Water Resources Management (2019) 33:3831–3849
https://doi.org/10.1007/s11269-019-02333-5

Modular Wavelet–Extreme Learning Machine: a New Approach for Forecasting Daily Rainfall

Aman Mohammad Kalteh[1]

Received: 17 October 2017 / Accepted: 5 August 2019 /
Published online: 16 August 2019

Abstract
A rainfall forecasting method based on coupling wavelet analysis and a novel artificial neural network technique called extreme learning machine (ELM) is proposed. In this way, the unique characteristics of each technique are combined to capture different patterns in the data. At first, wavelet analysis is used to decompose rainfall time series into wavelet coefficients, and then the wavelet coefficients are used as inputs into the ELM model to forecast rainfall. The accuracy of the model is further improved using a modular learning approach. In the modular learning, an innovative approach to determine the optimum number of clusters entitled threshold cluster number is introduced. The relative performances of the proposed models are compared with the single ELM model for three cases consisting of one daily rainfall series from Iran (Kharjeguil station), one daily rainfall series from India (Ajmer station) and one daily rainfall series from the United States (Barton Pond station). The correlation coefficient (*r*), root mean square errors (*RMSE*) and Nash–Sutcliffe efficiency coefficient (*NS*) statistics are used as the comparing criteria. The comparison results indicate that the proposed modular wavelet–ELM method could significantly increase the forecast accuracy and perform much better than both the wavelet–ELM and single ELM. Moreover, three case study results indicate the importance of determining the optimum number of clusters based on the new concept of threshold cluster number in order to achieve optimum forecast results.

Keywords Rainfall · Discrete wavelet analysis · Extreme learning machine · Modular learning · Threshold cluster number

1 Introduction

Accurate and reliable precipitation/rainfall forecasting is important in various water resources activities such as reservoir operation and flood mitigation. Especially, having accurate

✉ Aman Mohammad Kalteh
amanmkalteh@gmail.com

[1] Department of Range and Watershed Management, Faculty of Natural Resources, University of Guilan, P.O. Box 1144, Sowmehe Sara, Guilan Province, Iran

information on rainfall is a key factor in developing a rainfall-runoff relationship in order to reduce uncertainty associated with rainfall data and hence to improve the quality of river flow forecasts. However, rainfall is one of the most complex components of the water cycle to model because it varies severely both in space and time. Therefore, detailed modelling and forecasting of rainfall has become as a major research field in hydrology.

A significant number of research papers have been published for addressing the quantitative precipitation forecasting through different techniques including numerical weather prediction models and remote sensing observations (Yates et al. 2000; Ganguly and Bras 2003; Sheng et al. 2006; Diomede et al. 2008; He et al. 2013), statistical models (Chu and He 1994; Chan and Shi 1999; DelSole and Shukla 2002; Munot and Kumar 2007; Li and Zeng 2008; Nayagam et al. 2008), and soft computing methods including artificial neural networks (ANNs), support vector machine (SVM), and fuzzy inference system (French et al. 1992; Navone and Ceccatto 1994; Pongracz et al. 2001; Freiwan and Cigizoglu 2005; Marzano et al. 2006; El-Shafie et al. 2011; Hamidi et al. 2015).

Black-box models, such as artificial intelligence methods, basically rely on measured data which, of course, make them suitable to give an enhanced performance once the measured data with quality are provided. One approach that can be used to provide the measured data with quality is relying on a modular modelling (Furundzic 1998; Zhang and Govindaraju 2000; Jain and Srinivasulu 2006; Wu et al. 2010).

Over the last decade, another approach that extensively used to provide the measured data with quality is the application of wavelet analysis (e.g. Adamowski and Sun 2010; Kalteh 2013, 2015). Wavelet analysis was used in conjunction with soft computing techniques to improve the ability of these techniques in forecasting a process that is under investigation. Wavelet analysis gives useful decompositions of original time series, and the resulting data enhance the ability of a forecasting model by capturing valuable information on various decomposition levels.

Recently, the use of extreme learning machine (ELM) has been introduced in applications related to hydrology (e.g. Abdullah et al. 2015; Deo and Şahin 2015; Leuenberger and Kanevski 2015; Taormina and Chau 2015; Deo et al. 2016). For example, Abdullah et al. (2015) investigated the efficiency of ELM algorithm at predicting Penman-Monteith (P-M) reference evapotranspiration for Mosul, Baghdad, and Basrah meteorological stations, located at the north, mid, and southern part of Iraq. They compared the performance of ELM model with the empirical P-M equation and with ANN model. Their results proved the ELM model is efficient, simple in application, of high speed, and has very good generalization performance at predicting P-M reference evapotranspiration. Deo and Şahin (2015) applied the ELM algorithm for the prediction of monthly effective drought index (EDI) in eastern Australia. To demonstrate the effectiveness of the ELM model, a performance comparison in terms of the prediction capabilities and learning speeds was conducted between the proposed ELM algorithm and the conventional ANN algorithm trained with Levenberg–Marquardt back propagation. Their results indicated an excellent performance of the ELM over the ANN model. They found that the ELM model is executed with learning speed 32 times faster and training speed 6.1 times faster than the conventional ANN model. They concluded that out of the two machine learning algorithms tested, the ELM model is the more expeditious tool for prediction of drought and its related properties. Deo et al. (2016) developed a wavelet-based drought model using the ELM algorithm where the input data were first analyzed through the wavelet pre-processing technique for better accuracy to forecast the monthly EDI. The forecasting capability of coupled wavelet-ELM model was compared with ELM, ANN, LSSVM and their wavelet-equivalent (wavelet-ANN, wavelet-LSSVM) models. Their results indicated that wavelet-ELM outperformed other models.

Based on a review of the literature, previous studies on the application of ELM were generally devoted to predict reference evapotranspiration, effective drought index, sediments pollution by heavy metals and rainfall–runoff process. Moreover, they have mostly used single ELM models. The issue of modular modelling in conjunction with ELM has not yet been investigated in any application. Hence, the main contributions of this research which make it different from previous studies can be listed as follows:

- This study presents the first application of the ELM algorithm in daily rainfall forecasting.
- In daily rainfall forecasting, it is the first time that the wavelet–ELM conjunction model is developed.
- Also, it is the first time that the combination of modular learning with wavelet-ELM conjunction model in daily rainfall forecasting is introduced and investigated.
- Moreover, in modular learning, the effect of determining the optimum number of modules in modelling is investigated while previous studies when used modular learning in conjunction with other well-known methods such as ANN mostly relied on predefined number of modules which were generally determined by the user to be three or four modules. This study introduces a new concept entitled "threshold cluster number" in order to facilitate finding the optimum number of modules.

This study explored each of these issues using daily data of three rainfall stations from Iran (Kharjeguil station), India (Ajmer station) and the United States (Barton Pond station) in one-day ahead rainfall forecasting.

2 Methods

2.1 Extreme Learning Machine (ELM)

This section covers the theoretical concept behind the extreme learning machine (ELM) architecture. As shown in Fig. 1, ELM developed by Huang et al. (2006) is a new learning algorithm for single-hidden layer feedforward neural networks (SLFNs). Compared to the traditional neural network schemes, the ELM model is easy to implement, no parameters (input weights and hidden layer biases) need to be tuned, obtains the good generalization performance and runs extremely fast. SLFN randomly chooses the input weight matrix and hidden layer biases. It can be then simply considered as a linear system in which the output weights (linking the hidden layer to the output layer) can be analytically determined through simple generalized inverse operation of the hidden layer output matrices. This simple concept makes ELM very efficient and thousands of times faster than traditional feedforward network learning algorithms such as back-propagation algorithm.

