

Geomorphology-based genetic programming approach for rainfall–runoff modeling

Vahid Nourani, Mehdi Komasi and Mohamad Taghi Alami

ABSTRACT

Nowadays, artificial intelligence approaches such as artificial neural network (ANN) as a self-learn non-linear simulator and genetic programming (GP) as a tool for function approximations are widely used for rainfall–runoff modeling. Both approaches are usually created based on temporal characteristics of the process. Hence, the motivation to present a comprehensive model which also employs the watershed geomorphological features as spatial data. In this paper, two different scenarios, separated and integrated geomorphological GP (GGP) modeling based on observed time series and spatially varying geomorphological parameters, were presented for rainfall–runoff modeling of the Eel River watershed. In the first scenario, the model could present a good insight into the watershed hydrologic operation via GGP formulation. In the second scenario, an integrated model was proposed to predict runoff in stations with lack of data or any point within the watershed due to employing the spatially variable geomorphic parameters and rainfall time series of the sub-basins as the inputs. This ability of the integrated model for the spatiotemporal modeling of the process was examined through the cross-validation technique. The results of this research demonstrate the efficiency of the proposed approaches due to taking advantage of geomorphological features of the watershed.

Key words | Eel River watershed, genetic programming, geomorphology, rainfall–runoff modeling, spatiotemporal modeling

Vahid Nourani (corresponding author)
Department of Water Resources Engineering,
Faculty of Civil Engineering,
University of Tabriz,
Tabriz,
Iran
E-mail: vnourani@yahoo.com;
vnourani@umn.edu

Mehdi Komasi
Faculty of Civil Engineering,
University of Tabriz,
Tabriz,
Iran
and
University of Ayatollah Boroujerdi,
Boroujerd,
Iran

Mohamad Taghi Alami
Faculty of Civil Engineering,
University of Tabriz,
Tabriz,
Iran

NOTATION

$Q_{\text{outlet}(t)}$	outlet runoff at time t
$Q_i(t)$	runoff value at time t for i -th sub-basin
i	sub-basin index
$I_{(t-\alpha\Delta t)}$	rainfall value with $\alpha\Delta t$ lag time
Δt	modeling time scale
α	index of rainfall lag
$Q_{(t-\beta\Delta t)}$	runoff value with $\beta\Delta t$ lag time
β	index of runoff lag
$Q_i(t-\beta)$	runoff value with β day lag time at i -th sub-basin
$I_i(t-\alpha)$	rainfall value with α day lag time at i -th sub-basin
A	area
S	slope
CN	curve number
A_T	total area of watershed
\bar{S}	mean slope of watershed
$\overline{\text{CN}}$	mean curve number of watershed

ARIMA	auto regressive integrated moving average
GA	genetic algorithm
ANN	artificial neural network
GP	genetic programming
GGP	geomorphological genetic programming
SGGP	separated geomorphological genetic programming
IGGP	integrated geomorphological genetic programming
DC	determination coefficient
RMSE	root mean squared error

INTRODUCTION

Rainfall–runoff process is the most important and very complex phenomenon in the hydrology cycle. Since rainfall–runoff modeling involves a rather complex non-linear data

pattern, time varying, spatially distributed and not easily described by simple models, there are many novel modeling approaches to improve the modeling accuracy. A detailed classification of different approaches for rainfall–runoff modeling has been presented by [Nourani *et al.* \(2007\)](#).

Nowadays, black-box models based on artificial neural network (ANN) are developed and used for simulation of the rainfall–runoff process (e.g., [Tokar & Johnson 1999](#); [ASCE 2000](#); [Tokar & Markus 2000](#); [Sudheer *et al.* 2000](#); [Rajurkar *et al.* 2004](#); [Garbrecht 2006](#); [Nourani *et al.* 2009](#)). The ANN models are often based on the temporal characteristics of the process and the spatial patterns of the process are usually neglected. But it is obvious that there is a strong relationship between the geomorphological characteristics of a watershed and its hydrologic response. For this reason, several studies have argued for developing more cautious approaches that include considerations of relevant physics and statistical principles and to make ANN models more useful and practical tools. These models are developed in recognition of physical parameters of the watershed in the data-driven modeling to improve the modeling efficiency. For instance, [Wilby *et al.* \(2003\)](#) showed how the hidden layer neurons in an ANN can be related to different flow regimes or hydrological drivers. From this point of view, [Jain *et al.* \(2004\)](#) investigated the identification of physical processes inherent in the ANN-based rainfall–runoff models. The results of the research demonstrate that the hidden neurons in the ANN rainfall–runoff model approximate various components of the hydrologic system such as infiltration, base flow, delayed and quick surface flow, and other parameters. Also, [Zhang & Govindaraju \(2003\)](#) developed an ANN that explicitly describes within its architecture the geomorphological characteristics of the watershed. This geomorphology-based ANN (GANN) was utilized to estimate the runoff hydrographs from several storms over two Indiana watersheds. The architecture of the GANN as well as a part of the network connection strengths were determined by watershed geomorphology leading to a parsimonious modeling tool. It was concluded that GANN offers a promising step towards elevating the ANN model from purely empirical models to those based on geomorphology. Also, in another study, [Sarangi *et al.* \(2005\)](#) concluded that when the geomorphological parameters (such as slope, area, etc.) are associated with rainfall depth and peak runoff rate, the

prediction errors of statistical parameters are less for both ANN and linear regression models. Thus, associating the selected geomorphological parameters enhances the accuracy of the watershed's runoff predictions.

The published ANN models mentioned simulate the runoff only at the outlet of the watershed, as the sole output, but sometimes there is a need to predict runoff at intermediate points or stations within a watershed. Although the merit of ANN in temporal modeling of hydrological phenomena has already been confirmed by many researchers, only a few attempts have been made to assess the ANN ability in spatial and/or spatiotemporal modeling. For instance, three alternative models were presented by [Luk *et al.* \(2001\)](#) based on the use of ANNs to employ the pattern recognition methodology which attempts to forecast short-duration rainfall at specific locations within a catchment. Also, [Mutlu *et al.* \(2008\)](#) developed and compared ANN models to forecast daily flows at multiple gauging stations in an agricultural watershed. Feasibility of the ANN approach for spatiotemporal forecasting of groundwater level in aquifers was also evaluated by [Nourani *et al.* \(2008, 2011a\)](#). Recently, [Nourani & Kalantari \(2010\)](#) presented an integrated ANN model for spatial and temporal forecasting of daily suspended sediment load. The proposed model showed satisfactory performance in spatiotemporal simulation of daily suspended sediment load over the watershed by imposing the geomorphological characteristics of the basin into the input layer. In the context of the integrated model, [Demirel *et al.* \(2011\)](#) used the cluster analysis for validation of a multi-station ANN modeling. They concluded that a multi-station-based approach can be used as an additional validation criterion and might result in rejection of a model which initially passed a single-station validation criterion. So, the approach can be used to assess the inter-station relations based on observations and predictions.

Although ANNs are useful tools in hydrological time series modeling, the obvious disadvantage of ANNs is that they represent their knowledge in terms of weight matrix that is not accessible to human understanding at present. In addition, the number of inputs and/or hidden layer neurons is not clearly determined and should be obtained by a trial and error process which is usually a time-consuming procedure. Therefore, it is still necessary to develop an explicit model to overcome the limitations. From this point of

view, the genetic programming (GP) model is an evolutionary computing method that provides transparent and structured system identification (Savic *et al.* 1999). The motivation for this model is that the GP model has some advantages which allow mitigating some limitations of ANN models. These advantages can be summarized as follows (Falco *et al.* 2005): (1) generation of explicit model representations amenable to easy human comprehension, (2) automatic discovering of the model structure from the given data, (3) adaptive evolutionary search that escapes trapping in suboptimal, unsatisfactory local solutions, and (4) absence of specific knowledge.

The results of studies which employed the GP models indicate that the GP formulation performs quite well when compared with the obtained results by other data-driven approaches (e.g., ANN) in hydrological modeling (e.g., Dorado *et al.* 2003; Wang *et al.* 2009; Kisi & Guven 2010; Guven & Kisi 2011; Sreekanth & Datta 2011). For instance, Sivapragasam *et al.* (2007) found that there is no significant difference in the prediction accuracy between GP and ANN models for forecasting daily flows, but the GP model has the advantage of identifying the optimum inputs. Also, Makkeasorn *et al.* (2008) compared GP and ANN models for forecasting river discharges. The findings of the research indicated that GP-derived stream flow forecasting models were generally favored for forecasting over ANNs. Selle & Muttil (2011) showed that GP can be used to evaluate the structure of hydrological models and to gain insight into the dominant components of the hydrological systems. Also, some aspects of GP in hydraulic engineering were mentioned by Babovic *et al.* (2001), Xu *et al.* (2011), Sharifi *et al.* (2011), and Azamathulla (2012). Giustolisi (2004) determined Chezy resistance coefficient in the corrugated channels using GP. In the field of sediment transport modeling, Aytek & Kisi (2008) applied the GP approach to model daily suspended sediment–discharge relationship. It is apparent that there are a few studies in the literature on the use of GP in the field of rainfall–runoff modeling. Cousin & Savic (1997), Savic *et al.* (1999), Drecourt (1999), Whigham & Crapper (2001), Babovic & Keijzer (2002), Muttil & Liang (2004), Jayawardena *et al.* (2006), and Rodríguez-Vázquez *et al.* (2012) applied GP for rainfall–runoff modeling. Recently, Nourani *et al.* (2012a) used wavelet analysis and GP to determine the dominant inputs of the ANN model for rainfall–runoff

modeling. Hence, it is quite practical to use the GP model as well as other data-driven models in rainfall–runoff modeling.

As mentioned, there are only a few studies (e.g., Zhang & Govindaraju 2003; Sarangi *et al.* 2005; Nourani *et al.* 2012b) on GANN modeling of hydrological processes, but to the best of the authors' knowledge there is no study that has incorporated the spatially varying geomorphological features of a watershed into a GP structure to create an effective modeling tool. The geomorphological parameters of watershed, such as curve number, slope, and area can be used as terminals in a GP formulation. Therefore, in view of the importance of geomorphological parameters in rainfall–runoff modeling, the present study was undertaken to develop geomorphological GP (GGP) models that could be used to provide reliable and accurate estimation of watershed runoff with physical interpretation. Also, a model needs to be designed that can be used to predict runoff in the stations with lack of data or any desired point within the watershed. In this way, first, the conventional or classic GP model is used for daily rainfall–runoff modeling of the watershed. Then, two different GGP scenarios as separated geomorphological GP (SGGP) and integrated geomorphological GP (IGGP) are presented by imposing the geomorphological characteristics of the watershed into the GP structure.

