Journal of Hydrology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

A stepwise model to predict monthly streamflow

Anas Mahmood Al-Juboori [a,b], Aytac Guven [a,*]

[a] Civil Engineering Department, Gaziantep University, 27310 Gaziantep, Turkey
[b] Dams and Water Resources Research Center, Mosul University, Iraq

ARTICLE INFO

Article history:
Received 13 April 2016
Received in revised form 28 September 2016
Accepted 2 October 2016
Available online xxxx
This manuscript was handled by G. Syme, Editor-in-Chief, with the assistance of Hazi Azamathulla, Associate Editor

Keywords:
Monthly streamflow
Gene Expression Programming
Generalized Reduced Gradient Optimization
Markovian model
ARIMA

ABSTRACT

In this study, a stepwise model empowered with genetic programming is developed to predict the monthly flows of Hurman River in Turkey and Diyalah and Lesser Zab Rivers in Iraq. The model divides the monthly flow data to twelve intervals representing the number of months in a year. The flow of a month, t is considered as a function of the antecedent month's flow (t − 1) and it is predicted by multiplying the antecedent monthly flow by a constant value called K. The optimum value of K is obtained by a stepwise procedure which employs Gene Expression Programming (GEP) and Nonlinear Generalized Reduced Gradient Optimization (NGRGO) as alternative to traditional nonlinear regression technique. The degree of determination and root mean squared error are used to evaluate the performance of the proposed models. The results of the proposed model are compared with the conventional Markovian and Auto Regressive Integrated Moving Average (ARIMA) models based on observed monthly flow data. The comparison results based on five different statistic measures show that the proposed stepwise model performed better than Markovian model and ARIMA model. The R^2 values of the proposed model range between 0.81 and 0.92 for the three rivers in this study.

1. Introduction

Monthly streamflow prediction is an important issue in water resources management, reservoir operation, hydropower projects, water supply, etc. Many methodologies have been developed to improve monthly flow forecasting according to the past measurements. There is no single method that can perform well for all basins, therefore, for a given watershed; there are different techniques that model the different physical behavior of the watershed. In recent decades, artificial intelligence (AI) techniques have been widely used in modeling hydrological phenomena. A number of researches have been developed in order to find the accurate and applicable models (Yilmaz et al., 2011; Huo et al., 2012; Meshgi et al., 2015; Kisi and Parmar, 2016).

Gene Expression Programming (GEP) became popular among the AI techniques in various fields of water resources and geoscience. GEP is a symbolic regression algorithm to form mathematical functions alternative to traditional nonlinear regression techniques and autoregressive models (Guven, 2009; Guven and Talu, 2010; Traore and Guven, 2013; Karimi et al., 2015). GEP algorithm is an extension to the genetic programming (GP) that was invented by Ferreira (2001). The basic difference between GEP and GP is represented by computer programming. GP programs (individuals) are non-linear entities of different sizes and shapes (parse trees); and in GEP the programs are also non-linear entities of different sizes and shapes (expression trees), but these complex entities are encoded as simple strings of fixed length chromosomes (Ferreira, 2001, 2006). The form of GEP function is not fixed unlike the traditional linear and non-linear regression. GEP uses a genetic evolution algorithm to fit the data to obtain an optimum form of a mathematical function (Fernando et al., 2012).

The resultant GEP program (solution) for the corresponding problem is automatically generated by coding the expression as a tree structure with nodes (function) and leaves (terminal). A fitness function is used to evaluate the generated candidates to reproduce with modification, leaving progeny with new traits. The candidates of this new generation are, in their turn, subjected to the same developmental process: expression of the genomes, confrontation of the selection environment, and reproduction with modification. The process is repeated for a certain number of generations or until a solution has been found (Ferreira, 2001). The GEP code is very simple. The relation between the symbols of the nodes and chromosome is represented in the trees in one to one relation. GEP genes are composed of a head and a tail. The head contains symbols that represent both functions (+,−,*,/,power,x^2, etc.) and terminals (inputs or constants), whereas the tail contains

* Corresponding author.
E-mail address: aguven@gantep.edu.tr (A. Guven).

http://dx.doi.org/10.1016/j.jhydrol.2016.10.006
0022-1694/© 2016 Published by Elsevier B.V.

only terminals. Tree expression is translated to Karva language by reading the tree from left to right in the top line and from top to bottom (Ferreira, 2001). For example, consider the following algebraic expression (a − b) + (c/d), this can be translated to the K-expression as (+ − /, abcd) or expression tree diagram in Fig. 1.

