Improving capability of conceptual modeling of watershed rainfall–runoff using hybrid wavelet-extreme learning machine approach  
Kiyoumars Roushangar, Farhad Alizadeh and Vahid Nourani

ABSTRACT

Rainfall–runoff process identification, due to uncertainties and complexities, requires advanced modeling strategies. For this end, this study presented different strategies to explore spatio-temporal variation of rainfall–runoff process for the Ajichay watershed located in northwest Iran. Extreme learning machine (ELM) was used to predict the runoff in conceptual models. First, a geomorphology integrated ELM (G-ELM) was used to predict watershed runoff in multiple-stations form for the watershed. The spatial and temporal features of sub-basins were selected as input data wherein temporal features were pre-processed by wavelet transform (WT). Results confirmed the capability of G-ELM in successive prediction of watershed runoff. Afterwards, an integrated ELM (I-ELM) was developed based on conceptual reservoir modeling to predict monthly river runoff where the model had the semi-distributed specifications of ELM. This model was capable of exploring spatial variation of rainfall–runoff process without requiring physical characteristics of sub-basins. To meter sufficiency of the modeling strategies, cross-validation technique was performed for station 3 in which G-ELM performed better in comparison to I-ELMs. Furthermore, classic and wavelet-based modeling (W-ELM) of rainfall–runoff was performed for one-step-ahead predictions. Statistical evaluations confirmed the W-ELM, I-ELM, and G-ELM performance, respectively.

Key words | Ajichay watershed, conceptual modeling strategies, extreme learning machine, rainfall–runoff process, spatio-temporal features, wavelet transform

INTRODUCTION

Water management policy is a crucial global issue intended to govern water acquisition and determine proper ways to minimize the misapplication and spoil of water resources. River discharge is very important in the hydrological cycle and hydro-environmental management, and plays a basic designation in designing and operating irrigation schemes at a watershed scale. One of the most important problems in analysis and management of river discharge is the modeling/simulation. For this purpose, the complexity and uncertainty of the process has been widely analyzed using artificial intelligences (AI) by hydrologists and scientists.

AI approaches such as artificial neural networks (ANN) are black box modeling approaches that have been applied in a variety of areas of hydrology, including rainfall–runoff modeling (e.g., Hsu et al. 1995; Coulibaly et al. 1999; Anmala et al. 2000; Tokar & Markus 2000; Nilsson et al. 2006; Aqil et al. 2007). Likewise, machine learning approaches as AI approaches, are remarkable prediction approaches which in recent years have been implemented in the river modeling field (e.g., Azamathulla et al. 2009; Chen et al. 2010; Shiri & Kisi 2010; Sun et al. 2012; Roushangar et al. 2014a, 2014b, 2016a, 2016b; Roushangar & Alizadeh
2015; Roushangar & Koosheh 2015; Nourani et al. 2016). Also, some researchers have carried out reviews on good state-of-the-art application of AI and their advantages (Maier & Dandy 2000; Labat 2005; Abrahart et al. 2012; Sang 2013; Nourani et al. 2014). The extreme learning machines (ELMs) as a new AI approach is a fast learning technique with high generalization performance that basically uses single-hidden layer feed forward neural networks (SLFNNs) (Huang et al. 2006). This technique has been successfully applied in various fields of research (Abdullah et. al. 2015; Lima et al. 2016).

Likewise, wavelet transform (WT) provides remarkable vision into the physical form of the data by presenting information in both time and frequency domains of the time series (Farajzadeh & Alizadeh 2017; Roushangar & Alizadeh 2017). The WT is an appropriate temporal pre-processing method that can be employed to extract the diversity of features from the time series, such as short-term and long-term fluctuations, by decomposing the time series into different sub-components. A time series in the WT breaks down into a series of linearly independent detail signals and one approximation signal by using discrete WT with a specific wavelet function (Foufoula-Georgiou & Kumar 1995). Mallat (1998) presented a complete theory for wavelet multi-resolution signal decomposition (also mentioned as a pyramid decomposition algorithm). Researches confirmed that proper data pre-processing by applying the WT can lead the models to adequately illustrate the real specifications of the basic system. WT decomposes a non-stationary signal into a given quantity of stationary sub-signals. Then, AI approaches can be combined with WT to improve preciseness of the prediction. Precise hybrid models have been employed in recent years to forecast hydrological and hydrogeological processes (e.g., Partal 2009; Adamowski & Sun 2010; Tiwari & Chatterjee 2010).

In most former AI-based hydrological modeling techniques, temporal variables (as historical time series) of the process were used as inputs of the model to predict runoff (Anmala et al. 2000). Furthermore, in some studies, the AI-based and conceptual models have been combined to estimate river flow (Tokar & Markus 2000; Ashu & Sanaga 2006; Corzo et al. 2009; Huo et al. 2012). These models which combine conceptual hydrological models and AI-based models usually include two parts: one is a conceptual model for runoff of every sub-catchment and the other is the AI-based model to ensemble the runoff from each catchment. Nourani et al. (2013) developed a neuro-fuzzy-based geomorphological approach to handle the Eel River watershed runoff. The quality of the results supported the beneficial application of the approach in spatio-temporal modeling of hydrological processes. Multi-station modeling is a conceptual approach whose capability has been confirmed and, in some cases, its imposing geomorphological features caused an improvement in results (Zhang & Govindaraju 2003; Sarangi et al. 2005). Nourani et al. (2016) proposed a spatio-temporal framework in which SVM used conjugate geostatistics to monitor Ajichay watershed runoff-suspended sediment load process.

The objectives of present study are to: (1) use the classic modeling strategy and wavelet-based ELM (W-ELM) to predict one-step-ahead runoff for all hydrometric stations; (2) propose a WT-based spatio-temporal model to forecast the Ajichay watershed runoff with an ELM approach to improve the model ability; and (3) propose an integrated ELMs model conjugated to WT to forecast runoff and catch peak values of the process. Results of the proposed models were compared by using cross-validation technique for station 3 to find the suitability of the models. Finally, results of ELM networks and ANNs were compared to analyze the ELM’s capability.

