UC Santa Barbara
UC Santa Barbara Previously Published Works

Title

State-of-art of genetic programming applications in water-resources systems analysis

Permalink

https://escholarship.org/uc/item/25g8341n

Journal

Environmental Monitoring and Assessment, 192(2)

ISSN

0167-6369

Authors

Mohammad-Azari, Sahar
Bozorg-Haddad, Omid
Loáiciga, Hugo A

Publication Date

2020-02-01

DOI

10.1007/s10661-019-8040-9

Peer reviewed

State-of-art of genetic programming applications in water-resources systems analysis

**Sahar Mohammad-Azari · Omid Bozorg-Haddad ·
Hugo A. Loáiciga**

Abstract Evolutionary algorithms (EAs) have become competitive solvers of a wide variety of water-resources optimization problems. Genetic programming (GP) has become a leading EA since its inception in 1985. This paper reviews the state-of-the-art of GP and its applications in water-resources systems analysis. A comprehensive knowledge about GP's theory and modeling approach is essential for its successful application in water-resources systems analysis. This review presents variants of GP that have been proven useful in various applications to water resources problems. Several examples of applications of GP in water-resources systems analysis are herein presented. This review reveals GP's capability and superiority compared to other conventional methods, which makes it suitable for solving a wide variety of water-related problems including rainfall-runoff modeling, streamflow sediment prediction, flood prediction and routing, evaporation and evapotranspiration forecasting, reservoir operation, groundwater modeling, water quality modeling, water demand forecasting, and water distribution systems.

Introduction

The advent of evolutionary computation (EC) methods has revolutionized the field of water resources systems analysis and optimization. EC methods can tackle complex single-objective and multi-objective water resources systems problems that were previously intractable, as they may feature non-linear, discontinuous and non-differentiable, mixed-integer, and real variables of very large dimensionality (Koza 1994; Sreekanth and Datta 2010, 2011).

EC methods refer to a class of computational methods inspired by natural processes of evolution. EC is applied in the form of evolutionary algorithms (EAs) such as the genetic algorithm (GA), genetic programming (GP), evolutionary programming (EP), evolution strategy, and differential evolution (ESDE)

S. Mohammad-Azari
Department of Irrigation & Reclamation Engineering, Faculty of Agricultural Engineering & Technology, College of Agriculture & Natural Resources, University of Tehran, Karaj, Tehran, Iran
e-mail: Sahar.MAzari@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

H. A. Loáiciga
Department of Geography, University of California, Santa Barbara, CA 93016-4060, USA
e-mail: Hugo.Loaiciga@geog.ucsb.edu

(Babovic and Keijzer 2002). GP is a member of the EAs of relatively recent emergence. GP is applicable to a wide range of water-resources problems including rainfall-runoff prediction, evaporation and evapotranspiration modeling, streamflow and sediment modeling, water quality modeling, groundwater modeling, reservoir operation, flood routing, and water demand forecasting (Khu et al. 2001; Whigham and Crapper 2001; Liong et al. 2002; Rabunal et al. 2007; Aytek and Kisi 2008; Sivapragasam et al. 2008; Izadifar and Elshorbagy 2010; Kisi and Guven 2010; Arunkumar and Jothiprakash 2013; Danandeh Mehr et al. 2013; Lerma et al. 2013; Traore and Guven 2012; Orouji et al. 2014; Prakash and Datta 2014; Akbari-Alashti et al. 2015; Kasiviswanathan et al. 2016; Mirzaei-Nodoushan et al. 2016; Bozorg-Haddad et al. 2017), which feature unique conditions such that (1) the relations between system variables are poorly defined; (2) there are complex mathematics that defy classic treatment; (3) there is a wide range of data involved that require testing, compiling, and ranking; and (4) the problems' solutions are approximated and characterized by the average estimate and the standard deviation about a global optimal solution (Koza 1994; Babovic and Keijzer 2002; Orouji et al. 2013).

This paper's main goal is to present a review of GP applications in water-resources systems analysis. The characteristics of GP and its variants are discussed and evaluated to highlight its capabilities for solving complex water resources problems. The first section presents the theory of GP and its computational steps aided by an example. Next, different GP variants that have found application in water resources systems analysis are reviewed. Application areas include rainfall-runoff, evaporation and evapotranspiration, streamflow and sediment transport, floods, water supply, reservoir operation, water demand analysis, and groundwater management. A few applications include climate change, environmental sustainability, and greenhouse gas emissions to underline the breadth of range of GP. A conclusions section closes this work.

Materials and methods

The basis of GP is the Darwinian concept of survival of the fittest. According to this principle, those species that evolve and adapt in response to the conditions of their environments are the ones most likely to survive in the long term (Koza 1992). GP was introduced by Cramer (1985) and Koza (1992, 1994) developed GP into a practical tool.

GP applies a tree structure in its search for optimal solutions of a problem. The solution's tree structure features variables, operators, and functions. GP finds an appropriate tree of variables, operators, and functions for solving an optimization problem and for searching the best GP algorithmic parameters. Five steps are executed by GP in solving optimization problems:

1. Determination of terminal sets which include coefficients, the independent variables, and the state variables of an optimization problem. In other words, all the variables and constants of a problem are terminal sets.
2. Determination of the functional sets which contains arithmetic operations, logical and Boolean operators, or conditional statements organized in a tree structure to solve an optimization problem.
3. Determination of the fitness value of the trees of variables, operators, and functions applied to solve an optimization problem.
4. Determination of the GP parameters that control the solution runs, including the population size, the crossover rate, and the mutation rate.
5. Iterative improvement of the solution trees until satisfying a termination criterion that may be a predetermined number of generations of prospective solution trees, or a measure of the variation of solutions in consecutive generations, or a measure of the change of the fitness values of solution trees in consecutive iterations (Wang et al. 2009; Nasseri et al. 2011; Sarzaeim et al. 2017).

GP searches for optimal solutions by generating sets of trees randomly. There are several methods to generate the initial population of trees in the search space including the full method, the grow method, and the ramped half-and-half method. The full method generates full trees with all the leaves; it generates the tree nodes with the functional set, and only the tree terminals are optimized. The grow method allows the modeler to create trees of variable sizes and shapes. The ramped half-and-half method is a combination of the full method and the grow method (Koza 1992). The fitness function of each tree (individual solution) is calculated after generating the initial population. The fitness function (objective function) is the value of each tree, which is commonly

made equal to a norm of the differences between the predicted (GP's output) and observed values (target output). A suitable tree is one that has a negligible difference between the GP's output and the target output. The value of each tree can be calculated using several other methods.

GP is based on the principle that better individuals (solution trees) generate better children (improved solution tress). GP applies a selection process that concentrates the search for solutions in regions of the search space containing the superior solutions, which are the ones employed to generate the new, improved generation of solutions. Selection applies various operators for solution selection based on the trees' fitness. Better fitness improves the chance of current solutions to transfer its superior qualities to the next generation of solutions akin to the evolution and adaptation of successful species in nature (Koza 1992, 1994).

Methods for selecting the current superior solutions include the roulette wheel, tournament, and ranking methods. The next generation of solutions is produced by the crossover and mutation operators (Fallah-Mehdipour et al. b, c, 2016). Crossover selects two parents or solution trees and their sub-trees are crossed over randomly at cross points (these points are the nodes in the solutions organized as trees in GP). Two children (new solutions) are generated and replace the parent sub-trees. The mutation operators are applied at mutation points or nodes. Each node is chosen probabilistically and is replaced by an independent variable. The generated solution trees of mathematical operations are the inputs to the next generation of trees. This process is continued until reaching a termination criterion. The flowchart of the calculation steps of GP is shown in Fig. 1.

An example of GP computations follows to illustrate its basic algorithmic nature. Consider five points (X, Y) as follows: (0.1, 1.11), (0.13, 1.15), (0.16, 1.18), (0.18, 1.21), (0.2, 1.24). There are mathematical functions that could be identified representing a relation between X and Y. GP can be applied to search for the best relation between X and Y that minimizes the error between observed and predicted values as quantified by the root mean square error (RMSE) and correlation coefficient (R^2). Several relations are graphed in Fig. 2. It is evident in Fig. 2 that the quadric equation achieves the best relation between the points based on the RMSE and the R^2. GP generates set of functions and operators in a complex search process that best relate the input and outputs of a system.

