Multi-scale flood prediction based on GM (1,2)-fuzzy weighted Markov and wavelet analysis

Jinping Zhang, Yuhao Wang, Yong Zhao and Hongyuan Fang

ABSTRACT

In order to forecast flood accurately and reveal the relationship between rainstorm and flood at the micro level, a model which combines wavelet analysis, GM (1,2) and fuzzy weighted Markov is built. Taking the Jialu River of Zhengzhou City in China as study area, the GM (1,2) model is constructed between the components of rainfall and flood volume by wavelet decomposition to connect the two variables, then a fuzzy weighted Markov method is introduced to correct the predicted component of flood volume. The corrected results are superimposed to obtain the predicted value of flood. To verify the reliability of the model, the maximum daily, 3-, 5- and 7-day flood volume of the next five floods in Zhongmu and Jiangang hydrological stations are predicted in turn. The results show that the multi-scale flood forecasting model has high overall forecasting accuracy, with the average relative errors all less than 10%. The forecasting accuracy of maximum five-day flood volume is higher than other periods. On the micro level, the results indicate that the fluctuation trend and period of rainfall-flood volume in d1, d2 and d3 are basically the same. Among the components of forecasted flood, the impact of rainfall on flood volume is most significant in the d3 component.

Key words | flood forecast, fuzzy weighted Markov, GM (1,2), wavelet analysis

Jinping Zhang
Yuhao Wang
Hongyuan Fang
School of Water Conservancy Engineering, Zhengzhou University, High-Tech District, No. 100 Science Road, Zhengzhou City 450001, Henan Province, China

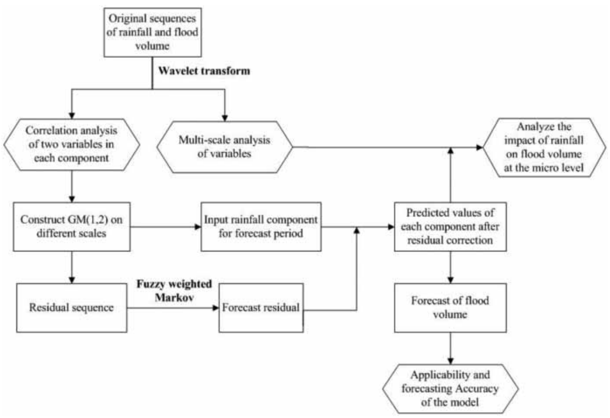
Yong Zhao (corresponding author)
State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China
E-mail: *iwhrzhy@sohu.com*

HIGHLIGHTS

- Combining GM (1,2) with wavelet analysis can reveal the relationship between rainfall and flood volume at the micro level, so as to better reflect the physical mechanism between them.
- The forecasted flood volume is reflected not only at the macro level but also at the micro level.
- Using the fuzzy weighted Markov method to correct the predicted components, then the prediction model has a favorable prediction effect.

doi: 10.2166/wcc.2021.289

GRAPHICAL ABSTRACT



INTRODUCTION

Flood caused by rainstorm is one of the severe and abrupt natural disasters, which poses a great threat to the national economy and people's security. With the continuous development of urbanization in China, cultivated land and grassland around the city is gradually being replaced by construction land, resulting in the change of relationship between runoff generation and confluence (Milly *et al.* 2002). The underlying surface variation has caused significant hydrological changes such as increased frequency of floods, increased peaks of floods, and increased flood volume. Therefore, flood prediction model is ever more important to be established (Lee *et al.* 2018).

The grey model is mainly suitable for forecasting problems with less sequence data and small volatility (Huang & Shen 2013), while the Markov chain is suitable for forecasting problems with large sequence volatility (Bonakdari *et al.* 2019). Combining the grey model with the weighted Markov chain can determine the relationship between variables and make preliminary predictions, and then it predicts the range of random fluctuation by the Markov chain method to optimize and adjust the grey predicted result, which can enhance the preciseness of the prediction (Kumar & Jain 2010). Therefore, some scholars combined the grey theory with the weighted Markov chain to predict runoff in recent years (Li *et al.* 2007). However, these studies were mainly based on the autocorrelation of the runoff sequence, which is GM (1,1). Nevertheless, flood is closely related to rainfall (Qin *et al.* 2013), so these studies cannot reflect the correlation between flood and rainstorm, they also cannot accurately predict the flood caused by extra heavy rainstorms. In addition, there is no uniform weighted Markov method, so different weighted Markov methods have a certain influence on forecast accuracy.

The basic purpose of wavelet analysis is to determine the frequency (or scale) content of the signal and to evaluate and determine the time variation of the frequency content (Labat 2005). As for discrete wavelet analysis, it is generally used to decompose a series into sub-signals given proper wavelet and decomposition levels, and then to guide various time series analysis, such as wavelet decomposition, wavelet denoising, and wavelet-assisted hydrological forecasting (Sang 2013; Ling *et al.* 2017). Wavelet analysis has been widely used in catastrophe detection and periodic change analysis of hydrological variables such as precipitation and runoff (Xu *et al.* 2009; Zhang *et al.* 2019). Moreover, wavelet analysis can also be combined with neural networks and other methods to predict runoff (Kalteh 2013; Sahay & Srivastava 2014).

It is more reasonable to construct a GM (1,2) model of rainstorm-flood to predict flood than a GM (1,1) model because of the good correlation between rainfall and flood volume. Besides, a fuzzy weighted Markov method which can theoretically improve forecast accuracy is introduced to predict the residual because it has large volatility. However, whether it is the weighted Markov model or the grey model, their prediction of variables is only reflected on the macro scale. The predicted results do not timely reflect the variation trend of variables on the micro scale, nor can they reveal the uncertainty relationship between variables at the micro level. Combining wavelet analysis with GM (1,2)-fuzzy weighted Markov can effectively solve this problem.

Based on the relationship between rainfall and flood volume, it is important to reflect the change of flood volume at the macro level, but the physical mechanism between them can be better reflected by analyzing the change of flood volume at the micro level, which needs to combine grey model and wavelet analysis. Therefore, the innovation of this paper is to combine the two methods and reveal the relationship between rainfall and flood volume at the micro level, then forecasted results are reflected not only at the macro level, but also at the micro level.

In this paper, wavelet transform is used to stratify rainfall-flood volume and analyze their changes at the micro level. The GM (1,2) model is constructed to link the rainfall and flood volume on different components, and the fuzzy weighted Markov method is used to predict residual and correct the forecast value of the grey model. The flood forecasting value is obtained by superimposing the forecasted component values. In order to verify the reliability of the model, the maximum daily, 3-, 5- and 7-day flood volume of the next five floods in Zhongmu and Jiangang hydrological stations are forecasted in this paper.

