



Research article

Runoff forecasting model based on variational mode decomposition and artificial neural networks

Xin Jing, Jungang Luo*, Shangyao Zhang and Na Wei

State Key Laboratory of Eco-hydraulics in Northwest Arid Region, Xi'an University of Technology, Shaanxi 710048, China

* **Correspondence:** Email: jgluo@xaut.edu.cn; Tel: +8613991890360.

Abstract: Accurate runoff forecasting plays a vital role in water resource management. Therefore, various forecasting models have been proposed in the literature. Among them, the decomposition-based models have proved their superiority in runoff series forecasting. However, most of the models simulate each decomposition sub-signals separately without considering the potential correlation information. A neoteric hybrid runoff forecasting model based on variational mode decomposition (VMD), convolution neural networks (CNN), and long short-term memory (LSTM) called VMD-CNN-LSTM, is proposed to improve the runoff forecasting performance further. The two-dimensional matrix containing both the time delay and correlation information among sub-signals decomposing by VMD is firstly applied to the CNN. The feature of the input matrix is then extracted by CNN and delivered to LSTM with more potential information. The experiment performed on monthly runoff data investigated from Huaxian and Xianyang hydrological stations at Wei River, China, demonstrates the VMD-superiority of CNN-LSTM to the baseline models, and robustness and stability of the forecasting of the VMD-CNN-LSTM for different leading times.

Keywords: runoff forecasting; variational mode decomposition; convolution neural networks; long short-term memory

1. Introduction

Accurate runoff forecasting is an essential part of water resources management [1]. Traditional runoff forecasting is based on physical processes and is performed using some developed physical

models. However, substantial computational resources are necessary to describe hydrological processes with complex variabilities [2]. Furthermore, due to regional disparity, developing models in different places is a challenging task. Accordingly, it is not easy to apply the physical models in practical runoff forecasting [2,3].

The big data analysis method can solve the problems of the cumbersome data processing and basic parameter calibration in the traditional hydrology model, find a certain regularity through the relationship between the relevant factors, and provide great help to guide the practical work. Therefore, big data analysis-based data-driven models have been utilized in runoff forecasting. Data-driven models mainly include time series models and machine learning (ML) models [4]. However, since the time series models are mainly based on the linear assumption, these models cannot be employed for non-stationary and nonlinear runoff forecasting. Therefore, ML models with stable nonlinear fitting abilities have been introduced into runoff forecasting. Chu et al. [5] compared the monthly runoff forecasting performance of the support vector regression model (SVR) with other models based on the Tangnaihahai station, The obtained results demonstrated the SVR's reliable forecasting ability. Hu et al. [6] employed artificial neural network (ANN) and long short-term memory (LSTM) to simulate the rainfall-runoff process in the Fen River basin, and their result indicated that the latter outperforms the former due to its special units, while the two are superior to the physical model.

However, traditional ML models cannot adequately learn the variation feature from the highly nonstationary and complex time series data. Thus, signal processing methods have been adopted for data preprocessing to convert nonstationary time series data into several relatively stable sub-signals with a recognizable variation feature and provide more information of the series for forecasting [7]. Neeraj et al. [8] employed the singular spectrum analysis (SSA) to eliminate the noisy components in electric load series and combined this method with LSTM to obtain superior forecasting results. He et al. [9] employed variational mode decomposition (VMD) and gradient boosting regression (GBRT), to establish a hybrid model, which is demonstrated to be an effective method for runoff series. Tan et al. [10] applied the ensemble empirical mode decomposition (EEMD) to attain cleaner signals of the monthly runoff data in the Yangtze River Basin and make a precise monthly prediction result by the proposed decomposition-ensemble framework.

Some decomposition-based models first decompose the single runoff series into several sub-signals and employ the constructed models for each sub-signals to perform sub-signals predictions. Finally, the prediction results of all sub-signals are summed to obtain the runoff forecasting results [9–12]. However, only the time delay information is considered by this forecasting scheme while predicting each subsequence, and the correlation information between the decomposed sub-signals is ignored, which may also contribute to improving the runoff forecasting. Moreover, these decomposition-ensemble models should repeat and cumbersome training processes for all sub-signals, which is impracticable in runoff forecasting.

This paper proposes a reliable runoff forecasting model based on VMD, convolution neural networks (CNN), and LSTM, namely, the VMD-CNN-LSTM model, to solve the inability of standard decomposition-based models to utilize the correlation information between each decomposition sub-signals and unnecessary work for sub-signals training. Firstly, the original runoff series is decomposed by VMD, and the partial autocorrelation function (PACF) is utilized to obtain a reasonable time delay and generate the input matrix. Then, the convolution neural networks (CNN) are introduced in the runoff series forecasting, which is proved to be effective in image feature

recognition [13]. The convolution and pooling layers of the CNN can convolute the width and height of the input matrix to extract the time delay and correlation information features from both dimensions. Finally, the features extracted by CNN are employed as the LSTM input for runoff forecasting. In order to verify the performance of the proposed VMD-CNN-LSTM, the monthly runoff data of the Huaxian and Xianyang hydrological stations in Wei River Basin, China, are investigated to evaluate the proposed and the baseline models.

2. Study area and data

2.1. Study area

As shown in Figure 1, the Wei River basin lies between 33.68–37.39°N and 103.94–110.03°E, originates from Niaoshu Mountain in Gansu Province, China, and flows east into the Yellow River in Shaanxi Province, with a total length of 818 km and a drainage area of 135,000 km², accounting for 17.9% of the total area of the Yellow River basin, making it the largest tributary of the Yellow River [14]. Although the Wei River basin supports industrial, agricultural, and domestic water use of the Guanzhong Plain, the total water resources in the basin are rare, and resource-based water shortages lead to a significant contradiction between supply and demand of water resources. The accurate prediction of monthly runoff in this region provides a reference for water resources management in the basin.

