Chapter 3
Application of Genetic Programming in Hydrology

E. Fallah-Mehdipour and O. Bozorg Haddad

3.1 Introduction

In the real world, there are several natural and artificial phenomenons which follow some rules. These rules can model in a mathematical and/or logical form considering simple or complex equation/s may be difficult in some systems. Moreover, it is sometimes necessary to model just some parts of system without considering whole system information. Data-driven models are a programming paradigm that employs a sequence of steps to achieve best connection between data sets.

The contributions from artificial intelligence, data mining, knowledge discovery in databases, computational intelligence, machine learning, intelligent data analysis, soft computing, and pattern recognition are main cores of data-driven models with a large overlap in the disciplines mentioned.

GP is a data-driven tool which applies computational programming to achieve the best relation in a system. This tool can set in the inner or outer of system modeling which makes it more flexible to adapt different system states.

E. Fallah-Mehdipour
Department of Irrigation & Reclamation Engineering, Faculty of Agricultural Engineering & Technology, College of Agriculture & Natural Resources, University of Tehran, Karaj, Tehran, Iran

National Elites Foundation, Tehran, Tehran, Iran
e-mail: Falah@ut.ac.ir

O. Bozorg Haddad (✉)
Department of Irrigation & Reclamation Engineering, Faculty of Agricultural Engineering & Technology, College of Agriculture & Natural Resources, University of Tehran, Karaj, Tehran, Iran
e-mail: OBHaddad@ut.ac.ir

In the water engineering, there are several successful metaheuristic algorithm applications in general (e.g. Yang et al. 2013a, b; Gandomi et al. 2013) and GP in particular. Sivapragasam et al. (2009), Izadifar and Elshorbagy (2010), Guven and Kisi (2011), and Traore and Guven (2012, 2013) applied different GP versions to find best evaporation or evapotranspiration values with minimum difference from real values. Urban water management is other GP application field in which monthly water demand has forecasted by lags of observed water demand. Nasseri et al. (2011) applied GP for achieving an explicit optimum formula. These results can help decision makers of water resources to reduce their risks of online water demand forecasting and optimal operation of urban water systems (Nasseri et al. 2011). Li et al. (2014) extracted operational rules for multi-reservoir system by GP out of mathematical model. They used following steps to find operational rules: (1) determining the optimal operation trajectory of the multi-reservoir system using the dynamic programming to solve a deterministic long-term operation model, (2) selecting the input variables of operating rules using GP based on the optimal operation trajectory, (3) identifying the formulation of operating rules using GP again to fit the optimal operation trajectory, (4) refining the key parameters of operating rules using the parameterization-simulation-optimization method (Li et al. 2014). Results showed the derived operating rules were easier to implement for practical use and more efficient and reliable than the conventional operating rule curves and ANN rules.

Hydrology is a field of water engineering that focuses on the quantity and quality of water on Earth and other planets. In the scientific hydrologic studies, formation, movement and distribution of water are considered in hydrologic cycle, water resources and environmental watershed sustainability. The Earth is often called "blue planet" because of water distribution on its surface that appears blue from space. The total volume of water on Earth is estimated at 1.386 billion km^3 (333 million cubic miles), with 97.5 % and 2.5 % being salt and fresh water, respectively. Of the fresh water, only 0.3 % is in liquid form on the surface (Eakins and Sharman 2010). Due to, the key role of freshwater in life and different limitations of available water on the Earth, appropriate accuracy on hydrology models is necessary. On the other hand, increasing accuracy needs more data and application of expand conceptual methods in the hydrology models. Thus, GP have been applied as a popular, simple and user-friendly tool. This tool can summarize complex methods in a black-box process without modeling all system details. The purpose of this chapter is to assess the state of the art in GP application in hydrology problems.