METHODS

Wavelet analysis

The basic idea of wavelet transform is to use a cluster of wavelet functions to represent or approximate a certain signal or function. Therefore, the wavelet function which can quickly decay to zero is the key to the wavelet transform (Mallat 1989), that is, the wavelet function satisfies this condition as follows:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \tag{1}$$

in which $\psi(t)$ is the basis wavelet function, and it can form a cluster of functions through scaling and translation on the time axis. The wavelet $\psi_{a,b}(t)$ is expressed as:

$$\psi_{a,b}(t) = |a|^{\frac{-1}{2}}\psi\left(\frac{t-b}{a}\right) \tag{2}$$

where a is the scale parameter, and b is the translation parameter. The former reflects the period length of the wavelet and the latter reflects the translation in time. They must both be real and positive (Cazelles *et al.* 2008). For a given energy-limited signal $f(t) \in L^2(R)$, continuous wavelet transform can be described as:

$$W_f(a, b) = |a|^{\frac{-1}{2}} \int f(t)\psi\left(\frac{t-b}{a}\right)dt \tag{3}$$

in which $\overline{\psi}(t-b/a)$ is a complex conjugate function of $\psi(t-b/a)$. Hydrological time series data are mostly discrete, set function $f(k\Delta t)$, where $1 \leq k \leq N$ and Δt is the sampling interval. The discrete wavelet transforms of Equation (3) are as follows:

$$W_f(a, b) = |a|^{\frac{-1}{2}}\Delta t \sum_{k=1}^{N} f(\Delta kt)\left(\frac{kt-b}{a}\right) \tag{4}$$

From Equation (4), we can determine the basic principle of wavelet transform, that is, obtaining low-frequency or high-frequency information of signals by increasing or decreasing a, and then analyzing the signal details to realize different time scale and spatial local characteristics of the signals (Seo *et al.* 2017).

GM (1,2) features

Deng (1989) proposed the grey system theory to study the uncertainty of the system. Grey models, including GM (1,1) and GM (1,2), were built based on grey system theory and are very effective to forecast linear and irregular data series. The meaning of the GM (1,2) model is to use a first-order differential equation to model two variables. The GM (1,2) characteristics were established based on the indicators that define the model, given the fact that these indicators are:

(a) The sequence of dependent variable:

$$X_1^{(0)} = (x_1^{(0)}(1), x_1^{(0)}(2), \cdots x_1^{(0)}(n))$$

(b) The sequence of independent variable:

$$X_2^{(0)} = (x_2^{(0)}(1), x_2^{(0)}(2), \cdots x_2^{(0)}(n))$$

(c) The sequence obtained by first-order accumulation of data series:

$$X_i^{(1)} = (x_i^{(1)}(1), x_i^{(1)}(2), \cdots x_i^{(1)}(n))$$

(d) The sequence consisting of the average of consecutive neighbors in the data series:

$$Z_1^{(1)} = (Z_1^{(1)}(1), Z_1^{(1)}(2), \cdots Z_1^{(1)}(n))$$

The specific formula is as follows:

$$x_i^{(1)}(k) = \sum_{k=1}^{n} x_i^{(0)}(k) \quad (5)$$

$$Z_1^{(1)}(k) = \frac{1}{2}[X_1^{(1)}(k) + X_1^{(1)}(k-1)] \quad (6)$$

Based on the above indicators that underline the GM (1,2) grey model, the model's differential equation can be written as:

$$x_1^{(0)}(k) + aZ_1^{(1)}(k) = bx_2^{(1)}(k) \quad (7)$$

where a is the development coefficient, b is the driving coefficient. The two parameters are obtained by the least square method (Hao et al. 2006; Bolos et al. 2016). Based on the above notations, the solution of the GM (1,2) model's differential equation $dx_1^{(1)}/dt + ax_1^{(1)} = bx_2^{(1)}$ is given by:

$$x_1^{(1)}(k+1) = \left[x_1^{(0)}(1) - \frac{b}{a}x_2^{(1)}(k+1)\right]e^{-ak} + \frac{b}{a}x_2^{(1)}(k+1) \quad (8)$$

$$x_1^{(0)}(k+1) = x_1^{(1)}(k+1) - x_1^{(1)}(k) \quad (9)$$

One-dimensional Markov chain

The Markov chain is a type of memoryless stochastic process in the state space that goes from one state to another. In other words, the next state is only affected by the current state, not the previous state. The Markov chain can be defined by the state transition probability matrix (Nielsen & Wakeley 2001). The dimension of the transition matrix is $m \times m$ if the data sequence is divided into m portions called states, and its contents can be described as:

$$p_{i,j} = \frac{S_{i,j}}{S_i} \quad (10)$$

$$\sum_{j=1}^{m} p_{i,j} = 1 \quad (11)$$

where $p_{i,j}$ is the probability of transition from state i to state j by one step; $S_{i,j}$ is the transition time from state i to state j by one step and S_i is the number of data belonging to the ith state.

The Markov chain occupies an important position in modern prediction methods. The state at the next moment can be predicted by the state transition probability matrix and the current state, and the prediction has high accuracy, scientificity and adaptability (Wang et al. 2007).

Fuzzy weighted Markov model

The traditional Markov chain can only make qualitative predictions for variables, but there is a large error in quantitative prediction. With the development of the Markov chain, many weighted Markov methods have been developed to make quantitative predictions (Alyousifi et al. 2020). In this paper, a method combining fuzzy mathematics with weighted Markov is introduced. According to the characteristics of the data at different times, the level eigenvalue method and multi-step state transition matrix are combined to quantitatively calculate the numerical value of the predicted target.

Assume that $X = \{x_1, x_2, \cdots x_{n-1}, x_n\}$, the autocorrelation coefficient r_k and weight w_k of each order are calculated based on the construction of the traditional Markov prediction model, where k is the number of steps of state transition selected for predicting system variables; u is the mean of the sequence and m is maximum order and its value is four, five or six generally:

$$r_k = \frac{\sum_{i=1}^{n-k}(x_i - u)(x_{i+k} - u)}{\sqrt{\sum_{i=1}^{n-k}(x_i - u)^2 \sum_{i=1}^{n-k}(x_{i+k} - u)^2}} \tag{12}$$

$$w_k = \frac{|r_k|}{\sum_{k=1}^{m}|r_k|} \tag{13}$$

The k-step state transition probability matrix P_k can be calculated according to a one-step state transition probability matrix, and the formula is as follows:

$$P_k = P_1^k \tag{14}$$

A predicted state transition probability matrix P is constructed if the state transition probability matrix of each order and the state of the lag period are combined. Probability of predictors in various states can be calculated, and the state with the highest transition probability is the state corresponding to the predicted target.

$$P = \{p_i^{(k)}\} \tag{15}$$

$$p_i = \sum_{k=1}^{m} w_k p_i^{(k)} \tag{16}$$

where k is step count, $p_i^{(k)}$ is the probability value of state i in order k, p_i is the probability of predictor in state i. The formulas for calculating the fuzzy number and level eigenvalue of each order are as follows:

$$d_i = \frac{p_i^{\varepsilon}}{\sum_{k=1}^{m} p_i^{\varepsilon}} \tag{17}$$

$$H = \sum_{i=1}^{m} i \times d_i \tag{18}$$

where ε is maximum probability index, and the role of maximum probability is more significant when the value is higher. Generally, the value is between two and four.

The following formula is used to calculate the predicted value of the variable:

$$x_i^{t+s} = \begin{cases} \dfrac{T_i H_i}{i + 0.5} & H_i > i \\ \dfrac{B_i H_i}{i - 0.5} & H_i < i \end{cases} \tag{19}$$

where T_i and B_i are the upper boundary and lower boundary of the ith state (Ran et al. 2006; Zhang et al. 2016).

Structure of the combined model

The sequences of rainfall and flood volume are decomposed into several component sequences by wavelet analysis. Based on the periodicity analysis and correlation analysis, the corresponding components of rainfall and flood volume are correlated by establishing GM (1,2) models at the micro level. The residual sequence is obtained by comparing the simulated sequence and the actual component sequence of each GM (1,2) model. Taking the rainfall component as an independent variable, the component value of flood volume is calculated. Based on residual sequence, fuzzy weighted Markov models are constructed to predict the residual error, thus, the predicted component value is corrected. Finally, flood volume is obtained by superposition. The specific process of flood prediction by the model is shown in Figure 1.

