Journal of Advances in Computer Research
Quarterly pISSN: 2345-606x eISSN: 2345-6078
Sari Branch, Islamic Azad University, Sari, I.R.Iran
(Vol. 9, No. 1, February 2018), Pages: 71-84
www.jacr.iausari.ac.ir

Modeling Ghotour-Chai River's Rainfall-Runoff Process by Genetic Programming

Mina Rouhnavaz, Abdolreza Hatamlou✉
Department of Computer Science, Khoy Branch, Islamic Azad University, Khoy, Iran
minarouhnavaz@yahoo.com; rezahatamloo@gmail.com

Received: 2015/02/05; Accepted: 2017/04/12

Abstract

Considering the importance of water and computing the amount of rainfall runoff resulted from precipitation in recent decades, using appropriate methods for predicting the amount of runoff from rainfall date has been really essential. Rainfall-runoff models are used to estimate runoff generated from precipitation in the catchment area. Rainfall-runoff process is totally a non-linear phenomenon. In the present study, it has been tried to Model Ghotour-Chai River rainfall-runoff, one of the studying sub-basins of Aras River with an area of 8544 square kilometers, by genetic programming and to analyze the results. In this study, the statistical data from Ghotour-Chai's daily rainfall-runoff, Marakan hydrometric station during the period 1386-1390 has been used. Data from events during the period 1386-1389 is used for training and data from 1390 for testing. In this modeling, 8 input models have been defined for the system. After applying input models in system, the results based on statistical measures of root-mean-square error and correlation coefficient were analyzed and evaluated. The findings show the success of genetic programming for rainfall-runoff process and this procedure can be suggested as a way for modeling this process.

Keywords: *Rainfall-Runoff, Genetic Programming, Forecasting, Ghotour-Chai River*

1. Introduction

Forecasting river runoff from the extent of rain is the most fundamental purpose of rainfall-runoff modeling [1-4]. The origin of river waters includes raining and melted snows, the water from melted snows can determine river's base flow and surface runoff caused by rate of precipitation. Part of rain can be evaporated or penetrated and the other part of rain changes into surface rainfall and flows in riverbed. As solving most of the complicated issues by means of customary techniques are really complex, difficult and costly, researchers try to solve such issues by optimization methods in a short period of time. Rainfall-runoff is also considered one of the non-linear and complex matters that face a great volume of data. Solving this issue is time-consuming and difficult by common techniques. In this paper, therefore, it is tried to solve it by one of the inspired methods from nature known as genetic programming. Although, it cannot guarantee reaching absolute optimization, but the response close to the optimization can be achieved by this technique. Genetic programming have been studied and analyzed by most of the scholars as one of the characteristics of genetic algorithm [6]. It has concentrated on genotype evolution just like genetic algorithms. The difference between these two models lies in the representation of schemes applied by them. Genetic algorithms make uses of strings or vector schemes, while in genetic programming scheme representation of tree is used. In genetic programming, in each generation, every individual's program is implemented to compute its efficiency in program framework, and the obtained results can determine individual's fitness. Most of the issues in artificial intelligence, symbolic and learning processes of machine need to discover a computer program in order to generate desired results for specific inputs. Genetic programming presents a technique for searching the best computer program. In genetic programming, in general, the population of hundreds or thousands is grown. The growth or development of population is done for breeding computer programs according to Darwinian reproduction and survival principle based on the fittest selection during appropriate genetic crossover operation. A computer program that solves a problem (or solves the problem relatively) can be appeared by Darwinean selection operations or genetic operations [5, 6]. In the recent years novel optimization algorithms have been proposed for solving complicated problems [12-21].

The main advantage of Genetic programming is its ability to create two new solutions from the same solution. In genetic programming identical parents can yield different offspring, while in genetic algorithms identical parents would yield identical offspring. Also it generally find nearly global optima in complex spaces.

In this paper we have chosen the Ghotour-Chai River for our experiments. This is first time that the genetic programming is applied for modeling the Rainfall-Runoff process of the Ghotour-Chai River.

