Research papers

Boost stream-flow forecasting model with extreme learning machine data-driven: A case study in a semi-arid region in Iraq

Zaher Mundher Yaseen, Othman Jaafar, Ravinesh C. Deo, Ozgur Kisi, Jan Adamowski, John Quilty, Ahmed El-shafie

PII: S0022-1694(16)30589-3
DOI: http://dx.doi.org/10.1016/j.jhydrol.2016.09.035
Reference: HYDROL 21531

To appear in: *Journal of Hydrology*

Received Date: 31 July 2016
Revised Date: 9 September 2016
Accepted Date: 12 September 2016

Please cite this article as: Mundher Yaseen, Z., Jaafar, O., Deo, R.C., Kisi, O., Adamowski, J., Quilty, J., El-shafie, A., Boost stream-flow forecasting model with extreme learning machine data-driven: A case study in a semi-arid region in Iraq, *Journal of Hydrology* (2016), doi: http://dx.doi.org/10.1016/j.jhydrol.2016.09.035

Boost stream-flow forecasting model with extreme learning machine data-driven: A case study in a semi-arid region in Iraq

Zaher Mundher Yaseen[1], Othman Jaafar[1], Ravinesh C. Deo[2], Ozgur Kisi[3], Jan Adamowski[4], John Quilty[4], Ahmed El-shafie[5]

[1] Civil and Structural Engineering Department, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia. Corresponding author: zahermundher@gmail.com

[2] School of Agricultural Computational and Environmental Sciences, Institute of Agriculture and Environment (IAg&E), University of Southern Queensland, Springfield, QLD 4300, Australia

[3] Civil Engineering Department, Faculty of Architecture and Engineering, Canik Basari University, 55080, Samsun, Turkey.

[4] Department of Bioresource Engineering, Faculty of Agricultural and Environmental Science, McGill University, Québec H9X 3V9, Canada

[5] Civil Engineering Department, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia

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 partial autocorrelation functions on 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 about 21.3% (relative to SVR) and by about 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.

Keywords: extreme learning machine; stream-flow forecasting; support vector regression; generalized regression neural network; semi-arid; Iraq

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.

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 or generalized 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 using machine learning algorithms 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; Moosavi et al., 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, 2016a; El-Shafie et al., 2012; Kisi, 2008a; Kişi, 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 machine-learning 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, 2016a; 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, echolocation, etc.) 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 SVR models) have also been applied in areas of stream-flow, drought, global solar radiation and evaporative loss modeling (Deo et al., 2016a, b; Kisi 2008b; 2011, Guo et al., 2011). 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. ELM) 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, support vector regression (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; Deo and Şahin, 2016a; Deo et al., 2015b). In 2014, the first attempt of applying ELM in stream-flow modeling was conducted by (Li and Cheng). They integrated the wavelet decomposition approach with ELM in forecasting monthly river flow in southwestern China. Recently, Deo and Sahin (2016) applied ELM model for stream-flow forecasting in Queensland to validate its superiority over artificial neural network (ANN) models. Online sequential extreme learning machine (OSELM) approach investigated in forecasting daily stream-flow as an online warning system in Canada by (Lima et al., 2016). Another study conducted most recently utilizing OSELM in forecasting river discharge in Germany (Yadav et al., 2016). In summary, ELM exhibited a robust and fast soft computing technique in comparison with AI data-driven models.

Since ELM is a promising but a relatively new approach for stream-flow forecasting, this paper investigates its potential for accurate monthly 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 experience 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 growing application of the ELM model in general forecasting problems where it has been shown to outperform traditional data-driven methods such as ANN (and SVR) (Huang et al., 2012), in addition to its limited applications in the field of hydrological forecasting (where ANN and SVR are ubiquitously implemented). The work of Deo and Sahin (2016) compared an ELM (with an ANN) model for stream-flow discharge and water level forecasting problems in Queensland (Australia) and found that the ELM produced more accurate forecasts than ANN with a lower computational expense. Other applications of the ELM model, broadly in the area of climate and hydrology, includes the forecasting of the effective drought index (Deo and Şahin, 2015b; Deo et al., 2015b) and the evaporative losses estimations in Australia (Deo et al., 2015a), downscaling of Global Climate Models (Acharya et al., 2013) in India, solar and wind forecasting problems in Turkey, Spain and the United States of America (Salcedo-Sanz et al., 2015) (Sahin et al., 2014). However, to the best of our knowledge, the ELM model has not yet been tested for stream-flow forecasting in the present study region, so exploring this technique carries significant merit.

In this paper, the potential usefulness of the ELM approach was compared with other data-driven methods including support vector regression (SVR) and the generalized regression neural network (GRNN) models in order to evaluate forecasting accuracy. In order to evaluate the performance of the proposed ELM approach, several quantitative performance indicators (section 3.5) are employed to validate the proposed research. The rest of this paper is organized as follows: section 2 describes the case study and the catchment properties. Section 3 provides a detailed introductory to the methodologies including a description of the data-driven approaches, model configuration and the performance assessment indicators. The application of the ELM, SVR and GRNN models to the study catchment and their forecasting results are discussed in Section 4. Finally, the main findings and concluding remarks are highlighted in the last section.

2. Case Study and Catchment Description

The Tigris River is one of the largest rivers in the Middle East. It spans approximately 1718 km. At the starting point, the Tigris River runs from Turkey toward the south alongside Iraq. About 85% of its basin lies in Iraq, covering a catchment area of approximately 253,000 km. The Tigris River shares its space with the Euphrates River as the main sources of water (e.g., domestic, agriculture, industrial, and other uses) for the major cities of Iraq. The climate in this region is characterized by a semi-arid environment. In the capital city, Baghdad, the average rainfall is estimated to be about 216 mm with seasonal characteristics (December to February). The mean flow in the Tigris river is estimated to be about 235 m^3/s. The maximum temperature is approximately 45 °C during the summer; the temperature drops to about 10 °C minimum in the winters (Al-Ansari, 2013; Salman et al., 2014). In this study, the monthly stream-flow time series data for the Baghdad station along the Tigris River in Iraq region were used. This data base was obtained from the USGS Data Series 540. The drainage area of this site is 134,000 km^2, which is located between (33° 24' 34") N Latitude and (44° 20' 32") E Longitude, as displayed in Fig.1.

Fig. 1

3. Theoretical Overview

3.1 Extreme Learning Machine

In this study, an extreme learning machine model (Figure 2) was developed that used a set of training data samples denoted as $\{(x_1, y_1),...,(x_t, y_t)\}$ with x_t as the explanatory variable and y_t as the response variable. The input vector (x_1, x_2… x_t) incorporated the predictor variable defined by the lagged combinations of historical monthly stream-flow from which the data patterns and attributes were extracted and the vector (y_1, y_2… y_t) was the response variable (i.e. the observed values of stream-flow used to validate the stream-flow forecasts). For the data points defined by $t = 1, 2, …, N$ that contains a set of N- training data samples, $x_t \in \Re^d$ and $y_t \in \Re$, the single layer feed-forward neural network (SLFN) with Λ hidden nodes can be mathematically expressed as (Huang et al., 2006a; Şahin et al., 2014):

$$\sum_{i=1}^{\Lambda} B_i g_i(\alpha_i . x_t + \beta_i) = z_t \tag{1}$$

where, $B \in \Re^{\Lambda}$ represents the weights that are to be estimated in the output layer between the nodes of the hidden layer (Λ nodes in total) and the ELM model's output Z ($z_t \in \Re$), G (α, β, x) is the hidden layer activation function, $\alpha_i \in \Re^d$ and $\beta_i \in \Re$ are the weights and biases in the ELM algorithm's randomised layers, respectively, i represents the index of a particular hidden neuron and d is the number of input neurons (Figure 2). The value of d is determined by the number of input time series that is used for forecasting the stream-flow values.

Fig. 2

For the present study, the number of input neurons was denoted as a value from $d = 1$ to 5 (i.e. the set of input combinations) and a logarithmic sigmoid activation function G (x) following previous studies (e.g. (Deo et al., 2016; Deo and Sahin 2015)) was applied:

$$G(x) = \frac{1}{1+\exp(-x)} \tag{2}$$

In the output layer where the forecasted values of stream-flow were generated, a linear transfer function was adopted as this is a common practice in hydrological time-series forecasting (Deo and Şahin, 2015a; Deo and Şahin, 2015b; Yonaba et al., 2010).

As described by Huang et al. (2006), Eq. (2) can be used to approximate a set of N training set samples via:

$$\sum_{t=1}^{N} \| z_t - y_t \| = 0 \tag{3}$$

The basic premise of the ELM modeling framework is that, in accordance with Eq. (3), the ELM model with a suitable number of hidden neurons and randomized input layer weights and hidden neurons biases (α and β) can yield a zero error, leading to the realization that the network's output weights (B) can be determined analytically for a training sample. It should be noted that the magnitudes of α and β are sampled from (one of many) continuous probability distributions (such as the uniform, triangular or normal distributions, for example).

Concordant with Eq. (3), one may estimate the values of B directly from the N input-output set of data samples with a system of linear equations (Huang et al., 2006b):

$$Y = G\mathrm{B} \tag{4}$$

such that:

$$G(\alpha,\beta,x)=\begin{bmatrix} g(x_1) \\ \vdots \\ g(x_N) \end{bmatrix}=\begin{bmatrix} g_1(\alpha_{1.}x_1+\beta_1) & \cdots & g_L(w_{\Lambda.}x_1+\beta_\Lambda) \\ \vdots & \cdots & \vdots \\ g_1(\alpha_{N.}x_N+\beta_1) & & g_L(w_{\Lambda.}x_N+\beta_H) \end{bmatrix}_{N\times\Lambda} \tag{5}$$

and

$$\mathrm{B}=\begin{bmatrix} \mathrm{B}_1^T \\ \vdots \\ \mathrm{B}_\Lambda^T \end{bmatrix}_{\Lambda\times 1} \text{ and } Y=\begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N\times 1} \tag{6}$$

Here, G is also known as the hidden layer output matrix (and T represents the transpose of the matrix/vector).

The output weights of the ELM network can be deduced by inverting the hidden layer matrix using the Moore-Penrose generalized inverse function (+) (Huang et al., 2006a):

$$\hat{\mathrm{B}}=G^{+}Y \tag{7}$$

where $\hat{\mathrm{B}}$ represents the estimated output weights from the N data samples used in the modelling process.

Finally, the forecasted values of the monthly stream-flow data, $\hat{y}$ can be obtained by the testing input vector (x_{test}) (outside of the training dataset) (Akusok et al., 2015):

$$\hat{y}=\sum_{i=1}^{\Lambda}\hat{B}_i\, g_i(\alpha_i.x_{test}+\beta_i) \tag{8}$$

For a more detailed evaluation of the ELM algorithm, the readers can refer to the seminal works of (Huang et al., 2012; Huang et al., 2015; Huang et al., 2006c).

3.2 Support Vector Regression

Compared to the earlier developed ANN models (McCulloch and Pitts, 1943), support vector regression (SVR) is a relatively new and improved data-driven model that is based on statistical learning theory (Vapnik, 1995). It was initially proposed to solve pattern recognition and classifications problems and was later adapted to solve regression problems.

