Study On Multiscale-Multivariate Attribution Analysis and Prediction of Urban Rainstorm Flood

Jinping Zhang
Zhengzhou University
Yuhao Wang (731270945@qq.com)
Zhengzhou University

Research Article

Keywords: Urban rainstorm flood, Multiscale attribution analysis, Flood prediction, Wavelet analysis method, Ridge regression model

Posted Date: November 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-929111/v1

Study on Multiscale-multivariate Attribution Analysis and Prediction of Urban Rainstorm Flood

Jinping Zhang[a,b,c], Yuhao Wang[a*]

a. *School of Water Conservancy Engineering, Zhengzhou University, Zhengzhou, 450001, China;*

b. *State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, 100000, China;*

c. *Yellow River Institute for Ecological Protection & Regional Coordinated Development, Zhengzhou University, Zhengzhou, 450001, China*

*Corresponding author. Tel.: +86-17538130612

E-mail addresses: wyh19971030@163.com (Yuhao Wang)

Abstract： In order to explore the impact of the changing environment on urban rainstorm flood, and reveal the relationship between flood volume and its influencing factors at the micro level, the rainfall and flood volume are decomposed by the wavelet analysis method to perform the multiscale attribution analysis. Then the multiscale-multivariate prediction model of urban rainstorm flood is constructed in the Jialu River Basin in Zhengzhou city of China. The results show that the main influencing factors of flood volume are rainfall and underlying surface, where the latter causes the mutation of flood volume in 1994 and 2005. At the micro level, there is a constant linear relationship between rainfall and flood volume in d1, d2 and d3, while the

impact of underlying surface on flood volume is mainly reflected in a3. The multiscale-multivariate prediction model has a good simulation effect on the flood volume of the first 45 rainstorm floods, *NSE*, R^2 and R_e are 0.966, 0.964 and 10.80%, respectively. Moreover, the model also has a good prediction effect, and the relative errors between the predicted and observed flood volume of 46th~50th rainstorm floods are all less than 20%.

Author keywords: Urban rainstorm flood; Multiscale attribution analysis; Flood prediction; Wavelet analysis method; Ridge regression model

1. Introduction

For recent years, the urban rainstorm flood disaster occurs frequently causing a great threat to the national economic development and people's lives and property. The continuous development of urbanization will induce the continuous changes of the urban rainstorm flood characteristics (Zhang et al. 2014). Therefore, many scholars have carried studies on the evolution of urban rainstorm flood by constructing models such as SWMM (Wu et al. 2017; Jamali et al. 2018), MIKE (Bisht et al. 2016; Li et al. 2018), InfoWorks ICM (Peng et al. 2015; Gong et al. 2018;). Moreover, Nigussie et al (2019) combined MIKE 21 model with the dynamic cellular automata-based urbanization model to predict urban flood under different scenarios in the future. However, most studies are focused on improving rainstorm flood hydrodynamic model or coupling with other methods, it is rare for the detailed study on the driving factors of urban rainstorm flood evolution. Meanwhile, most of the researches focus on the urban waterlogging simulation, it cannot truly reflect the overall impact of the changing environment on the urban rainstorm flood, while the study on the flood process at the urban

43 river outlet can effectively solve this problem (Ren et al. 2021).

44 Nowadays, statistical model and hydrological model have been widely used for predicting
45 hydrological variables. The hydrological model works well, but it requires lots of data and is
46 complicated to operate (Yan et al. 2008; Clark et al. 2015). In contrast, the statistical model is
47 simple to operate with an acceptable prediction accuracy (Wang and Huo, 2010), so more
48 scholars have used statistical models such as linear regression model (Madarang and Kang,
49 2014; Chu et al. 2017), grey model (Ho et al. 2015; Wu and Wang, 2020) and neural network
50 model (Taormina et al. 2015; Yaseen et al. 2016) to predict hydrological variables. Moreover,
51 the statistical models are often combined with each other to improve the model performance,
52 for example, Wang et al (2019) combined the gray model with BP neural network model to
53 predict runoff and used Markov chain to correct the results. However, these statistical models
54 can not exhibt the characteristic changes of variables with the multiscales, so it also can not
55 reveal their relations between variables and their influencing factors at the micro level.

56 Obviously, the knowledge of the driving factors makes to understand the evolution law of
57 urban rainstorm flood well. Meanwhile, constructing the statistical prediction model of urban
58 rainstorm flood can quantify the relationship between flood volume and its main influencing
59 factors, thus reflect the overall impact of the changing environment on the urban rainstorm
60 flood. Moreover, it is important to predict flood volume at the macro level, but the physical
61 mechanism between flood volume and its influencing factors can be better reflected by
62 predicting flood volume at the micro level, which can be achieved by combining with wavelet
63 analysis method. Therefore, the innovation of this paper is to construct the multiscale-

multivariate prediction model of urban rainstorm flood by combining wavelet analysis method and statistical models, thus the model can not only predict the urban flood at the macro and micro levels, but also quantify the multiscale attribution characteristics of flood volume to reflect the overall impact of the changing environment on urban rainstorm flood.

The objective of this paper is to determine the main influencing factors of flood volume firstly. Secondly, the rainfall and flood volume are decomposed by the wavelet analysis method to perform the multiscale attribution analysis. Thirdly, the prediction model for each flood volume component is constructed, then the multiscale-multivariate prediction model of urban rainstorm flood is constructed by superimposing them. Finally, the model performance including simulation effect and prediction effect is evaluated.

2. Materials and Methods

2.1 Study Area

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 (Sun et al. 2017). The Jialu River Basin is 929 km^2 in Zhengzhou central urban area, accounting for 94% of the total area of Zhengzhou central urban area (Zhang et al. 2019).

In the upstream of the Jialu River, there are Changzhuang reservoir, Jiangang reservoir and Hewang reservoir, thus it has no other natural water source except the main river course of Jialu River. In addition, the rainwater in Zhengzhou central urban area is mainly discharged into the nearby rivers, then with the Dongfeng Canal, Jinshui River, Xiong'er River, Qili River and other tributaries, the rainwater flows into the Jialu River (Zhou et al. 2015). The river system map of

the study area is shown in Figure 1.

