Journal of Hydrology 505 (2013) 240–249

Contents lists available at ScienceDirect

Journal of Hydrology

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

Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique

Ali Danandeh Mehr [a,*], Ercan Kahya [a,1], Ehsan Olyaie [b,2]

[a] Istanbul Technical University, Civil Engineering Department, Hydraulics Division, 34469 Maslak, Istanbul, Turkey
[b] Young Researchers & Elites Club, Hamedan Branch, Islamic Azad University, Hamedan, Iran

ARTICLE INFO

Article history:
Received 10 August 2013
Received in revised form 25 September 2013
Accepted 1 October 2013
Available online 10 October 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor

Keywords:
Feed forward neural networks
Wavelet transform
Linear genetic programming
Data pre-processing
Hydrologic models
Streamflow prediction

SUMMARY

Accurate prediction of streamflow is an essential ingredient for both water quantity and quality management. In recent years, artificial intelligence (AI) techniques have been pronounced as a branch of computer science to model wide range of hydrological processes. A number of research works have been still comparing these techniques in order to find more efficient approach in terms of accuracy and applicability. In this study, two AI techniques, including hybrid wavelet-artificial neural network (WANN) and linear genetic programming (LGP) technique have been proposed to forecast monthly streamflow in a particular catchment and then performance of the proposed models were compared with each other in terms of root mean square error (RMSE) and Nash–Sutcliffe efficiency (NSE) measures. In this way, six different monthly streamflow scenarios based on records of two successive gauging stations have been modelled by a common three layer artificial neural network (ANN) method as the primary reference models. Then main time series of input(s) and output records were decomposed into sub-time series components using wavelet transform. In the next step, sub-time series of each model were imposed to ANN to develop WANN models as optimized version of the reference ANN models. The obtained results were compared with those that have been developed by LGP models. Our results showed the higher performance of LGP over WANN in all reference models. An explicit LGP model constructed by only basic arithmetic functions including one month-lagged records of both target and upstream stations revealed the best prediction model for the study catchment.

1. Introduction

Accurate prediction of streamflow is an essential ingredient for both water quantity and quality management. Generally, there are two possible approaches to predict streamflow. The first approach is the process modelling that involves the study of rainfall-runoff processes in order to model the underlying physical laws (Kuchment et al., 1996). The rainfall–runoff process can be influenced by many factors such as weather conditions, land-use and vegetation cover, infiltration, and evapotranspiration. Therefore, it is subject to many simplification assumptions or excessive data requirements about the physics of the catchment.

The second approach to streamflow prediction is the pattern recognition methodology which attempts to recognize streamflow patterns based on their antecedent records. In this approach, thorough understanding of the physical laws is not required and the data requirements are not as extensive as for the process model (Nourani et al., 2011). The logic behind this approach is to find out relevant spatial and temporal features of historical streamflow records and to use these to predict the evolution of prospective flows. As inputs of the models in pattern recognition method are only time-lagged streamflow observations, this approach appears more useful for the catchments with no or sparse rain gauge stations (Besaw et al., 2010).

In recent years, artificial intelligence (AI) techniques such as artificial neural network (ANN) and genetic programming (GP) have been pronounced as a branch of computer science to model wide range of hydrological processes (Whigham and Crapper, 2001; Dolling and Varas, 2002). Following this, comparative studies between different AI techniques have been appeared in the relevant literature and still attempting to find out the most appropriate one (Ghorbani et al., 2010; Nourani et al., 2011; Abrahart et al., 2012). Hence we initially developed a hybrid wavelet-artificial neural network (WANN) model as an optimized ANN technique for monthly streamflow prediction in a particular catchment in this investigation. Then we, as a first time, compared the results of WANN with those of linear genetic programming (LGP)

* Corresponding author. Tel.: +90 553 417 8028.
E-mail addresses: danandeh@itu.edu.tr (A. Danandeh Mehr), kahyae@itu.edu.tr (E. Kahya), e.olyaie@iauh.ac.ir (E. Olyaie).
[1] Tel.: +90 212 2853002.
[2] Tel.: +98 811 4424189.

0022-1694/$ - see front matter
http://dx.doi.org/10.1016/j.jhydrol.2013.10.003

technique. The pattern recognition methodology is adopted as our prediction approach in this study.

ANN is an effective approach to manage large amounts of dynamic, non-linear and noisy data, especially when the underlying physical relationships are not necessary to fully understanding (Nourani et al., 2011). ANNs were widely used in various fields of hydrological predictions and successful results have also been reported in streamflow prediction (Abrahart et al., 2012; Besaw et al., 2010; Can et al., 2012; Dolling and Varas, 2002; Kisi and Cigizoglu, 2007; Nourani et al., 2011).

In the last decade, GP has been pronounced as a new robust method to solve wide range of modelling problems in water resources engineering such as rainfall-runoff modelling (Dorado et al., 2003; Nourani et al., 2012; Whigham and Crapper, 2001), unit hydrograph determination (Rabunal et al., 2007), flood routing (Sivapragasam et al., 2008), and sea level forecasting (Ghorbani et al., 2010). It was observed that a few studies existed in the literatures related to the comparison of the performance of GP and ANN in time series modelling of streamflow. Guven (2009) applied LGP, a variant of GP, and two versions of neural networks for prediction of daily flow of Schuylkill River in the USA and showed that the performance of LGP was moderately better than that of ANN. Wang et al. (2009) developed and compared several AI techniques include ANN, neural-based fuzzy inference system (ANFIS), GP and support vector machine (SVM) for monthly flow forecasting using long-term observations in China. Their results indicated that the best performance can be obtained by ANFIS, GP and SVM, in terms of different evaluation criteria. Londhe and Charhate (2010) used ANN, GP and model trees (MT) to forecast river flow one day in advance at two stations in Narmada catchment of India. The results showed the ANNs and MT techniques performed almost equally well, but GP performed better than its counterparts.