MATERIALS AND METHODS

ELM for SLFNNs

Learning principle

ELM is a recently developed learning scheme for SLFFN, proposed by Huang et al. (2004). Almost all learning algorithms for SLFFNs require adjustment of parameters that results in dependence between different layers of parameters like weights and biases. Therefore, many iterative tuning steps are required by traditional learning algorithms (Huang et al. 2006, 2012) whereas the ELM algorithm avoids slow iterative learning procedure and only requires a one-pass operation to learn SLFFNs. This is mainly due to the fact that there is no bias for the output layer.
neuron. In brief, to initiate one-pass learning operation, the hidden node network parameters (weights and biases) are randomly generated without any prior knowledge or training procedure. Consequently, the ELM turns into a system of linear equations and the unknown weights between the hidden and output layer nodes can be obtained analytically by only applying Moore–Penrose generalized inverse procedure. The learning scheme of ELM can be summarized in three steps (Algorithm 1).

**Algorithm 1.** Learning scheme of an ELM

Assume
- \( n \) inputs, \( m \) outputs, \( \tilde{N} \) hidden nodes (\( k = 1 \ldots \tilde{N} \))

**Require**
- \( N \) learning data samples \((x_i, t_i)\) (\( i = 1 \ldots N \))

\[ x_i = [x_{i1}, \ldots, x_{in}]^T \in \mathbb{R}^n, \quad t_i = [t_{i1}, \ldots, t_{im}]^T \in \mathbb{R}^m \]

1. Randomly assign parameters of hidden nodes, i.e., weights and bias \((w_h, b_{sh})\).
2. Obtain the hidden layer output matrix \( H \).
3. Find the output weight matrix \( \beta; \beta = H^T \), where \( H^T \) represents the Moore–Penrose generalized inverse solution for the hidden layer output matrix \( H \).

**Artificial neural network**

ANN as a black box model has been widely applied to simulate and forecast real world phenomena. The feed forward neural network (FFNN) as a special kind of ANN has been widely applied in water resource and hydrology studies as a time series forecasting approach. It has already been proved that a FFNN model trained by the back-propagation (BP) algorithm including three layers is satisfactory for simulating and forecasting hydro-environmental processes (Hornik et al. 1989; ASCE 2000). Three-layered FFNNs (one hidden layer), usually used in time series forecasting, provide a general framework for indicating nonlinear functional relationship between input and output parameters. Three-layered FFNN is created based on a linear combination of the input parameters, then transformed by a nonlinear activation function. In an ANN scheme, \( i, j, \) and \( k \) denote, respectively, the input, hidden, and output layer neurons, and \( w \) is the applied weight by the neuron. The explicit equation to compute the output value of FFNNs is given by:

\[
\hat{y}_k = f_0 \left[ \sum_{i=1}^{m} w_{kj} f_h \left( \sum_{l=1}^{n} w_{jl} x_l + w_{jo} \right) + w_{ko} \right]
\]  

(1)

where \( w_{ji} \) is weight in the hidden layer connecting \( i \)th neuron in the input layer to \( j \)th neuron in the hidden layer; \( w_{jk} \) is bias of \( j \)th hidden neuron; \( f_h \) is activation function of the hidden neuron; \( w_{kj} \) is weight in the output layer connecting \( j \)th neuron in the hidden layer to \( k \)th neuron in the output layer; \( w_{ko} \) is bias of \( k \)th output neuron; \( f_0 \) is activation function of the output neuron; \( x_i \) is \( i \)th input variable in the input layer; and \( \hat{y}, y \) are, respectively, computed and observed output variables. The weights in the hidden and output layers, and their values, can be altered during the calibration process of the network. The FFNN structure used in this study includes input layer, hidden layer (only one hidden layer), and output layer.

**Wavelet transform**

The WT has enlarged in application and prevalence in recent years since its establishment in the early 1980s, but widespread usage of the Fourier transform has yet to occur (Grossmann & Morlet 1984). Fourier analysis has a crucial disadvantage. In transforming to the frequency domain, time information is lost. Wavelet analysis allows the use of long time intervals where more precise low-frequency information and shorter regions are necessary when high-frequency information is wanted. The main property of WT is to provide a time-scale localization of the process, which derives from the compact support of its basic function. The WT searches for correlations between the signal and wavelet function. In real hydrological problems, the time series is usually in the discrete format rather than continuous and, therefore, discrete WT in the following form is usually used (Mallat 1998):

\[
g_{m,n}(t) = \frac{1}{\sqrt{a_0^n}} g \times \left( \frac{t - nb_0a_0^m}{a_0^n} \right)
\]  

(2)
where m and n are integers that control the wavelet dilation and translation respectively; \( a_0 \) is a specified fine dilation step greater than 1; and \( b_0 \) is the location parameter which must be greater than zero. The most common and simplest choice for parameters are \( a_0 = 2 \) and \( b_0 = 1 \). This power-of-two logarithmic scaling of the dilation and translation is known as the dyadic grid arrangement. The dyadic wavelet can be written in more compact notation as (Mallat 1998):

\[
g_{m,n}(t) = 2^{-m/2} g(2^{-m}t - n) \tag{3}
\]

For a discrete time series, \( x_i \), the dyadic WT becomes (Mallat 1998):

\[
T_{m,n}(t) = 2^{-m/2} \sum_{i=0}^{N-1} g(2^{-m}i - n)x_i \tag{4}
\]

where \( T_{m,n} \) is wavelet coefficient for the discrete wavelet of scale \( a = 2^m \) and location \( b = 2^m n \). Equation (4) considers a finite time series, \( x_i, i = 0, 1, 2, \ldots, N_i \); and \( N \) is an integer power of 2: \( N = 2M \). This gives the ranges of \( m \) and \( n \) as, \( 0 < n < 2^{M-m-1} \) and \( 1 < m < M \), respectively. The inverse discrete transform is given by Mallat (1998):

\[
x_i = T + \sum_{m=1}^{M} \sum_{n=0}^{2^m-1} T_{m,n} 2^{-m/2} g(2^{-m}i - n) \tag{5}
\]

Or in a simple format as (Mallat 1998):

\[
x_i = T(t) + \sum_{m=1}^{M} W_m(t) \tag{6}
\]

in which \( T(t) \) is called approximation sub-signal at level \( M \) and \( W_m(t) \) are details sub-signals at levels \( m = 1, 2, \ldots, M \). The wavelet coefficients, \( W_m(t) \) \( (m = 1, 2, \ldots, M) \), provide the detail signals, which can capture small features of inter-pretational value in the data; the residual term, \( T(t) \), represents the background information of data.