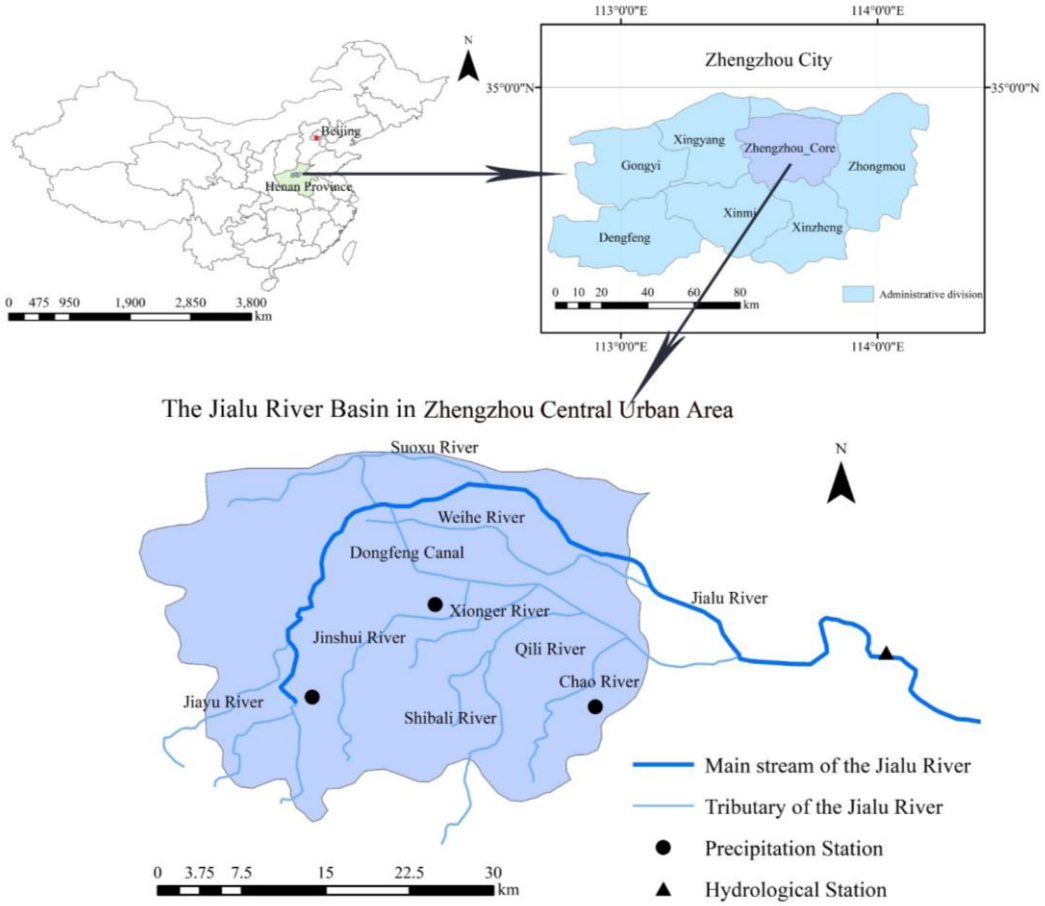


Figure 1. River system map of the study area

2.2 Data Source

The hourly rainfall data in this paper are from Zhengzhou Precipitation Station, Jiangang Precipitation Station and Sizhao Precipitation Station, while the flood data are from Zhongmu Hydrological Station. The geographical location of each station is shown in Figure 1. The selected rainstorm flood must meet the following two conditions: (1) The rainfall meets the standard of rainstorm magnitude in Table 1(Yang et al. 2014). For the two consecutive hours, its rainfall is less than 0.1mm, thus it is considered as the end of a rainstorm (Shao et al. 2018). Also, the rainfall of three precipitation stations is basically the same. (2) The flood process line

96 is a single-peak curve with a complete water recession process.

97 **Table 1. The standard of rainstorm magnitude**

Rainfall duration/h	1	2	3	6	12	24
Rainfall/mm	>15	>20	>22	>25	>30	>50

98 50 rainstorm floods are screened out from 1983 to 2018, and the corresponding rainfall is
99 the arithmetic average of the observed values at three precipitation stations. In this paper, the
100 straight-line segmentation method which connects the starting point on the rising process to the
101 turning point estimated on the recession process is used to divide flood process, and the
102 corresponding flood volume is above the connection line (Xu et al. 2011). Moreover, eight
103 phases (1980, 1990, 1995, 2000, 2005, 2010, 2015 and 2018) of the land use/cover data of study
104 area are obtained from the China Resources and Environmental Science and Data Center.

105 **2.3 Wavelet Analysis Method**

106 The basic idea of wavelet transform is to use a cluster of wavelet functions to represent or
107 approximate a certain signal or function. The corresponding wavelet function $\psi(t)$ is a function
108 having a wave shape and limited but flexible length with a mean value that is equal to zero, and
109 is localized in both time and frequency domains (Seo et al. 2015; Pathak et al. 2016). For a
110 timeseries, $\psi(t)$ is generally defined as:

111 $$\psi_{a,b}(t) = |a|^{-1/2}\psi(\frac{t-b}{a}) \quad (1)$$

112 where t represents time; b is the translation factor (time shift) of the wavelet over the time series,
113 and a ranging from 0 to +∞ denotes the wavelet scale (scale factor). For a given energy-limited
114 signal $f(t) \in L^2(R)$, the continuous wavelet transform (CWT) is defined as:

$$W_f(a,b) = |a|^{-1/2} \int f(t)\ \ \bar{\psi}(\frac{t-b}{a})dt \quad (2)$$

where $\bar{\psi}\left(\frac{t-b}{a}\right)$ is the complex conjugate function of $\psi(\frac{t-b}{a})$. In real hydrological problems, the time series are usually in the discrete format rather than continuous format (Roushangar et al. 2018), therefore, the discrete wavelet transform (DWT) in the following form is usually used:

$$W_f(a,b) = |a|^{-1/2} \Delta t \sum_{k=1}^{N} f(k\Delta t)\ \ \bar{\psi}(\frac{k\Delta t-b}{a}) \quad (3)$$

where Δt is the sampling interval.

2.4 Linear Regression Model and Ridge Regression Model

The core of the linear regression model is to establish a linear relationship between the dependent variable and one or more independent variables (Liu et al. 2016). When the number of independent variables is p and the number of samples is n, the linear regression model is formulated as follows:

$$Y = X\beta + \varepsilon \quad (4)$$

where Y ($n\times1$) is the vector of dependent variable, X ($n\times p$) is the regression matrix, ε ($n\times1$) is the vector of random error terms after removing the influence of independent variable on dependent variable, and β ($p\times1$) is the parameter vector which can be estimated by the least square method with the following equation (Zhao et al. 2020):

$$\beta = (X^T X)^{-1} X^T Y \quad (5)$$

When there is a good correlation between independent variables, the least squares estimator may lead to the ill-conditioned problem, that is, the calculated optimal parameter does not match the actual situation. The ridge regression model can effectively handle this problem.

The ridge regression is a biased estimation regression method for multicollinearity data

136 analysis. The motivation for the ridge estimator is to add a constant matrix kI (k>0) to the matrix
137 X^TX, which greatly reduces the probability of X^TX+kI approaching singularity (Rabiei et al.
138 2019). Therefore, the parameters of the ridge regression model can be obtained with the
139 following equation:

140 $$\hat{\beta} = \left(X^T X + kI_p\right)^{-1} X^T Y \quad (6)$$

141 where k≥0 is the ridge parameter and I_p is the p-dimensional identity matrix (Choi et al. 2019).

142 **2.5 Structure of the Combined Prediction Model**

143 The relationship between rainfall and flood volume is analyzed by double cumulative
144 curve method and T-test method, then the main influencing factors of flood volume can be
145 determined by combining the analysis of land use/cover change. Meanwhile, the sequences of
146 rainfall and flood volume are decomposed into several components by wavelet transform to
147 perform the multiscale attribution analysis of flood volume. According to the impact of each
148 influencing factors on the flood volume at different scales, the cumulative linear regression
149 model and ridge regression model are introduced to construct the prediction model for each
150 flood volume component. Finally, the multiscale-multivariate prediction model of urban
151 rainstorm flood is obtained by superimposing all component prediction models. The process for
152 constructing the combined prediction model is shown in Figure 2.

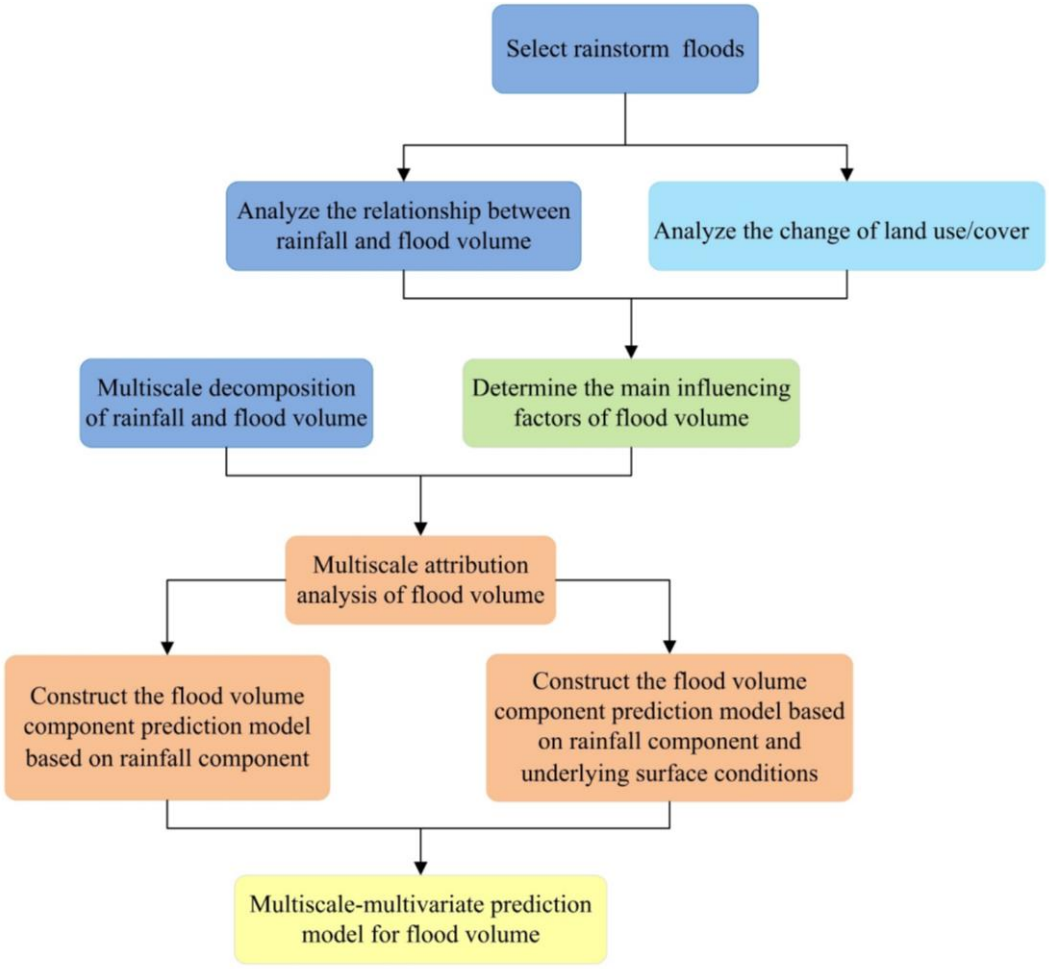


Figure 2. The process for constructing the combined prediction model

The Nash–Sutcliffe efficiency (*NSE*), the coefficient of determination (R^2) and the mean relative error (R_e) are used to evaluate the performance of the model. The closer the values of *NSE* and R^2 to 1.0 and the closer the value of R_e to zero, the better the performance of model can be achieved. These performance indexes can be written as:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i')^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (7)$$

