



Non-tuned machine learning approach for hydrological time series forecasting

Zaher Mundher Yaseen¹ · Mohammed Falah Allawi¹ · Ali A. Yousif² · Othman Jaafar¹ · Firdaus Mohamad Hamzah¹ · Ahmed El-Shafie³

Received: 19 September 2016/Accepted: 24 November 2016/Published online: 16 December 2016
© The Natural Computing Applications Forum 2016

Abstract Stream-flow forecasting is a crucial task for hydrological science. Throughout the literature, traditional and artificial intelligence models have been applied to this task. An attempt to explore and develop better expert models is an ongoing endeavor for this hydrological application. In addition, the accuracy of modeling, confidence and practicality of the model are the other significant problems that need to be considered. Accordingly, this study investigates modern non-tuned machine learning data-driven approach, namely extreme learning machine (ELM). This data-driven approach is containing single layer feedforward neural network that selects the input variables randomly and determine the output weights systematically. To demonstrate the reliability and the effectiveness, one-step-ahead stream-flow forecasting based on three time-scale pattern (daily, mean weekly and mean monthly) for Johor river, Malaysia, were implemented. Artificial neural network (ANN) model is used for comparison and evaluation. The results indicated ELM approach superior the ANN model level accuracies and time consuming in addition to precision forecasting in tropical zone. In measurable terms, the dominance of ELM model over ANN model was indicated in accordance

with coefficient determination (R^2) root-mean-square error (RMSE) and mean absolute error (MAE). The results were obtained for example the daily time scale $R^2 = 0.94$ and 0.90, RMSE = 2.78 and 11.63, and MAE = 0.10 and 0.43, for ELM and ANN models respectively.

Keywords Extreme learning machine · Artificial neural network · Stream-flow forecasting · Tropical environment · Multiple time horizons

1 Introduction

It is scientifically proved that stream-flow is characterized by high nonlinearity distribution and dynamic pattern [1–6]. Over the past several decades, stream-flow forecasting has been an important and challenging issue [7–9]. In practical management, stream-flow forecast is tremendously significant for water resources planning and operation. Real-time forecasting that can be addressed as short-term stream-flow can yield an important and reliable operation for flood control and mitigation protection, whereas long-term forecasting is essential for several water resources applications involving river sediment operation, reservoir and water demand sustainability, hydro power production, and several others uses [10, 11].

Since the early of 1970, the classical approaches based on mathematical and statistical models had been undertaken to solve this issue; for instance, multiple linear regression (MLR) model and autoregressive integrated moving average (ARIMA) [12–16]. The main drawback of the classical models is that they are limited with linear regression solution that is not really applicable in capturing the highly stochasticity of stream-flow pattern. Recently, a noticeable use of artificial intelligence (AI) techniques to

✉ Zaher Mundher Yaseen
zahermundher@gmail.com

¹ Civil and Structural Engineering Department, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

² Water Resources Engineering Department, College of Engineering, University of Duhok, Duhok, Iraq

³ Civil Engineering Department, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia

model and forecast river flow time series, such as classifiers and machine learning approaches [2, 17–28], fuzzy logic system [4, 29–33], evolutionary computation [34–38], wavelet complementary models [39–44]. Despite the flexibility and utility of AI techniques in modeling stream-flow, they still have some drawbacks and limitations (*e.g.*, over-fitting, slow learning speed, local minima, and difficulty to capture the high complexity, non-stationary, dynamism and nonlinearity of time series). Those mathematical models are much advanced over the physical models that are usually needed much efforts and information in order to module the hydrological variables of a specific watershed area [45–47].

In the last decade, a non-tuned machine learning entitled extreme learning machine algorithm is gained a noticeable efficiency in modeling the regression problem [48–52]. This approach is firstly proposed by [53] and has been broadly utilized in many applications due to its capabilities (*i.e.*, randomization of the internal network weights and less time consuming during the learning processes), for example, evapotranspiration prediction [54], fast object recognition and image classification [55, 56], landslide displacement prediction [57], sales prediction of fashion retailing [58], melting points prediction of organic compounds [59], big data classification [60], and use of prior knowledge [48]. Relatedly with the investigated application, monthly stream-flow prediction model conducted by integrating ELM model with wavelet decomposition approach for a case study in southwestern China region [61]. In 2016, Deo and Sahin [62] implemented ELM model for stream-flow forecasting in Queensland to validate its superiority over artificial neural network (ANN) models. Another version of ELM model which is based on online sequential methodology was investigated for short-term river flow forecasting for Canadian region [63]. In semi-arid region, the application of ELM model showed an optimistic satisfactory for modeling long-term stream-flow forecasting [64]. Generally, the ELM model as non-tuned approach showed an outstanding level of accuracies vis-à-vis state-of-the-art AI models.

By comparing with AI techniques, ELM method offers a highly modeling capabilities such as randomly assigned internal weights, much fast learning processes and very simple architecture of neural network. Moreover, it overcomes the shortcomings of the traditional popular gradient-based learning algorithms.

Since there are numerous neural networks architecture, researchers are still facing questions such as which neural network is precisely best fit or should be utilized for specific problem. Unfortunately, there is no general neural network satisfactory answering these questions because we are still in the stage of exploring methods that are capable for modeling hydrological processes. However, in this

research, an investigation of the proficiency of ELM method in forecasting one step ahead of hydrological data representing in time series fashion with acceptable level of accuracy overcome the highlighted drawbacks of the existing AI techniques (*e.g.*, time consuming of the learning processes and trivial human intervention). Daily, average weekly and average monthly river stream-flow data are employed to examine the proposed ELM method as alternative model for stream-flow forecasting in tropical environment, Johor River, Malaysia. Johor River is one of the essential rivers in Peninsular Malaysia that provides Johor state with water supply in addition to various domestic and agriculture usages. Studying the flow pattern of this river flow is extremely significant for its sustainability. In order to examine the proposed ELM method, a comparative analysis between the performance of the classical ANN method and the proposed ELM has been carried out. The following sections of the manuscript are established in the following manner. Description of the ANN method and its algorithm expressions have been presented. Comprehensive details for the proposed ELM method and approach are reported in Sect. 3. The case study and the data preparation with the model structure are presented in Sect. 4. Section 5 addresses the results achieved by the ELM method and introduced a details discussion on the comparison analysis with the classical ANN method. Finally, the conclusion of the current research has been highlighted in Sect. 6.

2 Artificial neural network (ANN) approach

ANN is an advance mathematical model that deals with problem similar to the human brain attitudes. Theoretically, ANN development based on several components was reported by Task and Neural (2000), which are: (i) the ANN structure has a single elements called nodes, the information processing occurs in each nodes, (ii) these connection links to transfer signals between the each of nodes, (iii) for each connection link has an associated weight that represents its connection strength, and (iv) the activation function is founded in each node to determine its output signal. In general, ANN architecture is designed in three main elements: input layer has one or more number of input nodes based on the number of parameters for the model; single hidden layer including the activation function; and output layer nodes.

Several algorithms have been used in the learning processes of ANN in the field of hydrological applications. However, according to the existing literature, radial basis function neural network (RBFNN) method superior to the other training methods. This is because RBFNN works with higher reliability, faster convergence and smaller

extrapolation [65–67]. RBFNN model was proposed by Lowe and Broomhead [52]. Mainly, the RBFNN is structured with three different layers, input, hidden and output layers as demonstrated in Fig. 1. Each layer has its own function in order to implement the proposed assignment, and in this research, it is forecasting stream-flow. The first layer is designed to transfer the input variables into the RBFNN process. The second layer is introduced to adapt the nonlinear transformation function connections between the input variables and neurons (radial basis function nodes in the hidden layer). Finally, the linear transformation has been implemented in order to transfer the hidden layer space information to the output layer which is considered as the desired variable.

The RBF functions $\varphi_1, \varphi_2, \dots, \varphi_N$ are known as basic hidden transfer functions, while $\{\varphi_i(x)\} i = 1N$ is termed as the intermediate hidden domain. There is one constraint of such architecture that the number of RBF functions that forward the input variables from the input layer to the hidden layer (N) is less than the number of the available data records that presenting the input–output pattern. The most popular RBF functions that usually used in such pattern recognition application is the Gaussian function; the following formula shows the Gaussian representation as one-dimensional domain:

$$\varphi(x, \mu) = e^{\frac{\|x - \mu\|^2}{2d^2}} \quad (1)$$

where μ indicates the center of the Gaussian function which presents the mean value of x and d indicates the distance from the center of $\varphi(x, \mu)$.

