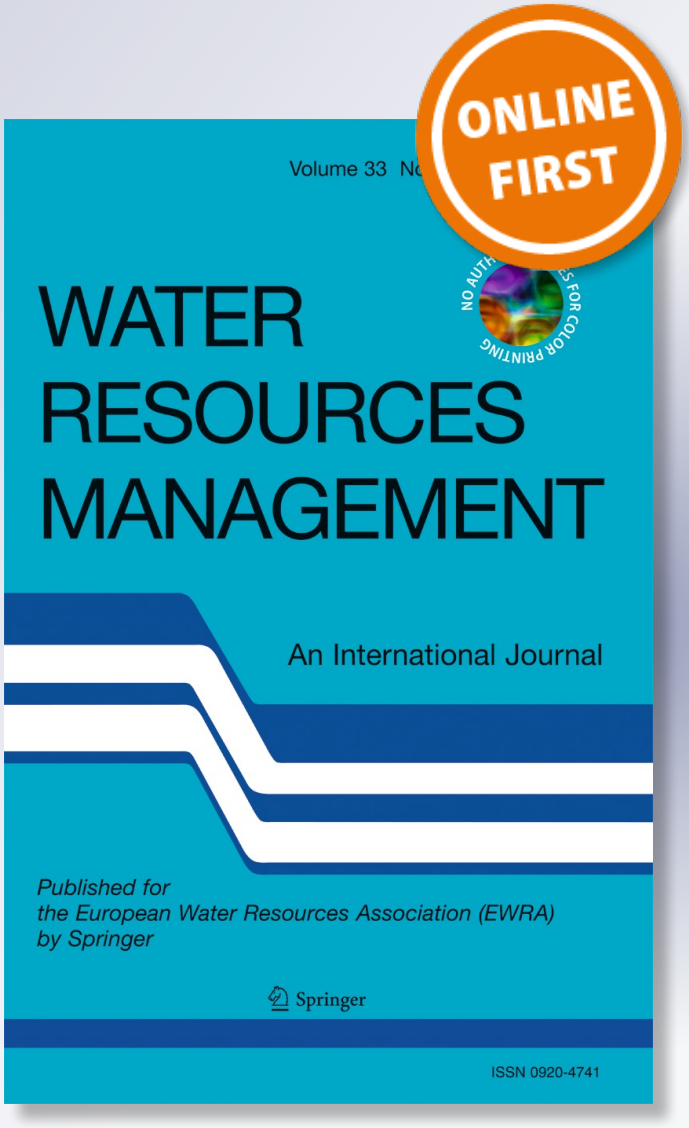
Daily Runoff Forecasting Using a Hybrid Model Based on Variational Mode Decomposition and Deep Neural Networks

Xinxin He, Jungang Luo, Ganggang Zuo & Jiancang Xie

Water Resources Management
An International Journal - Published for the European Water Resources Association (EWRA)

ISSN 0920-4741

Water Resour Manage
DOI 10.1007/s11269-019-2183-x



Springer

Water Resources Management
https://doi.org/10.1007/s11269-019-2183-x

Daily Runoff Forecasting Using a Hybrid Model Based on Variational Mode Decomposition and Deep Neural Networks

Xinxin He[1] · Jungang Luo[1] · Ganggang Zuo[1] · Jiancang Xie[1]

Received: 31 July 2018 / Accepted: 1 January 2019/Published online: 10 January 2019

Abstract
Accurate and reliable runoff forecasting plays an increasingly important role in the optimal management of water resources. To improve the prediction accuracy, a hybrid model based on variational mode decomposition (VMD) and deep neural networks (DNN), referred to as VMD-DNN, is proposed to perform daily runoff forecasting. First, VMD is applied to decompose the original runoff series into multiple intrinsic mode functions (IMFs), each with a relatively local frequency range. Second, predicted models of decomposed IMFs are established by learning the deep feature values of the DNN. Finally, the ensemble forecasting result is formulated by summing the prediction sub-results of the modelled IMFs. The proposed model is demonstrated using daily runoff series data from the Zhangjiashan Hydrological Station in Jing River, China. To fully illustrate the feasibility and superiority of this approach, the VMD-DNN hybrid model was compared with EMD-DNN, EEMD-DNN, and multi-scale feature extraction -based VMD-DNN, EMD-DNN and EEMD-DNN. The results reveal that the proposed hybrid VMD-DNN model produces the best performance based on the Nash-Sutcliffe efficiency (NSE = 0.95), root mean square error (RMSE = 9.92) and mean absolute error (MAE = 3.82) values. Thus the proposed hybrid VMD-DNN model is a promising new method for daily runoff forecasting.

Keywords Daily runoff forecasting · Hybrid model · Variational mode decomposition · Deep neural networks

✉ Jungang Luo
jgluo@xaut.edu.cn

Xinxin He
xinxinhe89@foxmail.com

Ganggang Zuo
zuoganggang@163.com

Jiancang Xie
jcxie@xaut.edu.cn

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

1 Introduction

Runoff forecasting, especially daily runoff forecasting is of great significance for water resource planning and management. Therefore, this topic has attracted the attention of many researchers, and several forecasting models have been developed over the past few decades.

Data-driven models, which have been widely used in runoff forecasting by construction of a "black-box", include two main categories: time-series models and artificial intelligence (AI) models. Time-series models (Box et al. 2015) such as the auto-regressive (AR), moving average (MA), auto-regressive moving average (ARMA), and auto-regressive integrated moving average (ARIMA) have supplied good forecasting accuracy for runoff time series (Toth et al. 2000), especially under the condition that the series are stationary or near-linear. However, these models encounter difficulty in capturing the nonlinear characteristics hidden in hydrological time series. Stationarity is a necessary condition for establishing a time-series model. In contrast, AI models, including artificial neural networks (ANNs) (Citakoglu et al. 2014), adaptive neural-fuzzy inference systems (ANFIS) (Moghaddamnia et al. 2009), genetic programming (GP) (Koza 1992), and support vector machine (SVM) (Lin et al. 2009), are able to address the non-stationarity of the runoff time series. Zhang et al. (2009) combined QPF into ANNs model for daily reservoir inflow prediction. Sattari et al. (2012) investigated the potential of time lag recurrent ANN for modelling of daily inflow into the Eleviyan Reservoir in Iran. Cheng et al. (2015) developed an ANN-QPSO model for daily runoff forecasting to enhance the generalization performance of the forecasting model. Nayak et al. (2004) presented the application of an adaptive NFIS (ANFIS) model used to forecast daily streamflow at the Baitarani River in India. Wang et al. (2009) conducted a comparative study of ARMA, ANN, ANFIS, GP and SVR for forecasting the monthly discharge time series, and the results indicated that ANFIS, GP, and SVM obtained the best performance during the training and validation phases.

With advances in research, the recent trend of data-driven runoff forecasting involves hybridizing the AI models with data pre-processing techniques for the purpose of further improving the forecasting accuracy. Various decomposition technique-based hybrid models were developed and widely used to address the nonlinear and nonstationarity present in runoff series. Three mainstream algorithms, i.e., wavelet transform (WT) (Dabuechies 1990), empirical mode decomposition (EMD) (Huang et al. 1998), and ensemble empirical mode decomposition (EEMD) (Wu and Huang 2009), are combined with various forecasting methods to form a hybrid forecasting model. Kisi and Cimen (2012) proposed a WA-SVM model for daily precipitation forecast in Turkey. Bai et al. (2016) introduced a multiscale deep feature learning method with hybrid models to address daily reservoir inflow forecasting in the TGD. Wang et al. (2018) developed a hybrid EMD/EEMD-ARIMA model for forecasting the daily streamflow time series for six years at Tangnaihai Station. Huang et al. (2014) proposed a modified EMD-SVM model for monthly streamflow prediction in the Wei River Basin. Wang et al. (2013) proposed an EEMD-based particle swarm optimization (PSO) and SVM for rainfall-runoff forecasting. Di et al. (2014) designed an EEMD-based RBF neural network for hydrologic time series forecasting. All of these applications indicate that data pre-processing techniques coupled with AI models outperform the conventional single models and the hybrid models without denoising or decomposition.