$$R^2 = \left[\frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)(y_i' - \bar{y}_i')}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2 \sum_{i=1}^{n}(y_i' - \bar{y}_i')^2}}\right]^2 \quad (8)$$

$$R_e = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i' - y_i}{y_i}\right| \times 100\% \quad (9)$$

where y_i and y_i' are the i-th observed and simulated value; $\bar{y}_i$ and $\bar{y}_i'$ are the average

observed and simulated value respectively, n is the number of observations.

3. Results and Discussion

3.1 Analysis of Relationship Between Rainfall and Flood Volume

The rainfall and flood volume of 50 rainstorm floods are shown in Figure 3. It can be seen from Figure 3 that the rainfall and flood volume of the 12th to 25th rainstorm floods are smaller, while those of other rainstorm floods fluctuate around the mean value. Moreover, the fluctuation trends of rainfall and flood volume are basically the same with the correlation coefficient of 0.876. In order to further explore the changes in the relationship between them, the double cumulative curve of rainfall and flood volume is constructed as shown in Figure 4.

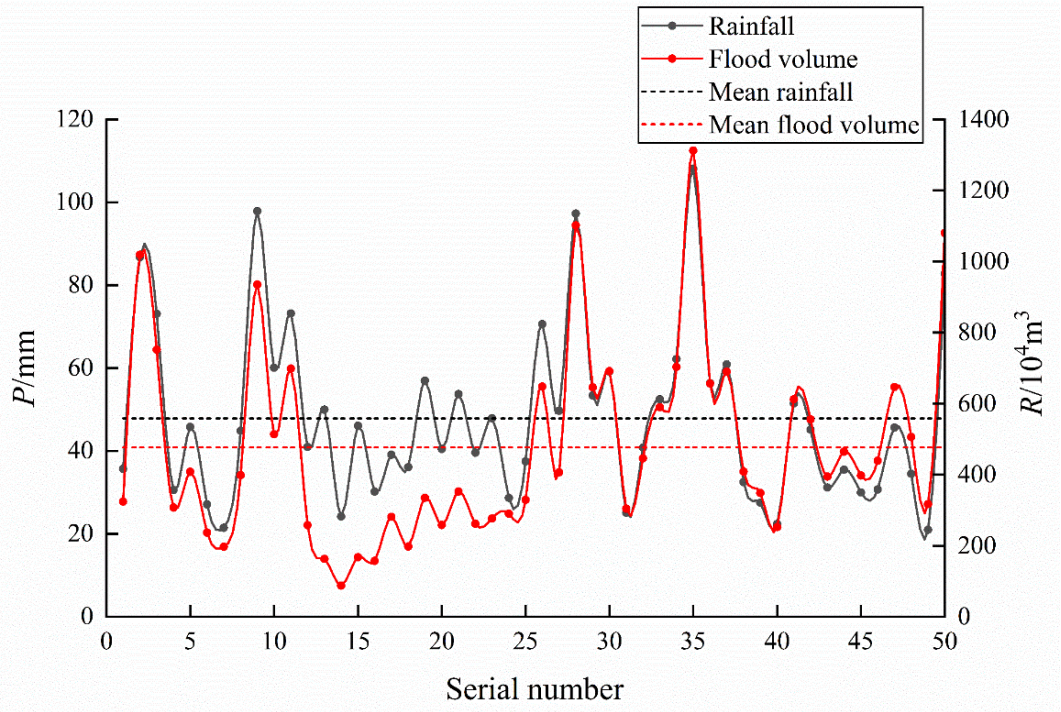
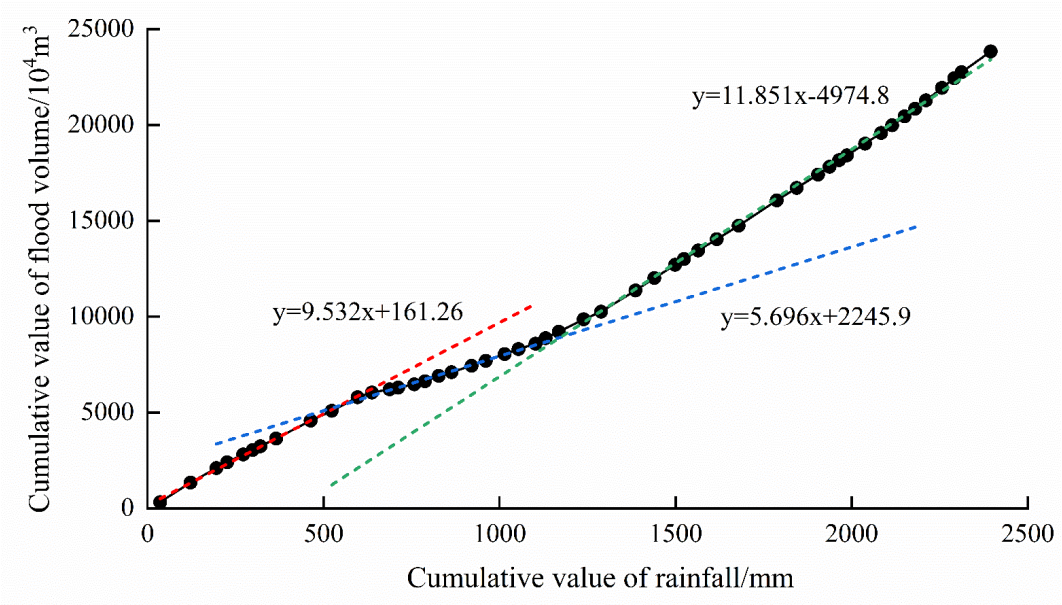


Figure 3. Rainfall and flood volume of 50 rainstorm floods



174

175 **Figure 4. Double cumulative curve of rainfall and flood volume**

176 According to Figure 4, it is obvious that the slope of the double cumulative curve changes
177 in the 12th and 24th rainstorm floods, dividing the 50 rainstorm floods into three periods
178 including the 1st rainstorm flood (September 22, 1983) ~ the 11th rainstorm flood (April 30,
179 1993), the 12th rainstorm flood (April 18, 1994) ~ the 23rd rainstorm flood (June 29, 2003),
180 the 24th rainstorm flood (June 21, 2005) ~ the 50th rainstorm flood (August 18, 2018).
181 Assuming that these two rainstorm floods are mutation points of the rainfall-flood volume
182 relationship, the significance T-test is respectively performed on the rainfall and flood volume,
183 the significance level is 0.01 and the results are shown in Table 2. If the absolute value of T
184 exceeds the critical value, it means that the significance test is passed.

185 **Table 2. Significance T-test of hypothetical mutation point**

No.	n_1	n_2	T value of rainfall	T value of flood volume	$t_{a/2}(a=0.01)$
12	11	12	1.530	3.452	2.831
24	12	27	-0.863	-4.310	2.715

It can be seen from Table 2 that the flood volume decreases abruptly in the 12th rainstorm flood and increases abruptly in the 24th rainstorm flood, the mutation of the latter is more significant. Moreover, the rainfall does not change abruptly in these two rainstorm floods. Therefore, the 12th and 24th rainstorm floods are mutation points of the relationship between rainfall and flood volume, that is, the characteristics of runoff generation and concentration in the study area change significantly in 1994 and 2005.

3.2 Determination of the Main Influencing Factors of Flood Volume

The land use/cover of study area is divided into 6 categories: cultivated land including dry land and paddy field, forest land, grassland, waters, impermeable land and unutilized land (shown in Figure 5), where the unutilized land mainly includes waste grassland, saline-alkali land, swamp, sand, bare land, bare rock, etc. Furthermore, the area of different underlying surface types is shown in Figure 6 (the unutilized land area is so small that it can be ignored).