2. Genetic Programming

In summary, genetic programming produces computer programs for solving the problem by implementing the following three processes. The pseudo code of the genetic programming is given in the following [6, 7]:

1) *Producing the primary population by random mixing of functions and problem's terminals (computer programs)*
2) *Repeating the implementation of the following processes till end condition:*
   a. *Implementing every program in population and allocating proper amount to it, based on how much it is good in solving the problem.*
   b. *Creating a new population from computer programs by applying the following two primary instructions. These instructions can be applied to computer programs in a population selected randomly based on fitness.*
      I. *Copying the available computer programs to the primary population*
      II. *Creating new computer programs by combining the parts chosen randomly from two available programs.*
3) *The best computer program that appears in every generation (i.e. the best individual) is determined as the result of genetic programming. The result can be a solution (or relative solution) for the problem.*

The set of possible structures in genetic programming is the set of all possible compositions of functions that can be composed recursively from the set of Nfunc functions from $F=\{f_1,f_2,\ldots,f_{Nfunc}\}$ and the set of Nterm terminal from $T=\{a_1,a_2,\ldots,a_{Nterm}\}$. Each particular function of f_i in the function set F takes a specified number z (f_i) of arguments $z(f_1)$, $z(f_2)$ ..., $z(f_{Nfunc})$. That is, function fi has an amount of

$z(f_i)$. $Z(f_i)$ is the number of taken arguments by f function. The functions in the function set may include arithmetic operations, mathematical functions, the Boolean operations, conditional operations, functions causing iteration, recursive functions, and any other domain-specific functions that may be defined. Terminals also include figures, variables and / or atoms (such as number 3 or fixed Boolean NIL). Sometimes terminals also include functions without arguments. The set of terminals and functions should be defined in a way to be able to give a solution for the problem. Genetic programming user should know that some combinations of functions and its terminals can yield a solution for the problem.

The primary structures in genetic programming include individuals in initial population with tree structure. Every generation has individuals with a tree made by creating a root randomly and then adding labeled ties with organized trees. We start by selecting function from F function sets randomly (normal random monotone distribution) to label the tree root. We limit the label selection for tree root to F functions sets. Because we want to create hierarchical structures not a structure on the decline containing only one terminal. Figure 1 show the starting point in the creation of tree program randomly. + function (takes two arguments) is chosen from a set of F function and marked as a tree root label. When a tie tree labeled with a f function from a set of F functions, then z(f) lines can be created from that point, z(f) is the number of arguments taken by f function. After that, for every one of these lines an element of combined set $C=F\cup T$ is chosen randomly from sets of functions and terminals for the end point of that line. If a function for labeling the end point line is chosen, production process continues recursively. For example, in Figure 2, X function of combined set $C=F\cup T$ was chosen from sets of functions and terminals for labeling non root point(point 2) at the end of line one(the left one) created from point 1, with + function (point 1). As one function is created for point 2, it became a linear point, non-root point of the tree that finally has been created. X function takes two arguments, as it has been shown in Figure 2, two lines exit from point 2. If one terminal was chosen for labeling every point, that point became the final point of tree and production process finalized for that point. For example, in Figure 3, a terminal of T terminals sets was selected for labeling the first line of x tie. Similarly, terminals B and C have been selected for labeling two other lines. This process continues recursively from left to right till the tree to be created and labeled completely, as we can see in Figure 3.

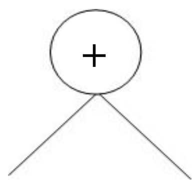

Figure 1. Starting point in the creation of tree program randomly, + function with two selected arguments for tree root

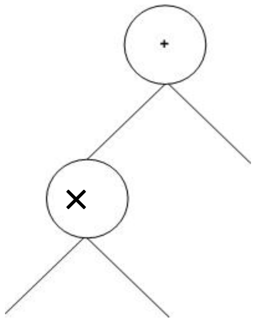

Figure 2. The rest of starting point in the creation of tree program randomly, X function with two arguments selected for point 2

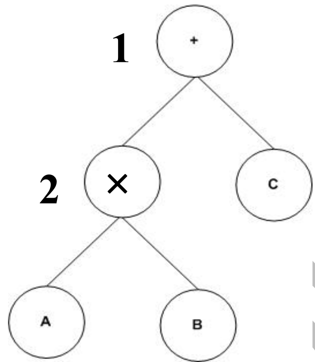

Figure 3. Completing tree program creation randomly, with selected terminals A,B,C

The production process can be done by different techniques that result in creating trees randomly with different sizes and shapes.

One of the difficulties we face in genetic programming is the programs' overgrowth that is called Bloating. To control the depth of tree different ways including determining the maximum depth for trees, penalizing bigger trees and applying Full and Grow ways for making the tree can be used.

3. Rainfall-Runoff

Rainfall is the most important factor involved directly in hydrologic cycle [8-11]. It is atmospheric precipitations to the ground surface, gathering moisture falling and entering the ground. When the intensity of rainfall is more than infiltration rate of soil, part of water remains in the basin surface. This amount of water fills the cavities on the ground and flows along the ground slope then can be escaped by watercourse channel and main rivers from the basin. This part of rainfall is known as surface runoff. Runoff and the relationship between rainfall-runoff are the most important and basic surface water hydrology subjects.

