Natural Hazards (2021) 108:799–817
https://doi.org/10.1007/s11069-021-04706-9

ORIGINAL PAPER

Rainfall prediction using optimally pruned extreme learning machines

Huajin Li[1] · Yusen He[2] · He Yang[1] · Yong Wei[3] · Songlin Li[4] · Jianqiang Xu[3]

Received: 2 September 2020 / Accepted: 16 March 2021 / Published online: 26 March 2021

Abstract
Rainfall impacts local water quantity and quality. Accurate and timely prediction of rainfall is highly desirable in water management and hydrogeologic hazards mitigation, which is critical for the environmentally sustainable development. Previous rainfall prediction processes are complex and computationally costly for its intrinsic high uncertainty and variability. In this paper, a data-driven approach is applied to predict the rainfall based on historic data via time-series modeling and optimally pruned extreme learning machine (OP-ELM). The rainfall datasets collected from six counties within Three Georges Reservoir are utilized as case studies in this research, first, the rainfall data is pre-processed with outlier removal and missing value imputation, and the monthly ahead difference is computed based on the instant average of monthly rainfall. Next, the autocorrelation function and partial autocorrelation function were computed and the Ljung–Box test statistic is utilized to explore the significance of the historic lagged-series. All the statistically significant historic lags are selected as inputs for prediction algorithms. Last, an OP-ELM algorithm is developed to predict the monthly rainfall with tenfold cross validation. Four activation functions: the sigmoid, sine, hardlim, and radial basis function, are considered in the OP-ELM. The prediction performance is evaluated with metrics including mean absolute error, mean absolute percentage error, root mean square error, and max error rate. Overall, the computational results indicate the proposed framework outperforms the other benchmarking machine learning algorithms through six case studies in Three Gorges Reservoir, China.

Keywords Rainfall prediction · Sustainable development · Time-series modeling · OP-ELM

1 Introduction

Rainfall prediction is highly relevant for improving the resiliency and well-being of communities as it directly impacts agriculture, the provision of drinking water; sanitation, electricity generation, and the occurrence of flooding and landslides (e.g. Mugume et al. 2018;

✉ Yusen He
yusen-he@uiowa.edu

Extended author information available on the last page of the article

Esteves et al. 2019; Li et al. 2020), which makes an obstacle for sustainable development of environment systems (Tian et al. 2019; Wu et al. 2020). However, efficient and reliable rainfall prediction is elusive, as it is the outcome of complex nonlinear processes within the atmosphere that are highly susceptible to initial conditions.

Remote sensing using radars and satellite images is a feasible approach to identify and investigates the short-term rainfall. However, their reliability decays quickly, particularly in mountainous regions (Toth et al. 2000). Time series analysis has been approved to be a better alternative solution if longer-term rainfall forecasts are needed. Pankratz (1983) pioneered the use of autoregressive integrated moving average models. Labat et al. (2000) explored patterns expanding diverse time scales through wavelet analysis. Pal et al. (2020) conducted a research on the temporal pattern of rainfall during the summer monsoon using hurst exponent, and a case study is carried out in northeast India. Ray et al. (2021) proposed a surface temperature prediction framework with a two-state Markov chain approach and an autoregressive approach, and discovered a strong correlation between the surface temperature and rainfall volume. Most recently, machine learning (ML) has become a predominant tool for solving nonlinear problems that are highly conditioned by initial values, as long extensive data is provided. Thus, it is well suited for developing models for rainfall estimation (He et al. 2017, 2018; He and Kusiak 2018; Xu et al. 2019; Ouyang et al. 2019).

Of the initial use of ML techniques for rainfall characterization can be traced back to the work of Sivapragasam et al. (2001) who proposed a hybrid model comprised of a support vector machine (SVM) coupled with singular spectrum analyses. Then Chattopadhyay and Chattopadhyay (2008) made further improvements as they introduced artificial neural networks (ANN) in rainfall prediction. These approaches have been further optimized by involving wavelets for steam flow predictions (Kişi 2009), and have been optimized through the inclusion of radial basis functions (Wu et al. 2015), Shannon entropy (Ray and Chattopadhyay 2021) and probability distribution (Chattopadhyay et al. 2020).

Albeit ANNs and SVMs have proven to be helpful for rainfall prediction, the latest models they produce have been observed to be extremely complex, involving several layers of neurons, and including several activations and/or kernel functions. These traits reduce their efficiency. These shortcomings can be addressed by extreme learning machines (ELM) which consider a single hidden feed-forward layer of neurons, allowing for widespread generalization while keeping a relatively simple structure (Huang et al. 2006). Thus, they can be implemented quickly and with lower computational demands, making their use widespread in forecast application (Suresh et al. 2009; Mohammed et al. 2011; Kasun et al. 2013; Deo and Şahin 2015; Li et al. 2018; He and Kusiak 2018).

This study goes a step forward and calibrates an enhanced form of ELM, an Optimally Pruned Extreme Learning Machine (OP-ELM) (Miche et al. 2010). OP-ELMs systematically rank and select best-performing algorithms while removing redundant neurons, leading to streamlined models by eliminating superfluous elements that only enhance accuracy slightly despite disproportionately increasing complexity and resource consumption.

This paper predicts monthly rainfall in terms of both short-term (less than a year) averages of monthly rainfall and the averages of the whole-time history. The proposed approach is tested considering six locations within the three Gorges Reservoir in China, finding that it overperforms ANNs, SVMs, and non-optimized ELMs.

In this framework, to obtain more characteristics of rainfall dataset, cis computed. In order to build a predictive framework, the monthly ahead difference (MAD) is acquired by the difference of instant rainfall and IAMR in the previous month. To confirm the feasibility of MAD based prediction model, a statistical analysis of rainfall and MAD are conducted, and ACF and PACF are selected to assess the periodicity.

2 Materials and methods

In this research, a novel data-driven predictive framework is proposed. As shown in Fig. 1, first, t the instant average of monthly rainfall (IAMR) is computed, and the monthly ahead difference (MAD) is acquired by the difference of instant rainfall and IAMR in the previous month. Then, the autoregressive moving average (ARMA) time series model is selected to analyze assess the periodicity of raw rainfall dataset and MAD, and autocorrelation factors (ACFs) of the ARMA model are computed to define all inputs. At last, an OP-ELM algorithm using tenfold cross-validation is employed for prediction, four activation function are considered for better performance.