Consider the same (X, Y) points introduced above, and suppose we seek the best relation between them using GP. To solve this problem, notice that (i) the population size equals four, (ii) we assume that the functions and terminals are (+, *) and (0, 1, 2), respectively, (iii) an individual's (solution tree's) fitness less than 0.01 represents the termination criterion. Considering (i) through (iii) above GP creates randomly initial solution trees of arithmetic operators, mathematical functions, and variables, and proceeds to determine the optimal mathematical expressions. The four initial populations generated with the grow method are shown in Fig. 3.

The value of each individual (solution tree) is calculated in the next step. The RMSE is considered in this example as the fitness function. The RMSE values for the four individuals (solution trees) shown in Fig. 3 are calculated based on their equations (**a** $x + 1$, **b** $x^2 + 1$, **c** $2x + 1$, and **d** $x^2 + x$). The calculated RMSE for the five points [(0.1, 1.11), (0.13, 1.15), (0.16, 1.18), (0.18, 1.21), and (0.2, 1.24)] introduced above equal 0.03, 0.16, 0.13, and 1 corresponding to individuals **a**, **b**, **c**, and **d**, respectively. The first individual is the fittest one (with RMSE = 0.03); therefore, it is advanced to the next generation without any alteration.

Two individuals are selected as parents to create offspring solutions (children). The fitter individuals have a better chance of being selected as parents. Next, crossover is performed to generate individuals for the next generation. The first crossover considers the right side x of the second individual (Fig. 3b) and the + function of the first individual (Fig. 3a). The children produced by this crossover are illustrated in Fig. 4a. It is shown in Fig. 4a that generated children of this crossover are x and $x^2 + x + 1$. The second crossover operation selects the function + of the first individual and the right side x of the fourth individual (Fig. 3d), to produce the children x and $x^2 + x + 1$. This process is shown in Fig. 4b.

One mutation is performed on the third individual (Fig. 3c), which changes its fitness value from 0.13 to 0.1. This mutation is illustrated in Fig. 5. The next generation includes four individuals, which are the two children resulting from parental crossover, one child from mutation, and the fittest individual (the first one or that shown in Fig. 3a) that is copied to next generation without alteration (Fig. 6). Among these individuals, there is an equation ($x^2 + x + 1$) which satisfies the termination criterion of the fitness lower than 0.01. This equation is the solution of this example problem.

Fig. 1 Flowchart of GP's
calculation steps
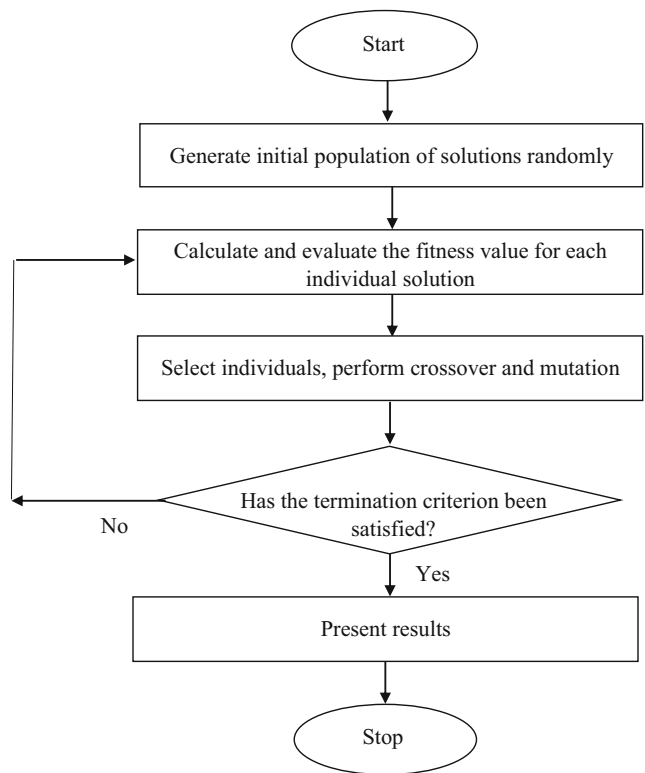
Several variants and new developments of GP have
emerged recently that attempt to overcome the limita-
tions of traditional, tree-based, GP. Linear genetic pro-
gramming (LGP), fixed-length gene genetic program-
ming (FLGGP), and gene expression programming

(GEP) are the leading three variants of GP applied in
water-resources-related problems. In fact, the mentioned
variants of GP can yield more accurate and efficient
structures and also mathematical relations compared to
traditional GP which is also simpler for interpretation.

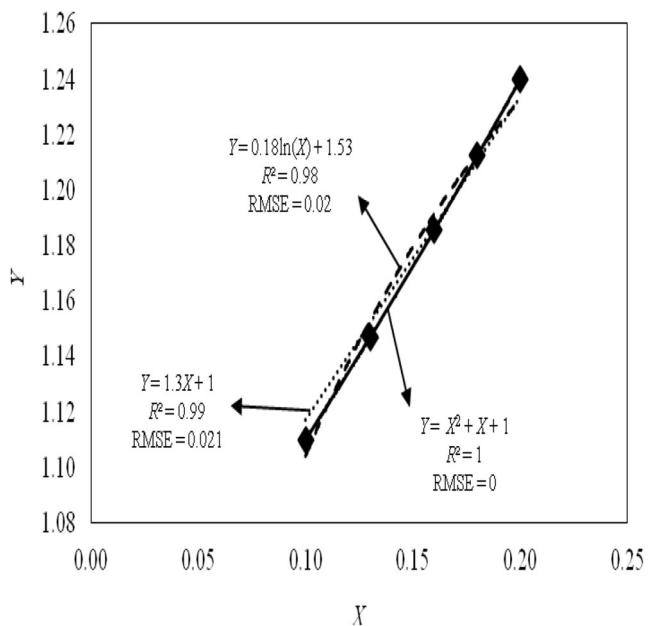Fig. 2 Several functions between
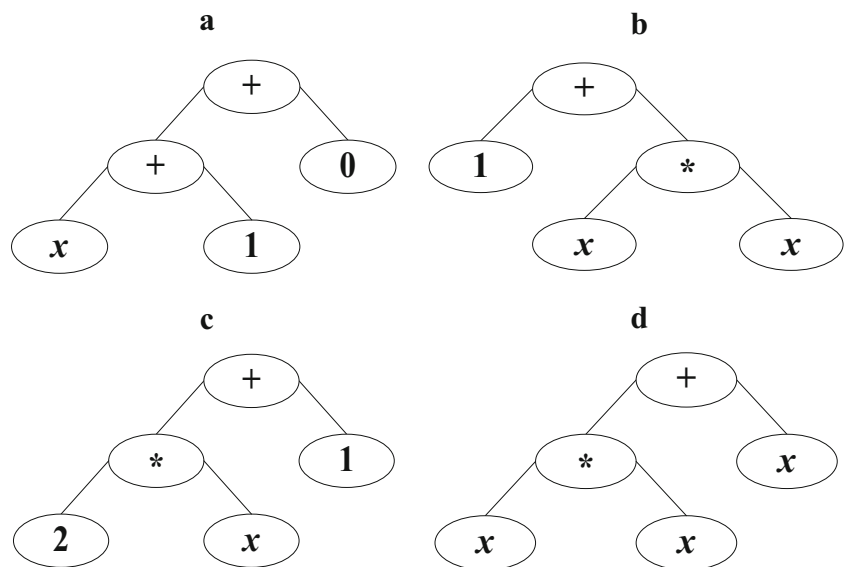five points (X, Y)

Fig. 3 Initial populations **a** $x + 1$, **b** $x^2 + 1$, **c** $2x + 1$, and **d** $x^2 + x$ in the first generation

On the other hand, traditional GP can present only one mathematical relation between input and output sets while some of its variants are capable of deriving more than one mathematical relation especially in water-resources systems with more than one subset. These variants are described in the following paragraphs.

The classical approach or tree-based GP (TGP) applies expressions by means of a functional programming language. In contrast, LGP is a linear variant of GP that substitutes expressions made by a functional programming language with programs of an imperative language. The main characteristic of LGP in comparison to conventional GP is the graph-based data flow that results from the multiple applications of indexed variables (Banzhaf et al. 1998).