4. Describing Study Scope

Aras basin has been studied diversely because of its extensive area and containing noticeable water. Identifying river system, investigating surface water resources in Aras basin, determining the extent of some of the most important hydrologic parameters in the study scope surface of this basin, and dividing plain areas and heights of these study scopes involved directly or indirectly in balancing basin’s surface and ground water resources are the most important factors for doing hydrologic studies in this scope.

Tributaries of Aras river originated from Turkey and Armenia soil and crosses from the extreme northwestern Iran in Iran-Azerbaijan common border and after connecting to Kura river, flows into Caspian river. It has an area of about 39896 square kilometer inside Iran. Aras in the southern part lies between longitudes 44-02 to 48-71 eastern and latitudes 37-78 to 39-79 southern. Aras river basin spreads from Zangmar river in north western to Darreh river in east. The present paper focuses on Ghotour-Chai River from Aras river basin. Its area is 8544 square kilometer. Considering area, it is the second basin in Aras river after Darreh river. The main origin of this river is Bazinga heights in Turkey that continues from west to east in Iran after joining a branch which originates from a mountain in the route of Omer and Panarik. Nearby Khoy, this river joins to Aland Chai River which originates from mountains Nazarbig and Foch in the heights 2600 to 2700 meters known as Piragloo river and it is directed from western north to eastern south.

5. Methods and Materials

Genetic programming technique is able to make a relationship between input and output [8-11]. The input model should be characterized in this programming. In this problem, Ghotour-Chai River’s daily rainfall and runoff data during 1386-1390 is available. We assume that tomorrow’s runoff depends on the last days’ runoff. In this problem, it has been tried to estimate the extent of tomorrow runoff from last days’ runoff. It should be explained that rainfall and runoff data from the years during the priod1386-1389 used for learning, and data from 1390 for testing. Data from rainfall and runoff in 1385 also have been applied for simulation process. In other words, the extent of rainfall and runoff from the last days in 1385 have been taught to the model, entered and the extent of tomorrow runoff have been predicted and finally the extent of predicted runoff compared with the extent of real runoff.

The input model should be characterized in this programming. To characterize the best input model, different input models should be defined for the system and every one of these input models applied for the system and finally among these input models, the model with the best efficiency selected as best input model. There is a relationship between input and output variables in genetic programming. In modeling Ghotour-Chai River’s rainfall-runoff process, the below input models have been used in daily intervals that these input models represent the relationship between problem’s input and output variables.

$$Q(t) = \{R(t)\}$$

$$Q(t) = \{Q(t-1), R(t-1), R(t)\}$$

$$Q(t) = \{Q(t-1), Q(t-2), R(t-1), R(t-2), R(t)\}$$

$$Q(t) = \{Q(t-1), Q(t-2), Q(t-3), R(t-1), R(t-2), R(t-3), R(t)\}$$

$$Q(t) = \{Q(t-1), Q(t-2), Q(t-3), Q(t-4), R(t-1), R(t-2), R(t-3), R(t-4), R(t)\}$$

$$Q(t) = \{Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5), R(t-1), R(t-2), R(t-3), R(t-4), R(t-5), R(t)\}$$

$$Q(t) = \{Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5), Q(t-6), R(t-1), R(t-2), R(t-3), R(t-4), R(t-5), R(t-6), R(t)\}$$

$$Q(t) = \{Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5), Q(t-6), Q(t-7), R(t-1), R(t-2), R(t-3), R(t-4), R(t-5), R(t-6), R(t-7), R(t)\}$$

In the present input models runoff Q is according to m^3s^{-1} and R is the intensity of rainfall according to $mmday^{-1}$. Entering raw data into the system decreases its speed and accuracy. To avoid this and also to unify data value for the system, normalizing the data can be done. In this research, normalizing the data is done in the intervals (0.1, 0.9). Equation (1) is used to normalize input data [6].

$$X_{normal} = 0.1 + 0.8 \times \left(\frac{X_0 - X_{min}}{X_{max} - X_{min}}\right) \qquad (1)$$

In this equation, X_{normal} is normalized data, X_0 is the real amount of data, X_{max}, X_{min} represent maximum and minimum amounts of the data respectively. It is necessary to explain that the sets in the process of rainfall-runoff assumed to be in the form of {+,-,×,/} and the sets of terminals are assumed to be rainfall data, runoff data and sets of real numbers. Ghotour-Chai River's rainfall data and daily flow from 1386 to 1390 used for the purpose of the present study. Statistical indexes of correlation coefficient and root mean square error used to evaluate the model for the process of Ghotour-Chai River's rainfall-runoff. In fact, the efficiency of programs is determined by these evaluations. You can use equation (2) and (3) to find Correlation coefficient and root mean square error [6]:

$$R = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}} \qquad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - x_i)^2}{N}} \qquad (3)$$

In these equations R is correlation coefficient, RMSE is root mean square error, x_i is an observed value in the i-th time step, y_i is the calculated value in the same time, N is the number of time steps, $\overline{x}$ is the mean of obseverd means and y is also the mean of calculation values. Figure 4 shows Ghotour-Chai River's rainfall-runoff data.