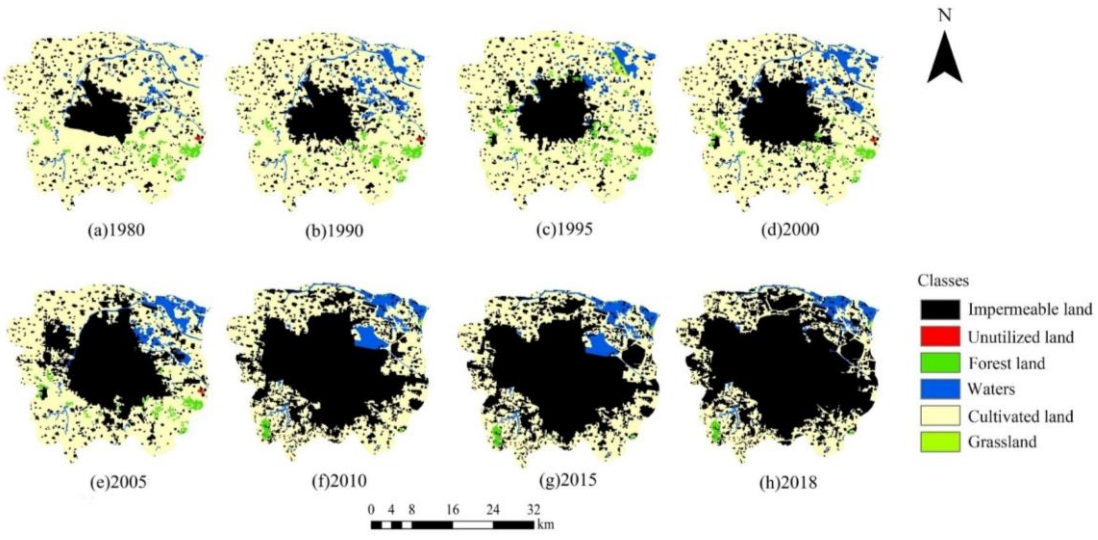


Figure 5. Land use distribution of study area in 1980, 1990, 1995, 2000, 2005, 2010, 2015 and 2018

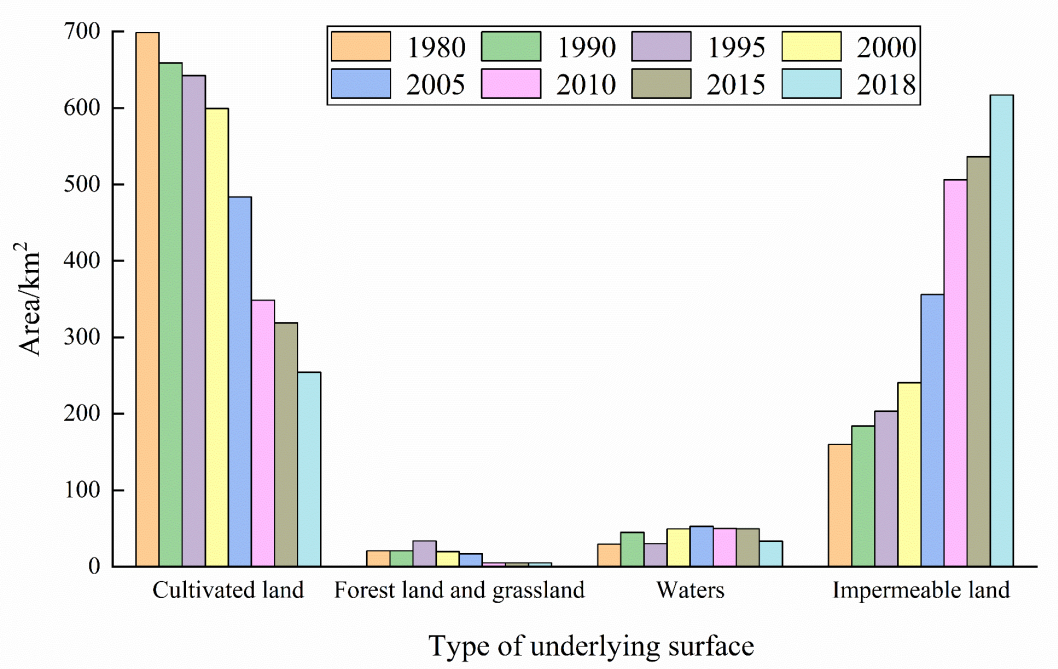


Figure 6. Area of different underlying surface types

From Figure 5 and Figure 6, it can be seen that the cultivated land area decreases continuously from 1980 to 2018, with the percentage decreasing from 76.8% to 28.0%, while the impermeable land area increases continuously, with the percentage increasing from 17.6% to 67.8%. Moreover, the area of forest and grassland accounts for 2.3% of the total area in 1990, increases to 3.7% in 1995, but then continues to decrease to 0.6 % in 2010. The water area basically remains unchanged from 2000 to 2015, with the percentage of 5.5%. The area changes of different underlying surface types in different periods are shown in Table 3.

Table 3. Area changes of different underlying surface types in different periods

(km^2/a)

Period/Type	Cultivated land	Forest and grassland	Water	Impermeable land
1980~1990	-3.95	0.00	1.56	2.40
1990~1995	-3.29	2.56	-3.00	3.87

1995~2000	-8.64	-2.81	3.87	7.46
2000~2005	-23.15	-0.54	0.67	23.03
2005~2010	-27.04	-2.36	-0.53	30.06
2010~2015	-5.94	0.00	-0.09	6.03
2015~2018	-21.44	0.00	-5.48	26.92

Table 3 shows that both cultivated land area and impermeable land area change sharply in the two periods of 2000~2010 and 2015~2018, while change slightly from 1980 to 1995. The water area changes most sharply from 2015 to 2018 and changes slightly from 2000 to 2015, the area of forest and grassland changes sharply in the two periods of 1990~2000 and 2005~2010, and basically remains unchanged in other periods.

Theoretically, the flood volume is affected by rainfall elements such as rainfall, rainfall duration and rainfall intensity. According to the selection principle of rainstorm floods and the analysis of the relationship between rainfall and flood volume, the rainfall is selected as the main influencing factor of flood volume among rainfall elements in this paper.

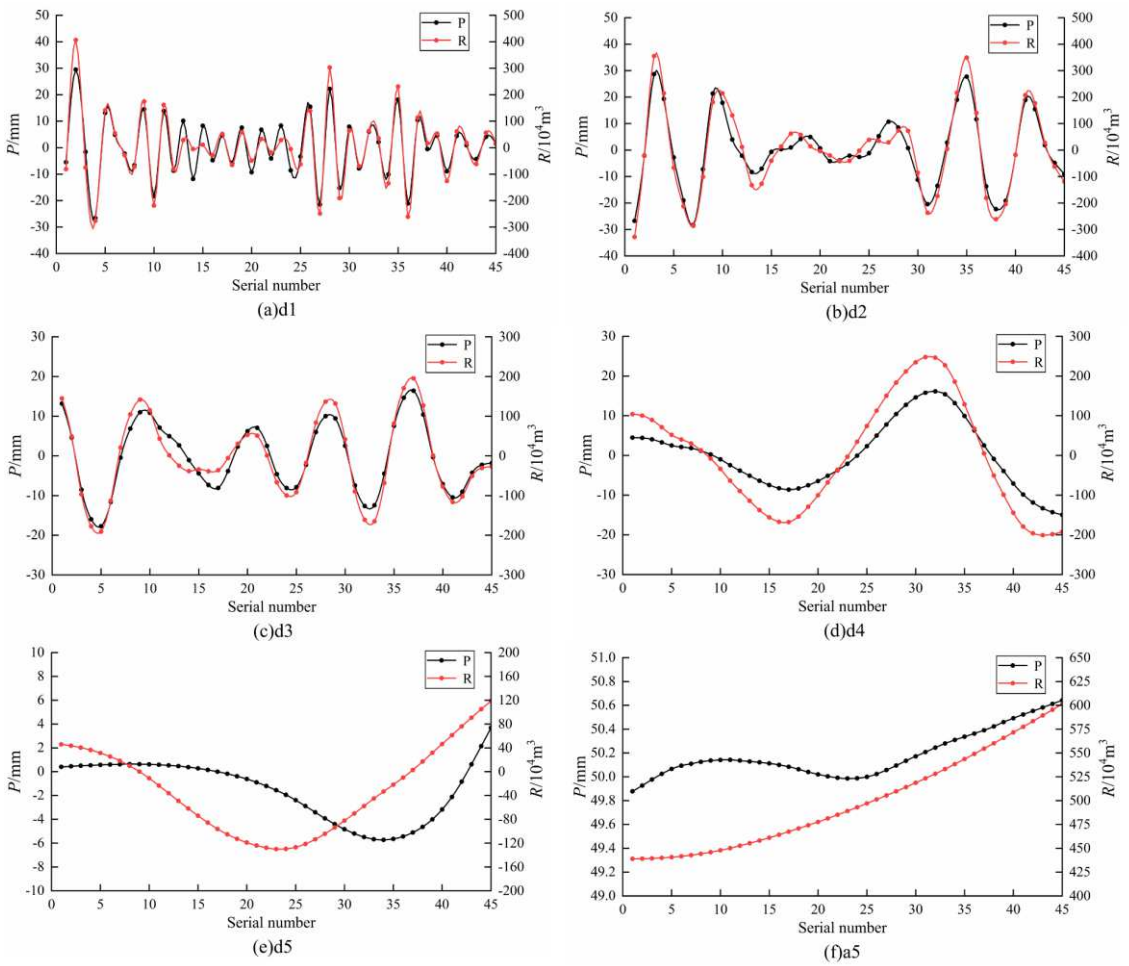
Furthermore, the increase of forest and grassland area or water area will lead to the decrease of runoff coefficient, reducing the flood volume generated by rainstorm of the same magnitude, while the increase of impermeable land area will lead to the opposite result. Therefore, according to the analysis of land/use change, it can be concluded the main reason for the mutation of the rainfall-flood volume relationship in 1994 is the change of forest and grassland area and water area under slight change of impermeable land area. Similarly, the main reason for the mutation of the rainfall-flood volume relationship in 2005 is the rapid