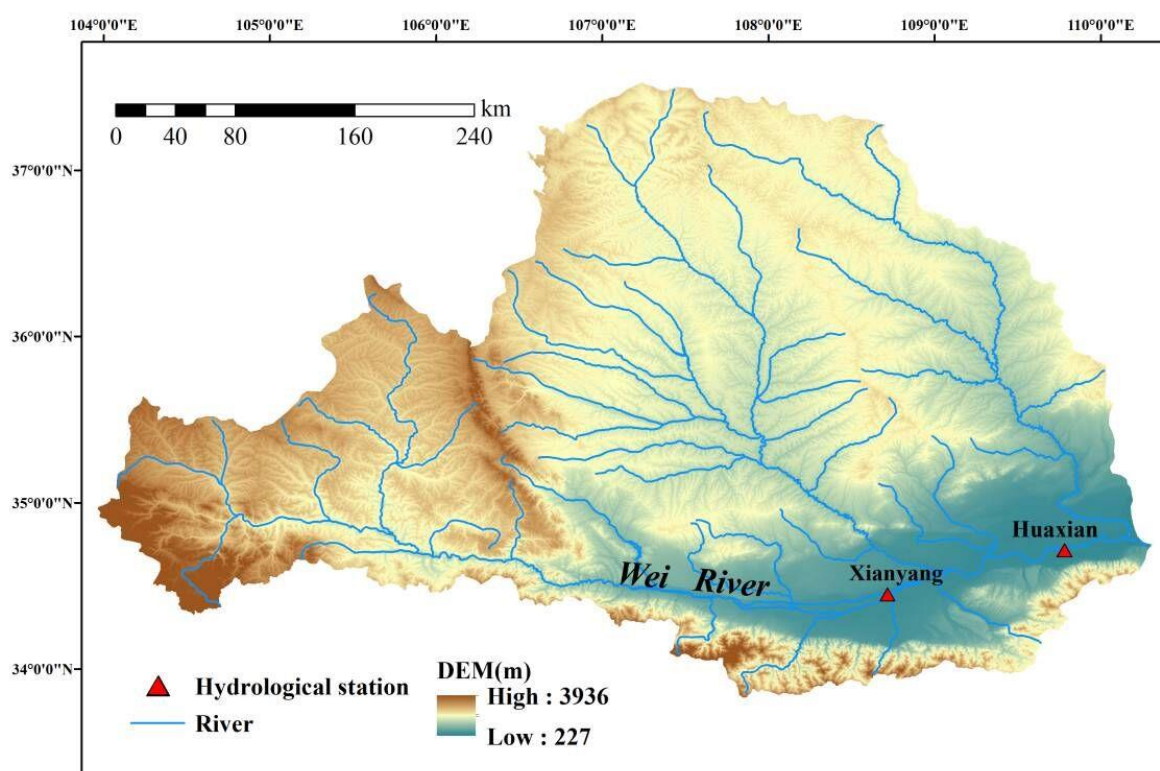


Figure 1. Location of the Huaxian and Xianyang Hydrological Stations in Jing River, China.

2.2. Data

Historical monthly runoff series from January 1953 to December 2018 in Huaxian and Xianyang hydrology stations in the Wei River basin are selected to evaluate the proposed forecasting model. The monthly runoff series data acquired from the Shaanxi Hydrological Information Center and the Water Resources Investigation Bureau are calculated through the daily monitored flow data. About 70% (January 1953~December 1998) and 30% (January 1999~December 2018) of the whole monthly runoff series are chosen as the training and validation-testing sets, respectively. Besides, half of the validation-testing set (January 1999~December 2008) is employed as the validation set, while the remaining (January 2009~December 2018) is chosen as the testing set to validate the forecasting model and test the optimal model, respectively [7].

3. Methods

3.1. Variational mode decomposition

The main idea of VMD is to construct and solve the variational problem. The constructed variational problem is expressed as follows [15]:

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s. t. } \sum k u_k(t) = f(t) \end{cases} \quad (1)$$

where k is the number of intrinsic mode functions (IMFs), and $f(t)$ is the input original signal, $\{u_k\}$ and $\{\omega_k\}$ are shorthand notations and center frequencies for the set of modes, respectively, t is the time, j is the imaginary unit, where j^2 is -1 , $*$ and δ are the convolution operator and Dirac delta function, respectively.

A Lagrangian multiplier (λ) and a quadratic penalty term (α) are utilized to convert the constrained optimization problem (1) into an unconstrained problem, where the augmented Lagrangian ℓ is expressed as follows [15]:

$$\begin{aligned} \ell(\{u_k\}, \{\omega_k\}, \lambda) &: \alpha \sum k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &+ \left\| f(t) - \sum k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum k u_k(t) \right\rangle \end{aligned} \quad (2)$$

In order to minimize the Lagrangian (2), the alternate direction method of multipliers is utilized to iteratively update the frequency-domain modes $u_k(\omega)$, the center frequencies ω_k , and the Lagrangian multiplier λ as follows [16]:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i < k} \hat{u}_i^{n+1}(\omega) - \sum_{i > k} \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (3)$$

$$\hat{\omega}_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \quad (4)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right) \quad (5)$$

where n is the iteration counter, τ is the iterative factor, while $\hat{u}_k^{n+1}(\omega)$, $\hat{f}(\omega)$ and $\hat{\lambda}^n(\omega)$ represent the Fourier transforms of $u_k^{n+1}(t)$, $f(t)$ and $\lambda^n(t)$, respectively.

3.2. Convolutional neural networks

Convolutional neural networks (CNN) can accelerate the training process, improve generalization performance, and extract the input feature with the identity of local connectivity by the weight sharing feature and convolutional computation [13]. As shown in Figure 2, a typical CNN mainly consists of a convolutional layer and a pooling layer.

The convolutional layer is employed for feature extraction and can minimize the training parameters and improve feature recognition via local perception and parameter sharing [17]. Each set of identical connections between the layers is often called a filter, which corresponds to the feature extraction and generates a feature mapping. Multiple filters employed in the convolutional layer can be expressed as follows [17]:

$$C = \sigma(X \otimes W + b) \quad (6)$$

where X is the input matrix, \otimes is the convolution operation, σ is the activation function (this paper selects the Relu function), W is the weight matrix, and b is the bias.

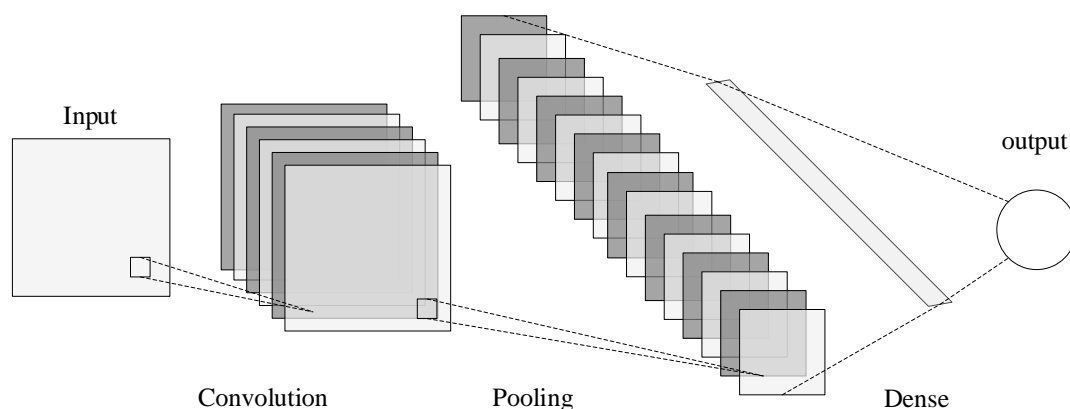


Figure 2. Diagram of a two-dimensional convolution neural network.