The arrangement of this article is as follows. In the next section, an in-depth introduction for the basic concept of GP is given. The following section then explains the architecture of the proposed GGP models. In the next two sections, the efficiency criteria and study area are introduced and then, the performance of the models is evaluated and compared. The final section concludes this paper by pointing out current problems and future directions.

GENETIC PROGRAMMING (GP)

GP is an evolutionary computing method that generates a transparent and structured representation of the system being studied (Koza 1992). The nature of GP allows the user to gain additional information on how the system performs, i.e., gives an insight into the relationship between input and output data. The GP is similar to genetic algorithm (GA) but unlike the latter, its solution is a computer program or an equation as opposed to a set of numbers in

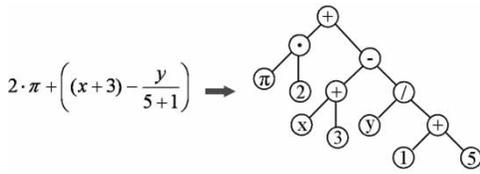


Figure 1 | Tree structure of GP for a mathematical expression.

the GA. So, GP is more attractive than traditional GA for problems that require the construction of explicit models (Savic et al. 1999). The GA and GP deal with two different structures but the solution procedure of both algorithms is similar. The GP is based on tree structure. For example, the outline of a simple GP structure for a sample mathematical expression is shown in Figure 1.

In GP, a random population of individuals is created, the fitness of individuals is evaluated and then, the parents are selected from these individuals. The parents are then made to yield offspring by following the process of reproduction, mutation, and crossover (Sivanandam & Deepa 2008). The creation of offspring continues (in an iterative manner) until a specified number of offspring in a generation are produced and further until another specified number of generations is

created. The resulting offspring at the end of these processes (an equation or a computer program) is the solution of the problem. The GP thus transforms one population of individuals into another in an iterative manner by applying operators. In evolutionary computation, it can distinguish between three different types of operators which are named crossover, reproduction, and mutation (Sivanandam & Deepa 2008). These operators are briefly described below.

- *Crossover operator*: Two parent individuals are selected and a sub-tree is selected on each one. Then, crossover exchanges the nodes and their relative sub-trees from one parent to the other (Figure 2).
- *Mutation operator*: The mutation operator can be applied to either a function node or a terminal node. A node in the tree is randomly selected. If the chosen node is a terminal, it is simply replaced by another terminal. If it is a function and point mutation is to be performed, it is replaced by a new function with the same arity (Figure 3).
- *Reproduction operator*: The reproduction operator simply chooses an individual in the current population and copies it without changes into the new population.

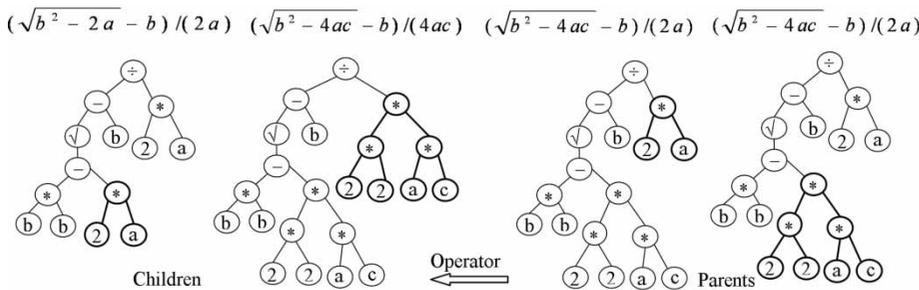


Figure 2 | Crossover operator.

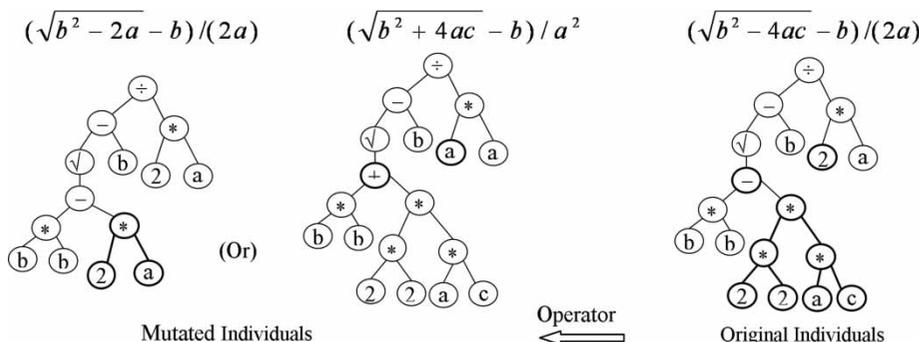


Figure 3 | Mutation operator.

The classic GP model is able to provide the modeling and prediction of a time series automatically. Given a fitness function, the forecasting problem becomes the search of the model which describes the essential characteristics of time series. This modeling process leads to a function which maps input values onto an output value. In the rainfall-runoff modeling, the input variables for the classic GP model are usually chosen as rainfall at current time step and some previous rainfall and runoff values (Figure 4).

The arithmetical functions such as four basic operators (+, -, ×, /) and some basic mathematical functions ($\sqrt{\quad}$, $\log(x)$, \exp , 10^x , power) can be used to create the initial population. Also, as shown in Figure 4, genetic operators (crossover, mutation, and reproduction) can be used to create a new generation in the GP process.

PROPOSED GEOMORPHOLOGICAL GP MODELS

The geomorphology-based GP model is created based on the observed rainfall and runoff time series and spatially varying geomorphological parameters. In other words, the input unit of the GGP model consists of the rainfall and runoff time series and geomorphological characteristics of the watershed's sub-basins. The schematic of the proposed GGP model is shown in Figure 5.

In Figure 5, i indicates the sub-basin. α and β are the lag times for rainfall (I) and runoff (Q) data, respectively. A_i , S_i , and CN_i are area, slope, and curve number for sub-basin i , respectively. Since the Eel River watershed has been made up of many smaller watersheds (sub-basins) and the GP-based models are developed based on these sub-basin data,

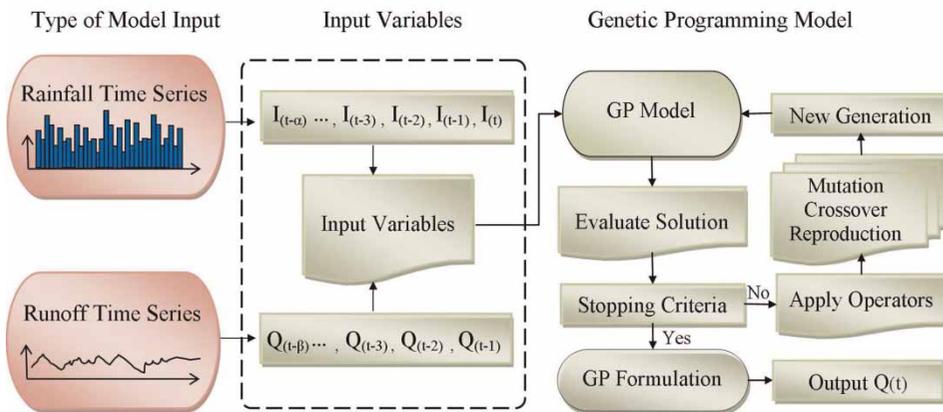


Figure 4 | The schematic structure of classic GP model for daily rainfall-runoff modeling.

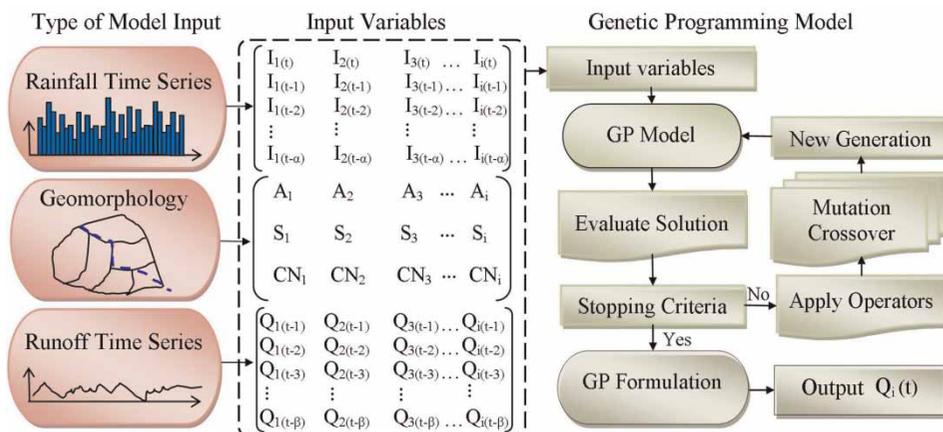


Figure 5 | Schematic representation of GGP model for daily rainfall-runoff modeling.

it is necessary to use the independent and important watershed characteristics such as area, slope, and curve number which can be defined for each sub-basin. In other words, the aim is to derive models for all over the watersheds. Hence, the considered input variables in the model should be defined for any desired point or area within the watershed. By considering the mentioned points, the area, slope, and curve number were used as model inputs. These parameters are the primary variables used by most of the regression equations. The drainage area (A) is probably the single most important watershed characteristic for hydrologic design. It reflects the discharge that can be generated from rainfall. Thus the drainage area is required as input to models ranging from simple linear prediction equations to complex computer models. Also, flood magnitudes reflect the momentum of the runoff and it is obvious that slope (S) is an important factor in the momentum. The curve number (CN) is often used for predicting direct runoff or infiltration from rainfall excess. Most of the conceptual models employ such geomorphological parameters to transfer rainfall to runoff (Beven & Kirkby 1979; Sarangi et al. 2005).

As shown in Figure 5, in addition to temporal variables (i.e., rainfall (I) and runoff (Q) time series), the spatial variables (i.e., geomorphological parameters) are considered as inputs of the model. Regardless of input variables, the selection of appropriate input structures for the GGP model is the most important step in the modeling process. From this point of view, two different scenarios, SGGP and IGGP,

are considered to take advantage of both spatial and temporal dynamic variables in time series modeling of the process. The main difference between the SGGP and IGGP models is related to the structure of the input variables imposed on the models.