However, the complex nonlinearity, high irregularity, and multi-scale variability present in the daily runoff series are always challenging aspects in forecasting, and the models that work on the original runoff series cannot handle these situations appropriately. In general, (1) AI

models mostly belong to the "shallow" learning category in which the distinct information is represented insufficiently (Bai et al. 2016), e.g., ANNs are often subject to slow learning speed, convergence to local minima (Okkan and Serbes 2012), and over-fitting, and SVMs are usually sensitive to parameter selection (Di et al. 2014); (2) data pre-processing techniques suffer from the drawbacks with different aspects in actual signal decomposition, e.g., the effectiveness of WT depends heavily on the choice of wavelet basis function (Chellali et al. 2010), the application of EMD is limited by the lack of an exact mathematical model, the frequent appearance of mode mixing, and sensitivity to both noise and sampling (Naik et al. 2018), and EEMD lacks a strict mathematical theory (Liu et al. 2018). To address the aforementioned challenges, deep learning techniques, which can produce better forecasting performance depending on "deeper" performances, are introduced to effectively overcome these drawbacks (Li et al. 2016). The deep neural network (DNN), one of the deep learning methods, has been proposed (Hinton et al. 2006; Bengio et al. 2007) and can capture highly abstracted features from the original runoff series and address the limitations of shallow-structured architectures. In recent years, DNN has facilitated great progress in analysis in fields such as speech recognition (Hinton et al. 2012), computer vision (Krizhevsky et al. 2017), text processing (Ciresan et al. 2010), and natural language processing (Collobert and Weston 2008). In addition, several new data pre-processing approaches have been proposed. Specifically, variational mode decomposition (VMD) has attracted much attention due to its sound mathematical framework, and it showed better noise robustness and more precise component separation (Ali et al. 2018). As a new multi-resolution technique, VMD has recently been applied in many research fields, such as biomedical signal processing (Lahmiri and Boukadoum 2015), international stock market analysis (Lahmiri 2015), and wind speed prediction (Liu et al. 2018).

In this paper, a novel hybrid VMD-DNN model is proposed to improve the accuracy of daily runoff forecasting. VMD is used to decompose the original daily runoff series into multiple intrinsic narrow-band components to facilitate their forecasting. For each sub-sequence, DNN is applied as a forecasting tool to establish the prediction models. The effectiveness of the proposed hybrid model is evaluated using daily runoff data from the Zhangjiashan Hydrological Station in Jing River, China, and is compared with five other models.

2 Methodology

2.1 Variational Mode Decomposition

Variational mode decomposition, a new non-recursive signal processing technique proposed by Dragomiretskiy and Zosso (Dragomiretskiy and Zosso 2014), is used to adaptively decompose a complicated signal into a number of band-limited intrinsic mode functions (BLIMFs).

An original signal$f(t)(t = 1, 2, ..., n)$ can be decomposed into a set of IMFs that are band limited. An IMF$u_k(t)$is considered as an amplitude-modulated-frequency-modulated (AM-FM) signal, which is expressed as follows:

$$u_k(t) = A_k(t)\cos(\phi_k(t)) \tag{1}$$

where$A_k(t)$is the instantaneous amplitude,$\phi_k(t)$denotes the instantaneous phase, and the instantaneous frequency$\omega_k(t) = \phi_k^{'}(t) = \partial\phi_k(t)/\partial t$is nonnegative and varies much slower than the phase$\phi_k(t)$.

To acquire the bandwidth of each mode$u_k(t)$, the constrained variational optimization problem can be solved by the following:

$$\left\{ \min_{u_k, w_k} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} s.t \sum_{k=1}^{K} u_k = f(t) \right. \tag{2}$$

where$w_k$is the centre frequency of the corresponding$k$th mode of the signal, $t$is the time script, $k$is the number of modes, and$\delta(t)$is the Dirac function. In addition, the expression $\left(\delta(t) + \frac{j}{\pi t}\right) \otimes u_k(t)$ is the Hilbert transform of$u_k(t)$, which can transform$u_k(t)$ into the analytical signal to form a one-side frequency spectrum with only positive frequencies. In this manner, using the index terms$e^{-jw_k t}$, the spectrum of each mode can be shifted to a baseband.

Two parameters, the quadratic penalty factor$\alpha$and Lagrange multipliersλ, are both considered to transform the above optimization problem into an unconstrained optimization problem. The expression of the augmented Lagrange function is given by the following:

$$\begin{aligned} L(u_k, \omega_k, \lambda) = \alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-jw_k t} \right\|_2^2 + \\ \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 + \left\langle \lambda(t) f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \end{aligned} \tag{3}$$

where$\left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2$is a quadratic penalty term to accelerate the convergence velocity.

According to the optimization technique known as alternate direction method of multipliers (ADMM),$u_k$and$w_k$can be updated in two directions to complete the analysis of the VMD. Therefore, the updated equations are given as follows:

$$\hat{u}_k^{n+1} = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{4}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k(\omega) \right|^2 d\omega} \tag{5}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right) \tag{6}$$

where$n$is the number of iterations,$\tau$is the iterative factor, and$\hat{u}_k(\omega)$,$\hat{u}_k^{n+1}$,$\hat{f}(\omega)$, and $\hat{\lambda}(\omega)$represent the Fourier transform of$u_i(\omega)$,u_k^{n+1},$f(\omega)$, and$\lambda(\omega)$, respectively.

In general, the VMD algorithm can be summarized as shown in Algorithm 1.

Algorithm 1 Process of VMD

Initialize $\hat{u}_k^1$, $\hat{\omega}_k^1$, $\hat{\lambda}^1$, and $n = 0$

repeat

$n = n + 1$

for $k = 1 : K$ **do**

Update $\hat{u}_k(\omega)$ for all $\omega \geq 0$ using Eq.(4)

Update ω_k using Eq.(5)

end for

Update $\hat{\lambda}(\omega)$ for all $\omega \geq 0$ using Eq.(6)

until convergence $\sum_{k=1}^{K} \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2 \Big/ \left\| \hat{u}_k^n \right\|_2^2 < \varepsilon$

Obtain $u_k^{n+1}(t)$ by fast Fourier transform of $\hat{u}_k^{n+1}(\omega)$.

return $u_k^{n+1}(t)$.

2.2 Deep Neural Networks

The deep neural network is a standard multi-layer feed-forward network consisting of one visible input layer, multiple intermediate hidden layers, and an output layer. The structure of the DNN is shown in Fig. 1. It can be clearly observed that the hidden layer consists of a deep belief network (DBN), which is a stack of restricted Boltzmann machines (RBM) (Bengio et al. 2013). The DNN training stage can be divided into two phases: (a) unsupervised pre-training and (b) supervised fine-tuning.

2.2.1 Deep Belief Network for Pre-Training

The purpose of pre-training is to obtain a set of initialized weights that can be used in later fine-tuning. In this stage, a DBN is applied to initialize or pretreat the hidden layer of DNN to lead the model into a space that is closer to the optimum (Mohamed et al. 2012). In practical application, a Gaussian-Bernoulli RBM (GRBM) (Cho et al. 2011) was established by normalizing real data into [0,1] to train the RBMs. The model defines the probability distribution on$v$,$h$through an energy function$E$:

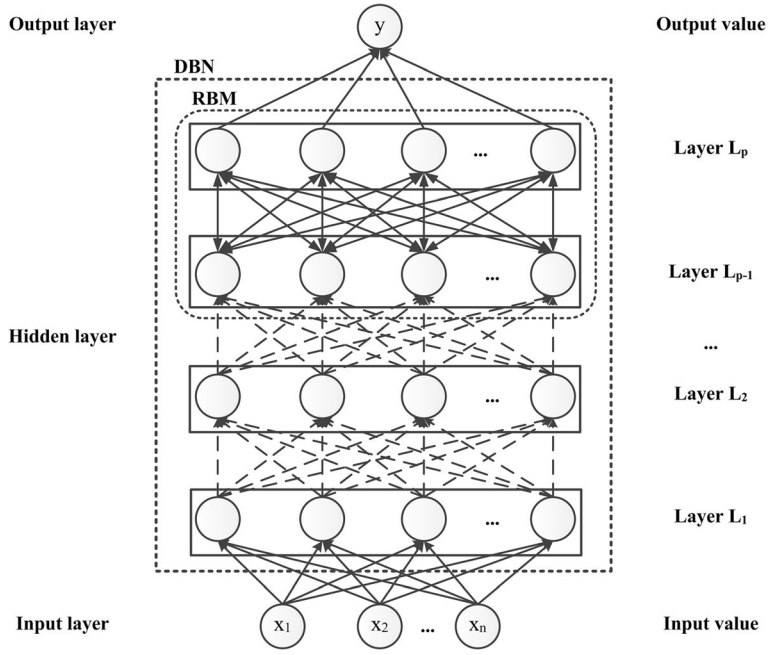


Fig. 1 Structure of the DNN

$$E(v,h|\theta) = -\sum_{i=1}^{V}\sum_{j=1}^{H} w_{ij}h_j\frac{v_i}{\sigma_i^2} - \sum_{i=1}^{V}\frac{(v_i-b_i)^2}{2\sigma_i^2} - \sum_{j=1}^{H} a_j h_j \tag{7}$$

where$\theta = (w, b, a)$is the parameter set,$w_{ij}$is the symmetric interaction weight between the units of hidden layers$l-1$and$l$,$b$and$a$are their respective bias terms,$V$and$H$represent the numbers of visible variables$v$and stochastic hidden units$h$, respectively, and$\sigma$is the standard deviation.