LGP manipulates individuals with binary machine codes that are executed without an interpreter during the fitness calculation (Nordin 1997; Banzhaf et al. 1998). LGP is relatively simple, yet it can develop complex functions whose evolution is carried out with simple arithmetic functions. The functional set of this method is composed of arithmetic operators, conditional branches, and functional calls. Each element of a functional set involves an assignment to a variable. This facilitates the application of multiple program inputs in LGP compared with conventional GP. Functions can operate on two variables or one variable and one constant. The terminal set of LGP is formed by variables and constants. On the other hand, each function is encoded in a four-dimensional vector. For each of two parent segments, random position and random length

are selected. If one of the children exceeds the maximum segment length, then crossover is restarted by exchanging segments that have equal size. The crossover points only take place between functions (Banzhaf et al. 1998). The mutation operation randomly replaces the function identifier inside functions, variables, or constants. The best individual of LGP is converted into a functional representation by successive replacements of its input and output starting with the last effective function (Brameier and Banzhaf 2001). The flowchart of computation steps of LGP is illustrated in Fig. 7.

FLGGP is one of GP's variants that have been employed in water-resources management studies. FLGGP attempts to find multiple mathematical equations simultaneously with appropriate accuracy by combining the genetic algorithm (GA) and GP's characteristics seeking to overcome the individual limitations of GP and GA. In fact, as stated above, mathematical expressions extracted by GP have a tree structure with various functions, operators, and variables. More precise expressions are increasingly complex, which may lead to very complex expressions not found in the real world. FLGGP was developed as a variant of GP to overcome the complexity of GP calculated expressions and to calculate more than one mathematical expression simultaneously (Fallah-Mehdipour et al. 2013b; Akbari-Alashti et al. 2015).

FLGGP generates sets of individuals or solutions with fixed length employing a uniform distribution as the initial population of solutions. These individuals include fixed numbers of genes, and these genes

Fig. 4 Crossover between **a** individuals $x + 1$ and $x^2 + 1$, **b** individuals $x + 1$ and $x^2 + x$
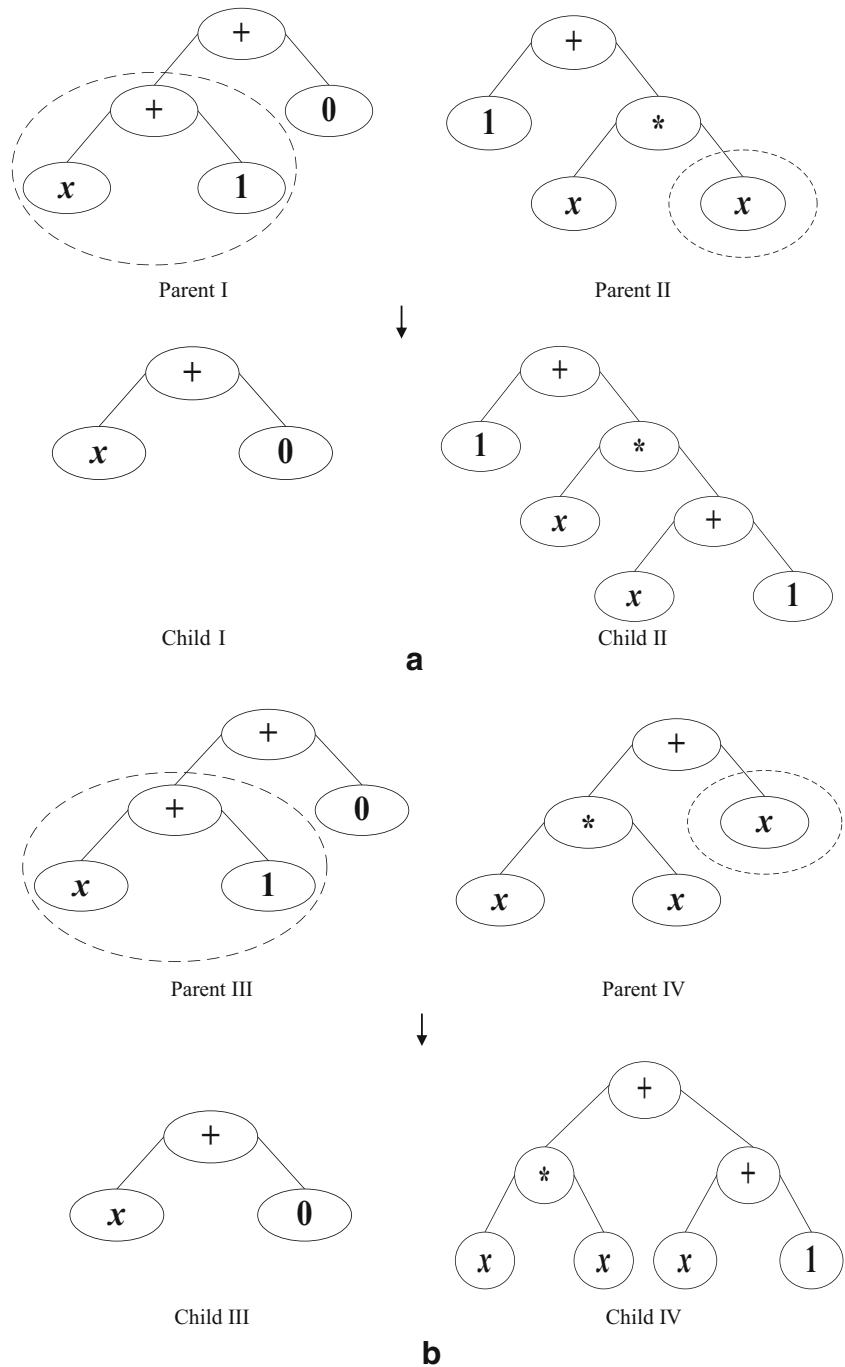
represent mathematical expressions related to variables present in the problem being tackled.

FLGGP generates a new population of solutions using selection, crossover, and mutation operators similarly as done by the GA. The generated individuals introduce new relations in the next generation, and the search proceeds until reaching a maximum predetermined number of iterations (Fallah-Mehdipour et al. 2013b). Several studies have been reported that have applied FLGGP for solving water resources problems, some of which are described in the following sections.

GEP was developed by Ferreira (2001). This method generates populations of solutions and ranks them according to their fitness, then implements genetic

Fig. 5 Tree structures, **a** before and **b** after mutation
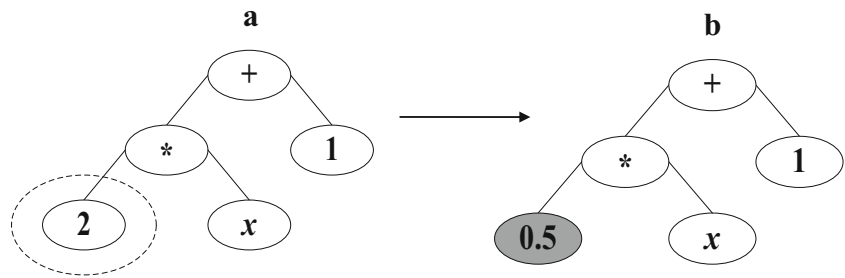
variations using one or more genetic operators to advance to the next population of improved solutions. The main difference between GA, GP, and GEP is related to the nature of the solutions. The GA applies linear strings of fixed length. GP relies on solutions that are non-linear and of different types and sizes (parse trees). GEP uses linear strings of fixed length (Ferreira 2001). The first stage of GEP is generating the initial population of solutions. This process starts randomly or using available information about the problem being solved. The solutions are represented as tree structures that are evaluated with a fitness function. The fitness function is usually made of specified objectives. The fittest solutions of an algorithmic iteration have a higher chance for generating new solutions. This process is repeated for several iterations. The search process for an optimal solution continues until reaching a termination criterion, at which point the current solution is reported (Ferreira 2001). The flowchart of the computational steps of GEP is illustrated in Fig. 8.

The cited variants and extensions of GP are the most commonly applied in water-related problems. The section describes several applications of GP to a variety of water-resources problems.

Results and discussion

The applications of GP in water resources include estimation, prediction, and simulation in hydrology and hydraulics, evapotranspiration, water quality, groundwater, risk assessment, sediment transport, water demand prediction, and reservoir operation, among the most common applications (Gandomi et al. 2015).