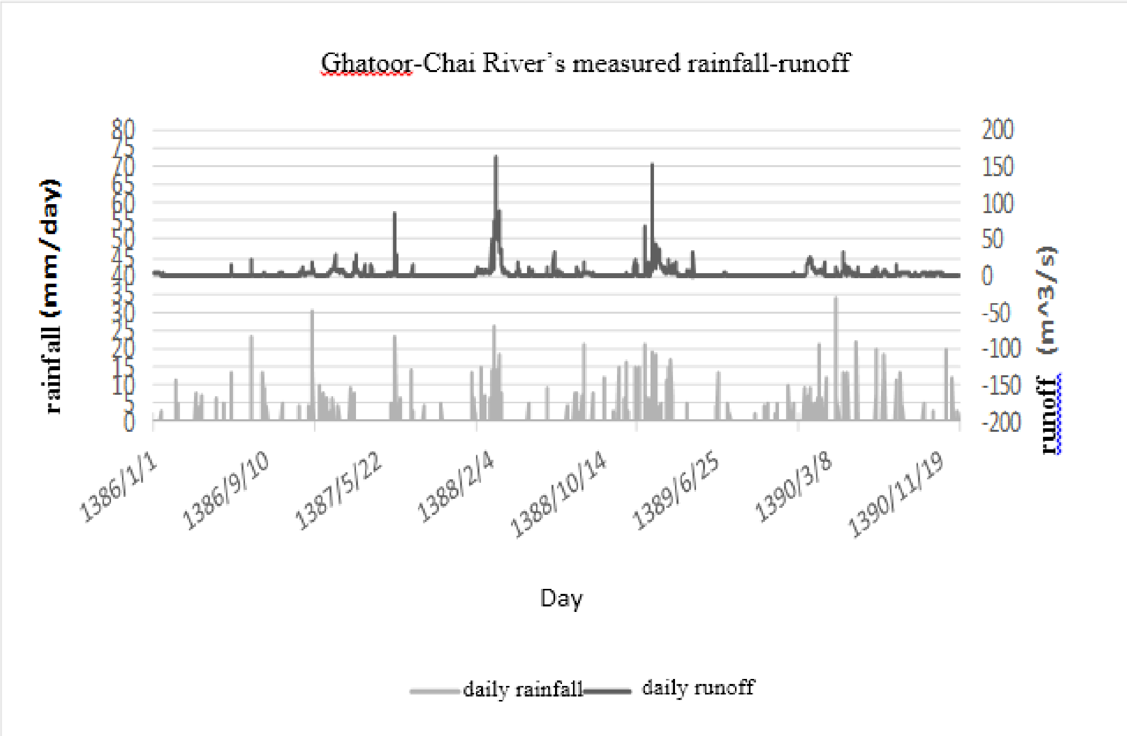

Figure 4. Ghotour-Chai River's rainfall-runoff data.

Statistical characteristics of rainfall-runoff data used in the present study in daily interval from 1386 to 1390is shown in Table 1.

Table 1: Statistical characteristics of rainfall-runoff data used in the present study in the daily interval from 1386 to 1390

Statistical charateristics	Daily rainfall (mmday)	Daily flow(m^3s^{-1})
Number of data	1826	1826
Mean	0.696360778	3.33182092
Variance	8.528677665	92.71964263
Maximum	30	164
Minimum	0	0.001

In modeling Ghotour-Chai River's rainfall-runoff process, normalized entered data according to equation (1) used for 8 input pattern model. The results obtained from the present models have been compared in Table 2 according to statistical fitting techniques.