The basic theory of the ELM model supposes that for N arbitrary distinct samples (x_i, t_i) where, $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in R^m$, the standard SLFNs with M hidden neurons and an activation function $g(x)$ for the given data are mathematically modelled as (Huang et al. 2006):

$$\sum_{i=1}^{M} \beta_i g_i(x_j) = \sum_{i=1}^{M} \beta_i g_i(w_i x_j + b_i) = y_j, \qquad j = 1, \ldots, N \tag{1}$$

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the randomly assigned weight vector, connecting the input nodes and *ith* hidden node and $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weight vector which connects the

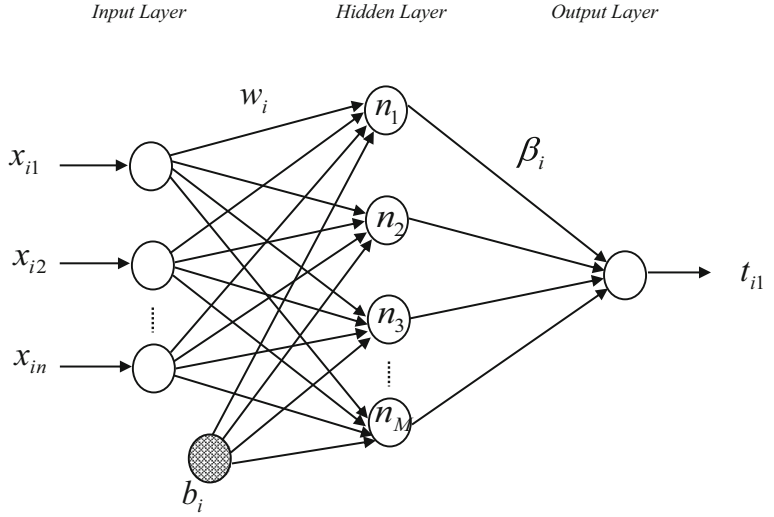

Fig. 1 Extreme learning machine topological structure. Input layer is denoted as $x_{i1}, x_{i2}, \ldots, x_{in}$, hidden nodes as n_1, $n_2, n_3, \ldots, n_M$ and the output layer as t_{i1} which generates the forecasted values of the daily rainfall $R^{'}_{t+1}$ in this study

output nodes with the *ith* hidden node. b_i denotes the threshold of the *ith* hidden node. If the standard SLFNs can approximate these N samples with zero error, then we have

$$\sum_{j=1}^{N} \left\| y_j - t_j \right\| = 0 \tag{2}$$

where y is the actual output value of the SLFN. There also exist β_i, w_i and b_i parameters such that

$$\sum_{i=1}^{M} \beta_i g_i\left(w_i x_j + b_i\right) = t_j, \qquad j = 1, \ldots, N \tag{3}$$

The above N equations can be written compactly as

$$H\beta = T \tag{4}$$

where H is the hidden-layer output matrix of the neural network.

$$H = \begin{bmatrix} g(x_1) \\ \ldots \\ g(x_N) \end{bmatrix} = \begin{bmatrix} g_1(x_1) & \ldots & g_M(x_1) \\ \ldots & \ldots & \ldots \\ g_1(x_N) & \ldots & g_M(x_N) \end{bmatrix} \tag{5}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \cdots \\ \ldots \\ \beta_M^T \end{bmatrix} \tag{6}$$

$$T = \begin{bmatrix} t_1^T \\ \cdots \\ \ldots \\ t_N^T \end{bmatrix} \tag{7}$$

As stated earlier, the input weights and hidden layer biases are randomly generated instead of being tuned. Thus, the output weights connecting the hidden layer to the output layer is as simple as finding the least-square solution to the given linear system. The smallest norm least-square solution to the linear system, Eq. (7), is

$$\hat{\beta} = H^{+}T \tag{8}$$

where H^{+} is the Moore-Penrose (MP) generalized inverse of the hidden layer matrix H.

2.2 Wavelet Analysis

The wavelet analysis is an advanced tool in signal processing and has been developed by the mathematics community over the last two decades. Since its theoretical development, it has attracted much attention in communications, image processing, optical engineering and hydrological applications. The wavelet analysis provides a timescale representation of time series and their relationships, and can be used to explore, denoise and smoothen time series which aid in forecasting time series that contain non-stationarities. Time series data can be decomposed into different components at different resolution levels using the wavelet transforming function (called the mother wavelet). Detailed coefficients are produced by high-pass filters and approximation series are produced by low-pass filters. Due to the multidimensional characters of the wavelets, they can adjust their scale to the nature of the signal features. Furthermore, wavelets can decompose a signal to provide dilations and translations parameters, and then information in the signal is presented by these parameters in the form of frequencies (Al-geelani et al. 2012; Venkata Ramana et al. 2013; Liu et al. 2014).

In applications, the wavelet analysis is of two types:

a) Continuous wavelet analysis (CWA) and
b) Discrete wavelet analysis (DWA).

The CWA of a signal $x(t)$ is defined as follows (Adamowski and Sun 2010):

$$CWA_{x}^{\Psi}(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} x(t)\Psi^{*}\left(\frac{t-\tau}{s}\right) dt \tag{9}$$

Where s is the scale parameter, τ is the translation parameter and '*' denotes the complex conjugate. The mother wavelet $\Psi(t)$ is the transforming function.

However, the CWA requires a significant amount of computation time and data. In contrast, the DWA requires less computation time and is simpler to develop compared to the CWA (Adamowski and Chan 2011). Therefore, DWA has been developed to overcome the drawbacks of the CWA and widely used in engineering applications.

The wavelet basis function for the DWA can be derived as:

$$\Psi_{m,n}(t) = a^{-m/2}\Psi\left(\frac{t - n\tau_0 a^m}{a^m}\right) \tag{10}$$

where, m and n integers that control, respectively, the scale and time, t is the time, a is a specified fixed dilation step greater than 1, and τ_0 is the location parameter that must be greater than zero. The term $a^{-m/2}$ in the above equation normalizes the functions.

One of the capabilities of DWA is that, it produces detailed coefficients to analyze high frequency information and approximation series to analyze low frequency information. In the wavelet analysis, the process of decomposing a signal into its sub signals involves the following steps. In the first step the captured signal is divided into high frequency and low frequency information in order to obtain detailed and approximation series. After that the approximation series is further subdivided into high frequency and low frequency information and sent to the following step. This procedure continues until the predefined number of levels is reached.

2.3 Modular Learning and the Concept of Threshold Cluster Number

To construct a modular model, the training data have to be decomposed into several clusters according to clustering techniques, and then a separate model is devoted to each cluster in order to map the inputs to the corresponding outputs. The self-organizing map (SOM) clustering technique is adopted in this study (Kalteh et al. 2008). It is a learning algorithm that was originally proposed by Kohonen (1982a, b). The SOM algorithm clusters data patterns or samples into a number of classes – the number of which is usually determined by the user.