RESULTS AND DISCUSSION

Study area and data source

The Jialu River originates in Xinmi City (affiliated to Zhengzhou City in China), and eventually flows into the Shaying River, which is the main tributary of the Huaihe River Basin. The Jialu River in Zhengzhou City has many tributaries, including the Jinshui River, Suoxu River, Xiong'er River, Qili River, and the Dongfeng Canal (Yang et al. 2012; Wang et al. 2018). Additionally, both Jiangang and

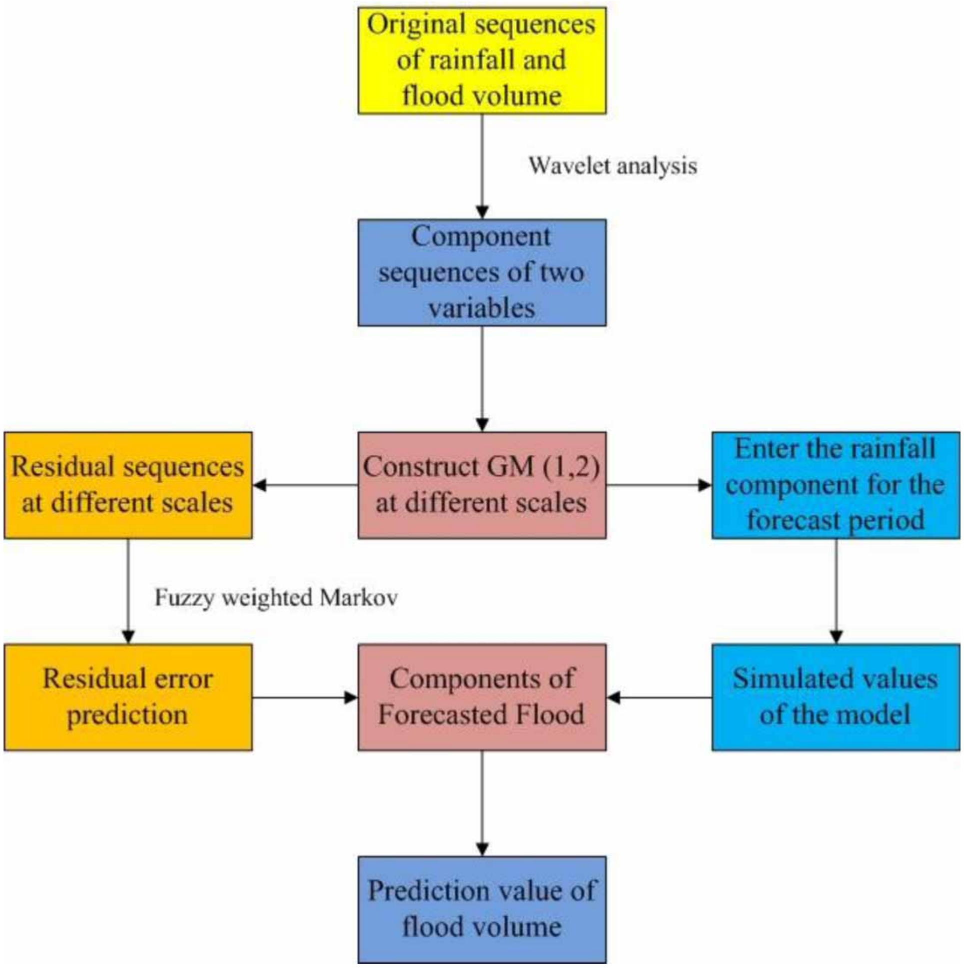


Figure 1 | Steps for predicting flood volume by model.

Zhongmu hydrological stations are located in the Zhengzhou section of the Jialu River, where the former is located upstream and the latter is located downstream, as shown in Figure 2.

Daily rainfall exceeding 50 mm is defined as rainstorm (Li *et al.* 2016). According to the daily rainfall data of four precipitation stations including Jiangang, Changzhuang, Sizhao and Sanli, the rainstorm process corresponding to the flood process of Zhongmu hydrological station was selected and compared. It can be concluded that the rainfall of Changzhuang rainfall station has a good correlation with the flood of Zhongmu hydrological station, and 51 rainstorm floods were screened out from 1980 to 2012. Similarly, 56 rainstorm floods were screened out according to the rainfall data of Jiangang rainfall station and flow data of Jiangang hydrological station (1980–2012). On this basis, the maximum daily, 3-, 5- and 7-day rainfall and flood volume of each storm flood can be calculated.

Multi-scale decomposition of rainstorm-flood

Taking the maximum daily rainfall and maximum daily flood volume of 51 rainstorm floods in Zhongmu hydrological station as an example, the change curves of rainfall and flood volume are shown in Figure 3.

Daubechies wavelet is generally abbreviated to dbN, where N is the order of the wavelet. In the wavelet analysis of hydrological series, db6 or db4 are usually used (Roushangar *et al.* 2018; Guo *et al.* 2019). In this paper, db6 wavelet function is selected to decompose the original