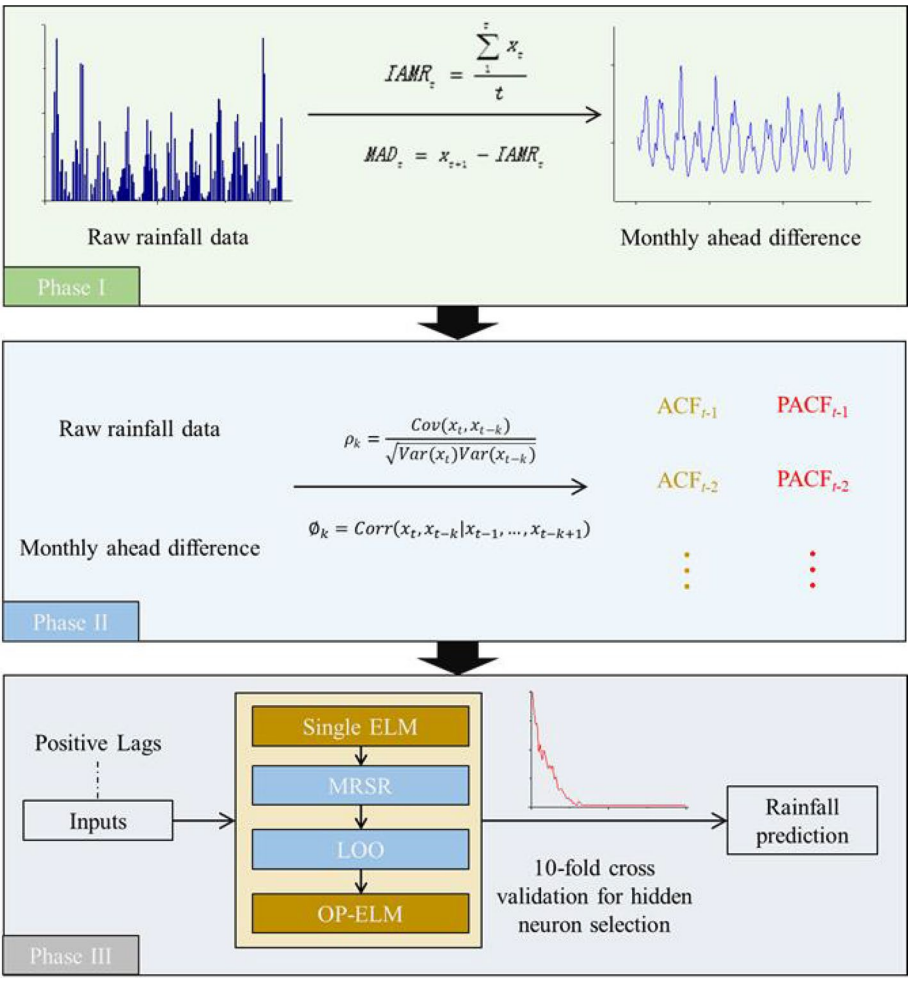

Fig. 1 Flow chart of the proposed framework

2.1 Data preprocess

Rainfall monitoring dataset has strong seasonal randomness, which is an obstacle for the operational prediction community. It is necessary to gain more characteristics based on rainfall dataset. The average monthly rainfall in different years approaching a value, and the difference between monthly rainfall and the mean has periodic characteristic. This knowledge indicates that the difference in rainfall prediction is highly promising.

The first step is to construct the instant average of monthly rainfall, and the function of IAMR can be described as (1):

$$\mathrm{IAMR}_t = \frac{\sum_1^t x_t}{t} \tag{1}$$

where x_t represents the monthly rainfall, tth is the number of months.

The second step is to build a data system of the predictive model. In this research, monthly ahead difference is computed by (2):

$$\mathrm{MAD}_t = x_{t+1} - \mathrm{IAMR}_t \tag{2}$$

Based on the MAD, the rainfall 1-month ahead of prediction can be obtained.

At last, a time series analysis is conducted to assess the temporal patterns of the rainfall data and MAD. The temporal correlation between the current data and the previous data has been computed based on autocorrelation function (ACF) and partial autocorrelation function (PACF). The details of ACF and PACF are as follows (McLeod and Li 1983):

$$\rho_k = \frac{\mathrm{Cov}(X_t, X_{t-k})}{\sqrt{\mathrm{Var}(X_t)\mathrm{Var}(X_{t-k})}} = \frac{\mathrm{Cov}(X_t, X_{t-k})}{\mathrm{Var}(X_t)} = \frac{\gamma_k}{\gamma_0} \tag{3}$$

$$\phi_k = \frac{\rho_k - \sum_1^{k-1} \phi_{k-1}\rho_{k-1}}{1 - \sum_1^{k-1} \phi_{k-1}\rho_k} \tag{4}$$

where ρ_k is the ACF of the kth lag of the variable, ϕ_k is the PACF of the kth lag of the variable, x_t is the ith data sample. In practice, both the ACF and PACF are nonzero for most of the case studies in time-series of problems, and the Ljung–Box statistic (Lee 2016) is applied to measure the statistical significance of the computed ACFs and PACFs. The significance level is set as 0.05 for all Ljung–Box tests.

2.2 Optimally pruned extreme learning machine

2.2.1 Extreme learning machine

The ELM is a machine learning algorithm making use of the single hidden layer feedforward neural network to solve the problem of classification and regression (Huang et al. 2006). Different from tradition neural network, only one parameter of ELM needs to set makes ELM with the advantages of generalization performance and fast learning speed.

During practical applications, the first step of ELM training process lies in the random initialization of the weights and biases of the SLFN. Then, the hidden layer output matrix

and the output weights can be calculated by input weights and biases. The network can be obtained by low computational cost with fewer steps (Miche et al. 2010). The SLFNs mathematically models as (5) and (6).

$$y_j = f_L(x_j) \tag{5}$$

$$f_L(x_j) = \sum_{i=1}^{L} \beta_i G(a_i, b_i, x_j), \quad 1 \le j \le N \tag{6}$$

where y_j represents the output. L represents the hidden neurons. a_i represents the weight vector connecting the ith hidden neuron and the input neurons. b_i represents the threshold of the ith hidden neuron. β_i represents the weight vector connecting the ith hidden neuron and the output neurons. x_j represents the inner product of inputs x_i, and the $G(\cdot)$ represents the activation function.

And the (6) can be writes compactly as (7).

$$H\beta = Y \tag{7}$$

where $\beta = \begin{pmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{pmatrix}^T$, $Y = \begin{pmatrix} y_1^T \\ \vdots \\ y_N^T \end{pmatrix}^T$. H represents the hidden layer output matrix.

In short, the ELM algorithm can be summarized as three steps:

Step 1 assign input weight a_i and biases b_i randomly, and $1 \le i \le L$.