Rainfall-runoff models can be generically categorized as black-box, conceptual, and physically based distributed models. The application of these models imposes limitations to process modeling and prediction. These types of models require a wide range of data for modeling purposes (parameters such as soil

Fig. 6 Population of the next generation **a** $x + 1$, **b** x, **c** $0.5x + 1$, and **d** $x^2 + x + 1$
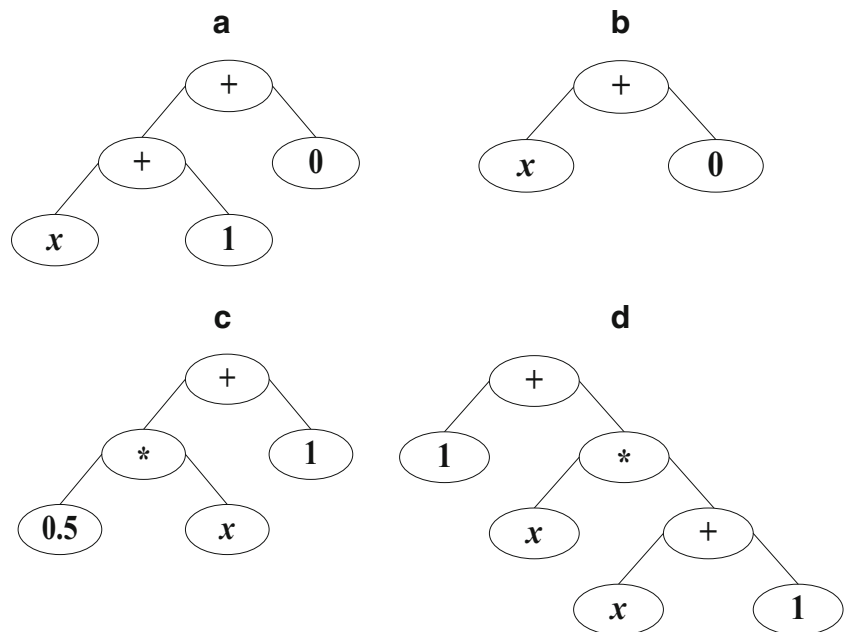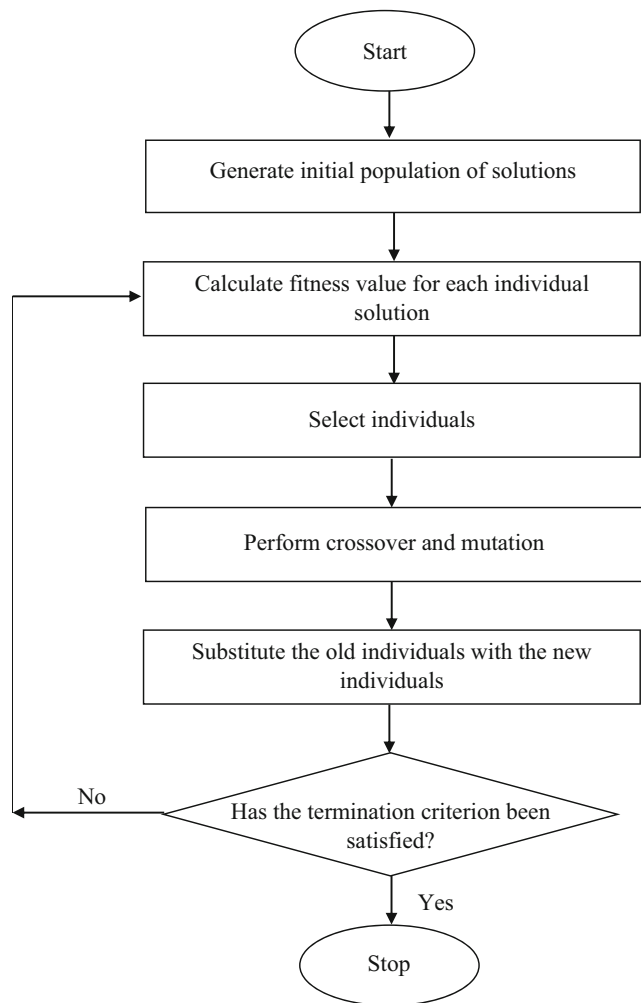
Fig. 7 Flowchart of LGP's computational steps

characteristics, basin characteristics, river networks, and other inputs). Conceptual models have limited capacity for handling non-linearity and non-stationary phenomena (Savic et al. 1999; Khu et al. 2001; Whigham and Crapper 2001).

The artificial neural network (ANN) is a type of black-box model that can model non-linear and complex hydrologic processes. Yet, the number of inputs and hidden neurons required by ANN must be obtained through a time-consuming trial–error process (Savic et al. 1999; Khu et al. 2001). The study of Savic et al. (1999) is one of the first studies that applied GP to rainfall-runoff modeling and prediction. The latter study compared ANN and GP in estimating runoff in a Scottish catchment. Its results indicated the superiority of GP over ANN based on the R^2.

Nourani et al. (2011) linked wavelet analysis to GP in order to form a hybrid model for detection seasonality

patterns in rainfall-runoff process. The results were also compared to ANN and GP based on RMSE and R^2 which indicated the capability of hybrid model in monitoring both short- and long-term patterns.

In other studies, Havlicek et al. (2013) and Adhikay et al. (2015) investigated the applicability of combined GP and basic hydrological models and GP-derived variogram model within ordinary kriging, respectively. In a former study, Havlicek et al. (2013) combined GP and basic hydrological modeling concepts in order to improve rainfall-runoff forecasts. The performance of the proposed model was also compared to ANN and GP model results which indicated the accuracy of combined model in simulation based on maximum absolute error (MAE), RMSE, and NSE. In the latter study, Adhikay et al. (2015) applied GP to derive a variogram model. They also investigated the applicability of GP-derived variogram model within ordinary kriging for spatial

Fig. 8 Flowchart of GEP's computational steps
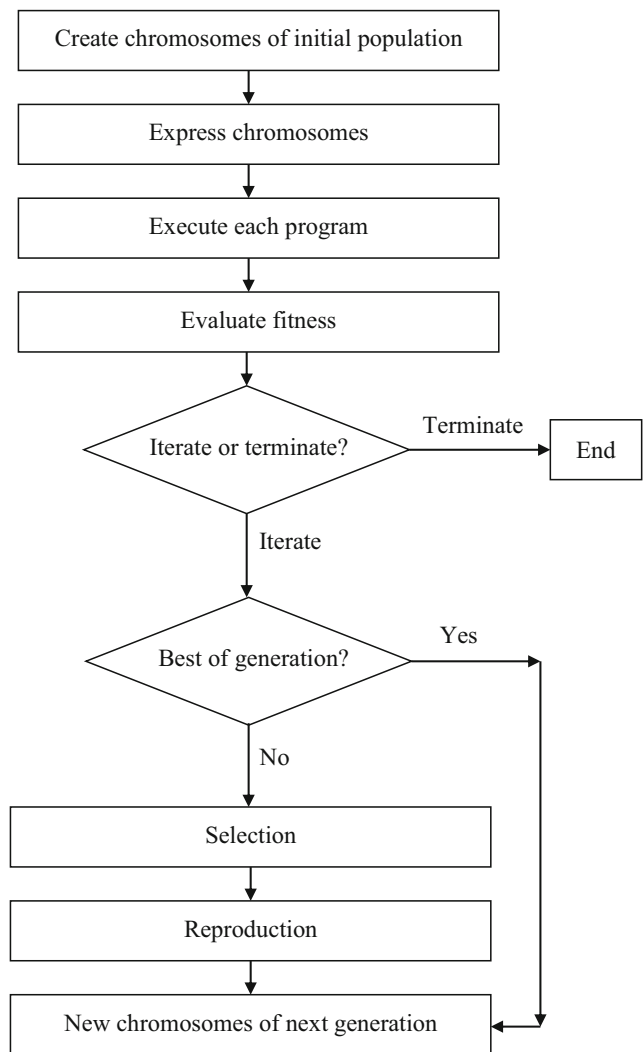
interpolation. The results indicated the superiority of GP-based ordinary kriging over traditional ordinary kriging and ANN-based ordinary kriging.