3.2   Genetic Programming

GP is a data-driven model which borrows a random iterative searching base from evolutionary algorithms and move toward optimal solution (optimal relation) using advantage of these algorithms. Evolutionary algorithm is a subfield of artificial intelligence that involves combinatorial optimization and uses in the different fields
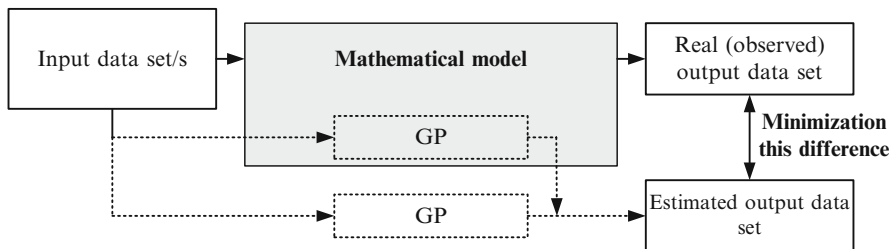
Fig. 3.1 GP presentation in the mathematical models

of water management considering single- and multi-objective. In the recent decades, there is a considerable growth in the development and improvement of evolutionary algorithms and application of hybrid algorithms to increase convergence velocity and find near-optimal solution.

Although, some new developed hybrid algorithms are capable to derive optimal solution, the decision variables have been considered only among the numerical variables. Thus, these algorithms present optimal value and not optimal equations. GP is one of the evolutionary algorithms, in which mathematical operators and functions are added to the numerical values as decision variables.

As shown in Fig. 3.1, GP equation can stand in or out of mathematical model to minimize difference between real (observed) and estimated output data set.

If GP equation presents in mathematical model, it will determine a constraint. In contrast, if GP equation is out of mathematical model, it will play a black-box role which can replace with mathematical model.

In evolutionary algorithms, each decision variable is called a gene, particle, frog and bee in the genetic algorithm (GA), particle swarm optimization (PSO), shuffled frog leaping algorithm (SFLA) and honey bees mating optimization (HBMO) algorithm and a set of aforementioned points with a fixed length is identified as solutions. However, in GP, the solutions have a tree structure which can include different numbers of decision variables and can produce a mathematical expression. Every tree node has an operator function and every terminal node has an operand, necessitating the evaluation of mathematical and logical expressions (Fallah-Mehdipour et al. 2012).

Figure 3.2a, b present two trees in the GP. As it is shown, in a tree structure, all the variables and operators are assumed to be the terminal and function sets, respectively.

Thus, $\{x, y, 47\}$ and $\{x, y\}$ are the terminal sets and $\{\sin, +, /\}$ and $\{\exp, \cos, /\}$ are the function sets of Fig. 3.2a, b, respectively. In the GP structure, the length of the tree creates the formula called depth of tree. The larger number of depth of tree, the more accuracy of the GP relation (Orouji et al. 2014). The GP searching process starts generating a random set of trees in the first iteration as same as other evolutionary algorithms. An error performance which is commonly assumed such as root mean squared error (RMSE) or mean absolute error (MAE) is then calculated. Thus, the error performance corresponds obtained objective function.

$$f(x,y) = \sin(x) + \frac{47}{y} \qquad f(x,y) = \frac{\exp(y)}{\cos(x)}$$
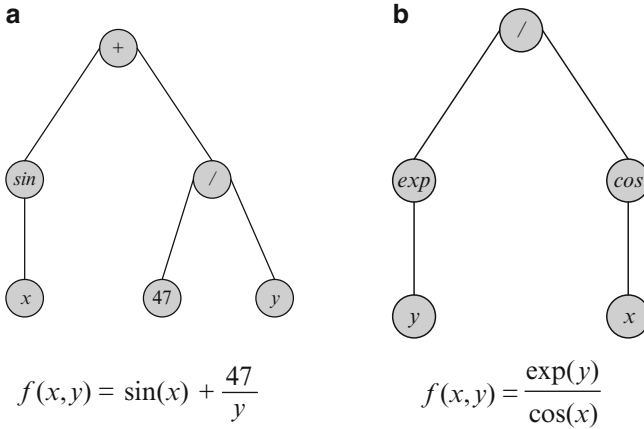
Fig. 3.2 Two GP expressions in the tree structure

To generate the next tree set, trees with the better fitness values are selected using techniques such as roulette wheel, tournament, or ranking methods (Orouji et al. 2014). In following, crossover and mutation as the two genetic operators as same as GA operators create new trees using the selected trees. In the crossover operator, two trees are selected and sub-tree crossover randomly (and independently) selects a crossover point (a node) in each parent tree. Then, two new trees are produced by replacing the sub-tree rooted at the crossover point in a copy of the first parent with a copy of the sub-tree rooted at the crossover point in the second parent, as illustrated in Fig. 3.3 (Fallah-Mehdipour et al. 2012).