To describe the SVR model, let us assume that we have a set of *N*-training data, $t = 1, 2, \ldots, N$, $x_t \in \Re^d$ and $y_t \in \Re$. The regression function of the SVR (Vapnik, 1995) can be expressed as:

$$f(x) = w^T \varphi(x) + b \tag{9}$$

where w is a weight vector, b is the bias and φ presents a non-linear transfer function. Here, the objective of the transfer function is to nonlinearly map the input variables to a high-dimensional feature space. The convex optimization formula with an e-insensitivity loss function is presented as follows:

$$\min \varphi(w, \xi) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{l} \xi_i\right) \tag{10}$$

subject to the following constraints:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \qquad \xi \geq 0, \qquad (i = 1, \ldots, N) \tag{11}$$

where ξ is a slack variable that penalizes training error by the loss function for the chosen error tolerance. C is a positive regularization parameter that shrinks the weight parameters while minimizing the empirical error in the optimization problem (see Fig. 3).

The SVR optimization problem is usually solved in its dual form using Lagrangian multipliers. The current study is conducted using Sequential Minimal Optimization (SMO) that was introduced by

(Platt, 1999). In this study, the main reason for employing the SMO algorithm in forecasting stream-flow over other possible algorithms is that SMO can provide an analytical solution for a subset that can be achieved without invoking a quadratic optimizer (He et al., 2014). This reduces the need for an expensive, third-party quadratic programming solver. For those interested in other optimization algorithms for SVR with applications in hydrology, one may refer to Raghavendra. N and Deka (2014), where several algorithms were proposed to solve the dual optimization problem.

The flexibility of the SVR model is reflected in the use of the kernel function that nonlinearly maps the training input data to a higher, and possibly an infinite dimensional feature space (Tripathi et al., 2006). In this paper, the Radial Basis Function (RBF) was employed as the kernel function, which was also utilized in a number of previous studies and is, in general, popular in hydrology (Deo et al., 2016a; Deo et al., 2016b). The RBF kernel width, regularization, and slackness parameters were solved through a grid-search.

Fig. 3

3.3 Generalized Regression Neural Networks

In order to further evaluate the ELM model for stream-flow forecasting, the GRNN model, which is a variation of RBF-based neural networks (similar to kernel regression), was used in this study (Cigizoglu and Alp, 2006). The main difference between the GRNN and the traditional FFBP-ANN is that the network's architecture (described in more detail below) is fixed for a given input-output dataset while the FFBP-ANN requires the determination of an optimal number of hidden layers and hidden nodes. In GRNN only a single parameter needs to be optimized, the RBF kernel spread parameter, σ, which is used to decipher the similarity between input parameters. A spread that it too large induces over-smoothing and will typically cause the majority of input patterns to appear similar while a spread that is too small will not provide a smooth regression surface, thus an intermediate value of the smoothing parameter should be sought to provide an optimal amount of smoothing so that the GRNN generalizes well to out-of-sample inputs.

The GRNN model is able to approximate any arbitrary continuous function mapping for a given input-output dataset, and draws the function estimates directly from the training input data (Kisi, 2006). As an additional advantage to the present forecasting problem, the GRNN model can generate consistent forecasts such that when the training data set size becomes large, the estimation error approaches zero, with only mild restrictions on the function (Cigizoglu and Alp, 2006). Based on these features, the GRNN was preferred over the back-propagation neural network approach (e.g. ANN model). The GRNN model consists of four layers that are designated as the input layer, pattern recognition layer, summation layer and the output layer as shown in Fig. 4.

Fig. 4

In the first layer, the input units pass the input variables provided to the network to the pattern layer through the input weights (which are equal to the input variables). The second layer, the "pattern layer", is connected to the first layer via the input weights and in this layer the similarity between input patterns is calculated using a distance function, formula (13). The third (summation) layer in the GRNN architecture computes the prediction for a given input data vector as the weighted sum of the outputs aligned with those input patterns that are closest to a given input, formula (12). The final (output) layer receives the result from the summation layer and represents the network prediction, or stream-flow forecast (since the transfer function in the layer is linear). According to the literature, the GRNN with RBF neurons can provide similar or better performance when compared to the other ANN-based algorithms and has the advantage that the network architecture is fixed and only a single parameter needs to be optimized (Yaseen et al., 2015b). Considering this, the RBF equation has been used to compute the distance metric between input patterns (Deo and Sahin, 2016; Deo et al., 2016).

To achieve the forecasted value $\hat{y}$ for an unknown input vector x_{test} from a training dataset defined by $t = 1, 2, \ldots, N$ that contains a set of N- training data samples, $x_t \in \Re^d$ and $y_t \in \Re$ it follows that (Firat, 2008):

$$\hat{y} = \frac{\sum_{i=1}^{N} y_i . \exp[-D(x_{\text{test}}, x_i)]}{\sum_{i=1}^{N} \exp[-D(x_{\text{test}}, x_i)]} \tag{12}$$

$$D(x, x_i) = \sum_{k=1}^{d} \left(\frac{x_i - x_{ik}}{\sigma}\right)^2 \tag{13}$$

where D denotes the distance function for the RBF kernel (Firat, 2008). In this study, we used the trial and error procedure to optimize the RBF spread parameter (Firat et al., 2010; Wang and Sheng, 2010).

3.4 Model Development

The stream-flow data spanned a period of 20 years (1991-2010). Table 1 displayed the statistical characteristics including training, testing and the complete data span. In order to partition the data into the model development (training) and model evaluation (testing) parts, the full set was split into 16 years (80% of the set) for the training phase and the remaining 4 years (20% of the set) was used for the testing phase.

Table 1.

As a prior and significant step in developing reasonable forecasting models using data-driven techniques, the selection of proper input variables are required to be supplied to the different models (i.e. ELM, SVR, and GRNN) (Graves and Pedrycz, 2009; Kisi, 2007; Kisi, 2008b; Kisi et al., 2012; Maier and Dandy, 2000; Nourani et al., 2012; Pramanik et al., 2010). In the literature, there are two common methods for input selection: (i) determining the number of sequential time series lagged data values that can provide the best forecast performance using the trial and error procedure or (ii) the most correlated lagged variable(s) that can be determined using the auto-correlation function (ACF) and the partial auto-correlation function (PACF) statistical methods. In this study, the input combinations were selected based on ACF and PACF, as done in other hydrological forecasting studies (Maier and Dandy, 2000; Maier et al., 2014; Sang, 2013; Yaseen et al., 2015b; Zhang et al., 1998).

To develop an accurate forecasting model, we analysed the patterns in the observed stream-flow data for the training set by using the correlation statistics via autocorrelation and partial autocorrelation functions to identify suitable predictors (Sudheer et al., 2002; Tiwari and Adamowski, 2013; Tiwari and Chatterjee, 2011). The statistical approach employed time-lagged information from Q time-series to analyze the monthly time periods between the current Q and the Q value at some specific point in the past (i.e. a time lag) to assess any temporal dependencies existing in the time-series. Subsequently, the optimum inputs for each (monthly) time lag were identified by statistically analyzing lagged combinations and the respective correlation coefficient (r).

Fig. 5

Fig. 5 shows auto-correlation and partial auto-correlation functions for monthly stream-flow data. It was evident that the PACF was useful for identifying the model inputs as it removed the dependence on intermediate elements (those within lags interpreted as a regression of time-series against its past lagged value) and identified the extent to which current stream-flow is correlated to past months. It was evident that the original signal (Q (t)) and the 1-month lagged signals Q (t – 1) where highly correlated (with $r \approx 0.8$). Furthermore, the PACF also exhibited statistically significant correlations for several monthly lags and most importantly, at five inputs at lags of 1, 2, 3, 4 and 5 months that were used for forecasting the stream-flow Q (t). This procedure aimed to develop a model that utilized memory (i.e. time-lagged or past values of stream-flow) to forecast the present monthly value. It should be noted that, one-month-ahead forecasting for stream-flow at Baghdad station is crucial for the decision-maker for water resources in order to have a vision for managing the available water for different users. For instance, irrigation schedule (re-adjusting the schedule for existing crops or tuning for the planning crop pattern), industry productions (giving proper decision on the amount of productions relying on the amount of water availability).

Subsequently, a set of five input sets were designed with lagged t of up to 5 months for one month lead time forecasting of stream-flow discharge using ELM, SVR, and GRNN methods, via:

$$\begin{aligned}
&M1: \quad Q(t) = f\left[Q(t-1)\right] \\
&M2: \quad Q(t) = f\left[Q(t-1), Q(t-2)\right] \\
&M3: \quad Q(t) = f\left[\, Q(t-1), Q(t-2), Q(t-3)\right] \\
&M4: \quad Q(t) = f\left[Q(t-1), Q(t-2), Q(t-3), Q(t-4)\right] \\
&M5: \quad Q(t) = f\left[Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5)\right]
\end{aligned} \tag{14}$$

where Q (t) = the target output or forecasted stream-flow value, Q (t – 1) = the first input (1-month lagged Q), Q (t – 2) = second input (2-month lagged Q), Q (t – 3) = the third input (3-month lagged Q), Q (t – 4) = the fourth input (4-month lagged Q), and Q (t – 5) = the fifth input (5-month lagged Q), and f represents the model type (i.e. ELM, SVR, or GRNN).

As a requirement in a problem with data-driven models, the scaling of the input variable(s) was undertaken to avoid data patterns and attributes with large numerical ranges dominating the role of the smaller numerical ranges via:

$$Q_{i,normalized} = \frac{Q_i - Q_{\min}}{Q_{\max} - Q_{\min}} \tag{15}$$

where $Q_{i,normalized}$ is the normalized value (between 0 and 1) for mean monthly stream-flow; Q_i, Q_{min} and Q_{max} are the current stream-flow value that is to be normalized and minimum and maximum stream-flows within the historical dataset.

In order to develop the optimal ELM model, the best neural network architecture (with the optimum number of hidden neurons) was identified to create an appropriate ELM structure for stream-flow forecasting, following earlier studies (Deo et al., 2016; Deo and Sahin 2015). Using the respective training datasets for different input combinations (of lagged stream-flows) defined in Eq. (14), the ELM model architecture that performed the best in the training period was determined as the optimal model. To identify the optimal ELM model, we followed the approach in Deo et al. (2016b) where the ELM network architecture was randomly executed 1.0×10^3 times with $1{:}2N{+}1$ hidden neurons (in increments of 1) to investigate how the randomized hidden layer weights and biases varied the model's output in the *N*-data training period. This process was performed to acquire the smallest value of the root mean square error (*RMSE*) between the network output (forecast) and the observed stream-flow in the train/target dataset. Consequently, the optimum ELM model which had a single hidden layer network (with an input, hidden and output layer), resulted in about 100 randomizations that were examined for a stable solution of the forecasted stream-flow values. For each model run, the modeling time was also recorded to examine the computational efficiency.