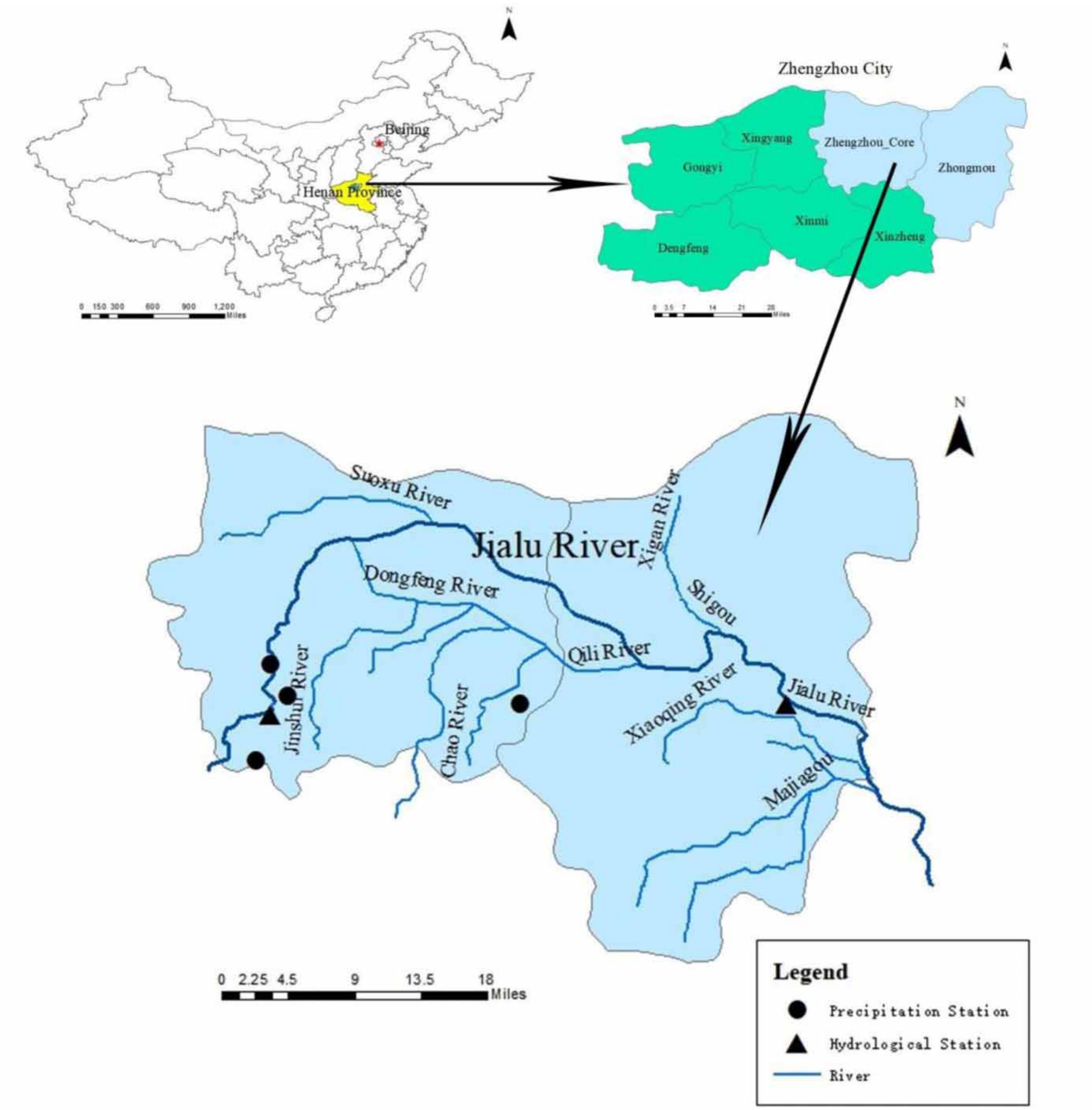


Figure 2 | Drainage map of the Jialu River in Zhengzhou City.

sequence. Moreover, the appropriate number of layers should be selected to decompose the variable in order to analyze its periodic variation. By the time the wavelet decomposition reaches the 5th layer, its high-frequency signal still fluctuates periodically and its low-frequency signal has no obvious periodic fluctuation, thus, the number of decomposition layers of wavelet transform is determined to be five layers. Five detail components and one trend component were obtained, that is d1, d2, d3, d4, d5 and a5, as shown in Figure 4. The original sequence consists of these six component sequences, and a5 reflects the overall change trend of the original sequence (Adewusi & Albedoor 2001).

It can be seen from Figure 4(a) that both P_{d1} and R_{d1} have fluctuation periods of quasi-2–4; (b) shows that both P_{d2} and R_{d2} have fluctuation periods of quasi-4–6; (c) shows that P_{d3} has fluctuation periods of quasi-8–15 while R_{d3} has fluctuation periods of quasi-9–14; (d) shows that P_{d4} has fluctuation periods of quasi-23 while R_{d4} has fluctuation periods of

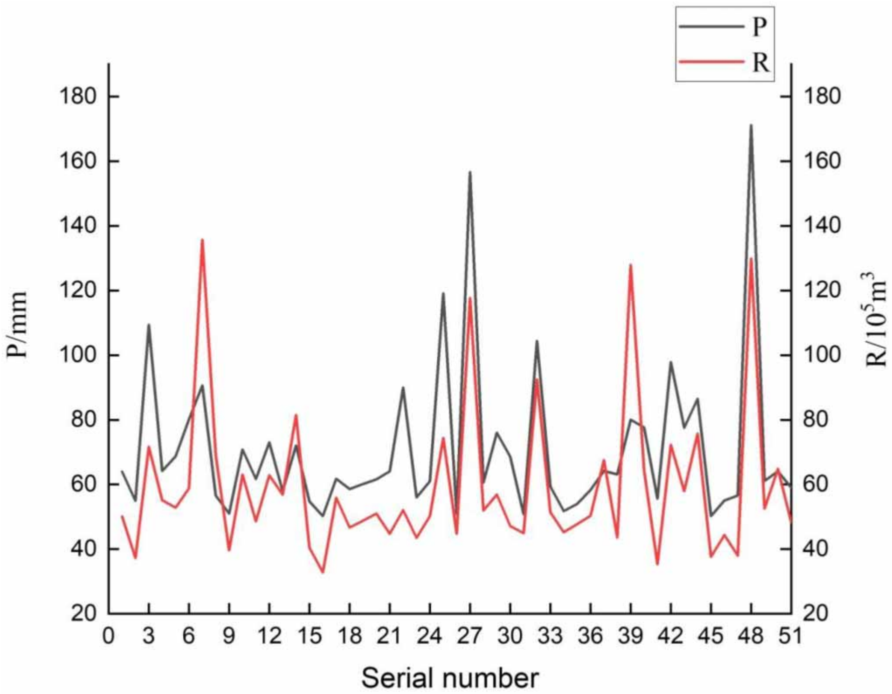


Figure 3 | Maximum daily rainfall-flood volume of 51 rainstorm floods.

quasi-17; (e) shows that P_{d5} has fluctuation periods of quasi-28. In addition, the fluctuation trend of the maximum daily rainfall and flood volume is basically the same in d1, d2 and d3 while the fluctuation trend of two variables is the opposite in d5; (f) shows that the sequence of the maximum daily rainfall decreased slightly and then increased slowly on the whole, while the sequence of the maximum daily flood volume shows a more obvious trend of decreasing first and then increasing on the whole. Combined with the actual situation, this may be due to the increasing urbanization of Zhengzhou City, the decrease of the permeable area and the acceleration of the confluence process in recent years (Wang *et al.* 2017).

In order to analyze the correlation between rainfall and flood volume in different components, the related coefficients between the components of two variables are calculated in Table 1. Moreover, the correlation coefficient represents the similarity of the changes of two variables, it also means that the correlation coefficient can measure the stability of the relationship between variables, so the magnitude of correlation coefficient can be used as a reference for reasonable construction of the GM (1,2) model.

It can be seen from Table 1 that the d1, d2, d3 and a5 of maximum daily rainfall-flood volume are closely and positively related, the d5 of two variables are closely and negatively correlated, while the correlation between them is the worst in the d4. The GM (1,2) model built by the two variables with poor correlation has a poor simulation effect. Moreover, the amounts of rainfall and flood on the d5 component are so small that it has little effect on the forecast of the flood. In order to improve the forecasting accuracy, d4, d5 and a5 are superimposed. Therefore, four GM (1,2) forecasting models are constructed for P_{d1}-R_{d1}, P_{d2}-R_{d2}, Pd_3-R_{d3}, $P_{d4+d5+a5}$–$R_{d4+d5+a5}$ respectively.

Determination of equivalent substitution value

Since the decomposed sequences such as d1, d2, and d3 obtained by wavelet transform have multiple negative values, the six data sequences are equivalently replaced to satisfy the non-negative requirements of the GM (1,2) model for data. That is, each item of the original sequence subtracts the minimum value *M* of the sequence to obtain new sequences including $P_{b1} \sim P_{b3}$, $R_{b1} \sim R_{b3}$. The minimum values of the sequences are shown in Table 2.