The conditional probability distributions$P$are defined as follows:

$$P\left(h_j = 1|v\right) = Sigm\left(\sum_{j=1}^{H} w_{ij}\frac{v_i}{\sigma_i^2} + a_j\right) \tag{8}$$

$$P(v_i = v|h) = Z\left(v|b_i + \sum_{j=1}^{H} w_{ij}h_j, \sigma_i^2\right) \tag{9}$$

where$Z(b, \sigma)$is a Gaussian probability density function.

The contrastive divergence algorithm (Hinton 2002) is used to update$w$as follows:

$$\Delta w_{ij} = \eta\left(\left\langle\frac{v_i}{\sigma_i^2}h_j\right\rangle - \left\langle\frac{v_i'}{\sigma_i^{2'}}h_j'\right\rangle\right) \tag{10}$$

where$\eta$is the learning rate, and$\langle\cdot\rangle$is the expectation of the training data.

In this manner, multiple GRBMS can be integrated into a DBN, and the process of feature learning continues until the required number of hidden layers has been obtained.

2.2.2 Fine-Tuning with Back-Propagation

After unsupervised pre-training, the overall DNN model can be fine-tuned using a standard back-propagation algorithm, i.e., back propagation (BP). BP can realize backwards propagation of the errors between the outputs and targets when the layers flow forward. Therefore, to minimize the loss function, a gradient is used in an optimization method to successively update the weight$w_i^l$and bias parameters$b_i^l$of the network. Because the network already knows the characteristics of the given training data, this fine-tuning is helpful to train DNN.

2.3 Performance Assessment Criteria

To quantitatively assess the prediction performance of the above methods, three commonly criteria were applied: the Nash-Sutcliffe efficiency (*NSE*), root mean square error (*RMSE*) and mean absolute error (*MAE*). The calculation equations of the three criteria are shown in Table 1.

where$X(t)$and$\hat{X}(t)$denote the recorded and forecasted runoff series, respectively,$\overline{X}(t)$is the average values of the records, and$N$is the number of runoff series.

3 Hybrid VMD-DNN Model for Daily Runoff Forecasting

To improve the prediction accuracy, a novel hybrid VMD-DNN model is proposed for daily runoff forecasting. The proposed hybrid model includes three important stages: (1) decomposition stage, (2) deep feature learning stage, and (3) forecasting stage. In the first stage, a new decomposition algorithm known as VMD is used to decompose the original runoff data into multiple IMFs to make the series stationary. During the second stage, an experimental method is used to confirm the optimal parameter settings of the DNN model for each decomposed IMF. In the final stage, the trained DNN model is applied as a forecasting tool to establish the prediction models. The ensemble prediction result is obtained by summing the predicted subresults of the modelled IMFs. The overall framework of the proposed model is shown in Fig. 2**.** The model algorithm flow is described as follows:

> **Step 1**: Decomposition. VMD is applied to decompose the original daily runoff series$X(t)(t = 1, 2, ..., n)$into multi-modes,$IMF_i(t)(i = 1, 2, ..., k)$. The extracted components represent a range from low frequency to high frequency and reveal various features hidden in the daily runoff. The details of the VMD algorithm are given in Section 2.1.

Table 1 Definition of assessment criteria

Performance assessment criteria	Formula
Nash-Sutcliffe efficiency	$NSE = 1-\frac{\sum_{t=1}^{N}\left(X(t)-\hat{X}(t)\right)^2}{\sum_{t=1}^{N}\left(X(t)-\overline{X}(t)\right)^2}$
Root mean square error	$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(X(t)-\hat{X}(t)\right)^2}$
Mean absolute error	$MAE = \frac{1}{N}\sum_{t=1}^{N}\lvert X(t)-\hat{X}(t)\rvert$

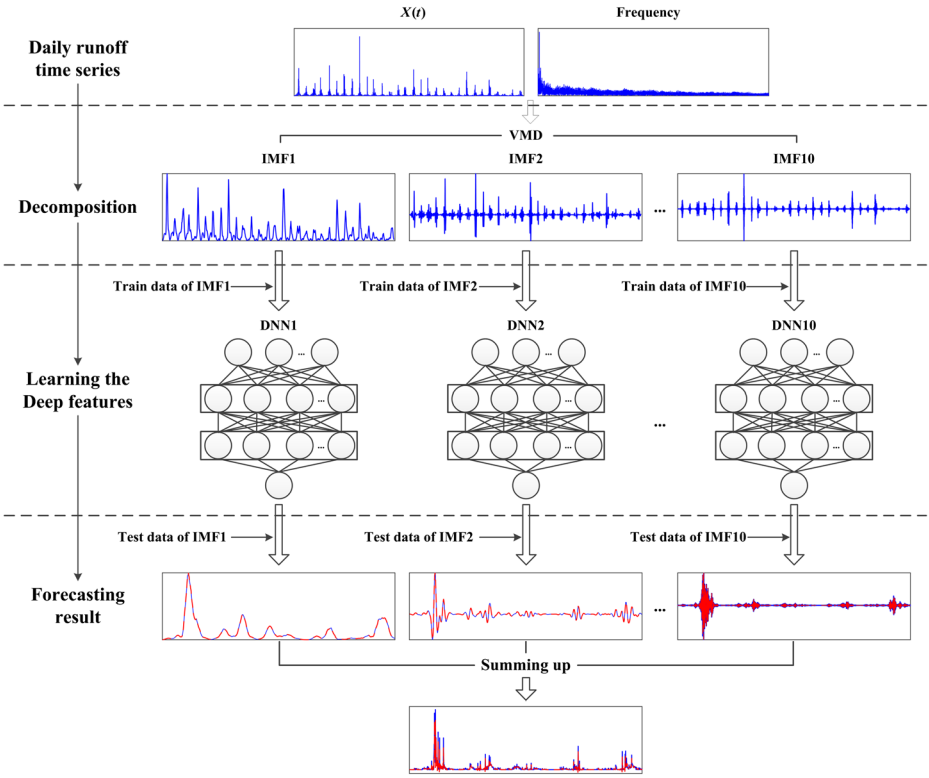


Fig. 2 Overall framework of the proposed hybrid VMD-DNN model for daily runoff forecasting

Step 2: Deep feature learning. After the decomposition process, each IMF is normalized by Eq. (11) to convert input data to the range [0,1]. An experimental method that decreases or increases the number of hidden layers and the number of neurons in each layer is subsequently used to confirm the optimal parameter settings of the DNN model for each of the normalized IMFs, and the number of inputs is obtained by analysing the resulting partial autocorrelation diagram.

(a) Standardization. Because the runoff series are highly nonlinear and time-varying, the extracted components$IMF_i(t)(i = 1, 2, ..., k)$must be normalized to avoid large fluctuations and numerical difficulties during the training process. The formula for normalization is given as follows:

$$x_i^{'} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{11}$$

where$x_{\max}$and$x_{\min}$denote the minimum and the maximum of the extracted components, respectively.