There are two different key parameters, namely the center and spread d . These two parameters are initiated at the commencement of the model process and then adjusted during training process. Generally, the hidden unit is more sensitive to data points near the center and

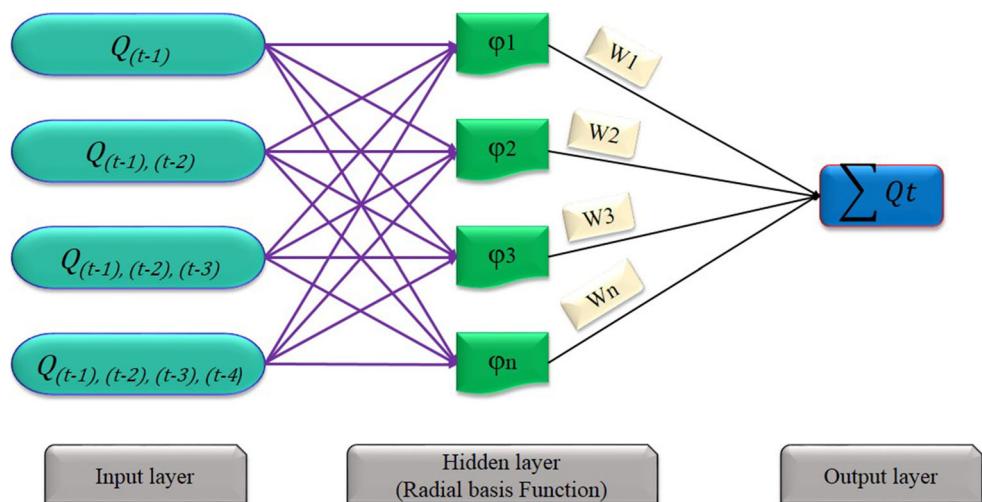
this is according to the Gaussian radial function. Such sensitivity could be controlled and adjusted by changing the initial values of the spread d . An example of the Gaussian radial basis is demonstrated in Fig. 2. It is obvious from Fig. 2 that the radial basis function is less sensitive to the input data pattern when the spread value is relatively large.

3 Non-tuned machine learning: ELM Approach

Extreme learning machine (ELM) was first proposed by [53]. The first proposed was with single hidden layer feedforward neural networks. After that, it was developed to the generalized SLFNs, the hidden layer after developed became need not be neuron alike [68, 69]. The main feature of the extreme learning machine method has strong potential as applicable alternative methods due that the hidden layer does not need to be tuned. In addition, it considers the minimum norm of the output layer weights and requires fewer parameter settings, the training processes are extremely fast compared with the gradient descent learning algorithms, and it indicates a good generalization performance [52].

This article investigates the capability of the ELM to forecast stream-flow using different time interval time series (daily, average weekly and average monthly). Different input combinations are supplied to the SLFNs-ELM, involving the antecedent records to forecast one step ahead. Mathematically, the random generation of the input weights consisting different lags time including $(Q(t-1), Q(t-1) Q(t-2), \dots, Q(t-1) Q(t-2) Q(-n))$ is mapped to L-dimension ELM random feature space, whereas the network output (forecasted Q_t) can be expressed as:

Fig. 1 Radial basis function algorithm structure in the artificial neural network



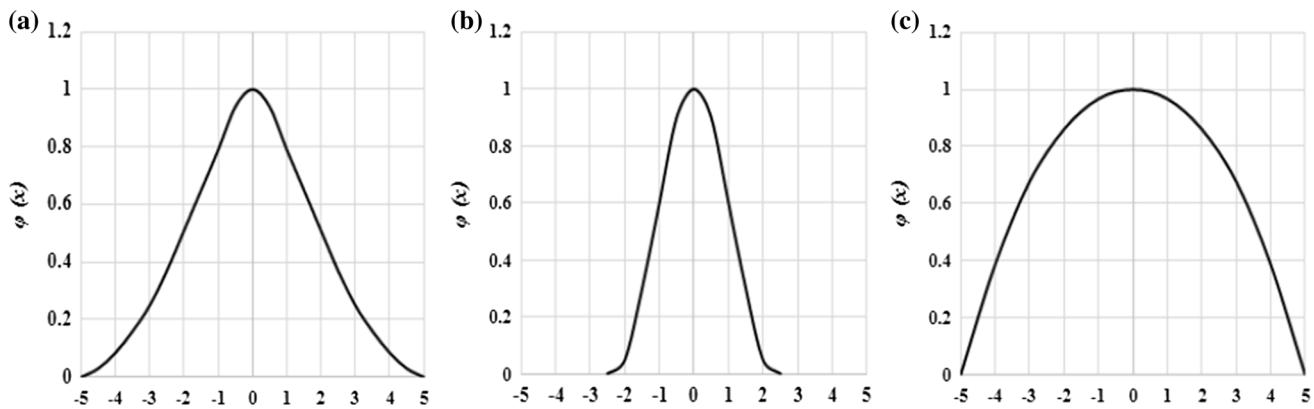


Fig. 2 RBFNN with different levels of spread. **a** Normal spread. **b** Small spread. **c** Large spread

$$fL(x) = \sum_{i=1}^L h_i \beta_i = h_{(x)} \beta \quad (2)$$

where $\beta = [\beta_1, \beta_2, \dots, \beta_L]^T$ represents the weight sandwiched between the hidden layer and output layer, whereas $h_{(x)} = [g_{1(x)}, \dots, g_{L(x)}]$ is the hidden weight output in which randomly generalization features for the input vectors. L is the number of the hidden neurons. $g_{i(x)}$ is the output of the i th hidden nodes. In this research, the development of the modeling conducted via sigmoid activation function, as best can be expressed:

$$g_i(x) = \text{SigAct}(x, a_i, b_i) = (1 + \exp(-(a_i x + b_i))) \quad (3)$$

where a_i and b_i are the random input weights and bias between input nodes and i th hidden nodes.

ELM model has the capability to resolve the learning problem $H\beta = T$. Here, $T = [T_1, \dots, T_N]$ is the target output matrix and $H = [h^T(X_1), h^T(X_2), h^T(X_3), \dots, h^T(X_N)]^T$. The β presents the output weight which is determined using $\beta = H^\dagger T$ where H^\dagger is the Moore–Penrose generalized inverse of matrix H .

While developing a stream-flow forecasting model, it is of importance to consider the implementation efficiency. The main point of the ELM approach is its potential to reduce the computational time for the training procedure noticeably because of its strong mathematical process capable to lessen the iterative and descent steps. Furthermore, the time needed for the calculation of the weights for the input variables and the desired output variable and their adjustment utilizing the least squares solution (linear system) which is computationally time consumption procedure. On the top of that, the approach of the ELM method employs the singular value decomposition (SVD) as adaptable and stabilization numerical technique for computing the Moore–Penrose generalization inverse. Finally, 30 nodes for the hidden layer are considered by utilizing

the trials-and-error process in order to achieve a balance in the statistical evaluation metrics in both model phases (training and testing).

4 Case study and data preparation

In the present research, daily, weekly and monthly data stream-flow were used belonging to Johor River Basin, Malaysia. Figure 3 illustrates the location of the case study. The drainage area of the selected river is around 2640 km^2 with total length 123 km. The observed inflow data are 20 years' time period in which equal (7300 days), (1040 weeks) and (240 months), for the observation period (1989–2008). The historical data as shown in (Fig. 4) at Rantau Panjang station were obtained from the Department of Irrigation and Drainage (DID). The first month of the year is January, and the last month of the year is December. Data division for the training and testing phases was assigned 90 and 10%, respectively. This division of data set was assigned using trial-and-error procedure until the best performance forecasting accuracies level obtained [11, 67, 70, 71].

In modeling regression problem, for example, stream-flow in the present study, the selection of the appropriate lag times as an input variables is one of the essential priorities [6]. Autocorrelation function (ACF) and partial autocorrelation function (PACF) have been applied on the selected time series data, for the purpose to determine the most correlated lead time and to perform effective modeling. Figure 5 indicates the ACF and the PACF for all the interval of stream-flow (daily, average weekly and average monthly) with various lags time. According to Fig. 5 based on these two behaviors of the autocorrelation and partial autocorrelation; three lags, four lags and two lags were considered to forecast one step ahead for daily, weekly and monthly, respectively. Hence, in the current study, we have