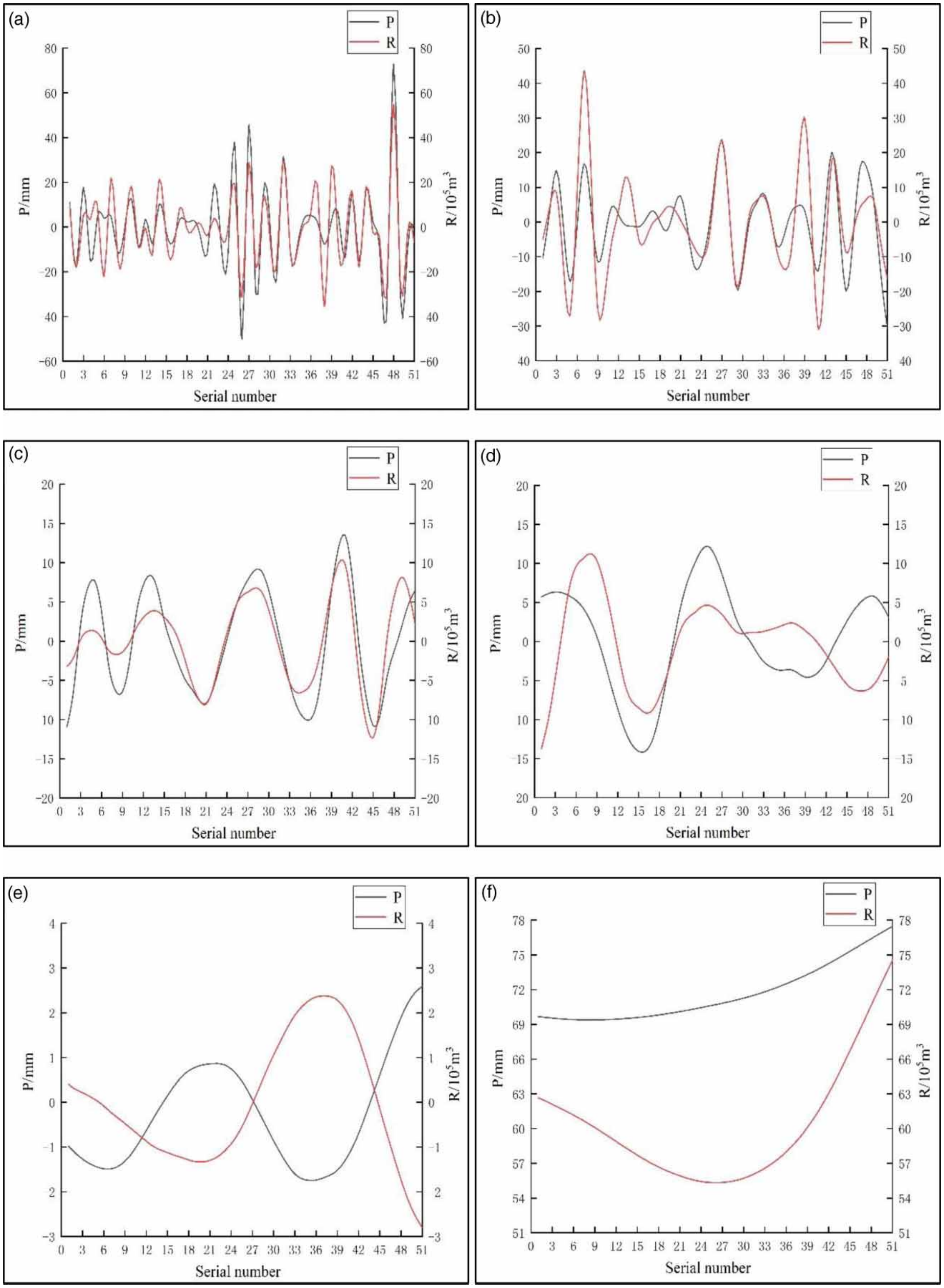


Figure 4 | The decomposed data sequences of rainfall and flood volume: (a) component d1; (b) d2; (c) d3; (d) d4; (e) d5; and (f) a5.

Table 1 | Related coefficients between components of two variables

Sequence	P–R	P_{d1}–R_{d1}	P_{d2}–R_{d2}	P_{d3}–R_{d3}	P_{d4}–R_{d4}	P_{d5}–R_{d5}	Pa_5–R_{a5}
r	0.770	0.840	0.727	0.865	0.343	–0.813	0.760

Construction of GM (1,2)

Taking P_{b1}–R_{b1} as a case, the model parameters $a = 1.7836$ and $b = 1.2604$ are obtained by using the least square method. According to Equation (8), the prediction function is as follows:

$$x_1^{(1)}(k+1) = [44.415 - 0.7067x_2^{(1)}(k+1)]e^{-1.7836k} + 0.7067x_2^{(1)}(k+1)$$

Similarly, the other models are as follows:

$$x_1^{(1)}(k+1) = [25.988 - 1.0212x_2^{(1)}(k+1)]e^{-1.0727k} + 1.0212x_2^{(1)}(k+1)$$

$$x_1^{(1)}(k+1) = [9.079 - 1.2742x_2^{(1)}(k+1)]e^{-0.2035k} + 1.2742x_2^{(1)}(k+1)$$

$$x_1^{(1)}(k+1) = [49.312 - 0.9443x_2^{(1)}(k+1)]e^{-0.2156k} + 0.9443x_2^{(1)}(k+1)$$

Combining Equation (9), the simulated values of each component of the maximum daily flood volume are obtained by inverse transformation of the equivalent substitution, and four groups of residual sequence are obtained by comparing with R_{d1}, R_{d2}, R_{d3} and $R_{(d4+d5+a5)}$. In addition, the component simulation values of maximum daily flood volume are obtained from the rainfall of the next five floods. Taking the 52nd flood as an example, the calculation results are shown in Table 3.

Table 2 | Minimum value *M* of the sequence

i	1	2	3
P_{di}	–50.044	–30.770	–10.980
R_{di}	–35.369	–31.026	–12.289

Fuzzy weighted Markov residual prediction model

According to Equation (19), when the upper and lower limit values appear negative, they cannot satisfy the requirements. Therefore, in order to avoid unnecessary calculation problems, Y_{d1}, Y_{d2}, Y_{d3} and $Y_{d4+d5+a5}$ are obtained from equivalent substitution of four residual sequences. The method is the same as above under 'Determination of equivalent substitution value'.

According to the mean value μ and standard deviation σ of the new sequence after the equivalent substitution, the state values of the variables are divided into five levels with $\mu - \sigma$, $\mu - 0.5\sigma$, $\mu + 0.5\sigma$, $\mu + \sigma$ as the boundaries. The specific criteria for division are shown in Table 4.

Taking Y_{d1} as an example, the one-step state transition probability matrix can be calculated. The value of the ith row and jth column in the one-step transition probability matrix is the number of times the state i passes to the state j after one step in the sample. When the state i appears in the last part of the data sequence, the last one is not counted in the total. The one-step state transition probability matrix P_1 is as follows:

$$P_1 = \begin{bmatrix} 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 3/7 & 2/7 & 2/7 \\ 1/25 & 4/25 & 13/25 & 1/5 & 2/25 \\ 1/8 & 3/8 & 1/2 & 0 & 0 \\ 1/3 & 1/6 & 1/3 & 1/6 & 0 \end{bmatrix}$$

According to Equation (14), multi-step state transition

Table 3 | Component simulation value of maximum daily flood volume for 52nd flood

Model number	Components of rainfall	Substitution value of rainfall component	Substitution value of flood component	Component of flood volume
1	23.653	73.697	52.079	16.710
2	–9.335	21.434	21.888	–9.138
3	7.871	18.851	24.024	11.735
4	81.911	81.911	77.365	77.365