(b) Training. During the training process, the numbers of hidden layers and hidden nodes are two critical parameters in the DNN model. Using the normalized training dataset, the number of hidden layers is initialized from 1 to 2 (the initialized stopping condition is

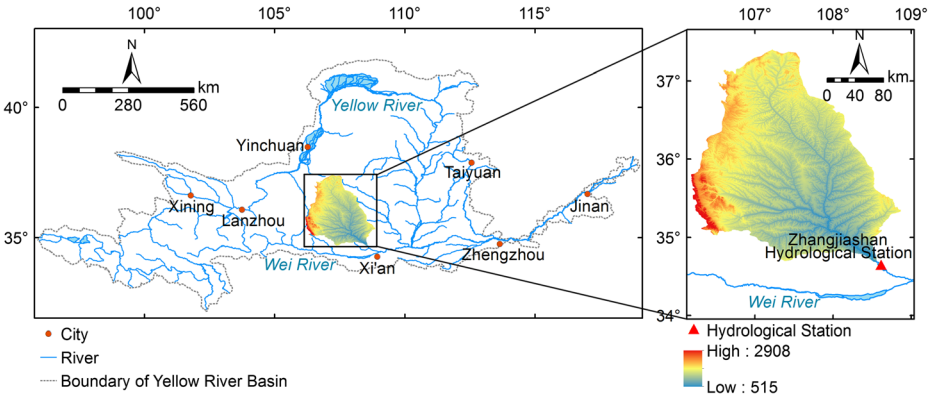


Fig. 3 Location of the Zhangjiashan Hydrological Station in Jing River, China

based on the DNN reaching convergence), and the number of hidden nodes is divided into 10 levels, ranging from 3 to 12 (interval of 1).

Step 3: Forecasting. The DNN model, trained in Step 2, is used as a forecasting tool to establish the corresponding prediction models for each decomposed IMF. The ensemble prediction results are subsequently obtained by summing up the predicted sub-results of the modelled IMFs.

4 Study Area and Data

The study area is located in the Jing River, which is the primary tributary of the Wei River and a second-level tributary of the Yellow River. Jing River originates from Liupan Mountain, flows through the provinces of Ningxia, Gansu, and Shaanxi, and traverses the Loess plateau of northwest China with a total length of 455.1 km. The basin covers an area of 45,421 km^2 and is located in the centre of the Loess Plateau (34°46′–37°19′N, 106°14′–108°42′E). A digital elevation model (DEM) of the region and the rivers located in the basin is presented in Fig. 3 (He et al. 2015). The altitude ranges from 515 m to 2908 m, and the southeastern portion of the basin is lower in altitude than the northwestern portion. According to the landform types,

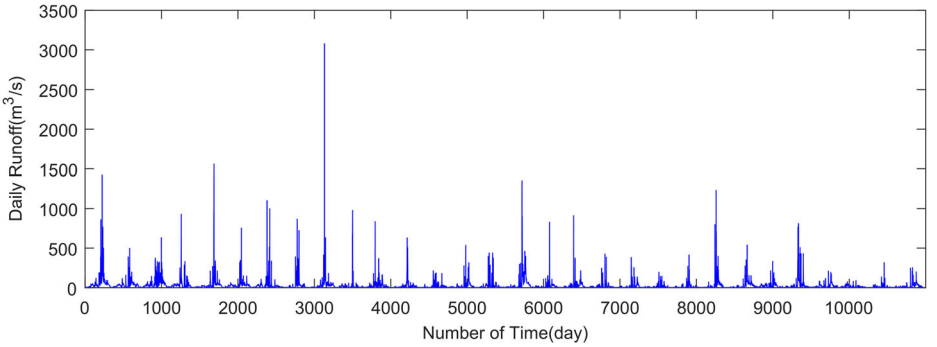


Fig. 4 Daily runoff hydrograph of ZHS ranged from 1/1/1988 to 31/12/2017

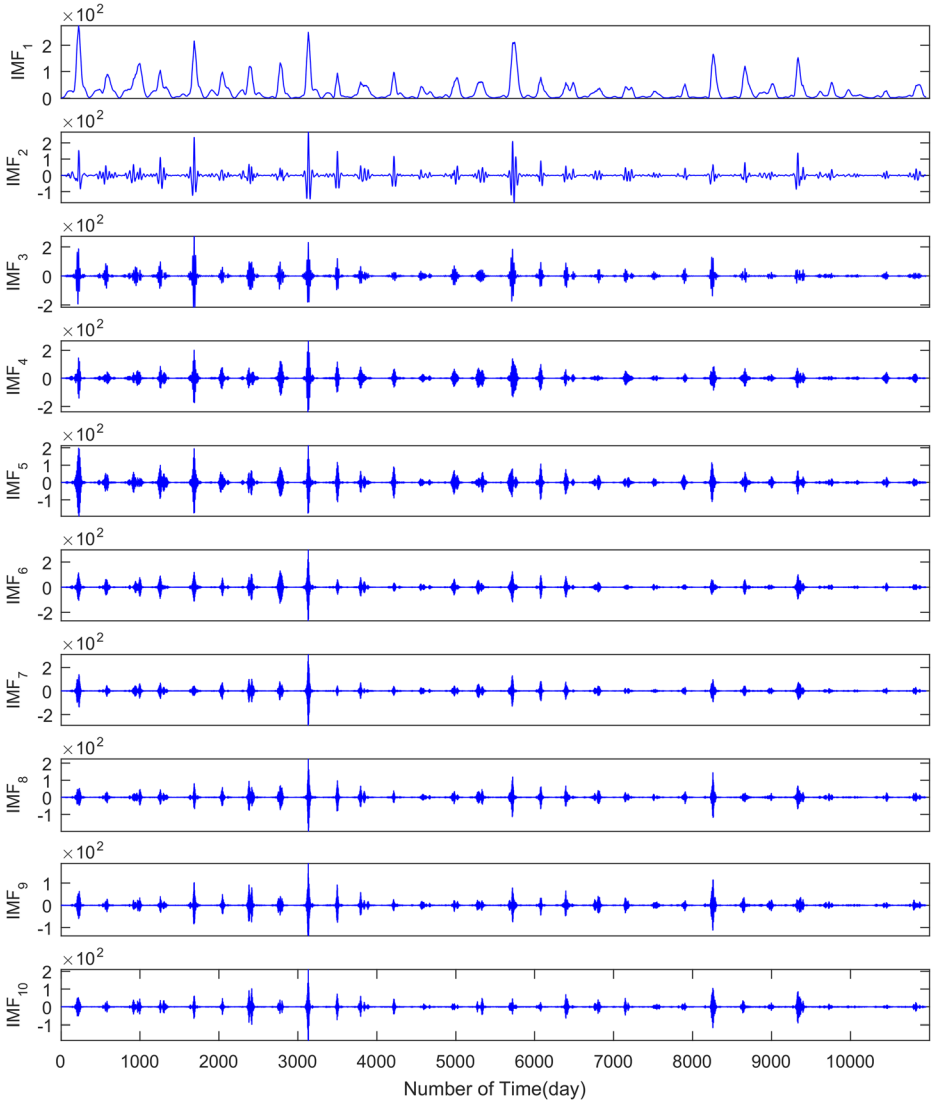


Fig. 5 Decomposition results of daily runoff data in ZHS

the basin can be divided into loess hilly and gully region, gully region of the loess plateau, earth-rocky mountainous area, loess hilly region and loess tableland region, of which more than 4.3% is earth-rocky mountainous, 41.7% is loess tableland and broken plateau, and 48.8% is loess hilly and gully regions. The total area of soil erosion reaches nearly 73% of the total land cover, making the basin one of the most soil-eroded areas of the Loess Plateau and an important source of coarse sediment delivered to the Yellow River. The soil is predominantly silt loam with silt content greater than 50%. The climate of the Jing River Basin is continental monsoonal, with an average annual precipitation that ranges from 400 to 600 mm and an average annual air temperature ranging from 8.0 to 10.0 °C (Ran et al. 2006). The basin