The SVR model was optimized using the SMO algorithm as mentioned in section 3.2. This is due to the features of the SMO algorithm including good computational speed and simple implementation. The theoretical detailed procedure of this optimization algorithm can be found in (Platt, 1999; Takahashi et al., 2008; Yang et al., 2007). The positive regularization C parameter and kernel function parameter γ were determined via the grid-search algorithm (He et al., 2014; Hsu et al., 2003). The GRNN model was developed using NeuroSolutions software (Principé and Lefebvre, 1998). NeuroSolutions is a tool that provide the ability to solve temporal problems by extracting how the time series pattern changes with time. RBF kernel spread parameter σ was determined for the GRNN model in accordance to the minimum *RMSE*.

3.5 Model Performance Indicators

To evaluate the performance of the three modeling approaches (Legates and McCabe, 1999a), the following statistical score metrics were used.

I. Correlation coefficient (r) expressed as:

$$r = \left(\frac{\sum_{i=1}^{N}\left(Q_{obs,i} - \overline{Q}_{obs,i}\right)\left(Q_{for,i} - \overline{Q}_{for,i}\right)}{\sqrt{\sum_{i=1}^{N}\left(Q_{obs,i} - \overline{Q}_{obs,i}\right)^2}\sqrt{\sum_{i=1}^{N}\left(Q_{for,i} - \overline{Q}_{for,i}\right)^2}} \right) \quad (16)$$

II. Willmott's Index of agreement (WI) (Willmott et al., 2012) expressed as:

$$WI = \left| 1 - \left[\frac{\sum_{i=1}^{N}\left(Q_{obs,i} - Q_{for,i}\right)^2}{\sum_{i=1}^{N}\left(\left|Q_{for,i} - \overline{Q}_{obs,i}\right| + \left|Q_{obs,i} - \overline{Q}_{obs,i}\right|\right)^2} \right] \right|, \; 0 \le WI \le 1 \quad (17)$$

III. Nash–Sutcliffe coefficient (E_{NS}), expressed as:

$$E_{NS} = \left| 1 - \left[\frac{\sum_{i=1}^{N}\left(Q_{obs,i} - Q_{for,i}\right)^2}{\sum_{i=1}^{N}\left(Q_{obs,i} - \overline{Q}_{obs,i}\right)^2} \right] \right|, \; 0 \le E_{NS} \le 1 \quad (18)$$

IV. Root mean square error ($RMSE$) expressed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Q_{for,i} - Q_{obs,i}\right)^2} \quad (19)$$

V. Mean absolute error (MAE) expressed as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\left(Q_{for,i} - Q_{obs,i}\right)\right| \quad (20)$$

where Q_{obs} and Q_{for} are the observed and forecasted i^{th} value of the stream-flow, $\bar{Q}_{obs}$ and $\bar{Q}_{for}$ are the average of observed and forecasted Q.

4. Results and Discussion

A direct comparison of the ELM with SVR and GRNN models is made in Table 2, which shows the models' performance metrics computed by Eq. (16-20) for each of the investigated models (i.e. with different time lagged stream-flow observations as model inputs) within the testing period phase. Here, we have used various performance metrics to assess the models with great detail. In particular, the correlation coefficient compares directly the observed and forecasted streamflow values in test period whereas the Willmott's Index and Nash-Sutcliffe efficiency for normalized metrics to measure the overall prediction skill. As correlation coefficient is largely based on linear agreements between observed and forecasted streamflow, *WI* and E_{NS} which are free from any assumptions were utilized. The E_{NS} provides a better assessment of model as it is sensitive to differences in the observed and forecasted means and variances (ASCE, 2000; NERC, 1975). However, the *WI* has an advantage over the r^2 and E_{NS} where differences are calculated as squared values so larger values can be overestimated whereas smaller values can be neglected (Legates and McCabe, 1999b). This insensitivity was overcome using WI (Willmott 1981) as the ratio of the MSE was considered instead of the square of the differences (Willmott, 1981). It can be noted that the ELM model outperformed the SVR and GRNN models in terms of each performance indicator and computational run-time (section 3.5). The best sets of model inputs were not the same for each of the explored predictive modeling techniques, indicating that the respective model type responds differently to different input variable sets and the data patterns/attributes in the historical input data. Overall, the most accurate input combinations were based on models M5, M1 and M2 for the ELM, SVR and GRNN, respectively. The assessment of the different modeling paradigms (ELM, SVR, and GRNN) with different input variable sets (i.e. M1-M5) illustrates that the most accurate forecasts depend on the model that is used and the optimization of the model parameters for the given input data - this is the reason why the different modeling techniques perform best with different input variable sets.

A closer assessment of the results showed that for the ELM model, the lowest magnitude of the absolute error metric (MAE = 71.544 $m^3 s^{-1}$) was acquired by the fifth antecedent value of stream-flow as an input variable. In this study, the modeling process was terminated at the fifth lag time in accordance with the PACF statistical results in which the first five lag times were the correlated attributes in forecasting one-step-ahead (Fig. 5). Using the efficiency criteria as the basis for identifying an optimal model for the ELM, SVR and GRNN (Nash-Sutcliffe efficiency (E_{ns}) 0.578, 0.378 and 0.144, respectively) yielded the correlation coefficient (r) (0.799, 0.761 and 0.468) and Willmott's index (WI) (0.853, 0.802 and 0.689). It should be noted that the WI was inspected for analyzing the trends on the modeling accuracy as it is a descriptive measure similar to the regression coefficient. However, the regression coefficient can be overly sensitive to extreme values as it is based on the differences in the observed and forecasted means and variances. Thus, larger values in forecasted stream-flow can be overestimated by r whereas smaller values can be neglected (Legates and McCabe, 1999a). This issue can be overcome using the WI metric where the ratio of the mean square error is considered instead of the square of the differences (Willmott, 1981). Therefore, WI was not established to be a measure of correlation, but rather applied to analyze the consistency in the modeling accuracy. The mean absolute error measures (e.g., $RMSE$ and MAE), which represented the raw error values of the forecasted stream-flow in the testing period were also evaluated for all predictive models. The ELM, SVR and GRNN yielded a $RMSE$ (m^3s^{-1}) value of 87.906, 106.749 and 125.187, respectively and the MAE (m^3s^{-1}) values were 71.544, 90.905 and 103.551, respectively. Taken together, it is noteworthy that several evaluation metrics indicate the merits of the ELM over the SVR and GRNN models.

The computational advantage of ELM in comparison to SVR and GRNN is a useful characteristic to consider when designing stream-flow forecasts for decision making and real-time forecasting applications. For monthly streamflow forecasts (the focus of this study), the computational efficiency of ELM would be useful for building an ensemble forecast, such as an ensemble of bootstrap-based forecasts (Tiwari and Adamowski, 2013); for short-term stream-flow forecasting, the ELM's computational efficiency would be well suited for flood-warning systems when compared with other models (e.g. SVR or GRNN). From the computational run-time results (Table 2), on average, we can run three ELM models for each SVR and GRNN model that is tested. This result, which agrees closely with the other investigations in terms of a shorter modeling time (Deo et al., 2016; Deo and Sahin, 2016; Deo and Sahin 2015; Sahin et al., 2014; Quilty et al., 2016), identifies a distinct advantage of the ELM over the SVR and GRNN models. The shorter modeling time is due to the non-tuned mechanism of the ELM algorithm (i.e. non-iterative-based solutions to the network parameters), and this led the ELM model to be much faster in comparison with the SVR and GRNN models. The current use of the ELM in forecasting hydrological time series is an offline application where the run-time is not significant. In real-time forecasting problems, the model's run-time for producing a forecast is of high importance, especially for complex problems in water resources planning and decision-making, where forecasts may be required at fine temporal and spatial units for flood forecasting, hydro-power generation, or forecasting water quality indicators to inform expert-(warning) systems.

To proceed with the further examination of the forecasting models, a series of scatter-plots for each input combination for the ELM, SVR and GRNN models are shown in Figure 6. The regression coefficient R^2 and the regression formula ($y=a_o x+a_1$) is also displayed in the scatter-plots. The ELM model was able to obtain the best fit line between the observed and the forecasted stream-flow values using the first input combination with an R^2 value of 0.66. With any similar input combinations, the SVR and the GRNN models had R^2 = 0.57 and 0.43, respectively.

Performing an assessment of model accuracy based on investigating the low-, medium- and high-flow forecasted stream-flow values is useful to give a deeper and more comprehensive analysis. Fig. 7 illustrates the patterns noted for the fluctuations between the observed and the forecasted stream-flow values for the three models within the testing period. This shows the percentage of the under- or over-forecasted stream-flow values. As shown in this figure, the low and medium values of the stream-flow were slightly overestimated, while the high values of the stream-flow were under-estimated. The relatively poor performance for these extreme values of stream-flow shows that there were possibly an insufficient number of peak flows in the training dataset used in estimating the model's parameters. However, despite this, the ELM-based forecasts were able to more closely match the actual stream-flow values within the testing period (2007-2010) compared to the SVR and GRNN models.

It is also useful to present a visualization of the distribution error for a forecasting model within the testing period. Fig. 8 illustrates the magnitude of the relative error (*RE*) indicator for the ELM, SVR and GRNN models. The figure shows that the maximum *RE* values are approximately (-40.3, -52.6 and -42.8) % for the ELM, SVR and GRNN, respectively. However, the ELM model's relative error percentages were between (-10% and 10%) for more than 71% of the testing dataset. Here, it can also be seen that the SVR and the GRNN models displayed consistency in their magnitude of the *RE* when compared with the ELM model. It was also not surprising to note that the maximum distribution error for all prescribed models appeared in the high peak stream flow values.

In predictive modelling especially for hydrological variables (e.g. streamflow) that exhibit significant chaotic (stochastic) behavior, it is important to cross-validate the model in terms of the statistical distribution of the observed and forecasted property. This can assist in deducting the level of agreement between the datum point pairs of observed and forecasted streamflow. We thus generated box-plots with the respective distributions to demonstrate how closely the ELM model forecasts compared with the original time series and the other data-driven approaches. In Fig. 9, the box-plots are utilized to indicate the degree of overall spread in the observed and predicted data in accordance with the respective quartile values and the whiskers. It should be noted that the lower quartile, Q_{25}, represents the 25th percentile and the upper quartile, Q_{75}, represents the 75th percentile, while the median is represented by, Q_{50}, the 50th percentile. Whiskers are stretched outwards from the lower and upper quartiles (Q_{25}-Q_{75}) up to the smallest and largest outliers' values, respectively. Based on the box-plots and whiskers, it was evident that the spread of the ELM model-based forecasts closely resembled the observed stream-flow, with a slightly larger interquartile range. The results showed strong high flow predictive capability of the ELM although the low flows tend to be slightly overestimated when compared to the observed time-series. On the other hand, the SVR and GRNN models showed reduced variance and poorer predictive coverage for the lower flows when compared to the observed time series and the ELM model. In terms of the practical advantage of the model, one may ascertain that the ELM model is expected to generate forecasted streamflow that exhibit closer resemblance to the observed values. Furthermore, the accuracy of the GRNN model is significantly limited as the distribution of the test data cover a much smaller range than the observed streamflow values.