urbanization of Zhengzhou City and the rapid increase of impermeable land area after 2000.

The above conclusion is consistent with the previous study on the annual runoff coefficient in the study area (Wang et al. 2017), that is, the underlying surface is the main driving factor for urban rainstorm flood evolution. Therefore, the rainfall and underlying surface are the main influencing factors of flood volume in the Jialu River Basin of Zhengzhou central urban area.

3.3 Multiscale Attribution Analysis of Flood Volume

In this paper, db6 wavelet (Daubechies wavelet of order 6) is selected to decompose the rainfall sequence and flood volume sequence of the first 45 rainstorm floods, then five detail components and one trend component are obtained respectively, that is d1, d2, d3, d4, d5 and a5 (shown in Figure 7). The original sequence consists of these six component sequences, and a5 reflects the overall change trend of the original sequence.



239

240 **Figure 7. Components of rainfall and flood volume for the first 45 rainstorm floods**

241 It can be seen from Figure.7 (a) that both P_{d1} and R_{d1} have fluctuation periods of quasi-
242 2~4; (b) shows that both P_{d2} and R_{d2} have fluctuation periods of quasi-5~7; (c) shows that both
243 P_{d3} and R_{d3} have fluctuation periods of quasi-8~12; (d) shows that R_{d4} has fluctuation period of
244 quasi-26; (f) shows that the sequence of rainfall shows a slight change trend of increase-
245 decrease-increase, while the sequence of flood volume shows a more obvious increase trend on
246 the whole. In addition, the fluctuation trends of rainfall and flood volume are basically the same
247 in d1, d2, d3 and d4.

248 In order to reasonably construct the multiscale prediction model to connect rainfall and
249 flood volume at the micro level, the correlation coefficients between rainfall and flood volume

in each component are calculated as 0.957, 0.977, 0.974, 0.983, 0.319, 0.872, respectively. In other words, the rainfall and flood volume are positively correlated and well correlated in d1, d2, d3 and d4, while the correlation between the two variables is the worst in d5. Combining with Figure 7, d4, d5 and a5 components are superimposed to obtain the trend component a3 to improve the prediction accuracy of the model.

Since the decomposed sequences including d1, d2 and d3 have several negative values, the six data sequences are equivalently replaced for ease of calculation, that is, each item of original sequence subtracts the minimum value of the sequence to obtain new sequences including P_{b1}, P_{b2}, P_{b3}, R_{b1}, R_{b2} and R_{b3}. Then the double cumulative curves of P_{b1}-R_{b1}, P_{b2}-R_{b2}, P_{b3}-R_{b3} and P_{a3}-R_{a3} are respectively constructed (shown in Figure 8).

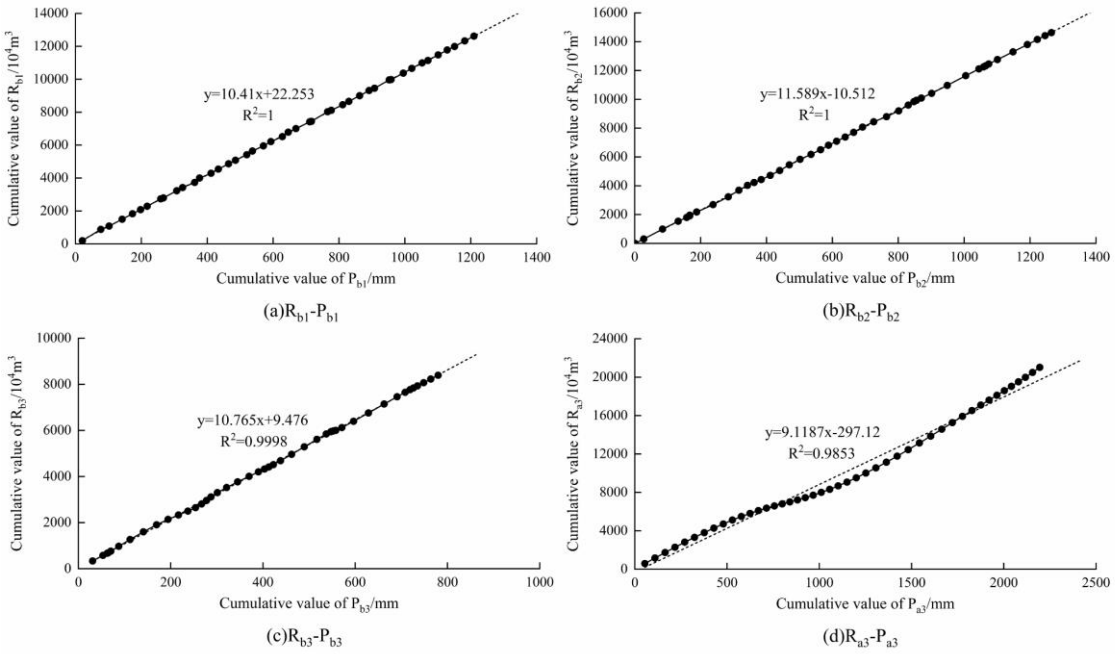


Figure 8. Double cumulative curve of rainfall and flood volume in each component

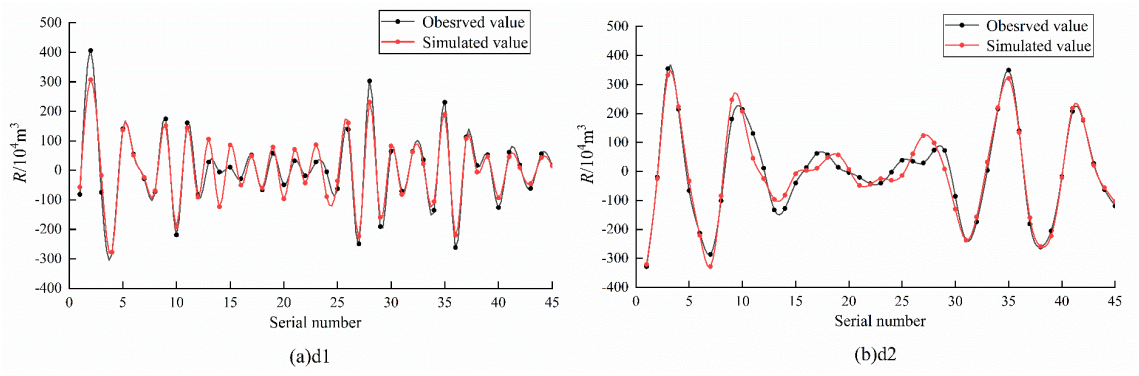
It can be seen from Figure 8(a), (b), (c) that the linear trend line basically coincides with the double cumulative curve, that is, there is a constant linear relationship between rainfall and

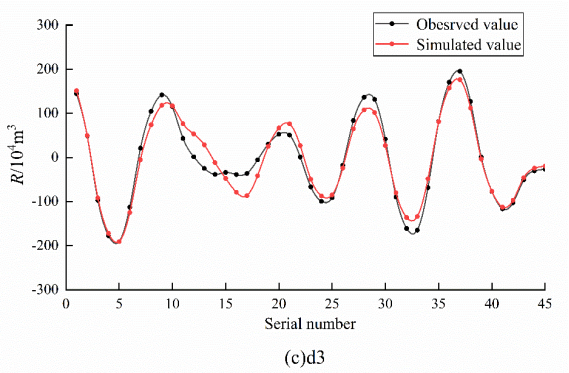
flood volume, while the change of underlying surface has little influence on flood volume in d1, d2 and d3. Moreover, Figure 8(d) shows that the slope of the double cumulative curve changes significantly in the 12th and 24th rainstorm floods, which is consistent with the analysis of the rainfall-flood volume relationship. Therefore, the impact of underlying surface on flood volume is mainly reflected in a3.