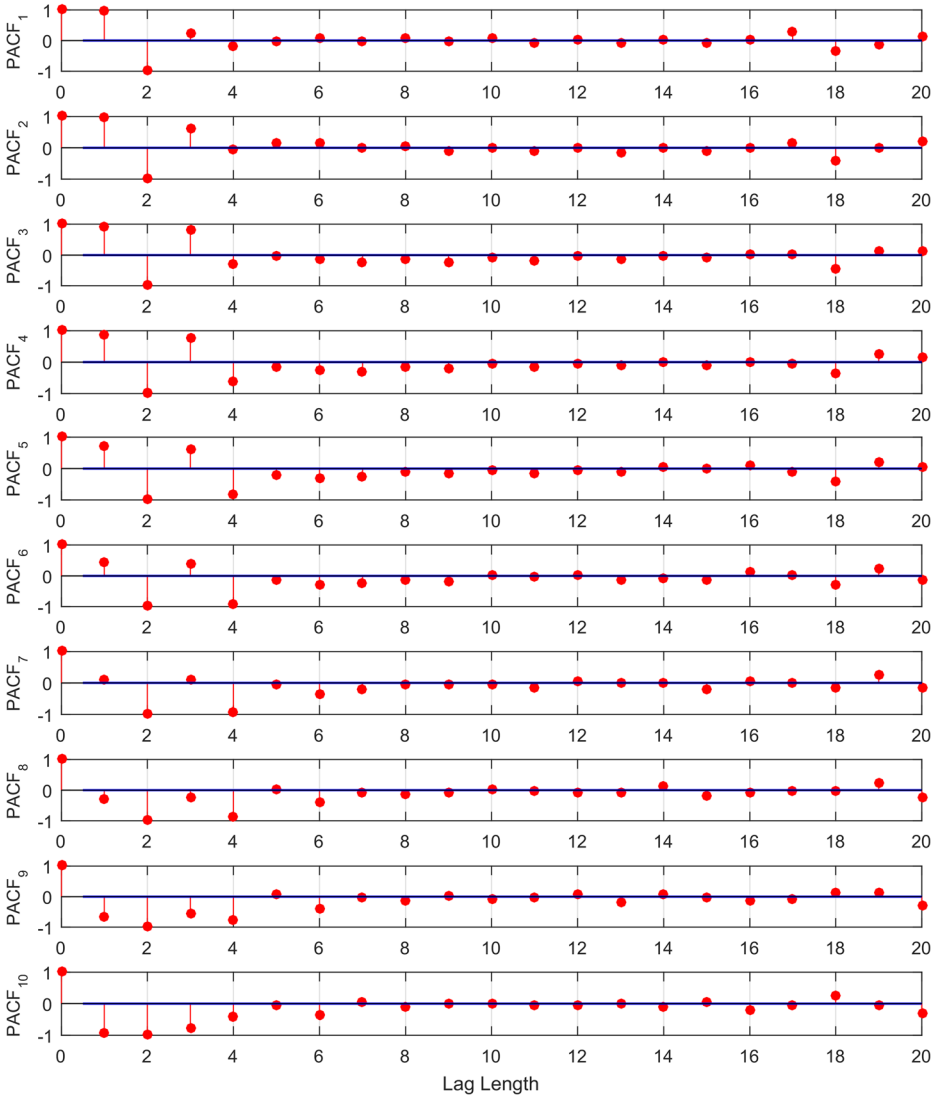


Fig. 6 PACF of each normalized IMFs decomposed by VMD

is located within a transition zone between semi-arid and semi-humid climates and exhibits high intra- and inter-annual variability of climate and runoff.

The data studied in this paper were retrieved from the Zhangjiashan Hydrological Station (ZHS, 34°38′N, 105°26′E), which is the main stream control station of the Jing River, with a total area of 43,216km^2, accounting for 95% of the total area of Jing River Basin. The daily runoff hydrograph of ZHS was measured from 1/1/1988 to 31/12/2017, as shown in Fig. 4**.** These measured values are collected from the hydrological information datacenter of Shaanxi Hydrographic and Water Resources Survey Bureau. In particular, the daily runoff value (m^3/s/day) was calculated by the instantaneous value (m^3/s) at 8 a.m. that is considered to be the average value in one day.

Table 2 Optimal input variables of each IMF decomposed by VMD

Decomposed IMFs	Input variables	Numbers of input	Output
IMF_1	$x_1(t-1)(t-2)(t-3)$	3	$x_1(t)$
IMF_2	$x_2(t-1)(t-2)(t-3)$	3	$x_2(t)$
IMF_3	$x_3(t-1)(t-2)(t-3)(t-4)$	4	$x_3(t)$
IMF_4	$x_4(t-1)(t-2)(t-3)(t-4)$	4	$x_4(t)$
IMF_5	$x_5(t-1)(t-2)(t-3)(t-4)$	4	$x_5(t)$
IMF_6	$x_6(t-1)(t-2)(t-3)(t-4)$	4	$x_6(t)$
IMF_7	$x_7(t-1)(t-2)(t-3)(t-4)$	4	$x_7(t)$
IMF_8	$x_8(t-1)(t-2)(t-3)(t-4)$	4	$x_8(t)$
IMF_9	$x_9(t-1)(t-2)(t-3)(t-4)$	4	$x_9(t)$
IMF_{10}	$x_{10}(t-1)(t-2)(t-3)(t-4)$	4	$x_{10}(t)$

5 Results and Discussion

The proposed VMD-DNN model was applied for daily runoff forecasting of the ZHS and was assessed using three common criteria. The prediction performance of the hybrid model was compared with that of five other hybrid models.

5.1 Decomposition Result

As introduced in Section 2, the number of modes*K*is the key parameter that affects the decomposed results of VMD. In this work, according to multiple pre-experiments, a suitable value of*K*was determined by analysing the iterative results of the centre frequency based on the mode*K*of different numbers. Finally, through extensive experiments, it was found that when the number of*K*increased to a certain value, an obvious convergence appeared in the iterative results, a phenomenon described as central frequency aliasing. Therefore, when$K = 10$, this critical point is chosen as the optimal value. Figure 5 shows 10 IMFs decomposed by VMD.

5.2 Learning the Features of the DNN

As described in Section 2, critical parameters such as the numbers of inputs, hidden nodes, and hidden layers are prerequisites for determining the structure of DNN. As a statistical tool, the partial autocorrelation function (PACF) (Huang et al. 2014) can be applied to overcome the limitation of the relationship between the inputs and outputs to obtain the input variables of DNN. The values of PACF corresponding to the lag length are calculated and presented in Fig. 6, of which $PACF_1$-$PACF_{10}$ denote the PACF of each normalized IMF. From Fig. 6, under the condition that the PACF at lag length falls within the confidence interval, the previous

Table 3 Evaluation of a DNN with 1 hidden layer for the IMF_1 of VMD

Predicted structure	NSE	RMSE	MAE	Predicted structure	NSE	RMSE	MAE
3–3-1	0.9999675	0.1363701	0.0531204	3–8-1	0.9999661	0.1393763	0.0545189
3–4-1	0.9999680	0.1353411	0.0527600	3–9-1	0.9999650	0.1415186	0.0569714
3–5-1	**0.9999684**	**0.1345273**	**0.0526057**	3–10-1	0.9999672	0.1369679	0.0540563
3–6-1	0.9999660	0.1394376	0.0528651	3–11-1	0.9999662	0.1391011	0.0573580
3–7-1	0.9999672	0.1371323	0.0527783	3–12-1	0.9999675	0.1363777	0.0528247

Table 4 Evaluation results of a DNN with 2 hidden layers for IMF_1 of VMD

Predicted structure	NSE	RMSE	MAE	Predicted structure	NSE	RMSE	MAE
3–3–7-1	**0.9999680**	**0.1353128**	**0.0515833**	3–8–6-1	0.9999667	0.1381517	0.0532335
3–4–3-1	0.9999679	0.1355256	0.0532438	3–9–3-1	0.9999678	0.1357162	0.0521475
3–5–10-1	0.9999679	0.1356389	0.0519252	3–10–5-1	0.9999667	0.1380241	0.0526726
3–6–6-1	0.9999671	0.1372905	0.0513703	3–11–4-1	0.9999671	0.1371802	0.0539445
3–7–7-1	0.9999669	0.1376938	0.0527946	3–12–7-1	0.9999673	0.1368899	0.0535196

value can be treated as the numbers of inputs. According to analysis of the PACF of the corresponding lag length in Fig. 6, the optimal input variables of each IMF can be obtained, and the results are shown in Table 2.

It remains a difficult task to determine the number of hidden nodes and the number of hidden layers, and no mature theoretical method is available to solve this problem. Instead, an experimental method that decreases or increases the number of layers and the number of neurons in each layer, is used to determine the optimal parameter settings. In this work, the daily runoff series can be divided into two portions: the training dataset (1/1/1988 to 31/12/2012) and the testing dataset (1/1/2013 to 31/12/2017). The first IMF of VMD, i.e., IMF_1, is used as an example to illustrate the experimental process.