Table 2. Results of input models

Input models	Number of implementation	Training				test				Program size	Number of variables
		Correlation coefficient	Root mean square error	Explanatory coefficient	Best fitness	Correlation coefficient	Root mean square error	Explanatory coefficient	Best fitness		
1	100	0.324	0.048	0.105	953.59	0.275	0.124	7.604	889.42	17	1
2	100	0.783	0.031	0.614	969.02	0.813	0.064	0.662	939.56	23	3
3	100	0.785	0.031	0.616	969.08	0.801	0.064	0.642	939.10	25	4
4	100	0.805	0.03.	0.648	969.97	0.817	0.063	0.668	940.57	21	5
5	100	0.821	0.029	0.674	971.33	0.822	0.062	0.676	941.51	25	7
6	100	0.819	0.029	0.671	971.06	0.804	0.064	0.647	939.32	21	7
7	100	0.811	0.030	0.658	970.55	0.786	0.067	0.618	936.82	33	7
8	100	0.801	0.030	0.643	970.13	0.444	0.132	0.197	883.33127	7	

6. Experimental Results

We conclude that the results improve little by little to input 5 and after that the results become weaker. From 8 input models, the fifth input model was introduced the best input model in Ghotour-Chai River's rainfall-runoff. It has the least root mean square error and the most correlation coefficient for training and test data. Rainfall-runoff relationship from genetic programming in modeling Ghotour-Chai River's rainfall-runoff process is as follow:

$$Q(t) = (d(0) * (d(7) * (((d(0) - d(4))
* (G1C1/d(1))) * G1C0)))
+ d(0) + ((d(8) * ((d(6) * d(5))
* d(8))) * (d(5)/d(7)))$$

The fixed values of equation made by GeneXpro Tools software are as follows:

$$G1C0 = -1.786133$$
$$G1C1 = 0.522095$$
$$G2C0 = -1.786133$$
$$G2C1 = 0.522095$$
$$G3C0 = -1.786133$$
$$G3C1 = 7.263397$$

$$d(0) = Q(t-1)$$
$$d(1) = Q(t-2)$$
$$d(2) = Q(t-3)$$
$$d(3) = Q(t-4)$$
$$d(4) = R(t)$$
$$d(5) = R(t-1)$$
$$d(6) = R(t-2)$$
$$d(7) = R(t-3)$$
$$d(8) = R(t-4)$$

By replacing the fixed values, Rainfall-runoff can be shown as follow:

$$Q(t) = (Q(t-1)\times(R(t-3)\times(((Q(t-1)-R(t))\times(\frac{0.522095}{Q(t-2)}))\times(-1.786133))))$$
$$+Q(t-1)+((R(t-4)\times((R(t-2)\times R(t-1))\times R(t-4)))\times(\frac{R(t-1)}{R(t-3)})$$

As can be seen, according to this equation, $Q(t)$ (runoff) is a function of 7 variables: *Q(t-1), Q(t-2), R(t), R(t-1), R(t-2), R(t-3), R(t-4)* at t time. As it was pointed out before the first step in genetic programming is the random production of the initial population, then the initial population are shown in a form of tree structure, this structure helps to show the initial population in every stage as tree structure and all changes can be done on the simple structures. This makes no need to the relatively complex structures for establishing in every stage. The tree structure of rainfall-runoff equation yielded from genetic programming in modeling Ghotour-Chai River's rainfall-runoff can be obtained from Figure 5.

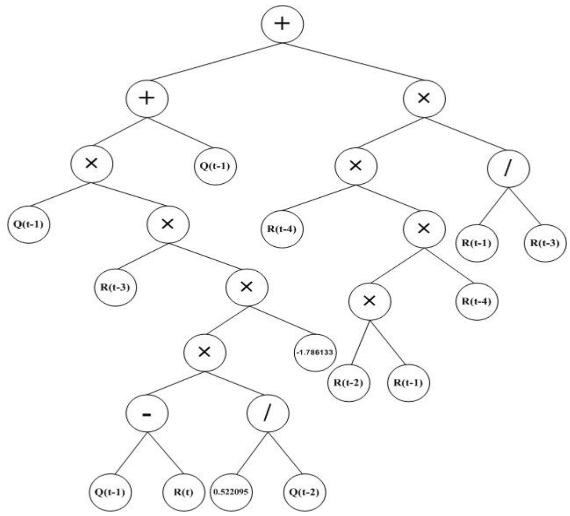

Figure 5. the tree structure of rainfall-runoff equation yielded from genetic programming in modeling Ghotour-Chai River's rainfall-runoff

To compare input models and proving that the results were improving up to 5 input model and after that the results became weaker, the following charts are drown.

