An Integrated Framework of GRU Based on Improved Whale Optimization Algorithm for Flood Prediction

Chunlei Ji  
Huaiyin Institute of Technology

Tian Peng (✉ husthydropt@126.com)  
School of hydropower and Engineering, Huazhong University of Science and Technology

Chu Zhang  
Huaiyin Institute of Technology

Lei Hua  
Huaiyin Institute of Technology

Wei Sun  
Huaiyin Institute of Technology

Research Article

Keywords: OVMD, random forests algorithm, GRU, WOA, flood prediction

Posted Date: December 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-947198/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
An integrated framework of GRU based on improved whale optimization algorithm for flood prediction

Chunlei Ji\(^a\), Tian Peng\(^a*,b\), Chu Zhang \(^a,b\), Lei Hua \(^a\), Wei Sun \(^a\)

\(^a\) College of Automation, Huaiyin Institute of Technology, Huaian 223003, China

\(^b\) Jiangsu Permanent Magnet Motor Engineering Research Center, Huaiyin Institute of Technology, Huaian 223003, China

Email addresses:

Chunlei Ji: jichunleii@163.com

Tian Peng: husthydropt@126.com

Corresponding author:

Name: Tian Peng

Telephone number: +86 15005238786

Complete postal address: College of Automation, Huaiyin Institute of Technology, Huaian 223003, China.
Abstract

Accurate prediction of floods is the first step in formulating flood control strategies and reducing flood disasters. This research proposes a deep learning model based on Gate Recurrent Unit (GRU), Random Forest Algorithm (RF), Whale Optimization Algorithm (WOA) and Optimal Variational Mode Decomposition (OVMD) for flood prediction. First, the random historical time series is decomposed using OVMD. Secondly, combined with the RF feature importance measurement, select features with high importance to obtain the optimal input set. Third, use the GRU model to predict all sub-models, and use the WOA algorithm to optimize the hyperparameters in the GRU model. This study also proposes a hybrid strategy to improve the traditional WOA algorithm and enhance the optimization ability of the WOA algorithm. Finally, the prediction results of all sub-models were aggregated to generate the final prediction result. The model was validated using data from three hydrological stations in the upper, middle and lower reaches of the Minjiang river basin in China. Through the results of the experiment, it can be seen that the proposed prediction model can effectively predict the flood time series, and has better accuracy than other models.

Keywords: OVMD; random forests algorithm; GRU; WOA; flood prediction

1 Introduction

Floods are one of the most common and destructive disasters in nature, causing irreversible damage to the environment, human life and construction facilities (Akbari et al. 2020). Since 2000, more than 1.6 billion people worldwide have been affected by
floods (Allahbakhshian-Farsani et al. 2020). The deaths caused by flood disasters account for about 84% of the deaths caused by all natural disasters in the world (Jamali et al. 2020). Flood disasters are more dangerous than other natural catastrophic disasters (Chapi et al. 2017). It is expected that the impact of flood disasters will continue to increase in the future (Furquim et al. 2016). Due to the great danger of floods, research on flood forecasting has become very important. In recent years, scholars and researchers have achieved excellent results in this regard. Flood prediction models can be divided into physical models and data-driven models (Masrur Ahmed et al. 2021).

The physical model is mainly modeled by simulating the internal physical mechanism of the runoff process. They are suitable for basins with a large number of physical parameters, but they usually need massive amounts of data to calibrate and validate the models (Huo et al. 2020; Hussein et al. 2019; Partington et al. 2012). Physical models usually require large amounts of hydrological and meteorological data, complex mathematical knowledge and accurate knowledge of the physical processes of runoff formation (Xie et al. 2019). There are many limiting factors in the application process, which leads to the uncertainty of the model (Yoon et al. 2011).

The data-driven model is built using machine learning algorithms and statistics, and does not require specific physical relationships within the runoff. It uses the historical runoff data to infer the relationship between input variables and output variables (Deo and Şahin 2015). Data-driven models can be considered as black-box models and are very useful in natural system modeling (Yaseen et al. 2016). Many nonlinear models are also used for hydrological prediction in the development of
machine learning. Nikoo et al. (2016) proposed a method of artificial neural network (ANN) to predict floods and used a sociality-based algorithm to train the weight of ANN. The prediction result of this model is more accurate than the prediction result of the statistical model. Tehrany et al. (2015) proposed a flood prediction model that integrates support vector machine (SVM) and frequency ratio, and compared with the decision tree model to verify the effectiveness of the method. Bui et al. (2019) proposed a PSO-ELM model for the spatial prediction of flash floods. Extreme learning machine (ELM) is used to generate the initial flood model, while PSO is used to optimize the model. Compared with SVM, ANN and DT, PSO-ELM shows better performance. Zhou et al. (2019) proposed a flood prediction model of a fuzzy inference system embedded with genetic algorithm and least squares estimator. Moreover, compared with two different structures of the adaptive network-based fuzzy inference system (ANFIS), the superiority of the proposed model is verified.

In recent years, deep learning has made great progress in many fields. It has a deeper nonlinear network structure than traditional machine learning methods (Li et al. 2020). Recurrent Neural Networks (RNN) can quickly solve the mapping relationship between input and output sequences using context-related information. Once the neural network and the time span continue to grow recursively, the input of the hidden layer will gradually decay the output of the neural network, causing the gradient to disappear and gradient explosion (Ni et al. 2020). Long short-term memory (LSTM) and gated recurrent unit (GRU) networks are variants of RNN and are proposed to improve the shortcomings of RNN. Their basic idea is to give the recurrent neural network the ability
to control the accumulation of its internal information through the gated unit. This function can selectively forget information to prevent overload in the learning process. Although GRU and LSTM perform very similarly on many questions, we chose GRU for our experiment. Because the GRU model has fewer parameters and a simpler structure, it has a faster learning curve (Zhang et al. 2018). The training time of the GRU model is shorter than that of the LSTM model (Stergiou and Karakasidis 2021).

Many scholars have applied GRU to hydrological prediction or established related models, and achieved satisfactory results (Ayzel and Heistermann 2021; Pan et al. 2020; Wang et al. 2020a). Some scholars will use optimization algorithms to optimize certain parameters of the GRU to improve the performance of the GRU. This study introduces the Whale Optimization Algorithm (WOA) to enhance the predictive ability of the GRU model. The WOA algorithm has a simpler structure, fewer internal parameters, and stronger search capabilities than other optimization algorithms (Du et al. 2021). However, the WOA algorithm is easy to fall into a local optimal state (Sun and Zhang 2018). This research improves the traditional WOA algorithm to reduce the impact of this problem.

Flood flow data is usually noisy and non-stationary, and direct use of the data for prediction will reduce the accuracy of the accuracy. In order to improve the prediction accuracy, the use of various signal decomposition techniques in the hydrological field has become a trend. Such as Wavelet Transform (Shoaib et al. 2014; Wu et al. 2021), Empirical mode decomposition (EMD) (Huang et al. 2014) and Variational Mode Decomposition (VMD). However, when EMD decomposes the signal, problems such
as end effects and mode mixing will occur. VMD has a better decomposing effect than EMD (Fu et al. 2019). Compared with traditional decomposition techniques, VMD can recursively convert non-stationary signals into a series of orthogonal intrinsic mode function (IMF) sequences with natural frequencies. At present, there are many applications of VMD technology in hydrological time series (Fu et al. 2018; Xie et al. 2019; Zuo et al. 2020). In order to better apply VMD to time series decomposition, this study uses the optimized variational mode decomposition (OVMD) method proposed by Zhang et al. (2017).