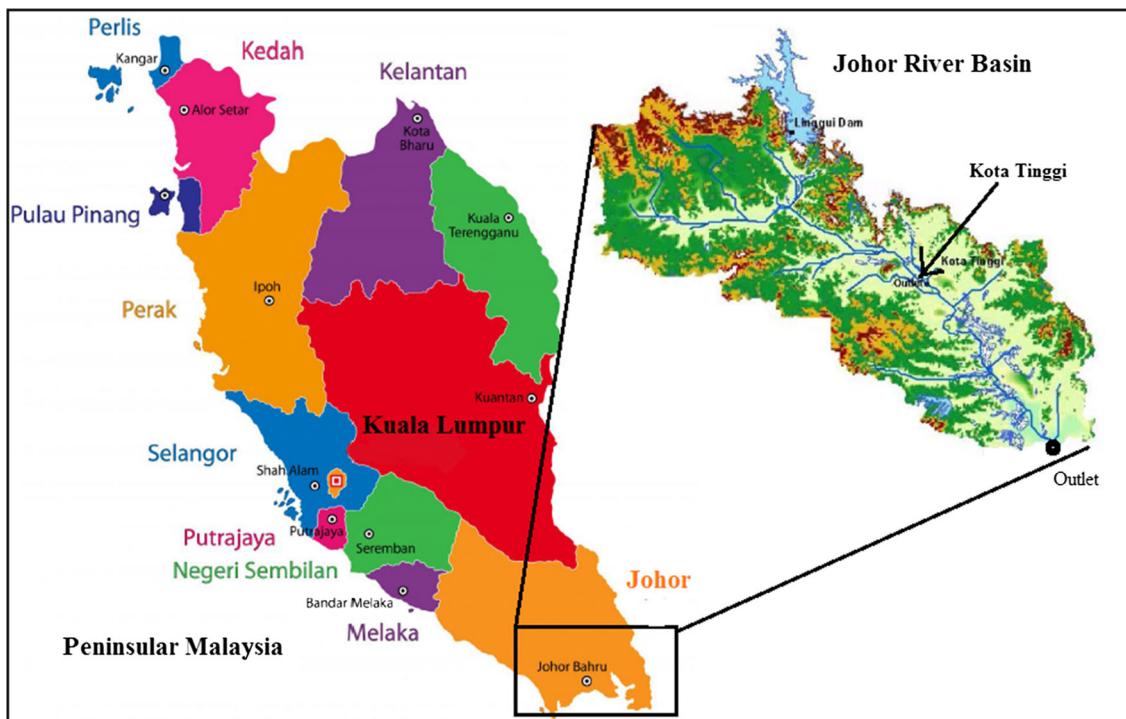


Fig. 3 Location of the case study (Johor River Basin, Peninsular Malaysia)

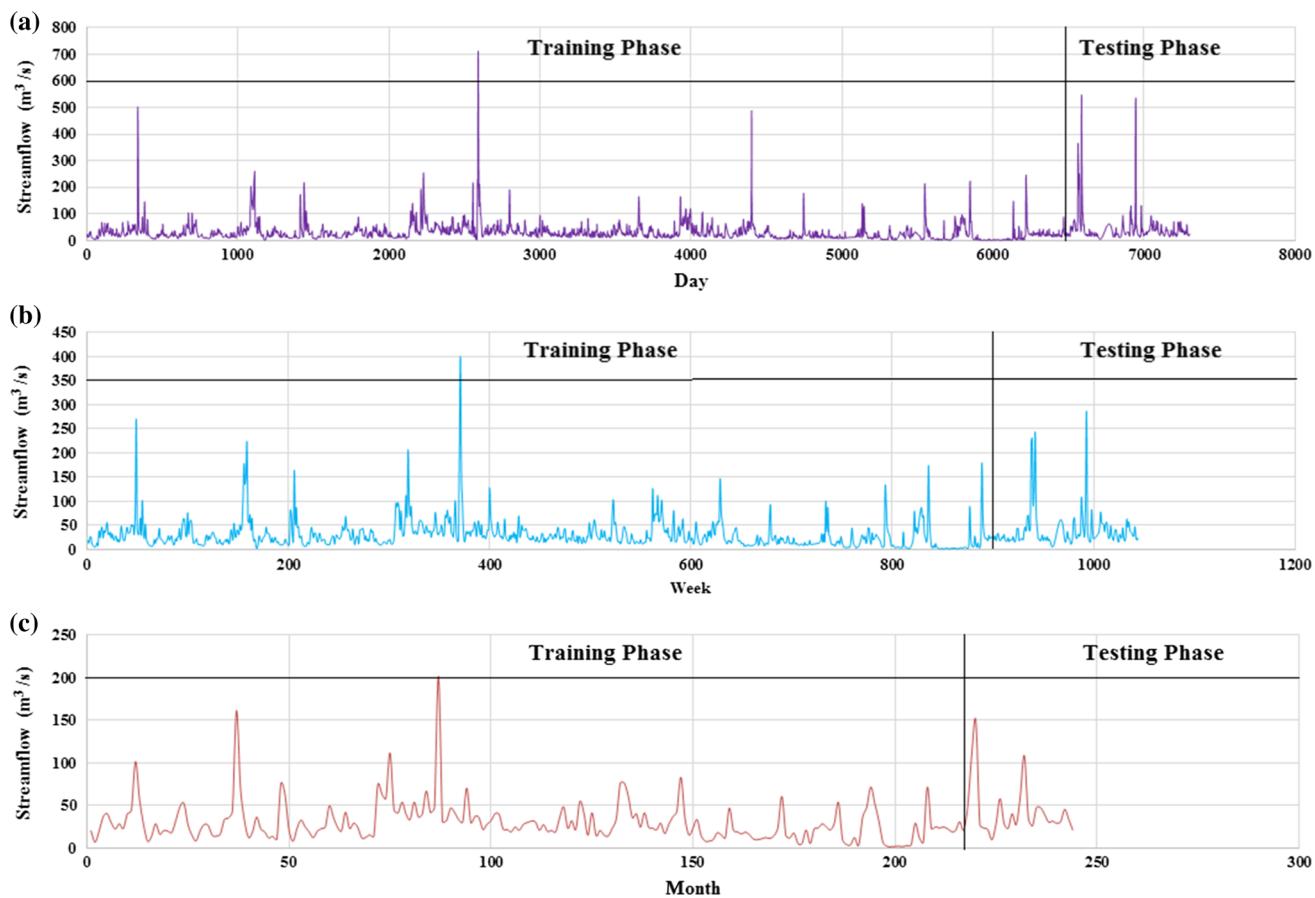


Fig. 4 Actual stream-flow records for Johor River Basin at Rantau Panjang station. **a** Daily. **b** Average weekly. **c** Average monthly

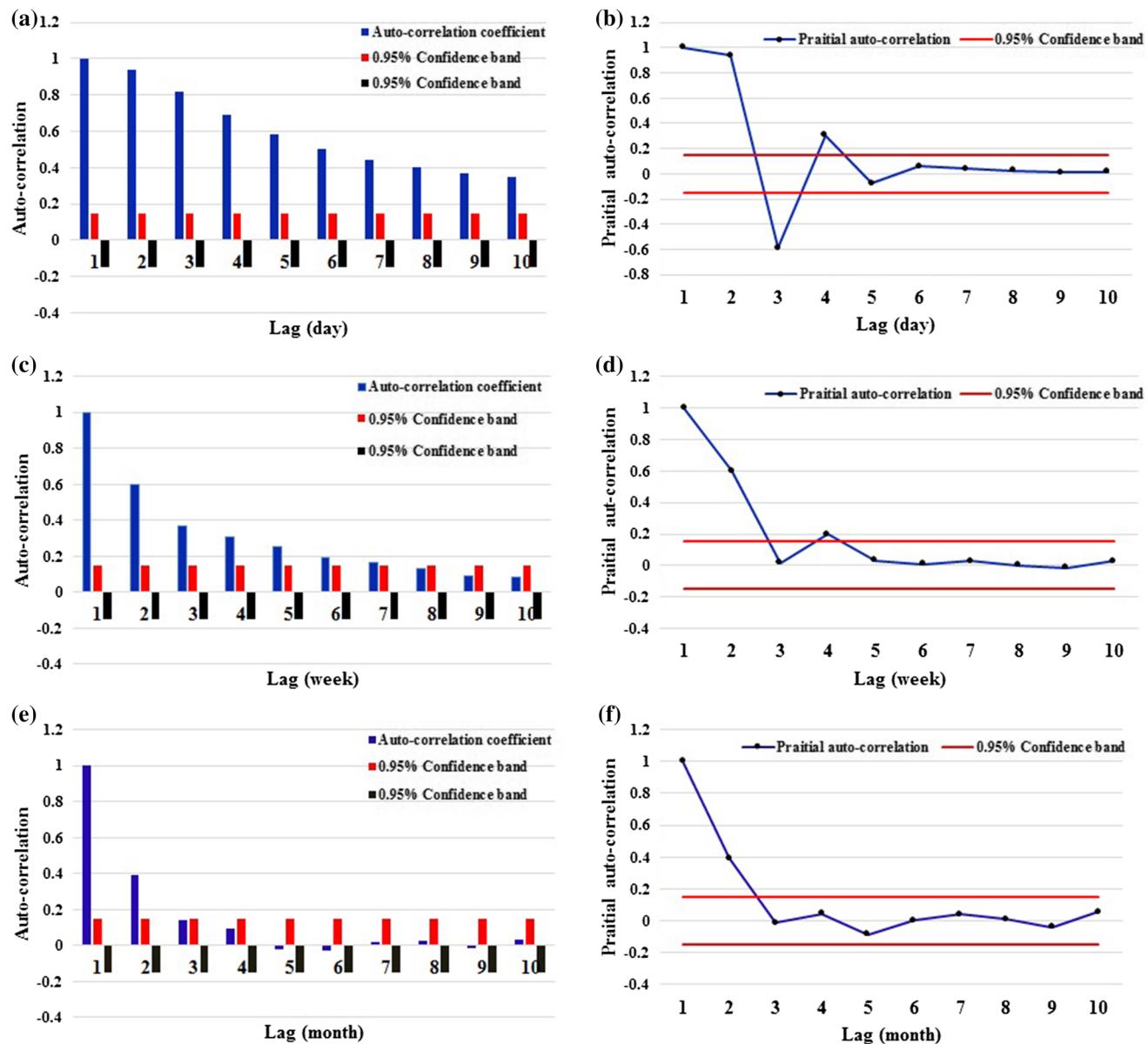


Fig. 5 Plots of ACF and PACF of the stream-flow time series with 95% confidence bounds (the red lines), (a, b) daily, (c, d) average weekly, and (e, f) average monthly

established varies models with various input combinations based on the time scales and lead times of stream-flows. The input combinations of all the time intervals are presented in Table 1. Another significant concern is the data preprocessing, the data were normalized for the purpose of regularity and balance data range. The data were normalized between (0–1) using the following formula:

$$x_{\text{new}} = (x - x_{\min}) / (x_{\max} - x_{\min}) \quad (4)$$

here, x value defines the actual records of the application. x_{\min} represents the minimum value of the data set, and x_{\max} represents the maximum value of the data set.