The pooling layer is employed for pooling the data after the convolutional operation. In this layer, the feature mapping obtained from the convolutional layer is divided into several small adjacent regions and reduced to a single value based on the average or maximum value. The pooling

layer can compress the data, remove unnecessary information, effectively enhance the network's generalization ability, and increase the computational speed. The maximum value pooling is chosen in this study.

3.3 Long short-term memory network

As a special kind of neural network, LSTM employs a gating mechanism for long-time series modeling and forecasting. The LSTM structure is shown in Figure 3 [18].

LSTM mainly controls the information sent to the cell state by the forget gate (f_t), input gate (i_t), and output gate (o_t). It outputs a vector (f_t) whose range is (0,1) to control the information from h_{t-1} and x_t flow within the cell. The detailed calculation formula is as follows [19].

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (7)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (8)$$

$$\tilde{C}_t = \tanh(W_{\tilde{C}} x_t + U_{\tilde{C}} h_{t-1} + b_{\tilde{C}}) \quad (9)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (10)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (11)$$

$$h_t = o_t * \tanh(C_t) \quad (12)$$

where W_f , W_i , $W_{\tilde{C}}$ and W_o are input weight matrices of f , i , \tilde{C}_t and o , respectively. b_f , b_i , $b_{\tilde{C}}$ and b_o are recurrent weight matrices of f , i , \tilde{C}_t and o , respectively. The last and the currently hidden cell states are indicated with the symbols h_{t-1} and h_t , respectively. The symbols C_{t-1} and C_t stand for the old and current cell information states, respectively. The symbol \tilde{C}_t is the candidate cell information through the \tanh layer. σ is the sigmoid function.

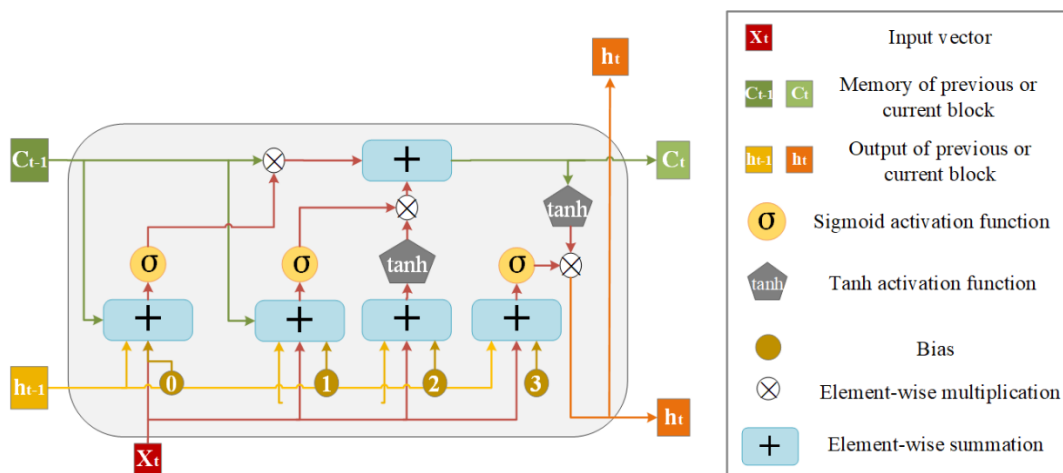


Figure 3. The cell state structure of LSTM.

3.4. The VMD-CNN-LSTM model

In order to solve the above problems, the VMD-CNN-LSTM model is established to predict the monthly runoff, and the design details can be summarized as follows (see Figure 4).

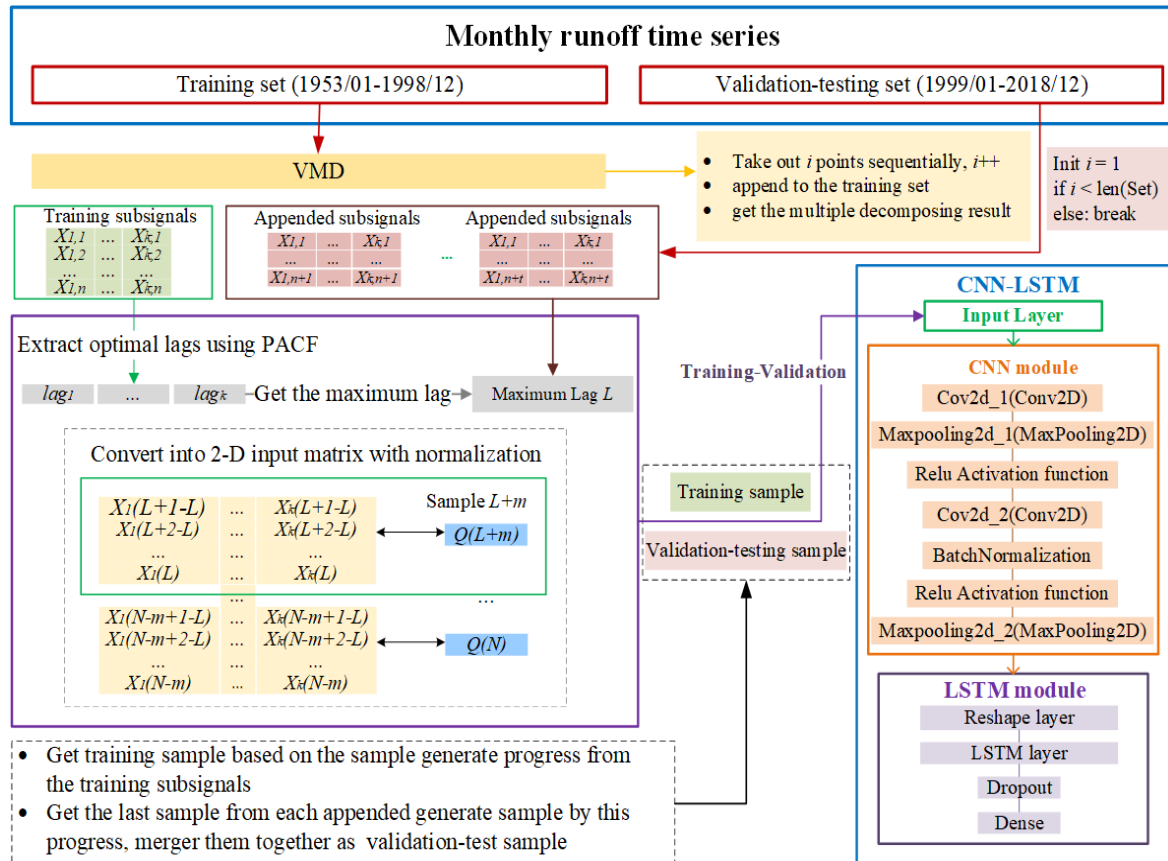


Figure 4. VMD-CNN-LSTM realization flow chart (The original set length is N ; training set length is n ; Max (lags) in every PACF of subsequences is L ; leading time is m ; the original runoff is Q ; $\text{len}(\text{Set})$ is the length of the validation-testing set).