### Strategies of modeling

With respect to the reviewed literature, most issues about runoff prediction could be related to the AI application. AI-based models which are fed only by temporal features cannot predict/interpolate the variable of interest in an ideal point across the watershed accurately. Such an AI-based model may not be able to cover the natural uncertainty of the hydrological process. Furthermore, the rainfall–runoff data include a broad domain of values and fluctuations; to resolve the mentioned deficiencies, in this research two strategies in rainfall–runoff modeling were proposed. Geomorphology integrated ELM (G-ELM) and integrated ELMs (I-ELMs) conjugated to WT are the strategies of modeling which are going to be discussed. In order to develop the models, m files were coded in the MATLAB environment.

### Geomorphology-based ELM (G-ELM)

In order to find the nonlinearity and uncertainty of the rainfall–runoff process over the watershed, the G-ELM was employed in this study. For this end, the G-ELM model was established according to the observed rainfall, runoff time series along with spatially varying physical and geomorphological parameters such as upstream drainage area and mean slope of sub-basins. In other words, the model inputs were determined using rainfall–runoff temporal features and geomorphological characteristics of the sub-basins. These physical parameters were selected as inputs of the model since most conceptual models take advantage of such physical/geomorphological parameters to convert independent variables to the variable of interest (e.g., Melesse et al. 2003). Likewise, these parameters are the preliminary features used by several experimental equations (e.g., SCS and rational equations). Area (upstream drainage area) and mean slope are two geomorphology-based parameters utilized in most physically based rainfall–runoff models. In order to examine the capability of the mentioned parameters in the modeling process, upstream drainage area \( (A_i) \) and mean slope \( (S_i) \) as spatial variables were selected to determine the input layer of ELM. These features were used as the physical characteristics of the watershed in the proposed model, in addition to hydro-climatologic parameters (i.e., rainfall and runoff temporal features).

By considering the rainfall–runoff temporal features and geomorphological parameters of sub-basins, the G-ELM was expanded in a way that both datasets could describe the characteristics of target data suitably \( (Q_i(t), i = 1, 2, 3, \ldots, n, \)
where $Q$ is the runoff values at station $I$ and $n$ is the number of
sub-basins) (Nourani et al. 2013). Equation (7) shows a sche-
matic of the proposed model’s input-output structure:

$$\begin{bmatrix}
I_1(t) & I_2(t) & I_3(t) & \ldots & I_n(t) \\
I_1(t-1) & I_2(t-1) & I_3(t-1) & \ldots & I_n(t-1) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
I_1(t-a) & I_2(t-a) & I_3(t-a) & \ldots & I_n(t-a)
\end{bmatrix}
\begin{bmatrix}
A_1 & B_1 \\
A_2 & B_2 \\
A_3 & B_3 \\
\vdots & \vdots \\
A_n & B_n
\end{bmatrix}
\begin{bmatrix}
Q_1(t) & Q_2(t) & Q_3(t) & \ldots & Q_n(t) \\
Q_1(t-1) & Q_2(t-1) & Q_3(t-1) & \ldots & Q_n(t-1) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
Q_1(t-\beta) & Q_2(t-\beta) & Q_3(t-\beta) & \ldots & Q_n(t-\beta)
\end{bmatrix}
$$

(7)

In Equation (7), $i$ indicates the sub-basin number, $\alpha$ and $\beta$
are the lag times for rainfall ($I$) and runoff ($Q$) data, respecti-
vely. $A_i$ and $S_i$ are upstream drainage area and slope for
sub-basin $i$, respectively. As shown in Equation (7), in
addition to temporal variables, spatial variables were con-
sidered as inputs of the model. The proposed G-ELM model
could be used for monthly rainfall–runoff modeling of the
entire watershed (Figure 1). In the G-ELM model, the input
variables consisted of different sets of antecedent and current
rainfall and runoff values and physical characteristics of the
sub-basins relevant to all stations to estimate the runoff time
series of the stations ($Q_i(t), i = 1, 2, 3, \ldots, n$, where $Q$
is the runoff values at station $i$, $n$ is the number of sub-basins).
As shown in 1, the G-ELM model exhibits a singular model
which is capable enough to be used instead of several
models (such as AI) for stations inside the watershed. Furth-
more, the proposed G-ELM model is able to predict runoff
values in stations located in any desired point within the
watershed (e.g., station 3 in Figure 1).