The results presented in this paper support the potential advantage of employing the ELM over the SVR and GRNN models for accurate and reliable stream-flow forecasts, indicating that this model may be further explored by hydrologists.

5. Conclusion

The primary basis of the present research was to investigate the applicability and capability of the extreme learning machine algorithm for forecasting one-month-ahead stream-flow in semi-arid environments, with the Tigris River in Iraq as an example. The effectiveness of the ELM model was investigated and compared to data-driven models widely used in hydrological forecasting application (i.e., support vector regression and generalized regression neural network) and evaluated in terms of several performance indicators.

In conclusion, the main findings of this research can be summarized as follows:

1. It was found that the ELM algorithm was a useful alternative to the SVR and GRNN models for forecasting stream-flow in the semi-arid environment of the Tigris River in Iraq.
2. The test dataset (2007-2010) indicated that ELM can potentially improve the predictive accuracy of the modeling process compared to the SVR and the GRNN models. Based on the absolute values of the error metrics, it was found that the *RMSE* and *MAE* were reduced significantly by about 17.44-29.78 % and 21.3-30.92 %, respectively when the ELM model was evaluated compared to the SVR and GRNN models.
3. Overall, the results showed a good improvement in the forecasting accuracies using the ELM model. The degree of matching between the observed and forecasted values of stream-flow was significantly enhanced by the ELM model over the other two data-driven models (SVR and GRNN). The ELM forecasts of the monthly stream-flow pattern was very close to the observed records. In addition, the relative error percentages did not exceed by ± 10% for over 70 percent of the tested data.
4. The computational efficiency of the ELM - due to the least-squares formulation used to estimate its model parameters – enables the rapid development of numerous models that can process large datasets which is very useful for ensemble learning approaches or forecasting at very fine temporal and spatial scales, both of which are very relevant problems/research topics in hydrology (and other environmental/engineering disciplines).

In the present study, the investigated ELM model yielded forecasts of stream-flow based on historical data of stream-flow itself, and therefore, did not incorporate the exogenous effects of hydrological variables (e.g. rainfall, temperature, humidity and evaporation) and large-scale climatic phenomena (e.g., sea surface temperatures or climate mode indices) (Deo and Şahin, 2016b) that could influence the evolution of the stream-flow data patterns. As this is considered a limitation of the present technique, in follow-up research, one could also enhance the ELM model performance by using multivariate modeling inputs by considering the various related variables that can be used to forecast streamflow. As streamflow model's input data can exhibit significant stationarity features such as periodic patterns, trends and other forms of stochasticity, the potential application of a multi-resolution algorithm where the cyclic behavior of the hydrological inputs are extracted by wavelet transformation algorithm (e.g. (Deo et al., 2016a; Tiwari and Adamowski, 2013)) can also be implemented into the ELM algorithm. In this paper, we have used time-lagged streamflow data as inputs, but in a follow-up study, multiple input can be used where a separate feature selection algorithm can be used to extract the pertinent useful features for an accurate estimation of stream-flow at multiple gauging locations and forecast lead times. In doing so, a variety of non-linear variable input selection methods such bootstrap rank-ordered CMI (broCMI) discussed in (Quilty et al. (2016)), evolutionary method of Salcedo-Sanz et al., (Salcedo-Sanz et al., 2014) and tree-based iterative input scheme (IIS) (Galelli and Castelletti, 2013) can be implemented to improve the forecasting efficiency. Moreover, this paper has validated the ELM model for monthly forecasting but it is also worthwhile to explore the effectiveness of the ELM model for short-term stream-flow forecasting (e.g. minute, hourly or daily scales) for its possibility in short-term prediction and the associated model uncertainties that may provide us with greater insights into its practicality for flood discharge forecasting problems.

References

Abrahart, R.J., See, L., 2000. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrological Processes, 14(July 1999): 2157-2172. DOI:10.1002/1099-1085(20000815/30)14:11/12<2157::AID-HYP57>3.0.CO;2-S

Acharya, N., Shrivastava, N.A., Panigrahi, B.K., Mohanty, U.C., 2013. Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine. Climate Dynamics, 43(5-6): 1303-1310. DOI:10.1007/s00382-013-1942-2

Afan, H.A. et al., 2014. ANN Based Sediment Prediction Model Utilizing Different Input Scenarios. Water Resources Management, 29(4): 1231-1245. DOI:10.1007/s11269-014-0870-1

Akusok, A., Bjork, K.-M., Miche, Y., Lendasse, A., 2015. High-performance extreme learning machines: a complete toolbox for big data applications. Access, IEEE, 3: 1011-1025.

Al-Ansari, N.A., 2013. Management of water resources in Iraq: Perspectives and prognoses.

Allawi, M.F., El-Shafie, A., 2016. Utilizing RBF-NN and ANFIS Methods for Multi-Lead ahead Prediction Model of Evaporation from Reservoir. Water Resources Management. DOI:10.1007/s11269-016-1452-1

Araghinejad, S., Burn, D.H., Karamouz, M., 2006. Long-lead probabilistic forecasting of streamflow using ocean-atmospheric and hydrological predictors. Water Resources Research, 42(3): n/a-n/a. DOI:10.1029/2004wr003853

ASCE, 2000. Artificial neural networks in hydrology. II: Hydrologic applications. Journal of Hydrologic Engineering`, 5(2): 124-137.

Aziz, K., Rahman, A., Fang, G., Shrestha, S., 2014. Application of artificial neural networks in regional flood frequency analysis: a case study for Australia. Stochastic Environmental Research and Risk Assessment, 28(3): 541-554.

Bray, M., Han, D., 2004. Identification of support vector machines for runoff modelling. Journal of Hydroinformatics: 265-280.

Chen, C.-S., Liu, C.-H., Su, H.-C., 2008. A nonlinear time series analysis using two-stage genetic algorithms for streamflow forecasting. Hydrological Processes, 22(18): 3697-3711. DOI:10.1002/hyp.6973

Chiang, Y.M., Chang, L.C., Chang, F.J., 2004. Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling. Journal of Hydrology, 290(3-4): 297-311. DOI:10.1016/j.jhydrol.2003.12.033

Chiang, Y.M., Chang, L.C., Tsai, M.J., Wang, Y.F., Chang, F.J., 2011. Auto-control of pumping operations in sewerage systems by rule-based fuzzy neural networks. Hydrology and Earth System Sciences, 15(1): 185-196. DOI:10.5194/hess-15-185-2011

Cigizoglu, H.K., 2005. Application of Generalized Regression Neural Networks to Intermittent Flow Forecasting and Estimation. Journal of Hydrologic Engineering, 10(4): 336-341. DOI:10.1061/(ASCE)1084-0699(2005)10:4(336)

Cigizoglu, H.K., Alp, M., 2006. Generalized regression neural network in modelling river sediment yield. Advances in Engineering Software, 37(2): 63-68.

Danandeh Mehr, a., Kahya, E., Şahin, a., Nazemosadat, M.J., 2014. Successive-station monthly streamflow prediction using different artificial neural network algorithms. International Journal of Environmental Science and Technology. DOI:10.1007/s13762-014-0613-0

Deo, R.C., Şahin, M., 2015a. Application of the Artificial Neural Network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydrometeorological parameters and climate indices in eastern Australia. Atmospheric Research, 161-162: 65-81. DOI:10.1016/j.atmosres.2015.03.018

Deo, R.C., Şahin, M., 2015b. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmospheric Research, 153: 512-525.

Deo, R.C., Şahin, M., 2016a. An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. Environmental monitoring and assessment, 188(2): 1-24.

Deo, R.C., Şahin, M., 2016b. An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. Environmental Monitoring and Assessment, DOI (10.1007/s10661-016-5094-9).

Deo, R.C., Samui, P., Kim, D., 2015a. Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. Stochastic Environmental Research and Risk Assessment. DOI:10.1007/s00477-015-1153-y

Deo, R.C., Tiwari, M.K., Adamowski, J.F., Quilty, J.M., 2015b. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. Stochastic Environmental Research and Risk Assessment: 1-30.

Deo, R.C., Tiwari, M.K., Adamowski, J.F., Quilty, J.M., 2016a. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. Stochastic Environmental Research and Risk Assessment. DOI:10.1007/s00477-016-1265-z

Deo, R.C., Wen, X., Qi, F., 2016b. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. Applied Energy, 168: 568-593.

Dorado, J. et al., 2003. Prediction and modeling of the rainfall-runoff transformation of a typical urban basin using ann and gp. Applied Artificial Intelligence, 17(4): 329-343. DOI:10.1080/713827142

El-Shafie, A., Abdin, A.E., Noureldin, A., Taha, M.R., 2009. Enhancing inflow forecasting model at aswan high dam utilizing radial basis neural network and upstream monitoring stations measurements. Water Resources Management, 23(11): 2289-2315. DOI:10.1007/s11269-008-9382-1

El-Shafie, A., El-Manadely, M., 2010. An integrated neural network stochastic dynamic programming model for optimizing the operation policy of Aswan High Dam. Hydrology research, 42(1): 50-67.

El-Shafie, a., Noureldin, a., Taha, M., Hussain, a., Mukhlisin, M., 2012. Dynamic versus static neural network model for rainfall forecasting at Klang River Basin, Malaysia. Hydrology and Earth System Sciences, 16(4): 1151-1169. DOI:10.5194/hess-16-1151-2012

El-Shafie, A., Taha, M.R., Noureldin, A., 2007. A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. Water Resources Management, 21(3): 533-556. DOI:10.1007/s11269-006-9027-1

Fijani, E., Nadiri, A.A., Moghaddam, A.A., Tsai, F.T.-C., Dixon, B., 2013. Optimization of DRASTIC method by supervised committee machine artificial intelligence to assess groundwater vulnerability for Maragheh–Bonab plain aquifer, Iran. Journal of Hydrology, 503: 89-100.

Firat, M., 2008. Comparison of artificial intelligence techniques for river flow forecasting. Hydrology and Earth System Sciences, 12(1): 123-139.

Firat, M., Turan, M.E., Yurdusev, M.A., 2010. Comparative analysis of neural network techniques for predicting water consumption time series. Journal of hydrology, 384(1): 46-51.

Galelli, S., Castelletti, A., 2013. Tree-based iterative input variable selection for hydrological modeling. Water Resour. Res., 49(7): 4295-4310.

Ghorbani, M.A., Zadeh, H.A., Isazadeh, M., Terzi, O., 2016. A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. Environmental Earth Sciences, 75(6). DOI:10.1007/s12665-015-5096-x

Graves, D., Pedrycz, W., 2009. Fuzzy prediction architecture using recurrent neural networks. Neurocomputing, 72(7-9): 1668-1678. DOI:10.1016/j.neucom.2008.07.009

Greco, R., 2012. A fuzzy-autoregressive model of daily river flows. Computers and Geosciences, 43: 17-23. DOI:10.1016/j.cageo.2012.02.031

Guo, J., Zhou, J., Qin, H., Zou, Q., Li, Q., 2011. Monthly streamflow forecasting based on improved support vector machine model. Expert Systems with Applications, 38(10): 13073-13081. DOI:10.1016/j.eswa.2011.04.114

Guven, A., 2009. Linear genetic programming for time-series modelling of daily flow rate. Journal of Earth System Science, 118(2): 137-146. DOI:10.1007/s12040-009-0022-9

He, Z., Wen, X., Liu, H., Du, J., 2014. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. Journal of Hydrology, 509: 379-386.

Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2003. A practical guide to support vector classification.

Hsu, K.-l., Gupta, H.V., Gao, X., Sorooshian, S., Imam, B., 2002. Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis. Water Resources Research, 38(12). DOI:10.1029/2001WR000795

Huang, G.-B., Zhou, H., Ding, X., Zhang, R., 2012. Extreme learning machine for regression and multiclass classification. IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society, 42(2): 513-29. DOI:10.1109/TSMCB.2011.2168604

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006a. Extreme learning machine: theory and applications. Neurocomputing, 70(1): 489-501.

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006b. Extreme learning machine: Theory and applications, pp. 489-501. DOI:10.1016/j.neucom.2005.12.126

Huang, G., Huang, G.B., Song, S., You, K., 2015. Trends in extreme learning machines: a review. Neural networks : the official journal of the International Neural Network Society, 61: 32-48. DOI:10.1016/j.neunet.2014.10.001

Huang, G.B., Chen, L., Siew, C.K., 2006c. Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Transactions on Neural Networks, 17(4): 879-892. DOI:10.1109/TNN.2006.875977

Katambara, Z., Ndiritu, J., 2009. A fuzzy inference system for modelling streamflow: Case of Letaba River, South Africa. Physics and Chemistry of the Earth, 34(10-12): 688-700. DOI:10.1016/j.pce.2009.06.001

Khatibi, R., Ghorbani, M.A., Kashani, M.H., Kisi, O., 2011. Comparison of three artificial intelligence techniques for discharge routing. Journal of hydrology, 403(3): 201-212.

Kisi, O., 2007. Streamflow Forecasting Using Different Artificial Neural Network Algorithms. Journal of Hydrologic Engineering, 12(5): 532-539. DOI:10.1061/(ASCE)1084-0699(2007)12:5(532)

Kisi, O., 2008a. The potential of different ANN techniques in evapotranspiration modelling. Hydrological Processes, 22(14): 2449-2460. DOI:10.1002/hyp.6837

Kisi, O., 2008b. Stream flow forecasting using neuro-wavelet technique, pp. 4142-4152. DOI:10.1002/hyp.7014

Kişi, Ö., 2011. A combined generalized regression neural network wavelet model for monthly streamflow prediction, pp. 1469-1479. DOI:10.1007/s12205-011-1004-4

Kisi, O., Shiri, J., Tombul, M., 2012. Modeling rainfall-runoff process using soft computing techniques. Computers & Geosciences, 51: 108-117. DOI:10.1016/j.cageo.2012.07.001

Legates, D.R., McCabe, G.J., 1999a. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. water resources research, 35(1): 233-241.

Legates, D.R., McCabe, G.J., 1999b. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. Water Resour. Res., 35(1): 233-241.

Li, B., Cheng, C., 2014. Monthly discharge forecasting using wavelet neural networks with extreme learning machine. Science China Technological Sciences, 57(12): 2441-2452. DOI:10.1007/s11431-014-5712-0

Lima, A.R., Cannon, A.J., Hsieh, W.W., 2016. Forecasting daily streamflow using online sequential extreme learning machines. Journal of Hydrology, 537: 431-443. DOI:10.1016/j.jhydrol.2016.03.017

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental modelling \& software, 15(1): 101--124.

Maier, H.R. et al., 2014. Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. Environmental Modelling & Software, 62: 271-299. DOI:10.1016/j.envsoft.2014.09.013

Makarynska, D., Makarynskyy, O., 2008. Predicting sea-level variations at the Cocos (Keeling) Islands with artificial neural networks. Computers and Geosciences, 34(12): 1910-1917. DOI:10.1016/j.cageo.2007.12.004

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(3-4): 336-354. DOI:10.1016/j.jhydrol.2008.01.023

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4): 115-133.

Mohammadi, K. et al., 2015. Extreme learning machine based prediction of daily dew point temperature. Computers and Electronics in Agriculture, 117: 214-225. DOI:10.1016/j.compag.2015.08.008

Moosavi, V., Malekinezhad, H., Shirmohammadi, B., 2014. Fractional snow cover mapping from MODIS data using wavelet-artificial intelligence hybrid models. Journal of Hydrology, 511: 160-170.

Moosavi, V., Vafakhah, M., Shirmohammadi, B., Behnia, N., 2013. A Wavelet-ANFIS Hybrid Model for Groundwater Level Forecasting for Different Prediction Periods. Water Resources Management, 27(5): 1301-1321. DOI:10.1007/s11269-012-0239-2

NERC, 1975. Flood Studies Report Vol. 1, Hydrological Studies. Natural Environment Research Council, London, UK.

Ni, Q., Wang, L., Ye, R., Yang, F., Sivakumar, M., 2010. Evolutionary Modeling for Streamflow Forecasting with Minimal Datasets: A Case Study in the West Malian River, China. Environmental Engineering Science, 27(5): 377-385. DOI:10.1089/ees.2009.0082

Nourani, V., Komasi, M., Alami, M.T., 2012. Hybrid Wavelet–Genetic Programming Approach to Optimize ANN Modeling of Rainfall–Runoff Process. Journal of Hydrologic Engineering, 17(6): 724-741. DOI:10.1061/(ASCE)HE.1943-5584.0000506

ÖZger, M., 2009. Comparison of fuzzy inference systems for streamflow prediction, pp. 261-273. DOI:10.1623/hysj.54.2.261

Palani, S., Liong, S.Y., Tkalich, P., 2008. An ANN application for water quality forecasting. Marine Pollution Bulletin, 56(9): 1586-1597. DOI:10.1016/j.marpolbul.2008.05.021

Platt, J.C., 1999. 12 fast training of support vector machines using sequential minimal optimization. Advances in kernel methods: 185-208.

Pramanik, N., Panda, R.K., Singh, A., 2010. Daily river flow forecasting using wavelet ANN hybrid models. Journal of Hydroinformatics, 13(1): 49-49. DOI:10.2166/hydro.2010.040

Principé, J., Lefebvre, C., 1998. NeuroSolutions v 7.1.1.0. Gainesville, FL: NeuroDimension. URL: http://www. nd. com.

Quilty, J., Adamowski, J., Khalil, B., Rathinasamy, M., 2016. Bootstrap rank-ordered conditional mutual information (broCMI)—A nonlinear input variable selection method for water resources modeling. Water Resources Research.

Raghavendra. N, S., Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. Applied Soft Computing, 19: 372-386. DOI:10.1016/j.asoc.2014.02.002

Rieker, J.D., Labadie, J.W., 2012. An intelligent agent for optimal river-reservoir system management. Water Resources Research, 48(9).
Şahin, M., Kaya, Y., Uyar, M., Yıldırım, S., 2014. Application of extreme learning machine for estimating solar radiation from satellite data. International Journal of Energy Research, 38(2): 205-212.
Salcedo-Sanz, S., Deo, R.C., Carro-Calvo, L., Saavedra-Moreno, B., 2015. Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms. Theoretical and Applied Climatology(Ml). DOI:10.1007/s00704-015-1480-4
Salcedo-Sanz, S., Pastor-Sánchez, A., Prieto, L., Blanco-Aguilera, A., García-Herrera, R., 2014. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization–Extreme learning machine approach. Energy Conversion and Management, 87: 10-18.
Saldarriaga, J., Gómez, R., Salas, D., 2004. Artificial intelligence methods applicability on water distribution networks calibration, Critical Transitions in Water and Environmental Resources Management. ASCE, pp. 1-11.
Salman, D.A., Amer, S.A., Ward, F.A., 2014. Water Appropriation Systems for Adapting to Water Shortages in Iraq. JAWRA Journal of the American Water Resources Association, 50(5): 1208-1225.
Sang, Y.F., 2013. Improved Wavelet Modeling Framework for Hydrologic Time Series Forecasting. Water Resources Management, 27(8): 2807-2821. DOI:10.1007/s11269-013-0316-1
Savic, D.A., Walters, G.A., Davidson, J.W., 1999. A genetic programming approach to rainfall-runoff modelling. Water Resources Management, 13(3): 219--231.
Shamim, M., Bray, M., Remesan, R., Han, D., 2015. A hybrid modelling approach for assessing solar radiation. Theoretical and Applied Climatology: 1-18.
Shu, C., Burn, D.H., 2004. Artificial neural network ensembles and their application in pooled flood frequency analysis. Water Resources Research, 40(9).
Singh, H., Sankarasubramanian, A., 2014. Systematic uncertainty reduction strategies for developing streamflow forecasts utilizing multiple climate models and hydrologic models. Water Resources Research, 50(2): 1288-1307.
Solomatine, D.P., Shrestha, D.L., 2009. A novel method to estimate model uncertainty using machine learning techniques. Water Resources Research, 45(12): n/a-n/a. DOI:10.1029/2008wr006839
Sudheer, K., Gosain, A., Ramasastri, K., 2002. A data-driven algorithm for constructing artificial neural network rainfall-runoff models. Hydrological Processes, 16(6): 1325-1330.
Takahashi, N., Guo, J., Nishi, T., 2008. Global convergence of SMO algorithm for support vector regression. IEEE Transactions on Neural Networks, 19(6): 971-982.
Taormina, R., Chau, K.-w., Sethi, R., 2012. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. Engineering Applications of Artificial Intelligence, 25(8): 1670-1676. DOI:10.1016/j.engappai.2012.02.009
Tezel, G., Buyukyildiz, M., 2015. Monthly evaporation forecasting using artificial neural networks and support vector machines. Theoretical and Applied Climatology(Terzi 2013). DOI:10.1007/s00704-015-1392-3
Tiwari, M.K., Adamowski, J., 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour. Res., 49(10): 6486-6507.
Tiwari, M.K., Chatterjee, C., 2011. A new wavelet-bootstrap-ANN hybrid model for daily discharge forecasting. Journal of Hydroinformatics, 13(3): 500-519.
Tripathi, S., Srinivas, V., Nanjundiah, R.S., 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. Journal of Hydrology, 330(3): 621-640.
Tsoukalas, I., Kossieris, P., Efstratiadis, A., Makropoulos, C., 2016. Surrogate-enhanced evolutionary annealing simplex algorithm for effective and efficient optimization of water resources problems on a budget. Environmental Modelling & Software, 77: 122-142.