Step 1: Collect monthly runoff time series data then divide it into the training and validation-testing sets (including 70% and 30% of the total series data, respectively).

Step 2: Extract k signal components of the intrinsic mode function from the training set to obtain the training sub-signals by VMD, where k is the decomposition number which depends on the phenomenon of the center frequency aliasing.

Step 3: Set an index i with an initial value of 1, and take out the first i values in the validation-testing set to obtain an appended set.

Step 4: Employ the VMD method to decompose the appended set to obtain and save the appended sub-signals.

Step 5: Add one to the value of i , repeat Step 3 and Step 4 until the value of i exceeds the length of the validation-testing set. Now, a set of appended decomposition sub-signals can be obtained for each point of the validation-testing set.

Step 6: Select the optimal time delay of all training sub-signals using the PACF method [20] to

obtain the maximum time delay L among the optimal time delays.

Step 7: Get the maximum and minimum values of each training subsequence, and employ the maximum-minimum (max-min) normalization method to normalize each subsequence of the training and validation-testing sets, respectively.

Step 8: Generate the input matrix ($L \times k$) according to the maximum delay L and decomposition number k , take the matrix as the model input, set the model output according to the leading time, then take them as a sample.

Step 9: Get a training sample through Step 8, and sequentially generate the sample for each set of the appended sub-signals, merge each of the last generated samples into the validation-testing sample.

Step 10: Divide the validation-testing sample into validation and testing samples, Concatenate training and validation samples to train and validate the model using cross-validation strategy [21] and test the model using testing sample.

3.5. Experiment setting

This paper investigates the monthly runoff data of the Huaxian and Xianyang hydrological stations in the Wei River Basin for the experiment. Then, the no-decomposition autoregressive integrated moving average (ARIMA) and LSTM models are chosen as the baseline models to compare with the VMD-CNN-LSTM model in the 1-month leading time. Moreover, the performance of the VMD-CNN-LSTM in different leading times is also verified in this experiment.

In these models, the optimal hyperparameters of the ARIMA, such as autoregressive and moving-average lags, are selected by grid search (GS) based on the Akaike Information Criterion, while the degree of difference is determined based on the augmented dickey-fuller (ADF) test [22]. The no-decomposition LSTM is trained by the monthly runoff series with the recent runoff data input determined by the PACF, while the Bayesian optimization (BO) algorithm [23] is employed to select the optimal hyperparameters, such as learning rate, the number of the hidden units, and the dropout rate. In the VMD-CNN-LSTM model, the BO algorithm is utilized to obtain the optimal values of the hyperparameters of the CNN and LSTM, such as the number of filters and the kernel size. The details of the mentioned hyperparameters of these models are shown in Table 1.

In Table 1, the max trial represents the maximum tuning times of each strategy; the formulas like [1: 1: 20] mean the beginning of the search scope, search step, and the end of the search scope, respectively, while these sample strategies employ uniform sampling. Especially logarithmic sampling is chosen as the search strategy of the learning rate. Besides, the hyperparameters of the LSTM in the VMD-CNN-LSTM are compatible with the baseline LSTM. Taking the monthly runoff series of Xianyang station as an example, a VMD-CNN-LSTM forecasting model is established based on the BO algorithm with 50 optimization times, and the optimization process is presented in Table 2.

As shown in Table 2, Index shows the order of each set of hyperparameters; while the Score is the mean square error (MSE), the smaller the score is, the better this set performs. The other columns describe the hyperparameters in the model. 1 hidden layer (NL) can get satisfactory predictions, and the number of layer units is a larger value, the pool size of Maxpooling2d_1 (KS_1) and the kernel size of the Cov2d_2 (KS_2) are consistent with the value of 4 and 2 in the top 10 sets of hyperparameters. However, other hyperparameters are not very sensitive to the model performance.

Table 1. The details of the hyperparameters mentioned in the forecasting models.

Forecasting model	Tuning strategy	Hyperparameter	Search scope
ARIMA	GS	Degree of difference(d)	Based on ADF test
		Autoregressive lags(p)	[1: 1: 20]
		Moving-average lags(q)	[1: 1: 20]
LSTM	BO (max trial = 50)	Number of hidden layer(NL)	1 or 2
		Number of layer units(NU)	[16: 16: 64]
		Drop rate(DR)	[0: 0.05: 0.5]
		Batch size(BS)	512
		Optimizer	Adam
		Learning rate(LR)	[10 ⁻⁴ : 10 ⁻¹]
		Activation	Relu
VMD-CNN-LSTM	BO (max trial = 50)	Number of filters(NF)	[8: 8: 64]
		Kernel size(KS)	[2, 3, 4]
		Pool size(PS)	[2, 3, 4]
		Padding strategy	Same padding
		LSTM hyperparameter	Consistent with LSTM

Table 2. Top 10 best sets of hyperparameters during the BO process in the VMD-CNN-LSTM model.

