Journal of Hydrology 542 (2016) 603–614

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq

Zaher Mundher Yaseen [a,*], Othman Jaafar [a], Ravinesh C. Deo [b], Ozgur Kisi [c], Jan Adamowski [d], John Quilty [d], Ahmed El-Shafie [e]

[a] Civil and Structural Engineering Department, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia
[b] School of Agricultural Computational and Environmental Sciences, Institute of Agriculture and Environment (IAg&E), University of Southern Queensland, Springfield, QLD 4300, Australia
[c] Center for Interdisciplinary Research, International Black Sea University, Tbilisi, Georgia
[d] Department of Bioresource Engineering, Faculty of Agricultural and Environmental Science, McGill University, Québec H9X 3V9, Canada
[e] Civil Engineering Department, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 31 July 2016
Received in revised form 9 September 2016
Accepted 12 September 2016
Available online 14 September 2016
This manuscript was handled by A. Bardossy, Editor-in-Chief

Keywords:
Extreme learning machine
Stream-flow forecasting
Support vector regression
Generalized regression neural network
Semi-arid
Iraq

ABSTRACT

Monthly stream-flow forecasting can yield important information for hydrological applications including sustainable design of rural and urban water management systems, optimization of water resource allocations, water use, pricing and water quality assessment, and agriculture and irrigation operations. The motivation for exploring and developing expert predictive models is an ongoing endeavor for hydrological applications. In this study, the potential of a relatively new data-driven method, namely the extreme learning machine (ELM) method, was explored for forecasting monthly stream-flow discharge rates in the Tigris River, Iraq. The ELM algorithm is a single-layer feedforward neural network (SLFNs) which randomly selects the input weights, hidden layer biases and analytically determines the output weights of the SLFNs. Based on the partial autocorrelation functions of historical stream-flow data, a set of five input combinations with lagged stream-flow values are employed to establish the best forecasting model. A comparative investigation is conducted to evaluate the performance of the ELM compared to other data-driven models: support vector regression (SVR) and generalized regression neural network (GRNN). The forecasting metrics defined as the correlation coefficient (r), Nash-Sutcliffe efficiency (E_{NS}), Willmott's Index (WI), root-mean-square error ($RMSE$) and mean absolute error (MAE) computed between the observed and forecasted stream-flow data are employed to assess the ELM model's effectiveness. The results revealed that the ELM model outperformed the SVR and the GRNN models across a number of statistical measures. In quantitative terms, superiority of ELM over SVR and GRNN models was exhibited by E_{NS} = 0.578, 0.378 and 0.144, r = 0.799, 0.761 and 0.468 and WI = 0.853, 0.802 and 0.689, respectively and the ELM model attained lower $RMSE$ value by approximately 21.3% (relative to SVR) and by approximately 44.7% (relative to GRNN). Based on the findings of this study, several recommendations were suggested for further exploration of the ELM model in hydrological forecasting problems.

1. Introduction

Accurate stream-flow modeling and forecasting are important tools for sustainable water resources planning and management. Accurate multiple-scale (e.g., weekly, monthly and seasonal) stream-flow forecasts are important for the efficient operation and planning of reservoirs, sediment transport in rivers, hydro-power generation, irrigation management decisions, scheduling reservoir releases and other hydrological applications (Araghinejad et al., 2006; Danandeh Mehr et al., 2014; Solomatine and Shrestha, 2009). Accurate short-term (real-time) forecasts (e.g., hourly or daily) are important for flood forecasting and developing early warning systems (Chiang et al., 2004; Guven, 2009; Yaseen et al., 2015a). This type of forecasting can be a valuable tool not only for providing advanced warning of an impending flood event to reduce and mitigate the impacts of flooding on infrastructure, property and public health, but can also yield significant information that can be used by hydrologists in the areas of water resources management, water quality assessments, water pricing and implementing sustainable agricultural practices.

* Corresponding author.
E-mail address: zahermundher@gmail.com (Z.M. Yaseen).

http://dx.doi.org/10.1016/j.jhydrol.2016.09.035
0022-1694/

Stream-flow time series forecasting is a challenging task. This is because the dynamics of stream-flow tends to be filled with chaotic disturbances, exhibiting complex non-linear dynamics and randomness phenomena (El-Shafie et al., 2009; Maier and Dandy, 2000; Singh and Sankarasubramanian, 2014). Enhancing the accuracy and reliability of forecasting such (potentially non-stationary) hydrological variables has always been an important research topic for hydrologists. To date, there has been no single universal approach that provides the most appropriate stream-flow forecasting results under all circumstances. This may be due to the fact that natural processes evolve uniquely through time while modeling approaches (which are based on finite-length datasets) are synthetic by construction and are controlled by parametric forms that differ from one modeling approach to the next (e.g., a family of models that may be skillful for predicting stream-flow may not be so useful for predicting urban water use, etc.).

Over the last two decades, Artificial Intelligence (AI) approaches have been broadly applied in the field of hydrological applications (Aziz et al., 2014; Chiang et al., 2011; El-Shafie and El-Manadely, 2010; Fijani et al., 2013; Khatibi et al., 2011; Moosavi et al., 2014, 2013; Rieker and Labadie, 2012; Saldarriaga et al., 2004; Shamim et al., 2015; Shu and Burn, 2004). AI (or statistical) models, which are classified as 'black box' models, are very useful in modeling natural systems. These models do not require complex physical equations and parametric assumptions often required in the case of deterministic 'white box' (or physically-based) models. Due to the simplicity in their design and implementation, including their relative accuracy in forecasting problems, numerous studies have successfully demonstrated their applicability in hydrological modeling/forecasting (Afan et al., 2014; Deo and Şahin, 2016; El-Shafie et al., 2012; Kisi, 2008a, 2011; Makarynska and Makarynskyy, 2008; Nourani et al., 2012; Palani et al., 2008; Salcedo-Sanz et al., 2015; Shu and Burn, 2004; Taormina et al., 2012; Tezel and Buyukyildiz, 2015). In this paper, we focus on the application of AI approaches for stream-flow forecasting in a semi-arid region.