3.4 Construction of Prediction Model for Each Flood Volume Component

According to the multiscale attribution analysis of flood volume, the univariate model between flood volume and rainfall should be constructed for d1, d2 and d3 components of flood volume, while the multivariate model between flood volume, rainfall and various underlying surface areas should be constructed for a3 component of flood volume.

The cumulative linear regression models of R_{b1}-P_{b1}, R_{b2}-P_{b2} and R_{b3}-P_{b3} are respectively constructed, as shown in the trend line equation in Figure 8 (a)(b)(c), then the simulated value of R_{d1}, R_{d2} and R_{d3} can be obtained by restoring the simulation result. The comparison between the simulated and observed flood volume in d1, d2 and d3 components is shown in Figure 9.





279

280 **Figure 9. Comparison between the simulated and observed flood volume in d1, d2 and**

281 **d3 components**

282 The characteristics of runoff generation and concentration of the basin can be characterized

283 by the rainfall-flood volume double cumulative curve slope, which is mainly affected by the

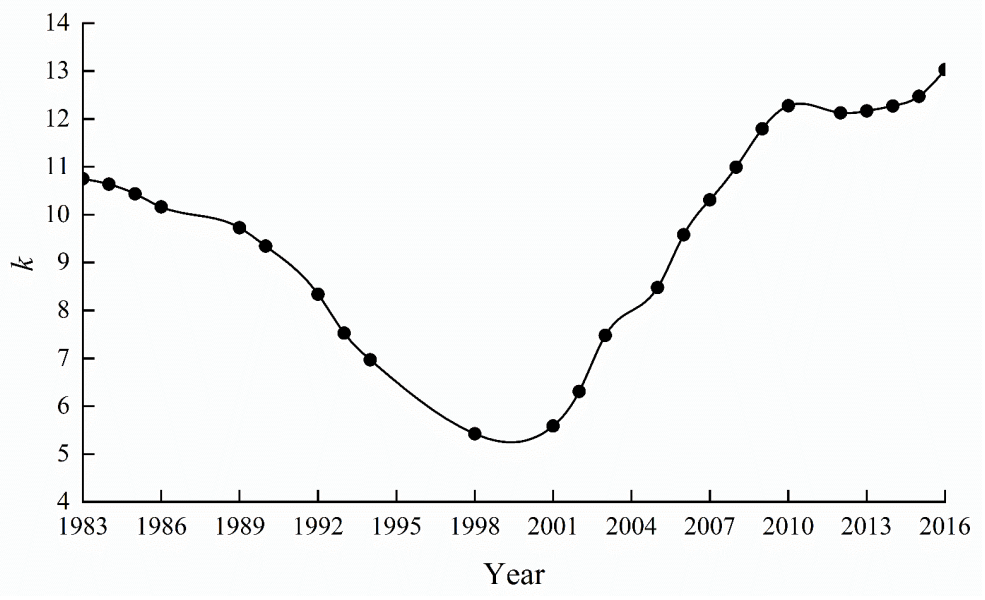
284 underlying surface, thus, the slope of two adjacent points on the P_{a3}-R_{a3} double cumulative

285 curve is calculated. Since the slopes of different rainstorm floods in the same year are not

286 significantly different, their mean value is represented by *k* and taken as the slope in the given

287 year, as shown in Figure 10 (where rainstorm floods are not screened out in some years such as

288 1995 and 2004).



289

290 **Figure 10. *k* value in each year**

291 It can be seen from Figure 10 that the *k* value shows a decreasing trend from 1983 to 2000,
292 which is mainly influenced by the change of water area and forest and grassland area; the *k*
293 value shows an increasing trend after 2000, which is mainly influenced by the change of
294 impermeable land area. Therefore, the model between the *k* value and various underlying
295 surfaces area should be constructed before 2000 and after 2000, respectively. Moreover, the
296 sum of cultivated land area(S1), forest and grassland area(S2), water area(S3) and impermeable
297 land area(S4) is basically a constant value, thus, S1 is discarded to avoid strong collinearity
298 among the independent variables and up to other three variables are selected as independent
299 variables.

300 Taking the year after 2000 as an example, the correlation coefficients of S2, S3 and S4 are
301 0.53, -0.98 and -0.55 by combining Table 3 and assuming that S2, S3 and S4 change evenly in
302 the periods of 2000~2005, 2005~2010, 2010~2015 and 2015~2018, that is, there is a strong
303 collinearity among the three. Based on the analysis of land use/cover change, SPSS software is
304 used to respectively construct the linear regression model dependent on S4 for *k* value (Model
305 1), the ridge regression model dependent on S2 and S4 for *k* value (Model 2), the ridge
306 regression model dependent on S2, S3 and S4 for *k* value (Model 3). The equations of these
307 three models are as follows.

308 $$k = (0.241S_4 - 1.782)/10 \quad (10)$$

309 $$k = (-1.444S_2 + 0.088S_4 + 80.018)/10 \quad (11)$$

310 $$k = (-1.404S_2 - 0.66S_3 + 0.086S_4 + 113.949)/10 \quad (12)$$

311 Similarly, for the year before 2000, the linear regression model dependent on S2 for *k* value

(Model 4), the linear regression model dependent on S3 for k value (Model 5) and the ridge regression model dependent on S2 and S3 for k value (Model 6) are respectively constructed. The equations of these three models are as follows.

$$k = (-3.677S_2 + 176.36)/10 \quad (13)$$

$$k = (-0.594S_3 + 112.01)/10 \quad (14)$$

$$k = (-4.595S_2 - 2.386S_3 + 289.739)/10 \quad (15)$$

The simulation results of each model are shown in Figure 11 and the performance indexes of each model are shown in Table 4.

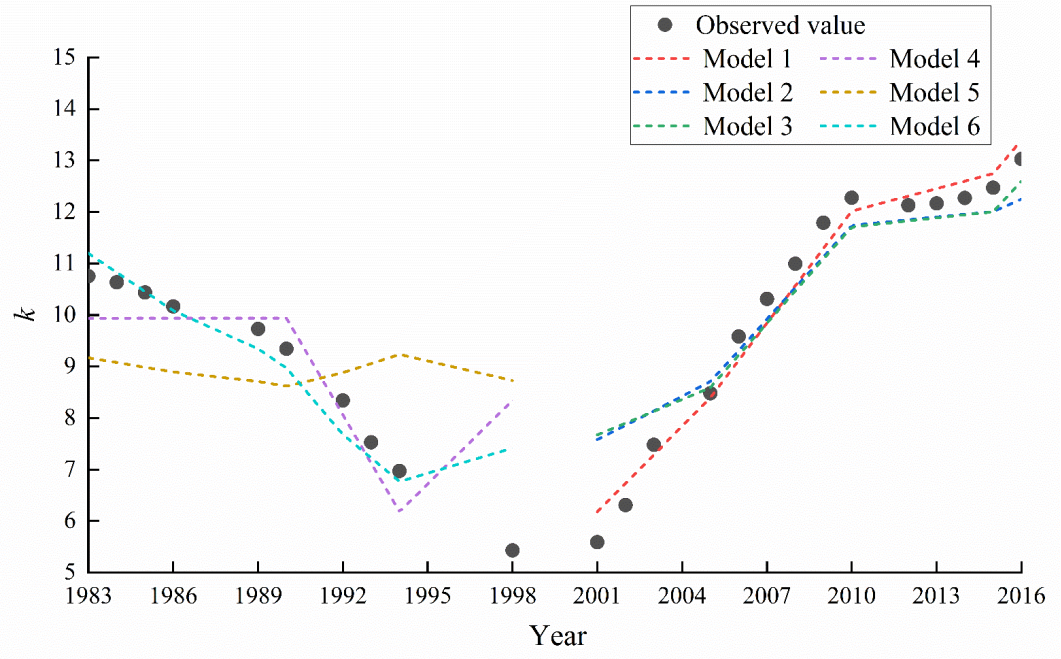


Figure 11. The simulation results of each model

Table 4. The performance indexes of each model

Index	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
NSE	0.976	0.886	0.881	0.616	0.013	0.827
R^2	0.976	0.965	0.951	0.616	0.013	0.829