5 Application and analysis

This section discussed the results of the application of ELM model for training and testing phases comparatively with ANN modeling. The modeling accuracy assessment is presented in terms of the error variation between the observed and the forecasted values. Throughout the study, several performance measures were used for the evaluation purposes [72, 73]. By referring to the established research in evaluation hydrological models, Legates and McCabe (1999) stated in their research that “goodness-of-fit” which exhibits the regression coefficient and absolute error

Table 1 The input combinations for all the time horizons

Time scale	Models	Input sets combinations
Daily	Model 1	$[Q(t-1)]$
	Model 2	$[Q(t-1), Q(t-2)]$
	Model 3	$[Q(t-1), Q(t-2), Q(t-3)]$
	Model 4	$[Q(t-1), Q(t-2), Q(t-3), Q(t-4)]$
Weekly	Model 5	$[Q(t-1)]$
	Model 6	$[Q(t-1), Q(t-2)]$
	Model 7	$[Q(t-1), Q(t-2), Q(t-3), Q(t-4)]$
Monthly	Model 8	$[Q(t-1)]$
	Model 9	$[Q(t-1), Q(t-2)]$

statistical measurements is advisable to inspect the degree of the accuracies [74]. Thus, in this research, coefficient determination (R^2), root-mean-square error (RMSE), mean absolute error (MAE) and relative error (RE) were used to evaluate the performance criteria of the propose approach. The formulas of the mentioned indicators can be expressed as:

$$R^2 = \frac{\sum_{t=1}^n [(S_o - \bar{S}_o)(S_p - \bar{S}_p)]}{\sqrt{\sum_{t=1}^n (S_o - \bar{S}_o)^2 \sum_{t=1}^n (S_p - \bar{S}_p)^2}} \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^n (S_o - S_p)^2} \quad (6)$$

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^n |S_o - S_p| \quad (7)$$

$$\text{RE} = \left[\frac{S_o - S_p}{S_o} \right] 100 \quad (8)$$

The definition of the formulas (5)–(8) variables is S_o (stream-flow observed records), S_p (stream-flow predicted records). \bar{S}_o and \bar{S}_p are the mean values. N is the number of the data set.

It is even worth to brief the formation structure of the modeling before proceeding with the discussion of the results. Since neural networks topology affects the complexity of the computational models and most importantly the level of the accuracies. Remarkably, RBFNN algorithm has been observed to be quite simple compared with the others (i.e. FFNN or MLP). The most significant parameters that should be obtained (as described in Sect. 2) are the spread values and the number of radial basis function. The spread values and the number of RBF are achieved by using trial-and-error procedure until the desired accuracies aim (MSE) is accomplished. This is for the reason that there is no general methodology or guideline to obtain them. The optimum spread values were established (0.35, 0.6, and 0.8) for daily, weekly, and monthly time scale,

respectively, whereas the number of the radial basis function structure was found to be 30 for all the intervals.

After establishing the forecasting models, the performance statistics of the ANN and the ELM models was compared over the training and testing phases. Table 2 indicates the performance indicators assessment of ANN forecasting model including the three time horizons. The best RMSE, MAE and R^2 values were obtained for daily time series forecasting with three lags time. The RMSE, MAE and R^2 values are $7.894 \text{ m}^3/\text{s}$, $0.311 \text{ m}^3/\text{s}$, and 0.914 (or $11.63 \text{ m}^3/\text{s}$, $0.4305 \text{ m}^3/\text{s}$, and 0.9076) for training (or testing) phases, respectively, whereas Table 3 presents the proposed ELM approach, which indicates the best statistical evaluation measures for daily stream-flow forecasting as well. The RMSE, MAE and R^2 values are $2.372 \text{ m}^3/\text{s}$, $0.084 \text{ m}^3/\text{s}$, and 0.967 (or $2.7804 \text{ m}^3/\text{s}$, $0.1029 \text{ m}^3/\text{s}$, and 0.9422) for training (or testing) phases, respectively. However, the results of the proposed approach showed a noticeable enhancement for all the time horizons accuracy and most specifically for daily flow forecasting. Besides, the results indicate that the models training phase performance is better than the testing phase performance. Another remarkable observation, it was expected that according to the statistical methods (i.e., ACF and PACF) that employed to determine dimension of the input vectors combinations. Tables 2 and 3 exhibited the best performance criteria of the models with the domain that determined in advanced. In addition, the best evaluation measures including RMSE, MAE and R^2 were obtained within three antecedents' values for daily time series, two antecedents' values for both weekly and monthly time scale. This observation indicated for both modeling ELM and ANN approaches. Another important observation is the time consuming for the testing period "validation phase". It can be seen a noticeable speed execution in comparison between ELM and ANN models. This remark was reported by [53] that the elapsed time using ELM modeling moderately fast in accordance with its tuning-free mechanism.

For better visualization of the performance accuracy, the forecasted stream-flow by ANN and ELM models are compared by presentable graph (see Fig. 6) with the observed data records. Figure 6 indicates the testing period (2007–2008), which presents the 10% of the whole time series (as mentioned earlier in Sect. 4). Both ANN and ELM forecasts show generally good agreement with the observed stream-flow in this study area, despite for some peak flow events, the two models did not perform very well.

To present the reliability and the effectiveness of the ELM model, we compute the relative error (formula 10) for the extreme events of peak flow for all the intervals. Table 4 shows the peak flow forecasting values for the all time scale using both models (ANN and ELM). From this

Table 2 ANN approach performances evaluation criteria for different time scale “daily, weekly and monthly”

Time scale	Models	Training phase			Testing phase			
		RMSE	MAE	R ²	RMSE	MAE	R ²	Time (s)
Daily	Model 1	10.152	0.425	0.872	12.052	0.671	0.852	20.22
	Model 2	8.472	0.344	0.864	11.811	0.5247	0.856	18.56
	Model 3	7.894*	0.311*	0.914*	11.63*	0.4305*	0.9076*	21.36
	Model 4	10.824	0.524	0.859	12.33	0.692	0.847	16.73
Weekly	Model 5	24.821	2.347	0.864	26.521	2.891	0.847	15.96
	Model 6	25.014*	2.221*	0.897*	25.242*	2.4752*	0.8756*	16.32
	Model 7	27.411	2.642	0.845	27.874	2.974	0.824	14.25
Monthly	Model 8	11.982	0.754	0.892	12.511	1.077	0.866	18.19
	Model 9	11.621*	0.684*	0.917*	12.232*	0.9224*	0.8937*	18.99

* Indicates the best performance of accuracy

Table 3 ELM approach performances evaluation criteria for different time scale “daily, weekly and monthly”

Time scale	Input sets combination	Training phase			Testing phase			
		RMSE	MAE	R ²	RMSE	MAE	R ²	Time (s)
Daily	Model 1	2.893	0.195	0.887	3.221	0.2414	0.864	8.59
	Model 2	2.723	0.187	0.896	3.012	0.2342	0.881	9.74
	Model 3	2.372*	0.084*	0.967*	2.7804*	0.1029*	0.9422*	7.62
	Model 4	3.154	0.262	0.873	3.452	0.2841	0.857	7.91
Weekly	Model 5	25.542	2.463	0.871	26.071	2.862	0.856	6.88
	Model 6	24.826*	2.213*	0.913*	25.224*	2.4734*	0.8805*	8.10
	Model 7	26.347	2.563	0.854	27.311	2.954	0.837	5.11
Monthly	Model 8	11.247	0.523	0.912	11.872	0.662	0.889	7.26
	Model 9	10.678*	0.344*	0.953*	11.524*	0.5443*	0.9304*	4.56

* Indicates the best performance of accuracy

table, the accuracy of the ELM seems to be better than ANN. The maximum daily peak flow is 488.327 m³/s instead of actual record 536.358 m³/s, with an underestimation of 8.955%, while the ANN results is 460.643 m³/s, with an underestimation 14.116%. The ELM forecasting of the maximum average weekly flow, 286.7233 m³/s is 233.47 m³/s, with an underestimation error of 18.573%, while the ANN model yields is 225.347 m³/s, with an underestimation of 21.406%. Finally, the maximum average monthly flow was underestimated by 7.285 and 21.316% regarding ELM and ANN models, respectively. Further observations from the obtained results, it can be seen that ELM model seems to perform better than ANN model for the all interval time series and accordingly displaying a better performance relatively.