Figure 2 illustrates the schematic diagram of the modular model where the training data is decomposed into several clusters using the SOM clustering algorithm. Since there is no a general rule to find an optimum number of clusters, here we propose the concept of "threshold cluster number", that is, a range of clusters from 2 up to the threshold cluster number is progressively considered, and for each cluster its respective sub models are developed and according to the Euclidean distance for each pattern in unseen testing input-output pairs we first determine of which cluster it belongs, and then the sub model of the given cluster is used to forecast the corresponding value, and finally the overall performance on testing set is evaluated. This procedure continues for the remaining clusters until the predefined number of clusters, which is equivalent to the threshold cluster number, reaches. The optimum number of clusters is the cluster number whose overall performance on testing set, in terms of the criterion used for evaluation, is the best among all clusters.

The threshold cluster number is determined in a way that while we consider a range of clusters, which starts from 2, if any module in the considered cluster number fails to locate any pattern in unseen testing input-output pairs (i.e. remain empty), this cluster number is discarded and the immediate cluster number prior to it is considered as the threshold cluster number. Assume we start to cluster a training data, which begins from 2 and increments by 1, once the cluster number reaches to 10 if any module of this cluster number fails to accept any pattern in testing data and remains empty, then the cluster number 10 is discarded and the threshold cluster number is considered to be 9. Therefore, the above modular approach for this example is conducted in the range from 2 up to 9.

3 Model Development

3.1 Description of Data and Case Studies

In this study, the performance of ELM was examined on daily rainfall. To insure wider application of the conclusions, three cases consisting of the daily rainfall time series of

Kharjeguil hydrology station of Iran, Ajmer station of India, and Barton Pond rain-gauge data of USA were investigated. The Kharjeguil station is located in northern Iran and observed data of which used here is 12-years long with an observation period between 1999 and 2010. The daily rainfall data of Ajmer station of the Rajasthan state of India have been made available by the water resources department of the Government of Rajasthan at the website http://www.indiawaterportal.org. The data, with period spanning from 1988 to 2007 (i.e. 20 years), were used for the Ajmer station. The daily rainfall data of the Barton Pond rain-gauge of USA, with period spanning from 2010 to 2013 (i.e. 4 years), are available at the website http://www3.a2gov.org/RainGauges/Default.aspx.

Here, each of data series is portioned into two parts as training/calibration set (approximate 80% of the data) and testing set (the remaining approximate 20% of the data). For the Kharjeguil station, a training data set consisting of years 2002–2010 (3287 days) and a testing data set of years 1999–2001 (1096 days) were used. For the Ajmer station, a data period from 1988 to 2003 (5844 days) as a training set and a data from 2004 to 2007 (1461 days) as a testing set were selected. A training set consisting of years from 2010 to 2012 (1096 days) and a testing set consisting of the year 2013 (365 days) were chosen for the Barton Pond rain-gauge data. Table 1 presents some descriptive statistics of the original dataset and two subsets, including mean (X_{mean}), standard deviation (S_x), coefficient of variation (C_v), minimum ($X_{\min}$), and maximum ($X_{\max}$).

The numbers of lags were selected according to the partial auto-correlation function (PACF) of daily rainfall data. The PACF of the daily rainfall data is shown in Fig. 3. It is seen from this figure that model inputs should take the previous 2-day, 7-day and 2-day rainfalls for Kharjeguil, Ajmer and Barton Pond stations, respectively. Thus, these lags were considered as inputs to the models that developed in this study.

Owing to the weak extrapolation ability of artificial intelligence methods, all input and output data were scaled to the range of [0 1].

3.2 ELM

For designing the ELM model three layers were used to build the architecture for forecasting daily rainfall with input data for each station determined by PACF and the output data which is considered to be rainfall forecasting at time $t+1$. For each case study, an optimum number of neurons as well as a suitable type of activation function in the hidden layer must be determined by the user. Here, three activation functions were tried one by one, which include "sigmoid", "sine" and "hard-limit". In each trial, for each station, the numbers of neurons in the hidden layer were increased gradually starting from 1 up to 100. The criteria to choose the optimum structure was the network's performance on testing set, that is, for each individual station the network with the best performance on testing set was considered as the optimal one. Then, the identified architectures of ELM were 2-74-1 with "hard-limit" activation function in the hidden

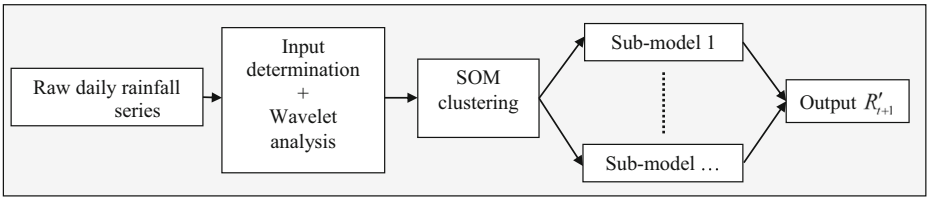

Fig. 2 Schematic diagram of modular model. "…" must be specified by the threshold cluster number

Table 1 Three case studies and the rainfall data information

Stations and datasets	Statistical parameters					Data period
	X_{mean}	S_x	C_v	$X_{\min}$	$X_{\max}$	
Kharjeguil: Iran						1999–2010
Original data	3.51	9.64	2.74	0.00	123.5	
Training	3.53	9.84	2.78	0.00	123.5	
Testing	3.43	9.04	2.63	0.00	107.3	
Ajmer: India						1988–2007
Original data	1.46	7.42	5.08	0.00	200.0	
Training	1.52	7.85	5.16	0.00	200.0	
Testing	1.19	5.32	4.47	0.00	58.40	
Barton Pond: USA						2010–2013
Original data	2.10	5.89	2.80	0.00	67.31	
Training	2.07	5.67	2.73	0.00	67.31	
Testing	2.20	6.51	2.95	0.00	64.00	

X_{mean}= mean daily rainfall (mm); S_x= standard deviation (mm); C_v= coefficient of variation; $X_{\min}$= minimum daily rainfall (mm); $X_{\max}$= maximum daily rainfall (mm)

layer for the Kharjeguil station, 7-14-1 with "sigmoid" activation function in the hidden layer for the Ajmer station and 2-49-1 with "hard-limit" activation function in the hidden layer for the Barton Pond rain-gauge data.

3.3 Proposed Wavelet–ELM Conjunction Models

The wavelet-ELM conjunction models were obtained combining two methods: the wavelet analysis and ELM. The wavelet–ELM conjunction models are ELM models, which use sub signals obtained by the application of wavelet analysis on original input data for each station determined by PACF as inputs, and the outputs are original and $t+1$ daily rainfall. To determine the number of wavelet decomposition levels, the general rule of $\log(N)$ was used