Section 2 introduces the theoretical background of this article in detail, including OVMD, RF, WOA, improved WOA (IWOA) and the proposed OVMD-RF-IWOA-GRU model; Section 3 gives a case study, which specifically includes the research area, data processing, determination of input variables and parameter settings; Section 4 compares the prediction results of different models; Section 5 gives the conclusion of this research.

2 Methods

2.1 Optimal Variational Mode Decomposition

Variational Mode Decomposition (VMD) is a new non-recursive adaptive decomposition processing method proposed by Dragomiretskiy and Zosso (2014), which is used to replace empirical mode decomposition or empirical wavelet transform. The purpose of VMD is to decompose the real-valued input signal \( f \) into different numbers of sub-signals with limited bandwidth, namely Intrinsic Mode Function (IMF),
which can reproduce the original input signal according to its sparsity characteristics (Mohanty et al. 2020).

For a signal $f(t)$, its corresponding constrained variational mode is as follows:

$$
\left\{ \begin{array}{l}
\min_{\{u_k\}, \{w_k\}} \left\{ \sum_{k=1}^{K} \left| \partial \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-j\omega_k t} \right|^2 \right\} \\
\text{s.t.} \sum_{k=1}^{K} u_k(t) = y(t)
\end{array} \right. 
$$

(1)

where $K$ is the total number of modal decomposition, $u_k(t)$ represents $K$ IMFs decomposed by VMD, $w_k(t)$ represents the corresponding central frequencies to $u_k(t)$, $\partial$ is the time derivative of the function, $\delta(t)$ is the unit impulse function.

The equilibrium constraint parameter $\alpha$ and the lagrangian multiplicative operator $\lambda(t)$ are introduced to change Eq. (1) into a non-constrained optimization problem represented by Eq. (2).

$$
L(\{u_k\}, \{w_k\}, \lambda(t)) = \alpha \sum_{k=1}^{K} \left| \partial \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-j\omega_k t} \right|^2 + \\
\left\| y(t) - \sum_{k=1}^{K} u_k(t) \right\|^2 + <\lambda(t), y(t) - \sum_{k=1}^{K} u_k(t) >
$$

(2)

where $\left\| y(t) - \sum_{k=1}^{K} u_k(t) \right\|^2$ is the quadratic penalty term, $< \cdot >$ is the inner product operation, $\alpha$ is used to ensure the accuracy of the reconstruction, and $\lambda(t)$ is the lagrangian multiplicative operator.

The OVMD decomposition in this paper is based on the method proposed by Zhang et al. (2017). The mode number and updating parameter $t$ of OVMD were determined by decomposing the mode center frequency and residual evaluation index.

2.2 Random Forest
Random forest (RF) is an integrated learning algorithm based on bagging algorithm thought proposed by Breiman (2001). RF can effectively analyze high-dimensional nonlinear data, and has good generalization ability and predictive performance (Hidayat and Astsauri 2021). RF has strong generalization ability and is not easy to overfit, the model is relatively simple, and the classification effect is good. RF uses bootstrap resampling to randomly extract data from the original sample to construct multiple samples, and randomly split nodes to construct multiple decision trees for each sample. The final prediction result is determined by the average of the output values of multiple decision trees in the random forest. The training set of each decision tree in RF is composed of $N$ samples randomly selected from the entire training sample set. This process is called self-service resampling, and the resulting sample set is called bootstrap set. The samples in the tree training set are called out-of-bag (OOB) data, and feature importance can be calculated from these OOB data.

RF has the characteristics of calculating the importance of a single feature variable, so this paper uses RF to perform feature selection on sample data. The detailed calculation steps of the importance of feature $x_i$ are as follows:

1. Initialize RF, set $k=1$, build a decision tree $T_k$ on the $k$th bootstrap sample set, and record the OOB data set corresponding to the bootstrap set as $L_k^{\text{ooob}}$;

2. Classify the OOB data $L_k^{\text{ooob}}$ based on $T_k$, and count the number of correct classifications, denoted as $R_k^{\text{ooob}}$;

3. Randomly change the value of the feature $x_i$ in $L_k^{\text{ooob}}$, and record the perturbed OOB data set as $L_{k,i}^{\text{ooob}}$, then use $T_k$ to classify it, and count the number of
correct classifications, and record it as $R_{k,i}^{\text{oob}}$.

(4) For $k = 2, 3, \cdots, K$, repeat steps (1) ~ (3), $K$ is the number of decision trees;

(5) Calculate the importance of feature $x_i$, the formula is as follows:

$$impot_i = \frac{1}{K} \sum_{k=1}^{K} (R_{k,i}^{\text{oob}} - R_{k,i}^{\text{oob}_k})$$

where $impot_i$ is the importance of the feature $x_i$, $K$ is the number of decision trees,
$R_{k,i}^{\text{oob}}$ and $R_{k,i}^{\text{oob}_k}$ are the number of correct statistical classifications before and after the feature value changes.

Sort $P$ in descending order to get the feature importance ranking, the higher the ranking, the higher the importance of the feature.

2.3 Gated Recurrent Unit Networks

Gated Recurrent Unit (GRU) was proposed by Cho et al. (2014). GRU model is an improved LSTM network algorithm. The basic unit of the LSTM hidden layer is a memory block structure composed of input gate, forget gate and output gate. Through the control of these three gates, update the history information. The Forget Gate controls the information retained to the current last moment; the input gate is responsible for regulating the current input; the output gate is responsible for regulating the current output. The function of GRU is the same as LSTM, but the structure of GRU is simpler than LSTM. GRU merges the forget gate and input gate of LSTM into the update gate, and the output gate becomes the reset gate. The update gate is used to choose to retain the state information at the previous moment. The reset gate mainly determines how much past information needs to be forgotten. This makes the structure of the GRU
network much simpler than that of the LSTM network, and does not add any additional internal state. Through this structure, the input information of the previous module is forgotten and added, and finally the new state and output of the current module are obtained through calculation. Continuously adjust the weight values in these structures through learning, until the output meets the requirements, and finally achieve the purpose of screening information, controlling data flow, determining whether information is transmitted, and simulating the reasoning process. The calculation process of the GRU network is as follows:

1. Calculate reset gate \( r_i \):
   \[
   r_i = \sigma(W_r \cdot X_i + Q_r \cdot h_{i-1} + b_r)
   \]
   \[
   (4)
   \]

2. Calculate the proposed hidden state \( \hat{h}_i \):
   \[
   \hat{h}_i = \tanh(W_h \cdot X_i + Q_h \cdot r_i \cdot h_{i-1} + b_h)
   \]
   \[
   (5)
   \]

3. Calculate update gate \( u_i \):
   \[
   u_i = \sigma(W_u \cdot X_i + Q_u \cdot h_{i-1} + b_u)
   \]
   \[
   (7)
   \]

4. Calculate the hidden state \( h_i \):
   \[
   h_i = u_i \cdot \hat{h}_i + (1 - u_i) \cdot h_{i-1}
   \]
   \[
   (8)
   \]

5. Calculate predicted value \( y_i \):
   \[
   y_i = h_i
   \]
   \[
   (9)
   \]

where in the above equations from Eq. (4) to Eq. (9), \( W_r \), \( W_h \), \( W_u \), \( Q_r \), \( Q_h \) and \( Q_u \) denote the weight matrixes for reset gate, the proposed hidden state, update gate and the hidden state, respectively while \( b_r \), \( b_h \) and \( b_u \) denote the corresponding bias.
vectors. \( h_{t-1} \) represent the previous hidden state. \( \sigma(x) \) represent the Sigmoid activation function.