In this experiment, a DNN training structure is initialized with 3 input nodes and 1 hidden layer, the number of hidden layer nodes is divided into 10 levels from 3 to 12 (interval of 1), and the number of hidden layers is initialized from 1 to 2 (the stopping condition is based on the DNN reaching convergence). Table 3 shows the DNN evaluation results with 3 input nodes and 1 hidden layer based on the three assessment criteria. It is obvious that the optimal learning performance is obtained, i.e., 3–5-1 (input layer: 3, hidden layer: 5, and output layer: 1).

The initialization of a DNN has 3 input nodes and 2 hidden layers, with 10 hidden nodes in the first hidden layer, and 100 hidden nodes in the second hidden layer. Table 4 shows the optimal prediction performance of DNN with the first hidden layer corresponding to the 10 hidden nodes in the second hidden layer.

The above results show that the best values of NSE, RMSE, and MAE appear for the architecture 3–3–7-1(input layer: 3, hidden layer: first 3, second 7, and output layer: 1). However, the prediction performance of the DNN with two hidden layers was not as good as that with one hidden layer, with values of NSE = 0.9999684, RMSE = 0.1345273, and MAE = 0.0526057. Therefore, it is necessary to prevent the experiment from selecting the hidden layer. The optimal predicted structure is a DNN with one hidden layer, i.e., 3–5-1.

Table 5 Evaluation results of different DNNs with different hidden layers for the 2nd-10th IMF of VMD

Decomposed IMFs	Predicted structure	NSE	RMSE	MAE
IMF_2	3–10–6-1	0.9998702	0.1841126	0.0515833
IMF_3	4–11-1	0.9986862	0.2694962	0.1269800
IMF_4	4–12–5-1	0.9978675	0.4108998	0.1790288
IMF_5	4–8–8-1	0.9921593	0.6250176	0.2809470
IMF_6	4–9–6-1	0.9870264	1.0773741	0.4591809
IMF_7	4–12–12-1	0.9786577	1.2430514	0.6457458
IMF_8	4–4–4-1	0.9758000	1.0846822	0.4676694
IMF_9	4–4–3-1	0.9812027	0.6358390	0.2868871
IMF_{10}	4–10–4-1	0.9987790	0.2897115	0.1248427

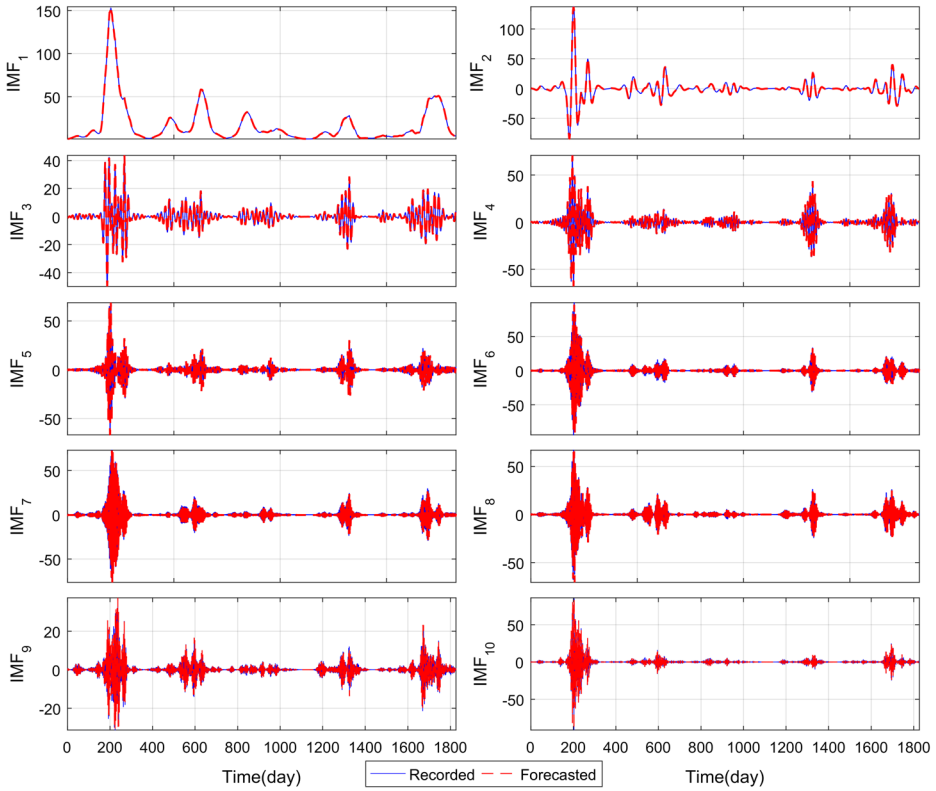


Fig. 7 Predicted and recorded results of IMFs decomposed by VMD

Thus, this analysis reveals that the learning performance of DNN with multiple hidden layers is not always better than a DNN with one hidden layer. In other words, excess hidden layers of the DNN might lead to an over-fitting problem.

According to the experimental steps described above, the 2nd-10th IMFs of VMD can be forecasted by the corresponding trained DNNs, and the evaluation results are summarized in Table 5.

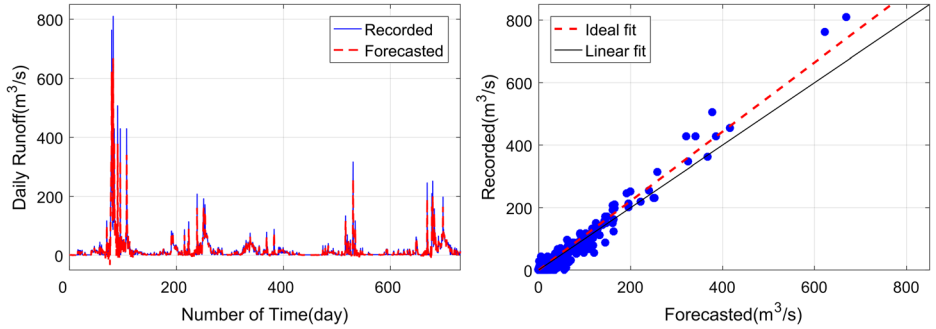


Fig. 8 Final predicted result and the scatters of VMD-DNN

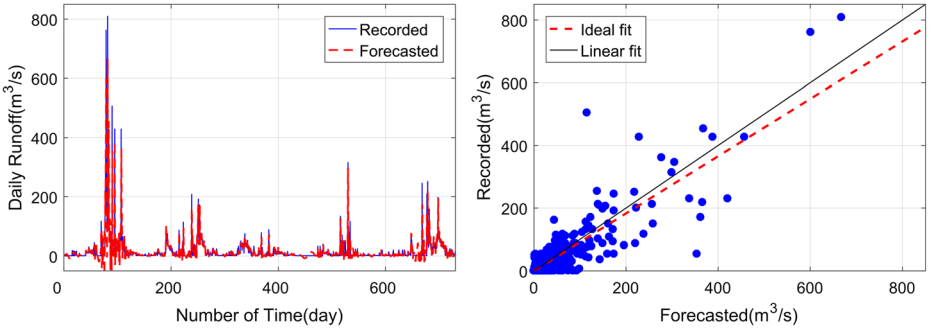


Fig. 9 Final predicted result and the scatters of EMD-DNN

From Table 5, the best predicted structure for IMF_2 is 3–10-1 in the first hidden layer. However, a DNN with 2 hidden layers (hidden nodes in the first and the second hidden layer set as 10 and 6, respectively) for IMF_2 showed the best overall performance, with corresponding performance assessment values of NSE = 0.9998702, RMSE = 0.1841126, and MAE = 0.0773231. Hence, the optimal predicted structure for IMF_2 is a DNN with two hidden layers, i.e., 3–10–6-1. Meanwhile, the best prediction results for the 3rd-10th IMF of VMD are indicated with bold text in Table 5. It can be clearly observed that the best predicted structure is distributed in one hidden layer or two hidden layers. This result further illustrates that for feature learning, additional hidden layers do not always correlate to better performance. The prediction results of IMFs decomposed by VMD were transformed into real data by inverse normalization and are plotted in Fig. 7, which shows the predicted and recorded results of each IMFs of the best prediction structure in the testing dataset.

5.3 Forecasting Results

The ensemble prediction results were obtained by summing the predicting sub-results of the modelled IMFs of the daily runoff data during the period 1/1/2013–31/12/2017. The final predicted results using the proposed model VMD-DNN are displayed in Fig. 8. In addition to the predicted and recorded results, scatter plots are presented to analyse the agreement between the predicted and recorded results.