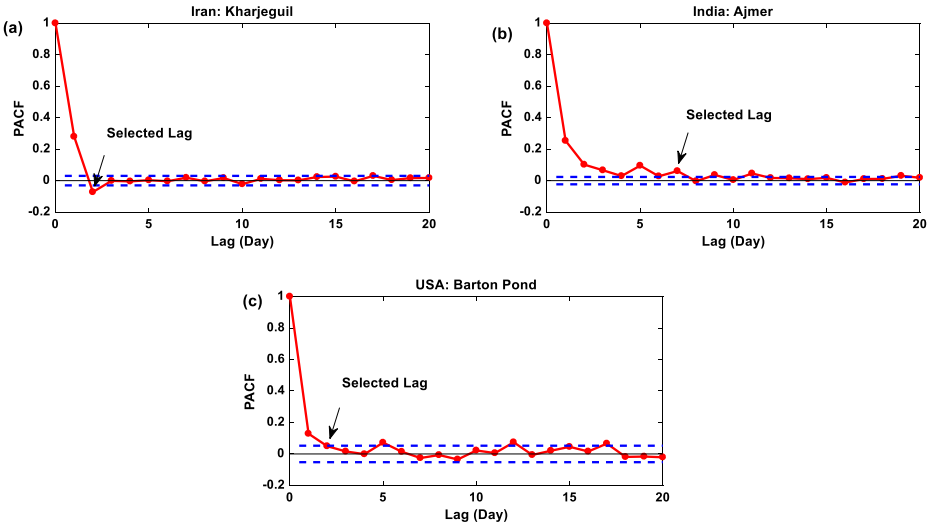

Fig. 3 The PACF of the daily rainfall series with the 95% confidence bands (the dashed lines), **a** for Kharjeguil, **b** for Ajmer and **c** for Barton Pond

(Wang and Ding 2003) in which N is the length of the time series. In this study, for the Kharjeguil, Ajmer and Barton Pond stations, N is 4383, 7305 and 1461 daily data, respectively, and their corresponding decomposition levels are approximately 4, 4 and 3. Another issue in the wavelet analysis that needs to be properly identified is the mother wavelet. There is no a practical guide on identifying a proper mother wavelet, and it is believed to be application dependent (Al-geelani et al. 2012). In this study, we used Daubechies-2(db2) as it is the most popular wavelet (Mallat 1989).

For each station, a three-layer ELM model was designed, and an optimum number of hidden layer neurons and a proper type of activation function in the hidden layer were determined by following the approach conducted as the single ELM model, that is, three activation functions were tried one by one, which include "sigmoid", "sine" and "hard-limit", but in this time, in each trial, the numbers of neurons in the hidden layer were increased gradually starting from 1 up to 150 since the number of input variables by the use of wavelet analysis increased. The criteria to choose the optimum structure was the network's performance on testing set, in a way that, for each individual station the network with the best performance on testing set was considered as the optimal structure. Then, the identified architectures of wavelet–ELM conjunction models were 10-104-1 with "sin" activation function in the hidden layer for the Kharjeguil station, 35-107-1 with "sigmoid" activation function in the hidden layer for the Ajmer station and 8-30-1 with "sigmoid" activation function in the hidden layer for the Barton Pond rain-gauge data.

3.4 Proposed Modular Wavelet–ELM Conjunction Models

The modular wavelet–ELM conjunction models were obtained combining the wavelet–ELM conjunction models with the modular learning approach, that is, for each station, several wavelet–ELM conjunction models, each of which corresponds to a particular module of the considered cluster number, were developed instead of devoting single wavelet–ELM conjunction model. In order to determine the optimum number of clusters, we used the concept of threshold cluster number. The identified threshold cluster numbers were 29 for the Kharjeguil station, 15 for the Ajmer station and 31 for the Barton Pond rain-gauge data, and then modular learning was conducted in the range cluster numbers from 2 up to 29, from 2 up to 15 and from 2 up to 31, respectively, and the cluster number with the best performance on testing set was considered as the optimal cluster number. In this study, the *RMSE* criterion used for the evaluation. Figure 4 shows the variations of *RMSE* criterion in the above ranges on testing sets of the stations. It is seen from the figure that the optimum numbers of clusters were 26, 14 and 28 for the Kharjeguil, Ajmer and Barton Pond stations, respectively. It should be noted that for each module in the given cluster number, an optimum number of neurons as well as a proper type of activation function in the hidden layer were determined as before, that is, three activation functions were tried one by one and in each trial, the numbers of neurons in the hidden layer were increased gradually starting from 1 up to 100. The criteria to choose the optimum structure was the network's performance on testing set of the module of under consideration. In other words, for each individual module of the cluster number of under consideration, the network with the best performance on testing set of the respective module was considered as the optimal structure for the module. This approach was repeated for all modules of the remaining clusters until the maximum number of clusters reached. The identified architectures, observed number of samples as well as activation functions for each module of the optimum modular wavelet–ELM conjunction models were shown in Table 2 for

the Kharjeguil station, in Table 3 for the Ajmer station and in Table 4 for the Barton Pond rain-gauge data.

3.5 Model Performance Comparison

The performance of the models was assessed using three statistical measures of goodness of fit: The correlation coefficient (r), root mean square errors ($RMSE$) and Nash–Sutcliffe model efficiency coefficient (NS).

The r criterion equation is:

$$r = \frac{\sum_{i=1}^{N}\left(R_i-\overline{R}\right)\left(R_i'-\overline{R'}\right)}{\sqrt{\sum_{i=1}^{N}\left(R_i-\overline{R}\right)^2\sum_{i=1}^{N}\left(R_i'-\overline{R'}\right)^2}} \tag{11}$$

The $RMSE$ is used to measure the average error magnitude. It is given by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(R_i-R_i'\right)^2}{N}} \tag{12}$$

The NS statistical measure is given by:

$$NS = 1-\frac{\sum_{i=1}^{N}\left(R_i-R_i'\right)^2}{\sum_{i=1}^{N}\left(R_i-\overline{R}\right)^2} \tag{13}$$

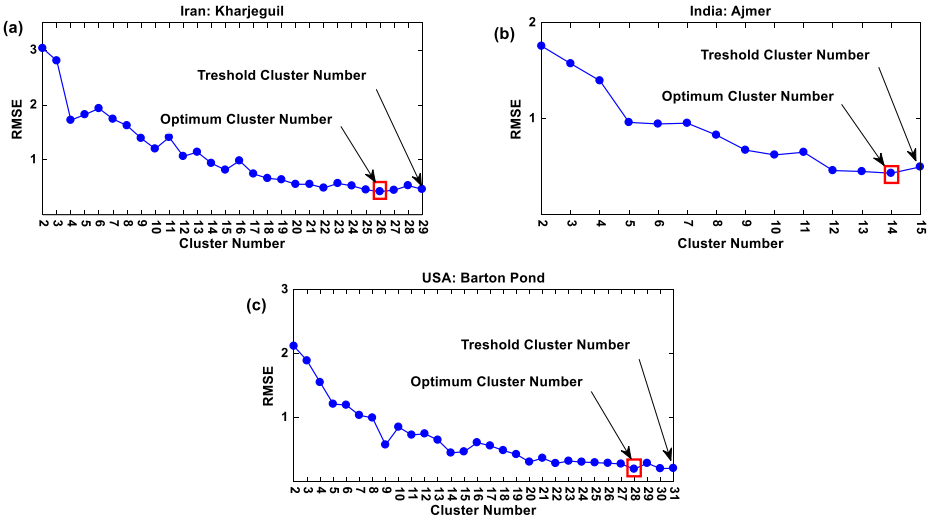

Fig. 4 The *RMSE* variations for various cluster numbers ranging from 2 up to the threshold cluster number, **a** for Kharjeguil, **b** for Ajmer and **c** for Barton Pond

In these equations, N is the number of observations, R_i represents observed rainfall, $R_i^{'}$ stands for forecasted rainfall, $\overline{R}$ denotes average observed rainfall, and $\overline{R'}$ is the average of the forecasted rainfall values.