GEP and other AI techniques were successfully applied in hydrologic engineering (Savic et al., 1999; Lopes and Weinert, 2004; Guven, 2009; Guven and Talu, 2010; Azamathulla et al., 2011; Guven and Kisi, 2011; Fernando et al., 2012; Seckin and Guven, 2012; Kisi et al., 2013; Traore and Guven, 2013; Terzi and Ergin, 2014). More recently, Aytek et al. (2014) predicted the monthly water level of Van Lake, Turkey by using GEP. Tofiq and Guven (2014) coupled LGP and statistical downscaling to predict the peak monthly discharges and also the impact of the global warming and climate change on estimating flood discharge by considering different scenarios. Hashmi and Shamseldin (2014) developed a parametric scheme of flow duration curve by using GEP to relate the flow duration curve characteristics to watershed characteristics. Zorn and Shameldin (2015) used GEP to predict the peak flood for the Auckland region of New Zealand.

Shoaib et al. (2015) utilized GEP and hybrid-wavelet-GEP for runoff forecasting. Most recently, Al-Juboori and Guven (2016) applied GEP in an integrated hydrological model for hydropower plant site assessment.

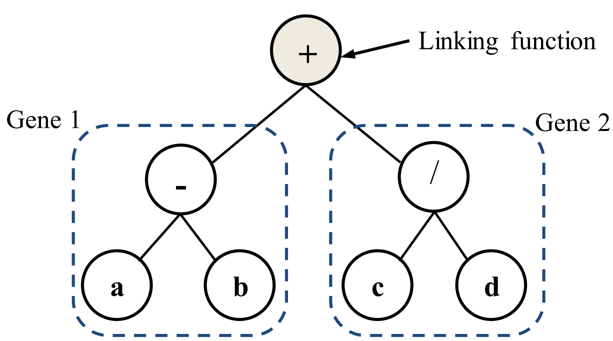

Fig. 1. Expression tree diagram for K-expression (+ − /, abcd).

The objective of this study is to propose an alternative model for monthly streamflow prediction. By this, we aim to present a stepwise model which couples GEP and NGRGO alternative to the traditional nonlinear regression. The results of the proposed model are compared to the conventional Markovian and ARIMA models, and the comparison results are illustrated as scatter plots and tables.

2. Study area and data collection

Three rivers are selected to evaluate the performance of the proposed model. Hurman River, one of major Ceyhan River tributaries in Turkey, Lesser Zab River and Diyalah River, two of major Tigris River tributaries in Iraq. Diyalah River has larger basin area in comparison with the Hurman and Lesser Zab. The Maximum recorded monthly flow is 38.5, 3891 and 1762 m^3/s for Hurman River, Lesser Zab and Diyalah River respectively. The basin areas with monthly flow time series characteristics for the three rivers are summarized in Table 1.

3. Model development

3.1. Model-1: Stepwise model

In this section, a stepwise model which couples GEP and NGRGO methods is developed to predict the monthly flow of permanently flowing rivers. The monthly flow data is divided to twelve intervals representing the number of months in the year (see Fig. 2). The proposed model considers monthly flow of a month t, Q_t, to be estimated as the product of a constant called K and flow of the antecedent month, Q_{t-1} as given in Eq. (1).

$$Q_t = KQ_{t-1} \tag{1}$$

where t denotes the sequence of month in the year. In this model, K is considered to be an optimal constant obtained by applying NGRGO method on the observed monthly flow and those predicted by GEP technique as the product of K as a model constant for each month and the antecedent month's flow (Q_{t-1}). The value of K is estimated by using the following procedure:

Table 1
Monthly flows time series characteristics.

River name	Basin area (km^2)	Time series	Monthly flow records length	Max flow (m^3/s)	Min flow (m^3/s)
Hurman	940	1963–2012	600	38.5	2.45
Lesser Zab	22,250	1978–2011	408	3891	62
Diyalah	31,896	1978–2011	408	1762	60

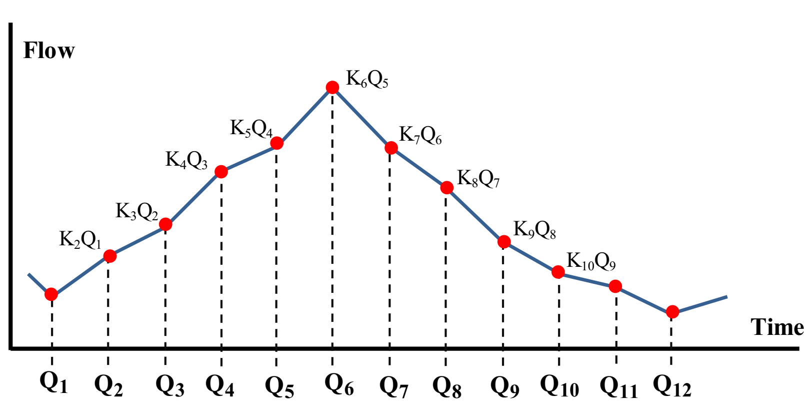

Fig. 2. Schematic representation of the proposed model for monthly flow prediction.

Please cite this article in press as: Mahmood Al-Juboori, A., Guven, A. A stepwise model to predict monthly streamflow. J. Hydrol. (2016), http://dx.doi.org/10.1016/j.jhydrol.2016.10.006