Integrated ELMs (I-ELMs)

In most previous studies, a simple AI approach was used to
predict the river runoff. In other words, some antecedent-
based or wavelet-based inputs were determined first in
which rainfall–runoff data or other relevant features were
provided as input data, then by using the input data, river
runoff could be predicted. Due to their simple structure,
they could not be applied to predict runoff at watershed out-
lets (Huo et al. 2012). To overcome this weak point, this study
planned a conceptual cascade reservoir model based on inte-
grated form of the ELM models. The I-ELMs strategy used
representing each of the sub-basins, the outflow from $n$th
reservoir could be determined based on the channel drainage
network (Figure 2). The integrated form of ELMs (I-ELMs) is
developed to take advantage of the semi-distributed models
considering the heterogeneity and spatial variation of the
rainfall in the watershed. Flow charts of the I-ELMs are illus-
trated in Figure 2. According to Figure 2, the I-ELMs were
designed to employ a meta-learning system; meta-learning is
a privileged technique to unify the results of multiple learners
(Liao & Feng 2014). I-ELMs model were achieved by forming
several base ELMs and one meta-learner. The meta-learner
learns from the outputs of the base ELMs, as shown in
Figure 2. Each base ELM generates a base-approximator and
the meta-learner generates a meta-approximator. In
other words, each base-learner is a trained and tested form
of ELM for each station located at location $i$. According to
Figure 2, the approximated results from runoff data were
aggregated and were imposed into meta-level training along
with decomposed rainfall data. Finally the meta-learner fore-
casts the runoff for station 13 (approximately located at
downstream of the watershed) by enjoying all temporal fea-
tures which were imposed into the system. As indicated in
Figure 2, the entire watershed was divided into sub-basins
based on the characteristics of the watershed so that several
ELMs could be trained to estimate runoff from the sub-
basins, respectively. The developed I-ELMs were formed in
a way that the runoff outputs of all base ELMs in the first
stage could be aggregated and used as the inputs of meta-lear-
er for forecasting runoff downstream according to the
conceptual reservoirs cascade model (Figure 2). The meta-
learner is a two-fold training system. In the first stage (base-
learners) the ELM was used to train the 13 stations located
at upstream of station 13. Next, in the meta-learning stage, the trained models are aggregated to form a unit time series of runoff and then train the meta-learner.

**Study area and dataset**

This research was extended to investigate the nonlinear relationship between rainfall–runoff processes in monthly scale for Ajichay watershed which is located in a semi-arid area of Iran. The monthly rainfall and runoff dataset used in this study (Table 1) were provided by East Azerbaijan regional water company. Ajichay watershed’s hydrometric stations and rain gauges are located in northwest Iran in Azerbaijan province (between 47°45’ and 45°30’ east longitude and 38°30’ and 37°45’ north latitude). Figure 3 demonstrates the geographical location and digital elevation map (DEM) of Ajichay watershed. The watershed drainage area is about 10,853 km² and covers about 25% of Urmia Lake. Watershed elevation varies between 1,228 m and 3,755 meters above sea level. River runoff is shed into Urmia Lake, and is momentous for survival of the lake. Mean daily temperature varies from 2.5°C in winter to 20°C in summer. Relative humidity varies...
from about 55 to 60% over the year. Wind speed in the western area is higher than in the eastern and is 17.5 and 11.5 km/h, respectively. Relative information about rainfall–runoff and watershed spatial information, such as statistical parameters of the data, are given in Table 1. For the purpose of forecasting, which includes the calibration and verification steps, the dataset was divided into two parts. The first 75% of total data was used in the training set and the remaining 25% of data was used for testing purposes. Runoff of station 3, which covers 12 years of rainfall–runoff data, was predicted by using cross-validation technique. Data duration is from 1993 to 2014 (except for station 3 which is from 2002 to 2014). Monthly streamflow time series from 1993 to 2014 for station 13 and summation of other stations were used in this study. Figure 4 illustrates the sum of runoff values for all stations except station 13 and respective linear trends. It is deduced from the figure that all stations have a behavior similar to station 13.

Table 1: Information of stations and statistical analysis of observed data

<table>
<thead>
<tr>
<th>Station no.</th>
<th>Area (%)</th>
<th>Slope (%)</th>
<th>Rainfall (mm)</th>
<th>Runoff (m³/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>10.4</td>
<td>26.18</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>14.1</td>
<td>24.3</td>
<td>0.87</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>13.58</td>
<td>32.84</td>
<td>2.5</td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>14</td>
<td>29.9</td>
<td>1.1</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>14</td>
<td>23.87</td>
<td>0.65</td>
</tr>
<tr>
<td>6</td>
<td>49</td>
<td>13.6</td>
<td>22.1</td>
<td>6.1</td>
</tr>
<tr>
<td>7</td>
<td>47</td>
<td>15.43</td>
<td>21.97</td>
<td>1.1</td>
</tr>
<tr>
<td>8</td>
<td>69</td>
<td>8.02</td>
<td>20.93</td>
<td>0.5</td>
</tr>
<tr>
<td>9</td>
<td>77</td>
<td>8.42</td>
<td>27.14</td>
<td>1.2</td>
</tr>
<tr>
<td>10</td>
<td>81</td>
<td>10.05</td>
<td>25.3</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>75</td>
<td>12.01</td>
<td>24.95</td>
<td>0.3</td>
</tr>
<tr>
<td>12</td>
<td>90</td>
<td>13.3</td>
<td>23.2</td>
<td>0.9</td>
</tr>
<tr>
<td>13</td>
<td>97</td>
<td>5.38</td>
<td>20.55</td>
<td>8.5</td>
</tr>
<tr>
<td>14</td>
<td>88</td>
<td>14.2</td>
<td>18.02</td>
<td>0.83</td>
</tr>
<tr>
<td>15</td>
<td>76</td>
<td>18.6</td>
<td>24.36</td>
<td>0.5</td>
</tr>
</tbody>
</table>
In this study, the total data were separated into calibration and verification sets. The performance of proposed models are commonly assessed by computing a number of indices/measures of accuracy. It is important to apply a number of indices, which should be chosen according to the particular needs of each individual application (Legates & McCabe 1999; Roushangar & Koosheh 2015; Nourani et al. 2016). Moreover, the adopted criteria should not be redundant (e.g., Adamowski & Chan 2011) and be sensitive to different types of errors (e.g., peak errors, delay, small value errors, etc.). However, often only redundant indices, which are linked to each other, such as the root mean square error (RMSE), correlation coefficient ($R$), and coefficient of determination ($DC$) are applied. Moreover, the ‘best’ model is usually selected by comparing the values of the performance criteria. Two of these measures are redundant ($R$ and $DC$ up to 1) while the other is not ($RMSE$ from zero to a positive value). Redundant indices are chosen to provide a mutual validation of the results as well as stressing the importance of using non-redundant indices. Equations (8)–(10)
represent the statistics used in this study.