Index	NL	NU_0	DR_0	LR	NF	KS_1	KS_2	PS_1	PS_2	Score
40	1	48	0.3	0.0011	32	3	4	2	3	0.0193
23	1	64	0.45	0.0047	16	4	4	2	3	0.0195
37	1	48	0.5	0.0009	40	4	4	2	2	0.0199
9	1	48	0.45	0.0022	16	3	4	2	4	0.0203
24	1	64	0.5	0.0003	40	4	4	2	4	0.0218
46	1	64	0.2	0.0012	8	4	4	2	4	0.0235
28	1	48	0.3	0.0187	16	2	4	2	4	0.0240
27	1	32	0.5	0.0039	56	4	4	3	3	0.0247
44	1	48	0.45	0.0015	48	4	3	3	2	0.0252
29	1	48	0.15	0.0004	64	4	2	3	2	0.0295

3.6. Evaluation criteria

In this paper, four error analysis criteria are utilized to evaluate the forecasting capacity of the proposed model and the baseline models. These evaluation criteria are expressed in the following equations. The normalized root mean square error (NRMSE) [7] reflects the prediction error directly by calculating the normalized difference between the prediction and recorded values. Similar to mean absolute error (MAE), mean absolute percentage error (MAPE) [9] converts MAE into the percentage value to identify the model performance quickly. If the MAPE value is greater than 1, the model's performance is not satisfactory. Nash-Sutcliffe efficiency (NSE) [7] is widely utilized in time series forecasting; the closer the NSE is to the 1, the more accurate the model prediction is. Peak percent threshold statistics (PPTS) [26] can evaluate the forecasting precision of the peak

runoff by comparing the prediction error of the peak value.

$$NRMSE = \frac{\sqrt{\sum_{t=1}^N (x(t) - \hat{x}(t))^2 / N}}{\sum_{t=1}^N x(t) / N} \quad (13)$$

$$MAPE = \frac{100}{N} \sum_{t=1}^N \left| \frac{x(t) - \hat{x}(t)}{x(t)} \right| \quad (14)$$

$$NSE = 1 - \frac{\sum_{t=1}^N (x(t) - \hat{x}(t))^2}{\sum_{t=1}^N (x(t) - \bar{x}(t))^2} \quad (15)$$

$$PPTS(\gamma) = \frac{100}{\gamma} \frac{1}{N} \sum_{t=1}^G \left| \frac{x(t) - \hat{x}(t)}{x(t)} \cdot 100 \right| \quad (16)$$

where $x(t)$ and $\hat{x}(t)$ are the original time series and the predicted time series, respectively; $\bar{x}(t)$ is the average of the original time series; N is the length of the time series; γ is the percentage of the data selected from the top of the arranged data sequence, namely the threshold level; G is the number of the values higher the threshold level.

4. Results and discussion

4.1. Preparation of the VMD-CNN-LSTM input

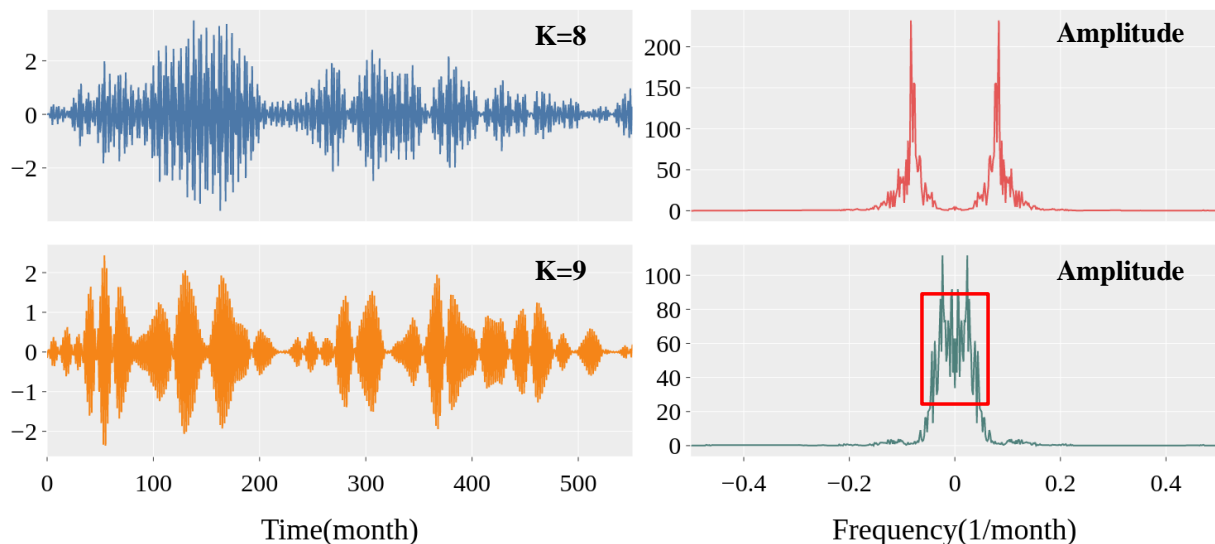


Figure 5. VMD mode components of monthly runoff with corresponding spectra in $K = 8$ and $K = 9$ at the Huaxian station.

In order to improve the precision of the prediction, the VMD method is proposed to extract the sub-signals of the time series, and the suitable number of IMFs k is determined by finding the center frequency aliasing phenomenon in the last component [15]. In this experiment, the k values from 2 to 12 are evaluated at the Huaxian station, and the center frequency aliasing phenomenon is found in the k value of 9, as shown in Figure 5. Thus, the 8 is selected as the suitable k . Now, the decomposition result of the time series is shown in Figure 6. Then, the max-min normalization [24] is employed to normalize the decomposition subsequences into $[-1, 1]$. Finally, the input matrix is constructed based on the maximum delay among the selected time delays. The input preparation of the Xianyang station is the same as the above.

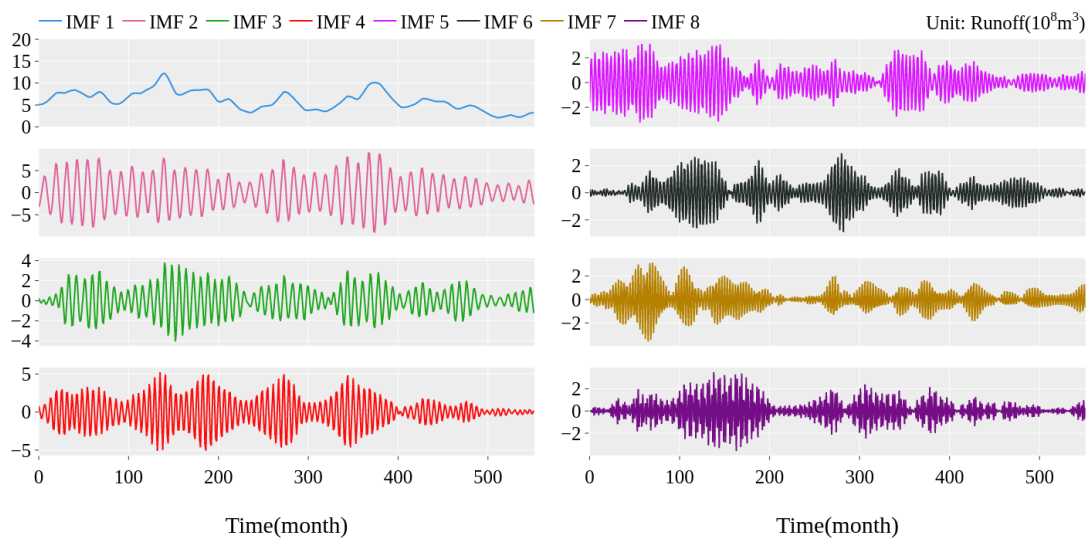


Figure 6. VMD decomposition result of the Huaxian station.