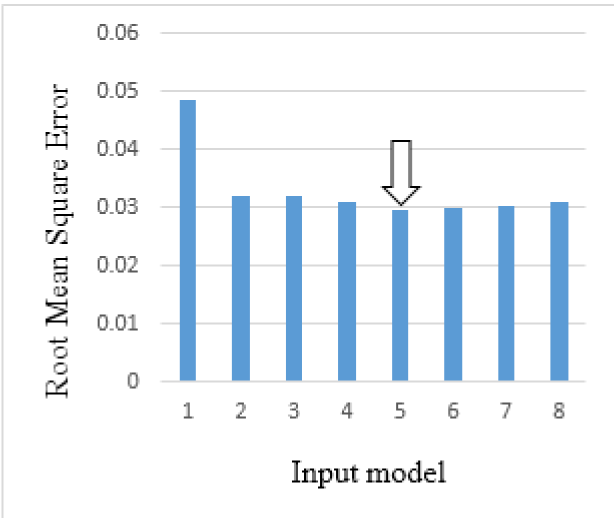

Figure 6. Root Mean Square Error in Training

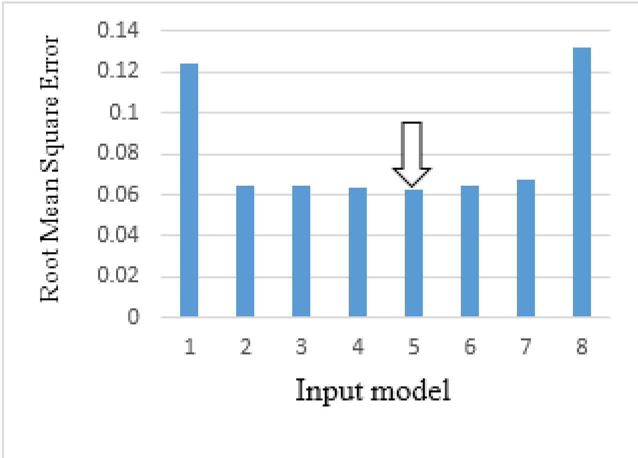

Figure 7. Root Mean Square Error in test

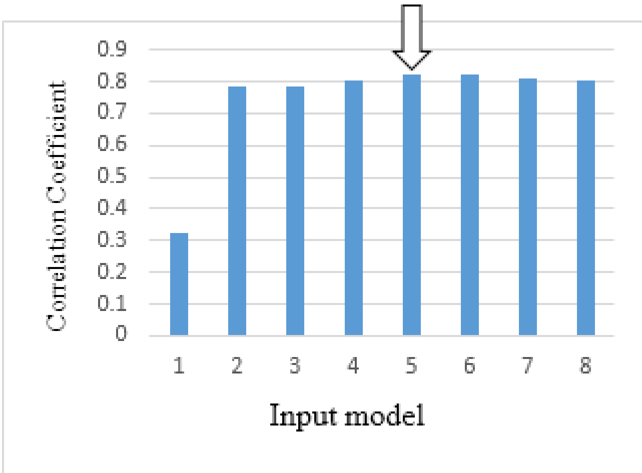

Figure 8. Correlation Coefficient in Training

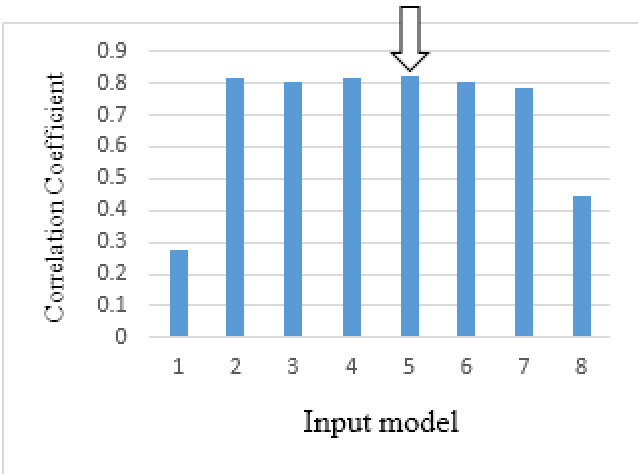

Figure 9. Correlation Coefficient in Test

Regarding the least error, in the present charts, the figure shown by arrow is known as the best model among 8 input models. As it is clear from the charts, the results are going down after 5 input model. Figures 10 and 11 show observable figures' values against computational values by genetic programming for the best input model (5).