Vapnik, V., 1995. The Nature of statistical Learning Theory.
Wang, Z.-l., Sheng, H.-h., 2010. Rainfall prediction using generalized regression neural network: case study Zhengzhou, Computational and Information Sciences (ICCIS), 2010 International Conference on. IEEE, pp. 1265-1268.
Whigham, P.A., Crapper, P.F., 2001. Modelling rainfall-runoff using genetic programming. Mathematical and Computer Modelling, 33(6-7): 707-721. DOI:10.1016/S0895-7177(00)00274-0
Willmott, C.J., 1981. On the validation of models. Physical geography, 2(2): 184-194.
Willmott, C.J., Robeson, S.M., Matsuura, K., 2012. A refined index of model performance. International Journal of Climatology, 32(13): 2088-2094. DOI:10.1002/joc.2419
Yadav, B., Ch, S., Mathur, S., Adamowski, J., 2016. Discharge forecasting using an Online Sequential Extreme Learning Machine (OS-ELM) model: A case study in Neckar River, Germany. Measurement, 92: 433-445. DOI:10.1016/j.measurement.2016.06.042
Yang, J.-F., Zhai, Y.-J., Xu, D.-P., Han, P., 2007. SMO algorithm applied in time series model building and forecast, 2007 International Conference on Machine Learning and Cybernetics. IEEE, pp. 2395-2400.
Yaseen, Z.M. et al., 2015a. RBFNN versus FFNN for daily river flow forecasting at Johor River, Malaysia. Neural Computing and Applications. DOI:10.1007/s00521-015-1952-6
Yaseen, Z.M., El-shafie, A., Jaafar, O., Afan, H.A., Sayl, K.N., 2015b. Artificial intelligence based models for stream-flow forecasting: 2000–2015. Journal of Hydrology, 530: 829-844. DOI:10.1016/j.jhydrol.2015.10.038
Yonaba, H., Anctil, F., Fortin, V., 2010. Comparing Sigmoid Transfer Functions for Neural Network Multistep Ahead Streamflow Forecasting. Journal of Hydrologic Engineering, 15(4): 275-283. DOI:10.1061/(ASCE)HE.1943-5584.0000188
Zhang, G.P., Patuwo, E.B., Michael Y, H., 1998. Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14(1): 35-62. DOI:10.1016/S0169-2070(97)00044-7

Highlights

Non-tuned data-driven approach is investigated for monthly stream-flow forecasting

The model is examined for river flow located in semi-arid environment

A comprehensive assessment and comparative analysis have been carried out

Further enhancements are proposed for similar hydrological forecasting problem

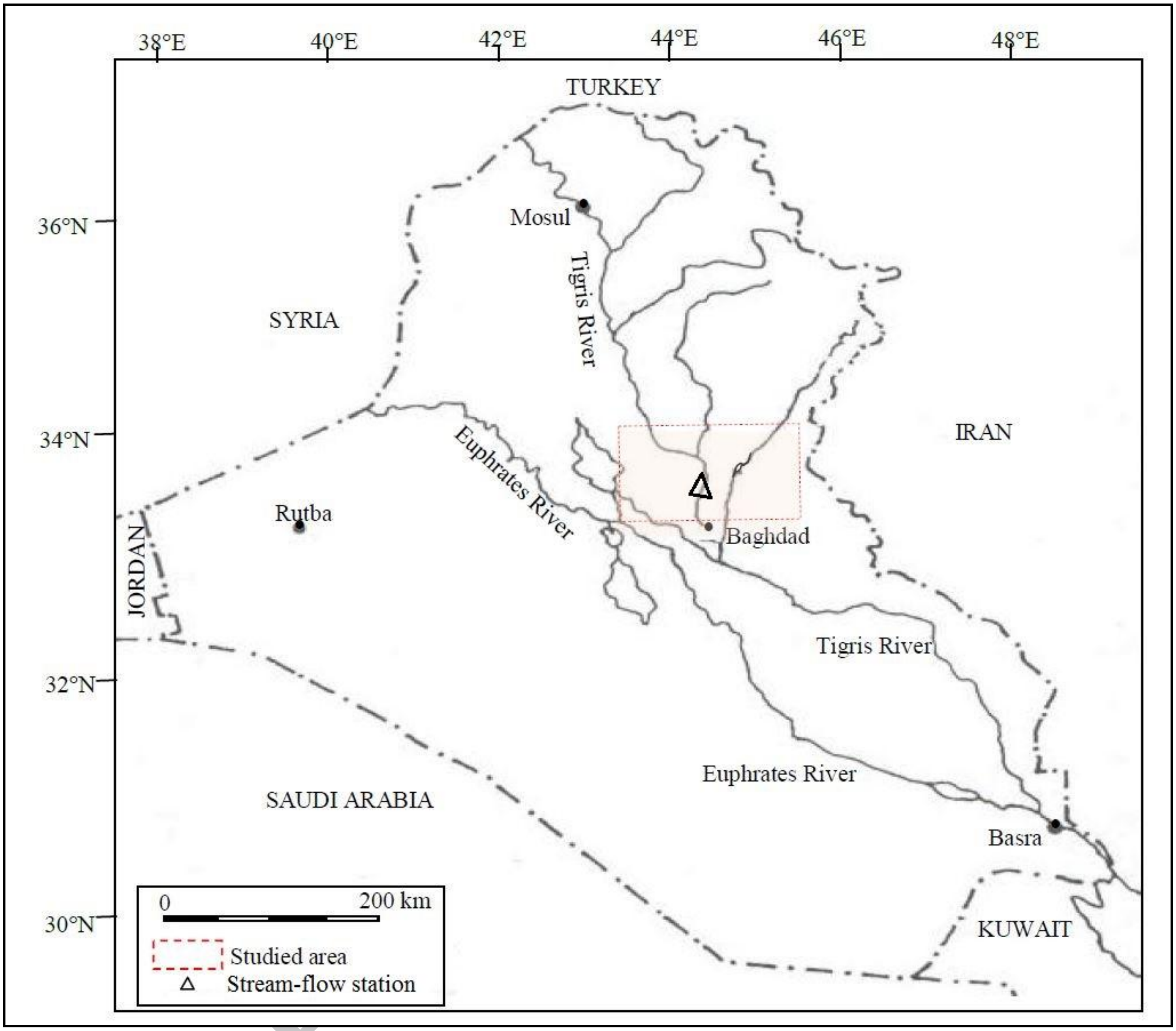

Fig. 1: Case study site: Baghdad stream-flow station located in Iraq region.

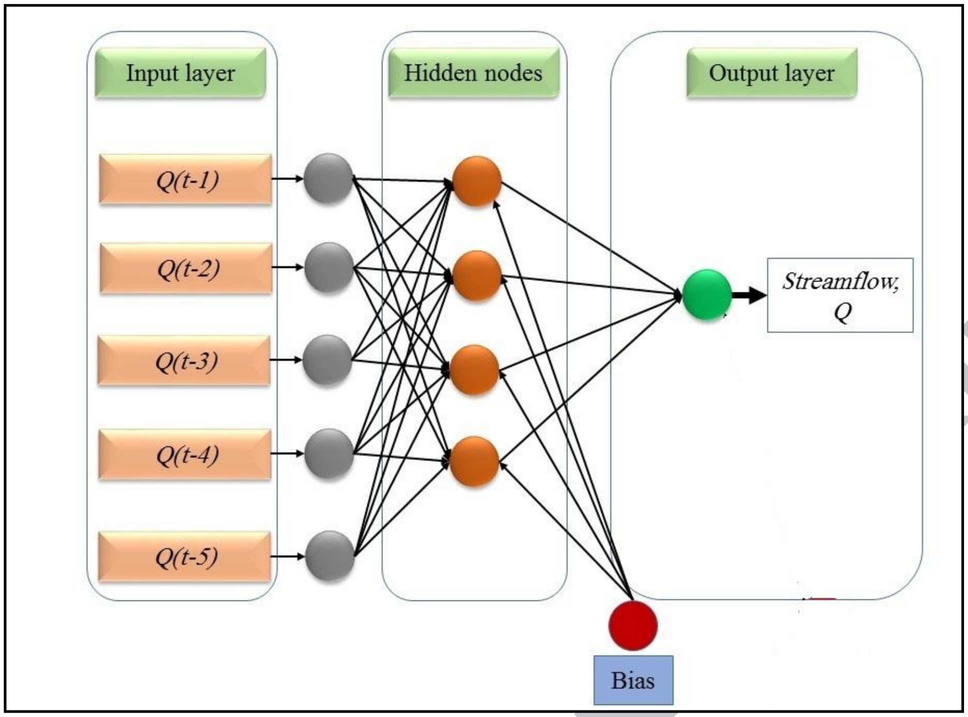

Fig. 2: The topological structure of extreme learning machine network used in this study. Inputs layer are denoted the lag times of the streamflow, hidden nodes presents the hidden layer that randomly generated and the output layer generates the predicted values of the streamflow, *Q*.

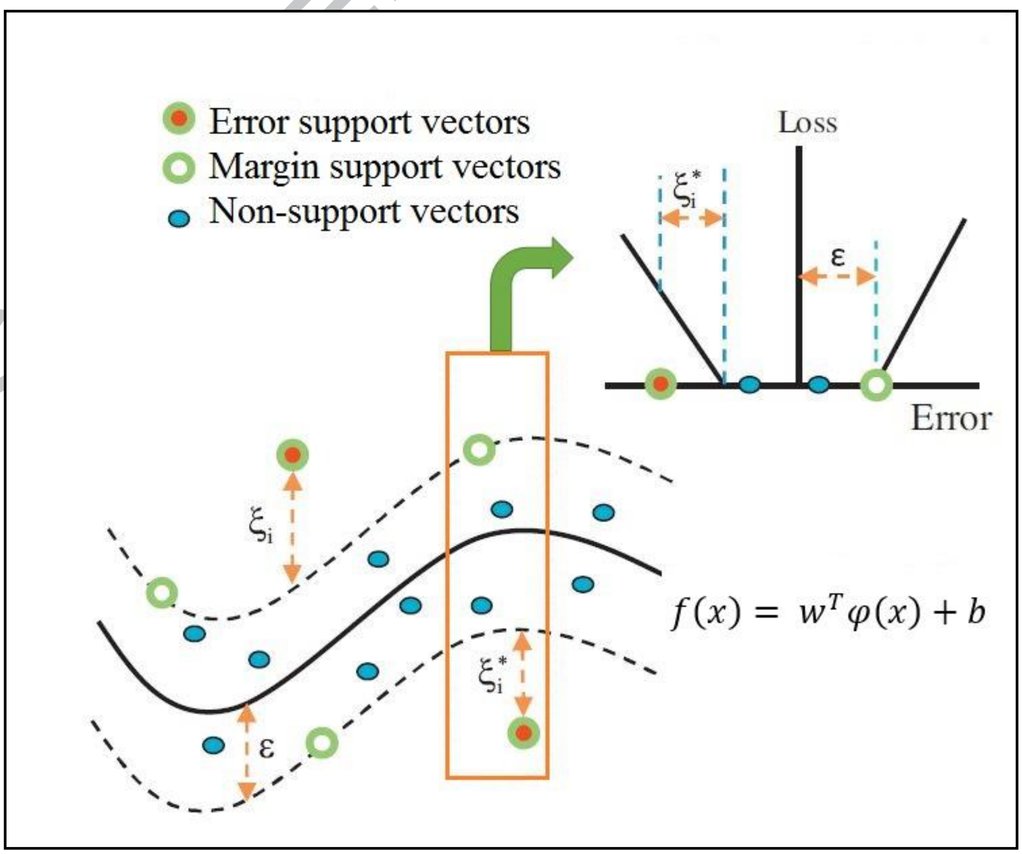

Fig. 3: Nonlinear support vector machine with Vapnik's -insensitive loss function (Raghavendra. N and Deka, 2014).