In the mutation operator, point mutation is applied on a per node basis. That is, some node/s are randomly selected, it is exchanged by another random terminal or function, as it is presented in Fig. 3.4. The produced trees using genetic operators are the input trees for the next iteration and the GP process continues up to a maximum number of iterations or minimum of error performance.

3.3 GP Application in Hydrology Problems

GP is a data-driven model based on a tree-structured approach presented by Cramer (1985) and Koza (1992, 1994). This method belongs to a branch of evolutionary algorithm, based on the GA, which presents the natural process of struggle for existence. There are two approaches to apply GP in water problems: (1) outer and (2) inner mathematical model. In the first approach, GP extracts system behavior by using some or all characteristics without focus on the system modeling. In contrast, in the second approach, the derived equation by GP uses in system modeling as same as other basic equations. In this section, some applications of aforementioned approaches have been considered.
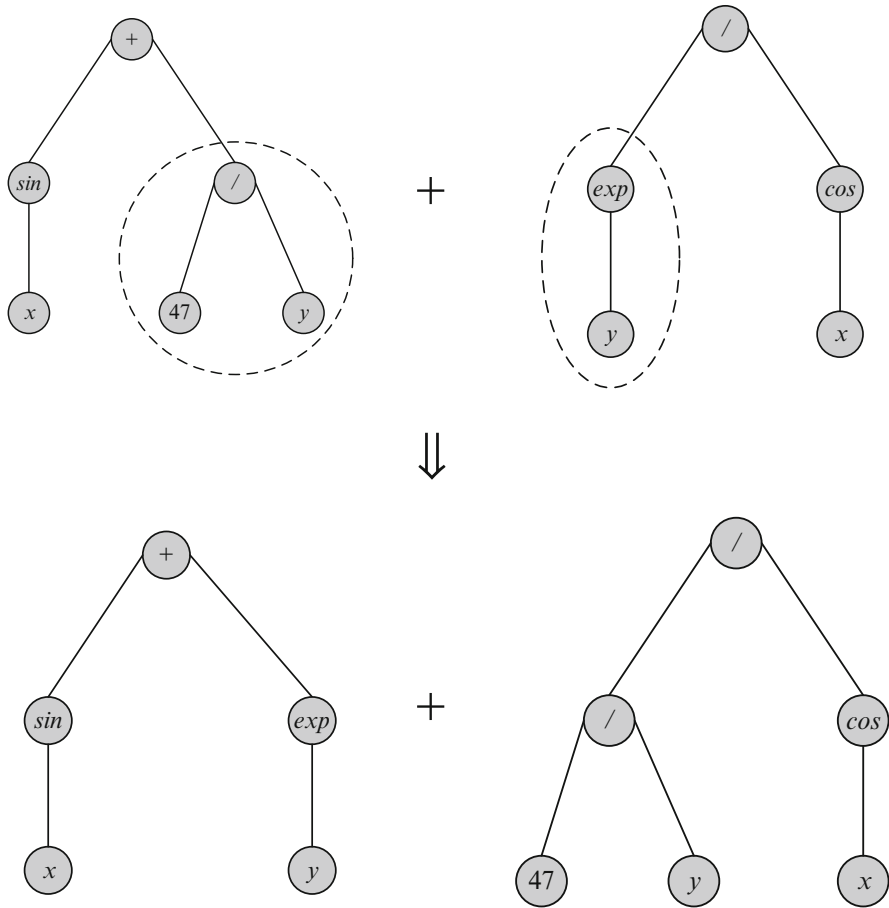
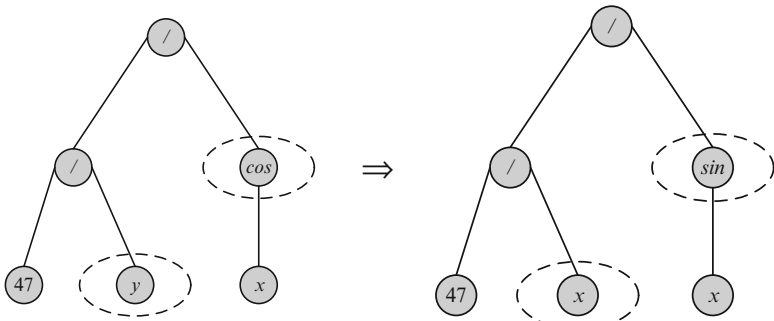**Fig. 3.3** Crossover operator in GP structure

Fig. 3.4 Mutation operation in GP structure

3.3.1 GP Application Outer Mathematical Model

In this section, a common GP application as a modeling tool in the natural and artificial phenomenon is presented. This type of GP applications which is used outer mathematical model to extract the best equation in a system without considering whole details.