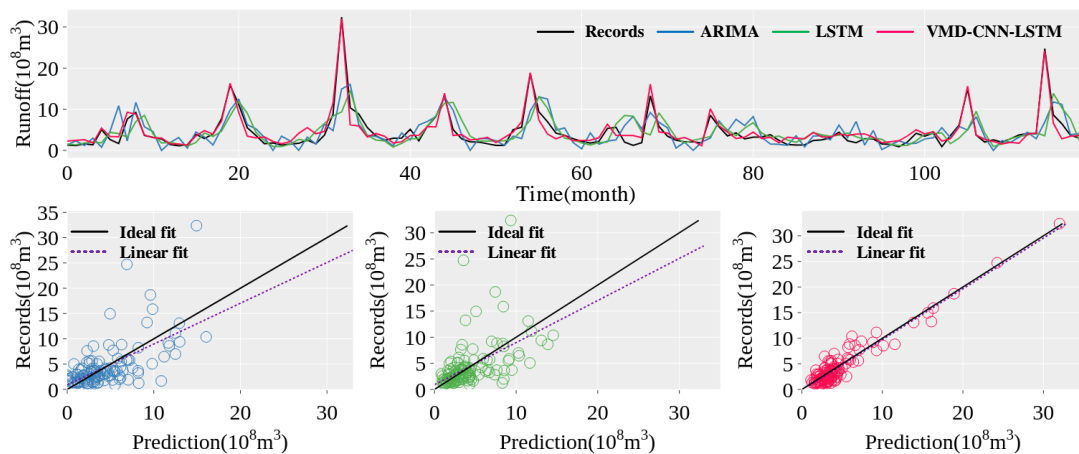


Figure 7. The predicted results and the scatters of different models at the Huaxian station.

4.2. Comparing the prediction results of different models

The comparisons of the 1-month prediction between ARIMA, LSTM, and VMD-CNN-LSTM at Huaxian and Xianyang stations are presented in Figures 7 and 8 and Table 3. According to the line

chart in Figures 7 and 8, the CNN-VMD-LSTM prediction is generally consistent with the change of the record runoff compared with the ARIMA and LSTM. However, the peak values predicted by ARIMA and LSTM models are generally lower than the record runoff, while there is a significant delay in these predictions, which means that these models cannot forecast the appearance time of the peak value. As shown in each scatter diagram of the models, the scatter of the VMD-CNN-LSTM predictions and records are mainly concentrated near the ideal fit, indicating a higher consistency compared with the ARIMA and LSTM models. Simultaneously, it can be concluded from each scatter diagram that the VMD-CNN-LSTM also has the smallest angle, reflecting the best prediction ability.

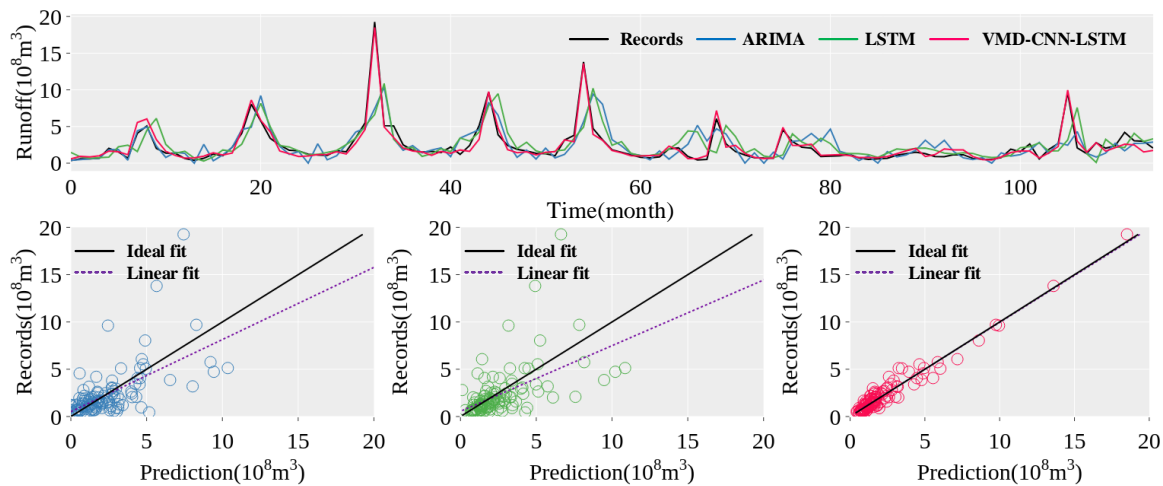


Figure 8. The predicted results and the scatters of different models at the Xianyang station.

Moreover, the quantitative evaluations of these models in Table 3 demonstrate that the VMD-CNN-LSTM model has a lower NRMSE, MAPE, and PPTS, while the NSE value is close to 1, which indicates that the model has a prediction ability superior to the other models. Besides, the PPTS of the VMD-CNN-LSTM model is far below the MAPE, which means the model can well fit the peak value in the runoff. However, due to the complexity of the model, it needs massive computing time and computing resources while selecting its hyperparameters.

Table 3. Comparison of prediction performances using different models.

Station	models	NRMSE	MAPE	NSE	PPTS	Time cost(s)
Huaxian	ARIMA	0.81	69.36	0.35	51.62	11.697
	LSTM	0.90	62.16	0.21	58.86	2632.43
	VMD-CNN-LSTM	0.20	24.69	0.96	3.78	6439.86
Xianyang	ARIMA	0.90	73.94	0.33	47.21	13.256
	LSTM	0.97	71.99	0.22	56.91	2872.16
	VMD-CNN-LSTM	0.24	25.04	0.95	6.04	7821.28

In summary, the prediction ability can be ranked from high to low as VMD-CNN-LSTM > ARIMA ≈ LSTM. Since the no-decomposition model can only employ the original data, it is difficult to capture the nonlinear characteristics. Accordingly, the ARIMA and LSTM models cannot

give good prediction results. The proposed VMD-CNN-LSTM model can extract the information in the original data via VMD and CNN and apply it into the LSTM to obtain a more precise prediction result for both the Huaxian and Xianyang stations compared with the no-decomposition model.