Figure 7 shows the results of the scatter diagrams for one-step-ahead stream-flow forecasting using ANN and ELM models for the testing phase. The figure presents the three time horizons forecasted models.

Authors explored a deep and comprehensive detailed analysis between ANN and ELM models, the relative errors distribution (RE, Eq. 10) have been studied over the testing data phase period. Figure 8 demonstrates the

accomplished results for the three time scales stream-flow forecasting using ANN and ELM models. For daily basis, as confirmed in Fig. 8b, the residual error was strongly enhanced over the testing phase comparatively with ANN modeling. The maximum value of RE was decreased up to (30%). For weekly and monthly basis, Fig. 8d, f indicates the distribution of the RE, if we carefully examine those figures, it could be noticed that the pattern of the error using ELM model is similar to the ANN model, but the relative error value is relatively improved.

In the light of the above discussion, it could be remarked in general, and the performance of the daily stream-flow forecasting is outperforming the other time horizons (weekly and monthly). This is due to the sufficiency of historical time series records on the daily basis which provides more information of the nature phenomena of the flow. Thus, the modeling could capture most of the non-linearity of the stream-flow patterns, which provide lower forecasting error. In addition, long-term stream-flow was influenced by several unsystematic hydrological variables that cause an uncertainty in the time series modeling. Furthermore, 123 km long river is short time traveling river flow; hence, modeling weekly or monthly stream-flow is

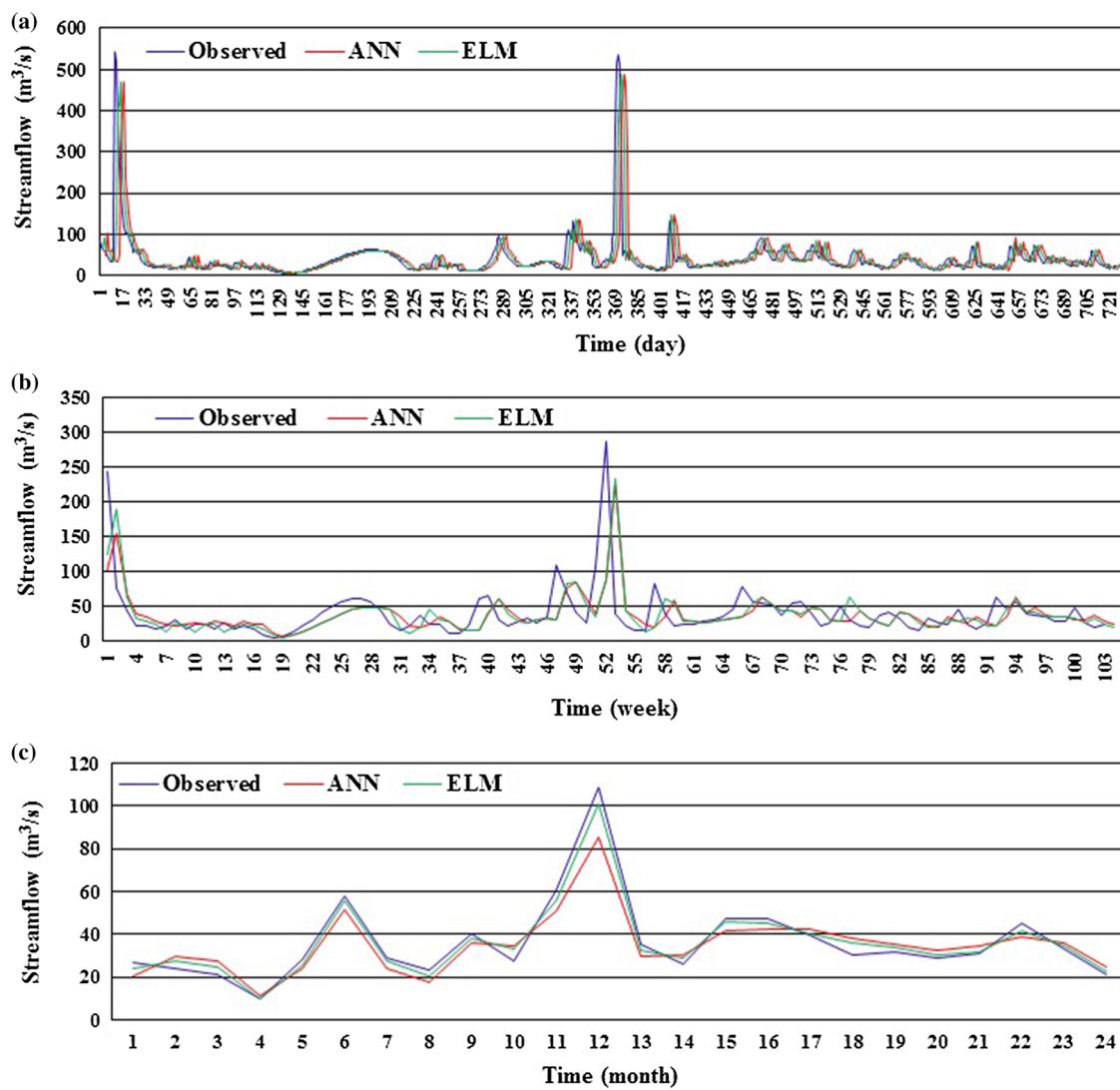


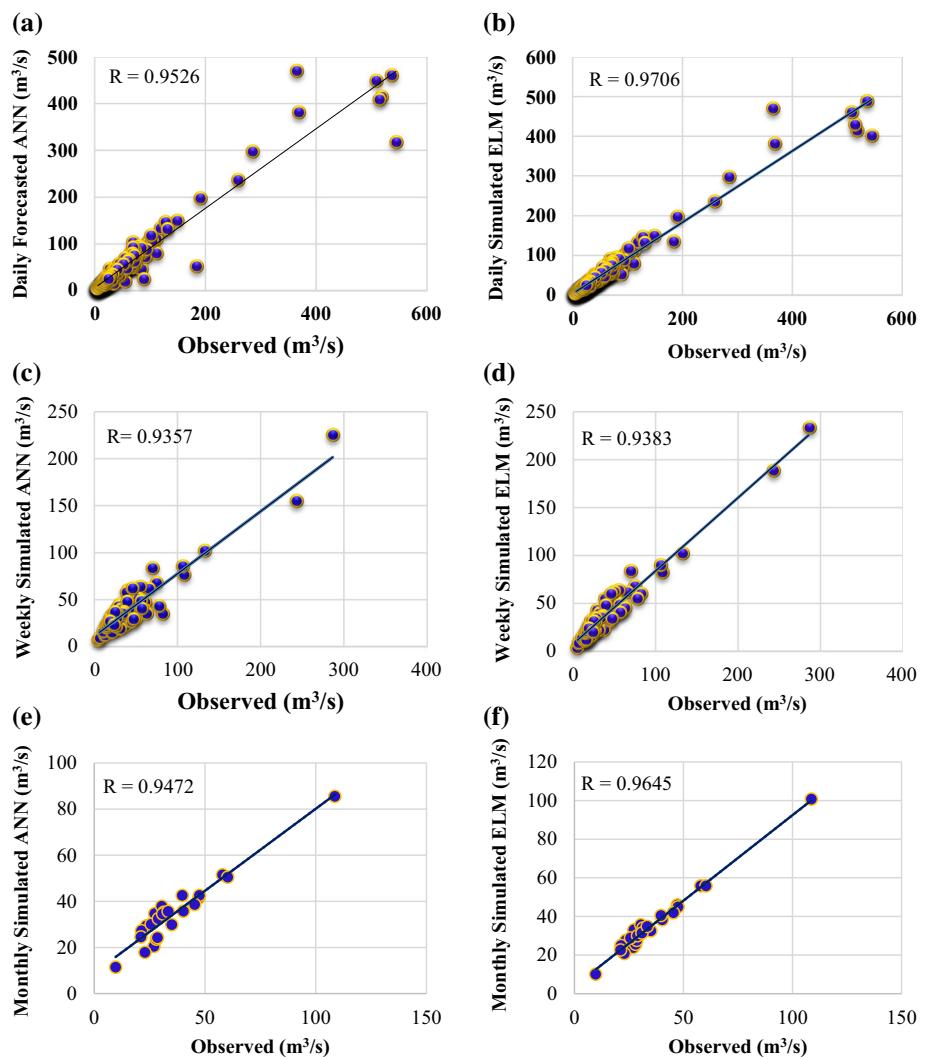
Fig. 6 Comparison between observed and forecasted stream-flow for one-step-ahead (testing phase) using ANN and ELM methods. **a** Daily. **b** Average weekly. **c** Average monthly