\[
R = \left[ \frac{\sum_{t=1}^{n} (Q_t - \tilde{Q}_t)(Q_p - \tilde{Q}_p)}{\sqrt{\sum_{t=1}^{n} (Q_t - \tilde{Q}_t)^2 + \sum_{t=1}^{n} (Q_p - \tilde{Q}_p)^2}} \right]^2
\]

(8)

\[
DC = 1 - \frac{\sum_{t=1}^{n} (Q_t - \tilde{Q}_t)^2}{\sum_{t=1}^{n} (Q_t - \tilde{Q}_p)^2}
\]

(9)

\[
RMSE = \sqrt{\frac{\sum_{t=1}^{n} (Q_t - Q_p)^2}{R}}
\]

(10)

where \( N \) represents the number of test data, \( Q_o \) is the observed runoff, and \( Q_p \) is the predicted runoff.

Input and output variables are normalized by scaling between 0 and 1 to eliminate their dimensions and to ensure that all variables receive equal attention during calibration of the model. Therefore, RMSE values are to be transformed to between 0 and 1 (Nourani et al. 2016).

\section*{RESULTS AND DISCUSSION}

\textbf{Results of classic modeling strategy}

First, a SLFNN model was trained to predict one-step-ahead runoff with normalized rainfall and runoff features for all stations. ELM is a fast learning emerging technique with high generalization performance. It basically uses single-hidden layer (feature mapping) feed forward neural networks (SLFNNs), with number of hidden nodes that can be randomly generated; in the traditional FFNNs, all the parameters are required to be tuned, which generates a case of dependency between different layers of parameters (weights and biases) (Huang et al. 2006). The significance of this kind of ELM is that the hidden layer does not need to be tuned; learning speed is faster than FFNNs, in addition to better generalization performance.

In order to determine the best classic ELM structure, temporal features were used in different combinations. For this end, lag-1 autocorrelation coefficients (ACF) of rainfall–runoff was performed for all stations. It was observed that generally all stations had satisfactory correlation of rainfall–runoff values with the prior 1, 2, and 12 months (\( t - 1, t - 2, \) and \( t - 12 \)). Hence, the following input combinations (1-5) which included different numbers of input variables \( Q \) and \( R \) were considered in determining the input matrix to predict the watershed runoff.

Comb. 1. \( Q(t) \)

Comb. 2. \( Q(t), R(t) \)

Comb. 3. \( Q(t), Q(t-1), R(t), R(t-1) \)

Comb. 4. \( Q(t), Q(t-12), R(t), R(t-12) \)

Comb. 5. \( db(4,3) Q(t), db(4,3) R(t) \)

In all cases, \( t \) demonstrates previous time step, \( Q \) represents runoff features, and \( R \) represents rainfall features. The output consists of a sole variable, i.e., \( Q \) at future time steps (\( Q_{(t+1)} \)).

In order to get a precise prediction of runoff, the input matrix should be set in a way that can it use all related information about the target data. Based on the selected input data, the pattern of the input layer was arranged with only the most important time memories of rainfall and runoff time series. In this regard, remembered combinations including sliding window on runoff time series with different lag time steps were considered as model input. Comb. 5 was designed to take advantage of WT since rainfall–runoff processes have highly non-stationary fluctuations of time series. To handle such features, wavelet-based temporal data were fed into the ELM in order to improve modeling accuracy. The hybrid model, wavelet-ELM (W-ELM) was designed to catch the capability of wavelet implication in nonlinear runoff modeling and ELM’s capability simultaneously. The selection of an appropriate mother wavelet plays an important role in wavelet-based modeling. As discussed by Nourani et al. (2014), and due to the structure of the Daubechies-4 (\textit{db}4) mother wavelet, which is almost similar to the hydrological properties signal, it could capture the signal’s features, especially peak values, well and was selected as the mother wavelet for decomposition of the runoff time series in this study. In addition to the type of mother wavelet, the selection of decomposition level is another important task in W-ELM modeling. The decomposition of the runoff time series at level L yields \( L + 1 \) sub-signals (one approximation sub-signal, \( a(t) \) and \( L \) detailed sub-signals, \( di(t) \) \((i = 1, 2, ..., L)\)). Decomposition level 3 was...
considered as the optimum decomposition level, according to the following equation (Aussem et al. 1998):

\[ L = \log N \]  

(11)

in which \( L \) and \( N \) are decomposition level and time series length, respectively. For the study, \( L = 2.4 - 2.57 \) in which integer number of 3 was considered in this study. In Comb. 5, \( \text{db}(4,3) Q_i(t) \) means the selected mother wavelet is \( \text{db}4 \) and decomposition level is 3 for the runoff time series.

In order to determine the ELM outcome, the performance evaluation results are presented for different combinations in Figure 5 (DC and R). According to the results of one-step-ahead from Figure 5, although Combs. 3 and 4 might have shown adequate efficiency, Comb. 5 showed the best performance. Also, Combs. 1 and 2 could not predict the runoff accurately. From this point of view, it could be concluded that Comb. 5 was the best choice as input combination. Incorporating WT to ELM caused a great increase in prediction accuracy compared to the outcome of Comb. 1. As an illustration, Figure 6 shows the scatterplot of predictions against observed values for stations 2 and 10. It could be inferred that Comb. 5 (W-ELM) had the best outcome.