The r measure describes the degree of collinearity between forecasted and observed data, which ranges from −1 to 1, is an index of the degree of linear relationship between observed and forecasted data. If $r = 0$, no linear relationship exists. If $r = 1$ a perfect positive linear relationship between observed and forecasted data exists. If $r = -1$ a perfect negative linear relationship between observed and forecasted data exists. The smaller *RMSE* value, the better the performance of the model. The *NS* indicates how well the plot of the observed values versus the forecasted values fits the 1:1 line. *NS* ranges from $-\infty$ to 1, with larger *NS* values signifying better model performance.

4 Results and Discussions

4.1 ELM for Rainfall Forecasting

The results of ELM models, for all case studies, were summarized in Table 5. It is apparent that the results are not satisfactory and hence there is a need to invoke some approaches in order to

Table 2 The identified architectures of the modular wavelet–ELM conjunction models for each module of the optimum cluster number of "26" – Kharjeguil station

Module number	Observed number of samples		Activation function	Architecture
	Training	Testing		
1	577	240	"sigmoid"	10-20-1
2	440	123	"sine"	10-5-1
3	372	128	"sine"	10-46-1
4	301	92	"hard-limit"	10-46-1
5	291	124	"sine"	10-57-1
6	83	21	"hard-limit"	10-77-1
7	165	25	"hard-limit"	10-87-1
8	129	35	"sigmoid"	10-40-1
9	95	30	"hard-limit"	10-60-1
10	107	28	"hard-limit"	10-14-1
11	82	23	"sigmoid"	10-2-1
12	63	29	"hard-limit"	10-41-1
13	83	28	"hard-limit"	10-67-1
14	67	22	"hard-limit"	10-39-1
15	62	20	"sine"	10-19-1
16	58	22	"hard-limit"	10-37-1
17	53	19	"sine"	10-10-1
18	48	15	"hard-limit"	10-92-1
19	40	11	"hard-limit"	10-36-1
20	39	15	"hard-limit"	10-77-1
21	32	18	"hard-limit"	10-16-1
22	26	9	"hard-limit"	10-16-1
23	27	6	"hard-limit"	10-86-1
24	21	6	"sigmoid"	10-8-1
25	7	4	"hard-limit"	10-73-1
26	19	3	"hard-limit"	10-65-1

Table 3 The identified architectures of the modular wavelet–ELM conjunction models for each module of the optimum cluster number of "14" − Ajmer station

Module number	Observed number of samples		Activation function	Architecture
	Training	Testing		
1	3375	885	"sine"	35-1-1
2	806	192	"sine"	35-1-1
3	508	131	"sine"	35-1-1
4	517	114	"hard-limit"	35-1-1
5	85	11	"sigmoid"	35-37-1
6	121	22	"sigmoid"	35-37-1
7	91	22	"hard-limit"	35-30-1
8	73	21	"hard-limit"	35-54-1
9	68	18	"hard-limit"	35-71-1
10	61	19	"hard-limit"	35-42-1
11	46	7	"hard-limit"	35-25-1
12	36	11	"sine"	35-8-1
13	29	7	"hard-limit"	35-77-1
14	28	1	"hard-limit"	35-84-1

Table 4 The identified architectures of the modular wavelet–ELM conjunction models for each module of the optimum cluster number of "28" − Barton Pond station

Module number	Observed number of samples		Activation function	Architecture
	Training	Testing		
1	205	87	"sine"	8-1-1
2	120	29	"sine"	8-1-1
3	132	36	"sine"	8-1-1
4	115	30	"sine"	8-1-1
5	70	28	"sine"	8-1-1
6	63	19	"sine"	8-1-1
7	37	20	"sine"	8-9-1
8	43	19	"hard-limit"	8-12-1
9	16	6	"hard-limit"	8-35-1
10	33	11	"hard-limit"	8-35-1
11	28	8	"hard-limit"	8-97-1
12	19	3	"sine"	8-22-1
13	26	6	"hard-limit"	8-64-1
14	22	6	"hard-limit"	8-89-1
15	17	7	"hard-limit"	8-15-1
16	19	6	"hard-limit"	8-49-1
17	21	2	"sine"	8-69-1
18	11	10	"hard-limit"	8-14-1
19	12	5	"hard-limit"	8-86-1
20	15	3	"sine"	8-6-1
21	17	2	"hard-limit"	8-31-1
22	11	2	"hard-limit"	8-52-1
23	10	7	"hard-limit"	8-19-1
24	6	5	"hard-limit"	8-35-1
25	10	2	"hard-limit"	8-99-1
26	7	3	"hard-limit"	8-32-1
27	5	1	"hard-limit"	8-49-1
28	6	2	"hard-limit"	8-67-1

improve the forecasting accuracy of ELM model, the issue that we would investigate in upcoming subsections.

4.2 ELM and Proposed Wavelet-ELM Conjunction Models for Rainfall Forecasting

In this subsection, for all case studies, the results of the wavelet-ELM conjunction models were investigated and presented in Table 6. Furthermore, these results were compared with those of obtained by the use of single ELM (see Table 5) in order to highlight the importance of using wavelet analysis in providing useful and meaningful input information to the ELM model. As evident in Table 6, for the Kharjeguil station, the wavelet-ELM conjunction model illustrated the best result in terms of $r = 0.899$, $RMSE = 4.307$ and $NS = 0.808$ in the training period and the best result in terms of $r = 0.872$, $RMSE = 4.424$ and $NS = 0.760$ in the testing period while these criteria for the single ELM model were $r = 0.357$,$RMSE = 6.892$ and $NS = 0.128$, and $r = 0.370$,$RMSE = 7.298$ and $NS = 0.131$, respectively. For the Ajmer station, the inclusion of wavelet analysis results to ELM also indicated a significant improvement compared to the single ELM model. In this case, the wavelet-ELM conjunction model outperformed with $r = 0.908$, $RMSE = 3.280$ and $NS = 0.825$ in the training period and with $r = 0.892$, $RMSE = 2.413$ and $NS = 0.794$ in the testing period. These performance statistics for the single ELM model were of $r = 0.327$,$RMSE = 7.423$ and $NS = 0.107$ in the training period and of $r = 0.348$,$RMSE = 5.001$ and $NS = 0.118$ in the testing period. For the Barton Pond rain-gauge data, it can also be seen that the best model was the wavelet-ELM conjunction model, which had a training R of 0.872, a training $RMSE$ of 2.769, a training NS of 0.761, a testing r of 0.888, a testing $RMSE$ of 3.048 and a testing NS of 0.780. The single ELM model of this station had a training R of 0.217, a training $RMSE$ of 5.532, a training NS of 0.047, a testing r of 0.204, a testing $RMSE$ of 6.370 and a testing NS of 0.040.

Overall, it can be seen that for 1–day ahead forecasting the proposed wavelet-ELM conjunction models provided more accurate forecasting results than the single/regular ELM models.

4.3 Proposed Modular Wavelet-ELM Conjunction Models

In this subsection, the results obtained from the modular wavelet–ELM conjunction models were compared with wavelet–ELM conjunction models to examine the effect of proposed modular learning approach of this study in forecasting daily rainfall. Table 7 shows the modular wavelet–ELM conjunction models performance statistics (r, $RMSE$ and NS) for the best (or optimum) modular wavelet–ELM conjunction models for three case studies for $1-$ day ahead rainfall forecasting. It should be noted that this illustration is solely given based on testing data sets of respective stations.