It can be clearly observed from the data presented in Fig. 8 that the predicted values are generally consistent with the change trend of the recorded results, although the predicted effect

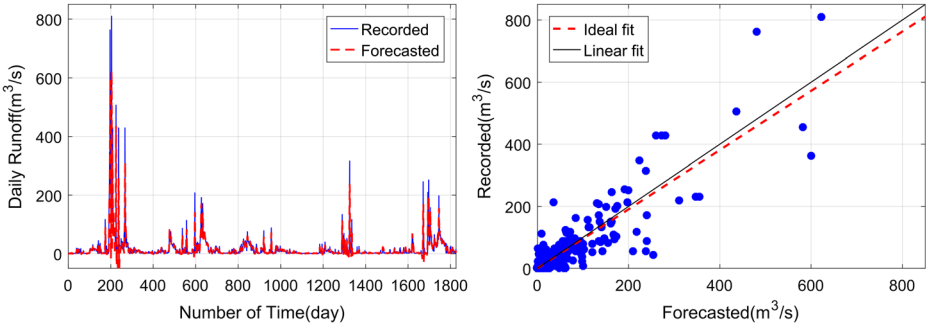


Fig. 10 Final predicted result and the scatters of EEMD-DNN

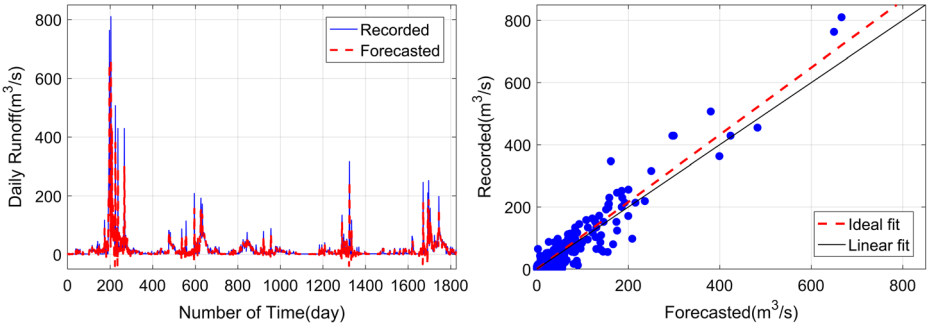


Fig. 11 Final predicted result and the scatters of MFE-VMD-DNN

is not particularly ideal at each individual peak point. Additionally, in the scatter diagram, the predicted and recorded values are mostly concentrated near the ideal fit, which shows a higher consistency. Additionally, the small angle between the linear fit and ideal fit further indicates that the model adopted in this paper has good prediction ability. The quantitative assessment results of NSE = 0.95, RMSE = 9.92, and MAE = 3.82 show good precision.

5.4 Comparison Analysis

To further test the feasibility and superiority of the VMD-DNN hybrid model, this model was compared with EMD-DNN, EEMD-DNN, and multi-scale feature extraction -based VMD-DNN, EMD-DNN and EEMD-DNN (MFE-VMD-DNN, MFE-EMD-DNN, MFE-EEMD-DNN). Specifically, ensemble members$M$and the amplitude of the added white noiseε are key parameters that affect the decomposition results of EEMD. In this paper, according to reference (Wu and Huang 2009) and pre-experiments,Mandεwere set to 100 and 0.2, respectively.

Figures. 9, 10, 11, 12 and 13 display the forecasting results using the other five models. From Fig. 9, it is clear that the forecasted data exhibits a stronger tracking ability for the recorded data, indicating that EMD-DNN has better performance. As shown in Fig. 10, EEMD-DNN presents a positive fitting ability for runoff changes, and the scatters of the forecasted and recorded data are more concentrated close to the ideal fit. The MFE–VMD-DNN model is shown in Fig. 11 and reveals that most of the forecasted points are close to the recorded data, but this model offers only a poor capacity for prediction of peak points. MFE-

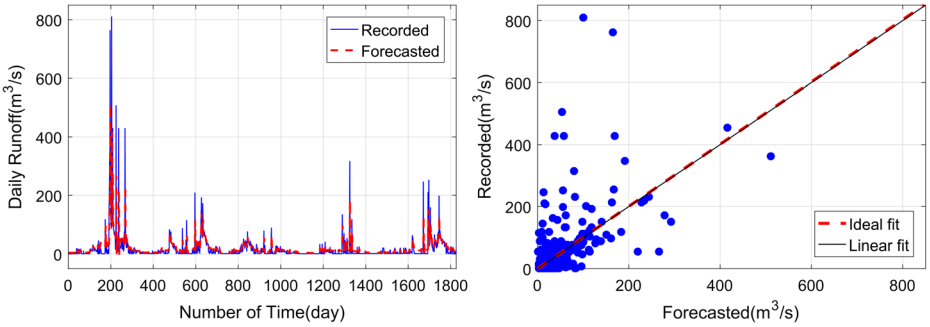


Fig. 12 Final predicted result and the scatters of MFE-EMD-DNN

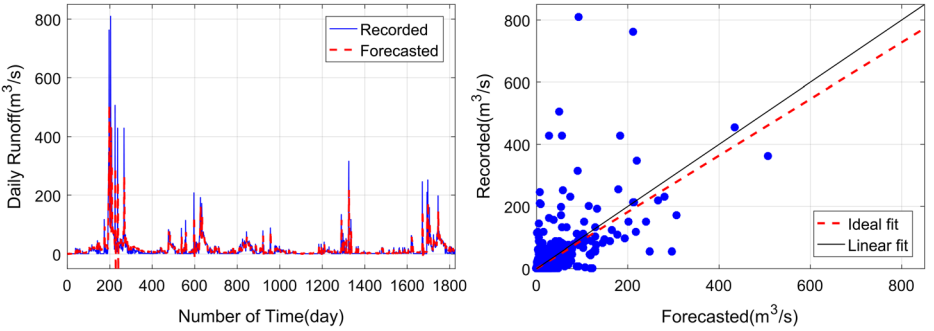


Fig. 13 Final predicted result and the scatters of MFE-EEMD-DNN

EMD-DNN and MFE-EEMD-DNN were also tested and the results are shown in Fig. 12 and 13**.** Limited fitting and tracking ability are displayed, and the scatter plots are more dispersed than the above three models.

Comparing the results of the above models with the results of Fig. 8, on the one hand, Fig. 8 shows a better ability to fit the trend and periodicity than the other five peer models, especially during peak inflow, and on the other hand, the scatter plot of Fig. 8 also reveals a higher correlation between the predicted and recorded results. Therefore, the qualitative comparisons of the six models indicate that VMD-DNN shows the best prediction performance for both overall prediction and peak inflow prediction capability.

In addition, quantitative assessment can be performed. The NSE, RMSE, and MAE values were used to evaluate the prediction performances of the six models, as shown in Table 6**.** The following three observations can be concluded from the results: (1) For non-multiscale feature extraction, VMD-DNN shows better prediction performance for the same prediction method. The methods fall in the order of VMD-DNN>EEMD-DNN>EMD-DNN; (2) For multiscale feature extraction, MFE-VMD-DNN showed better prediction performance, based on the same prediction method. For comparison of this method with the others, the methods fall in the order of MFE-VMD-DNN>MFE-EEMD-DNN>MFE-EMD-DNN; (3) By comparing the results of (1) and (2), it is clear that non-multiscale feature extraction has a higher prediction performance than multiscale feature extraction, i.e., VMD-DNN>MFE-VMD-DNN.

Via the above comparative analysis, two constructive conclusions can be reached: (1) Among the three decomposition methods described in this article, i.e., EMD, EEMD, and VMD, VMD exhibited the best noise reduction performance for the original runoff series; (2) The final forecasting result obtained by the non-multiscale feature extraction was obviously

Table 6 Comparison of prediction performances using different models

Forecasting Approach	Prediction Model	Performance assessment criteria		
		NSE	RMSE	MAE
Non-multiscale feature extraction	EMD-DNN	0.76	22.49	9.21
	EEMD-DNN	0.80	20.71	6.99
	VMD-DNN	**0.95**	**9.92**	**3.82**
Multiscale feature extraction	MFE-EMD-DNN	0.40	44.99	16.59
	MFE-EEMD-DNN	0.42	35.12	11.73
	MFE-VMD-DNN	0.91	14.08	5.61

better than that obtained by multiscale feature extraction. To summarize, the comparative results indicate that the proposed hybrid model VMD-DNN has better forecasting performance than the other models.