Table 4 The comparison of ANN and ELM relative error indicator for the peak flows for the testing phase Johor River, Malaysia

Time scale	Observed peak (m^3/s)	ANN		Relative error %	
		ANN	ELM	ANN	ELM
Daily	544.76	317.337	401.325	-41.74	-26.329
	518.19	390.435	414.55	-24.654	-19.99
	365.11	305.74	330.47	-16.26	-9.48
	368.68	293.86	320.45	-20.29	-13.08
	514.325	408.741	429.421	-20.528	-16.507
	536.358	460.643	488.327	-14.116*	-8.955*
	508.61	448.855	460.248	-11.748	-9.508
	243.26	155.427	188.668	-36.106	-22.441
Weekly	108.19	85.68	92.72	-20.8	-14.29
	106.49	82.52	89.83	-22.5	-15.64
	286.7233	225.347	233.47	-21.406*	-18.573*
	108.634	85.478	100.72	-21.316*	-7.285*

* Indicates the best performance of accuracy

Fig. 7 The observed versus the simulated stream-flow for the testing phase using ANN and ELM models, (a, b) daily, (c, d) average weekly, and (e, f) average monthly, respectively



not accurately performed with the non-existence exogenous predictors (e.g., rainfall, humidity, wind speed, temperature and etc.). Finally, the results of the application exhibited very well harmony of goodness in comparison with most recently researches conducted using non-tuned machine learning approach (i.e., ELM) [61–64].

6 Conclusion and future research

In this study, the accuracy of ELM model has been investigated for forecasting one-step-ahead short-term and long-term stream-flow in tropical environment Johor River, Malaysia. According to the statistical measures (R^2 , RMSE, MAE, and RE) that have been carried out to evaluate the forecasting model, authors conclude that the proposed ELM approach is outperformed the ANN approach. This is much agreeable with several researches that have been conducted in the literature and that ELM approach can yield a much better performance in

comparison with the existing predictive models in stream-flow forecasting. In addition, this investigation establishes modern methodology that offers a very optimistic and positive alternative for the hydrological applications.

Future research efforts should be devoted:

1. Model development that involves the data preprocessing utilizing wavelet transfer [75] or fast orthogonal research [76] might be applied and examine on ELM method for accuracy improvement purposes.
2. Seeking for new computational method as an alternative to compute the Moore–Penrose generalization inverse could be a good step to further research in this field. However, complete orthogonal decomposition method (COD) that proposed by [77], which is characterized by frivolous and reliable alternative to SVD, might give a promising to improve the computation efficiency of ELM approach.
3. External variables that have a correlated or even causal relationship with the stream-flow time series might have

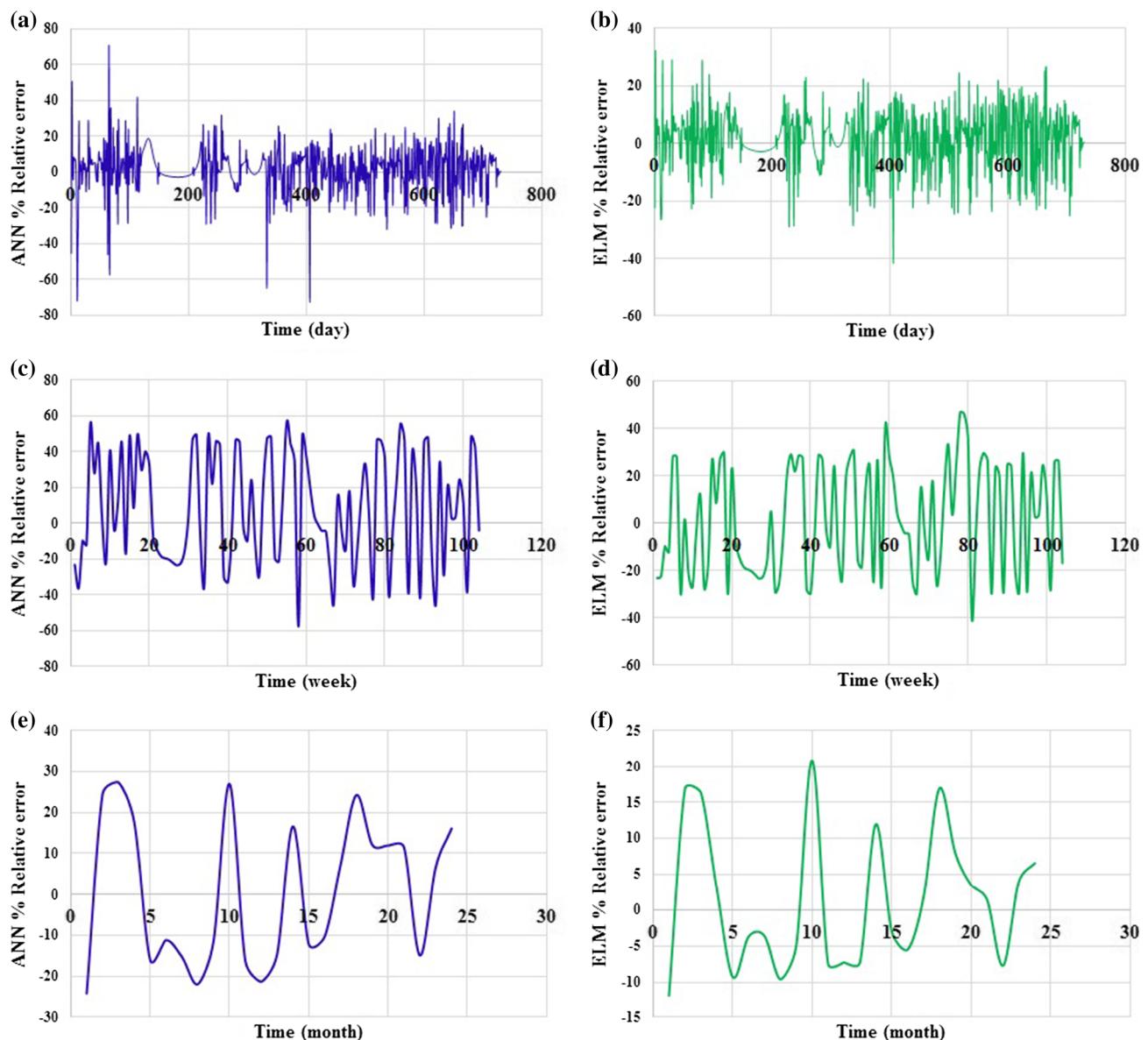


Fig. 8 Relative error distribution for the testing phase using ANN and ELM methods, (a, b) daily, (c, d) average weekly, and (e, f) average monthly, respectively

an essential influence to improve the accuracy of the modeling. For instant, climatological data (e.g., rainfall, humidity, and weather temperature) need to be investigate as an inputs parameter to predict stream-flow.

Acknowledgements Authors would like the acknowledge their gratitude and appreciate for the Department of Irrigation and Drainage (DID), Malaysia, for providing the river flow data set of the studied case study and their admirable cooperation. We are also grateful to the Editor and three anonymous referees for their helpful comments and suggestions.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interests regarding publishing this article.

References

1. Alvisi S, Franchini M (2011) Fuzzy neural networks for water level and discharge forecasting with uncertainty. Environ Model Softw 26:523–537. doi:[10.1016/j.envsoft.2010.10.016](https://doi.org/10.1016/j.envsoft.2010.10.016)
2. Asefa T, Kembrowski M, McKee M, Khalil A (2006) Multi-time scale stream flow predictions: the support vector machines approach. J Hydrol 318:7–16. doi:[10.1016/j.jhydrol.2005.06.001](https://doi.org/10.1016/j.jhydrol.2005.06.001)
3. Ch S, Anand N, Panigrahi BK, Mathur S (2013) Streamflow forecasting by SVM with quantum behaved particle swarm optimization. Neurocomputing 101:18–23. doi:[10.1016/j.neucom.2012.07.017](https://doi.org/10.1016/j.neucom.2012.07.017)
4. Chang FJ, Chen YC (2001) A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. J Hydrol 245:153–164. doi:[10.1016/S0022-1694\(01\)00350-X](https://doi.org/10.1016/S0022-1694(01)00350-X)
5. Maier HR, Kapelan Z, Kasprzyk J et al (2014) Evolutionary algorithms and other metaheuristics in water resources: current