![Fig. 1. The model structure of GRU](image)

### 2.4 Improved Whale Optimization Algorithm

#### 2.4.1 Standard whale optimization algorithm

Whale optimization algorithm (WOA) is a meta-heuristic optimization algorithm proposed by Mirjalili and Lewis (2016). WOA achieves the optimization goal by simulating the unique bubble net foraging strategy of whales. The mathematical model is established by imitating its foraging. In order to reduce the control variables in the algorithm, the velocity vector is deleted and the position vector is retained. In this way, the search capability of WOA is enhanced.

In the whale optimization algorithm, there are three strategies for whale position updating: encircling prey, bubble-net attacking method and search for prey. Assuming that the population number of whales is \( N \), the position of whale in the \( D \)-dimensional search space can be expressed as \( (X_{ij}) \), where \( i = 1, 2, \ldots, N \). \( j = 1, 2, \ldots, D \). The three strategies are described as follows:

1. **Encircling prey**
\[
\begin{align*}
X(t+1) &= X^*(t) - A \cdot D \\
D &= |C \cdot X^*(t) - X(t)|
\end{align*}
\] (10)

where \( t \) is the number of current iterations, \( X^*(t) \) is the current best whale position vector, \( X(t) \) is the current position vector of the whale, \( A \) and \( C \) are the coefficient vectors, and they are defined as follows:

\[
\begin{align*}
A &= 2a \cdot r - a \\
C &= 2r
\end{align*}
\] (11)

where \( r \) is a random number in the interval \([0,1]\), and \( a = 2 - \frac{2t}{T_{\text{max}}} \cdot T_{\text{max}} \) is the maximum number of iterations.

(2) Bubble-net attacking method

Humpback whales hunt by spiral motion, and a mathematical model of spiral position updating can be built. The distance between the whale group and the prey is calculated, and the spiral swimming behavior of the humpback whale is imitated to establish a mathematical model of the spiral movement. The expression is as follows:

\[
\begin{align*}
X(t+1) &= X^*(t) + D_p \cdot e^{bl} \cos(2\pi l) \\
D_p &= |X^*(t) - X(t)|
\end{align*}
\] (12)

where \( D_p \) is the distance between prey and whale, \( b \) is the logarithmic spiral coefficient, and \( l \) is a random number in the interval of \([-1,1]\).

It is worth noting that while the whale swims toward its prey in a spiral shape, it also shrinks its enclosure. Therefore, this behavior is manifested in the model as a 50% probability to choose the shrinking envelopment mechanism and a 50% probability to choose the spiral model to update the position of the whale. The mathematical model is as follows:
\[ X(t+1) = \begin{cases} X^*(t) - A \cdot D, & p < 0.5 \\ X^*(t) + D_p \cdot e^{bl} \cos(2\pi t), & p \geq 0.5 \end{cases} \]  \tag{13}

where \( p \) is a random number in \([0,1]\).

(3) Search for prey

Whales use \(|A|\) to control whether they are in the search for prey or in the encircling prey. When \(|A| >1\), the whales are unable to get useful information to the prey and they need to keep trying to find their prey by using random clues of information. The mathematical model for the search of the prey is as follows:

\[
\begin{cases} 
X(t+1) = X_{rand}(t) - A \cdot D \\
D = |C \cdot X_{rand}(t) - X(t)|
\end{cases} \tag{14}
\]

where \( X_{rand}(t) \) is a random position vector of the whale population.

2.4.2 Improvement of whale optimization algorithm (IWOA)

The standard whale optimization algorithm has shortcomings such as slow convergence speed, low convergence accuracy, and easy to fall into local optimum, so this study improves it. The three improvement strategies are as follows:

(1) Chaos population initialization

The traditional WOA algorithm solves the optimization problem by randomly generating the initial population, which makes the initial population unevenly distributed and reduces the optimization speed. The chaotic motion has the characteristics of randomness, regularity and ergodicity (Paul et al. 2020). Using these advantages can generate an initial population with better diversity and improve the global search ability.

The Tent mapping expression is as follows:
\[ x_{r+1} = \begin{cases} \frac{x_r}{u}, & 0 \leq x_r \leq u \\ \frac{1-x_r}{1-u}, & u < x_r \leq 1 \end{cases} \quad (15) \]

When \( u = 0.5 \), the generated sequence is uniformly distributed, it is not sensitive to the changes of different parameters, and has an approximately uniform distribution density. The most typical Tent mapping is expressed as follows:

\[ x_{r+1} = \begin{cases} 2x_r, & 0 \leq x_r \leq 0.5 \\ 2(1-x_r), & 0.5 < x_r \leq 1 \end{cases} \quad (16) \]

The Tent map is used to generate the initial population for the WOA algorithm in this study.

(2) Nonlinear convergence factor and inertia weight

The function of parameter A in the standard WOA is to adjust the global exploration and local development capabilities of the algorithm. The convergence factor \( a \) decreases linearly during the iteration process, which easily makes the search speed of the algorithm unable to adapt to the actual situation. The global search and local search capabilities of WOA are affected by the convergence factor \( a \). In this paper, the convergence factor \( a \) is changed to nonlinear, and the formula is shown in eq. (17). As the number of iterations increases, the change rate of the convergence factor is firstly slower and then faster. At the beginning, the attenuation rate of \( a \) is slower, so that the algorithm can optimize globally. In the later stage of optimization, the attenuation of \( a \) becomes faster, which corresponds to the precise local optimization of the algorithm. The different attenuation levels of \( a \) correspond to the different search performances of WOA, which effectively improves the development capability of WOA.
where $t$ is the current number of iterations, and $\text{max} \_ \text{iter}$ is the maximum number of iterations.

In the iterative process, the standard WOA algorithm does not take into account the possible differences in the guiding force of the prey to guide the whale to update its position. Therefore, combined with the idea of inertial weight guiding the population optimization in the PSO algorithm, the WOA algorithm introduces adaptive parameters as the inertial weight in the position update process in Eq. (12) ~ (14). Dong et al. (2021) and Yan et al. (2020) proposed a nonlinear decreasing inertia weight and both improved the search ability of WOA. The large initial weight will lead to a wider search range, while the latter option value is small, so that the search can be limited to a local area. When the whale is close to the target, a smaller weight is used to change the position of the best whale. This facilitates fast global search and fine local search. This study proposes a new nonlinear decreasing inertia weight to improve WOA. The inertial weight is as follows:

$$\omega(t) = \frac{1}{2} \cdot \cos \left( \frac{t}{\text{max} \_ \text{iter}} \cdot \frac{\pi}{2} \right) + \frac{2}{5} \quad (18)$$

The specific improvement for Eq. (12) ~ (14) of the WOA are as follows:

$$\begin{align*}
X(t + 1) &= \omega \cdot X^*(t) - A \cdot D \\
X(t + 1) &= (1 - \omega) \cdot X^*(t) + D_p \cdot e^{\beta t} \cos (2\pi t) \\
X(t + 1) &= \omega \cdot X_{\text{rand}}(t) - A \cdot D
\end{align*} \quad (19)$$

where $t$ is the current number of iterations, and $\text{max} \_ \text{iter}$ is the maximum number of iterations. With the change of $t$, the inertia weight gradually decreases, realizing the
dynamic change of the weight, so that the global search and local development of WOA are better balanced.