Step 2 compute the hidden layer output matrix H from (8);

$$H = \begin{pmatrix} a_1, b_1, x_1 & \cdots & a_L, b_L, x_1 \\ \vdots & \ddots & \vdots \\ a_1, b_1, x_N & \cdots & a_L, b_L, x_N \end{pmatrix} \tag{8}$$

Step 3 compute the output weights matrix β from (9).

$$\beta = H^{\dagger} Y \tag{9}$$

where $H^{\dagger}$ represents Moore–Penrose generalized inverse of the matrix H(Liang et al. 2006; Petković, and Stanimirović 2011).

In addition, five activation functions (Camps-Valls et al. 2004; Ouyang et al. 2016) namely sigmoid (10), sine (11), hardlim (12), radial basis function (13), and tanh (14) are selected in this study which can be computed as follows:

$$G(z) = {}^{1}/_{1 + e^{-z}} \tag{10}$$

$$G(z) = \sin(z) \tag{11}$$

$$G(z) = \begin{cases} 1, & z > 0 \\ 0, & z = 0 \\ -1, & z < 0 \end{cases} \tag{12}$$

$$G(z, z') = \exp\left(\frac{-z - z'}{2\sigma^2}\right) \tag{13}$$

$$G(z) = \tanh(z) \tag{14}$$

where z is the vector of the input data, σ denotes the standard deviation of the input data.

2.2.2 Optimally pruned extreme learning machine

Based on classical ELM, combing the OP-ELM with the multiresponse sparse regression (MRSR) using the leave-one-out (LOO) validation theory (Cawley and Talbot 2003) to get rid of uselessness hidden neurons. As a function approximator, OP-ELM includes three important steps is summarized as below:

Step 1 SLFN construction with extreme learning machine;
Step 2 rank the best hidden neurons of ELM with MRSR algorithm;
The MRSR adds each column of the regressor matrix $X = [x_1 \cdots x_n]$ one by one to construct $Y^{nz} = X\beta^{nz}$. Where Y^{nz} is the target approximation of the algorithm. β^{nz} is the weight matrix has nz nonzero rows at nzth step of the MRSR. A fresh nonzero row and a fresh column of regressor matrix will be added to the algorithm in each fresh step. And more details of the MRSR algorithm is described in Similä and Tikka (2005).
Step 3 prune the useless hidden neurons with LOO validation technique.
Owing to the single ranking function of MRSR, the final decision over the optimal number of hidden neurons for OP-ELM can resort to a LOO error technique. And the LOO error will be computed by (15).

$$\varepsilon^{\text{PRESS}} = \left\| D(y - H(H^T H)^{-1} H^T y) \right\|_2^2 \tag{15}$$

where $\varepsilon^{\text{PRESS}}$ is called PREdiction Sum of Squares statistic. D represents diagonal matrix with $D_{ii} = \frac{1}{1-(H(H^T H)^{-1} H^T)_{ii}}$ (Grigorievskiy et al. 2014).

2.3 Benchmarking machine learning algorithms

In this research, three benchmarking machine learning algorithms including artificial neural network, support vector machine and classical extreme learning machine are compared with the OP-ELM in rainfall prediction.

The artificial neural network with back-propagation (BP) to optimize performance of ANN (Horikawa et al. 1992) is selected in this research. The structure of the ANN algorithm includes the input layer, the hidden layer, and the output layer. Activation function of this ANN select sigmoid function. The number of hidden layers with values of 1, 2, 3, 4 and the number of hidden neurons in each hidden layer with values of 5, 10, 15, 20, 25, and 30 are all evaluated during training process via tenfold cross validation to determine the optimal structure.

The support vector machine algorithm provides the unique maximal margin separating hyperplane based on support vector (Hearst et al. 1998). The parameters of SVM include capacity factor C and $\gamma = 1/2\sigma^2$. Kernel function of the SVM selected Gaussian kernel function. C with the values of 1, 10, 100, 1000 and σ with the values of 0.001, 0.01, 0.1,1. All

the parameters evaluated during training process via tenfold cross validation to determine the optimal setting of parameters.

The classical extreme learning machine algorithm is introduced in Sect. 2.2. The parameters of ELM and OP-ELM include activation function and number of hidden neurons. Five kinds of activation function, namely sigmoid, sine, hardlim, radial basis function, and tanh and the number of hidden neurons with values of 3, 6, 9, …, 60 are all evaluated via tenfold cross validation to determine the optimal setting of parameters.

2.4 Model evaluation metrics

In this research, the mean absolute error [MAE (16)], mean absolute percentage error [MAPE (17)], root mean square error [RMSE (18)] and max error rate [MER (19)] are selected in this paper to assess the accuracy of the proposed framework.

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|\overline{y}_t - y_t\right| \tag{16}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\overline{y}_t - y_t}{y_t}\right| \tag{17}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\overline{y}_t - y_t)^2} \tag{18}$$

$$\mathrm{MER} = \frac{\max(\overline{y}_t - y_t)}{y_{\mathrm{MER}}} \tag{19}$$

where y_t represents the actual rainfall value; $\overline{y}_t$ represents the predicted rainfall value; n represents the number of test dataset; y_{MER} represents the actual rainfall value at where max error happened.

3 Study area and dataset description

The Tree Gorges Dam is the largest hydroelectricity project worldwide as it is capable of generating 22,500 MW. However, despite the hydro power generation, the main purpose of the dam construction is to enhance the flood control capacity in order to prevent catastrophic flooding during the rainy season in the summer time. This leads to the formation of a large reservoir which has a total volume of 39 billion m^3 widely spreading over an area of 1080 km^2, while spanning 660 km backward from the dam (Fig. 2).

The geological and geomorphological profile of the Three Gorges Reservoir (TGR) are complex. First, the relief in TGR is usually steep as it on average includes terrain above 2300 masl while the riverbank sections are always at sea level. Moreover, most (72%) of the banks of the TRG reservoir are underlain by red strata which is a geological unit comprised of interbedded layers of sandstone and mudstone deposited during the Jurassic and the Triassic. Tectonism and a humid environment together resulted the composition of these geomaterials prone to weathering (Wang et al. 2017).

The predominant climate within the TGR is humid subtropical according to Kopen's classification. The largest annual average rainfall is 1930 mm, observed at Wanzhou, which is a city about 200 km distant from the dam. The precipitation on both far west and far east directions are smaller in composition. The smallest precipitation is 995 mm per year at Zigui which is close to the dam site.

The engineering operation conditions of the reservoir are challenging. The water level oscillates between 145 and 175 m from its datum, allowing for a systematic variation of 2.5 billion m^3 of storage within the reservoir. Most of the reservoirs' capacity has been fulfilled during the rainy season as the water level within the reservoir shapely increases.