- status, research challenges and future directions. *Environ Model Softw* 62:271–299. doi:[10.1016/j.envsoft.2014.09.013](https://doi.org/10.1016/j.envsoft.2014.09.013)
6. Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ Model Softw* 15:101–124. doi:[10.1016/S1364-8152\(99\)00007-9](https://doi.org/10.1016/S1364-8152(99)00007-9)
 7. Chang TJ (1990) Effect of drought on streamflow characteristics. *Eng J Irrig Drain* 116:332–341
 8. Mohseni O, Stefan HG (1998) A monthly streamflow model. *Water Resour Res* 34:1287–1298. doi:[10.1029/97WR02944](https://doi.org/10.1029/97WR02944)
 9. Sogbedji JM, McIsaac GF (2002) Modeling streamflow from artificially drained agricultural watersheds in Illinois. *J Am Water Resour Assoc* 38:1753–1765. doi:[10.1111/j.1752-1688.2002.tb04379.x](https://doi.org/10.1111/j.1752-1688.2002.tb04379.x)
 10. Bourdin DR, Fleming SW, Stull RB (2012) Streamflow modelling: a primer on applications, approaches and challenges. *Atmos Ocean* 50:507–536. doi:[10.1080/07055900.2012.734276](https://doi.org/10.1080/07055900.2012.734276)
 11. Yaseen ZM, Kisi O, Demir V (2016) Enhancing long-term streamflow forecasting and predicting using periodicity data component: application of artificial intelligence. *Water Resour Manag*. doi:[10.1007/s11269-016-1408-5](https://doi.org/10.1007/s11269-016-1408-5)
 12. Box GEP, Jenkins GM (1970) Time series analysis, forecasting and control, 1st edn. Holden-Day, San Francisco
 13. Salas JD (1980) Applied modeling of hydrologic time series. Water Resources Publication, Littleton CO
 14. Valipour M, Banihabib ME, Behbahani SMR (2013) Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *J Hydrol* 476:433–441. doi:[10.1016/j.jhydrol.2012.11.017](https://doi.org/10.1016/j.jhydrol.2012.11.017)
 15. Valipour M, Banihabib M, Behbahani S (2012) Monthly inflow forecasting using autoregressive artificial neural network. *J Appl Sci* 12:2139–2147
 16. Valipour M (2015) Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorol Appl* n/a-n/a. doi:[10.1002/met.1491](https://doi.org/10.1002/met.1491)
 17. Hsu K, Gupta HV, Gao X et al (2002) Self-organizing linear output map (SOLO): an artificial neural network suitable for hydrologic modeling and analysis. *Water Resour Res*. doi:[10.1029/2001WR000795](https://doi.org/10.1029/2001WR000795)
 18. Cigizoglu HK (2005) Application of generalized regression neural networks to intermittent flow forecasting and estimation. *J Hydrol Eng* 10:336–341. doi:[10.1061/\(ASCE\)1084-0699\(2005\)10:4\(336\)](https://doi.org/10.1061/(ASCE)1084-0699(2005)10:4(336))
 19. Wu JS, Han J, Annambhotla S, Bryant S (2005) Artificial neural networks for forecasting watershed runoff and stream flows. *J Hydrol Eng* 10:216–222. doi:[10.1061/\(ASCE\)1084-0699\(2005\)10:3\(216\)](https://doi.org/10.1061/(ASCE)1084-0699(2005)10:3(216))
 20. Kiş Ö (2007) Streamflow forecasting using different artificial neural network algorithms. *J Hydrol Eng* 12:532–539. doi:[10.1061/\(ASCE\)1084-0699\(2007\)12:5\(532\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:5(532))
 21. Ahmed JA, Sarma AK (2007) Artificial neural network model for synthetic streamflow generation. *Water Resour Manag* 21:1015–1029. doi:[10.1007/s11269-006-9070-y](https://doi.org/10.1007/s11269-006-9070-y)
 22. Kagoda PA, Ndiritu J, Ntuli C, Mwaka B (2010) Application of radial basis function neural networks to short-term streamflow forecasting. *Phys Chem Earth* 35:571–581. doi:[10.1016/j.pce.2010.07.021](https://doi.org/10.1016/j.pce.2010.07.021)
 23. Yonaba H, Anctil F, Fortin V (2010) Comparing sigmoid transfer functions for neural network multistep ahead streamflow forecasting. *J Hydrol Eng* 15:275–283. doi:[10.1061/\(ASCE\)HE.1943-5584.0000188](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000188)
 24. Dibike Yonas B, Velickov Slavco, Solomatine Dimitri, Abbott MB (2001) Model induction with support vector machines: introduction and applications. *J Comput Civ Eng* 15:208–216
 25. Behzad M, Asghari K, Eazi M, Palhang M (2009) Generalization performance of support vector machines and neural networks in runoff modeling. *Expert Syst Appl* 36:7624–7629. doi:[10.1016/j.eswa.2008.09.053](https://doi.org/10.1016/j.eswa.2008.09.053)
 26. Noori R, Karbassi AR, Moghaddamia A et al (2011) Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. *J Hydrol* 401:177–189. doi:[10.1016/j.jhydrol.2011.02.021](https://doi.org/10.1016/j.jhydrol.2011.02.021)
 27. Kalra A, Ahmad S (2009) Using oceanic-atmospheric oscillations for long lead time streamflow forecasting. *Water Resour Res* 45:1–18. doi:[10.1029/2008WR006855](https://doi.org/10.1029/2008WR006855)
 28. 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. *J Hydrol* 509:379–386. doi:[10.1016/j.jhydrol.2013.11.054](https://doi.org/10.1016/j.jhydrol.2013.11.054)
 29. El-Shafie A, Taha MR, Noureldin A (2007) A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. *Water Resour Manag* 21:533–556
 30. Nayak PC, Sudheer KP, Jain SK (2007) Rainfall-runoff modeling through hybrid intelligent system. *Water Resour Res* 43:1–17. doi:[10.1029/2006WR004930](https://doi.org/10.1029/2006WR004930)
 31. Pramanik N, Panda RK (2009) Application of neural network and adaptive neuro-fuzzy inference systems for river flow prediction. *Hydrol Sci J* 54:247–260. doi:[10.1623/hysj.54.2.247](https://doi.org/10.1623/hysj.54.2.247)
 32. Sanikhani H, Kisi O (2012) River flow estimation and forecasting by using two different adaptive neuro-fuzzy approaches. *Water Resour Manag* 26:1715–1729. doi:[10.1007/s11269-012-9982-7](https://doi.org/10.1007/s11269-012-9982-7)
 33. Sharma S, Srivastava P, Fang X, Kalin L (2015) Performance comparison of Adoptive Neuro Fuzzy Inference System (ANFIS) with Loading Simulation Program C ++ (LSPC) model for streamflow simulation in El Niño Southern Oscillation (ENSO)-affected watershed. *Expert Syst Appl* 42:2213–2223. doi:[10.1016/j.eswa.2014.09.062](https://doi.org/10.1016/j.eswa.2014.09.062)
 34. Whigham PA, Crapper PF (2001) Modelling rainfall-runoff using genetic programming. *Math Comput Model* 33:707–721. doi:[10.1016/S0895-7177\(00\)00274-0](https://doi.org/10.1016/S0895-7177(00)00274-0)
 35. Makkeasorn A, Chang NB, Zhou X (2008) Short-term streamflow forecasting with global climate change implications—a comparative study between genetic programming and neural network models. *J Hydrol* 352:336–354. doi:[10.1016/j.jhydrol.2008.01.023](https://doi.org/10.1016/j.jhydrol.2008.01.023)
 36. Guven A (2009) Linear genetic programming for time-series modelling of daily flow rate. *J Earth Syst Sci* 118:137–146. doi:[10.1007/s12040-009-0022-9](https://doi.org/10.1007/s12040-009-0022-9)
 37. Kashid SS, Ghosh S, Maity R (2010) Streamflow prediction using multi-site rainfall obtained from hydroclimatic teleconnection. *J Hydrol* 395:23–38. doi:[10.1016/j.jhydrol.2010.10.004](https://doi.org/10.1016/j.jhydrol.2010.10.004)
 38. Kisi O, Shiri J, Tombul M (2013) Modeling rainfall-runoff process using soft computing techniques. *Comput Geosci* 51:108–117. doi:[10.1016/j.cageo.2012.07.001](https://doi.org/10.1016/j.cageo.2012.07.001)
 39. Sahay RR, Srivastava A (2014) Predicting monsoon floods in rivers embedding wavelet transform, genetic algorithm and neural network. *Water Resour Manag* 28:301–317. doi:[10.1007/s11269-013-0446-5](https://doi.org/10.1007/s11269-013-0446-5)
 40. Nourani V, Kisi Ö, Komasi M (2011) Two hybrid Artificial Intelligence approaches for modeling rainfall-runoff process. *J Hydrol* 402:41–59. doi:[10.1016/j.jhydrol.2011.03.002](https://doi.org/10.1016/j.jhydrol.2011.03.002)
 41. Danandeh Mehr A, Kahya E, Bagheri F, Deliktas E (2013) Successive-station monthly streamflow prediction using neuro-wavelet technique. *Earth Sci Inform*. doi:[10.1007/s12145-013-0141-3](https://doi.org/10.1007/s12145-013-0141-3)
 42. Kisi O, Cimen M (2011) A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *J Hydrol* 399:132–140. doi:[10.1016/j.jhydrol.2010.12.041](https://doi.org/10.1016/j.jhydrol.2010.12.041)