Results of proposed G-ELM model

The proposed G-ELM model’s input data were determined using rainfall–runoff time series and spatially varying parameters (Figure 2). Since the geomorphological input variables (i.e., \( A_i \) as upstream area of sub-basins) have inconsistent dimensions, it is necessary to normalize the data. One standard approach is to use dimensionless values to avoid the potential conflicts with incorrect dimensionality of derived formulations. This is a standard scientific practice, since units of measurements are effectively eliminated (Babovic et al. 2001). In this way, the dimensionless geomorphological parameters were employed in the model. Therefore, the model structure for each sub-basin could be expressed as (Nourani et al. 2013):

\[ Q_{i}(t) = f\left(R_{i}(t), \text{db}(4,3)R_{i}(t), \ldots, Q_{i}(t), Q_{i}(t-1), \text{db}(4,3)Q_{i}(t), \ldots, \frac{A_{i}}{A_{T}}, \frac{S_{i}}{S}\right) \]  

(12)

In Equation (12), \( i \) refers to the sub-basin number (\( i = 1, 2, 3, \ldots, 15 \)). \( R_{i}(t) \) and \( \text{db}(4,3)R_{i}(t) \) are rainfall values at \( i \)th sub-basin, and \( Q(t), Q(t - 1) \) and \( \text{db}(4,3)Q(t) \) are runoff values at different lag times; \( \text{db}(4,3) \) was designed according to Equation (11). Also, \( A_{i} \) and \( S_{i} \) are drainage area and slope for each sub-basin, respectively. \( A_{T} \) and \( S \) are the total watershed area and mean slope for the whole watershed, respectively. Since the runoff is extremely affected by current and previous conditions of the watershed, WT could increase the G-ELM abilities to catch seasonality of the time series. The results presented in Table 2 show the eligibility of the proposed G-ELM model for estimating the stations’ runoff.

Figure 5 | Performance evaluation of classic and W-ELM (Comb. 5) modeling via ELM by the means of DC and R.
Performance evaluation was performed to determine the efficient proposed G-ELM architecture. For this purpose, the input combinations (1–6) which include different numbers of input variables were considered in the input matrix (Table 2). Table 2 shows the G-ELM performance for different combinations of input matrix. As shown in Table 2, in order to find out the effect of physical and geomorphological parameters, Comb. 1 was defined with only temporal variables (i.e., only rainfall and runoff data). According to Table 2, the impact of imposing dimensionless physical parameters along with the temporal variables was eligible. In general, Combs. 2–5 showed higher efficiency with respect to Comb. 1. Furthermore, substituting spatially varying parameters (i.e., A and S) instead of rainfall or runoff variables (i.e., \( I(t-1)/Q(t-2) \)) in the input combination led to better performance. By comparing the obtained results for Combs. 1–6 (Table 2), it could be concluded that Comb. 6 proved to be the best structure of G-ELM. The input structure of this combination consisted of two temporal variables (\( db(4,3)Q(t) \) and \( db(4,3)R(t) \)) and two spatial variables (i.e., A and S). The drainage area (A) is an important physical characteristic of the watershed. Drainage area reflects the volume of water that can be generated from rainfall. Mean slope of watershed (S) is a significant factor for runoff yield from rainfall. By considering these facts, the volume of available water for runoff would be the product of rainfall depth and the drainage area and move along the watershed affected by the slope. As well, flood magnitudes reflect the momentum of the runoff and it is obvious that slope is an important factor in

Table 2 | Results of proposed G-ELM model

<table>
<thead>
<tr>
<th>Combination no.</th>
<th>Input matrix variables</th>
<th>Calibration</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( Q(t), Q(t-1), R(t) )</td>
<td>0.85 0.88 0.023</td>
<td>0.62 0.76 0.040</td>
</tr>
<tr>
<td>2</td>
<td>( Q(t), Q(t-1), R(t), A )</td>
<td>0.87 0.89 0.019</td>
<td>0.67 0.77 0.036</td>
</tr>
<tr>
<td>3</td>
<td>( Q(t), Q(t-1), R(t), S )</td>
<td>0.87 0.9 0.019</td>
<td>0.69 0.8 0.038</td>
</tr>
<tr>
<td>4</td>
<td>( Q(t), Q(t-1), R(t), A, )</td>
<td>0.89 0.93 0.017</td>
<td>0.64 0.74 0.037</td>
</tr>
<tr>
<td>5</td>
<td>( Q(t), R(t), A, S )</td>
<td>0.86 0.9 0.019</td>
<td>0.72 0.8 0.032</td>
</tr>
<tr>
<td>6</td>
<td>( db(4,3)Q(t), db(4,3)R(t), A, S )</td>
<td>0.91 0.93 0.018</td>
<td>0.82 0.87 0.025</td>
</tr>
</tbody>
</table>

Figure 6 | Comparison of predicted runoff values in calibration and validation periods for classic and W-ELM for stations 2 and 10.
the momentum. Therefore, the importance of the roles of drainage area and slope in the rainfall–runoff process could be reconfirmed by the outcome of the current modeling. Conjugating WT to temporal features in Comb. 6 gave an incredible insight into temporal features. Comb. 6 had a supreme performance in the G-ELM.

Figure 7 presents decomposed times series of runoff values by $db(4,3)$ for all stations formed by the input matrix. In Figure 7, S refers to the main signal which was decomposed to approximation and detail signals. $a_3$ represents approximation sub-series at level 3; $d_1$, $d_2$, and $d_3$ represents detail sub-series at level 1, 2, and 3 respectively.

Figure 8 shows the computed time series versus the observed time series for stations 1–15 (except station 3) in the calibration and validation steps. In Figure 8, St represents an abbreviation for the word station. Each time series was separated into calibration (189 months) and verification (65 months) periods. The consecutive form of multiple-stations time series of Ajichay watershed and location of stations is demonstrated in Figure 8. The figure proves the ability of the proposed G-ELM model for predicting runoff at different parts of the watershed. The reason for this ability could be due to the detection of interaction among the stations when their data were imposed onto a unique framework. Additionally, the spatial relationship among the stations could be detected as well as their temporal pattern. As a result, the model should be classified as a time–space model. This model can operate as a nonlinear time–space regression approach rather than a fully lumped model. However, Figure 8 also shows that the proposed model could not catch the peak values. Therefore, the potential risk of flooding could be very high when using the model to predict flow. According to the presented results in Table 2, the G-ELM led to sufficient performance. The main reason for such a suitable efficiency was probably related to the input structure of the G-ELM. Since the runoff variables (i.e., $db(4,3)Q(t)$) were considered in the input structure, the autoregressive and Markovian property and seasonality of runoff time series could be caught accurately. It is worth noting that implication of geomorphological parameters, in addition to the temporal variables, caused an improvement in the model performance (Table 2).