Table 5 The performance statistics of the single ELM models during the training and testing periods

Stations	Training			Testing		
	r	RMSE	NS	r	RMSE	NS
Kharjeguil, Iran	0.357	6.892	0.128	0.370	7.298	0.131
Ajmer, India	0.327	7.423	0.107	0.348	5.001	0.118
Barton Pond, USA	0.217	5.532	0.047	0.204	6.370	0.040

For the Kharjeguil station, it can be seen that the best model overall for 1 − day lead time forecasting was the best modular wavelet–ELM conjunction model, which had a testing *r* of 0.998, a testing *RMSE* of 0.423, and a testing *NS* of 0.997. As seen in Table 6, the wavelet–ELM conjunction model (i.e. without modular learning approach) had a testing *r* of 0.872, a testing *RMSE* of 4.424, and a testing *NS* of 0.760. Figure 5 shows the hyetographs and scatter plots of the results at 1–day ahead forecasting of ELM, wavelet−ELM and modular wavelet–ELM using the testing rainfall data set of Kharjeguil, where the hyetograph is zoomed in selected ranges for better visual inspection. ELM seriously underestimates high-intensity rainfalls. Wavelet–ELM improves noticeably the accuracy of forecasting although high-intensity rainfalls (or peak values) are still underestimated. From the representative hyetograph, it can be seen that the modular wavelet–ELM model reproduces perfectly the corresponding observed rainfall data and provides highly accurate forecasts, which is further revealed by the scatter plot with perfect match of the diagonal.

For the Ajmer station, it can be seen that the best model overall for forecasting was the best modular wavelet–ELM conjunction model, which had a testing *r* of 0.996, a testing *RMSE* of 0.433, and a testing *NS* of 0.993. These performance statistics for the wavelet–ELM conjunction model in Table 6 were a testing *r* of 0.892, a testing *RMSE* of 2.413, and a testing *NS* of 0.794, which again indicate the superiority of the modular wavelet–ELM conjunction model to the wavelet–ELM conjunction model. One-step lead forecasts of ELM, wavelet–ELM and wavelet–ELM with the help of modular learning approach are shown in Fig. 6 in the form of hyetographs and scatter plots (the former is zoomed in selected ranges for better visual inspection). Compared with ELM, scatter plot of wavelet–ELM is closer to the exact fit line, which means that the daily rainfall process is fitted appropriately. Nevertheless, some peak values still remain underestimated although wavelet–ELM shows a better ability to capture the peak values than ELM. Regarding the modular wavelet–ELM conjunction model, the scatter plot with an extremely low dispersion around the exact fit line indicates that the rainfall process is almost perfectly reproduced. The representative hyetograph shows that the peak times and peak values are also accurately forecasted.

For the Barton Pond rain-gauge data, good accuracy of forecasting was also made by the best modular wavelet–ELM conjunction model, which had a testing *r* of 0.999, a testing *RMSE* of 0.196, and a testing *NS* of 0.999. As shown in Table 6, the above statistics for the wavelet–ELM conjunction model were 0.888, 3.048, and 0.780, respectively. These statistics indicate the noticeable accuracy of the proposed modular wavelet–ELM conjunction model. Figure 7 shows the hyetograph and scatter plots of both the observed data and the forecasted obtained by using ELM, wavelet–ELM and modular wavelet–ELM of the testing period. It was

Table 6 The performance statistics of the proposed wavelet–ELM conjunction models during the training and testing periods

Stations	Training			Testing		
	r	RMSE	NS	r	RMSE	NS
Kharjeguil, Iran	0.899	4.307	0.808	0.872	4.424	0.760
Ajmer, India	0.908	3.280	0.825	0.892	2.413	0.794
Barton Pond, USA	0.872	2.769	0.761	0.888	3.048	0.780

Table 7 The performance statistics of the proposed modular wavelet–ELM conjunction models during the training and testing periods

Stations	Training			Testing		
	r	RMSE	NS	r	RMSE	NS
Kharjeguil, Iran	0.999	0.367	0.998	0.998	0.423	0.997
Ajmer, India	0.998	0.433	0.997	0.996	0.433	0.993
Barton Pond, USA	0.999	0.132	0.999	0.999	0.196	0.999

obviously seen from the hyetographs and scatter plots that the wavelet–ELM conjunction model forecasts were closer to the corresponding observed rainfall values than those of the ELM model which seriously underestimates high-intensity rainfalls. Nevertheless, some high-intensity rainfalls still remain overestimated by wavelet–ELM. Regarding the modular wavelet–ELM conjunction model, the scatter plot with perfect match of the exact fit line indicates that the corresponding observed rainfall values is perfectly reproduced. The representative hyetograph shows that the peak times and peak values are also forecasted with a very high accuracy.

Overall, although both wavelet–ELM and modular wavelet–ELM could give highly accurate and reliable forecasting performance; nevertheless, the results of this study suggest that modular wavelet–ELM is a better choice since it could forecast the observed hyetographs in all cases almost perfectly.

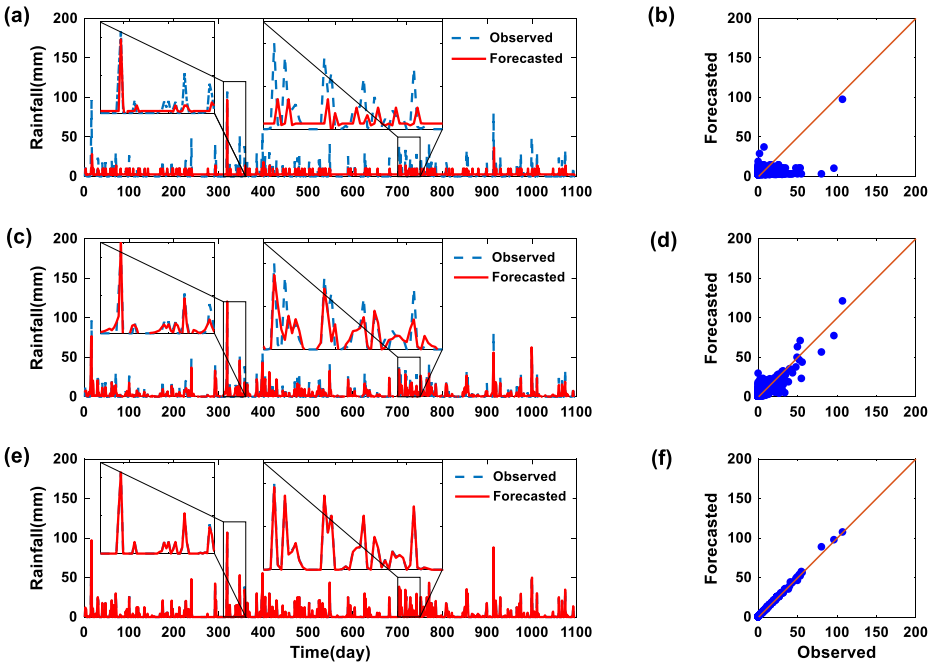

Fig. 5 Hyetographs and scatter plots of one-step-ahead forecast using ELM, wavelet−ELM and modular wavelet −ELM in the test period for Kharjeguil (**a** and **b** from ELM, **c** and **d** from wavelet−ELM, and **e** and **f** from modular wavelet−ELM)