SGGP model

The SGGP model should be utilized for each sub-basin of the watershed, separately. The schematic structure of the input and output units is presented for the SGGP model in Figure 6. In addition to the rainfall and runoff, the geomorphological parameters are also considered as the model inputs. The temporal variables (i.e., rainfall and runoff values) are arranged in the input matrix according to the Markovian property of the process. Since GP has the ability to select variables that contribute beneficially to the model and to disregard those that do not, it is expected that the GP evolved equations would contain the most significant rainfall and runoff variables. In the SGGP model, the geomorphological parameters are constant for each sub-basin. Hence, the SGGP model should not be categorized as a spatiotemporal model.

IGGP model

The proposed IGGP model can be used for rainfall-runoff modeling of the entire watershed (Figure 7). In the IGGP

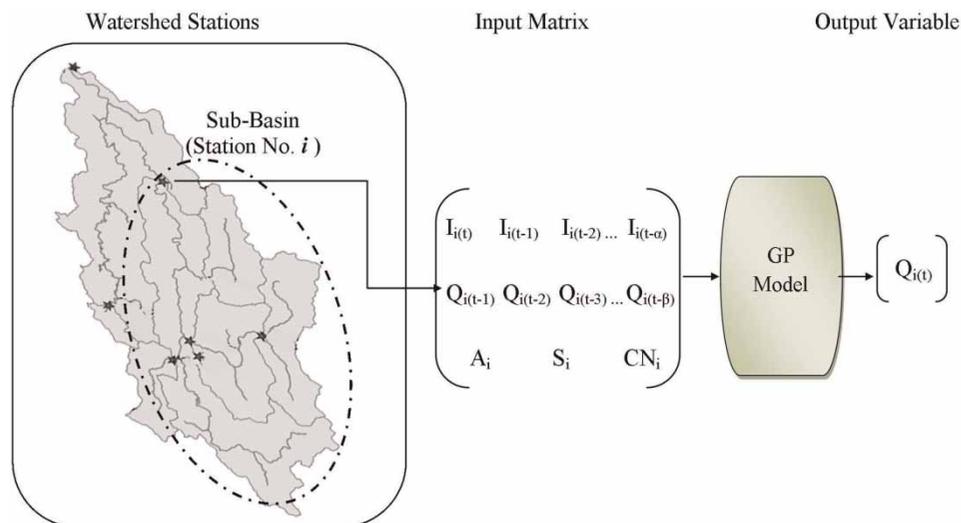


Figure 6 | Schematic structure of SGGP model.

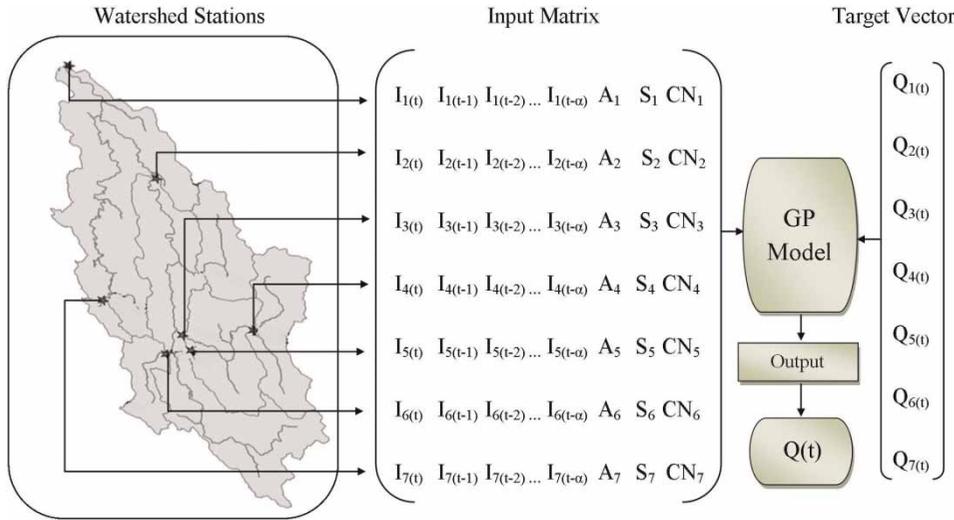


Figure 7 | Schematic structure of IGPP model.

model, the input variables consist of different sets of previous and current rainfall values and geomorphological characteristics of the sub-basins relevant to all stations to predict the runoff time series of the stations ($Q_{i(t)}$, $i = 1, 2, 3, \dots$). As shown in Figure 7, the IGPP model aims to present a single general model that can perform well enough to be used instead of several SGGP models for any station inside the watershed. Furthermore, the proposed IGPP model can be used to predict runoff in stations with lack of data or any desired point within the watershed because of employing spatially variable parameters and rainfall time series of the sub-basins as the model inputs.

EFFICIENCY CRITERIA

In time series modeling, it is common to split the total data into training (calibration) and test (verification) sub-sets. For this purpose, the data set was divided into two parts: the first 75% of total data was used as the training set and the second 25% for verifying the models. In an evolutionary model such as the GP model, when an individual emerges whose sum of absolute errors is less than a given value, the training of the model must be stopped. In this process, in order to evaluate the model performance, different efficiency criteria are used in the training process. Finally, the

prediction accuracy of the optimized model is verified using the verification data.

The model that yields the best results in terms of determination coefficient (DC; Nash & Sutcliffe 1970) and root mean squared error (RMSE) in the training and verifying steps can be determined through a process of trial and error. For this purpose, the following measures of evaluation are used to compare the performance of the different models (Nourani 2009):

$$DC = 1 - \frac{\sum_{i=1}^N (Q_{obs_i} - Q_{com_i})^2}{\sum_{i=1}^N (Q_{obs_i} - \bar{Q}_{obs})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Q_{obs_i} - Q_{com_i})^2}{N}} \quad (2)$$

where N , Q_{obs_i} , Q_{com_i} and \bar{Q}_{obs} are the number of observations, observed data, computed values and mean of observed data, respectively. The RMSE is used to measure forecast accuracy which produces a positive value by squaring the errors. The RMSE increases from zero for perfect forecasts through large positive values as the discrepancies between forecasts and observations become increasingly large. Obviously, high value for DC (up to 1) and small value for RMSE indicate high efficiency of the model. Legates & McCabe (1999) indicated that a hydrological model can be sufficiently evaluated by these two statistics.

STUDY AREA AND DATA

The Eel River watershed in California, USA, was selected for this study. The Eel River watershed with a drainage area of 8,051 km² is located between 39°14'25" and 40°29'43"N latitude, and 122°39'58" and 124°08'31" W longitude. It drains a rugged area between the Sacramento valley and the ocean, crossing Humboldt, Lake, Mendocino, and Trinity counties. For most of its course, the river flows northwest, parallel to the coast. The Eel River has three major tributaries: South Fork Eel River to the left, and Middle Fork Eel River and North Fork Eel River to the right. Figure 8 shows the map of the study area. The length of the longest flow path is 300 km. The major land use/cover of the area is evergreen forest land (75%). The minimum and maximum daily air temperatures are -10 and 35 °C, respectively. The mean daily temperature varies from -1 up to 25 °C over a year for the basin.

The historical daily discharge data are available for nine stations inside the watershed (USGS, <http://co.water.usgs.gov/sediment/seddatabase.cfm>). Each station was considered as an outlet for the sub-basin. Extra information about the stations and related sub-basins are presented in Table 1.

Also, daily rainfall data are available for three rain gauges (HydroLab, <http://hydrolab.arsusda.gov/nicks/nicks.htm>) in which the distance-weighted average of recorded rainfall values in these gauges were considered as the daily rainfall over each sub-basin. Table 2 shows the statistical analysis of the available data. Figure 9 shows the time series of daily rainfall-runoff relevant to the station located at the outlet of the basin (USGS ID = 11477000) as an example.

In the current research, the models were calibrated using the data from December 1966 to September 1969 and were verified by the data set gathered during October 1969 to September 1970. Since station 8 is close to station 4, this station was used for purposes of cross-validation. The runoff data for station 8 were not available for 10 months (i.e., from December 1966 to September 1967), hence the cross-validation was performed for the available time range (i.e., October 1967 to September 1970).

RESULTS AND DISCUSSION

In an evolutionary model such as GP, the model output depends on the input variables which must be investigated

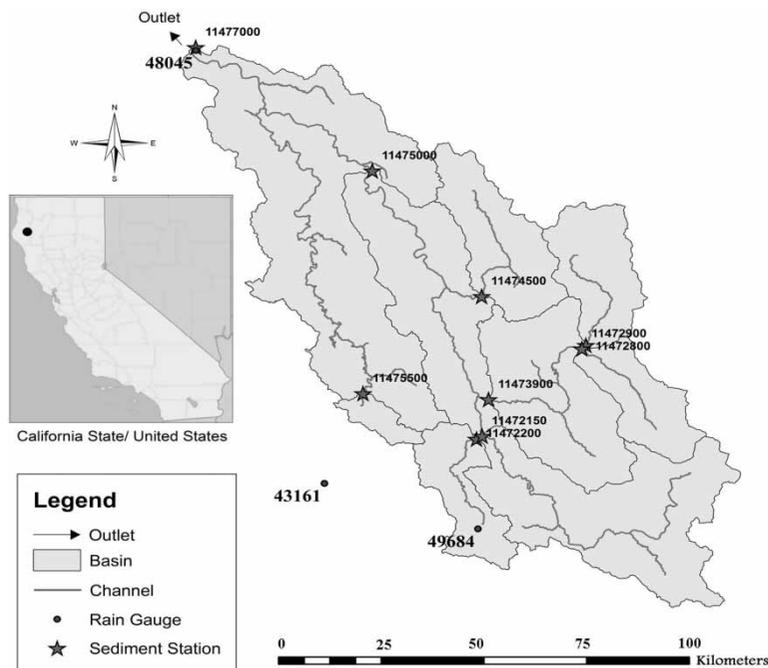


Figure 8 | The Eel River watershed.