**Integrated ELMs**

The temporal data used in I-ELMs were first pre-processed by WT. Such a pre-processing can cause an improvement in the performance of I-ELMs. Next, each base-ELM was used to
forecast the relative target runoff (see Figure 3). Outcome results of these models were aggregated to form a single time series. In the next stage, the decomposed time series of runoff (aggregated time series) and decomposed rainfall features of station 13 were imposed into meta-learner to forecast station 13’s runoff time series. Such a modeling strategy has the privilege of enjoying all datasets so that the meta-learner can handle station 13’s time series’ specification. One of the great differences between the G-ELM and I-ELMs is the use of spatial characteristics of the watershed as input data. I-ELMs did not employ spatial data, whereas G-ELM was designed based on employing spatial data. In other words, I-ELMs did not apply spatial data of the watershed; in exchange, it was designed in a way that it can explore spatial variation in the rainfall–runoff process and did not require any physical features of the watershed. In order to train the I-ELM, each base-learner was trained using $db(4,3)Q_i(t)$, $db(4,3)R_i(t)$ as input data. In other words, decomposed rainfall and runoff values via $db(4,3)$ were used to calibrate the base models. Next, $db(4,3)\sum Q_i(t)$, $db(4,3)R_{13}(t)$ was used as input data to train the meta-learner. Since the meta-learner is trained by using the base-learners, the outcome of the base-learners were aggregated to create a united time series of runoff ($\sum Q_i(t)$). The runoff time series was then decomposed via $db(4,3)$ and was used to train the I-ELM along with $R_{13}(t)$ (rainfall time series of station 13). Results demonstrated that in the calibration period $DC = 0.95$, $R = 0.95$, and $RMSE = 0.028$ and for the verification period $DC = 0.88$, $R = 0.91$, and $RMSE = 0.42$ was obtained. According to the obtained results, I-ELMs performed well and in comparison to G-ELM, outperformed it relatively. However, both these strategies are unique and applicable. It means that these strategies could be set for different spatial and temporal conditions.

**Results of cross-validation**

As previously mentioned, station 3 was selected to be predicted through cross-validation technique. Cross-validation is a way of qualifying the performance of proposed strategies. For cross-validating the conceptual models, temporal features of station 3 were omitted from the calibration process and it was predicted in the verification process. In order to apply this technique in this study, the time series of station 3 was imposed into the trained system of the models. The G-ELM, which utilized spatio-temporal features in its structure, was set in a way to interpolate station 3’s runoff accurately. For this purpose, calibration dataset was arranged according to the G-ELM structure; the only difference was the input dataset, in which station 6’s spatio-temporal features were entered into the input matrix because of its approximately similar topographical position. The relation and interaction among the stations’ data, learnt through the calibration phase and collaborated with the model to have appropriate predictions for the cross-validation procedure (Nourani et al. 2013). Moreover, for verifying the capability of I-ELMs in the cross-validation
procedure for station 3, I-ELMs were set as proposed. The only diversity was at meta-learning formation, which was set for 12 years of monthly time steps (calibration and verification, from December 2002 to October 2014). The results of predictions were compared with the available observed data to evaluate performance of the developed models for spatial predictions. In order to evaluate the proposed G-ELM model and I-ELMs model, the predicted values are shown in Figure 9. According to the results, a reliable performance was obtained ($DC = 0.8$, $R = 0.83$, and $RMSE = 0.1$ for G-ELM and $DC = 0.78$, $R = 0.82$, and $RMSE = 0.12$ for I-ELMs) by considering limitations in the modeling procedure. It is obvious that if the G-ELM and I-ELMs were trained using more stations, the model accuracy would be increased in the validation and cross-validation steps. According to the obtained results, G-ELM performed better than I-ELMs. It can be stated that G-ELM, which employed spatial data and similar temporal features, led to better performance. The results of cross-validation in terms of computed versus observed time series and scattered plot are shown in Figure 9, which demonstrates the ability of the G-ELM model for spatiotemporal prediction of the rainfall–runoff process of the Ajichay watershed.

Limitations, issues and requirements

Generally, the ELM algorithm shows faster learning speed over traditional algorithms for SLFNs. For example, it learns faster than the support vector machine by a factor of up to thousands. ELM does not suffer from problems like imprecise learning rate, presence of local minima, etc. This underlines the suitability of ELM for real complex applications that need fast prediction and response capabilities. In addition, the ELM algorithm requires less human involvement, because it does not have any control parameters to be manually tuned, which shows its better applicability for real world applications (Bhat et al. 2008). Finally, recent works confirm the advantages of ELM over earlier approaches for ANN (Malathi et al. 2011). Nevertheless, three key aspects should be pointed out:

- Random initialization of hidden node parameters may affect the performance of ELM (Bhat et al. 2008) that also require high complexity of SLFFNs for improved performance. This may lead to unfavorable conditions, which means that an ELM may not be robust enough to capture variations in data.
- Considering the network complexity issue, it is essential to carefully choose hidden neuron activation functions that show good ability to handle complexity, improve convergence of algorithm, and result in a compact structure of network (Huang & Chen 2008). These aspects are taken into account in the proposed ELM-based model.
- ELMs have been widely applied to hydrology to predict runoff (Abdullah et al. 2015; Lima et al. 2016). One limitation of ELMs is that the prediction can only be made based on historical rainfall–runoff relations. Therefore ELMs will show weakness when trying to predict some high peaks that have not yet been covered in the training
data. In this study, pre-processing by WT recovered some weaknesses. In proposed models (i.e., W-ELM and I-ELM) peak values were predicted accurately. However, in G-ELM in which a wide range of fluctuation in data was observed, the model could not catch the peak values.