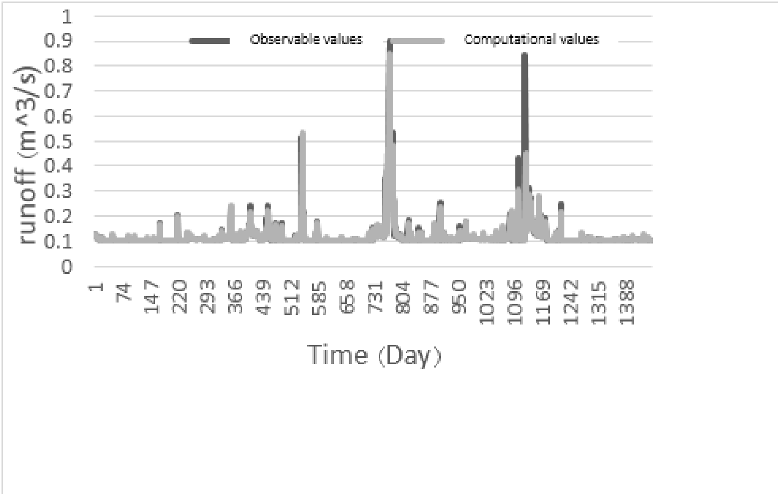

Figure 10. Observable-Computational values figure compared with time in the best learning genetic programming input model

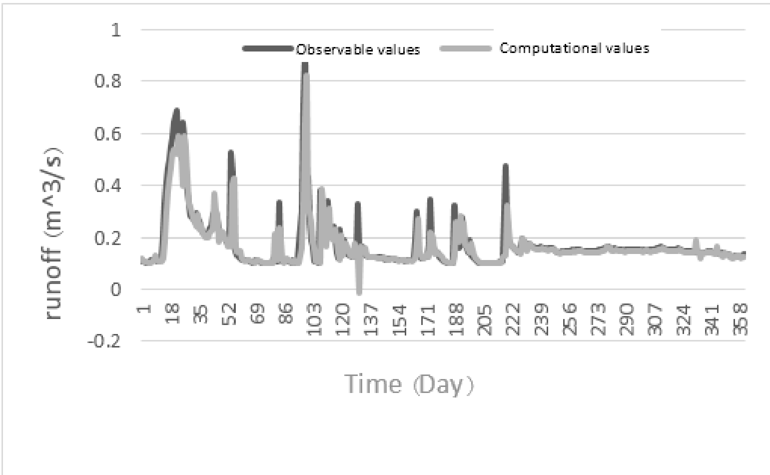

Figure 11. Observable-Computational values figure compared with time in the best testing genetic programming input model

Here, we select the best input model (model 5) as ideal model and demodulate simulation process by this model. In this phase new defined data to the program are about Ghotour-Chai River's daily rainfall-runoff related to Marakand hydrometric station in 1385. During simulation normalized data were used. In fact, the 5 last days rainfall data and the last four days runoff data in 1385 were taught, entered into the system as an input and tomorrow's runoff has been forecasted and finally the predicted runoff has been compared with real runoff. Figure 8 shows the values of predicted runoff and real runoff in comparison with the time in the simulation process of data related to 1385.

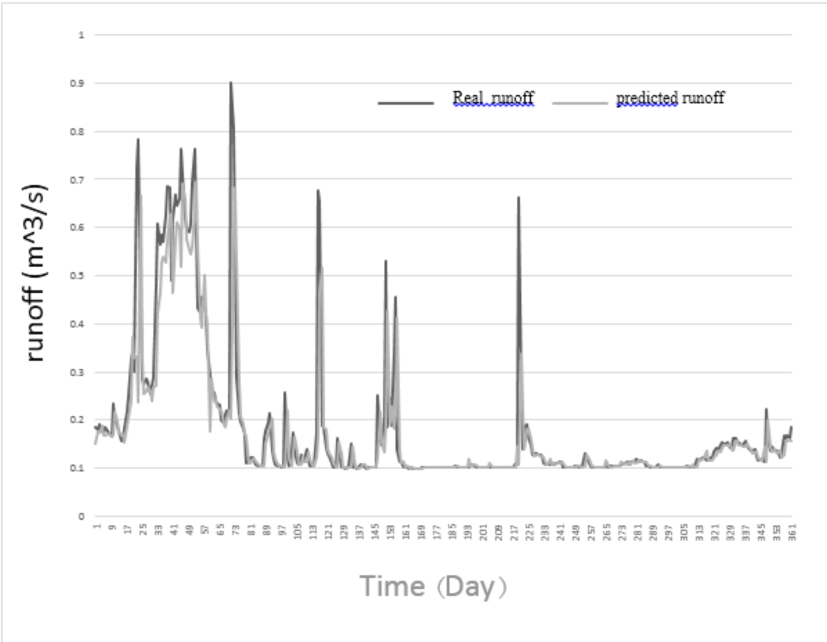

Figure 12. The values of predicted runoff-real runoff in comparison with simulation obtained time

7. Conclusion

In this research we have applied the genetic programming for modeling and predicting the rainfall-runoff process in the Ghotour-Chai River. The results show the success of genetic programming method for modeling rainfall-runoff process and it is suggested as way for modeling this process. By modeling Ghotour-Chai River's rainfall-runoff process in five years interval and obtaining the results, we can make use of the model in predicting the extent of runoff resulted from rainfall during next days, avoiding the injuries of flood, managing region water resources, water frugality, optimal management of water consumption in agricultural and urban drinking water. As the results from genetic programming is close to reality considering modeling rainfall-runoff process and the error extent of this method is low, it is suggested to apply the present model for modeling rainfall-runoff in different places' basin instead of using difficult and complicated techniques.