Table 4 | Criteria for dividing the state of variables

State	Y_{d1}	Y_{d2}	Y_{d3}	$Yd_{4+d5+a5}$
1	0 ~ 20.558	0 ~ 7.377	0 ~ 3.501	0 ~ 3.139
2	20.558 ~ 25.181	7.377 ~ 12.261	3.501 ~ 5.762	3.139 ~ 8.207
3	25.181 ~ 34.426	12.261 ~ 22.030	5.762 ~ 10.283	8.207 ~ 18.345
4	34.426 ~ 39.049	22.030 ~ 26.914	10.283 ~ 12.543	18.345 ~ 23.414
5	39.049 ~ +∞	26.914~ +∞	12.543 ~ +∞	23.414~ +∞

probability matrix is obtained:

$$P_2 = \begin{bmatrix} 14/75 & 49/300 & 32/75 & 11/60 & 1/25 \\ 311/2100 & 67/300 & 242/525 & 2/15 & 6/175 \\ 364/5023 & 1263/7363 & 727/1497 & 428/2625 & 380/3541 \\ 1/50 & 2/25 & 475/983 & 29/140 & 587/2800 \\ 41/1200 & 139/1200 & 1039/2100 & 4/35 & 253/1050 \end{bmatrix}$$

$$P_3 = \begin{bmatrix} 213/3995 & 323/2248 & 2451/5000 & 52/375 & 653/3750 \\ 349/7500 & 250/1931 & 1978/4057 & 283/1750 & 213/1219 \\ 376/4975 & 163/1040 & 857/1787 & 517/3152 & 428/3449 \\ 298/2589 & 883/4649 & 734/1565 & 1859/12037 & 298/4167 \\ 333/2911 & 283/1745 & 467/1012 & 544/3159 & 433/4824 \end{bmatrix}$$

$$P_4 = \begin{bmatrix} 233/2453 & 409/2565 & 399/848 & 387/2302 & 159/1487 \\ 245/2501 & 480/2861 & 2133/4525 & 224/1369 & 215/2166 \\ 358/4417 & 301/1894 & 418/975 & 239/1481 & 807/6673 \\ 240/3877 & 892/6157 & 661/1366 & 515/3219 & 1129/7560 \\ 226/3233 & 323/2106 & 906/1877 & 781/5085 & 867/6173 \end{bmatrix}$$

$$P_5 = \begin{bmatrix} 253/3352 & 569/3644 & 1443/3005 & 423/2686 & 749/5731 \\ 337/4655 & 203/1324 & 705/1466 & 505/3181 & 479/3558 \\ 250/3141 & 682/4341 & 1165/2437 & 449/2787 & 310/2497 \\ 178/1997 & 180/1109 & 826/1741 & 1553/9524 & 451/4061 \\ 411/4817 & 663/4190 & 1433/3015 & 207/1264 & 505/4302 \end{bmatrix}$$

The coefficients and weights of each order obtained by Equations (12) and (13) are shown in Table 5. In addition, the prediction probability matrix is constructed to predict the sequence value of the 52nd flood, as shown in Table 6.

The sum of the product of the fuzzy number and the state level is used as the characteristic value, the calculation result is as follows:

$$H = \sum_{i=1}^{5} i \times d_i = 3.4087$$

Table 5 | Coefficients and weights of each order

Order	1	2	3	4	5
r	−0.648	−0.061	0.233	−0.216	0.190
w	0.4812	0.0452	0.1727	0.1599	0.1409

From Table 6, the transition probability of state three is the largest. Therefore, the sequence value of the 52nd flood is predicted to be in state three. Combining Equation (19), the predicted value of new series is 33.528. After restoring, the residual correction value in d1 of the maximum daily flood volume is 3.594. Similarly, the residual correction values of other components are obtained.

Residual correction is carried out on the simulation results of each layer of grey model and the corrected results are superimposed to obtain the maximum daily flood forecasting value of the 52nd to 56th floods in Zhongmu hydrological station, which is compared with the actual measured value. The results are shown in Table 7.

The maximum daily flood components of the original sequence and the predicted components are combined (Figure 5), and it can be concluded that during the forecasting period, the d1 of maximum daily flood volume has a quasi-2 fluctuation period with slightly reduced fluctuation amplitude when the d2 presents a quasi-6 fluctuation period with increased fluctuation amplitude compared with the previous period; the fluctuation amplitude of d3 component has obvious change compared with the previous period with a short sudden change in fluctuation period; the fluctuation amplitude of d4 + d5 + a5 increases slightly, but generally tends to be flat. In a word, the influence of rainfall on the flood volume is the most significant in the d3 component during the process of using the model to predict the maximum daily flood volume of Zhongmu hydrological station.

Table 6 | Prediction probability matrix

Serial number	Order/State	Weight	Status 1	Status 2	Status 3	Status 4	Status 5
47	5/3	0.1409	250/3,141	682/4,341	1,165/2,437	449/2,787	310/2,497
48	4/3	0.1599	358/4,417	301/1,894	418/975	239/1,481	807/6,673
49	3/3	0.1727	376/4,975	163/1,640	857/1,787	517/3,152	428/3,449
50	2/4	0.0452	1/50	2/25	475/983	29/140	587/2,800
51	1/2	0.4812	0	0	3/7	2/7	2/7
Weighted sum			0.0381	0.0782	0.455	0.224	0.205
Fuzzy number			0.0047	0.0200	0.6746	0.1633	0.1374

Table 7 | Maximum daily flood forecast of 52nd to 56th floods in Zhongmu

Serial number	R_{d1}	R_{d2}	R_{d3}	$R_{(d4+d5+a5)}$	Predictive value	Measured value	Relative error (%)
52	20.304	–7.455	7.363	70.851	91.062	91.584	–0.570
53	–24.669	18.311	7.064	70.181	70.888	67.478	5.054
54	24.135	15.262	0.959	69.633	109.990	111.110	–1.008
55	–16.520	7.844	−8.342	69.263	52.245	60.653	–13.863
56	5.878	4.338	−14.477	68.664	64.402	76.118	–15.391

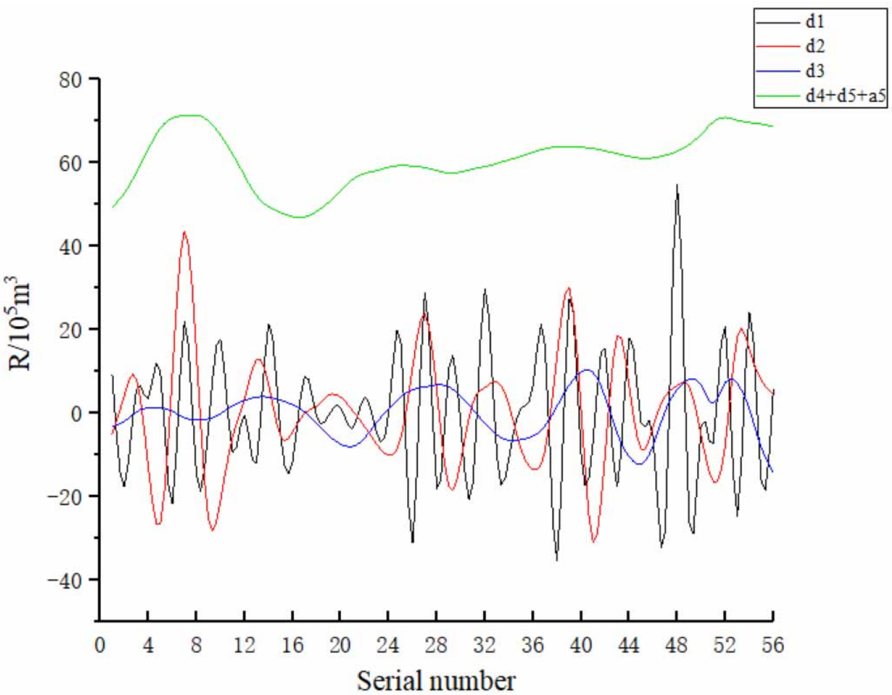


Figure 5 | Components of maximum daily flood volume.