Table 1 | Information about stations and related sub-basins at the Eel River watershed

Station No.	Station USGS ID	Longitude	Latitude	Area (km ²)	Slope (%)	Curve Number	Duration of available rainfall and runoff data
1	11477000	124°05'55"	40°29'30"	8,051	37.51	67.01	Dec 1966–Sep 1970
2	11475000	123°37'53"	40°13'06"	5,448	36.85	67.95	Dec 1966–Sep 1970
3	11473900	123°19'27"	39°42'23"	1,923	32.32	68.28	Dec 1966–Sep 1970
4	11472900	123°04'50"	39°49'15"	414	37.88	64.25	Dec 1966–Sep 1970
5	11472150	123°20'25"	39°37'30"	1,368	37.71	68.04	Dec 1966–Sep 1970
6	11475500	123°39'06"	39°43'09"	113	37.05	63.08	Dec 1966–Sep 1970
7	11472200	123°21'20"	39°37'05"	416	29.01	70.07	Dec 1966–Sep 1970
8	11472800	123°04'11"	39°49'45"	527	39.23	64.71	Oct 1967–Sep 1970
9	11474500	123°20'36"	39°56'18"	641	39.97	64.38	Dec 1972–Sep 1975

Table 2 | Statistical analysis of observed data at the Eel River watershed

Runoff data (cm)											
Station No.	Calibration data					Verification data					
	Mean	Min	Max	Standard deviation	Skewness coefficient	Mean	Min	Max	Standard deviation	Skewness coefficient	
1	251.6	2	5,380	544.6	4.7	298.7	2	7,560	815.7	4.7	
2	148.2	1	3,090	316.8	4.8	160.2	0	4,810	457.2	5.5	
3	56.0	0	1,310	115.7	5.0	58.2	0	1,490	169.5	5.4	
4	11.7	0	270	25.3	5.0	13.3	0	286	33.2	4.5	
5	31.2	0	793	77.9	5.2	39.8	0	1,270	129.2	5.9	
6	4.8	0	134	10.4	5.1	5.9	0	168	17.5	5.3	
7	13.3	0	408	36.0	5.3	15.6	0	416	46.9	4.9	
8	22.5	0	861	62.9	8.3	22.9	0	951	73.4	7.2	
9	28.6	0	572	57.7	4.8	2.1	0	16	3.6	2.2	
Rainfall data (mm)											
Gauge (1)	13.3	0	432	39.4	5.1	16.5	0	321	43.8	3.6	
Gauge (2)	11.1	0	358	32.1	4.6	14.0	0	240	36.9	3.5	
Gauge (3)	13.9	0	521	44.9	5.5	19.0	0	436	52.3	4.0	

Gauge (1): USGS NO. 48045; Gauge (2): USGS NO. 43161; Gauge (3): USGS NO. 49684.

by sensitivity analysis. Hence, first, two types of classic GP models only based on temporal variables (i.e., rainfall and runoff time series) were applied to identify the relationship between rainfall and runoff over the Eel River watershed (Figure 4). In the second step, in order to consider the geomorphological parameters of the watershed in the rainfall-runoff modeling, the SGGP models were designed by considering the temporal and geomorphological characteristics of the sub-basins (Figure 6). For this purpose, seven SGGP models for the seven stations (i.e., stations 1 to 7) of the Eel River watershed were calibrated using the data

from December 1966 to September 1969 and verified by the data set gathered during October 1969 to September 1970. In the third step, in order to develop a unique GGP model for the entire watershed, the data of seven sub-basins were imposed into the IGGP model (Figure 7). It is notable that the observed rainfall and runoff time series of station 9 are completely out of the study period. So, this station was not considered in the current modeling. Moreover, the data of station 8 were excluded from the data set to be used for the cross-validation of the proposed IGGP model.

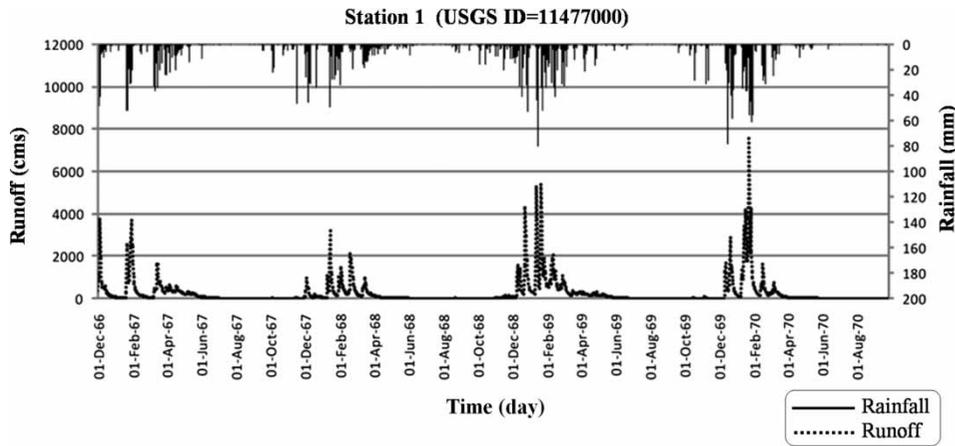


Figure 9 | Daily time series of rainfall and runoff at station 1 (USGS ID = 11477000).

Results of classic GP model

At the first step, the classic GP model was employed for rainfall-runoff modeling of the whole basin to predict runoff discharge at the outlet station of the watershed (i.e., station 1, USGS ID = 11477000). The input variables of the GP model were considered as rainfall at the current and some previous times to predict the outlet runoff at current time step ($Q_{outlet(t)}$). The mathematical relationship could be expressed as a function of f :

$$Q_{outlet(t)} = Q_{1(t)} = f(I(t), I_{(t-\Delta t)}, I_{(t-2\Delta t)}, \dots, I_{(t-\alpha\Delta t)}) \quad (3)$$

In Equation (3), $I_{(t-\alpha\Delta t)}$ is the rainfall value over watershed with $\alpha\Delta t$ lag time. Δt is the modeling time scale which should be substituted by 1 (i.e., $\Delta t = 1$) for daily rainfall-runoff modeling. Since the GP model has the ability to select some variables contributing beneficially to the model and disregard those that do not, the number of input variables can be arbitrarily selected in Equation (3). In other words, the α value can be substituted by high numbers and it is expected that the GP evolved equations would contain the most significant rainfall variables. It is notable that more input variables may lead to a complex and huge formulation. Furthermore, it causes long run time to complete the evolutionary process of the program. In the current research, eight variables were used as inputs for the GP model (i.e., $\alpha = 7$).

In an evolutionary model such as GP, the model output depends not only on the input variables but also initial

population size, sampling method, and the functions type which must be calibrated by sensitivity analysis, correctly. Since the evolution process is a non-deterministic process, it does not end with a successful solution in each program run. So, these parameters must be set by a trial and error method. For this purpose, the program must be processed in several independent runs to obtain a successful solution for the problem. The number of required program runs for the satisfactory solution depends on the difficulty of the problem or model. In the current research, different sets of evolutionary parameters were examined for the GP model and the best combination of the value/method for evolutionary parameters is listed in Table 3.

As shown in Table 3, the arithmetical functions were set up to the GP model. In order to determine which type of function has high fitness and suitable structure, the introductory test runs in the GP model were executed for two different combinations of functions. The basic arithmetical functions (addition, subtraction, multiplication, and division) and the mathematical functions (natural exponential, power, sine and cosine functions) were used in the first and second combinations, respectively. For each of the function

Table 3 | The GP parameters for rainfall-runoff modeling

Parameter	Value/Method
Initial population number	2,000
Sampling method	Tournament
Functions type	Arithmetical
Fitness function	RMSE

combinations, several independent runs were executed with the same condition (i.e., variables, generation number, initial population number, etc.). The results of standard deviation percentage for RMSE in both combinations are presented in Table 4.

As shown in Table 4, the analysis of both combinations shows that the probability of successful solutions will be greatest if the basic arithmetical functions are used as the function genes. Furthermore, the arithmetical functions lead to simple explicit formulations whereas mathematical functions may lead to a complex formulation. Consequently, the arithmetical function genes were used in all subsequent runs in the presented GP and GGP (i.e., SGGP and IGGP) models.

In addition to the parameter selection, the main concern in using a GP model is the training process. When a GP model is trained iteratively in order to improve its performance on the training data, it is possible that the GP formulation finally ‘memorizes’ the training samples and does not ‘learn’ the underlying pattern. This is called an over-fitting (over-training) problem in the artificial intelligence models. This is more likely to happen in the GP model with a large number of generations. For this reason, the GP model with suitable generations is preferred just enough to provide an adequate fit to avoid over-training. Also, the GP formulation should be adjusted only on the account of the training (calibration) set, but the error should be monitored on the validation data set, simultaneously. The error on the validation data will normally decrease during the initial iterations together with the error on the training set and when the GP formulation begins to over-fit the training data, the error on the validation data will begin to rise. For this reason, the calibration and verification errors must be considered in the training process simultaneously. This technique was applied in the modeling process of this paper and it was found that 1,000 generations is a good choice to prevent the over-fitting issue.

Table 4 | Results of GP model for two function combinations

Combination	Function	Percentage of standard deviation for RMSE	
		Calibration step (%)	Verification step (%)
(1)	Arithmetical	4.3	6.8
(2)	Mathematical	9.5	11.3

By considering the factors mentioned about the GP modeling, various runs were conducted and the optimum formulations were selected according to the obtained evaluation criteria in terms of DC and RMSE. The best GP formulation for the outlet of the Eel River watershed was:

$$Q_{1(t)} = 6.22I_{(t)} + 5.82I_{(t-1)} + 3.5I_{(t-2)} + I_{(t-3)} + 0.53I_{(t-4)} \quad (4)$$

As shown in Equation (4), the complex relationship of the rainfall-runoff process can be represented by a few variables with arithmetical operators. When only rainfall data are available, this model can be used. The model efficiency criterion in terms of DC is 0.87 and 0.82 for the calibration and verification steps, respectively. This type of classic GP model has no satisfactory performance, which may be attributed to the insufficient variables in the model structure. From this point of view, the input variables of a classic GP model were chosen as rainfall and runoff at current and some previous times to predict the outlet runoff ($Q_{(t)}$) as a Markovian process. Hence, the mathematical relationship could be expressed as:

$$Q_{1(t)} = f(I_{(t)}, I_{(t-\Delta t)}, I_{(t-2\Delta t)}, \dots, I_{(t-\alpha\Delta t)}, Q_{(t-\Delta t)}, Q_{(t-2\Delta t)}, \dots, Q_{(t-\beta\Delta t)}) \quad (5)$$

In Equation (5), $Q_{(t-\beta\Delta t)}$ denotes the observed runoff value at the outlet station with $\beta\Delta t$ lag time. Since the rainfall and runoff time series have almost the same seasonality and pattern, the α and β values could be substituted by equal value (Nourani et al. 2011b). For this purpose, 15 variables were used as inputs in the GP model consisting of eight rainfall and seven runoff variables (i.e., $\alpha = \beta = 7$). The modeling process led to Equation (6) as:

$$Q_{1(t)} = Q_{(t-1)} - \frac{Q_{(t-1)} + Q_{(t-2)}}{5.76} + \frac{Q_{(t-1)}I_{(t-2)}}{Q_{(t-3)} + I_{(t-2)} + I_{(t-1)}} + 5.76I_{(t)} + I_{(t-1)} - I_{(t-2)} - 12.75 \quad (6)$$

As seen in Equation (6), the variables of the equation indicate that the rainfall and runoff lag times for the basin are almost 2 and 3 days, respectively; thus, examination of the variables with a lag of more than seven is not necessary.