**Comparison of results**

In order to compare the performance of the presented models, the obtained results via classic models (W-ELM), G-ELM and I-ELMs for station 13 are shown in Figure 10 (as mutual outcome). A zoom window is presented for the first 75 months because of intense fluctuations in these months in comparison with the whole. Also, calculated data were plotted against observed data in order to judge the obtained results more carefully. As indicated from Figure 10, the G-ELM showed less efficiency with respect to the I-ELMs and W-ELM. W-ELM outperformed proposed strategies in this research by catching maximum and minimum values with better precision. As a general conclusion, it must be stated that ELM is a powerful and fast forecasting approach. For example, by considering W-ELM in which ELM benefits from WT, the modeling procedure possesses a strong insight which makes it supreme in predicting shorter times series (here 252 months for one station). Therefore, such a compound of approaches for short-term modeling performed better than other proposed models. The G-ELM model, which took advantage of spatial characteristics of watershed, was designed in a way to forecast imposed data for all stations (14 stations). Adding WT also decreased error in modeling, but the very wide range of fluctuation caused some deflection in some levels of forecasting. The I-ELMs strategy, which benefited from temporal features by taking advantage of WT also could not perform as well as W-ELM. This is because this model did not use the target data in the modeling procedure but took advantage of other station features in two stages including rainfall and runoff data.

Generally, the results of different wavelet-based models were satisfactory. This is due to the capability of WT and nonlinear ability of ELM and mostly the proposed strategies which were designed in a way to catch the peak and least values in modeling procedure.

FFNN is a powerful model in prediction of hydrological variables and was employed as a benchmark to analyze the performance of ELM for station 13. Since Comb. 5 in classic modeling led to the best outcome among all input data, it was used as input for FFNN. According to Table 3, FFNN with eight input data, one hidden layer and one output (8-1-1) used 200 epochs to achieve the best structure. Both wavelet-based FFNN (W-NN) and W-ELM forecast the runoff values of station 13 accurately (see Table 3). However, results indicated that W-ELM outperformed W-NN by means of $DC$ and $RMSE$. The FFNN’s structure was optimized via trial and error process, which means FFNN was optimized by several times running in MATLAB environment.

**SUMMARY AND CONCLUDING REMARKS**

Numerous recently developed rainfall–runoff studies have focused on AI-based modeling in which AI approaches were used along with WT. The present study took advantage of different rainfall–runoff modeling strategies to improve their performances. Calibrating and designating sufficient inputs and the method of employing them are the key to taking advantage of them in a way that they can operate as semi-distributed models or physical-based models but with precision. Among different AI-based models, the ELM model has been used by researchers very recently because of its unique susceptibility. To this end, in this study different strategies of modeling were employed and examined (i.e., geomorphology integrated model and integrated form of ELMs) which were designed according to spatial and temporal features. At first, the lumped form of ELM was developed by using input combinations for one-step-ahead predictions. Comb. 5, which was conjugated to WT and utilized rainfall–runoff features (W-ELM via $db4$), led to the best outcome for all hydrometric stations. Subsequently, two approaches were proposed, namely, G-ELM and I-ELMs. G-ELM presented a multi-station modeling of the Ajichy watershed runoff in continuous form which used the geomorphological properties of sub-basins along with decomposed rainfall and run off time series. The presented results for the G-ELM model indicated that the proposed integrated geomorphological approach is an
Figure 10 | Comparison of results obtained from different strategies via ELM (calibration and verification periods).
efficient way to predict runoff in multi-station form for the watershed. A conceptual cascade of the reservoir model was developed according to the integrated form of ELM. This strategy did not employ any physical features of the sub-basins. However, the structure of I-ELMs was formed in a way to be able to explore spatial variation of rainfall–runoff process. I-ELMs took advantage of all correlated temporal information and performed better than G-ELM in terms of DC, R, and RMSE. W-ELM, I-ELM, and G-ELM performance was ranked, respectively, based on statistical evaluations. The cross-validation technique was employed to predict the runoff values for station 3. Based on the results from cross-validation, G-ELM predicted the runoff values of station 3 better than I-ELMs. This could be due to the structure of the proposed models. The I-ELMs were designed to forecast runoff values at the outlet of the watershed; the spatial pattern and relationship among the stations could be detected better through the G-ELM model.

Finally, the W-ELM and conventional W-NN-based modeling was performed for station 13. Results indicate that W-ELM was more suitable than the W-NN model, which could not cope as well with the nonlinear characteristics of the rainfall–runoff process as W-ELM.

The presented models were tuned according to the physical properties and time series of the case study. The presented methodologies can be adopted for other watersheds in order to take advantage of the watershed geomorphological information for AI-based modeling. It is also suggested for future studies to focus on utilizing the spatial and temporal clustering approach. Evidently, long-term data for the watershed will be needed to achieve this goal. Based on the results, W-ELM outperformed I-ELMs and G-ELM; however, it must be considered that each strategy has weak points and advantages, therefore the unique structure of these strategies must be utilized for appropriate situations.

**ACKNOWLEDGEMENTS**

The authors would like to thank East Azerbaijan regional water company for data preparation. The authors also would like to express their gratitude to the anonymous reviewers for their constructive criticism.

**REFERENCES**


Ashu, J. & Sanaga, S. 2006 Integrated approach to model decomposed flow hydrograph using artificial neural network


Roushangar, K. & Alizadeh, F. 2015 Suitability of different modelling strategies in predicting of solid load discharge of an alluvial river. In: *36th World Congress of IAHR*, The


Sang, Y. F. 2013 A review on the applications of wavelet transform in hydrology time series analysis. *Atmospheric Research* 122, 8–15.


First received 7 January 2017; accepted in revised form 16 September 2017. Available online 2 November 2017