6 Conclusions

In this paper, a proposed model based on variational mode decomposition and deep neural networks was used to forecast the daily runoff of the Zhangjiashan Hydrological Station in Jing River, China. The three steps of this model were applied for runoff forecasting: (1) Decomposition. The runoff original series were decomposed into multiple IMFs by VMD; (2) Deep feature learning. Using an experimental method, the deep feature values of the DNN were fully learned for forecasting each decomposed IMF; (3) Forecasting. The trained DNN model was applied as a forecasting tool to establish the prediction models, and the ensemble prediction results were obtained by summing the predicting sub-results of the modelled IMFs. This model was compared with five hybrid models, i.e., EMD-DNN, EEMD-DNN, MFE-VMD-DNN, MFE-EMD-DNN, and MFE-EEMD-DNN. Of these, VMD-DNN exhibited the best prediction performance among all comparable methods in terms of both overall prediction and peak inflow prediction capability.

Overall, in addressing the forecasting of the daily runoff series, this work successfully demonstrates that the performance of daily runoff forecasting can be significantly enhanced using the proposed VMD-DNN model. The three stages of decomposition by VMD, deep feature learning with an experimental method, and forecasting by DNN were established for a hybrid model to solve the daily runoff forecasting. It should be noted that the proposed hybrid model offers two main advantages: (1) Reduction of the number of decomposed IMFs, thereby shortening the time required to build the predictive model and decreasing the forecasting workload; (2) Reduction of accumulative error during multiscale feature extraction. Overall, the results indicate that the proposed VMD-DNN hybrid model is a useful tool for predicting complex time series data that are complicated by non-stationary and nonlinearity, making this approach a promising new method for daily runoff forecasting. Further research should be considered to solve the following two problems: (1) The decomposed IMFs could be split into high-frequency terms and low-frequency terms, of which the low-frequency terms can be predicted by time-series models, and the high-frequency terms can be predicted using more advanced artificial intelligence techniques; (2) The proposed hybrid model can consider selected factors that affect hydrological time series such as climate change instead of univariate time series analysis.

Acknowledgments This work was supported by the National Key R&D Program of China under Grant No. 2016YFC0401409 and the National Natural Science Foundation of China under Grant Nos. 51679186 and 51679188.

Compliance with Ethical Standards

Conflict of Interest The authors declare no conflicts of interest.

References

Ali M, Khan A, Rehman NU (2018) Hybrid multiscale wind speed forecasting based on variational mode decomposition. Int Trans Electr Energ Syst 28(1):e2466

Bai Y, Chen ZQ, Xie JJ, Li C (2016) Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. J Hydrol 532:193–206

Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828

Bengio Y, Lamblin P, Popovici D, Larochelle H (2007) Greedy layer-wise training of deep networks. Adv Neural Inf Proces Syst 19:153–160

Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) Time Series Analysis: Forecasting and Control. John Wiley & Sons

Chellali F, Khellaf A, Belouchrani A (2010) Wavelet spectral analysis of the temperature and wind speed data at Adrar. Algeria Renew Energ 35(6):1214–1219

Cheng CT, Niu WJ, Feng ZK, Shen JJ, Chau KW (2015) Daily reservoir runoff forecasting method using artificial neural network based on quantum-behaved particle swarm optimization. Water 7(8):4232–4246

Cho KH, Ilin A, Raiko T (2011) Improved learning of Gaussian-Bernoulli restricted Boltzmann machines. Lect Notes Comput Sci 6791:10–17

Ciresan DC, Meier U, Gambardella LM, Schmidhuber J (2010) Deep big simple neural nets excel on handwritten digit recognition. Neural Comput 22(12):3207–3220

Citakoglu H, Cobaner M, Haktanir T, Kisi O (2014) Estimation of long-term monthly mean reference evapotranspiration in Turkey. Water Resour Manag 28(1):99–113

Collobert R, Weston J (2008) A unified architecture for natural language processing: deep neural networks with multitask learning. Proceedings of the 25th international conference on. Mach Learn:160–167

Dabuechies I (1990) The wavelet transform, time-frequency localization and signal analysis. IEEE Trans Inf Theory 36:6–7

Di CL, Yang XH, Wang XC (2014) A four-stage hybrid model for hydrological time series forecasting. PLoS One 9(8):e104663

Dragomiretskiy K, Zosso D (2014) Variational mode decomposition. IEEE Trans Signal Process 62(3):531–544

He Y, Wang F, Mu X, Yan H, Zhao G (2015) An assessment of human versus climatic impacts on Jing River Basin, Loess Plateau, China. Advances in Meteorology. Article ID 478739

Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B (2012) Deep neural networks for acoustic modeling in speech recognition. IEEE Sigrurl Processing Magazine 29(6):82–97

Hinton GE (2002) Training products of experts by minimizing contrastive divergence. Neural Comput 14(8): 1771–1800

Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7): 1527–1554

Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng QA, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond A 454:903–995

Huang SZ, Chang JX, Huang Q, Chen YT (2014) Monthly streamflow prediction using modified EMD-based support vector machine. J Hydrol 511:764–775

Kisi O, Cimen M (2012) Precipitation forecasting by using wavelet-support vector machine conjunction model. Eng Appl Artif Intell 25:783–792

Koza JR (1992) Genetic Programming: On the programming of computers by means of natural selection. MIT Press 33:69–73

Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. Commun ACM 60(6):84–90

Lahmiri S (2015) Long memory in international financial markets trends and short movements during 2008 financial crisis based on variational mode decomposition and detrended fluctuation analysis. Phys A, Stat Mech Appl 437:130–138

Lahmiri S, Boukadoum M (2015) Physiological signal denoising with variational mode decomposition and weighted reconstruction after DWT thresholding. In Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), Lisbon, Portugal pp 806–809

Li C, Bai Y, Zeng B (2016) Deep feature learning architectures for daily reservoir inflow forecasting. Water Resour Manag 30(14):5145–5161

Lin GF, Chen GR, Huang PY, Chou YC (2009) Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods. J Hydrol 372(1–4):17–29

Liu H, Mi XW, Li YF (2018) Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis. LSTM network and ELM Energy Convers Manage 159:54–64

Moghaddamnia A, Ghafari M, Piri J, Amin S, Han D (2009) Evaporation estimation using artificial nctworlcs and adaptive ncuro-fuzzy inference system techniques. Adv Water Resour 32:88–97

Mohamed AR, Dahl GE, Hinton G (2012) Acoustic modeling using deep belief networks. IEEE Trans on Audio Speech Language Process 20(1):12–22

Naik J, Dash S, Dash PK, Bisoi R (2018) Short term wind power forecasting using hybrid Variational mode decomposition and multi-kernel regularized Pseudo inverse neural network. Renew Energy 118:180–212

Nayak PC, Sudheer KP, Rangan DM, Ramasastri KS (2004) A neuro-fuzzy computing technique for modeling hydrological time series. J Hydrol 291(1):52–66

Okkan U, Serbes ZA (2012) Rainfall-runoff modeling using least squares support vector machines. Environmetrics 23(6):549–564

Ran DC, Liu B, Wang H, Luo QH, Ma Y (2006) Soil and water conservation measures and their benefits in runoff and sediment reductions of typical tributary in the middle of Yellow River. The Yellow River Water Conservancy Press, Zhengzhou, China

Sattari MT, Yurekli K, Pal M (2012) Performance evaluation of artificial neural network approaches in forecasting reservoir inflow. Appl Math Model 36(6):2649–2657

Toth E, Brath A, Montanari A (2000) Comparison of short-term rainfall prediction models for real-time flood forecasting. J Hydrol 239(1):132–147

Wang WC, Chau KW, Cheng CT, Qiu L (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J Hydrol 374(3):294–306

Wang WC, Xu DM, Chau KW, Chen SY (2013) Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. J Hydroinf 15(4):1377–1390

Wang ZY, Qiu J, Li FF (2018) Hybrid models combining EMD/EEMD and ARIMA for Long-term streamflow forecasting. Water 10(7):853

Wu ZH, Huang NE (2009) Ensemble empirical mode decompostion: a noise-assisted data analysis method. Adv Adaptive Data Anal 1(1):1–41

Zhang J, Cheng CT, Liao SL, Wu XY, Shen JJ (2009) Daily reservoir inflow forecasting combining QPF into ANNs model. Hydro Earth Syst Sci 6:121–150