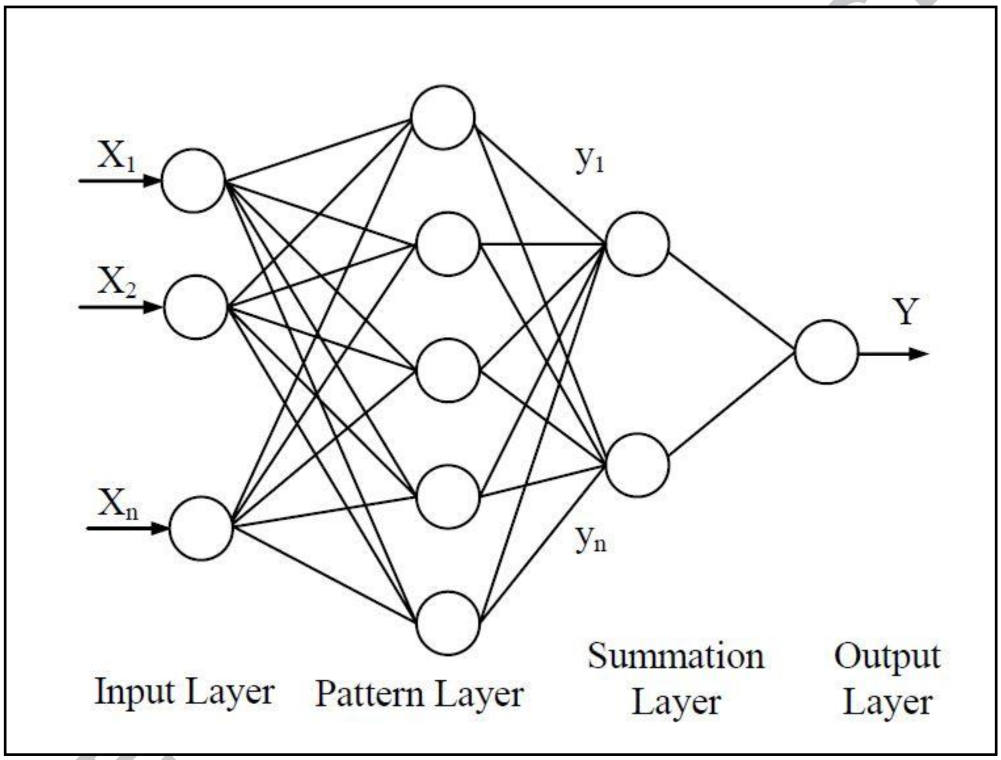

Fig. 4: Schematic of the generalized regression neural network architecture (Firat, 2007).

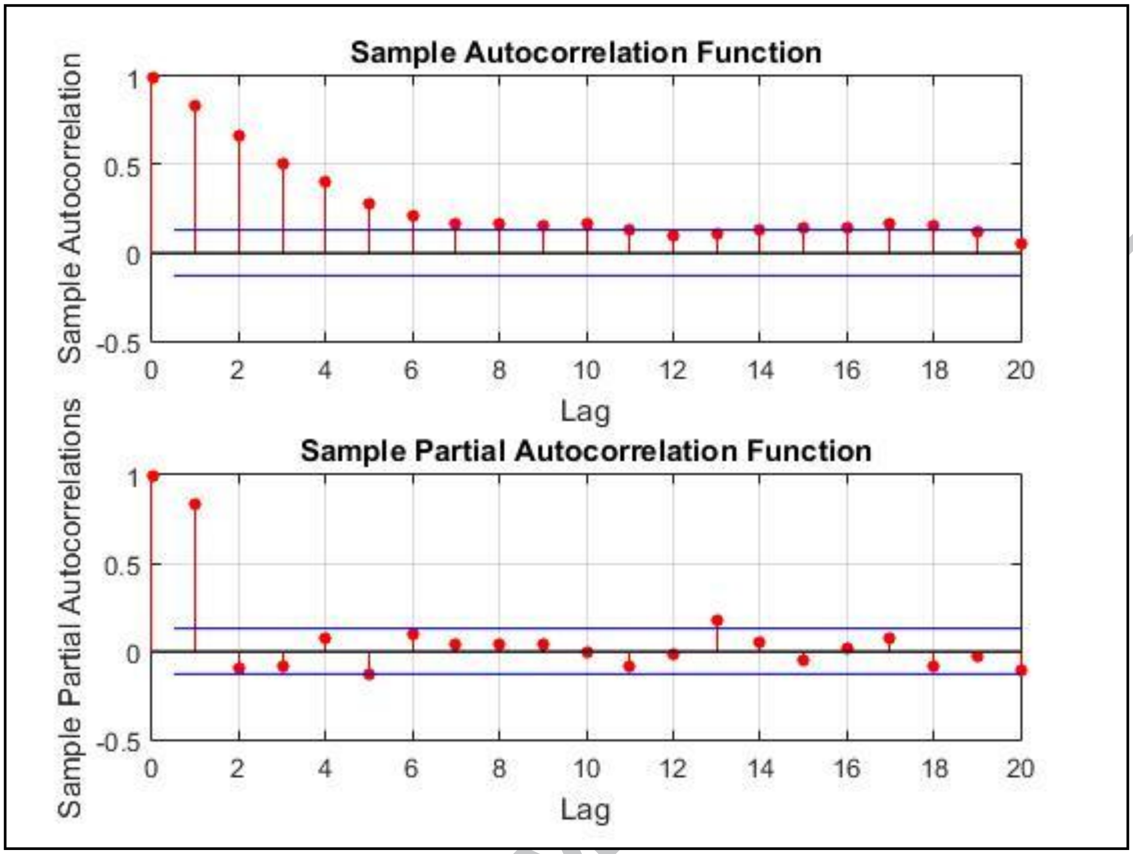

Fig. 5: Auto-correlation and partial auto-correlation function for the monthly stream-flow time series at the Baghdad (Iraq) stream-flow station.

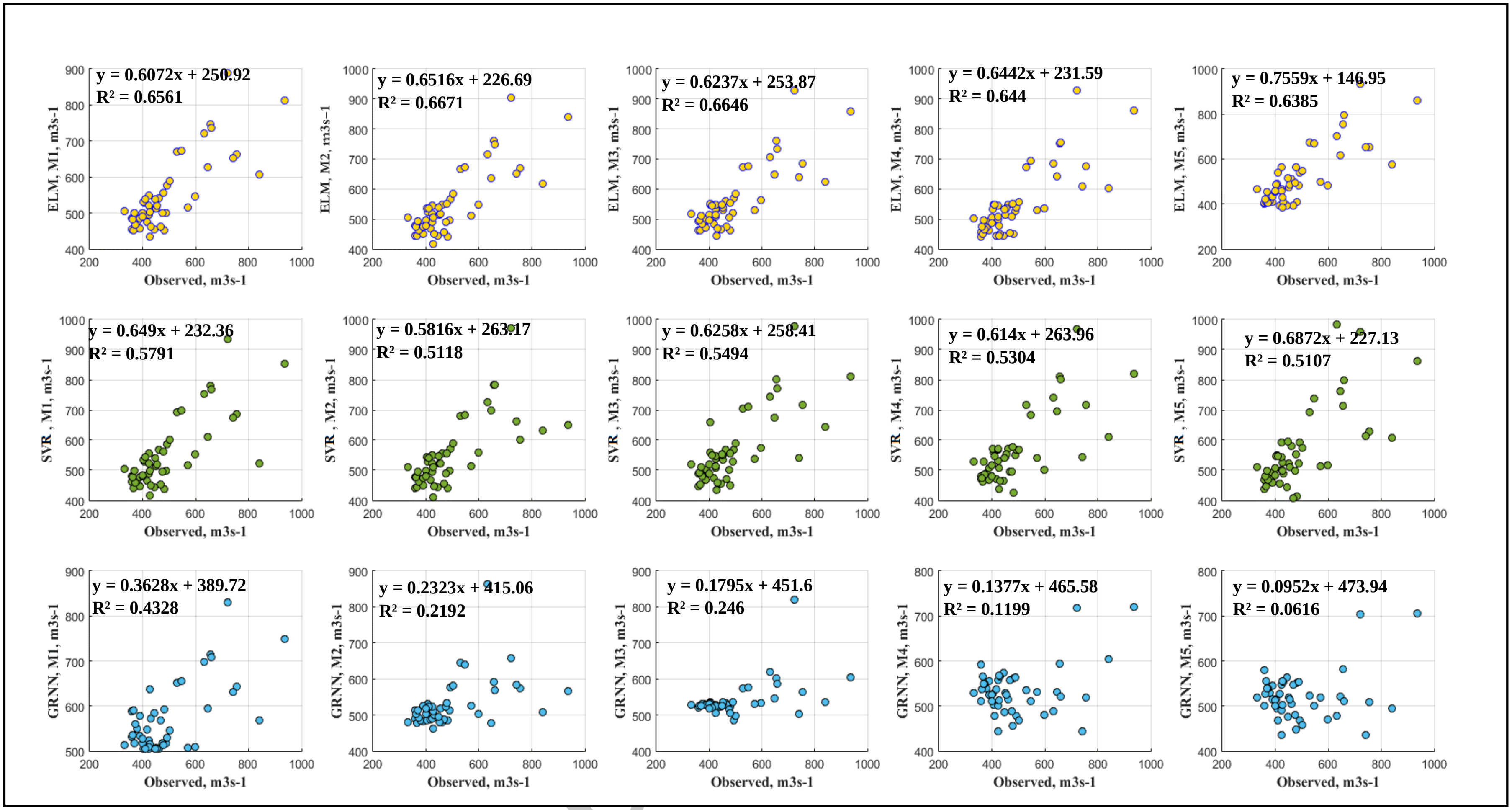

Fig. 6: Scatter-plots of observed and simulated mean streamflow, Q (t) using the ELM, SVR and GRNN model with five sets of input combination (Model 1: Q (t – 1); Model 2: Q (t – 1, t – 2); Model 3: Q (t – 1, t – 2; t – 3); Model 4: Q (t – 1, t – 2, t – 3, t – 4) and Model 5 (t – 1, t – 2, t – 3, t – 4, t – 5). m^3s^{-1}