4.3. Comparing the prediction results for different leading times

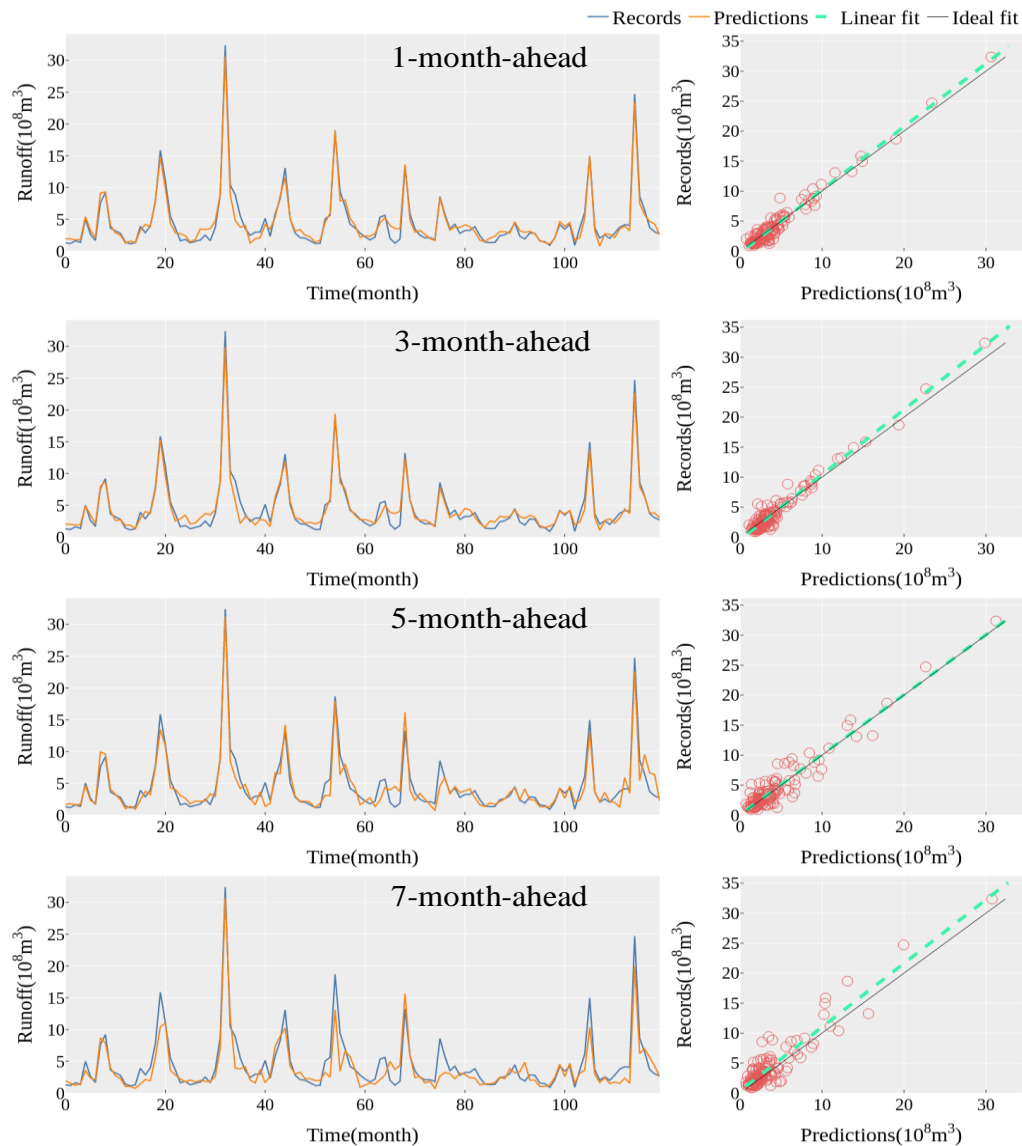


Figure 9. The predicted results and scatters of different leading times.

We set 1, 3, 5 and 7 month-ahead runoff forecastings at the Huaxian station to evaluate the prediction ability of the VMD-CNN-LSTM model in various leading times. Figure 9 shows the prediction line chart and scatter diagram in different leading times, and Table 4 shows the corresponding quantitative evaluations. As shown in the line chart in Figure 9, the predictions can vary consistently with the records in different leading times. However, in the 7-month-ahead runoff forecast, the model underestimates the record values in the peak runoff and overestimates them in the valley runoff. As shown in the scatter chart in Figure 9, the scatter of the predictions and records

gradually deviates from the ideal fit, while the linear fits in the 3 and 7 month-ahead predictions are above the ideal fit, indicating that the model generally underestimates the records in these leading times. In Table 4, the NSEs of 1, 3 and 5 month-ahead predictions are all over 0.9, demonstrating a good prediction ability. However, in the 7-month-ahead, the precision of the predictions is significantly decreased. According to the PPTS and MAPE, the PPTS and MAPE increase with the increase of the leading time, while the PPTS gradually approaches the MAPE, which indicates that the prediction error in the 7-month-ahead is mainly in the peak runoff value.

Table 4. Comparison of prediction performances in different leading times.

Leading time	NRMSE	MAPE	NSE	PPTS
1-month-ahead	0.20	24.69	0.96	3.783
3-month-ahead	0.22	27.75	0.95	6.262
5-month-ahead	0.31	30.30	0.91	10.94
7-month-ahead	0.41	32.91	0.83	22.91

In conclusion, the VMD-CNN-LSTM model provides more accurate predictions in different leading times. However, the performance of the 7-month-ahead prediction has decreased, which is mainly reflected in the peak runoff value prediction. Moreover, it can be concluded that since the correlation between the predicted and target values is decreasing, the model cannot obtain enough information to give a precise prediction.

5. Conclusions

The VMD-CNN-LSTM model employs the correlation information among the decomposed sub-signals for forecasting to improve monthly runoff forecasting performance and accelerate the training process. In this model, VMD decomposes the single time series and provides more information for forecasting, which makes the model different from the baseline model. CNN accelerates the training process and utilizes the potential correlation information by extracting the features from the sub-signals, while LSTM is employed to attain precise forecasting with adequate input information by VMD and CNN. The monthly runoff data of the Huaxian and Xianyang hydrological stations in the Wei River Basin are employed to evaluate the proposed model.

Compared with the ARIMA and LSTM in the 1-month leading time, the VMD-CNN-LSTM model provides the best prediction accuracy, which can reflect the general tendency and the peak value of the monthly runoff and help to support the flood emergency management.

Comparing the VMD-CNN-LSTM model predictions in different leading times indicates that the VMD-CNN-LSTM model provides an acceptable prediction precision. However, due to the decreasing correlation between the predictors and targets, its performance decreases in the 7-month-ahead, underestimating the peak runoff values.