In this process, some characteristic/s are selected as the input data and one corresponding data set is used as the real or observed output data set. The main goal is finding the best appropriate equation between these input and output data that yield the minimum difference from observation values. As it is presented in Fig. 3.5, this GP application has a black-box framework in which there is no direct relation with system modeling and equations. In other words, in this type of application, GP can be viewed solely in terms of its input, output and transfer characteristic without any knowledge of its internal working.

3.3.1.1 Rainfall-Runoff Modeling

A watershed is a hydrologic unit in which surface water from rain, melting snow and/or ice converges to a single point at a lower elevation, usually the exit of the basin. Commonly, water that moves to external point and join another water body, such as river, lake or sea. Figure 3.6 presents schematic of a watershed.

When rain falls on watershed, water that called runoff, flows on it. A rainfall-runoff model is a mathematical model describing relations between rainfall and runoff for a watershed. In this case, conceptual models are usually used to obtain both short- and long-term forecasts of runoff. These models are applied several variables such as climate parameters, topography and land use variables to determine runoff volume. Thus, that volume depends directly on the accuracy of each aforementioned variable estimation. On the other hand, some global circulation model (GCM) that is used for runoff calculation apply for large scale and runoff volume for smaller scale should be extracted by extra processes.
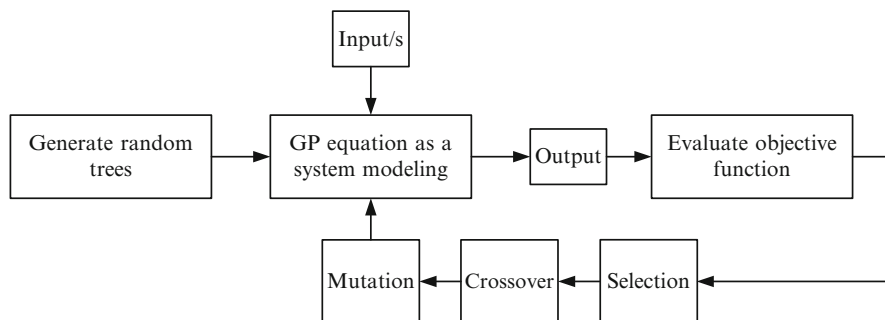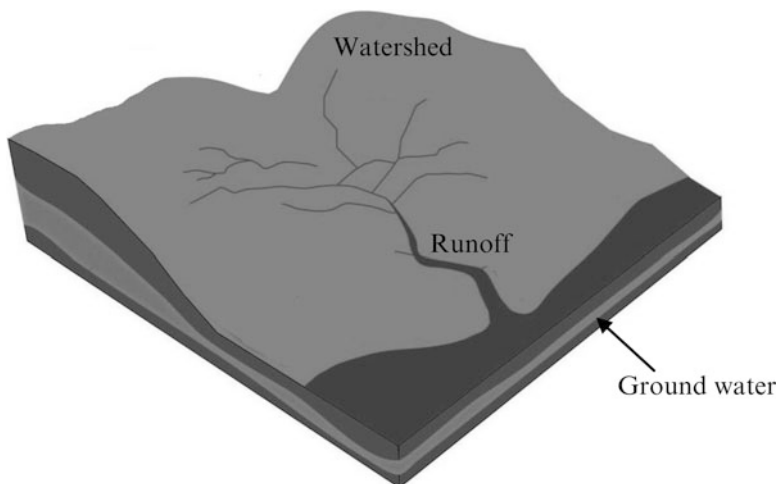

Fig. 3.5 GP framework in the outer mathematical model

Fig. 3.6 Schematic of a watershed

Although conceptual models can calculate runoff for a watershed, their processes are long and expensive. Therefore, to overcome these problems, Savic et al. (1999) applied GP to estimate runoff volume for Kirkton catchment in Scotland.