A rainfall-runoff study featuring a GP application was reported by Danandeh Mehr and Nourani (2018). The rainfall-runoff model was integrated with multigene-GP to enhance timing accuracy of GP-based rainfall-runoff models. They evaluated the timing and prediction accuracy of the proposed model based on RMSE and NSE efficiency criteria. The results indicated the superiority of multigene-GP compared to monolithic GP for identifying the underlying structure of the rainfall-runoff process.

The cited applications and others are listed in Table 1 (Khu et al. 2001; Whigham and Crapper 2001; Liong et al. 2002; Rabunal et al. 2007). All the aforementioned studies stressed the capability and superiority of GP over

other methods that have been applied to rainfall-runoff modeling and forecasting.

The accurate predictions of streamflow and sediment transport are important in water-resources problems. Streamflow prediction generally is made by two methods. One is focused on the study of rainfall-runoff processes to model underlying physical laws; the other method is the pattern recognition method in which the streamflow patterns are recognized based on antecedent records. Both methods required a wide range of catchment data and they require many simplifying assumptions (Danandeh Mehr et al. 2013).

Sediment transport estimation features two general methods, which are physically based models or simplified partial differential equations and rating curves. Although the cited methods are commonly employed in

Table 1 Summary of application of GP to rainfall-runoff modeling

Application	Year	Authors	Comparison	Description
Rainfall-runoff modeling	2001	Khu et al.	Autoregression and Kalman filter	GP is a better updating tool for real-time flow forecasting
	2001	Whigham and Crapper		GP is a useful tool for developing hydrological models, especially when surface water movement and water losses are poorly understood
	2002	Liong et al.	Regression	GP-induced rainfall runoff functions are a viable alternative to traditional rainfall runoff models
	2007	Rabunal et al.	ANN	GP predicts the flow rate resulting from rain
	2011	Nourani et al.	ANN and GP	Wavelet analysis was linked to GP. The capability of hybrid model was proved in the case of monitoring short- and long-term rainfall-runoff patterns
	2013	Havlicek et al.	ANN and GP	GP and basic hydrological modeling were combined to improve rainfall-runoff process. The combined model increased the accuracy in simulation
	2015	Adhikay et al.	Ordinary kriging and ANN-based ordinary kriging	GP-derived variogram model within ordinary kriging is applicable for spatial interpolation
	2015	Sivapragasam et al.	Linear, polynomial, and logarithmic model	GP detects non-linearity in what is thought to be a linear process
	2018	Danandeh Mehr and Nourani	Monolithic GP	Multigene-GP was integrated with season algorithm to enhance timing accuracy of GP-based rainfall-runoff models

sediment estimation studies, they have some limitations which introduce estimation inaccuracies (Aytek and Kisi 2008). GP has emerged as a powerful tool that overcomes the limitations of streamflow prediction and sediment estimation methods. Makkeasorn et al. (2008) and Guven (2009) pioneered the application of GP to streamflow forecasting. Garg and Jothiprakash (2009) employed GP to estimate the volume of sediment production. Their results indicated GP captured the trend and magnitude of sediment transport well. These studies demonstrated GP can effectively capture the non-linearity of streamflow and sediment production. Other studies were those by Danandeh Mehr et al. (2013). The latter authors applied LGP to forecast monthly streamflow and compared the performance of GP to wavelet-artificial neural network (WANN). The results indicated a superior performance of LGP over WANN based on the Nash–Sutcliffe efficiency and the RMSE. A study by Danandeh Mehr (2018) applied the genetic algorithm in combination with GEP as a hybrid model for streamflow forecasting in intermittent streams. The proposed hybrid model was compared to GP, GEP. The results indicated the suitability of the hybrid model in such studies. A summary of streamflow and sediment prediction studies is listed in Table 2.

Two approaches are applied to hydrograph prediction in river reaches, namely, the hydraulic and hydrologic approaches. Hydraulic approaches' calculations are time consuming. For this reason, the hydrologic approach is frequently used, although it relies on simplifying assumptions between river-reach input, output, and storage (Sivapragasam et al. 2008; Orouji et al. 2014). GP has been applied to overcome the shortcomings of hydraulic and hydrologic approaches. Among those studies are those by Fallah-Mehdipour et al. (2013a, b, c, d) and (Fallah-Mehdipour et al. 2016), Hakimzadeh et al. (2014), Orouji et al. (2014), and Hu et al. (2016).

Fallah-Mehdipour et al. (2013b) estimated the stage hydrograph of compound channels with GP. Their results indicated that GP reduced the computational burden and had better accuracy of hydrograph estimation than the coupled characteristic-dissipative-Galerkin procedure in one-dimension (CCDG-1D) hydraulic method. Fallah-Mehdipour et al. (2016) applied GP for flow routing in simple and compound channels. Results indicated that GP yields acceptable predicted hydrographs and the computational burden was decreased compared to the Muskingum model. Hakimzadeh et al. (2014) applied GP to simulate outflow hydrographs. GP-obtained values for outflow were

Table 2 Summary of application of GP and its variants to streamflow and sediment prediction

Application	Year	Authors	Comparison	Description
Streamflow and sediment prediction	2008	Aytek and Kisi	Rating curves and multi-linear regression techniques	GP formulation performs quite well in comparison to sediment rating curves and multi-linear regression models. GP is practical for use
	2008	Makkeason et al.	Neural network (NN)	The superiority of GP in streamflow forecasting was proven in comparison to NN
	2009	Guven	NN	Performance of LGP is better than NN. LGP is a powerful tool in predicting river flow data
	2011	Azamathulla et al.	Stage rating curve (SRC) and regression techniques	GEP was applied to model the stage–discharge relationship. The performance of the GEP model was substantially superior to both GP and conventional models
	2012	Kisi et al.	ANFIS, ANN, SVM	GP is superior to the ANFIS, ANN, and SVM models for estimating daily suspended sediment load
	2013	Maheswaran and Khosa		Multi-resolution GP (MRGP) was applied to resolve non-stationeries with a hybrid of wavelets-based multi-resolution decomposition and GP. MRGP is a promising approach for flow forecasting
	2018	Danandeh Mehr	Multi-linear regression, GP and GEP-linear regression	Genetic algorithm applied in combination with GEP is a hybrid model for streamflow forecasting in intermittent streams. The proposed hybrid model was compared to GP, GEP and results indicated the suitability of model in such studies

in good agreement with observed values, and were more than results calculated with other empirical methods.

Orouji et al. (2014) applied an extended version of the Muskingum hydrologic method and GP for flood routing in branching rivers. The latter authors compared the results of their study to those obtained with the Saint–Venant hydraulic method. Results indicated the objective function's improvement with GP compared to the extended Muskingum method for routing floods with return periods ranging from 10 to 100 years. These results established the effectiveness of GP for flood routing in branching rivers. Hu (2016) reported an application of GP to solve a symbolic regression problem for flood risk assessment in Beijing. Results indicated that GP could meet the requirements for risk assessment in an artificially intelligent manner.

Evaporation and evapotranspiration are poorly understood components of the hydrologic cycle despite their importance at all spatial and temporal scales (Brutsaert 1982). There are inherent non-linearities and complexities in these two processes (Soucha et al. 1996). Unlike precipitation and river flow, which can be measured directly, evaporation and evapotranspiration are estimated by pan-evaporimeter, lysimeter, mass transfer, energy balance, combination (mass transfer and energy balance), and water budget methods. Traditional measurements of evaporation and evapotranspiration are subject to several assumptions that may not be

appropriate for large-scale studies (Soucha et al. 1996; Drexler et al. 2004). Micrometeorological methods such as the energy-balance–Bowen-ratio (EBBR) and eddy-covariance (EC) have found widespread applications for estimating actual evaporation (Drexler et al. 2004).