Evaluation of flood prediction effect

The flood volume for each period of the next five floods in two hydrological stations is forecast. In the process, we can conclude that both rainfall and flood volume of storm floods in two areas decreased first and then increased, and the change trend of flood volume was more obvious. In addition, the flood volume corresponding to the same level of rainfall gradually increases. In all periods rainfall and flood volume have high consistency in these components including d1, d2 and d3. The relative errors of predicted flood are shown in Figures 6 and 7.

From Figures 6 and 7 it can be concluded that the multi-scale flood forecasting model based on GM (1,2)-fuzzy weighted Markov and wavelet analysis has great prediction accuracy. The relative error meets the requirement of less than 20%.

Comparing the forecasted flood volume in each period of the next five floods with the actual measurement, the average relative errors in Zhongmu hydrological station are 7.177, 7.554, 5.240 and 9.937%, while the average relative errors in Jiangang hydrological station are 5.239, 7.967, 4.544 and 5.803%. In addition, the forecast accuracy of the maximum 5-day flood volume is higher than that of other periods in both hydrological stations.

CONCLUSIONS

Using wavelet analysis, a GM (1,2) model and the fuzzy weighted Markov method, the model is constructed which not only reveals the relationship between rainfall and flood at the micro level, but also forecasts flood volume at the micro and macro level.

The multi-scale flood forecasting model based on GM (1,2)-fuzzy weighted Markov and wavelet analysis has high forecasting precision. Whether it is Zhongmu hydrological station or Jiangang hydrological station, the average relative errors of the maximum daily, 3-, 5- and 7-day flood volume are all less than 10% when compared with the actual value. In addition, the forecasting accuracy of maximum 5-day flood volume is higher than other periods, with average relative errors of 5.240 and 4.544%. Because of the time lag between rainstorm and flood, the model can forecast the downstream flood according to the rainfall at the upstream rain station, and then relevant departments can take appropriate measures to avoid unnecessary losses.

In the process of forecasting flood, it can be concluded that the flood volume of two hydrological stations decreased first and then increased, while the corresponding flood volume of the same rainfall level increased gradually. At the micro level, the fluctuation trend and period of rainfall-flood volume in d1, d2 and d3 are basically the same. Among the components of forecasted flood, the fluctuation

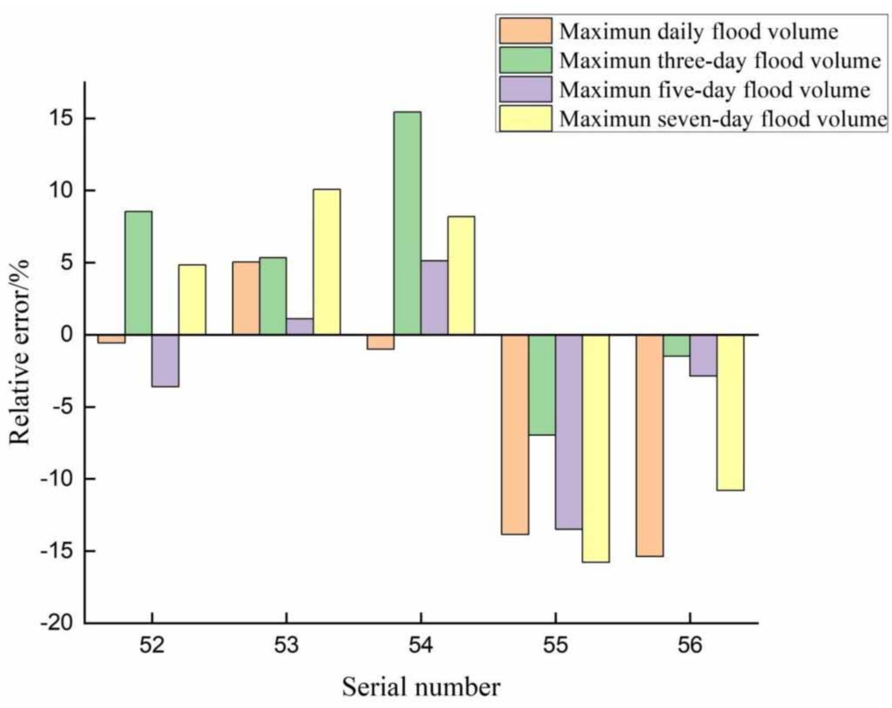


Figure 6 | Relative error of flood forecasting in Zhongmu hydrological station.

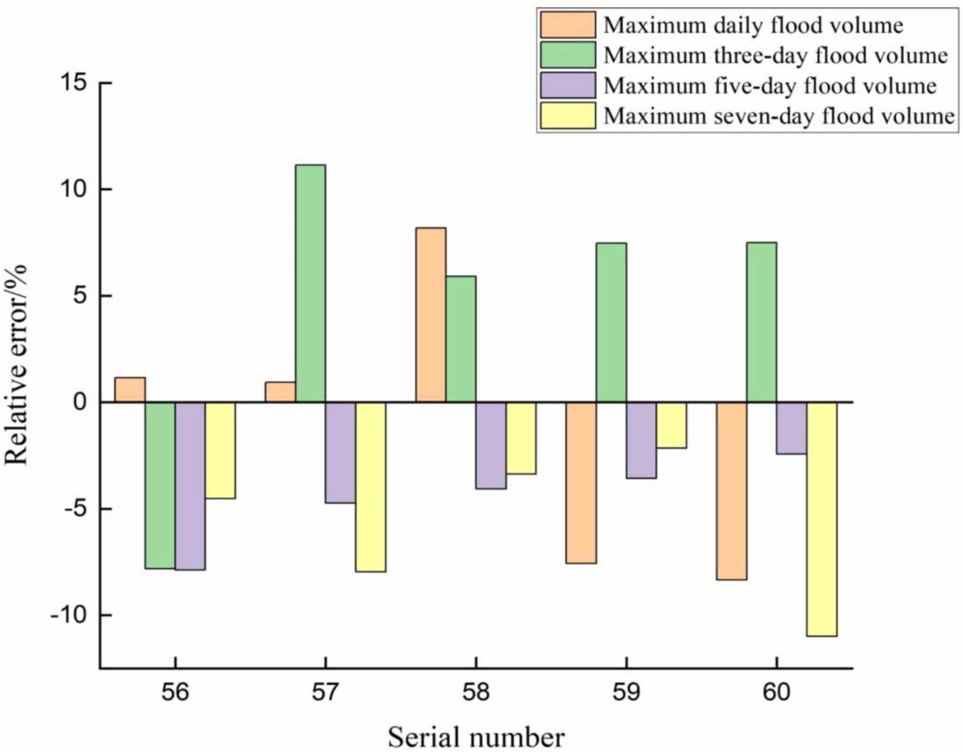


Figure 7 | Relative error of flood forecasting in Jiangang hydrological station.

periods and amplitudes of d1 and d2 are basically the same as those of the actual measuring period, while d3 has an obvious abrupt change, that is, the impact of rainstorm on flood is more significant in the d3 component.