(3) Gauss perturbation

Gaussian disturbance is a random disturbance with Gaussian distribution, in which the Gaussian distribution is centered on a random variable. In each iteration, Gaussian perturbation is performed on the current global optimal individual, which can increase the probability of obtaining a better solution. Gaussian mutation originated from Gaussian normal distribution, which has been proved in many algorithm improvements, such as grey wolf optimizer algorithm (Peng and Zhou 2019), artificial bee colony algorithm (Xiao et al. 2021), ant colony optimization algorithm (Tuani et al. 2020), PSO (Wang et al. 2020b), and also used by researchers in WOA (Luo et al. 2019). It is more likely to produce new offspring near the primary parent, which will cause the current optimal individual to update its position in small steps. Gaussian mutation can make the WOA algorithm converge faster and enhance the search ability of the algorithm (Luo et al. 2019). The probability density function of the Gaussian distribution is expressed as follows:

\[
f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]  

(20)

It can be seen from Eq. (20) that the larger \( \mu \) is, the smaller the probability value of position \( x \) is, indicating that the curve is flatter and the probability distribution is more dispersed. Conversely, the smaller \( \mu \) is, the steeper the curve is and the more concentrated the probability distribution is. According to the characteristics of the distribution curve, the \( \mu \) value in this paper is 0; that is, the peak of the distribution
curve is located at 0. In the later iterations of the algorithm, when Gaussian perturbation is applied to the current optimal individual, the nearby area can be explored and the search ability of WOA can be improved. If $\sigma$ value is too large, the variation amplitude will be too large, and the implementation of Gaussian perturbation will easily lead to the deviation of the algorithm. It is difficult to find the global optimal advantage. If the value of $\sigma$ is too small, the disturbance to the search individual is insufficient, unable to help the individual successfully get rid of the shackles of the local extreme value, and unable to achieve the purpose of increasing the diversity of the search population. Therefore, the value of $\sigma$ in this paper is 1, which can get the best effect.

When $\mu = 0$ and $\sigma = 1$, the density function is as follows:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$  \hspace{1cm} (21)

The Gaussian perturbation formula for the optimal individual in WOA is as follows:

$$x'(t) = x^*(t) + x^*(t) \cdot normrnd(0,1)$$  \hspace{1cm} (22)

where $normrnd(0,1)$ represents a random number that generates a Gaussian distribution with a mean of 0 and a variance of 1.

### 2.5 Training procedures of GRU using IWOA

The basic idea of the training procedures of GRU using IWOA is as follows: using the IWOA algorithm to optimize the learning rate and the number of hidden layer units of the GRU model, taking the GRU training error as the individual fitness value, and selecting the optimal GRU model parameters. The specific process is as follows:
**Step 1:** Collect flood flow data, preprocess the original data, eliminate outliers, and use cubic spline interpolation to fill in missing values.

**Step 2:** Use RF to measure the importance of the decomposed data. Data with high importance were selected as input variables of GRU. The entire data set is normalized to \([0,1]\) and divided into training set and test set.

**Step 3:** Initialize GRU model parameters, including batch size, model training times, and dropout value.

**Step 4:** Initialize IWOA, set the number of populations \(N\), the maximum number of iterations \(max\_iter\), the search range. The initial population of the model is generated by chaos mapping in the feasible region.

**Step 5:** The fitness value was calculated and sorted to find the global optimal solution. According to Eq. (21), Gaussian perturbation is performed on the global optimal solution to improve the position of the optimal solution.

**Step 6:** According to the values of \(A\) and \(P\), different formulas are used to update the position of the next generation of whales.

**Step 7:** \(t = t+1\). If \(t < max\_iter\), go to Step 5, otherwise go to Step 8;

**Step 8:** Assign the current parameters to the GRU model, and make predictions based on the obtained parameters and the test data set.
3 Case study

3.1 Study area

The Minjiang River Basin in China is chosen as the study area. The Minjiang River is an important tributary of the upper reaches of the Yangtze River, located in the middle of Sichuan Province. The source of the Minjiang River is divided into east and west
sources, both of which are located in Songpan County. It merges into the Yangtze River in Yibin, with a total length of 735 kilometers. The Minjiang River has a total drop of 3,560 meters and a drainage area of 135,881 square kilometers, of which Sichuan Province accounts for 126,280 square kilometers. The upper reaches belong to a plateau climate zone, and the middle and lower reaches belong to a subtropical climate zone. The temperature and rainfall increase with the terrain from north to south. The average annual flow of the Minjiang River Basin is 2850 m$^3$/s, and the annual runoff is 90 billion cubic meters. The runoff mainly comes from rainfall and alpine snowmelt. May to October is the wet season, the water volume accounts for about 80% of the year, and the rest is the dry season. In July and August each year, the maximum flow in the Minjiang River Basin can reach 20000 m$^3$/s, causing great damage to lives and properties along the river basin, especially in the lower Chengdu Plain.
3.2 Data description and preprocessing

In order to show the accuracy and comprehensiveness of the model proposed in this study for flood forecasting, this paper selects three stations in the Minjiang River Basin for verification. The Zhenjiangguan hydrological station in the upper reaches of Minjiang River (No. 60601100, 103° 44’E, 32° 17’N), the Dujiangyan hydrological station in the middle reaches (No. 60602480, 103° 36’E, 31° 01’N) and the Gaochang hydrological station in the lower reaches (No. 60603300, 104° 25’E, 28° 48’N) were selected respectively. The daily runoff data of hydrological stations from May 2009 to April 2020 were collected respectively at three hydrological stations. Each station provided 4020 data, and a total of 12060 discharge data of Minjiang River were
collected.

Among the 4020 flood flow data of each station, the first 3220 observation data are used as the training set, and the remaining 800 observation data are used to test the model. The statistical information of the whole data set, training data set and test data set is shown in Table 1. The statistical information includes the overall maximum value (Max.), minimum value (Min.), median value (Mid.), mean value (Me.), standard deviation (CV) and skewness (CS) of the data set.