Parasuraman et al. (2007) evaluated GP's capacity to model the evapotranspiration process. They compared the performance of GP to ANN models and the Penman–Monteith combination (mass transfer/energy balance) method. Results indicated GP-evolved relations are understandable and well suited to modeling the dynamics of evapotranspiration. Guven et al. (2008) implemented GP to estimate the reference evapotranspiration. Their results indicated GP-evolved equations provided satisfactory results and can be applied as an alternative to conventional models including the Penman–Monteith, Jensen–Haise, and Hargreaves–Samani methods. Kisi and Guven (2010) applied LGP to daily reference evapotranspiration modeling. The accuracy of LGP was compared to support vector regression (SVR), ANN, and empirical models for evapotranspiration modeling. The efficiency criteria including RMSE, mean-absolute errors, and R^2 were applied to compare the accuracy of the models' predictions. The findings indicated superiority of LGP compared to SVR and ANN techniques. Applications of GP to evaporation and evapotranspiration estimation are listed in Table 3, and they are the works by Izadifar and Elshorbagy

Table 3 Summary of application of GP and its variants to evaporation and evapotranspiration forecasting

Application	Year	Authors	Comparison	Description
Evaporation and evapotranspiration forecasting	2010	Izadifar and Elshorba-gy	ANN, multi-regression	GP-derived results had less error compared to ANN and multi-regression
	2011	Guven and Kisi	GEP, multilayer perceptrons (MLP), radial basis neural networks (RBNN), generalized regression neural networks (GRNN)	LGP was employed successfully in modeling evaporation
	2012	Traore and Guven	–	GEP is applied to evapotranspiration modeling. Statistically, GEP is an effective modeling tool for computing evapotranspiration successfully
	2013	Arunkumar and Jothipra-kash	ANN and mode tree (MT)	GP predicts the reservoir evaporation better than ANN and MT models

(2010), Guven and Kisi (2011), Arunkumar and Jothiprakash (2013), and Traore and Guven (2012).

Water quality modeling is an important water-related problem. There are many studies involving applications of genetic algorithms in water quality modeling (Osman and Badr 2010), and there are applications of PIKAIA to the calibration of water quality parameters (Pelletier et al. 2006; Lerma et al. 2013). Yet, applications of GP to water quality studies are scarce. Chen (2003) applied GP to reservoir water quality monitoring. The results indicated the better performance of GP compared to traditional regression methods. Orouji et al. (2013) simulated water quality parameters (sodium, potassium, magnesium, sulfates, chloride, pH, electrical conductivity, and total dissolved solids) at the Astane station in the Sefidrood river in Iran with GP and with the adaptive network-based fuzzy inference system (ANFIS). Their results indicated GP is an effective tool for quality parameter determination in the training (calibration) and testing steps compared with ANFIS.

Mirzaei-Nodoushan et al. (2016) applied GP to long-term prediction of streamflow and riverine total dissolved solids (TDS) in the Karoon River, Iran. They compared the results with observed and short-term predicted values. Results confirmed the applicability and suitability of GP for predicting Karoon river's streamflow and TDS.

Large-scale numerical simulation models and complex decision-making models have been applied to groundwater management. Such applications require a wide range of data and rigorous model calibration. The measurement of model parameters such as the hydraulic conductivity, storage coefficient, and porosity is elaborate and expensive. Moreover, there are considerable uncertainties related to these parameter estimates (Shiri and Kisi 2011; Fallah-Mehdipour et al. 2014).

GP has been proposed as a suitable tool for groundwater characterization relying on available data. GP provides a non-physical analysis for natural phenomena that can be effective in groundwater resources management. Sreekanth and Datta (2010, 2011) applied GP to saltwater intrusion management in coastal aquifers. The latter authors compared GP with a modular neural network (MNN). Their results indicated the less uncertainty of estimates by GP compared to the MNN model due to fewer parameters used in GP. The GP-based models were better suited for groundwater optimization. Fallah-Mehdipour et al. (2014) applied GP and ANFIS to extract governing groundwater flow equations. Their results showed the flexibility of GP over ANFIS in time-series modeling of groundwater variables. Recent applications of GP to groundwater modeling are those by Prakash and Datta (2014) and Kasiviswanathan et al. (2016). Several GP applications in groundwater modeling have been reported in Table 4.

Common tasks in reservoir operation are inflow prediction and the extraction of rule curves for reservoir releases (Fallah-Mehdipour et al. 2013a; Ashofteh et al. 2014). Ashofteh et al. (2015) implemented multiobjective GP to extract operation rule curves. Results indicated the capability of GP in extraction operation rules in a system with one, two, or more objectives. Recent applications of GP to reservoir operation were reported by Akbari-Alashti et al. (2015), Ashofteh et al. (2017), and Bozorg-Haddad et al. (2017). In the former study, GP and FLGGP were compared to extract static

Table 4 Summary of application of GP and its variants to groundwater modeling

Application	Year	Authors	Comparison	Description
Groundwater modeling	2011	Shiri and Kisi	ANFIS	GP is superior to ANFIS in giving explicit expressions for water table depth prediction
	2013b	Fallah-Mehdipour et al	ANFIS	GP is an effective tool for determining groundwater levels
	2014	Prakash and Datta		The capability of GP was proven in groundwater pollution sources identification
	2016	Kasiviswanathan		GP is a useful tool for effective planning and management of groundwater resources

and dynamic operation rules. Their results indicated that FLGGP is a powerful tool without the limitation of the classic GP. The comparison also demonstrated the superiority of dynamic operation rules over static operation rules.

Ashofteh et al. (2017) applied logical GP to derive optimal hedging rules of reservoir under baseline and climate change conditions. Their findings indicate the improvement of the objective function in logical GP compared to traditional GP under baseline and climate change conditions. Furthermore, calculated results with LGP approach and standard operation policy (SOP) were compared, and the results indicated the better performance of logical GP compared to traditional GP based on higher R^2 values.

Bozorg-Haddad et al. (2017) applied GP to calculate optimal monthly water allocation downstream of the Zarrineh-Roud Dam, Iran. Results indicated that GP obtained rule curves and water allocations that are very close to the optimal allocations obtained by constrained non-linear programming. A recent GP application was reported by Ashofteh et al. (2017). The latter authors introduced logical GP by adding logical functions and operators to traditional GP to calculate reservoir operation hedging rules for agricultural water supply. Results demonstrated the superiority of logical GP compared to traditional GP under baseline and climate change condition. Studies reporting applications of GP are listed in Table 5.

Water demand forecasting has been approached with various times series models and multivariate regressions. Methods for studying the performance of pipeline networks are physically based and computationally burdensome (Wu and Yan 2010; Xu et al. 2011). GP is a powerful tool for water demand forecasting and for the analysis of water distribution networks.

Nasseri et al. (2011) applied GP forecast water demand with an explicit optimal formula. Their

results obtained with GP and hybrid models of the extended Kalman filter GP (EKFGP) demonstrated the effect of observational accuracy on water demand prediction and online water demand forecasting. Wu and Yan (2010) applied TGP and GEP to construct demand forecasting models for water systems. Their results indicated TGP and GEP are effective for demand forecasting. Xu et al. (2011) applied three methods to model the failure of pipeline networks relying on statistical models coupled with GP. GP was shown to accurately predict the performance of water distribution networks.

Shabani et al. (2018) reported an application of GP to water demand forecasting. They proposed an approach based on GEP coupled with unsupervised learning for short-term water demand forecasting. Results indicated that coupling GEP with the unsupervised learning is a promising emerging non-linear modeling technique.

Other studies dealing with GP applications to water-related problems, climate, climate change, and CO_2 emission are those by Azamathulla (2012), Azamathulla and Ahmad (2012), Baareh (2018), Puente et al. (2019), and Liu and Shi (2019) which are discussed next.

Azamathulla (2012) applied GEP for scour prediction obtaining satisfactory results compared to ANN in predicting scour depth at an abutment. Azamathulla and Ahmad (2012) proposed GP to predict critical submergence which produced satisfactory results compared to existing predictors.