Moreover, the catchment response is highly influenced by the recent rainfall and runoff (i.e., $I_{(t)}$ and $Q_{(t-1)}$) values of the catchment. Since the current rainfall variable ($I_{(t)}$) has been magnified in Equation (6), it can be concluded that previous rainfall conditions that could extend for several days in the past are not significant as relative as to the current rainfall condition in the predicted runoff value. In this step, the model efficiency criterion in terms of DC is 0.93 and 0.89 for the calibration and verification data sets, respectively. This high efficiency can be related to the autoregressive properties of runoff time series which is captured via runoff variables (i.e., $Q_{(t-1)}$ and $Q_{(t-2)}$). However, the major shortcoming in applying such classic GP to rainfall-runoff modeling is that the network does not incorporate geomorphological elements, so does not take into account the spatial dependence. Hence, the model cannot give true insight into the watershed's physical characteristics in the rainfall-runoff modeling. Such a drawback can be removed by the SGGP model which contains the geomorphological features of the basin. Furthermore, in some stations, a major limitation of classic GP is the lack of availability of adequate data to predict runoff value accurately. Also, this model is not able to calculate the runoff in a sub-basin or area which has no station. To circumvent this problem the integrated geomorphological GP model (i.e., IGGP) is proposed.

Results of proposed SGGP model

In the first GGP scenario (i.e., SGGP model), in addition to the temporal variables (i.e., rainfall and runoff time series), geomorphological parameters such as area (A), slope (S) and curve number (CN) of sub-basins were considered as geomorphological inputs of the input unit for the GP model (Figure 6). It is notable that in the conventional data-driven approach such as ANN and GP models, the measurement units can be removed using the normalized data. Since the geomorphological input variables of GGP models (i.e., SGGP and IGGP models) have inconsistent dimensions, one standard approach for avoiding the potential conflicts with incorrect dimensionality of induced formulations is to use dimensionless values. This is a standard scientific practice, as units of measurements are effectively eliminated through the introduction of dimensionless ratios (Babovic et al. 2001). In this way, the dimensionless geomorphological

parameters were imposed onto the SGGP models which can reflect the watershed sub-basins' characteristics. Therefore, the SGGP models for sub-basins could be expressed as:

$$Q_{i(t)} = f\left(I_{i(t)}, I_{i(t-1)}, I_{i(t-2)}, \dots, I_{i(t-\alpha)}, Q_{i(t-1)}, Q_{i(t-2)}, Q_{i(t-3)}, \dots, Q_{i(t-\beta)}, \frac{A_i}{A_T}, \frac{S_i}{\bar{S}}, \frac{CN_i}{\bar{CN}}\right) \quad (7)$$

In Equation (7), i refers to the sub-basin number ($i = 1, 2, 3, \dots, 7$). $Q_{i(t-\beta)}$ is runoff value with β day lag time at i -th sub-basin. $I_{i(t-\alpha)}$ is rainfall value with α day lag time at i -th sub-basin. A , S , and CN are area, slope, and curve number for each sub-basin, respectively. A_T , \bar{S} and \bar{CN} are the total watershed area, mean slope, and mean curve number for the entire watershed, respectively.

Since the runoff is extremely affected by recent watershed conditions in daily time scale modeling (see Equation (6)), only four lag times were used for rainfall and runoff variables (i.e., $\alpha = \beta = 4$) in the GP model. By considering the factors mentioned in the GP modeling, seven SGGP models were executed and the best GP formulations for sub-basins are listed in Equations (8–14) for sub-basins 1–7, respectively. The equations include the temporal variables and geomorphological parameters of the sub-basins.

$$Q_{1(t)} = Q_{1(t-1)} + 5.2I_{(t)} + I_{(t-1)} - \frac{0.80I_{(t-1)}}{6.73 + I_{(t-1)}} + 0.3\frac{A_1}{A_T} \quad (8)$$

$$Q_{2(t)} = Q_{2(t-1)} + I_{(t)} + \frac{A_2}{A_T}I_{(t-1)} - \frac{A_2}{A_T} + \frac{CN_2}{\bar{CN}} \quad (9)$$

$$Q_{3(t)} = I_{(t)} + \left(1 - \frac{1}{\frac{CN_3}{\bar{CN}} + 2\frac{A_3}{A_T}}\right)Q_{3(t-1)} + \frac{A_3}{A_T} \quad (10)$$

$$Q_{4(t)} = \frac{A_4}{A_T}I_{(t)} + Q_{4(t-1)}\left(\frac{S_4}{\bar{S}} - \frac{A_4}{A_T}\right) + 2\frac{S_4}{\bar{S}} + \frac{CN_4}{\bar{CN}} \quad (11)$$

$$Q_{5(t)} = 3Q_{5(t-1)}\left(\frac{1 - \frac{S_5}{\bar{S}}}{\frac{S_5}{\bar{S}}}\right) + \frac{I_{(t)}}{2} + 3\frac{S_5}{\bar{S}} + \frac{CN_5}{\bar{CN}} + \frac{A_5}{A_T} \quad (12)$$

$$Q_{6(t)} = Q_{6(t-1)} \left(1 - \frac{S_6 A_6}{\bar{S} A_T} \right) + I_{(t)} + 3 \frac{A_6}{A_T} + \frac{S_6}{\bar{S} \left(1 + \frac{\frac{S_6}{\bar{S}}}{-Q_{6(t-1)} - 2 \frac{CN_6}{CN}} \right)} \quad (13)$$

$$Q_{7(t)} = Q_{7(t-1)} \left(\frac{S_7}{\bar{S}} - 1 \right) + \frac{Q_{7(t-1)}}{\frac{A_7}{A_T} + Q_{7(t-1)}} + \frac{S_7}{\bar{S}} I_{(t)} + \frac{A_7}{A_T} + \frac{S_7}{\bar{S}} \quad (14)$$

As shown in Equations (8–14), the dimensionless geomorphological parameters of equations yield some interesting physical interpretations regarding the structure of runoff prediction formulation. By comparing Equations (8–14) with Equation (6), it can be concluded that geomorphological parameters were substituted instead of real random constants. In other words, physically meaningless constants could be related to the geomorphological characteristics of the sub-basins. These GGP formulations provide good insight into the watershed. Hence, although the performance of the SGGP models is almost equal to the second classic GP model (see Table 5), the superiority of the SGGP model is related to its significant ability in substituting the terminals with geomorphological parameters in the structure of the classic GP model. This result is in agreement with the research that showed the connection weights of neurons in a GANN model can reflect the geomorphologic properties of the watershed (i.e., Zhang & Govindaraju 2003).

As seen in Equations (8–14), although up to 4 days' lagged data were used in the input unit of the SGGP model

(i.e., $I_{(t-1)}$, $I_{(t-2)}$, $I_{(t-3)}$, $I_{(t-4)}$, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$, $Q_{(t-4)}$), only one day's lagged data (i.e., $I_{(t-1)}$ and $Q_{(t-1)}$) has appeared in the structures of the final formulations and the others have been substituted by geomorphological variables via the evolutionary process of GP. Also, the rainfall with 1 day lag time (i.e., $I_{(t-1)}$) has appeared only in the GP formulations of downstream stations (i.e., stations 1 and 2). This may be attributed to long concentration time of runoff to reach downstream stations compared with the upstream stations. By considering Equations (8–14), it can be concluded that in the upstream sub-basins (i.e., 5, 6, and 7) of the Eel River watershed, numerous geomorphological variables have appeared in the structure of formulations. In these sub-basins, the slope parameter is the main component of formulations. Since the upstream sub-basins have steeper slopes compared with the downstream sub-basins, it could be expected that this parameter plays an important role in the rainfall-runoff process of upstream sub-basins. Table 5 presents an inter-comparison of SGGP model results. As shown in Table 5, the sixth sub-basin (sub-basin 6) has the highest efficiency compared with the other sub-basins. For instance, the comparison of computed and observed runoff time series has been shown for station 6 in Figure 10.