Table 1 Statistics for the entire data set, training data set, and test data set for the three sites.

| Station       | Data Set | Data Length | Max.  | Min.   | Mid.   | Me.    | CV    | CS    |
|---------------|----------|-------------|-------|--------|--------|--------|-------|-------|
| Gaochang      | All      | 4020        | 20284 | 549.08 | 1751.46| 2557.69| 2038.92 | 2.54  |
|               | Training | 3220        | 20284 | 550.77 | 1707.71| 2479.08| 1907.04| 2.51  |
|               | Testing  | 800         | 15738.46 | 549.08 | 1884.58| 2874.11| 2475.22| 2.37  |
| Dujiangyan    | All      | 4020        | 2011.2 | 133.46 | 421.38 | 487.27 | 269.58 | 1.30  |
|               | Training | 3220        | 1862.08 | 133.46 | 424.43 | 496.76 | 275.01 | 1.26  |
|               | Testing  | 800         | 2011.2 | 151.77 | 401.16 | 449.01 | 242.81 | 1.50  |
| Zhenjiangguan | All      | 4020        | 391.79 | 10.74  | 42.63  | 54.70  | 40.12  | 1.88  |
|               | Training | 3220        | 361.32 | 10.74  | 41.67  | 53.44  | 38.90  | 1.65  |
|               | Testing  | 800         | 391.79 | 18.13  | 46.48  | 59.77  | 44.34  | 2.43  |

OVMD is used to decompose the data of three hydrological stations. The center frequency of each component after OVMD is decomposed under different K values is shown in Table 2. For high-field sites, when $K = 7$, similar frequency components begin
to appear, that is, excessive decomposition occurs. Therefore, the total number of patterns at the Gaochang site $K_{GC} = 6$, the total number of patterns at the Zhenjiangguan site $K_{ZJG} = 7$, and the total number of patterns at the Dujiangyan site $K_{DJY} = 6$. In general, the update parameter $s$ is set to 0, so that the Lagrangian multiplier remains at 0.

**Table 2** Center frequencies corresponding to different $K$ values at the Gaochang station

| Station  | Modal number | Center frequency |
|----------|--------------|------------------|
| Gaochang | $K=2$        | $7.35 \times 10^{-4}$ 0.25 |
|          | $K=3$        | $7.35 \times 10^{-4}$ 0.16 0.33 |
|          | $K=4$        | $7.35 \times 10^{-4}$ 0.12 0.25 0.37 |
|          | $K=5$        | $7.35 \times 10^{-4}$ 0.10 0.20 0.30 0.40 |
|          | $K=6$        | $7.35 \times 10^{-4}$ 0.08 0.17 0.25 0.33 0.42 |
|          | $K=7$        | $7.35 \times 10^{-4}$ 0.07 0.14 0.22 0.28 0.41 0.43 |
| Zhenjiang | $K=2$        | $7.26 \times 10^{-4}$ 0.24 |
| ngguan   | $K=3$        | $7.26 \times 10^{-4}$ 0.16 0.33 |
|          | $K=4$        | $7.26 \times 10^{-4}$ 0.12 0.24 0.37 |
|          | $K=5$        | $7.26 \times 10^{-4}$ 0.09 0.20 0.30 0.40 |
|          | $K=6$        | $7.26 \times 10^{-4}$ 0.08 0.16 0.25 0.33 0.42 |
|          | $K=7$        | $7.26 \times 10^{-4}$ 0.07 0.14 0.21 0.29 0.36 0.43 |
|          | $K=8$        | $7.26 \times 10^{-4}$ 0.06 0.13 0.19 0.26 0.33 0.40 0.44 |
| Dujiangyan | $K=2$       | $9.94 \times 10^{-4}$ 0.27 |
| yan      | $K=3$        | $9.94 \times 10^{-4}$ 0.16 0.33 |
|          | $K=4$        | $9.94 \times 10^{-4}$ 0.12 0.25 0.37 |
|          | $K=5$        | $9.94 \times 10^{-4}$ 0.10 0.20 0.30 0.40 |
|          | $K=6$        | $9.94 \times 10^{-4}$ 0.08 0.17 0.25 0.32 0.42 |
|          | $K=7$        | $9.94 \times 10^{-4}$ 0.07 0.15 0.22 0.29 0.39 0.42 |

The changes of the VMD decomposition residual index REI of the three sites are
shown in Fig.3. The smallest possible residual error in signal decomposition can improve the prediction accuracy. When the REI of the high field site is the smallest, \( \tau=0.95 \); when the REI of the Zhenjiangguan site is the smallest, \( \tau=0.57 \); when the REI of the Dujiangyan site is the smallest, \( \tau=0.96 \).

Fig. 4. The REI values of data from three hydrological stations

After determining the optimal K and s values of the time series, the time series of the three stations can be decomposed by OVMD. As shown in Fig. 5 and Fig. 7, the time series of the Gaochang site and the Dujiangyan site are decomposed into 6 sub-series through OVMD. From Figs. 5-7, it can be found that the frequency of the subsequences after VMD decomposition is getting larger and larger, indicating that the impact on the final prediction result is small.
Fig. 5. The original time series of the Gaochang site and the sub-series obtained after the decomposition using the OVMD method.
Fig. 6. The original time series of the Dujiangyan site and the sub-series obtained after the decomposition using the OVMD method.
Fig. 7. The original time series of the Zhenjiangguan site and the sub-series obtained after the decomposition using the OVMD method.

3.3 Determination of the input variables

Before using the hybrid model to predict flood volume, the input matrix needs to
be determined. Choosing the appropriate input matrix is very important for the entire model. The original data of the three hydrological stations are decomposed by OVMD to obtain different IMFs. The RF method is used to calculate the importance of the antecedent time series of different IMFs from lag 0 to 20. Figs. 8 and 9 show the importance of IMF lags between 0 and 20 at different sites. It can be seen from the figure that for different lags, the importance calculated by RF is different. According to the importance of the time series with different time lags for each IMF, when the cumulative importance of RF reaches 85%, the observations of all time lags before this time are selected as the input of the model. As can be seen from the figure, generally select the first 10 to 14 previous data. Using RF to calculate and select features in the original sequence can improve the accuracy of prediction.

Fig. 8. The importance of RF of the decomposition modes of the Gaochang site
and the Dujiangyan site

![Graphs showing decomposition modes](image)

Fig. 9. The importance of RF of the decomposition modes of the Zhenjiangguan site

### 3.4 Parameter settings

In order to highlight the effectiveness of the proposed OVMD-RF-IWOA-GRU model, this study constructed ten models for comprehensive comparison, including ELM, LSTM, GRU, PSO-GRU, WOA-GRU, PCA-WOA-GRU, RF-WOA-GRU, VMD-RF-WOA-GRU, OVMD-RF-WOA-GRU and OVMD-RF-IWOA-GRU. In order to improve the performance of the prediction model, it is necessary to determine the parameters of the algorithm and the model. The settings of OVMD-RF-IWOA-GRU proposed in this study are as follows. For OVMD, for the three site $K$ methods 6, 6, and 7, $\tau$ are 0.95, 0.57, and 0.96, respectively. The confidence band of RF importance is set to 85%. For the IWOA algorithm, set the number of search agents to 10 and the maximum number of iterations to 10.

The optimizer chosen in this paper for GRU training is the Adam optimization
function. The learning rate of GRU is set to 0.002. The number of hidden layer nodes is set to 100. The batch size is set to 36. In other words, the model will update the parameters after processing 36 sets of data. This operation makes the direction of gradient descent in neural network learning more accurate, and the resulting training fluctuations reach a relatively stable state (Huang et al. 2021). In order to prevent the model from overfitting, the dropout parameter is set to 0.2, which makes 20% of the random neuron nodes in each hidden layer fail. For using the optimization algorithm to optimize the learning rate and the number of hidden layer nodes of the GRU model, it is necessary to set the range of optimized parameters, where the range of the learning rate is set to $[0.001, 0.02]$, and the number of hidden layer nodes is set to $[1,500]$.