Baareh (2018) applied GP for carbon gas emission estimation. The results indicated the effectiveness and robustness of GP in solving and dealing with climate pollution problems. In one of the recent studies of GP application, Puente et al. (2019) applied GP to calculate new Vegetation Indices (VIs). Results indicated that the

Table 5 Summary of application of GP and its variants to reservoir operation

Application	Year	Authors	Comparison	Description
Reservoir operation	2012	Fallah-Mehdipour et al.		Obtained rule by GP based on deterministic variables is effective in determining optimal rule curves for reservoir and inflow prediction
	2013b	Fallah-Mehdipour et al	GA	A fixed-length gene genetic programming (FLGGP) rule was developed and calculated based on GP. FLGGP is flexible and effective for determining optimal rule curves
	2015	Akbari-Alashti et al.	GP	FLGGP compared to GP to extract statistic and dynamic operation rules
	2017	Ashofteh et al.	GP	Logical GP compared to traditional GP in order to derive operation hedging rules for agricultural water supply
	2017	Bozorg-Haddad et al.	Non-linear programming	GP applied to derive optimal monthly water allocation. GP-obtained rule curves were very close to constrained non-linear programming

synthetic indices calculated by GP produce better approximation to the soil cover factor in comparison to state-of-the-art indices like NDVI and EVI. In another study, Liu and Shi (2019) developed a recursive approach to long-term prediction of monthly precipitation using GP. The results indicated that GP can improve the more accurate predictions of monthly precipitation compared to statistical models.

It is evident that GP and its variants are applicable to a wide range of water, climate, and environmental-related problems. Based on our survey of published studies, the superiority of GP and its variants has been proven over some other common physical and statistical methods and has been applied to various types of water-resources-related problems.

The main advantage of GP is its ability to simulate complex processes efficiently. Another advantage over other methods is its clear and structured representation of a system being modeled without the need for system identification. The number of GP algorithmic parameters is small in comparison with those extant in other models. On the other hand, one criticism of GP is that it generates equations for the management or prediction of complex systems that are difficult to interpret. Another challenge in the application of GP is selecting appropriate parameters that control algorithmic execution. These parameters control the convergence of the GP algorithm to global optima.

Conclusions

This study reviewed GP applications to solve water-resources problems. A review of published studies indicates that the scope of GP and its variant applications is predominant on rainfall-runoff prediction, evaporation modeling, and flood routing. On the other hand, the applications of GP have proven its efficiency as a computational algorithm. GP is also a suitable tool for simulation of complex phenomena in water-resources problems given its dynamic and evolutionary behavior. Furthermore, GP has been shown to outperform a wide range of data-driven models applied in water-resources systems including rating curve estimation, unit hydrograph method, linear regression, autoregressive moving average, and autoregressive integrated moving average, ANN, and SVR. It overcomes the limitations of other competing models. The latter models are black-box models whereby input and outputs are known without understanding the processes which transforms inputs into outputs. In contrast, GP plays a key role in finding appropriate relations for the quantitative description of physical phenomenon. GP can overcome the limitations of other models by evolving its model structure. GP functional form does not assume prior solutions, which constitutes GP's key advantage over competing models. GP can find suitable mathematical relations between independent and dependent variables in a water-resources system.

Many variants of traditional GP have emerged and have been applied successfully to solve water-resources systems problems. The wide range of water-related problems and their complexity has called for the application of other variants of GP by hybridizing GP with other evolutionary algorithms. The hybrid variants have proven effective and more efficient than traditional GP.

Compliance with ethical standards

Conflict of interests   None.

References

Adhikay, S. K., Muttil, N., & Yilmaz, G. (2015). Genetic programming-based ordinary kriging for spatial interpolation of rainfall. *Journal of Hydrology Engineering*. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001300.

Akbari-Alashti, H., Bozorg-Haddad, O., & Mariño, M. A. (2015). Application of fixed length gene genetic programming (FLGGP) in hydropower reservoir operation. *Water Resources Management, 29*, 3357–3370.

Arunkumar, R., & Jothiprakash, V. (2013). Reservoir evaporation prediction using data-driven techniques. *Journal of Hydrologic Engineering, 18*, 40–49.

Ashofteh, P. S., Bozorg-Haddad, O., & Loáiciga, H. A. (2014). Determination of irrigation allocation policy under climate change by genetic programming. *Journal of Irrigation and Drainage Engineering*. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000807.

Ashofteh, P. S., Bozorg-Haddad, O., & Loáiciga, H. A. (2015). Evaluation of climatic-change impacts on multi-objective reservoir operation with multiobjective genetic programming. *Journal of Water Resources Planning and Management*. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000540.

Ashofteh, P. S., Bozorg-Haddad, O., & Loáiciga, H. A. (2017). Logical genetic programming development for irrigation water supply hedging under climate change conditions. *Irrigation and Drainage, 66*(4), 530–541.

Aytek, A., & Kisi, O. (2008). A genetic programming approach to suspended sediment modelling. *Journal of Hydrology, 351*, 288–298.

Azamathulla, H. M. (2012). Gene-expression programming to predict scour at a bridge abutment. *Journal of Hydro Informatics, 14*, 324–331.

Azamathulla, H. M., & Ahmad, Z. (2012). GP approach for critical submergence of intakes in open channel flows. *Journal of Hydroinformatics, 14*, 937–943.

Azamathulla, H. M., Ghani, A. A., Leow, C. S., & Chang, C. K. (2011). Gene-expression programming for the development of a stage-discharge curve of the Pahang River. *Water Resources Management, 25*, 2901–2916.

Baareh, A. K. (2018). Evolutionary design of a carbon dioxide emission prediction model using genetic programming. *International Journal of Advanced Computer Science and Applications, 9*, 298–303.

Babovic, V., & Keijzer, M. (2002). Rainfall runoff modelling based on genetic programming. *Nordic Hydrology, 33*, 331–346.

Banzhaf, W., Nordin, P., Keller, R., & Francone, F. (1998). *Genetic programming: an introduction on the automatic evolution of computer programs and its application*. San Francisco: Morgan Kaufmann.

Bozorg-Haddad, O., Athari, E., Fallah-Mehdipour, E., & Loáiciga, H. A. (2017). Real-time water allocation policies calculated with bankruptcy games and genetic programming. *Water Science and Technology*. https://doi.org/10.2166/WS.2017.102.

Brameier, M., & Banzhaf, W. (2001). A comparison of linear genetic programming and neural networks in medical data mining. *IEEE Transactions on Evolutionary Computation, 5*, 17–26.

Brutsaert, W. H. (1982). *Evaporation into the atmosphere*. Dordrecht: Springer.

Chen, L. (2003). A study of applying genetic programming to reservoir trophic state evaluation using remote sensor data. *International Journal of Remote Sensing, 24*, 2265–2275.

Cramer, N.L. (1985). A representation for the adaptive generation of simple sequential programs. Proceedings of an international conference on genetic algorithms and their applications, Carnegie-Mellon University, Pittsburgh, USA, 24–26 July.

Danandeh Mehr, A. (2018). An improvement gene expression programming model for streamflow forecasting in intermittent streams. *Journal of Hydrology, 563*, 669–678.

Danandeh Mehr, A., & Nourani, V. (2018). Season algorithm-multigene genetic programming: a new approach for rainfall-runoff modeling. *Water Resources Management, 32*, 2665–2679.

Danandeh Mehr, A., Kahya, A., & Olyaei, E. (2013). Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. *Journal of Hydrology, 505*, 240–249.

Drexler, J. Z., Snyder, R. L., Spano, D., & Paw, K. T. (2004). A review of models and micrometeorological methods used to estimate wetland evapotranspiration. *Hydrological Processes, 18*, 2071–2101.

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

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

Fallah-Mehdipour, E., Bozorg-Haddad, O., & Mariño, M. A. (2013b). Prediction and simulation of monthly groundwater levels by genetic programming. *Journal of Hydro-Environment Research, 7*, 253–260.

Fallah-Mehdipour, E., Bozorg-Haddad, O., & Mariño, M. A. (2013c). Extraction of optimal operation rules in an aquifer–dams system: genetic programming approach. *Journal of Irrigation and Drainage Engineering, 139*. https://doi.org/10.1016/(ASCE)IR.1943-4774.0000628.

Fallah-Mehdipour, E., Bozorg-Haddad, O., Orouji, H., & Mariño, M. A. (2013d). Application of genetic programming in stage hydrograph routing of open channels. *Water Resources Management, 27*, 3261–3272.

Fallah-Mehdipour, E., Bozorg-Haddad, O., & Mariño, M. A. (2014). Genetic programming in groundwater modeling.

Journal of Hydrologic Engineering, 19. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000987.

Fallah-Mehdipour, E., Bozorg-Haddad, O., Orouji, H., & Mariño, M. A. (2016). Application of genetic programming to flow routing in simple and compound channels. *Journal of Irrigation and Drainage Engineering, 142.* https://doi.org/10.1061/(ASCE)IR.1943-4774.0001109.