References

[1] Whigham PA and Crapper PF, Modeling rainfall-runoff using genetic programming, Mathematical and Computer Modeling, 33, 70-721, 2001.

[2] David G. Tarboton, Rainfall-Runoff Processes, A workbook to accompany the Rainfall-Runoff Processes Web module, Utah State University,2003.

[3] Anthony j.Calandro, Rainfall-Runoff Relations for Southeastern Louisiana And Southwestern Mississippi,Louisiana Department Of Public Works Baton Rouge,LA.1967.

[4] Aytek A, Asce M and Alp M, An application of artificial intelligence for rainfallrunoff modeling. J Hydrol Earth Syst Sci 117(2): 145-155, 2008.

[5] Andries P. Engelbrecht, Computational Intelligence AnIntroduction, John Wiley & Sons Ltd, Second Edition,University of Pretoria,South Africa, 2007.

[6] Sette S, Boullart L, Genetic programming: principles and applications. Engineering Applications of Artificial Intelligence 14: 727–736, 2001.

[7] John R. Koza, Genetic Programming on the Programming of Computers by Means of Natural Selection, A Bradford Book The MIT Press Cambridge, Massachusetts London, England, Sixth printing, 1998.

[8] Marco Franchini,Giorgio Galeati,Comparing several genetic algorithn schemes for the calibration of conceptual rainfall-runoff models,Hydrological Sciences-Journal-des Sciences Hydrologiques, 42(3), 1997.

[9] Jay awardena AW, Muttil N and Fernando TMKG, Rainfall-Runoff Modelling using Genetic Programmin. . International Congress on Modelling and Simulation Society of Australia and New Zealand, pp.1841-1847, 2005.

[10] Khu ST, Liong SY, Babovic V, Madsen H and Muttil N, Genetic programming andits application in real- time runoff forming. J AmWater Res Assoc 37(2): 439-451, 2001.

[11] Liong SY, Gautam TR, Khu ST, Babovic V, Keijzer M and Muttil N, Geneticprogramming: A new paradigm in rainfall runoff modeling. J Am Water Res Assoc 38(3): 705-718, 2002.

[12] A. Hatamlou, Heart: a novel optimization algorithm for cluster analysis, Prog.Artif. Intell. 2 (2014) 167–173.

[13] A. Hatamlou, Black hole: a new heuristic optimization approach for data clustering, Inf. Sci. 222 (2013) 175–184.

[14] M Farahmandian, A Hatamlou, Solving optimization problems using black hole algorithm, Journal of Advanced Computer Science & Technology 4 (1), 68-74, 2015

[15] A Bouyer, A Hatamlou, AH Abdullah, An optimized clustering algorithm using genetic algorithm and rough set theory based on kohonen self organizing map, LJS Publisher and IJCSIS Press, 2010

[16] A Hatamlou, M Hatamlou, PSOHS: an efficient two-stage approach for data clustering, Memetic Computing 5 (2), 155-161, 2013

[17] A Hatamlou, M Hatamlou, Hybridization of the gravitational search algorithm and Big Bang-Big Crunch algorithm for data clustering, Fundamenta Informaticae 126 (4), 319-333, 2013

[18] B Javidy, A Hatamlou, S Mirjalili, Ions motion algorithm for solving optimization problems, Applied Soft Computing 32, 72-79, 2015

[19] A Hatamlou, S Abdullah, H Nezamabadi-Pour, A combined approach for clustering based on K-means and gravitational search algorithms, Swarm and Evolutionary Computation 6, 47-52, 2012

[20] A Hatamlou, E Ghaniyarlou, Solving knapsack problems using heart algorithm, International Journal of Artificial Intelligence and Soft Computing 5 (4), DOI: http://dx.doi.org/10.1504/IJAISC.2016.081347, 2016

[21] K Mohrechi, A Hatamlou, Locating Optimal Places for Emergency Medical Centers Using Artificial Bee Colony Algorithm, Journal of Advances in Computer Research 6 (1), 115-124, 2015.