In summary, the VMD-CNN-LSTM model is superior to the baseline models in prediction tasks considering the feature boost by the VMD method and the feature extraction by CNN. However, increasing the leading time leads to inevitable missing information in the time series, which may degrade the prediction performance. Moreover, CNN increases the model's training time cost while the LSTM input features are significant. As future work, the attention layer can be located before the LSTM layer, which is widely utilized in image recognition and natural language processing.

Furthermore, the training time should be reduced.

Acknowledgments

This work was supported by Natural Science Basic Research Program of Shaanxi Province (Grant No. 2019JLZ-15, 2019JLZ-16, 2017JQ5076), Special Scientific Research Program of Shaanxi Provincial Education Department (Grant No. 17JK0558), and Science and Technology Program of Shaanxi Province (Grant No. 2018slkj-4, 2019slkj-13, 2020slkj-16). The authors thank the editor for their comments and suggestions.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Y. Huang, L. Yang, S. Liu, G. Wang, Multi-step wind speed forecasting based on ensemble empirical mode decomposition, long short term memory network and error correction strategy, *Energies*, **12** (2019), 18–22. doi: 10.3390/en12101822.
2. G. K. Devia, B. P. Ganasri, G. S. Dwarakish, A review on hydrological models, *Aquat. Proc.*, **4** (2015), 1001–1007. doi: 10.1016/j.aqpro.2015.02.126.
3. J. W. Kirchner, Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, **42** (2006). doi: 10.1029/2005WR004362.
4. B. Shirmohammadi, M. Vafakhah, V. Moosavi, A. Moghaddamnia, Application of several data-driven techniques for predicting groundwater level, *Water Resour. Manag.*, **27** (2013), 419–432. doi: 10.1007/s11269-012-0194-y.
5. H. Chu, J. Wei, T. Li, K. Jia, Application of support vector regression for mid-and long-term runoff forecasting in “Yellow river headwater” region, *Proc. Eng.*, **154** (2016), 1251–1257. doi: 10.1016/j.proeng.2016.07.452.
6. C. Hu, Q. Wu, H. Li, S. Jian, N. Li, Z. Lou, Deep learning with a long short-term memory networks approach for rainfall-runoff simulation, *Water*, **10** (2018), 1543. doi: 10.3390/w10111543.
7. G. Zuo, J. Luo, N. Wang, Y. Lian, X. He. Two-stage variational mode decomposition and support vector regression for streamflow forecasting, *Hydrol. Earth Syst. Sci.*, **24** (2020), 5491–5518. doi: 10.5194/hess-24-5491-2020.
8. N. Neeraj, J. Mathew, M. Agarwal, R. K. Behera, Long short-term memory-singular spectrum analysis-based model for electric load forecasting, *Electr. Eng.*, **103** (2021), 1067–1082. doi: 10.1007/s00202-020-01135-y.
9. X. He, J. Luo, P. Li, G. Zuo, J. Xie, A hybrid model based on variational mode decomposition and gradient boosting regression tree for monthly runoff forecasting, *Water Resour. Manag.*, **34** (2020), 865–884. doi: 10.1007/s11269-020-02483-x.

10. Q. F. Tan, X. H. Lei, X. Wang, H. Wang, X. Wen, Y. Ji, et al., An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach, *J. Hydrol.*, **567** (2018), 767–80. doi: 10.1016/j.jhydrol.2018.01.015.
11. H. Niu, K. Xu, W. Wang, A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network, *Appl. Intell.*, **50** (2020), 4296–4309. doi: 10.1007/s10489-020-01814-0.
12. H. Liu, X. Mi, Y. Li, Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM, *Energy Convers. Manag.*, **159** (2018), 54–64. doi: 10.1016/j.enconman.2018.01.010.
13. Y. Chen, H. Jiang, C. Li, X. Jia, P. Ghamisi, Deep feature extraction and classification of hyperspectral images based on convolutional neural networks, *IEEE Trans. Geosci. Remote Sens.*, **54** (2016), 6232–6251. doi: 10.1109/TGRS.2016.2584107.
14. J. Chang, Y. Wang, E. Istanbuluoglu, T. Bai, Q. Huang, D. Yang, et al., Impact of climate change and human activities on runoff in the Weihe River Basin, China, *Quat. Int.*, **380** (2015), 169–179. doi: 10.1016/j.quaint.2014.03.048.
15. K. Dragomiretskiy, D. Zosso, Variational mode decomposition, *IEEE Trans. Signal Process.*, **62** (2013), 531–544. doi: 10.1109/TSP.2013.2288675.
16. W. Liu, S. Cao, Y. Chen, Applications of variational mode decomposition in seismic time-frequency analysis, *Geophysics*, **81** (2016), 365–378. doi: 10.1190/geo2015-0489.1.
17. G. Zhao, G. Liu, L. Fang, B. Tu, P. Ghamisi, Multiple convolutional layers fusion framework for hyperspectral image classification, *Neurocomputing*, **339** (2019), 149–160. doi: 10.1016/j.neucom.2019.02.019.
18. F. Kratzert, D. Klotz, C. Brenner, K. Schulz, M. Herrnegger, Rainfall-runoff modelling using long short-term memory (LSTM) networks, *Hydrol. Earth Syst. Sci.*, **22** (2018), 6005–6022. doi: 10.5194/hess-22-6005-2018.
19. Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, J. Liu, LSTM network: a deep learning approach for short-term traffic forecast, *IET Intell. Transp. Syst.*, **11** (2017), 68–75. doi: 10.1049/iet-its.2016.0208.
20. A. Berryman, P. Turchin, Identifying the density-dependent structure underlying ecological time series, *Oikos*, **92** (2001), 265–270. doi: 0.1034/j.1600-0706.2001.920208.x.
21. S. Hwang, Cross-validation of short-term productivity forecasting methodologies, *J. Constr. Eng. Manag.*, **136** (2010), 1037–46. doi: 10.1061/(ASCE)CO.1943-7862.0000230.
22. M. S. Voss, X. Feng, ARMA model selection using particle swarm optimization and AIC criteria. *IFAC Proc. Vol.*, **35** (2002), 349–354. doi: 10.3182/20020721-6-ES-1901.00469.
23. J. Wu, X. Y. Chen, H. Zhang, L. D. Xiong, H. Lei, S. H. Deng, Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.*, **17** (2019), 26–40. doi: 10.11989/JEST.1674-862X.80904120.
24. S. Yang, J. Liu, Time-series forecasting based on high-order fuzzy cognitive maps and wavelet transform. *IEEE Trans. Fuzzy Syst.*, **26** (2018), 3391–402. doi: 10.1109/TFUZZ.2018.2831640.