R_e	3.70%	7.81%	7.92%	10.36%	19.41%	6.64%

From Figure 11 and Table 4, it can be seen that the Model 1 has the best simulation effect after 2000, while the Model 6 has the best simulation effect before 2000. Moreover, the Model 6 has a poor simulation effect on the *k* value in 1998, which may be caused by uneven change of the underlying surface area from 1995 to 2000. All in all, the multivariate model for a3 component of flood volume before and after 2000 are shown in Eq. (16) and Eq. (17), respectively. The simulated flood volume component is shown in Figure 12.

$$R_{a3} = (-4.595S_2 - 2.386S_3 + 289.739)P_{a3}/10 \quad (16)$$

$$R_{a3} = (0.241S_4 - 1.782)P_{a3}/10 \quad (17)$$

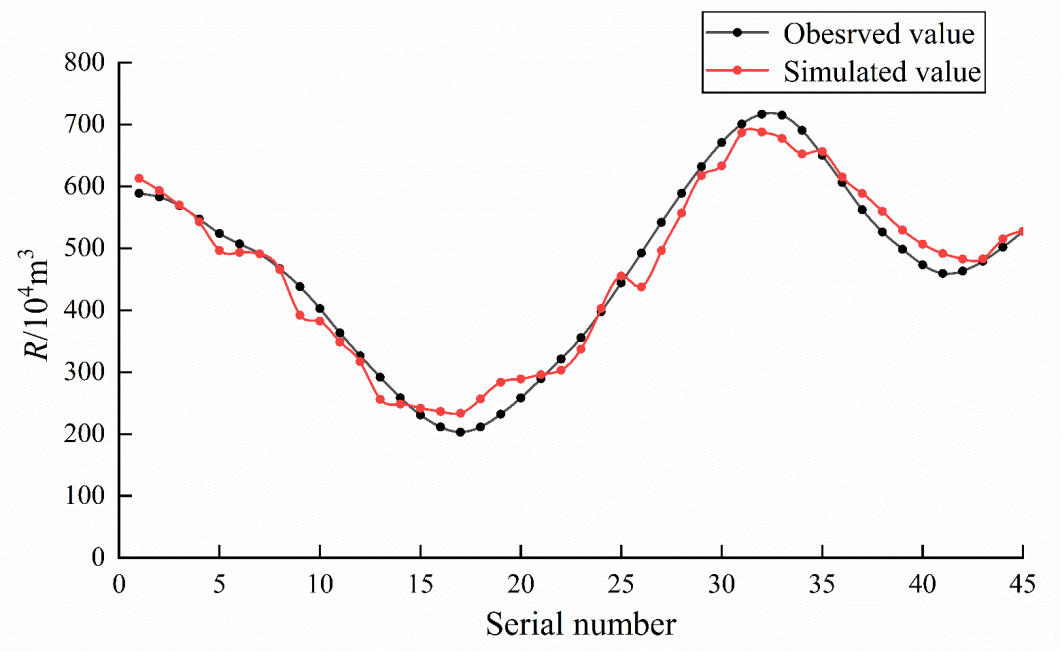


Figure 12. Comparison between the simulated and observed flood volume in a3 component

3.5 Effect Evaluation of the Multiscale-multivariate Prediction Model

The multiscale-multivariate prediction model of urban rainstorm flood can be obtained by

superimposing all the component prediction models. The simulated flood volume of the first 45 rainstorm floods is shown in Figure 13, where *NSE*, R^2 and R_e are 0.966, 0.964 and 10.80%, respectively. That is, the multiscale-multivariate prediction model has a good simulation effect on the flood volume of the first 45 rainstorm floods.

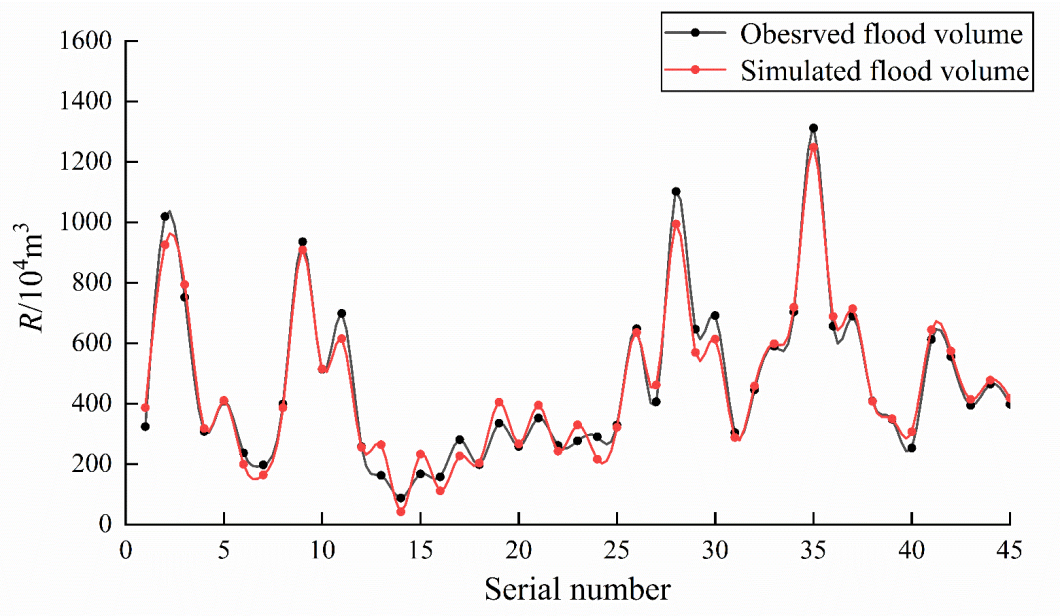


Figure 13. The simulated flood volume of the first 45 rainstorm floods

The rainfall of the 46th~50th rainstorm floods is decomposed and input into the model, in which the 46th~48th rainstorm floods occur in 2017 and the 49th~50th rainstorm floods occur in 2018. Then the flood volume can be predicted by combining the impermeable land area, the predicted flood volume of each rainstorm flood is shown in Table 5.

Table 5. The predicted flood volume of the 46th~50th rainstorm floods

No.	P_{d1}	P_{d2}	P_{d3}	P_{a3}	k	Predicted value	Observed value	Relative error
46	-3.76	-3.17	-2.63	40.26	14.04	461.2	439.9	4.84%

47	3.70	4.79	-4.28	41.49	14.04	630.6	647.0	-2.53%
48	-2.42	-3.38	-3.37	43.67	14.04	512.6	506.2	1.26%
49	-14.88	-10.38	-0.46	46.72	14.69	375.9	337.0	11.54%
50	24.28	3.80	4.02	50.70	14.69	1085.0	1080.6	0.41%

From Table 5, the relative errors between the predicted and observed flood volume of 46th~50th rainstorm floods are all less than 20%, indicating that the multiscale-multivariate prediction model has a good prediction effect on flood volume.

4. Conclusion

Using wavelet analysis method, cumulative linear regression model and ridge regression model, the multiscale-multivariate prediction model of urban rainstorm flood is constructed which can not only reveal the relationship between flood volume and its influencing factors at the micro level, but also predict flood volume at the micro and macro levels.

The analysis of rainfall-flood volume relationship and land use/cover change show that the main influencing factors of flood volume are rainfall and underlying surface, and the change of the underlying surface causes the mutation of characteristics of runoff generation and confluence in 1994 and 2005.

At the micro level, the fluctuation trend and period of rainfall-flood volume in d1, d2 and d3 are basically the same. The multiscale attribution analysis of flood volume shows that there is a constant linear relationship between rainfall and flood volume in d1, d2 and d3, while the change of underlying surface has little influence on flood volume. Moreover, the impact of underlying surface on flood volume is mainly reflected in a3.

The multiscale-multivariate prediction model has a good simulation effect on the flood volume of the first 45 rainstorm floods, *NSE*, R^2 and R_e are 0.966, 0.964 and 10.80%, respectively. Moreover, the model also has a good prediction effect, and the relative errors between the predicted and observed flood volume of the 46th~50th rainstorm floods are all less than 20%.

