Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Study on the Application of Clustering Method in the Determination of Uncertainty Parameters of SWMM Model

Chengshuai Liu
Zhengzhou University
Bingyan Ma
Zhengzhou University
Caihong Hu (hucaihong@zzu.edu.cn)
Zhengzhou University
Qiang Wu
Zhengzhou University
Yue Sun
Zhengzhou University
Xiang Li
Zhengzhou University

Research Article

Keywords: Urban flood simulation, SWMM, Parameter sample, Model uncertainty parameters, City functional area, K-means clustering algorithm

Posted Date: October 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-811017/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Study on the Application of Clustering Method in the Determination of Uncertainty Parameters of SWMM Model

Chengshuai Liu[1], Bingyan Ma[1], Caihong Hu[1, ✉], Qiang Wu[1], Yue Sun[1], Xiang Li[1]

1 School of Water Conservancy Science and Engineering Zhengzhou University, Zhengzhou, 450001, China

✉ Corresponding Author: Caihong Hu, School of Water Conservancy and Engineering, Zhengzhou University, 450001, China

Email address: hucaihong@zzu.edu.cn

Abstract

Storm Water Management Model (SWMM) is one of the most commonly used models in urban flood simulation. However, because the calibration and verification of the model's uncertainty parameters are extremely dependent on the measured flood data, it is difficult to apply the model in areas lacking data. This study proposes a parameter sample clustering method based on peer research results to determine the uncertainty parameters of SWMM, and compares the simulation results with the simulation results of the manual adjustment method based on measured data. The research shows that the Absolute error (*AE*), Relative error (*RE*), Nash efficiency coefficient (*NSE*), and Coefficient of determination (R^2) of the water depth simulated by the parameter sample clustering method are 0.040m, 9.08%, 0.949, 0.967 compared with the measured value, respectively. The value of *AE*, *RE*, *NSE*, and R^2 of the manual tuning method during the calibration simulation period are 0.066m, 15.95%, 0.881 and 0.924, respectively. Therefore, the parameter sample clustering method has a better simulation effect than

manual tuning method, and it can be further promoted in areas without flood data.

Keywords

Urban flood simulation; SWMM; Parameter sample; Model uncertainty parameters; City functional area; K-means clustering algorithm

1 Introduction

Urban flood disasters threaten the safety of people's lives and property and hinder the sustainable development of society and economy (Z. N. Wu et al. 2021; Xu et al. 2020). Just in the past, the 7·20 rainstorm and waterlogging event in Zhengzhou City shocked the world. The extreme rainstorm caused 292 deaths in the city and the economic loss was as high as 65.5 billion yuan（Hu et al. 2021）. Accurate flood forecasting is an important non-engineering measure for flood prevention and disaster reduction (J. Y. Wang et al. 2020; Fu et al. 2020). In terms of non-engineering measures, through real-time high-precision urban rain and flood models or systems, accurate and effective early warning and forecasting of urban storm and flood processes have become the key to reducing disaster losses (Kastridis et al. 2020; Sarauskiene et al. 2020). The measured precipitation-runoff data with high temporal and spatial resolution is the basis for reliable and accurate simulation and forecasting of urban rain and flood models (Luan et al. 2019). The number of existing urban hydrological stations, monitoring capabilities and station network density, especially the serious lack of real-time monitoring of urban pipeline flow, is far from meeting the needs of rain and flood warning and forecasting in the context of rapid urbanization (Fu et al. 2020). Because the calibration and verification problems in the modeling process of urban

45 storm flood model are directly related to the reliability of urban flood simulation
46 (Hapuarachchi et al. 2011; Li et al. 2020). How to conduct rapid and efficient flood
47 simulation in urban areas where data is lacking or without data has always been a
48 research hot-spot in urban rain and flood simulation and forecasting (Luan et al. 2017).
49 In the application of urban flood simulation, the SWMM model is one of the most
50 widely used representative models (Hu et al. 2020; Du, 2020). This study focuses on the
51 parameter selection process of the SWMM model. SWMM model parameters are
52 divided into deterministic parameters and uncertain parameters (de Souza et al. 2018;
53 Liwanag et al. 2018). Among them, deterministic parameters have clear physical
54 meanings. For example, the parameters of each sub-catchment include area,
55 characteristic width, impermeability, slope, etc. (Liwanag et al. 2018); uncertainty
56 Parameters such as depression storage volume in impervious area, depression storage
57 volume in permeable area, impervious area Manning coefficient, permeable area
58 Manning coefficient, etc (Mei, 2019). Uncertainty parameter model identification and
59 verification of uncertainty parameters is the key content of SWMM model application
60 (Mei et al. 2017).
61 In the actual application of the SWMM model, it is a recognized method to
62 calibrate and verify model parameters based on a large number of measured flood
63 events (Hu et al. 2020). However, this method needs to continuously adjust each
64 sub-catchment according to the flood results, which is very time-consuming and
65 complicated (Ma et al. 2020). In the absence of data, the method proposed by (Zaghloul,
66 1983) to calibrate the parameters of the SWMM model based on the internal connection

and mutual influence between the runoff calculation parameters of the system analytical model and the spatial distribution of the catchment area (Zaghloul, 1983). Liu (2009) pointed out an urban rainfall runoff model verification method based on runoff coefficients was put forward under conditions; The model parameter calibration method combining surface runoff field observations and flood investigations has also been affirmed (Luan et al. 2017); With the help of physical processes and rainfall-runoff evolution mechanism research, Fu et al. (2020) obtained a verification method by generalizing the geometry of stagnant water. These methods are mainly used to deduce the model parameters through parameter transplantation, correlation analysis, empirical formulas, etc., and have achieved good results, but the complexity of manual parameter adjustment is still unavoidable. Nor can it take into account the differences in the value of the same parameter under the underlying conditions of complex cities.