### 3.5 Performance metrics

This paper selects four commonly used indicators to verify the effect of the proposed combination forecasting model. Including: root mean square error (RMSE), correlation coefficient (R), mean absolute error (MAE) and mean absolute percentage error (MAPE). The calculation formulas of the three indicators are as follows:

\[
RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (OB_t - PR_t)^2} \tag{23}
\]

\[
R = \frac{\sum_{t=1}^{N} (PR_t - \overline{PR})(OB_t - \overline{OB})}{\sqrt{\sum_{t=1}^{N} (PR_t - \overline{PR})^2} \times \sqrt{\sum_{t=1}^{N} (OB_t - \overline{OB})^2}} \tag{24}
\]

\[
MAE = \frac{1}{N} \sum_{t=1}^{N} |OB_t - PR_t| \tag{25}
\]

\[
MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{PR_t - OB_t}{OB_t} \right| \tag{26}
\]
where \( OB_t \) and \( PR_t \) are observed and predicted data at time \( t \); \( \overline{OB}_t \) and \( \overline{PR}_t \) represent the average of observed and predicted data; \( N \) is the length of the time series.

For the three indicators RMSE, MAE and MAPE, the closer the value of the indicator is to 0, the closer the predicted value is to the observed value. The closer the value of \( R \) is to 1, the closer the predicted value is to the observed value.

4 Comparisons results

This section compares the method in this paper with a single prediction model and a combination of different prediction models. This article uses a total of 10 prediction models to predict and compare the water volume of three stations, and these models use the same data set. The single and combined models compared with the models in this article include ELM, LSTM, GRU, PSO-GRU, WOA-GRU, PCA-WOA-GRU, RF-WOA-GRU, OVMD-WOA-GRU, OVMD-RF-WOA-GRU and OVMD-RF-IWOA-GRU.

4.1 Comparison of predicted results of the Gaochang site

This section first compares and analyzes the results predicted by the ten models. In order to show the error metric more intuitively, the RMSE, \( R \), MAE, and MAPE values of ten models and three data sets are illustrated using horizontal histograms and radar charts, respectively. Because Gaochang, Dujiangyan and Zhenjiangguan are located in the lower, middle and upper reaches of the Minjiang River, respectively. Therefore, it can be found from the table that the data set of the Gaochang site has the largest error, and then the data set of the Dujiangyan site, and the data set of the
Zhenjiangguan site is the smallest. The prediction results of the ten models are shown in Table 3.

Table 3 Comparison of Evaluation Indexes of ten prediction models for the Gaochang site

| Model          | RMSE   | R      | MAE   | MAPE  |
|----------------|--------|--------|-------|-------|
| ELM            | 1277.4 | 0.8608 | 582.13| 16.77 |
| LSTM           | 1193.7 | 0.8777 | 538.82| 16.09 |
| GRU            | 1141.9 | 0.8889 | 518.74| 14.99 |
| PSO-GRU        | 1078.8 | 0.9004 | 505.94| 15.40 |
| WOA-GRU        | 1021.7 | 0.9117 | 482.79| 14.39 |
| PCA-WOA-GRU    | 974.8  | 0.9200 | 445.95| 13.13 |
| RF-WOA-GRU     | 928.1  | 0.9285 | 428.10| 12.95 |
| VMD-RF-WOA-GRU | 480.8  | 0.9830 | 220.71| 6.67  |
| OVMD-RF-WOA-GRU| 442.0  | 0.9853 | 202.42| 5.98  |
| OVMD-RF-IWOA-GRU| 379.1  | 0.9885 | 199.94| 6.50  |

Table 3 records the RMSE, R, MAE, and MAPE results of all comparison models at the Gaochang hydrological observatory. The horizontal histogram and radar chart in Fig. 10 intuitively show the size of the performance indicators of the ten models. It can be concluded from the four indicators of three single prediction models that the results of each indicator of the GRU model are better than the traditional ELM model and LSTM. The GRU model has good performance in flood forecasting. By comparing WOA-GRU with PSO-GRU and GRU, it can be concluded that the WOA algorithm is
effective for the optimization of the GRU model and that the WOA algorithm is more suitable for the GRU model than the PSO. Based on the WOA-GRU model, the feature selection method is used to select the input data to improve the prediction accuracy of the model. It can be seen from the table that the model prediction performance has been improved, but the RF method is better than the traditional PCA method. Based on the RF-WOA-GRU model, using the modal decomposition method for the original input, it can be seen that various indicators have been greatly improved. After adding the VMD and OVMD decomposition algorithms, the prediction accuracy of the model has been significantly improved. Compared with VMD, OVMD is more suitable for the model of this paper. Compared with the RF-WOA-GRU model, the OVMD-RF-WOA-GRU model has a 50.7% increase in RMSE index, a 53.8% increase in MAPE index, and an improvement in both MAE and R indexes.

![Graph showing RMSE, R, MAE, and MAPE for ten models to predict the Gaochang sites](image)

**Fig. 10.** RMSE, R, MAE, and MAPE for ten models to predict the Gaochang sites

This article also improves the basic WOA algorithm to increase its optimization
capabilities. After using the improved WOA to replace the traditional WOA, all indicators of the model have been improved. The following conclusions can be drawn from this article: The improvement strategy proposed by WOA is effective. In order to show the prediction results more intuitively and further verify the effectiveness of the proposed model, the proposed OVMD-RF-IWOA-GRU model is compared with the ELM, GRU, WOA-GRU and RF-WOA-GRU models. The prediction results and the observation results of the four modes are compared in the line chart at the top of Fig. 11. The middle of Fig. 11. shows the absolute errors of the five models. Finally, the scatter plots and error density plots of the predicted and observed values are given at the bottom of the graph. It can be found from the figure that the absolute error of the proposed model is significantly smaller than the other three models. The OVMD-RF-IWOA-GRU model predicted and observed value scatter plots have the smallest wide deviation, and the 95% prediction interval is also the narrowest. The final error density graph also clearly shows that the curve of the proposed model is better than other models. In short, it can be concluded from the above analysis that the proposed OVMD-RF-IWOA-GRU model can better predict floods.
Fig. 11. The prediction results of ELM, GRU, WOA-GRU, RF-WOA-GRU and OVMD-RF-IWOA-GRU models at the Gaochang site

### 4.2 Comparison of predicted results of the Dujiangyan site and the Zhenjiangguan site

This section shows the prediction results of the proposed OVMD-RF-IWOA-GRU model at the Dujiangyan site and the Zhenjiangguan site. The performance of the proposed model is also compared with other nine comparative models. Table 5 and Table 6 show the RMSE, $R$, MAE, and MAPE values of the two sites, respectively. The corresponding horizontal histogram and radar chart are shown in Fig. 12 and Fig. 14, respectively. Fig. 13 and Fig. 15 show the comparison between the OVMD-RF-IWOA-GRU model and the ELM, GRU, WOA-GRU, and RF-WOA-GRU models for
the two sites, respectively. From Table 4-5 and Fig. 12-15, the same conclusions as in section 4.1 can be drawn.