Ferreira, C. (2001). Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems, 13,* 87–129.

Gandomi, A. H., Alavi, A. H., & Ryan, C. (2015). *Handbook of genetic programming applications*. Basel: Springer.

Garg, V., & Jothiprakash, V. (2009). *Reservoir sedimentation estimation using genetic programming technique.* Kansas City: World Environmental and Water Resources Congress.

Guven, A. (2009). Linear genetic programming for time-series modelling of daily flow rate. *Journal of Earth System Science, 118,* 137–146.

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

Guven, A., Aytek, A., IshakYuce, M., & Aksoy, H. (2008). Genetic programming-based empirical model for daily reference evapotranspiration estimation. *Clean, 36,* 905–912.

Hakimzadeh, H., Nourani, V., & Babaeian-Amini, A. (2014). Genetic programming simulation of dam breach hydrograph and peak outflow discharge. *Journal of Hydrologic Engineering, 19,* 757–768.

Havlicek, V., Hanel, M., Maca, P., Kuraz, M., & Pech, P. (2013). Incorporating basic hydrological concepts into genetic programming for rainfall-runoff forecasting. *Computing, 95,* 363–380.

Hu, H. (2016). Rainstorm flash flood risk assessment using genetic programming: a case study of risk zoning in Beijing. *Natural Hazards, 83,* 485–500.

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

Kasiviswanathan, K. S., Saravanan, S., Balamurugan, M., & Saravanan, K. (2016). Genetic programming based monthly groundwater level forecast models with uncertainty quantification. *Modeling Earth Systems and Environment, 2,* 1–11. https://doi.org/10.1007/s40808-016-0083-0.

Khu, S. T., Liong, S.-Y., Babovic, W., Madsen, H., & Muttil, N. (2001). Genetic programming and its application in real-time runoff forecasting. *Journal of the American Water Resources Association, 37,* 439–451.

Kisi, O., & Guven, A. (2010). Evapotranspiration modeling using linear genetic programming technique. *Journal of Irrigation and Drainage Engineering, 136,* 715–723.

Kisi, O., Hossein-Zadehdalir, A., Cimen, M., & Shiri, J. (2012). Suspended sediment modeling using genetic programming and soft computing techniques. *Journal of Hydrology, 450–451,* 48–58.

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

Koza, J. R. (1994). *Genetic programming II: automatic discovery of reusable programs*. Cambridge: MIT Press.

Lerma, N., Paredes-Arquiola, J., Andreu, J., & Solera, A. (2013). Development of operating rules for a complex multi-reservoir system by coupling genetic algorithms and network optimization. *Hydrological Sciences Journal, 58,* 797–812.

Liong, S. Y., Gautam, T. R., Khu, S. T., Babovic, V., Keijzer, M., & Muttil, N. (2002). Genetic programming: a new paradigm in rainfall runoff modeling. *Journal of the American Water Resources Association, 28,* 705–718.

Liu, S., & Shi, H. (2019). A recursive approach to long-term prediction of monthly precipitation using genetic programming. *Water Resources Management, 33,* 1103–1121.

Maheswaran, R., & Khosa, R. (2013). *Multi resolution genetic programming approach for stream flow forecasting.* International Conference on Swarm, Evolutionary, and Memetic Computing, 19-21 December, Bhubaneswar, India.

Makkeasorn, A., Chang, N. B., & Zhou, X. (2008). Short-term streamflow forecasting with global climate change implications: a comparative study between genetic programming and neural network models. *Journal of Hydrology, 352,* 336–354.

Mirzaei-Nodoushan, F., Bozorg-Haddad, O., Fallah-Mehdipour, E., & Loáiciga, H. A. (2016). Application of data mining tools for long-term quantitative and qualitative prediction of streamflow. *Journal of Irrigation and Drainnage Engineering, 142.* https://doi.org/10.1061/(ASCE)WR.1943-4774.0001096.

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

Nordin, P. (1997). Evolutionary program induction of binary machine code and its applications. Ph.D. dissertation, Department of Computer Science, University of Dortmund, Dortmund, Germany. 1997.

Nourani, V., Komasi, M., & Alami, M. T. (2011). Hybrid wavelet-genetic programming approach to optimize ANN modeling of rainfall-runoff process. *Journal of Hydrologic Engineering, 17,* 724–741.

Orouji, H., Bozorg-Haddad, O., Fallah-Mehdipour, E., & Mariño, M. A. (2013). Modeling of water quality parameters using data-driven models. *Journal of Environmental Engineering, 139,* 947–957.

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

Osman, M.E. & Badr, N.B.E. (2010). Lake Edku water-quality monitoring, analysis and management model for optimizing drainage water treatment using a genetic algorithm. *International Journal of Environmental and Waste Management, 5,* 152–162.

Parasuraman, K., Elshorbagy, A., Carey, K., & S. (2007). Modelling the dynamics of the evapotranspiration process using genetic programming. *Journal of Hydrological Science, 52,* 563–578.

Pelletier, P. J., Chapra, S. C., & Tao, K. (2006). A framework for modelling water quality in streams and rivers using genetic algorithm for calibration. *Environmental Modelling and Software, 21,* 419–465.

Prakash, O., & Datta, B. (2014). Multiobjective monitoring network design for efficient identification of unknown groundwater pollution sources incorporating genetic programming-

based monitoring. *Journal of Hydrologic Engineering*. https://doi.org/10.1061/(ASCE)(HE).1943-5584.0000952.

Puente, C., Olague, G., Trabucchi, M., Arjona-Villicana, P. D., & Soubervielle-Montalvo, C. (2019). Synthesis of vegetation indices using genetic programming for soil erosion estimation. *Remote Sensing, 11*, 156–180.

Rabunal, J. R., Puertas, J., Suarez, J., & Riverto, D. (2007). Determination of the unit hydrograph of a typical urban basin genetic programming and artificial neural networks. *Hydrological Processes, 21*, 476–485.

Sarzaeim, P., Bozorg-Haddad, O., Bozorgi, A., Loáiciga, A., & H. (2017). Runoff projection under climate change conditions with data-mining methods. *Journal of Irrigation and Drainage Engineering*. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001205.

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

Shabani, S., Candelieri, A., Archetti, F., & Naser, B. (2018). Gene expression programming coupled with unsupervised learning: a two-stage learning process in multi-scale short-term water demand forecasts. *Water, 10*, 142–155.

Shiri, J., & Kisi, O. (2011). Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. *Computer and Geoscience, 37*, 1692–1701.

Sivapragasam, C., Maheswaran, R., & Venkatesh, V. (2008). Genetic programming approach for flood routing in natural channels. *Hydrological Processes, 22*, 623–628.

Sivapragasam, C., Muttil, N., Jeselia, M.C., & Visweshwaran, S. (2015). Infilling of rainfall information using genetic programming. *Aquatic Procedia, 4*, 1016–1022.

Soucha, C., Wolfe, C. P., & Grimmond, C. S. B. (1996). Wetland evaporation and energy partitioning: Indiana Dunes National Lakeshore. *Journal of Hydrology, 184*, 189–208.

Sreekanth, J., & Datta, B. (2010). Multi-objective management of saltwater intrusion in coastal aquifers using genetic programming and modular neural network based surrogate models. *Journal of Hydrology, 393*, 245–256.

Sreekanth, J., & Datta, B. (2011). Coupled simulation-optimization model for coastal aquifer management using genetic programming-based ensemble surrogate models and multiple realization optimization. *Water Resources Research*. https://doi.org/10.1029/2010WR009683.

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

Wang, W.-C., Chau, K.-W., & Cheng, C.-T. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology, 374*, 294–306.

Whigham, P. A., & Crapper, P. F. (2001). Modelling rainfall-runoff relationships using genetic programming. *Mathematical and Computer Modeling, 33*, 707–721.

Wu, Z.Y., & Yan, X. (2010). Applying genetic programming approaches to short-term water demand forecast for district water system. 12th Annual Conference on Water Distribution Systems Analysis, Tucson, United States of America, September 12–15.

Xu, Q., Chenn, Q., & Li, W. (2011). Application of genetic programming to modeling pipe failure in water distribution systems. *Journal of Hydroinformatics, 13*, 419–428.