Rainfall on the Kirkton catchment is estimated using a network of 11 period gauges and 3 automatic weather stations at different altitudes. The daily average rainfall is calculated from weighted domain areas for each gauge. Stream flow is measured by a weir for which the rating has been adjusted after intensive current metering (Savic et al. 1999). They compared obtained results with HYRROM, one conceptual model by Eeles (1994) that applied 9 and 35 parameters for runoff estimation considering different land use variables. Moreover, GP employed different combinations rainfall, runoff and evaporation for one, two and three previous periods and rainfall at current period as the input data to estimate runoff of current period as the output data. Results showed that GP can present better solution even by fewer input data sets than other conceptual models by Eeles (1994).

3.3.1.2 Groundwater Levels Modeling

When rain falls, extra surface water and runoff moves under earth and forms groundwater. In groundwater, soil pore spaces and fractures of rock formations fill from water and called an aquifer. The depth at which soil pores and/or fractures become completely saturated with water is water table or groundwater level.

Groundwater contained in aquifer systems is affected by various processes, such as precipitation, evaporation, recharge, and discharge. Groundwater level is typically measured as the elevation that the water rises in, for example, a test well.

Two-dimensional groundwater flow in an isotropic and heterogeneous aquifer is approximated by the following equation (Bozorg Haddad et al. 2013):

$$\frac{\partial}{\partial x}\left(T\frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(T\frac{\partial h}{\partial y}\right) \pm W = Sy\frac{\partial h}{\partial t} \tag{3.1}$$

in which, T = aquifer transmissivity; h = hydraulic head; Sy = storativity; W = the net of recharge and discharge within each a real unit of an aquifer model, e.g., a cell in a finite-difference grid; W is positive (negative) if it represents recharge (discharge) in the aquifer; and x, y = spatial coordinates, and t = time.

Based on Eq. (3.1), mathematical models are used to simulate various conditions of water movement over time. However, mathematical simulation necessitates values of several parameters which may not be measured or their measurements incur considerable expenses (Fallah-Mehdipour et al. 2013a). Thus, to overcome those expenses and increase calculation accuracy in groundwater modeling, Fallah-Mehdipour et al. (2013a) applied GP in both prediction and simulation of groundwater levels. Results of the prediction and simulation process respectively help determining unknown and missed data in a time series. In order to modeling, three observation well of Karaj aquifer with water level variation in a 7-year (84-month) period have been considered. This aquifer is recharged from precipitation and recharging wells. To judge fairly about GP capabilities in groundwater modeling, results of the GP have been compared with adaptive neural fuzzy inference system (ANFIS). Results showed that GP yields more appropriate results than ANFIS when different combinations of input data sets have been employed in both prediction and simulation processes.

3.3.2 GP Application in Inner Mathematical Model

In this section, reservoir presents as an example of hydro systems in which GP is applied in mathematical model. In this model, GP is extracted operational rule as a constraint that illustrates when and how release water from reservoir.

Reservoirs are one of the main water structures which operate for several purposes, such as supplying downstream demands, generating hydropower energy, and flood control. There are several investigations in the short, long, and integrating short and long term (e.g., Batista Celeste et al. 2008) reservoir operation without considering any operational decision rules (Fallah-Mehdipour et al. 2013b). In these investigations, released water from reservoir is commonly identified as the decision variable.

The result of this type of operation is only determined for the applied time series. In order to operate a reservoir system in real-time, an operational decision rule can be used in reservoir modeling which helps the operator to make an appropriate decision to calculate how much (amount) and when (time) to release water from the reservoir.

To determine a decision rule, a general mathematical equation is usually embedded in the simulation model:

$$R_t = F_1 (S_t, Q_t) \tag{3.2}$$

in which, R_t, S_t and Q_t are release, storage and inflow at t^{th} period. Moreover, F_1 is linear or nonlinear function for transferring storage volume and inflow to the released water from the reservoir at each period.

The common pattern of aforementioned decision rule which is a linear decision rule that a, b and c are the decision variables (e.g., Mousavi et al. 2007; Bolouri-Yazdeli et al. 2014):

$$R_t = a \times Q_t + b \times S_t + c \tag{3.3}$$

Although, application of Eq. (3.3) as a decision rule is useful in real-time operation, this rule has a pre-defined linear pattern. It is possible to exist some decision rules with other mathematical frame (not just linear). GP can extract an embed equation in this reservoir model without any assumed pattern which is adapted with storage and inflow and their fluctuations at each period.