Based on a recent review conducted by Yaseen et al. (2015b), stream-flow forecasting using AI techniques can be divided into four different categories: (i) classification and regression-based data-driven approaches, (ii) fuzzy sets, (iii) evolutionary computation, and (iv) conjunction AI models (e.g., those based on wavelet filter or other hybrid-models). The artificial neural network (ANN) algorithm has been applied in numerous studies for stream-flow modeling (and forecasting) using its supervised and non-supervised capabilities (Abrahart and See, 2000; Allawi and El-Shafie, 2016; Bray and Han, 2004; Cigizoglu, 2005; Danandeh Mehr et al., 2014; Deo and Şahin, 2016; Ghorbani et al., 2016; Hsu et al., 2002). In order to consider the uncertainty in time series modeling, Chang and Chen (2001) proposed the earliest research using fuzzy set neural networks. This study was followed by many investigations of the fuzzy logic approach (El-Shafie et al., 2007; Graves and Pedrycz, 2009; Greco, 2012; Katambara and Ndiritu, 2009; Özger, 2009). Inspired by the Darwinian theory of evolution, the evolutionary-class of optimization algorithms have been used to solve challenging hydrological and water resources optimization problems (Tsoukalas et al., 2016). The main advantage of evolutionary optimization methods when compared to traditional gradient-based optimization algorithms stems from their ability to gradually search for solutions for the examined problem (using evolutionary concepts such as bacterial foraging, particle swarming, and echolocation) instead of giving direct solutions (i.e., through partial derivatives of the model parameters, with respect to the input and output data). Many authors have examined the robustness of evolutionary optimization in stream-flow forecasting and modeling (Chen et al., 2008; Dorado et al., 2003; Guven, 2009; Kisi et al., 2012; Makkeasorn et al., 2008; Ni et al., 2010; Savic et al., 1999; Whigham and Crapper, 2001). Likewise, conjunction AI models (e.g., wavelet-hybrid support vector regression models) have also been applied in areas of stream-flow, drought, global solar radiation and evaporative loss modeling (Deo et al., 2016a, 2016b; Kisi, 2008b, 2011; Guo et al., 2011; Belayneh et al., 2016). While evolutionary optimization algorithms and AI conjunction models are worthwhile research endeavors to explore, they are beyond the scope of this work. In this work we focus on the evaluation of a newer data-driven algorithm (i.e., extreme learning machine) and its comparison with traditional data-driven approaches for monthly stream-flow forecasting with an application to a semi-arid region. However, in future studies, evolutionary optimization algorithms and conjunction models can be adapted for use with the newer data-driven algorithm explored in this work.

Despite the growing applications and usefulness of AI techniques in modeling stream-flow data (and other hydrological time series), the forecasts produced by some of these methods (e.g., ANN) still suffer from several shortcomings (e.g., over-fitting, slow learning speed, and local minima). An emerging data-driven algorithm for single hidden layer feed-forward networks (SLFNs), the extreme learning machine (ELM) model, was proposed by Huang et al. (2006a) and overcomes the disadvantages of the traditional feed-forward backpropagation ANN (FFBP-ANN) (i.e., over-fitting, slow learning speed, and local minima). In the last decade, the ELM algorithm has been applied in a diverse range of applications due to its high-performance and innovative design features (i.e., random generation of the parameters of hidden nodes without the need for iteratively tuning the algorithm, determining the output weights analytically by solving a least squares problem and yielding significantly faster solutions compared to traditional neural network models (e.g., FFBP-ANN)). Some recent applications of the ELM model in diverse fields of research include: the prediction of evapotranspiration (Abdullah et al., 2015), dew point prediction (Mohammadi et al., 2015), fast object recognition and image classification (Samat et al., 2014; Bencherif et al., 2015), land displacement prediction (Lian et al., 2012), sales prediction (Sun et al., 2008), melting point prediction of organic compounds (Bhat et al., 2008), big data classification (Wang et al., 2015), and the use of prior knowledge (Soria-Olivas et al., 2011). In comparison with other AI techniques (e.g., ANN, SVR, fuzzy logic, etc.), the ELM method has important advantages due to its improved (or, at least, comparable) generalization performance and faster learning speed (Deo and Şahin, 2015b, 2016; Deo et al., 2016a). In 2014, the first attempt of applying ELM in stream-flow modeling was conducted by Li and Cheng (2014). They integrated the wavelet decomposition approach with ELM in forecasting monthly river flow in southwestern China. Recently, Deo and Sahin (2016) applied ELM flow forecasting in Queensland to validate its superiority over artificial neural network (ANN) models. The online sequential extreme learning machine (OSELM) approach was investigated in forecasting daily stream-flow as an online warning system in Canada by Lima et al. (2016). Another study conducted more recently used OSELM in forecasting river discharge in Germany (Yadav et al., 2016).

Since ELM is a promising but relatively new approach for stream-flow forecasting, this paper investigates its potential for accurate long-term stream-flow forecasting in a semi-arid environment (the Tigris river in Iraq) and compares its performance against traditional data-driven methods. In the last decade, Tigris river has experienced a negative deterioration in water resources management and sustainability due to climate changes and diplomatic issues in the region. Thus, establishing the current model comes with the important motive of developing an accurate expert system for this river system and other eco-hydrological systems in arid and semi-arid regions. This study was inspired also by the