As shown in Figure 10, the SGGP model has underestimated the peak values of time series as similar to the other black box models such as GP, ANN, ANFIS (adaptive neural-fuzzy inference system), ARIMA (auto regressive integrated moving average), etc. (Nourani et al. 2008). These extreme values might have been created because of instantaneous imposition of the extreme inputs to the hydrological system, such as a heavy storm event. Therefore, fully regressive models which use just current and some previous

Table 5 | The SGGP model results for sub-basins

Station No.	Station USGS ID	DC		RMSE (m ³ /s)	
		Calibration	Verification	Calibration	Verification
(1)	11477000	0.94	0.93	82.41	104.28
(2)	11475000	0.93	0.92	49.28	62.46
(3)	11473900	0.94	0.93	10.40	15.86
(4)	11472900	0.91	0.90	5.39	6.58
(5)	11472150	0.94	0.93	5.88	11.52
(6)	11475500	0.94	0.93	1.94	3.79
(7)	11472200	0.94	0.93	3.62	4.53

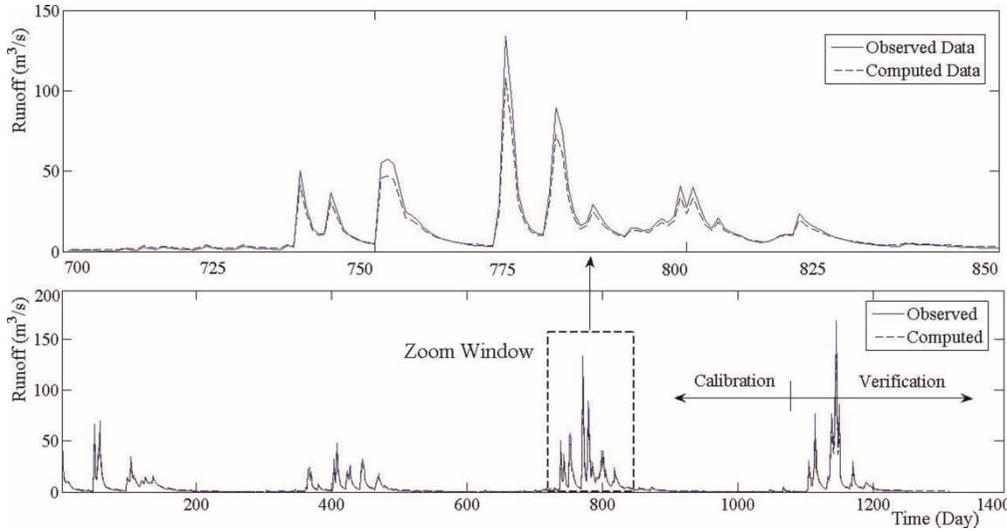


Figure 10 | Computed versus observed runoff time series for station 6.

time steps' data as the model inputs, could not manage the reasonable detection of the instantaneous jumps in the time series.

Results of proposed IGGP model

In this step, in order to develop the second GGP scenario (i.e., IGGP model) for the entire watershed, the data of all seven stations were imposed onto the GP framework as:

$$Q_{(t)} = f \left(\text{Mat} \left[I_{(t)} I_{(t-1)} \dots I_{(t-\alpha)} \frac{A}{A_T} \frac{SCN}{\overline{SCN}} \right] \right) \quad (15)$$

Here, the matrix *Mat*[] involves rainfall data and geomorphological variables of all seven sub-basins in order to train the GP formulation using the target vector of observed runoff time series at stations within a unique and integrated framework (see Figure 7). As mentioned earlier, the arithmetic operators were used and since the runoff is extremely affected by current and previous conditions of the watershed, a four-day window of rainfall data were used to consider the time dependence of the phenomenon (i.e., $\alpha = 3$). Since the runoff variables are not considered in the IGGP model, this model can be applied to stations which have no runoff data or any desired point within the watershed.

Again, several GP runs were examined to determine the best formulation using optimum GP parameters (i.e.,

Tables 3 and 4). After calibration and verification steps and according to the evaluation criteria, the best performance of the model was obtained with approximately 1,000 generations. The best-selected IGGP formulation could be defined by the following equation:

$$\frac{Q_{(t)}}{A_T} = \frac{9(I_{(t)} + I_{(t-1)} + I_{(t-2)})A}{A_T} + 3 \frac{S}{\overline{S}(I_{(t)} + I_{(t-1)})A} \quad (16)$$

where $Q_{(t)}$ is forecasted runoff at any station. $I_{(t)}$, $I_{(t-1)}$, and $I_{(t-2)}$ are rainfall values for current, and 1 and 2 days ago over the related upstream portion of the watershed, respectively.

Equation (16) demonstrates how geomorphologic parameters can be incorporated into the IGGP formulation. It is most pronounced in the conventional GP formulation where the model is trained and simulated based on only the temporal rainfall and runoff time series data, the formulation architecture is determined in terms of time-dependent variables and some physically meaningless random constants as terminals while the IGGP model provides watershed geomorphology parameters as terminals of the formulation. In Equation (16), the curve number (CN) variable does not appear. By referring to Table 1, it can be seen that the curve number value is not significantly variable over the basin. In the other expression, curve number is approximately constant for the entire watershed. Hence it does not

have a significant effect on the evolutionary computation process. Table 6 shows the IGGP model performance using two statistics (i.e., DC and RMSE between observed and computed runoff values) for each sub-basin. In this table, the calibration and verification data periods for the IGGP model are similar to the SGGP model. The results presented in Table 6 show the merit of the proposed model for estimating runoff of stations by employing just one IGGP model instead of seven SGGP models. The reason for this ability may be related to the interaction of stations to each other when their data are imposed to a unique framework. Therefore, the spatial and physical relationship among the stations could be captured through

the GP formulation and the spatial pattern of the phenomenon could be learned and recognized as well as its temporal pattern. Consequently, the model should be classified as a time-space model. This model can operate as a non-linear time-space regression tool rather than a fully lumped model.

According to the results presented in Tables 5 and 6, the IGGP model has a lower efficiency in comparison with the SGGP model. The main reason for this low efficiency is probably related to the input structure of the IGGP model. Since the runoff variables (i.e., $Q_{i(t-1)}$, $Q_{i(t-2)}$, $Q_{i(t-3)}, \dots$) have been disregarded in the input variables, the autoregressive property of runoff time series could not be captured accurately. Nevertheless, the IGGP model is superior to the SGGP model because the IGGP model cannot only be applied for runoff estimation at the sub-basins (or stations) but also to any other desired points within the watershed. For this purpose, the rainfall data and geomorphological characteristics of the upstream sub-basins should be inserted into the IGGP model formulation. As was previously stated, station 8 was selected for the cross-validation of the IGGP model. In order to evaluate the proposed IGGP model by the presented methodology, the available rainfall and runoff data from 1966 to 1967 and geomorphological properties of sub-basin 8 were imposed onto

Table 6 | Evaluation criteria of proposed IGGP model

Sub-basin (station) No.	Calibration step		Verification step	
	DC	RMSE (m ³ /s)	DC	RMSE (m ³ /s)
(1)	0.82	112.16	0.79	151.87
(2)	0.80	99.68	0.80	114.29
(3)	0.84	64.51	0.81	48.82
(4)	0.79	17.33	0.78	20.41
(5)	0.83	24.29	0.80	37.08
(6)	0.82	13.89	0.81	14.25
(7)	0.81	18.45	0.79	20.48

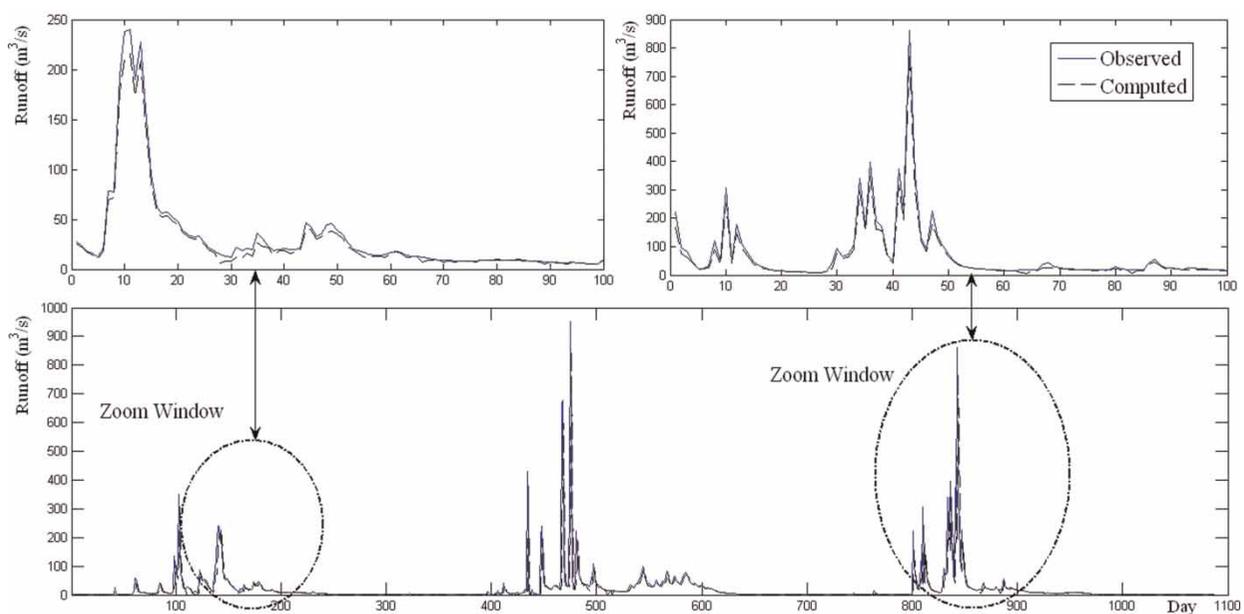


Figure 11 | Computed versus observed runoff time series at cross-validation step for station 8.

the trained IGGP model formulation (i.e., Equation (16)) to simulate the runoff values at this station. The result of cross-validation in term of computed versus observed time series can be seen in Figure 11, which demonstrates the ability of the IGGP model for spatiotemporal forecasting. Also, Figure 12 shows the scatter plots of the computed versus observed runoff values for stations 1–7 and the cross-validation station (i.e., station 8) in the verification step. The scatter plots denote the flexibility of the proposed IGGP model for forecast runoff for different parts of the watershed.

The daily rainfall-runoff is a Markovian process, so that consideration of the process's state at previous times ($Q_{(t-1)}$, $Q_{(t-2)}$, ...) may help the model, as an autoregressive model (see Equations (5–14)), to prompt a good performance even by using the rainfall data from only a few stations. However, with regard to the IGGP model where the $Q_{(t-i)}$

values are not imposed onto the input unit of the model, the interaction and relationship between the discharge values of different stations, imposed onto the output unit (see Figure 7), are learnt via the model. Consequently, as stated in the cross-validation step in the paper, the IGGP can be applied to simulate the discharge values of a point within the watershed using the learnt relationship between the discharge values and geomorphic parameters of sub-basins and, of course, the rainfall data from a few gauges.

It is obvious when the IGGP model is trained using more stations, the model ability will increase in the validation and cross-validation steps. In addition, although the spatial pattern and relationship between the stations could be detected through the IGGP model, the level of this relationship may be different for different stations.