Rainfall prediction is of good interest at the Tree Gorges Area for assessing potential geological risks before occurrence, and for improving reservoir water level operation. This would lead to better administration of water resources on the lower basin of the Yangtze river, benefiting millions of residents. Data for modeling and validation of the framework was collected in six weather stations obtaining notes that span at least 15 years. It was obtained using tipping bucket rain gauge pluviometers through aggregation of monthly measurements. Obtained time histories are presented in Fig. 3. While main statistical descriptors of the dataset are summarized in Table 1.

4 Computing results

The main objective of this paper is to propose a statistical based machine learning technique for predicting monthly rainfall using MAD and OP-ELM. As discussed earlier, the three cases are considered. A comprehensive analysis of the case from Zigui county is shown in Sects. 4.1, 4.2, and 4.3, and a model validation is conducted in Sect. 4.4 with other cases.

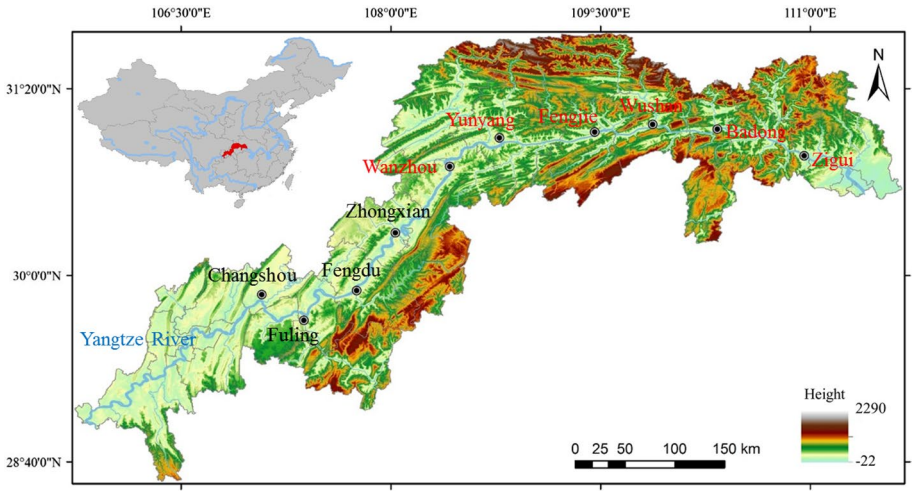

Fig. 2 Overview of study area

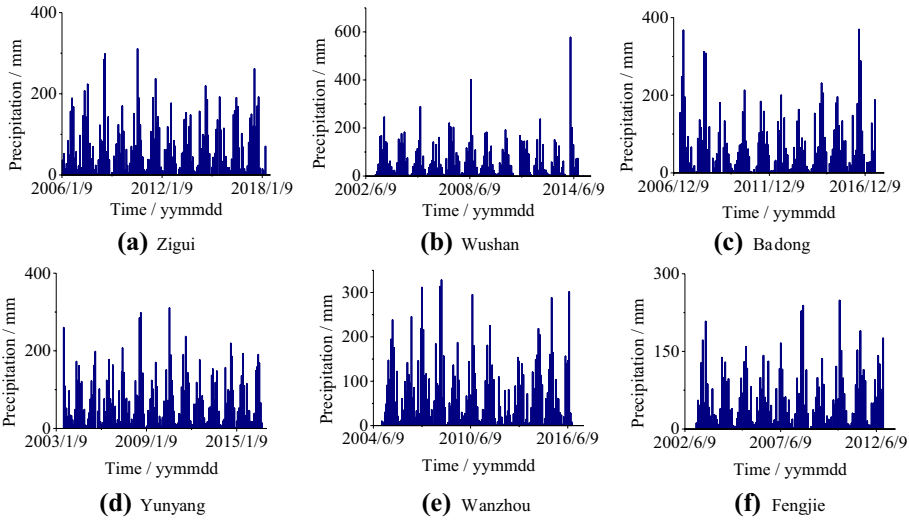

Fig. 3 Rainfall time histories considered in this study

Table 1 Statistical descriptors of rainfall datasets

County	Time period	Max (mm)	Min (mm)	Mean (mm)	Standard deviation (mm)
Zigui	Jan 2006–Mar 2018	298.6	1.5	83.41	68.57
Wushan	Jan 2003–Jan 2015	577.9	1.0	71.49	81.72
Badong	Apr 2007–Jun 2017	370.1	1.5	84.16	78.56
Yunyang	Jul 2003–Sep 2016	311.5	1.5	81.19	65.76
Wanzhou	Dec 2004–Sep 2016	328.5	1.5	88.77	76.33
Fengjie	Jan 2003–Oct 2012	249	1.0	69.44	55.14

4.1 Statistical analysis of rainfall dataset

The prediction dataset used in this research is computed from the rainfall dataset, named monthly ahead difference (MAD). Autocorrelation function (ACF) and partial autocorrelation function (PACF) (Liu et al. 2012) were obtained for original rainfall as well as for MAD to analyses the month-to-month correlation. ACF or PACF is a valuable quantitative method for assessment of the presence of autocorrelation at lags. To detect the autocorrelation in the original rainfall and MAD, the time series tests have been carried out by considering lags 20 months (for more than 1 year). In this paper, the Ljung–Box test is applied to measure the statistical significance of its autocorrelation. And the significance level is set as 0.05. The computing results are shown in Fig. 4.

Figure 4 illustrates the ACF and PACF of rainfall and MAD in terms of autoregression for 20 months of lag. The AFC of MAD has more value than ACF of rainfall (see Fig. 3a, c), and the PAFC of MAD convergence into the interval of twice the standard deviation at the lag of 7th month, and MAD is larger than the PAFC of rainfall at the lag of 8th month.

