for evolution of parsimonious streamflow forecasting model. Theor Appl Climatol 139(1–2):549–564
32. Reddy MJ, Maity R (2007) Regional rainfall forecasting using large scale climate teleconnections and artificial intelligence techniques. J Intell Syst 16(4):307–322

33. Roulin E (2007) Skill and relative economic value of medium-range hydrological ensemble predictions. Hydrol Earth Syst Sci 11(2):725–737

34. Samantaray S, Tripathy O, Sahoo A, Ghose DK (2020) Rainfall forecasting through ANN and SVM in Bolangir Watershed, India. In: Satapathy SC, Bhateja V, Das S (eds) Smart intelligent computing and applications. Springer, Singapore, pp 767–774

35. Sivapragasam C, Liong SY, Pasha MFK (2001) Rainfall and run-off forecasting with SSA–SVM approach. J Hydroinformatics 3(3):141–152

36. Tür R (2020) Maximum wave height hindcasting using ensemble linear-nonlinear models. Theor Appl Climatol 141:1151–1163

37. Tyralis H, Papacharalampous G, Langousis A (2019) A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water 11(5):910

38. Vano JA, Miller K, Dettinger MD, Cifelli R, Curtis D, Dufour A et al (2019) Hydroclimatic extremes as challenges for the water management community: lessons from Oroville Dam and hurricane harvey. Bull Am Meteorol Soc 100(1):S9–S14

39. Van den Bergh J, Roulin E (2016) Postprocessing of medium range hydrological ensemble forecasts making use of reforecasts. Hydrology 3(2):21

40. Wang D, Hagen SC, Alizad K (2013) Climate change impact and uncertainty analysis of extreme rainfall events in the Apalachicola River basin, Florida. J Hydrol 480:125–135

41. Wu J, Liu M, Jin L (2010) A hybrid support vector regression approach for rainfall forecasting using particle swarm optimization and projection pursuit technology. Int J Comput Intell Appl 9(02):87–104

42. Yaseen ZM, Ebtehaj I, Kim S, Sanikhani H, Asadi H, Ghareb MI et al (2019) Novel hybrid data-intelligence model for forecasting monthly rainfall with uncertainty analysis. Water 11(3):502

43. Yevjevich V (1987) Stochastic models in hydrology. Stoch Hydrol Hydraul 1(1):17–36

44. Zhang CJ, Zeng J, Wang HY, Ma LM, Chu H (2020) Correction model for rainfall forecasts using the LSTM with multiple meteorological factors. Meteorol Appl 27(1):e1852

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.