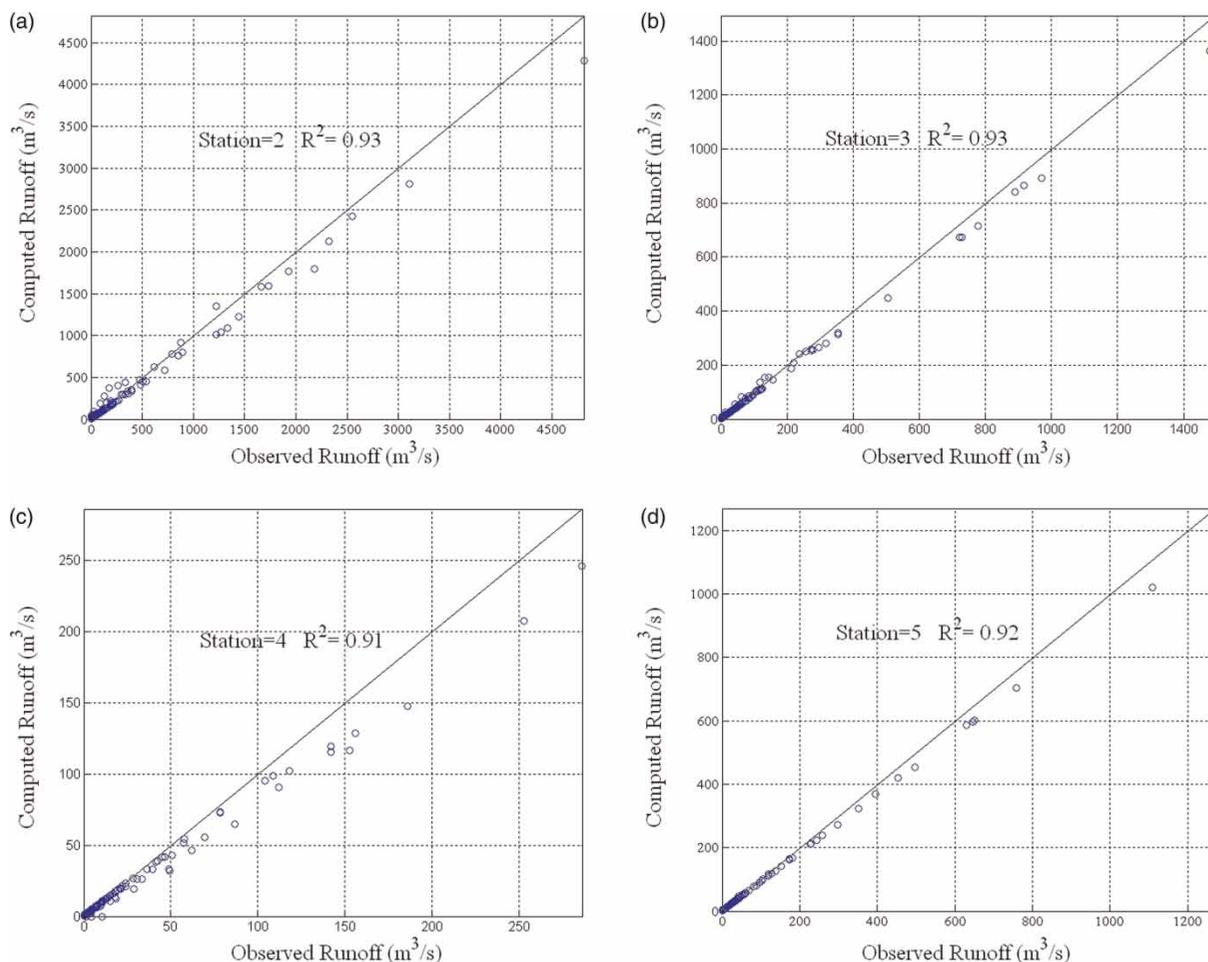


Figure 12 | Scatter plots of computed versus observed runoff time series for stations: (a) station 1, (b) station 2, (c) station 3, (d) station 4, (e) station 5, (f) station 6, (g) station 7, (h) station 8.

SUMMARY AND CONCLUSIONS

During the last decade, hydrologists have focused on data-driven modeling techniques. Many of these approaches are derived from computational intelligence and machine learning, such as neural networks, fuzzy logic, and GP as well as hybrid combinations of different approaches. It is certainly true that artificial intelligence was not originally conceived to deepen the physical analysis of hydrological processes, but rather to take advantage of advances in information and communication technology as applied to hydrological issues. However, there is an increasing trend towards opening up the black box and trying to understand how these models work and, more importantly, how they can be related to watershed geomorphological parameters and process knowledge. Among artificial intelligence techniques, the

GP model has been successfully applied to modeling a wide range of hydrologic processes including rainfall-runoff. Regarding the need to predict the quantitative amount of runoff, GP can be trained to learn the formulation requiring no *a priori* knowledge of catchment characteristics.

In the current research, three different GP-based approaches were presented as classic GP, SGGP, and IGGP models for rainfall-runoff modeling. The classic GP model was constructed based on only temporal rainfall and runoff data whereas in the second and third models, in addition to the temporal variables, the geomorphological properties of the watershed were also considered as inputs. In this way, first, the GP parameters were calibrated through sensitivity analysis for the entire watershed. For instance, it was concluded that a GP with arithmetical functions which lead to an explicit and simple equation is more suitable for

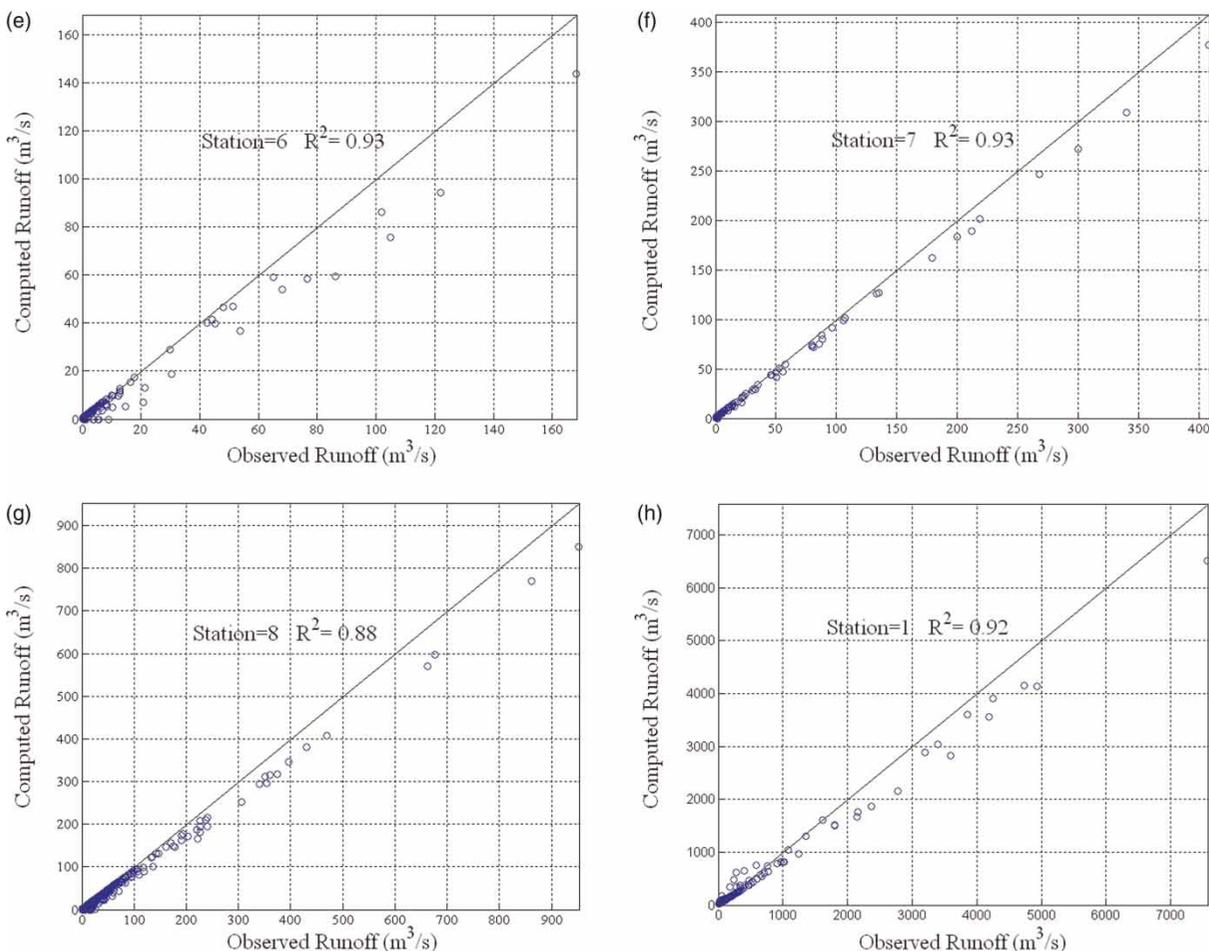


Figure 12 | (Continued)

the rainfall-runoff modeling of the Eel River watershed. Then, two scenarios (i.e., SGGP and IGGP) were presented to investigate the effect of geomorphological features of the sub-basins in the modeling process. The obtained results of SGGP models for the sub-basins showed that the geomorphological parameters could be substituted instead of some physically meaningless real random terminals in the structure of formulations. Also, the SGGP model could distinguish between the dominant variables of the sub-basins in the process.

The results obtained from the IGGP model indicated that the proposed IGGP model can be a reliable tool for spatial and temporal interpolation of runoff through the watershed. Also, in order to affirm this claim, the cross-validation lemma was employed for an internal station (i.e., station 8). The result of cross-validation showed that the IGGP model could estimate the runoff value in the selected period. Hence, the superiority of IGGP model is related to its significant ability in filling the station's data gaps.

Although the duration of the available time series was short, the application of spatial geomorphologic data in the proposed model could compensate for the lack of temporal data.

In conclusion, it is very important to emphasize that the geomorphological-temporal GP model provides the most comprehensive model. Hence, it can be concluded that the proposed temporal-geomorphological data-driven approach has clear advantages over other existing data-driven approaches in rainfall-runoff modeling to produce the explicit formulation based on temporal and spatial geomorphological characteristics of the watershed. Although the GP formulations developed are site-specific just like any other data-driven method, the presented methodology may be easily adopted by the data of other watersheds in order to take advantage of watershed geomorphological information via the modeling. Finally, as a future research plan, it is suggested that the ANFIS is linked to the proposed geomorphological approach in order to handle the uncertainty effects in the modeling (Mirbagheri et al. 2010; Nourani et al. 2011b).