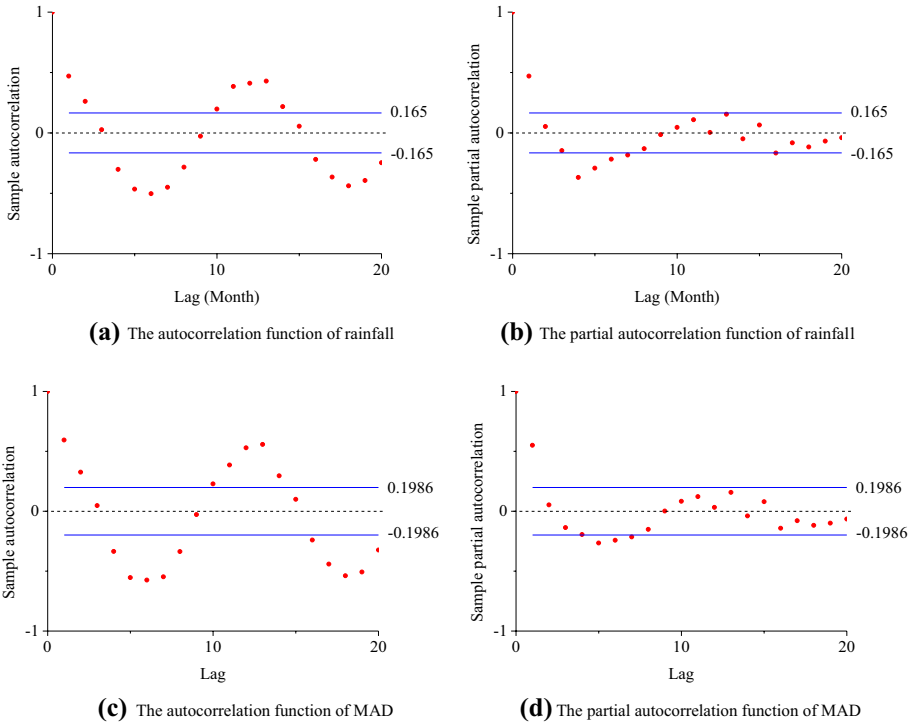

Fig. 4 AFC and PACF of rainfall and monthly ahead difference (MAD)

In addition, the ACF value of MAD at 12th month lag is 0.51, which is larger than 0.41 from rainfall at the same lag. As a consequence, better performance indices the MAD dataset has stronger periodicity than original rainfall dataset, and the MAD based prediction can acquire more approaching results.

4.2 OP-ELM with four activation functions

The proposed OP-ELM is applied in modeling and predicting the future rainfall. The 12-month period is selected as the time window considering the annual seasonal patterns of the dataset in the temporal domain. Based on the ACFs of MAD, all the positive lags (e.g., $t-1$, $t-2$, $t-3$, $t-10$, and $t-11$) are selected as inputs to predict the current value (see Fig. 4). In order to reach the optimal prediction performance, five different activation functions namely sigmoid, sine, hardlim, radial basis function (RBF), and tanh are selected in this research. In detail, the optimal number of hidden neurons within the single hidden layer are determined via tenfold cross validation according to Sect. 2.3. The MAD dataset has been spilt into training set (from January 2007 to September 2015) and testing set (from October 2015 to March 2018). By tenfold cross validation, the number of hidden neurons of four activation function-based OP-ELM are presented by Table 2.

Figure 5 shows the predated results of OP-ELM with four different activation functions, and Table 3 illustrates the testing performance of all five activation functions. From Table 3, the sigmoid activation function offers the lowest MAE, MAPE, RMSE, and MER,

Table 2 Optimal OP-ELM structure of various activation functions

Activation function	Number of hidden neurons
Sigmoid	15
Sine	21
Hardlim	39
RBF	18
Tanh	12

which is outstanding at comprehensive viewpoint. Since MAD is a dataset with periodicity, sine can produce good result. The RBF is an ideal choice for its robustness, but local large error affects its overall accuracy.

4.3 Comparison of proposed model with benchmarking algorithms

In this section, comparative analysis is conducted with OP-ELM utilizing sigmoid and three benchmarking machine learning algorithms including SVM, ANN, and classical ELM using sigmoid. Specifically, the structure of machine learning algorithms is determined via tenfold validation. The parameters selected for prediction are presented by Table 4.

It can be observed from Fig. 6 and Table 5 that the algorithm OP-ELM using sigmoid outperform all other algorithms investigated in this paper. The scatter points from Fig. 5a are gathered near the line '$y=x$', and the scatters from Fig. 5c are most dispersed. In terms of MAE, algorithm SVM performs worst (34.99 mm) while OP-ELM using sigmoid performs the best (13.76 mm). MAPE for OP-ELM with sigmoid was 4.43%, whereas the value of MAPE for SVM is 11.25%. In terms of RMSE, OP-ELM with sigmoid performs best (17.96 mm) while SVM performs worst (47.32 mm). Most importantly, the MER for OP-ELM using sigmoid (16.22%) has the most advantages than any other.

4.4 Model validation

The approach proposed in this study was applied to other 5 locations within the study area consigning 6 to 8-year training spans; while validation was performed over 15–36 months, as presented in Table 6. Model input variables, activations functions and number of hidden neurons vary among all assessed cases. However, general remarks can be made: the sigmoid activation function leads to best results; the number of hidden neurons ranged between 20 and 27 in most cases, up to 33 for the Wanzhou site; and, it was observed that the average in rainfall along the preceding 2 months, and the average of rainfall in the immediately previous 10 months were the most predominant parameters for rainfall prediction in all cases; hinting a lagging effect spanning the past two seasons. This is in agreement with what was found after computing the ACFs.

Average relative errors are less than 5% for the OP-ELM methodology developed in this study. This is a large improvement considering the output of previously proposed approaches in other studies, like SVMs, ANNs, and ELMs without optimization (Table 5). Therefore, there is a systematically lower divergence of OP-ELM predictions from measured rainfall (Fig. 5). All of the benchmarking frameworks lead to relative