**Table 4** Comparison of Evaluation Indexes of ten prediction models for the Dujiangyan site

| Model               | RMSE  | R     | MAE   | MAPE  |
|---------------------|-------|-------|-------|-------|
| ELM                 | 108.18| 0.8969| 56.490| 11.820|
| LSTM                | 94.66 | 0.9212| 47.503| 9.471 |
| GRU                 | 91.24 | 0.9268| 46.970| 9.486 |
| PSO-GRU             | 86.25 | 0.9371| 45.117| 9.566 |
| WOA-GRU             | 84.00 | 0.9388| 43.723| 8.992 |
| PCA-WOA-GRU         | 80.18 | 0.9507| 45.512| 9.810 |
| RF-WOA-GRU          | 74.49 | 0.9522| 39.043| 8.090 |
| OVMD-WOA-GRU        | 45.93 | 0.9826| 24.149| 5.051 |
| OVMD-RF-WOA-GRU     | 41.10 | 0.9862| 21.391| 4.438 |
| OVMD-RF-IWOA-GRU    | 36.59 | 0.9890| 19.354| 4.066 |
Fig. 12. RMSE, R, MAE, and MAPE for ten models to predict the Dujiangyan sites.

Fig. 13. The prediction results of ELM, GRU, WOA-GRU, RF-WOA-GRU and OVMD-RF-IWOA-GRU models at the Dujiangyan site.
Table 5 Comparison of Evaluation Indexes of ten prediction Models for the Zhenjiangguan site

| Model             | RMSE  | R      | MAE   | MAPE  |
|-------------------|-------|--------|-------|-------|
| ELM               | 13.31 | 0.9649 | 5.347 | 7.773 |
| LSTM              | 12.95 | 0.9566 | 5.044 | 6.454 |
| GRU               | 12.33 | 0.9616 | 4.823 | 6.333 |
| PSO-GRU           | 10.89 | 0.9697 | 4.325 | 5.771 |
| WOA-GRU           | 10.16 | 0.9737 | 4.079 | 5.426 |
| PCA-WOA-GRU       | 9.92  | 0.9767 | 4.387 | 7.183 |
| RF-WOA-GRU        | 9.39  | 0.9778 | 3.817 | 5.202 |
| OVMD-WOA-GRU      | 4.91  | 0.9943 | 1.983 | 2.648 |
| OVMD-RF-WOA-GRU   | 4.62  | 0.9949 | 1.845 | 2.492 |
| OVMD-RF-IWOA-GRU  | 4.29  | 0.9956 | 1.741 | 2.337 |
Fig. 14. RMSE, R, MAE, and MAPE for ten models to predict the Zhenjiangguan sites

Fig. 15. The prediction results of ELM, GRU, WOA-GRU, RF-WOA-GRU and OVMD-RF-IWOA-GRU models at the Zhenjiangguan site

5 Conclusions

Accurately predicting the arrival of floods can reduce the loss of life and property caused by flood disasters. In this study, a deep learning model based on OVMD, RF, WOA and GRU (OVMD-RF-IWOA-GRU) was constructed to predict the flood time series. In order to prove the excellence the OVMD-RF-IWOA-GRU model, a total of ten models were constructed, namely ELM, LSTM, GRU, PSO-GRU, WOA-GRU, PCA-WOA-GRU, RF-WOA-GRU, OVMD-WOA-GRU, OVMD-RF-WOA-GRU and OVMD-RF-IWOA-GRU. The conclusions obtained are as follows:
(1) The OVMD-RF-IWOA-GRU model obtains the smallest MAE, MAPE, RMSE and the largest R, which proves that the predictive ability of the OVMD-RF-IWOA-GRU model is superior than the comparison model.

(2) The GRU model is better than the ELM and LSTM models, which proves the effectiveness of the GRU network in processing hydrological time series data.

(3) The GRU models optimized by the PSO algorithm and the WOA algorithm are better than the separate GRU model, and WOA-GRU is better than the PSO-GRU, which proves that the optimization algorithm can improve the predictive ability of the GRU model. The WOA algorithm is more suitable for GRU than the PSO algorithm.

(4) PCA and RF are used to perform feature selection on the input data to improve the predictive ability of the model, and the RF is more excellent.

(5) On the basis of the standard WOA, three improvement strategies are added to improve the search ability of WOA. The OVMD-RF-IWOA-GRU model is better than the OVMD-RF-WOA-GRU model, which proves the effectiveness of the improved strategy added to WOA. The OVMD-RF-IWOA-GRU model is better than the OVMD-RF-WOA-GRU model, which proves the effectiveness of the improvement to the WOA algorithm.

In general, the proposed OVMD-RF-IWOA-GRU model can be regarded as a competitive technology in improving the accuracy of flood prediction. Since flood forecasting is affected by many factors and sudden factors, the limited training data used in this article leads to limited model prediction accuracy, which needs to be combined with more relevant influencing factors for research and processing.
**Ethical Approval**

Relevant research content in this study was in accordance with the ethical standards of the institutional and national research committee.

**Consent to Participate**

All of the authors consent to participate in the relevant research content in this paper.

**Consent for Publish**

The authors give their full consent for the publication of this manuscript.

**Authors Contributions**

**Chunlei Ji:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Validation, Visualization. **Tian Peng:** Data curation, Supervision, Validation, Writing – review & editing, Visualization. **Chu Zhang:** Conceptualization, Software, Supervision. **Lei Hua:** Methodology. **Wei Sun:** Visualization.

**Funding**

This work is supported by the Natural Science Foundation of Jiangsu Province (No. BK20191052), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 19KJB470012), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 19KJB480007), the Construction
system of Jiangsu Province of China (No.20200037).

Conflicts of Interest

The authors declare that they have no conflict of interest.

Availability of data and materials

The data and code that support the study are available from the corresponding author upon reasonable request.

References

Akbari R, Hessami-Kermani M-R and Shojaee S (2020) Flood Routing: Improving Outflow Using a New Non-linear Muskingum Model with Four Variable Parameters Coupled with PSO-GA Algorithm. Water Resources Management 34(10): 3291-3316.

Allahbakhshian-Farsani P, Vafakhah M, Khosravi-Farsani H and Hertig E (2020) Regional Flood Frequency Analysis Through Some Machine Learning Models in Semi-arid Regions. Water Resources Management 34(9): 2887-2909.

Ayzel G and Heistermann M (2021) The effect of calibration data length on the performance of a conceptual hydrological model versus LSTM and GRU: A case study for six basins from the CAMELS dataset. Computers & Geosciences 149: 104708.

Breiman L (2001) Random Forests. Machine Learning 45(1).

Bui D T, Ngo P-T T, Pham T D, Jaafari A, Minh N Q, Hoa P V and Samui P (2019) A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. CATENA 179: 184-196.

Chapi K, Singh V P, Shirzadi A, Shahabi H, Bui D T, Pham B T and Khosravi K (2017) A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environmental Modelling & Software 95: 229-245.

Cho K, Merrienboer B V, Gulcehrre C, Ba Hdana D, Bougares F, Schwenk H and Bengio Y (2014) Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. Computer Science.

Deo R C and Şahin M (2015) Application of the Artificial Neural Network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydrometeorological parameters and climate indices in eastern Australia. Atmospheric Research 161-162: 65-81.

Dong W, Zhang S, Jiang A, Jiang W, Zhang L and Hu M (2021) Intelligent fault diagnosis of rolling bearings based on refined composite multi-scale dispersion q-complexity and adaptive whale algorithm-extreme learning machine. Measurement 176: 108977.