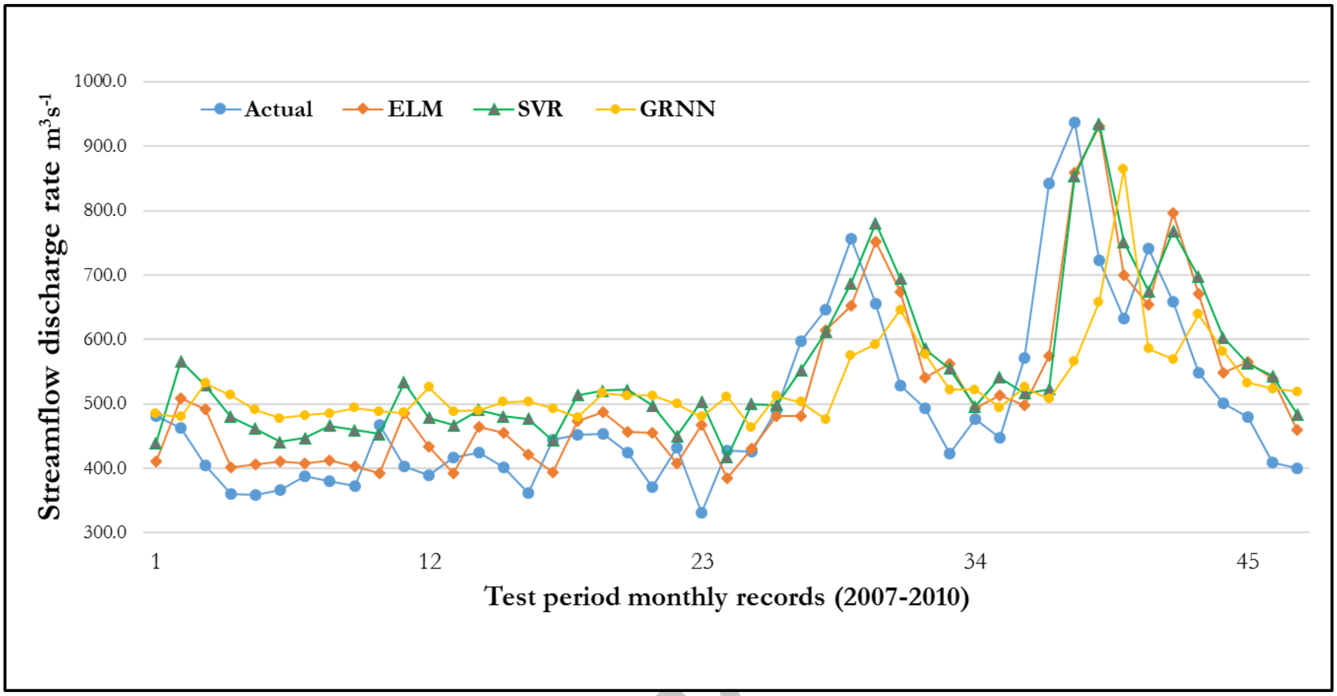

Fig. 7: Optimal models and observed streamflow for testing period (2007-2010) using ELM, SVR and GRNN.

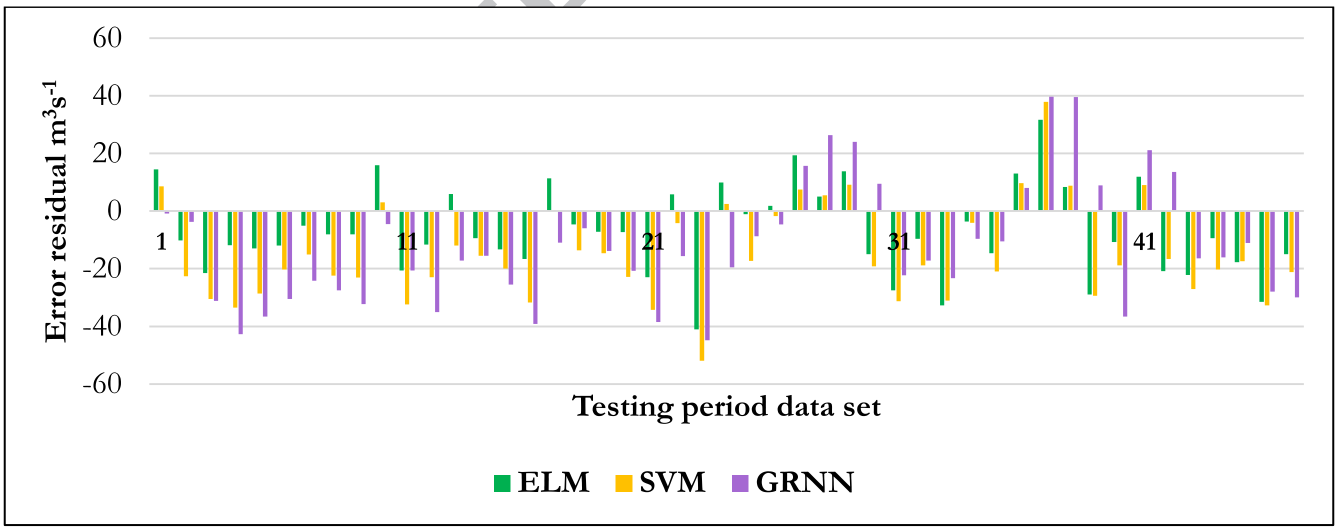

Fig. 8: The relative distribution error (RE) for the testing period phase (2007-2010) using ELM, SVR and GRNN models.

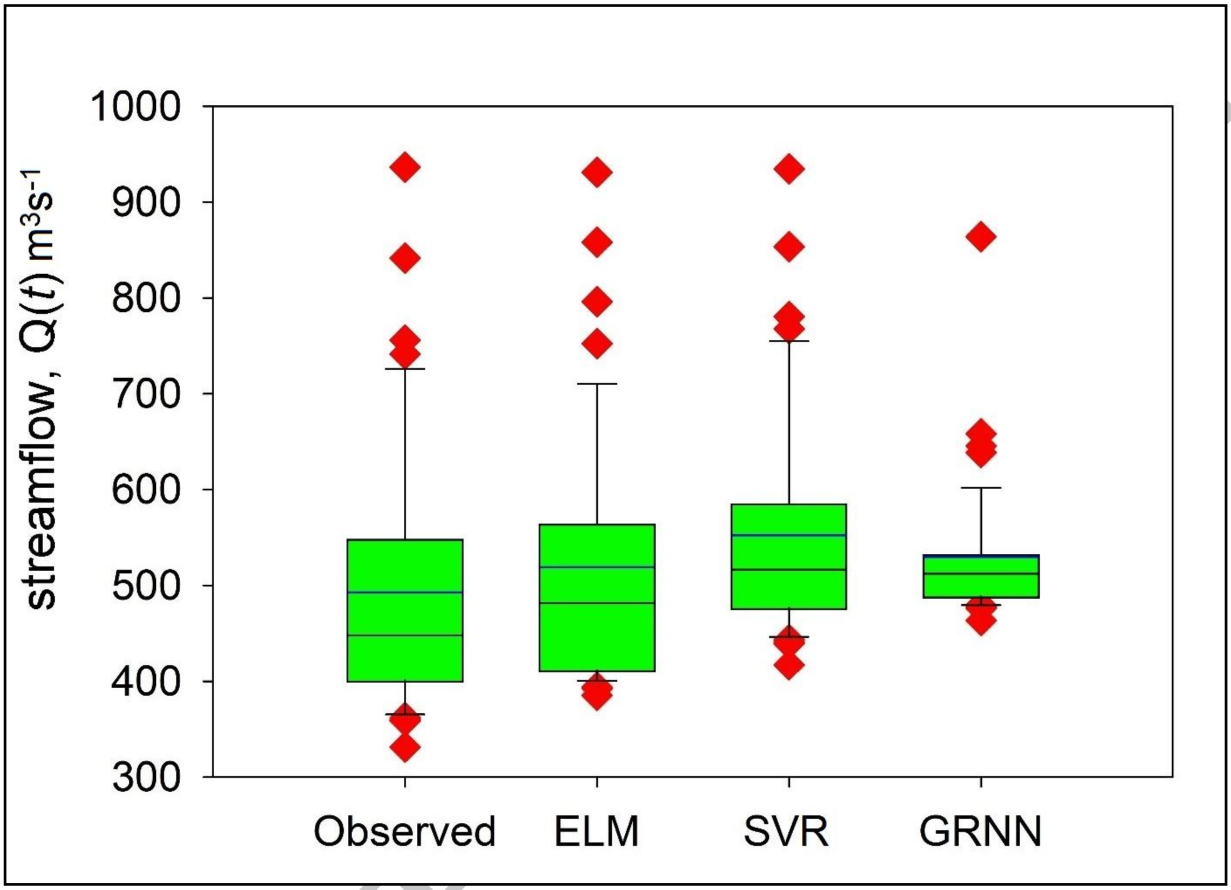

Fig. 9: Box-plots of the observed streamflow compared with forecasted streamflow from ELM, SVR and GRNN models.

Table 1: Descriptive statistics for mean monthly stream-flow for Tigris River at Baghdad (Iraq) (1991-2010).

Partition	Time Period	No. Records	Q (m^3s^{-1})				
			Mean	**St. Dev.**	**Median**	**Minimum**	**Maximum**
Training	Jun. 1991 - Dec. 2007	187	780.099	379.712	674.700	298.100	2651.000
Testing	Jan. 2007 - Dec. 2010	48	489.879	136.746	445.900	331.400	936.400
Complete	Jun. 1991 - Dec. 2010	235	720.820	363.469	636.900	298.100	2651.000

Table 2: Performance indicators using correlation coefficient (*r*), Willmott's Index (*WI*), Nash-Sutcliffe efficiency (E_{NS}), Root mean square error (*RMSE*), mean absolute error (*MAE*), and time consumption (seconds) for the ELM, SVR and GRNN models evaluated in the testing period.'

Models	Correlation Coefficient, *r*	Willmott's Index *WI*	Nash-Sutcliffe Coefficient, E_{NS}	Root mean square error, m^3s^{-1} *RMSE*	Mean Absolute Error, m^3s^{-1} *MAE*	Model Run-Time, *s*
			ELM			
M1	0.810	0.820	0.473	98.272	87.012	2.620
M2	0.817	0.834	0.501	95.571	83.899	2.749
M3	0.815	0.814	0.406	104.307	92.544	3.129
M4	0.803	0.826	0.471	98.432	85.117	3.176
M5	0.799	0.853	0.578	87.906	71.544	3.617
			SVR			
M1	0.761	0.802	0.378	106.749	90.905	9.732
M2	0.715	0.769	0.325	111.228	95.140	11.279
M3	0.741	0.776	0.240	118.020	101.850	11.014
M4	0.728	0.769	0.220	119.539	106.585	12.935
M5	0.715	0.768	0.158	124.155	108.367	10.148
			GRNN			
M1	0.658	0.689	0.108	127.818	113.698	13.519
M2	0.468	0.540	0.144	125.187	103.551	8.781
M3	0.496	0.549	0.106	127.988	107.989	10.812
M4	0.346	0.507	0.027	133.522	114.063	15.034
M5	0.248	0.416	0.000	135.350	112.606	12.361

Highlights

Non-tuned data-driven approach is investigated for monthly stream-flow forecasting

The model is examined for river flow located in semi-arid environment

A comprehensive assessment and comparative analysis have been carried out

Further enhancements are proposed for similar hydrological forecasting problem