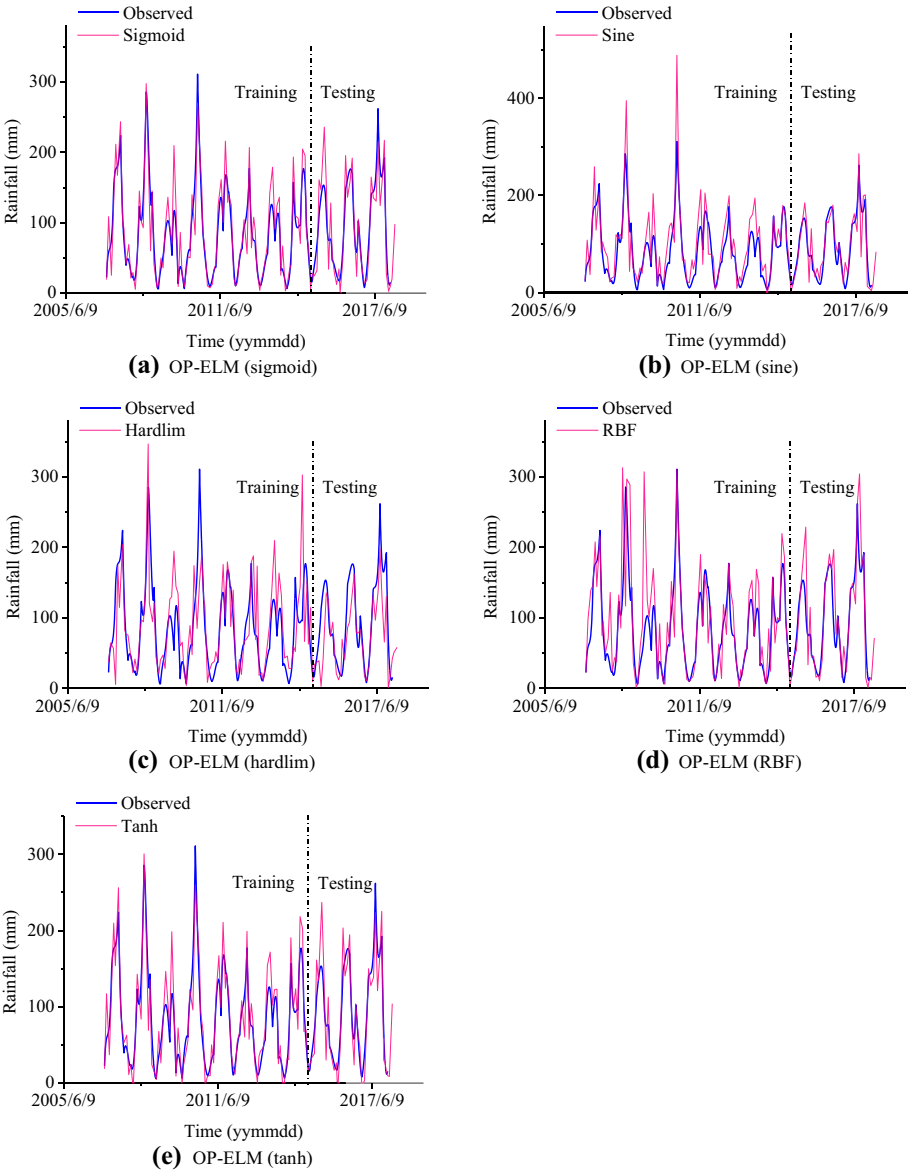

Fig. 5 Rainfall prediction performance with various activation functions using OP-ELM

errors twice or thrice. High-accuracy prediction of outliers remains elusive as maximum relative errors range between 16 and 27% for the OP-ELM implementation developed in this work (Table 6). However, this result is a large improvement compared to what is noticed for previous approaches. SVMs, ANSSs, and ELMs developed for benchmarking yield errors over 50% (Table 7). Consequently, the method proposed in this study greatly outperforms other commonly used tools for estimating rainfall.

Table 3 Testing performance of OP-ELM using various activation functions

Activation functions	MAE (mm)	MAPE (%)	RMSE (mm)	MER (%)
Sigmiod	13.76	4.43	17.96	16.22
Sine	23.21	7.46	33.72	37.19
Hardlim	33.64	10.82	50.82	46.54
RBF	18.73	6.02	49.89	74.54
Tanh	21.49	6.81	31.21	24.57

Table 4 Parameters selected via tenfold validation

Algorithm	Parameter
SVM	Compacity factor: 10; gamma: 0.01
ANN	Hidden layer: 2; hidden neuron: 10
Classical ELM using sigmoid	Hidden neuron: 15

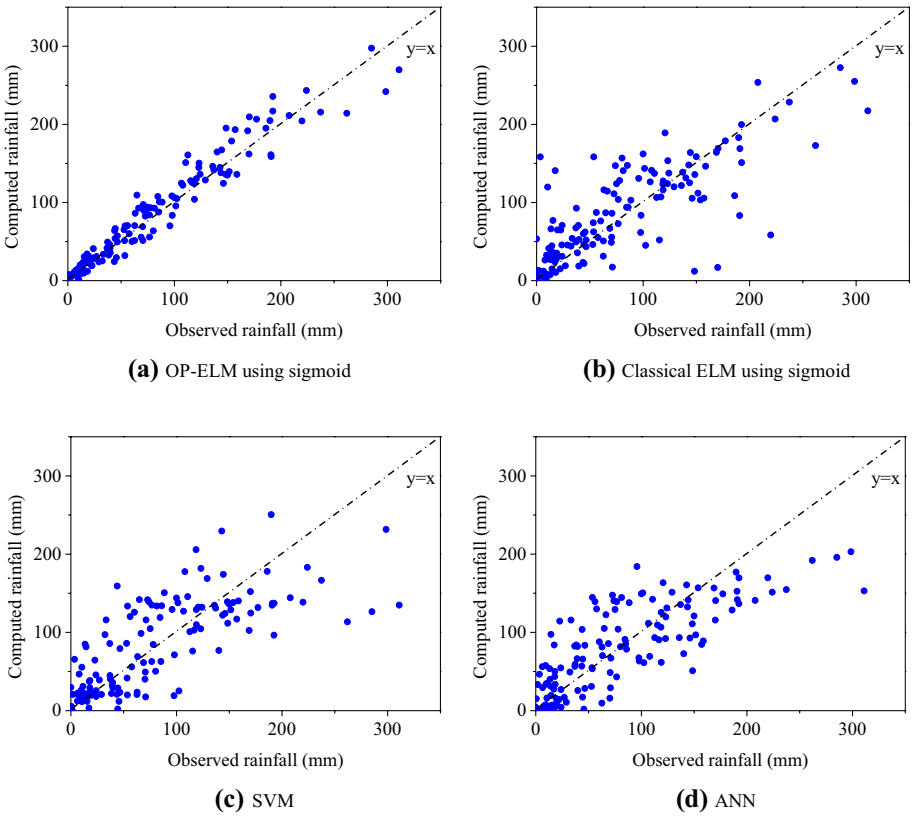

Fig. 6 Scatter plot for the computed result of rainfall

Table 5 Testing evaluation of all algorithms

Algorithms	MAE (mm)	MAPE (%)	RMSE (mm)	MER (%)
SVM	34.99	11.25	47.32	56.62
ANN	34.43	11.07	44.18	50.79
Classical ELM using sigmoid	30.47	9.81	44.95	51.91
OP-ELM using sigmoid	13.76	4.43	17.96	16.22

5 Discussion

The proposed model presents an advance in assessing the behavior or monthly rainfall, by overperforming by a factor or two and three previous approaches. Its results allow to start a discussion about how to better manage resources, land, and prepare for emergencies months ahead following continuous feedback. For example, development of awareness plans for increased flooding and landslides can be considered in detail a few months before while being revisited constantly; instead of just adopting strategies based solely on highly- variable seasonal trends. Likewise, dynamic management of reservoirs and upgraded planning of farming to dynamically adapt to drought is feasible; being one example, more accurate crop insurance pricing (Matsuda and Kurosaki 2019) Therefore, rainfall estimation should develop in tandem with initiatives for improving management of water resources and disaster prevention, allowing for valuable synergies between them.