Moreover, the aforementioned rule involves Q_t needs commonly a prediction model may be coupled with decision rule to estimate inflow as a stochastic variable. Inappropriate selection of this prediction model increases calculations and impacts the reservoir operation efficiency (Fallah-Mehdipour et al. 2012). To overcome this inappropriate selection, GP can find a flexible decision rule which develops a reservoir operation policy simultaneously with inflow prediction. In this state, GP which presented its capability in inflow prediction, has been used as the reservoir simulation tool and two operational rule curves including water release, storage volume, and previous inflow/s (not in the current period (t)) are extracted.

Fallah-Mehdipour et al. (2012, 2013b) applied the GP application considering inflow of the current and previous periods. In these investigations, GP tries to close released water from reservoir to the demand by using different functions and terminals in the decision rule. Thus, GP rules presented a considerable improvement compare to the common linear decision rule.

Figure 3.7 presents GP framework in the real-time operation of reservoir. As it is shown, the random trees are generated in the first iteration. These trees are decision rules which explain a mathematical function including inflow, storage and release.

Accordingly, decision rule is embedded in the reservoir operation model and the released water from reservoir is calculated using continuity equation and limited constraint storage volume between minimum and maximum allowable storage ($S_{Min} < S_t < S_{Max}$). Then, the objective function yields considering minimization of deficit and maximization of generated energy in the supplying downstream demand and hydropower energy generation purpose, respectively. To find released water and storage in a feasible range, the constraints are considered in the optimization
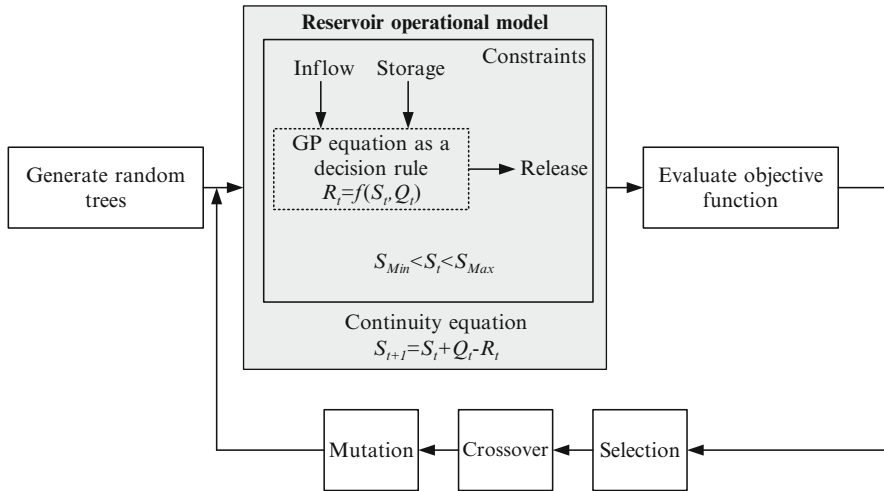
Fig. 3.7 GP framework in the real-time operation of reservoir

process by penalty. This penalty is added and subtracted in the minimization and maximization objective for each violation unit from feasible bound. The other GP process (selection, crossover and mutation) are continues to satisfy stopping criteria.

3.4   Concluding Remarks

There are many investigations that present successful application, development and adaptation of GP in the water engineering and hydrology. This chapter reviewed these investigations considering different aspects of GP application in the mathematical models that can be inner and outer of system modeling. Inner system modeling such as decision operational rule uses GP equation in the modeling process as same as other system equations. Thus, the output which is released water in reservoir system is adapted to the GP equation. In contrast, the outer mathematical model is widely used for developing an optimal existing relation between input and output data in water resources in a black-box method. In both aforementioned methods, GP illustrated appropriate solution and can be recommended for the future studies, because some highlight reasons:

- Appropriate capability to use in and out of models.
- Predict and simulate some phenomenon with a considerable fluctuation especially in the extreme bounds.
- Easy link with other models, softwares, and optimization techniques.

References

Batista Celeste, A., Suzuki, K., and Kadota, A. (2008). "Integrating long-and short-term reservoir operation models via stochastic and deterministic optimization: case study in Japan." Journal of Water Resources Planning and Management (ASCE), 134(5), 440–448.

Bolouri-Yazdeli, Y., Bozorg Haddad, O., Fallah-Mehdipour, E., and Mariño, M.A. (2014). "Evaluation of real-time operation rules in reservoir systems operation." Water Resources Management, 28(3), 715–729.