All aforementioned researches show that GP models result in higher accuracy than regular ANN based modelling approaches. The fact behind this is that the ANN models are not very satisfactory in terms of precision when a time series is highly non-stationary and hydrologic process being operated under a large range of time scales (Nourani et al., 2012). To improve the results of the ANN models, input and/or output data pre-processing by wavelet decomposition technique (hybrid wavelet-ANN models) are suggested and successful results have been reported (Kisi, 2008; Labat 2005; Nourani et al., 2009a, 2011). In recent years, this hybrid model was also compared with some other classic models such as multiple linear regressions and regular ANNs. The findings deduced that the hybrid WANN model can be considered as an effective tool in modelling complex hydrological processes (Anctil and Tape, 2004; Adamowski and Sun, 2010; Partal and Kisi, 2007; Rajaee et al., 2010).

To the best of our knowledge, there is no research examining the performance of WANN and LGP models in monthly streamflow prediction. Thus, in this study we initially developed WANN and LGP models to predict monthly streamflow at a particular river and then compared the prediction results with the observations. Following this, based upon two successive gauging stations records we put forward six black-box ANN structures as reference models for monthly streamflow prediction on Çoruh River located in eastern Black Sea region (Turkey). Then we applied wavelet transform to our ANN-based reference models. The ANN component of the models can handle the nonlinearity and non-stationary elements, while the wavelet component can deal with seasonal (cyclic) non-stationary elements of the phenomenon. In the second step, we developed monthly streamflow prediction models based on explicit LGP technique. Ultimately, we discussed the both accuracy and applicability of ANN, WANN, and LGP techniques via the comparison of their performances. Such a comparison has also been done by Wang et al. (2009) among GP and ANN for monthly discharge forecasting but they did not consider wavelet transform in their neural networks.

The remainder of this paper is organized as follow. In Sections 2 and 3 the concepts of wavelet transforms (WT) and GP are briefly reviewed, respectively. In Sections 4 and 5, study area, applied data and selected efficiency criteria are introduced, respectively. Section 6 presents the formulations and application of the proposed models and techniques. In Section 7, the results and model's performances are evaluated, discussed and compared with each other. Concluding remarks are presented in the last section of the paper.

2. Discrete wavelet transform (DWT)

Wavelet transforms (WT) have recently begun to be explored as a tool for the analysis, de-noising and compression of signals (e.g. time series) and images. WT separates a signal into shifted and scaled version of the original (or mother) wavelet. In other words, by a WT, a signal decomposes into multiple levels of details, sub-signals, which provide an interpretation of the time series structure and history in both the time and frequency domains using a few coefficients (Rajaee et al., 2010). Wavelets are wave-like mathematical functions with amplitude that begins at zero, increases, and then decreases back to zero. Unlike the sine waves, they generally tend to be irregular and asymmetric (Ozger, 2010). WT allows the use of long-time intervals for low frequency signals and shorter intervals for high frequency signals and is able to reveal some statistical features of time series like trend and shift that other signal analysis techniques such as Fourier transform might miss. Another advantage of WT is the flexible choice of mother wavelet according to the characteristics of the investigated time series (Adamowski and Sun, 2010). Mallat (1998) can be referred for more information about wavelet functions.

Hydrologic data usually are recorded in discrete time intervals. Hence, the discrete wavelet transform is usually preferred in hydrological time series decomposition (Rathinasamy and Khosa, 2012). The wavelet function (or mother wavelet) in its discrete form can be represented as Mallat, 1998:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-a}{b}\right) \tag{1}$$

where t is time, a is position parameter, and b is scaling (or dilation) factor of the mother wavelet.

In DWT, wavelet coefficients are commonly calculated at every dyadic step, i.e., the operation of WT is carried out at dyadic dilation ($a = 2^m$) and integer translations ($b = 2^m n$); therefore, the dyadic wavelet function can be obtain by Eq. (2) and DWT coefficients, $T_i\,(m,n)$, for a time series such as $f(t)$ can be defined as Eq. (3) with the integers m and n:

$$\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m_t} - n) \tag{2}$$

$$T_i(m,n) = \sum_{i=0}^{N-1}\psi_{i}.(t)f_i(t) = 2^{-m/2}\sum_{i=0}^{N-1}\psi(2^{-m_i} - n)f_i(t) \tag{3}$$

3. Genetic programming

GP is a heuristic evolutionary modelling technique that automatically solves problems without requiring the user to know or specify the form or structure of the solution in advance. At the most abstract level GP is a systematic, domain-independent method for getting computers to solve problems automatically starting from a high-level statement of what needs to be done (Poli et al., 2008). Unlike statistical techniques such as ANN, decision trees and the like, GP is self-parameterizing that builds models without