43. Pramanik N, Panda RK, Singh A (2010) Daily river flow forecasting using wavelet ANN hybrid models. *J Hydroinformatics* 13:49. doi:[10.2166/hydro.2010.040](https://doi.org/10.2166/hydro.2010.040)
44. Adamowski J, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J Hydrol* 390:85–91. doi:[10.1016/j.jhydrol.2010.06.033](https://doi.org/10.1016/j.jhydrol.2010.06.033)
45. Babovic V, Keijzer M (2002) Rainfall runoff modelling based on genetic programming. *Nord Hydrol* 33:331–346
46. Kashani MH, Ghorbani MA, Dinpashoh Y, Shahmorad S (2016) Integration of Volterra model with artificial neural networks for rainfall-runoff simulation in forested catchment of northern Iran. *J Hydrol* 540:340–354. doi:[10.1016/j.jhydrol.2016.06.028](https://doi.org/10.1016/j.jhydrol.2016.06.028)
47. Fayaed S, El-Shafie A, Jaafar O (2013) Integrated artificial neural network (ANN) and stochastic dynamic programming (SDP) model for optimal release policy. *Water Resour Manag* 27:3679–3696. doi:[10.1007/s11269-013-0373-5](https://doi.org/10.1007/s11269-013-0373-5)
48. Soria-Olivas E, Gómez-Sanchis J, Martín JD et al (2011) BELM: Bayesian extreme learning machine. *IEEE Trans Neural Networks* 22:505–509. doi:[10.1109/TNN.2010.2103956](https://doi.org/10.1109/TNN.2010.2103956)
49. Chang F-J, Chen P-A, Lu Y-R et al (2014) Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. *J Hydrol* 517:836–846. doi:[10.1016/j.jhydrol.2014.06.013](https://doi.org/10.1016/j.jhydrol.2014.06.013)
50. Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70:489–501. doi:[10.1016/j.neucom.2005.12.126](https://doi.org/10.1016/j.neucom.2005.12.126)
51. Nourani V, Sayyah Fard M (2012) Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. *Adv Eng Softw* 47:127–146. doi:[10.1016/j.advengsoft.2011.12.014](https://doi.org/10.1016/j.advengsoft.2011.12.014)
52. Huang G-B, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern B Cybern* 42:513–529. doi:[10.1109/TSMCB.2011.2168604](https://doi.org/10.1109/TSMCB.2011.2168604)
53. Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70:489–501. doi:[10.1016/j.neucom.2005.12.126](https://doi.org/10.1016/j.neucom.2005.12.126)
54. Abdullah SS, Malek MA, Abdullah NS et al (2015) Extreme Learning Machines: a new approach for prediction of reference evapotranspiration. *J Hydrol* 527:184–195. doi:[10.1016/j.jhydrol.2015.04.073](https://doi.org/10.1016/j.jhydrol.2015.04.073)
55. Samat A, Du P, Member S et al (2014) Ensemble extreme learning machines for hyperspectral image classification. *Sel Top Appl Earth Obs Remote Sensing, IEEE J* 7:1060–1069
56. Bencherif MA, Bazi Y, Member S et al (2015) Fusion of extreme learning machine and graph-based optimization methods for active classification of remote sensing images. *Geosci Remote Sens Lett IEEE* 12:527–531
57. Lian C, Zeng Z, Yao W, Tang H (2012) Displacement prediction model of landslide based on ensemble of extreme learning machine. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 7666 LNCS:240–247. doi:[10.1007/978-3-642-34478-7_30](https://doi.org/10.1007/978-3-642-34478-7_30)
58. Sun Z-L, Choi T-M, Au K-F, Yu Y (2008) Sales forecasting using extreme learning machine with applications in fashion retailing. *Decis Support Syst* 46:411–419. doi:[10.1016/j.dss.2008.07.009](https://doi.org/10.1016/j.dss.2008.07.009)
59. Bhat AU, Merchant SS, Bhagwat SS (2008) Prediction of melting points of organic compounds using extreme learning machines. *Ind Eng Chem Res* 47:920–925. doi:[10.1021/ie0704647](https://doi.org/10.1021/ie0704647)
60. Wang B, Huang S, Qiu J et al (2015) Parallel online sequential extreme learning machine based on MapReduce. *Neurocomputing* 149:224–232. doi:[10.1016/j.neucom.2014.03.076](https://doi.org/10.1016/j.neucom.2014.03.076)
61. Li BJ, Cheng CT (2014) Monthly discharge forecasting using wavelet neural networks with extreme learning machine—Springer. *Sci China Technol Sci* 57:2441–2452. doi:[10.1007/s11431-014-5712-0](https://doi.org/10.1007/s11431-014-5712-0)
62. Deo RC, Şahin M (2016) An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. *Environ Monit Assess*. doi:[10.1007/s10661-016-5094-9](https://doi.org/10.1007/s10661-016-5094-9)
63. Lima AR, Cannon AJ, Hsieh WW (2016) Forecasting daily streamflow using online sequential extreme learning machines. *J Hydrol* 537:431–443. doi:[10.1016/j.jhydrol.2016.03.017](https://doi.org/10.1016/j.jhydrol.2016.03.017)
64. Yaseen ZM, Jaafar O, Deo RC et al (2016) Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. *J Hydrol*. doi:[10.1016/j.jhydrol.2016.09.035](https://doi.org/10.1016/j.jhydrol.2016.09.035)
65. Moradkhani H, Hsu K, Gupta HV, Sorooshian S (2004) Improved streamflow forecasting using self-organizing radial basis function artificial neural networks. *J Hydrol* 295:246–262. doi:[10.1016/j.jhydrol.2004.03.027](https://doi.org/10.1016/j.jhydrol.2004.03.027)
66. Mehr AD, Kahya E, Şahin A, Nazemosadat MJ (2014) Successive-station monthly streamflow prediction using different artificial neural network algorithms. *Int J Environ Sci Technol*. doi:[10.1007/s13762-014-0613-0](https://doi.org/10.1007/s13762-014-0613-0)
67. Yaseen ZM, El-Shafie A, Afan HA et al (2015) RBFNN versus FFNN for daily river flow forecasting at Johor River. *Neural Comput Appl, Malaysia*. doi:[10.1007/s00521-015-1952-6](https://doi.org/10.1007/s00521-015-1952-6)
68. He L, Huang GH, Lu HW (2008) A simulation-based fuzzy chance-constrained programming model for optimal groundwater remediation under uncertainty. *Adv Water Resour* 31:1622–1635. doi:[10.1016/j.advwatres.2008.07.009](https://doi.org/10.1016/j.advwatres.2008.07.009)
69. Bin Huang G, Chen L (2007) Convex incremental extreme learning machine. *Neurocomputing* 70:3056–3062. doi:[10.1016/j.neucom.2007.02.009](https://doi.org/10.1016/j.neucom.2007.02.009)
70. Elzwayie A, El-shafie A, Yaseen ZM et al (2016) RBFNN-based model for heavy metal prediction for different climatic and pollution conditions. *Neural Comput Appl*. doi:[10.1007/s00521-015-2174-7](https://doi.org/10.1007/s00521-015-2174-7)
71. Afan HA, El-Shafie A, Yaseen ZM et al (2014) ANN based sediment prediction model utilizing different input scenarios. *Water Resour Manag* 29:1231–1245. doi:[10.1007/s11269-014-0870-1](https://doi.org/10.1007/s11269-014-0870-1)
72. Yaseen ZM, El-shafie A, Jaafar O et al (2015) Artificial intelligence based models for stream-flow forecasting: 2000–2015. *J Hydrol* 530:829–844. doi:[10.1016/j.jhydrol.2015.10.038](https://doi.org/10.1016/j.jhydrol.2015.10.038)
73. Nourani V, Hosseini Baghanam A, Adamowski J, Kisi O (2014) Applications of hybrid wavelet-artificial intelligence models in hydrology: a review. *J Hydrol* 514:358–377. doi:[10.1016/j.jhydrol.2014.03.057](https://doi.org/10.1016/j.jhydrol.2014.03.057)
74. Legates DR Jr, McCabe GJ (1999) Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 35:233–241
75. Grossmann A, Morlet J (1984) Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM J Math Anal* 15:723–736. doi:[10.1137/0515056](https://doi.org/10.1137/0515056)
76. Korenberg MJ (1989) A robust orthogonal algorithm for system identification and time-series analysis. *Biol Cybern* 60:267–276. doi:[10.1007/BF00204124](https://doi.org/10.1007/BF00204124)
77. Hough PD, Vavasis SA (1997) Complete orthogonal decomposition for weighted least squares. *SIAM J Matrix Anal Appl* 18:369–392