REFERENCES

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000 [Artificial neural networks in](#)

- hydrology. II: Hydrologic applications. *J. Hydrol. Eng.* **5** (2), 124–137.
- Aytek, A. & Kisi, O. 2008 [A genetic programming approach to suspended sediment modeling](#). *J. Hydrol.* **351**, 288–298.
- Azamathulla, H. Md. 2012 [Gene-expression programming to predict scour at a bridge abutment](#). *J. Hydroinf.* **14** (2), 324–333.
- Babovic, V. & Keijzer, M. 2002 Rainfall-runoff modelling based on genetic programming. *Nordic Hydrol.* **33** (5), 331–334.
- Babovic, V., Keijzer, M., Aguilera, D. R. & Harrington, J. 2001 [An evolutionary approach to knowledge induction: Genetic programming in hydraulic engineering](#). World Water and Environmental Resources Congress, Orlando, FL, USA.
- Beven, K. J. & Kirkby, M. J. 1979 [A physically based, variable contributing area model of basin hydrology](#). *Hydrol. Sci. J.* **24** (1,3), 43–69.
- Cousin, N. & Savic, D. A. 1997 [A rainfall-runoff model using genetic programming](#). Centre for Systems and Control Engineering. Report No. 97/03, School of Engineering, University of Exeter, Exeter, UK, 70 pp.
- Demirel, M. C., Booij, M. J. & Kahya, E. 2011 [Validation of an ANN flow prediction model using a multi-station cluster](#). *J. Hydrol. Eng.* **16** (2), 262–271.
- Dorado, J., Rabunal, J. R., Pazos, A., Rivero, D., Santos, A. & Puertas, J. 2003 [Prediction and modelling of the rainfall-runoff transformation of a typical urban basin using ANN and GP](#). *Appl. Artif. Intell.* **17** (4), 329–343.
- Drecourt, J. P. 1999 [Application of neural networks and genetic programming to rainfall-runoff modeling](#). D2K Technical Report 0699-1-1, Danish Hydraulic Institute, Denmark.
- Falco, D., Cioppa, A. D. & Tarantino, E. 2005 [A genetic programming system for time series prediction and its application to El Niño forecast](#). *Appl. Soft Comput.* **32**, 151–162.
- Garbrecht, J. D. 2006 [Comparison of three alternative ANN designs for monthly rainfall-runoff simulation](#). *J. Hydrol. Eng.* **11** (5), 502–505.
- Giustolisi, O. 2004 [Using genetic programming to determine Chezy resistance coefficient in corrugated channels](#). *J. Hydroinf.* **6**, 157–173.
- Guven, A. & Kisi, O. 2011 [Estimation of suspended sediment yield in natural rivers using machine-coded linear genetic programming](#). *Water Resour. Manage.* **25** (2), 691–704.
- Jain, A., Sudheer, K. P. & Srinivasulu, S. 2004 [Identification of physical processes inherent in artificial neural network rainfall-runoff models](#). *Hydrol. Process.* **18**, 571–581.
- Jayawardena, A. W., Muttill, N. & Lee, J. H. W. 2006 [Comparative analysis of data-driven and GIS-based conceptual rainfall-runoff model](#). *J. Hydrol. Eng.* **11** (1), 1–89.
- Kisi, O. & Guven, A. 2010 [A machine code-based genetic programming for suspended sediment concentration estimation](#). *Adv. Eng. Softw.* **41** (7–8), 939–945.
- Koza, J. R. 1992 [Genetic Programming on the Programming of Computers by Means of Natural Selection](#). MIT Press, Cambridge, MA, USA.

- Legates, D. R. & McCabe Jr, G. J. 1999 Evaluating the use of goodness-of-fit measures in hydrologic and hydro-climatic model validation. *Water Resour. Res.* **35** (1), 233–241.
- Luk, K. C., Ball, J. E. & Sharma, A. 2001 An application of artificial neural networks for rainfall forecasting. *Math. Comput. Modell.* **33**, 683–693.
- Makkeasorn, A., Chang, N. B. & Zhou, X. 2008 Short-term stream flow forecasting with global climate change implications – a comparative study between genetic programming and neural network models. *J. Hydrol.* **352**, 336–354.
- Mirbagheri, S. A., Nourani, V., Rajaei, T. & Alikhani, A. 2010 Neuro-fuzzy models employing wavelet analysis for suspended sediment concentration prediction in rivers. *Hydrol. Sci. J.* **55** (7), 1175–1189.
- Mutlu, E., Chaubey, I., Hexmoor, H. & Bajwa, S. G. 2008 Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed. *Hydrol. Process.* **22**, 5097–5106.
- Muttill, N. & Liang, S. Y. 2004 Physically interpretable rainfall-runoff models using genetic programming. *6th International Conference on Hydroinformatics*, Singapore, pp. 1655–1662.
- Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I. A discussion of principles. *J. Hydrol.* **10** (3), 282–290.
- Nourani, V. 2009 Reply to comment on ‘An ANN-based model for spatiotemporal groundwater level forecasting.’ *Hydrol. Process.* **24** (3), 370–371.
- Nourani, V. & Kalantari, O. 2010 Integrated artificial neural network for spatiotemporal modeling of rainfall-runoff-sediment processes. *Environ. Eng. Sci.* **27** (5), 411–422.
- Nourani, V., Monadjemi, P. & Singh, V. P. 2007 Liquid analog model for laboratory simulation of rainfall-runoff process. *J. Hydrol. Eng.* **12** (3), 246–255.
- Nourani, V., Mogaddam, A. A. & Nadiri, A. O. 2008 An ANN based model for spatiotemporal groundwater level forecasting. *Hydrol. Process.* **22**, 5054–5066.
- Nourani, V., Komasi, M. & Mano, A. 2009 A multivariate ANN-wavelet approach for rainfall-runoff modeling. *Water Resour. Manage.* **23**, 2877–2894.
- Nourani, V., Ejlali, R. G. & Alami, M. T. 2011a Spatiotemporal groundwater level forecasting in coastal aquifers by hybrid artificial neural network-geostatistics model: a case study. *Environ. Eng. Sci.* **28** (3), 217–228.
- Nourani, V., Kisi, Ö. & Komasi, M. 2011b Two hybrid artificial intelligence approaches for modeling rainfall-runoff process. *J. Hydrol.* **402** (1–2), 41–59.
- Nourani, V., Komasi, M. & Alami, M. T. 2012a A hybrid wavelet-genetic programming approach to optimize ANN modelling of rainfall-runoff process. *J. Hydrol. Eng.* **17**(6), 724–741.
- Nourani, V., Kalantari, O. & Hosseini Baghanam, A. 2012b Two semi-distributed ANN-based models for estimation of suspended sediment load. *J. Hydrol. Eng.* Published online.
- Rajurkar, M. P., Kothiyari, U. C. & Chaube, U. C. 2004 Modeling of the daily rainfall-runoff relationship with artificial neural network. *J. Hydrol.* **285**, 96–113.
- Rodríguez-Vázquez, K., Arganis-Juárez, M. L., Cruickshank-Villanueva, C. & Domínguez-Mora, R. 2012 Rainfall-runoff modelling using genetic programming. *J. Hydroinf.* **4** (1), 108–121.
- Sarangi, A., Madramootoo, C. A., Enright, P., Prasher, S. O. & Patel, R. M. 2005 Performance evaluation of ANN and geomorphology-based models for runoff and sediment yield prediction for a Canadian watershed. *Curr. Sci.* **89** (12), 2022–2033.
- Savic, D. A., Walters, G. A. & Davidson, J. 1999 A genetic programming approach to rainfall-runoff modeling. *Water Resour. Manage.* **13**, 219–231.
- Selle, B. & Muttill, N. 2011 Testing the structure of a hydrological model using genetic programming. *J. Hydrol.* **397** (1–2), 1–9.
- Sharifi, S., Sterling, M. & Knight, D. W. 2011 Prediction of end-depth ratio in open channels using genetic programming. *J. Hydroinf.* **13** (1), 36–48.
- Sivanandam, S. N. & Deepa, S. N. 2008 *Introduction to Genetic Algorithms*. Springer, Berlin Heidelberg, New York.
- Sivapragasam, C., Vincent, P. & Vasudevan, G. 2007 Genetic programming model for forecast of short and noisy data. *Hydrol. Process.* **21**, 266–272.
- Sreekanth, J. & Datta, B. 2011 Comparative evaluation of genetic programming and neural network as potential surrogate models for coastal aquifer management. *Water Resour. Manage.* **25** (13), 3201–3218.
- Sudheer, K. P., Gosain, A. K. & Ramasastri, K. S. 2000 A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrol. Process.* **16** (6), 1325–1330.
- Tokar, A. S. & Johnson, P. A. 1999 Rainfall-runoff modeling using artificial neural networks. *J. Hydrol. Eng.* **4**, 232–239.
- Tokar, A. S. & Markus, M. 2000 Precipitation-runoff modeling using artificial neural networks and conceptual models. *J. Hydrol. Eng.* **5** (2), 156–161.
- Whigham, P. A. & Crapper, P. F. 2001 Modeling rainfall-runoff using genetic programming. *Math. Comput. Modell.* **33**, 707–721.
- Wilby, R. L., Abraham, R. J. & Dawson, C. W. 2003 Detection of conceptual model rainfall-runoff processes inside an artificial neural network. *Hydrol. Sci. J.* **48** (2), 163–181.
- Wang, W., Xu, D., Qiu, L. & Ma, J. 2009 Genetic programming for modelling long-term hydrological time series. *Fifth International Conference, Natural Computation, Tianjin, China*, 4, pp. 265–269.
- Xu, Q., Chen, Q. & Li, W. 2011 Application of genetic programming to modeling pipe failures in water distribution systems. *J. Hydroinf.* **13** (3), 419–428.
- Zhang, B. & Govindaraju, R. S. 2003 Geomorphology-based artificial neural networks (GANNs) for estimation of direct runoff over watersheds. *J. Hydrol.* **273**, 18–34.

First received 24 January 2012; accepted in revised form 3 August 2012. Available online 10 October 2012