In summary, the model is effective in forecasting flood and researching the uncertainty relationship between rainfall and flood volume, so it provides a new idea and method for flood prediction. Moreover, many factors, such as the stationary of the original data series, the fineness of the state partition in the Markov chain and the fuzzy mathematics method, may affect the accuracy of the model, thus, how to coordinate these factors may be a research direction.

ACKNOWLEDGEMENTS

This research was supported by the National Key R&D Program of China (Grant No. 2018YFC0406501), Program for Innovative Talents (in Science and Technology) at University of Henan Province (Grant No. 18HASTIT014), and Foundation for University Youth Key Teacher of Henan Province (Grant No. 2017GGJS006).

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Adewusi, S. A. & AlBedoor, B. O. 2001 Wavelet analysis of vibration signals of an overhang rotor with a propagating transverse crack. *Journal of Sound and Vibration* **246** (5), 777–793.
Alyousifi, Y., Othman, M., Faye, I., Sokkalingam, R. & Silva, P. C. L. 2020 Markov weighted fuzzy time-series model based on an optimum partition method for forecasting air pollution. *International Journal of Fuzzy Systems* **22** (5), 1468–1486.
Bolos, M., Bradea, I. & Delcea, C. 2016 Adjusting the errors of the GM (1, 2) grey model in the financial data series using an

adaptive fuzzy controller. *Grey Systems-Theory and Application* **6** (3), 341–352.
Bonakdari, H., Zaji, A. H., Binns, A. D. & Gharabaghi, B. 2019 Integrated Markov chains and uncertainty analysis techniques to more accurately forecast floods using satellite signals. *Journal of Hydrology* **572**, 75–95.
Cazelles, B., Chavez, M., Berteaux, D., Menard, F., Vik, J. O., Jenouvrier, S. & Stenseth, N. C. 2008 Wavelet analysis of ecological time series. *Oecologia* **156** (2), 287–304.
Deng, J. L. 1989 Introduction to grey system theory. *Journal of Grey System* **1** (1), 1–24.
Guo, B. T., Sun, S. Y., Zhang, J. P. & Li, J. Y. 2019 Study of irrigation water amount in irrigation area based on multi-scale wavelet transform WNN. *Yellow River* **41** (11), 154–158.
Hao, Y. H., Yeh, T. C. J., Gao, Z. Q., Wang, Y. R. & Zhao, Y. 2006 A gray system model for studying the response to climatic change: the Liulin karst springs, China. *Journal of Hydrology* **328** (3–4), 668–676.
Huang, L. M. & Shen, B. 2013 Low-flow runoff prediction using the grey self-memory model. *Advances in Environmental Technologies* **726–731**, 3272–3278.
Kalteh, A. M. 2013 Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Computers & Geosciences* **54**, 1–8.
Kumar, U. & Jain, V. K. 2010 Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy* **35** (4), 1709–1716.
Labat, D. 2005 Recent advances in wavelet analyses: part I. A review of concepts. *Journal of Hydrology* **314** (1–4), 275–288.
Lee, E. H., Kim, J. H., Choo, Y. M. & Jo, D. J. 2018 Application of flood nomograph for flood forecasting in urban areas. *Water* **10** (1), 53. doi:10.3390/w10010053.
Li, X., Wang, X. Y., Shao, W., Xia, L. Y., Zhang, G. S., Tian, B., Li, W. D. & Peng, P. 2007 Forecast of flood in Chaohu Lake basin of china based on grey-Markov theory. *Chinese Geographical Science* **17** (1), 64–68.
Li, H. Y., Bao, S. S., Wang, X. J. & Lv, H. 2016 Storm flood characteristics and identification of periodicity for flood-causing rainstorms in the second Songhua River Basin. *Water* **8** (12), 529. doi:10.3390/w8120529.
Ling, H. B., Deng, X. Y., Long, A. H. & Gao, H. F. 2017 The multi-time-scale correlations for drought-flood index to runoff and North Atlantic Oscillation in the headstreams of Tarim River, Xinjiang, China. *Hydrology Research* **48** (1), 253–264.
Mallat, S. G. 1989 A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11** (7), 674–693.
Milly, P. C. D., Wetherald, R. T., Dunne, K. A. & Delworth, T. L. 2002 Increasing risk of great floods in a changing climate. *Nature* **415** (6871), 514–517.
Nielsen, R. & Wakeley, J. 2001 Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158** (2), 885–896.
Qin, H. P., Li, Z. X. & Fu, G. T. 2013 The effects of low impact development on urban flooding under different rainfall characteristics. *Journal of Environmental Management* **129**, 577–585.
Ran, J. J., Zhao, X. J. & Liang, C. 2006 Research on application of forecasting precipitation based on weighted Markov chain. *Yellow River* **28** (4), 32–34.
Roushangar, K., Nourani, V. & Alizadeh, F. 2018 A multiscale time-space approach to analyze and categorize the precipitation fluctuation based on the wavelet transform and information theory concept. *Hydrology Research* **49** (3), 724–743.
Sahay, R. R. & Srivastava, A. 2014 Predicting monsoon floods in rivers embedding wavelet transform, genetic algorithm and neural network. *Water Resources Management* **28** (2), 301–317.
Sang, Y. F. 2013 A review on the applications of wavelet transform in hydrology time series analysis. *Atmospheric Research* **122**, 8–15.
Seo, Y., Choi, Y. & Choi, J. 2017 River stage modeling by combining maximal overlap discrete wavelet transform, support vector machines and genetic algorithm. *Water* **9** (7), 525. doi:10.3390/w9070525.
Wang, W. S., Jin, J. L. & Li, Y. Q. 2007 Advances in stochastic simulation of Hydrology. *Advances in Water Science* **18** (5), 768–775.
Wang, H. L., Wu, Z. N. & Sun, R. C. 2017 Effect of urbanization in Zhengzhou on river hydrological process in Jialu River watershed. *Science Technology and Engineering* **17** (31), 316–321.
Wang, X. F., Zhao, G. P., Wang, H. Q., Liang, J. T., Xu, S. M., Chen, S. P., Xu, A. & Wu, L. J. 2018 Assessment of the cytotoxic and mutagenic potential of the Jialu River and adjacent groundwater using human-hamster hybrid cells. *Journal of Environmental Sciences* **70**, 133–143.
Xu, L. G., Zhou, H. F., Liang, C. & Wu, A. Q. 2009 Multi-time scale variability of precipitation in the desert region of North China. *Journal of Hydraulic Engineering* **40** (8), 1002–1011.
Yang, L., Song, X., Zhang, Y., Han, D., Zhang, B. & Long, D. 2012 Characterizing interactions between surface water and groundwater in the Jialu River basin using major ion chemistry and stable isotopes. *Hydrology and Earth System Sciences* **16** (11), 4265–4277.
Zhang, J., Tao, W. X. & Wang, Q. 2016 Application of weighted Markov chain in forecasting rainfall in Jinan City. *Yellow River* **38** (9), 13–16.
Zhang, J. P., Xiao, H. L., Zhang, X. & Li, F. W. 2019 Impact of reservoir operation on runoff and sediment load at multi-time scales based on entropy theory. *Journal of Hydrology* **569**, 809–815.

First received 6 October 2020; accepted in revised form 2 February 2021. Available online 22 March 2021