4.4 The Investigation of Optimum Cluster Number through the New Concept of "Threshold Cluster Number"

Figure 4 shows the *RMSE* variations of each cluster number (in the range of 2 up to the threshold cluster number) for each case study. It is evident that the optimum cluster number of each case is located in the vicinity of its respective threshold cluster number. For example, for Kharjeguil station, the optimum cluster number is the third cluster number (i.e. cluster number of "26") prior to the threshold cluster number. For Ajmer station, the optimum cluster number is the first cluster number (i.e. cluster number of "14") prior to the threshold cluster number. Finally, for Barton Pond station, the optimum cluster number is the third cluster number (i.e. cluster number of "28") prior to the threshold cluster number. These results indicate that the optimum cluster number is located in the vicinity of the threshold cluster number while in all former studies the search for the optimum cluster number has been conducted in the beginning of the range introduced in this study, that is, from 2 up to 4. We would call this range as sub-optimal range. For example, Furundzic (1998) employed cluster number of 3 to decompose the rainfall–runoff process input–output space. Zhang and Govindaraju (2000) used cluster number of 3 representing low-, medium, and high-magnitude flows. Jain and Srinivasulu (2006) developed modular models using cluster number of 4. Wu et al. (2010) used cluster number of 3 based on an assumption that three magnitudes of rainfall (i.e. low, medium, and high) exist. Although these studies obtained improved performance; however, the results of this study indicate that their results are suboptimal since they relied on the beginning of the range

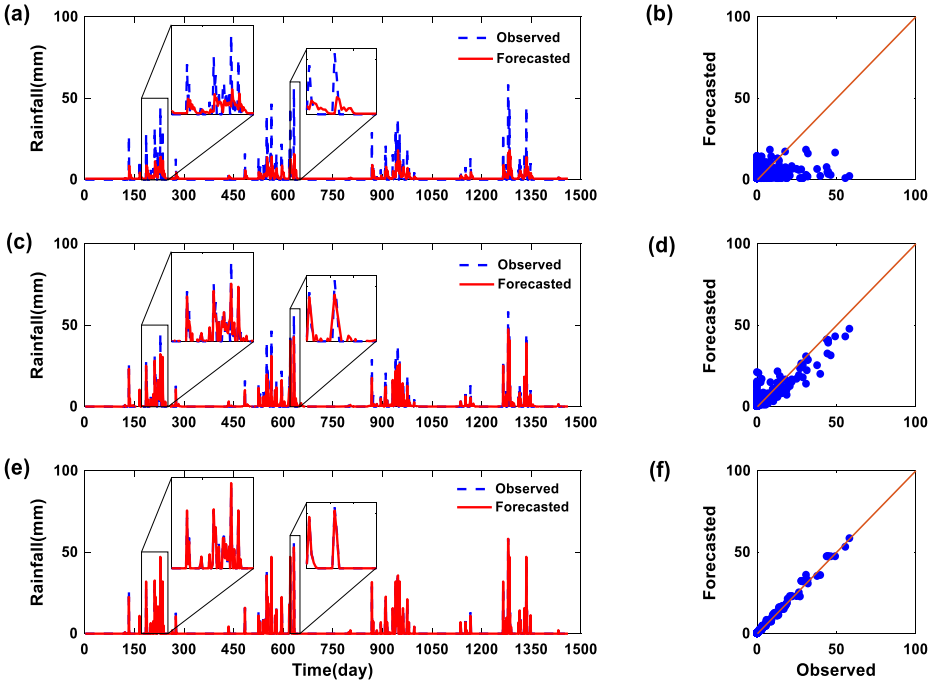

Fig. 6 Hyetographs and scatter plots of one-step-ahead forecast using ELM, wavelet−ELM and modular wavelet −ELM in the test period for Ajmer (**a** and **b** from ELM, **c** and **d** from wavelet−ELM, and **e** and **f** from modular wavelet−ELM)

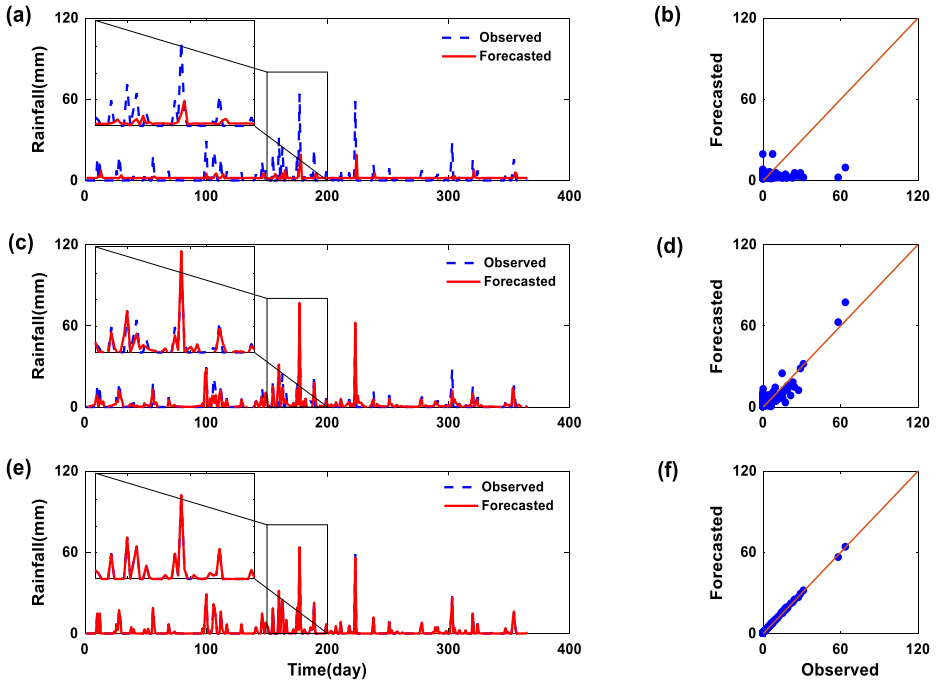

Fig. 7 Hyetographs and scatter plots of one-step-ahead forecast using ELM, wavelet–ELM and modular wavelet –ELM in the test period for Barton Pond (**a** and **b** from ELM, **c** and **d** from wavelet–ELM, and **e** and **f** from modular wavelet–ELM)

introduced here while the optimum cluster number is located somewhere at the end of the range introduced here.

Based on the results of this study, it can be recommended that once the threshold cluster number through the approach introduced in this study was obtained the search for optimum cluster number should be conducted in the last parts of the range introduced here (e.g. within the last five cluster numbers of the range including the threshold cluster number itself) in order to ensure achieving the optimum cluster number.

5 Conclusions

In this study, a new hybrid wavelet-ELM model was proposed for rainfall time series forecasting for three rain-gauge stations, the Kharjeguil, Ajmer and Barton Pond, located in Iran, India and the United States, respectively. For this purpose, the discrete wavelet analysis, which capture the multi–scale features of a signal, was first used for decomposing the time series into several wavelet coefficients. Then, these wavelet coefficients were imposed to the ELM model as inputs so as to forecast the rainfall for one–day ahead. The Daubechies mother wavelet in order 2 (db2) was used for decomposing rainfall time series. The accuracy of the wavelet–ELM conjunction model was further improved using the modular learning approach. To do this, the SOM clustering algorithm was used. An approach to determine the optimum number of clusters in modular learning was introduced based on the concept of "threshold cluster number". This

study indicated that the optimum forecast results were obtainable through the optimum cluster number defined by this concept.