In summary, the model has a good performance in exploring the overall impact of the changing environment on urban rainstorm flood, quantifying the multiscale attribution characteristics of flood volume, and predicting the flood volume. At the same time, the model provides a new idea and method for flood prediction, and also provide an important basis for urban flood management.

Acknowledgments

This study is supported by the Natural Sciences Foundation of Henan Province (No. 212300410404), the Open Grants of the State Key Laboratory of Severe Weather (No. 2021LASW-A15), Scientific and Technological Research Program of Henan Province (No. 192102310508).

Availability of Data and Materials: The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Authors Contributions: Conceptualization, Jinping Zhang, and Yuhao Wang; Methodology,

Jinping Zhang, and Yuhao Wang; Validation, Jinping Zhang, and Yuhao Wang; Writing-Original Draft Preparation, Yuhao Wang; Writing-Review & Editing, Jinping Zhang, and Yuhao Wang; Supervision, Jinping Zhang.

Declarations

Ethical Approval: The authors will comply with all academic norms by the journal of Water Resources Management.

Consent to Participate: All authors agreed to join this research.

Consent to Publish: All authors agreed with the content and that all gave explicit consent to submit.

Competing Interests: There is no conflict of interest

References

Bisht DS, Chatterjee C, Kalakoti S, Upadhyay P, Sahoo M, Panda A (2016) Modeling urban floods and drainage using SWMM and MIKE URBAN: a case study. Natural Hazards 84(2):749-776.

Clark MP, Nijssen B, Lundquist JD, Kavetski D, Rupp DE, Woods RA, Freer JE, Gutmann ED, Wood AW, Brekke LD, Arnold JR, Gochis DJ, Rasmussen RM (2015) A unified approach for process-based hydrologic modeling: 1. Modeling concept. Water Resources Research 51(4):2498-2514.

Choi SH, Jung HY, Kim H (2019) Ridge Fuzzy Regression Model. International Journal of Fuzzy Systems 21(7):2077-2090.

Chu HB, Wei JH, Li JY, Qiao Z, Cao JW (2017) Improved Medium- and Long-Term Runoff

406 Forecasting Using a Multimodel Approach in the Yellow River Headwaters Region Based
407 on Large-Scale and Local-Scale Climate Information. Water 9(8).
408 Gong YW, Li XN, Zhai DD, Yin DK, Song RN, Li JQ, Fang X, Yuan DH (2018) Influence of
409 Rainfall, Model Parameters and Routing Methods on Stormwater Modelling. Water
410 Resources Management 32(2):735-750.
411 Ho JY, Lee KT (2015) Grey Forecast Rainfall with Flow Updating Algorithm for Real-Time
412 Flood Forecasting. Water 7(5):1840-1865.
413 Jamali B, Lowe R, Bach PM, Urich C, Arnbjerg-Nielsen K, Deletic A (2018) A rapid urban
414 flood inundation and damage assessment model. Journal of Hydrology 564:1085-1098.
415 Li JK, Zhang B, Li YJ, Li HE (2018) Simulation of Rain Garden Effects in Urbanized Area
416 Based on Mike Flood. Water 10(7).
417 Liu JY, Zhang Q, Deng XY, Ci H, Chen XH (2016) Quantitative analysis the influences of
418 climate change and human activities on hydrological processes in Poyang Basin. Journal
419 of Lake Sciences 28(2):432-443.
420 Madarang KJ, Kang JH (2014) Evaluation of accuracy of linear regression models in predicting
421 urban stormwater discharge characteristics. Journal of Environmental Sciences
422 26(6):1313-1320.
423 Nigussie TA, Altunkaynak A (2019) Modeling the effect of urbanization on flood risk in
424 Ayamama Watershed, Istanbul, Turkey, using the MIKE 21 FM model. Natural Hazards
425 99(2):1031-1047.
426 Pathak P, Kalra A, Ahmad S, Bernardez M (2016) Wavelet-Aided Analysis to Estimate Seasonal

427 Variability and Dominant Periodicities in Temperature, Precipitation, and Streamflow in
428 the Midwestern United States. Water Resources Management 30(13):4649-4665.
429 Peng HQ, Liu Y, Wang HW, Ma LM (2015) Assessment of the service performance of drainage
430 system and transformation of pipeline network based on urban combined sewer system
431 model. Environmental Science and Pollution Research 22(20):15712-15721.
432 Rabiei MR, Arashi M, Farrokhi M (2019) Fuzzy ridge regression with fuzzy input and output.
433 Soft Computing 23(23):12189-12198.
434 Ren MF, Xu ZX, Pang B (2021) Driving mechanisms of urban floods under the changing
435 environment: case study in the Wenyu River basin. Advances in Water Science 32(3):345-
436 355.
437 Roushangar K, Nourani V, Alizadeh F (2018) A multiscale time-space approach to analyze and
438 categorize the precipitation fluctuation based on the wavelet transform and information
439 theory concept. Hydrology Research 49(3):724-743.
440 Seo Y, Kim S, Kisi O, Singh VP (2015) Daily water level forecasting using wavelet
441 decomposition and artificial intelligence techniques. Journal of Hydrology 520:224-243.
442 Shao DN, Liu GS (2018) Up-to-date urban rainstorm intensity formulas considering spatial
443 diversity in China. Environmental Earth Sciences 77(14).
444 Sun J, Zhang R, Qin L, Zhu HX, Huang Y, Xue YG, An SQ, Xie XC, Li AM (2017)
445 Genotoxicity and cytotoxicity reduction of the polluted urban river after ecological
446 restoration: a field-scale study of Jialu River in northern China. Environmental Science
447 and Pollution Research 24(7):6715-6723.

448 Taormina R, Chau K, Sivakumar B (2015) Neural network river forecasting through baseflow
449 separation and binary-coded swarm optimization. Journal of Hydrology 529:1788-1797.
450 Wang HL, Wu ZN, Sun RC (2017) Effect of Urbanization in Zhengzhou on River Hydrological
451 Process in Jialu River Watershed. Science Technology and Engineering 17(31):316-321.
452 Wang WC, Li WJ, Xu DM, Li QM (2019) Runoff prediction based on GM-BP model calibration
453 against Markov chain. South-to-North Water Transfers and Water Science & Technology
454 17(5):44-49.
455 Wu GZ, Wang C (2020) Research on runoff grey prediction based of improved fitting. Yellow
456 River 42(10):34-36&41.
457 Wu XS, Wang ZL, Guo SL, Liao WL, Zeng ZY, Chen XH (2017) Scenario-based projections
458 of future urban inundation within a coupled hydrodynamic model framework: A case study
459 in Dongguan City, China. Journal of Hydrology 547:428-442.
460 Xu LL, Liu JL, Jin CJ, Wang AZ, Guan DX, Wu JB, Yuan FH (2011) Baseflow separation
461 methods in hydrological process research: A review. Chinese Journal of Applied Ecology
462 22(11):3073-3080.
463 Wang FQ, Huo FL (2010) Summary of research on medium and long-term hydrological
464 forecasting methods. Yellow River 32(3):25-28.
465 Yan HF, Wang CH, Wen P (2008) Overview of Studies on Distributed Hydrological Model.
466 Water Resources and Power 26(6):1-4.
467 Yang MN, Xu YP, Pan GB, Han LF (2014) Impacts of Urbanization on Precipitation in Taihu
468 Lake Basin, China. Journal of Hydrologic Engineering 19(4):739-746.

469 Yaseen ZM, Jaafar O, Deo RC, Kisi O, Adamowski J, Quilty J, El-Shafie A (2016) Stream-flow
470 forecasting using extreme learning machines: A case study in a semi-arid region in Iraq.
471 Journal of Hydrology 542:603-614.
472 Zhang JY, Song XM, Wang GQ, He RM, Wang XJ (2014) Development and challenges of urban
473 hydrology in a changing environment: Ⅰ: Hydrological response to urbanization.
474 Advances in Water Science 25(4):594-605.
475 Zhang KZ, Shen JQ, Han H, Jia YZ (2019) Urban River Health Analysis of the Jialu River in
476 Zhengzhou City Using the Improved Fuzzy Matter-Element Extension Model. Water 11(6).
477 Zhao SW, Liao J, Yu DL (2020) Model averaging estimator in ridge regression and its large
478 sample properties. Statistical Papers 61(4):1719-1739.
479 Zhou ZM, Liu JX, Guo W (2015) Comparison and Evaluation of Drainage Pattern and
480 Connectivity in Zhengzhou City. Yellow River 10:54-57.