Du W, Zhang Q, Chen Y and Ye Z (2021) An urban short-term traffic flow prediction model based on
wavelet neural network with improved whale optimization algorithm. Sustainable Cities and Society 69: 102858.

Fu W, Wang K, Li C, Li X, Li Y and Zhong H (2018) Vibration trend measurement for hydropower generator based on optimal variational mode decomposition and LSSVM improved with chaotic sine cosine algorithm optimization. Measurement Science and Technology.

Fu W, Wang K, Li C and Tan J (2019) Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM. Energy Conversion and Management 187: 356-377.

Furquim G, Pessin G, Faïçal B S, Mendiondo E M and Ueyama J (2016) Improving the accuracy of a flood forecasting model by means of machine learning and chaos theory. Neural Computing and Applications 27(S): 1129-1141.

Hidayat F and Astsauri T M S (2021) Applied random forest for parameter sensitivity of low salinity water Injection (LSWI) implementation on carbonate reservoir. Alexandria Engineering Journal.

Huang G, Li X, Zhang B and Ren J (2021) PM2.5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition. Science of The Total Environment 768: 144516.

Huang S, Chang J, Huang Q and Chen Y (2014) Monthly streamflow prediction using modified EMD-based support vector machine. Journal of Hydrology 511: 764-775.

Huo W, Li Z, Zhang K, Wang J and Yao C (2020) GA-PIC: An improved Green-Ampt rainfall-runoff model with a physically based infiltration distribution curve for semi-arid basins. Journal of Hydrology 586: 124900.

Hussein S, Abdelkareem M, Hussein R and Askalany M (2019) Using remote sensing data for predicting potential areas to flash flood hazards and water resources. Remote Sensing Applications: Society and Environment 16: 100254.

Jamali B, Bach P M and Deletic A (2020) Rainwater harvesting for urban flood management – An integrated modelling framework. Water Research 171: 115372.

Li C, Tang G, Xue X, Chen X, Wang R and Zhang C (2020) The short-term interval prediction of wind power using the deep learning model with gradient descend optimization. Renewable Energy 155.

Luo J, Chen H, Heidari A A, Xu Y, Zhang Q and Li C (2019) Multi-strategy Boosted Mutative Whale-inspired Optimization Approaches. Applied Mathematical Modelling 73(SEP.): 109-123.

Masrur Ahmed A A, Deo R C, Feng Q, Ghahramani A, Raj N, Yin Z and Yang L (2021) Deep learning hybrid model with Boruta-Random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity. Journal of Hydrology 599: 126350.

Mirjalili S and Lewis A (2016) The Whale Optimization Algorithm. Advances in Engineering Software 95.

Mohanty M, Biswal P and Sabut S (2020) Machine learning approach to recognize ventricular arrhythmias using VMD based features. Multidimensional Systems and Signal Processing 31(1): 49-71.

Ni L, Wang D, Singh V P, Wu J, Wang Y, Tao Y and Zhang J (2020) Streamflow and rainfall forecasting by two long short-term memory-based models. Journal of Hydrology 583: 124296.

Nikoo M, Ramezani F, Hadzima-Nyarko M, Nyarko E K and Nikoo M (2016) Flood-routing modeling with neural network optimized by social-based algorithm. Natural Hazards 82(1): 1-24.

Pan M, Zhou H, Cao J, Liu Y and Chen C H (2020) Water Level Prediction Model Based on GRU and CNN. IEEE Access PP(99): 1-1.

Partington D, Brunner P, Simmons C T, Werner A D, Therrien R, Maier H R and Dandy G C (2012)
Evaluation of outputs from automated baseflow separation methods against simulated baseflow from a physically based, surface water-groundwater flow model. Journal of Hydrology 458-459: 28-39.

Paul C, Roy P K and Mukherjee V (2020) Chaotic whale optimization algorithm for optimal solution of combined heat and power economic dispatch problem incorporating wind. Renewable Energy Focus 35: 56-71.

Peng T and Zhou B (2019) Hybrid bi-objective gray wolf optimization algorithm for a truck scheduling problem in the automotive industry. Applied Soft Computing 81: 105513.

Shoaib M, Shamseldin A Y and Melville B W (2014) Comparative study of different wavelet based neural network models for rainfall–runoff modeling. Journal of Hydrology 515: 47-58.

Stergiou K and Karakasidis T E (2021) Application of deep learning and chaos theory for load forecasting in Greece. Neural Computing and Applications.

Sun W and Zhang C (2018) Analysis and forecasting of the carbon price using multi—resolution singular value decomposition and extreme learning machine optimized by adaptive whale optimization algorithm. Applied Energy 231: 1354-1371.

Tehrany M S, Pradhan B and Jebur M N (2015) Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. Stochastic Environmental Research & Risk Assessment 29(4): 1149-1165.

Tuani A F, Keedwell E and Collett M (2020) Heterogenous Adaptive Ant Colony Optimization with 3-opt local search for the Travelling Salesman Problem. Applied Soft Computing 97: 106720.

Wang Q, Liu Y, Yue Q, Zheng Y and Yu J (2020a) Impact of Input Filtering and Architecture Selection Strategies on GRU Runoff Forecasting: A Case Study in the Wei River Basin, Shaanxi, China. Water 12(12): 3532.

Wang S, Liu G, Gao M, Cao S, Guo A and Wang J (2020b) Heterogeneous comprehensive learning and dynamic multi-swarm particle swarm optimizer with two mutation operators. Information Sciences 540: 175-201.

Wu C, Zhang X, Wang W, Lu C, Zhang Y, Qin W, Tick G R, Liu B and Shu L (2021) Groundwater level modeling framework by combining the wavelet transform with a long short-term memory data-driven model. Science of The Total Environment 783: 146948.

Xiao S, Wang H, Wang W, Huang Z, Zhou X and Xu M (2021) Artificial bee colony algorithm based on adaptive neighborhood search and Gaussian perturbation. Applied Soft Computing 100: 106955.

Xie T, Zhang G, Hou J, Xie J, Lv M and Liu F (2019) Hybrid forecasting model for non-stationary daily runoff series: A case study in the Han River Basin, China. Journal of Hydrology 577: 123915.

Yan C, Li M and Liu W (2020) Prediction of bank telephone marketing results based on improved whale algorithms optimizing S_Kohonen network. Applied Soft Computing 92: 106259.

Yaseen Z M, Jaafar O, Deo R C, Kisi O, Adamowski J, Quilty J and El-Shafie A (2016) Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq. Journal of Hydrology 542: 603-614.

Yoon H, Jun S-C, Hyun Y, Bae G-O and Lee K-K (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. Journal of Hydrology 396(1): 128-138.

Zhang C, Zhou J, Li C, Fu W and Peng T (2017) A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed
Zhang D, Lindholm G and Ratnaw Ee Ra H (2018) Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. Journal of Hydrology 556: 409-418.

Zhou Y, Guo S and Chang F-J (2019) Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts. Journal of Hydrology 570: 343-355.

Zosso D and Dragomiretskiy K (2014) Variational Mode Decomposition. IEEE Transactions on Signal Processing: A publication of the IEEE Signal Processing Society.

Zuo G, Luo J, Wang N, Lian Y and He X (2020) Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting. Journal of Hydrology 585: 124776.