The proposed modular wavelet–ELM conjunction model significantly increased the forecast accuracy and performed much better than both the wavelet–ELM and single ELM. The proposed wavelet–ELM conjunction model results demonstrate that, it has better forecasts than the single ELM model because of the utilization of multi–resolution time series sub-signals as inputs. The proposed modular wavelet–ELM conjunction model results demonstrate that, it has better forecasts than the wavelet–ELM conjunction model because of the utilization of a separate model for each cluster. The highly accurate forecast results for all three stations illustrated that, the modular wavelet–ELM conjunction model is a very useful new model for rainfall forecasting. As a conclusion, it is clear that the modular wavelet–ELM conjunction model which was created by consolidating three routines, the ELM, wavelet and modular learning approach, has obvious advantages over both the wavelet–ELM and single ELM models in rainfall forecasting by providing a highly accurate and reliable model.

In order to strengthen the findings of this study, it is suggested that, for future studies, one ought to endeavor the utilization of the modular wavelet–ELM conjunction model in the rainfall time series modelling for other stations in other periods with longer lead times such as two or three days ahead. Also, it can be investigated with other hydrological data.

References

Abdullah SS, Malek MA, Abdullah NS, Kisi O, Yap KS (2015) Extreme learning machines: a new approach for prediction of reference evapotranspiration. J Hydrol 527:184–195

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

Adamowski J, Chan HF (2011) Awavelet neural network conjunction model for groundwater level forecasting. J Hydrol 407:28–40

Al-geelani NA, Piah MAM, Shaddad RQ (2012) Characterization of acoustic signals due to surface discharges on H.V. glass insulators using wavelet radial basis function neural networks. Appl Soft Comput 12:1239–11246

Chan JCL, Shi JE (1999) Prediction of the summer monsoon rainfall over South China. Int J Climatol 19(11): 1255–1265

Chu PS, He YX (1994) Long-range prediction of Hawaiian winter rainfall using canonical correlation analysis. Int J Climatol 14(6):659–669

DelSole T, Shukla J (2002) Linear prediction of Indian monsoon rainfall. J Clim 15(24):3645–3658

Deo RC, Şahin M (2015) Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in eastern Australia. Atmos Res 153:512–525

Deo RC, Tiwari MK, Adamowski JF, Quilty JM (2016) Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. Stoch Env Res Risk A. https://doi.org/10.1007/s00477-016-1265-z

Diomede T, Davolio S, Marsigli C, Miglietta MM, Moscatello A, Papetti P, Paccagnella T, Buzzi A, Malguzzi P (2008) Discharge prediction based on multi-model precipitation forecasts. Meteorog Atmos Phys 101(3–4): 245–265

El-Shafie A, Jaafer O, Seyed A (2011) Adaptive neuro-fuzzy inference system based model for rainfall forecasting in Klang River, Malaysia. Int J Phys Sci 6(12):2875–2888

Freiwan M, Cigizoglu HK (2005) Prediction of total monthly rainfall in Jordan using feed forward backpropagation method. Fresenius Environ Bull 14(2):142–151

French MN, Krajewski WF, Cuykendal RR (1992) Rainfall forecasting in space and time using a neural network. J Hydrol 137:1–37

Furundzic D (1998) Application example of neural networks for time series analysis: rainfall-runoff modelling. Signal Process 64:383–396
Ganguly AR, Bras RL (2003) Distributed quantitative precipitation forecasting (DQPF) using information from radar and numerical weather prediction models. J Hydrometeorol 4(6):1168–1180
Hamidi O, Poorolajal J, Sadeghifar M, Abbasi H, Maryanaji Z, Faridi HR, Tapak L (2015) A comparative study of support vector machines and artificial neural networks for predicting precipitation in Iran. Theor Appl Climatol 119(3):723–731
He S, Raghavan SV, Nguyen NS, Liong SY (2013) Ensemble rainfall forecasting with numerical weather prediction and radar-based nowcasting models. Hydrol Process 27(11):1560–1571
Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1):489–501
Jain A, Srinivasulu S (2006) Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques. J Hydrol 317:291–306
Kalteh AM (2013) Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. Comput Geosci 54:1–8
Kalteh AM (2015) Wavelet genetic algorithm-support vector regression (wavelet GA-SVR) for monthly flow forecasting. Water Resour Manag 29(4):1283–1293
Kalteh AM, Hjorth P, Berndtsson R (2008) Review of the self-organizing map (SOM) approach in water resources: analysis, modelling and application. Environ Model Softw 23(7):835–845
Kohonen T (1982a) Analysis of a simple self-organizing process. Biol Cybern 44:135–140
Kohonen T (1982b) Self-organized formation of topologically correct feature maps. Biol Cybern 43:59–69
Leuenberger M, Kanevski M (2015) Extreme learning machines for spatial environmental data. Comput Geosci 85:64–73
Li F, Zeng QC (2008) Statistical prediction of east Asian summer monsoon rainfall based on SST and sea ice concentration. J Meteorol Soc Jpn 86(1):237–243
Liu Z, Zhou P, Chen G, Guo L (2014) Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting. J Hydrol 519(27):2822–2831
Mallat SG (1989) A theory for multi resolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 11(7):674–693
Marzano FS, Fionda E, Ciotti P (2006) Neural-network approach to ground- based passive microwave estimation of precipitation intensity and extinction. J Hydrol 328:121–131
Munot AA, Kumar KK (2007) Long range prediction of Indian summer monsoon rainfall. J Earth Syst Sci 116(1):73–79
Navone HD, Ceccatto HA (1994) Predicting indian monsoon rainfall: a neural network approach. Clim Dyn 10: 305–312
Nayagam LR, Janardanan R, Mohan HSR (2008) An empirical model for the seasonal prediction of southwest monsoon rainfall over Kerala, a meteorological subdivision of India. Int J Climatol 28(6):823–831
Pongracz R, Bartholy J, Bogardi I (2001) Fuzzy rule-based prediction of monthly precipitation. Phys Chem Earth Part B 26(9):663–667
Sheng C, Gao S, Xue M (2006) Short-range prediction of a heavy precipitation event by assimilating Chinese CINRAD-SA radar reflectivity data using complex cloud analysis. Meteorog Atmos Phys 94(1–4):167–183
Taormina R, Chau KW (2015) Data-driven input variable selection for rainfall–runoff modeling using binary-coded particle swarm optimization and extreme learning machines. J Hydrol 529:1617–1632
Venkata Ramana R, Krishna B, Kumar SR, Pandey NG (2013) Monthly rainfall prediction using wavelet neural network analysis. Water Resour Manag 27:3697–3711
Wang W, Ding J (2003) Wavelet network model and its application to the prediction of the hydrology. Nat Sci 1(1):67–71
Wu CL, Chau KW, Fan C (2010) Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. J Hydrol 389:146–167
Yates DN, Warner TT, Leavesley GH (2000) Prediction of a flash flood in complex terrain. Part II: a comparison of flood discharge simulations using rainfall input from radar, a dynamic model, and an automated algorithmic system. J Appl Meteorol 39(6):815–825
Zhang B, Govindaraju RS (2000) Prediction of watershed runoff using Bayesian concepts and modular neural networks. Water Resour Res 36(3):753–762