Bozorg Haddad, O., Rezapour Tabari, M.M., Fallah-Mehdipour, E., and Mariño, M.A. (2013). "Groundwater model calibration by meta-heuristic algorithms." Water Resources Management, 27(7), 2515–2529.

Cramer, N.L. (1985). "A representation for the adaptive generation of simple sequential programs." In Proceedings of an International Conference on Genetic Algorithms and the Applications, Grefenstette, John J. (ed.), Carnegie Mellon University. 24-26 July, 183–187.

Eakins, B.W. and Sharman, G.F. (2010). Volumes of the World's Oceans from ETOPO1, NOAA National Geophysical Data Center, Boulder, CO, 2010.

Eeles CWO: (1994) Parameter optimization of conceptual hydrological models, PhD Thesis, Open University, Milton Keynes, U.K.

Fallah-Mehdipour, E., Bozorg Haddad, O., and Mariño, M. A. (2012). "Real-time operation of reservoir system by genetic programming." Water Resources Management, 26(14), 4091–4103.

Fallah-Mehdipour, E., Bozorg Haddad, O., and Mariño, M. A. (2013a). "Prediction and Simulation of Monthly Groundwater level by Genetic Programming." Journal of Hydro-environment Research, 7(4), 253–260.

Fallah-Mehdipour, E., Bozorg Haddad, O., and Mariño, M. A. (2013b). "Developing reservoir operational decision rule by genetic programming." Journal of Hydroinformatics, 15(1), 103–119.

Gandomi, A.H., Yang, X.S., Talatahari, S., and Alavi, A. H. (2013). "Metaheuristic Applications in Structures and Infrastructures" Elsevier. 568 pages.

Guven, A., and Kisi, O. (2011). "Daily pan evaporation modeling using linear genetic programming technique." Irrigation Science, 29(2), 135–145.

Izadifar, Z., and Elshorbagy, A. (2010). "Prediction of hourly actual evapotranspiration using neural network, genetic programming, and statistical models." Hydrological Processes, 24(23), 3413–3425.

Koza, J. R. (1992). Genetic programming: on the programming of computers by means of natural selection. MIT Press, Cambridge, MA.

Koza, J. R. (1994). Genetic Programming II: Automatic Discovery of Reusable Programs. MIT 293 Press. Cambridge, MA.

Li, L., Liu, P., Rheinheimer, D.E., Deng, C., and Zhou, Y. (2014). "Identifying explicit formulation of operating rules for multi-reservoir systems using genetic programming." Water Resources Management, 28(6), 1545–1565.

Mousavi, S. J., Ponnambalam, K., and Karray, F. (2007). "Inferring operating rules for reservoir operations using fuzzy regression and ANFIS." Fuzzy Sets and Systems, 158(10), 1064–1082.

Nasseri, M., Moeini, A., and Tabesh, M. (2011). "Forecasting monthly urban water demand using extended Kalman filter and genetic programming." Expert Systems with Applications, 38(6), 7387–7395.

Orouji, H., Bozorg Haddad, O., Fallah-Mehdipour, E., and Mariño, M.A. (2014). "Flood routing in branched river by genetic programming." Proceedings of the Institution of Civil Engineers: Water Management, 167(2), 115–123.

Savic, D. A., Walters, G. A., and Davidson, J. W. (1999). "A genetic programming approach to rainfall-runoff modeling." Water Resources Management, 13(3), 219–231.

Sivapragasam, C., Vasudevan, G., Maran, J., Bose, C., Kaza, S., and Ganesh, N. (2009). "Modeling evaporation-seepage losses for reservoir water balance in semi-arid regions." Water Resources Management, 23(5), 853–867.

Traore, S., and Guven, A. (2012). "Regional-specific numerical models of evapotranspiration using gene-expression programming interface in Sahel." Water Resources Management, 26(15), 4367–4380.

Traore, S., and Guven, A. (2013). "New algebraic formulations of evapotranspiration extracted from gene-expression programming in the tropical seasonally dry regions of West Africa." Irrigation Science, 31(1), 1–.10.

Yang, X.S., Gandomi, A.H., Talatahari, S., and Alavi, A. H. (2013a). "Metaheuristis in Water, Geotechnical and Transportation Engineering" Elsevier. 496 pages

Yang, X.S., Cui, Z., Xiao, R., Gandomi, A.H., and Karamanoglu, M. (2013b). "Swarm Intelligence and Bio-Inspired Computation: Theory and Applications", Elsevier. 450 pages.