As climate change becomes even more predominant and disrupts established seasonal rainfall trends (Ohba and Sugimoto 2019), models need to be adopted quickly in shorter time resolutions, as variability could increase severely. Thus, differentiation between short- and long-term forecasting will become less defined, probably by involving data from continuous remote sensing into time-history analyses using artificial intelligence. Model calibration and validation must be done quickly as conditions become more challenging while remaining practical. An interesting opportunity for this is deep learning (Yen et al. 2019), allowing for development adaptive neural networks that can be superimposed to deliver precise estimations for upcoming days using extremely detailed arrangements that can also feedback into lighter structures capable of making predictions for longer time-spans; while feeding on diverse types remote sensed data.

Finally, nonparametric approaches have been explored, where a prediction algorithm has been developed without considering a deep understanding of physical processes explaining rainfall. Uses of ANNs and other machine learning algorithms are the first step only, as they don't provide a rationale for results obtained. The next step is proposing models based on the physical and chemical processes that can shed light on the fundamentals that explain how they work. Getting only results without understanding why the outcomes are observed is not an ideal approach. Consequently, physical models should be revisited and adapted to take advantage of AI, being one interesting option is Genetic Programing (Adhikary et al. 2016), which can propose mathematical expressions better rooted on theoretical frameworks to estimate this phenomenon.

Table 6 Details of modeling parameters for rainfall prediction

	Training set	Test set	Inputs	Activation function	Number of hidden neurons
Wushan	Jan 2004–Apr 2012	May 2012–Jan 2015	$t-1, t-2, t-9, t-10, t-11$	Sigmoid	21
Badong	Apr 2008–Apr 2015	May 2015–Jun 2017	$t-1, t-2, t-3, t-4, t-10, t-11$	Sigmoid	27
Yunyang	Jul 2004–Apr 2013	May 2013–Sep 2016	$t-1, t-2, t-3, t-9, t-10, t-11$	Sigmoid	24
Wanzhou	Dec 2005–March 2014	Apr 2014–Sep 2016	$t-1, t-2, t-10, t-11$	Sigmoid	33
Fengjie	Jan 2004–Dec 2010	Jan 2011–Oct 2012	$t-1, t-2, t-3, t-9, t-10$	Sigmoid	27

Table 7 Performance of all algorithms for rainfall prediction in other cases

Performance measure	MAE (mm)	MAPE (%)	RMSE (mm)	MER (%)
Wushan				
SVM	45.49	14.63	65.61	63.71
ANN	46.76	14.98	54.47	60.27
Classical ELM	37.61	11.97	54.58	61.83
OP-ELM	16.88	6.15	21.28	26.81
Badong				
SVM	38.89	15.27	50.52	63.28
ANN	38.77	12.81	45.98	59.86
Classical ELM	35.71	11.19	47.04	57.01
OP-ELM	16.53	5.38	19.76	18.74
Yunyang				
SVM	38.02	12.15	56.15	64.96
ANN	37.48	11.99	44.71	58.42
Classical ELM	30.96	10.48	48.06	56.28
OP-ELM	14.88	4.84	18.69	17.16
Wanzhou				
SVM	40.85	13.29	54.48	66.11
ANN	39.96	12.79	50.87	59.85
Classical ELM	35.45	12.28	51.96	60.69
OP-ELM	15.84	5.15	21.45	19.03
Fengjie				
SVM	31.97	10.99	42.41	56.28
ANN	30.28	10.67	38.88	49.52
Classical ELM	28.13	9.86	39.56	49.80
OP-ELM	12.88	4.34	16.81	16.37

6 Conclusions

The complexity and nonlinear nature of rainfall generation phenomena makes its reliable estimation challenging. The most promising approach for achieving it is through the use of artificial intelligence. This study shows promising results by using an Extreme Learning Machine with Optimized Pruning that is capable of reaching less than 5% relative errors in rainfall prediction. Outliers are estimated with divergences of less than 20%. This performance is a large improvement from previous approaches considering vector support machines; neural networks and non-optimized extreme learning machines. All of these frameworks yield relative errors for both averages and peak values that are twice or thrice what is observed for the scheme developed in this work.

Models obtained showcase the complexity of the problem as 20–35 hidden neurons are required to yield reliable results. It is observed that current rainfall is highly dependent on both short-term (less than 3 months) and medium-term (up to 10 months before) preceding rainfall. Thus, indicating seasonability. In all test cases considered in this study, it was found that the sigmoid activation function leads to the best performance.

Acknowledgements We are immensely grateful to the suggestions and help from Andrés Felipe Alonso Rodriguez.

References

Adhikary S, Muttil N, Yilmaz A (2016) Genetic-programming based ordinary kriging for spatial interpolation of rainfall. J Hydrol Eng ASCE 14:04015062

Camps-Valls G, Martín-Guerrero JD, Rojo-Alvarez JL, Soria-Olivas E (2004) Fuzzy sigmoid kernel for support vector classifiers. Neurocomputing 62:501–506

Cawley GC, Talbot NL (2003) Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. Pattern Recognit 36(11):2585–2592

Chattopadhyay S, Chattopadhyay G (2008) Identification of the best hidden layer size for three-layered neural net in predicting monsoon rainfall in India. J Hydroinform 10(2):181–188

Chattopadhyay G, Midya SK, Chattopadhyay S (2020) Information theoretic study of the ground-level ozone and its precursors over Kolkata, India, during the summer monsoon. Iran J Sci Technol Trans Sci. https://doi.org/10.1007/s40995-020-01007-x

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

Esteves JT, de Souza Rolim G, Ferraudo AS (2019) Rainfall prediction methodology with binary multilayer perceptron neural networks. Clim Dyn 52(3–4):2319–2331

Grigorievskiy A, Miche Y, Ventelä AM, Séverin E, Lendasse A (2014) Long-term time series prediction using OP-ELM. Neural Netw 51:50–56

He Y, Kusiak A (2018) Performance assessment of wind turbines: data-derived quantitative metrics. IEEE Trans Sustain Energy 9(1):65–73

He Y, Kusiak A, Ouyang T, Teng W (2017) Data-driven modeling of truck engine exhaust valve failures: a case study. J Mech Sci Technol 31(6):2747–2757

He Y, Fei F, Wang W, Song X, Sun Z, Baek S (2018) Predicting manufactured shapes of a projection microstereolithography process via convolutional encoder-decoder networks. In: Proceedings of ASME 2018 international design engineering technical conferences & computers and information in engineering conference, Quebec City, Quebec, Canada

Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (1998) Support vector machines. IEEE Intell Syst Their Appl 13(4):18–28

Horikawa SI, Furuhashi T, Uchikawa Y (1992) On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm. IEEE Trans Neural Netw 3(5):801–806

Huang GB, Chen L, Siew CK (2006) Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans Neural Netw 17(4):879–892

Kasun LLC, Zhou H, Huang GB, Vong CM (2013) Representational learning with extreme learning machine for big data. IEEE Intell Syst 28(6):31–34

Kişi Ö (2009) Wavelet regression model as an alternative to neural networks for monthly streamflow forecasting. Hydrol Process Int J 23(25):3583–3597

Labat D, Ababou R, Mangin A (2000) Rainfall–runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses. J Hydrol 238(3–4):149–178

Lee T (2016) Wild bootstrap Ljung–Box test for cross correlations of multivariate time series. Econ Lett 147:59–62

Li H, Xu Q, He Y, Deng J (2018) Prediction of landslide displacement with an ensemble-based extreme learning machine and copula models. Landslides 15(10):2047–2059

Li H, Xu Q, He Y et al (2020) Modeling and predicting reservoir landslide displacement with deep belief network and EWMA control charts: a case study in Three Gorges Reservoir. Landslides 17(3):693–707

Liang NY, Huang GB, Saratchandran P, Sundararajan N (2006) A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Trans Neural Netw 17(6):1411–1423

Liu H, Tian HQ, Li YF (2012) Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. Appl Energy 98:415–424

Matsuda A, Kurosaki T (2019) Demand for temperature and rainfall index insurance in India. Agric Econ 50(3):353–366

McLeod A, Li W (1983) Diagnostic checking ARMA time series models using squared residual autocorrelations. J Time Ser Anal 4(4):269–273

Miche Y, Sorjamaa A, Bas P, Simula O, Jutten C, Lendasse A (2010) OP-ELM: optimally pruned extreme learning machine. IEEE Trans Neural Netw 21(1):158–162

Mohammed AA, Minhas R, Wu QJ, Sid-Ahmed MA (2011) Human face recognition based on multidimensional PCA and extreme learning machine. Pattern Recognit 44(10–11):2588–2597
Mugume I, Mesquita M, Bamutaze Y, Ntwali D, Basalirwa C, Waiswa D, Reuder J, Twinomuhangi R, Tumwine F, Ngailo TJ, Ogwang B (2018) Improving quantitative rainfall prediction using ensemble analogues in the tropics: case study of Uganda. Atmosphere 9(9):328
Ohba M, Sugimoto S (2019) Differences in climate change impacts between weather patterns: possible effects on spatial heterogeneous changes in future extreme rainfall. Clim Dyn 52:4177–4191
Ouyang T, Zha X, Qin L, Xiong Y, Xia T (2016) Wind power prediction method based on regime of switching kernel functions. J Wind Eng Ind Aerodyn 153:26–33
Ouyang T, He Y, Li H, Sun Z, Baek S (2019) Modeling and forecasting short-term power load with copula model and deep belief network. IEEE Trans Emerg Top Comput Intell 3(2):127–136
Pal S, Dutta S, Nasrin T et al (2020) Hurst exponent approach through rescaled range analysis to study the time series of summer monsoon rainfall over northeast India. Theor Appl Climatol 142:581–587
Pankratz A (1983) Forecasting with univariate Box–Jenkins method. Wiley, New York
Petković MD, Stanimirović PS (2011) Iterative method for computing the Moore–Penrose inverse based on Penrose equations. J Comput Appl Math 235(6):1604–1613
Ray SN, Chattopadhyay S (2021) Analyzing surface air temperature and rainfall in univariate framework, quantifying uncertainty through Shannon entropy and prediction through artificial neural network. Earth Sci Inf. https://doi.org/10.1007/s12145-020-00555-5
Ray S, Bose S, Chattopadhyay S (2021) A markov chain approach to the predictability of surface temperature over the northeastern part of India. Theor Appl Climatol 143:861–868. https://doi.org/10.1007/s00704-020-03458-z
Similä T, Tikka J (2005) Multiresponse sparse regression with application to multidimensional scaling. In: International conference on artificial neural networks. Springer, Berlin, pp 97–102
Sivapragasam C, Liong SY, Pasha MFK (2001) Rainfall and runoff forecasting with SSA-SVM approach. J Hydroinform 3(3):213–217
Suresh S, Babu RV, Kim HJ (2009) No-reference image quality assessment using modified extreme learning machine classifier. Appl Soft Comput 9(2):541–552
Tian X, Veldhuis M, Schleiss M et al (2019) Critical rainfall thresholds for urban pluvial flooding inferred from citizen observations. Sci Total Environ 689:258–268
Toth E, Brath A, Montanari A (2000) Comparison of short-term rainfall prediction models for real-time flood forecasting. J Hydrol 239(1–4):132–147
Wang Y, Cheng C, Xie Y et al (2017) Increasing trends in rainfall–runoff erosivity in the source region of the Three Rivers, 1961–2012. Sci Total Environ 592:639–648
Wu J, Long J, Liu M (2015) Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm. Neurocomputing 148:136–142
Wu Z, Zhou Y, Wang H, Jiang Z (2020) Depth prediction of urban flood under different rainfall return periods based on deep learning and data warehouse. Sci Total Environ 716:137077
Xu Q, Li H, He Y, Liu F, Peng D (2019) Comparison of data-driven models of loess landslide runout distance estimation. Bull Eng Geol Environ 78(2):1281–1294
Yen M, Lui D, Hsin Y, Lin C, Chen C (2019) Application of deep learning for the prediction of rainfall in southern Taiwan. Sci Rep 9:12774

Authors and Affiliations

Huajin Li[1] · Yusen He[2] · He Yang[1] · Yong Wei[3] · Songlin Li[4] · Jianqiang Xu[3]

Huajin Li
Huajinlee@yahoo.com

He Yang
yanghexq@163.com

Yong Wei
ceweiyong@hotmail.com

Songlin Li
lsl68@foxmail.com

Jianqiang Xu
xujq163@163.com

[1] School of Architecture and Civil Engineering, Chengdu University, Chengdu 610106, Sichuan, China

[2] Department of Mechanical and Industrial Engineering, The University of Iowa, Iowa City, IA 52242, USA

[3] State Key Laboratory of Geo-hazard Prevention and Geo-environment Protection, Chengdu University of Technology, Chengdu 610059, Sichuan, China

[4] Department of Natural Resources of Guizhou Province, Guiyang 550004, Guizhou, China