For this reason, this study proposes a new method for the value of SWMM parameters. First, combining the land use type and the nature of land planning to divide the sub-catchment with the nature of a single urban functional area. Then use the K-means clustering method, a large number of peers' existing research results, combined with the principle of single city functional area value of the same parameter to assign parameters to the model, and finally compare and verify with the actual measured value. This can avoid the cumbersome and complicated adjustment of model parameters repeatedly, and at the same time provide a new model construction method for areas with no data.

2 Methodology

As shown in **Figure 1,** the focus of this research is based on the SWMM model, and the principle of parameter values for a single urban functional area is proposed. Based on this principle, a large number of peer research results parameters are selected as cluster calculation samples to obtain urban flood simulation calculations. The value of the model parameter. This method is called parameter sample clustering method, and compared with the traditional manual parameter adjustment method, it aims to improve the efficiency of model building and simulation accuracy.

2.1 SWMM model

The SWMM model was developed by the U.S. Environmental Agency in 1971 and has been continuously improved (Xu et al. 2020). It is a widely used storm-water runoff simulation software (Gironas et al. 2010). It is widely used to simulate a single precipitation event in a city or long-term water volume, water quality simulation and sponge city research (X. Wu et al. 2017; Xiang et al. 2017; Zhang et al. 2017).

The SWMM model has the advantages of simple operation, short calculation time, and relatively reliable calculation results (van der Sterren et al. 2014), but it has disadvantages such as rough division of sub-catchments and complex parameter adjustment (Z. N. Wu et al. 2020). In particular, the calibration and verification of its uncertainty parameters need to rely on measured flood data (Hu et al. 2020), which makes it difficult for SWMM to be applied in urban areas without data.

SWMM parameters can be divided into deterministic parameters and uncertain parameters (As shown in **Table 1**). The deterministic parameters include eight

parameters: Sub-catchment area (SA), Characteristic width (CW), Impervious area rate (IAR), Slope (SL), Node elevation (NE), Pipeline shape (PS), Pipeline length (PL), and Pipeline inner bottom elevation (PBE). These parameters can be obtained by field measurement or software analysis. The uncertainty parameters of the SWMM model mainly include the Impervious area depression (DI), Impervious area Manning coefficient (NI), Permeable area depression (DP), Permeable area Manning coefficient (NP), Maximum infiltration rate (MAX), Minimum infiltration rate (MIN), Attenuation coefficient (DE) and Pipeline Manning Ning coefficient (MC) 8 parameters (Du. 2020).

The sensitivity of model uncertainty parameters is related to factors such as rainfall characteristics and underlying surface conditions (Gironas et al. 2010; Xu et al. 2020). Different studies have different values of DI, NI, DP, and NP. The surface depression storage volume reflects the depression storage depth of the hydrological response unit, and the surface Manning coefficient reflects the resistance of precipitation when it passes through the hydrological response unit (X. Wu et al. 2017; Xiang et al. 2017). Under the conditions of low slope and small topography in plain cities, surface depression storage and surface Manning coefficient become more significant sensitive parameters, and their correlation with urban surface characteristics is relatively large. Therefore, we select four uncertainty parameters DI, NI, DP, and NP for sample clustering (Li. 2016).

2.2 Divide the sub-catchment

The division of sub-catchments is one of the main steps of SWMM model modeling, and the quality of the division has a relatively large impact on the accuracy of

the results (Chen et al. 2018; de Souza et al. 2018; Gironas et al. 2010; Xu et al. 2020). We refer to previous research and refine the following division steps：1）The surface of the city has both natural and social attributes, and the two attributes are superimposed to divide the city's functions；2）Analyze the topography and confluence characteristics of the study area, determine the drainage area, and use the distribution of the drainage network trunk pipes and trunk roads as the division skeleton；3）Control the number and spatial scale, and the ratio of the number of pipe sections and nodes to the number of sub-catchments is mostly 0.6-1.4；4）Divide the sub-catchment area so that there is only one urban functional area in the hydrological response unit.

2.3 SWMM model parameter clustering calculation

The K-means clustering algorithm in the statistical analysis method is used to calculate the model parameters. It is an iterative solution clustering analysis algorithm (C. Wang et al. 2021). It often selects *K* objects randomly as the initial cluster centers, and then calculates the various sub-clusters of each object distance The distance between centers assigns each object to the cluster center closest to it (Liu et al.2015; Qian et al.2012).

In the process of clustering, determining the number of classes is a common problem faced by all clustering methods. The K-Means clustering analysis method selects the optimal number of classifications through analysis of variance, that is, defines an F statistic:

$$F = [S_{SA} / (k-1)] / [S_{SE} / (n-k)]$$

152
$$S_{SA} = \sum_{i=1}^{k} n_i (x_i - \overline{x_i})^2$$
$$S_{SE} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \overline{x_i})^2$$

153 In the formula, k is the number of clusters, n_i is the sample size of the i-th
154 class, S_{SA} is the sum of squared deviations between groups, S_{SE} is the sum of squared
155 deviations within the group, n is the total number of samples, and x_i is the size of the
156 i-th class. The sample value, $\overline{x_i}$ represents the sample average of the i-th class, and
157 x_{ij} represents the observation value of the j-th index of the i-th class.

158 According to the research purpose, the appropriate clustering index is selected,
159 and the sample data is standardized to eliminate the difference in dimensions. There are
160 m sample groups, and each sample group has p index data. These $m \times p$ data
161 constitute a parameter clustering observation matrix, namely:

162
$$x = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ & \dots. & & \\ & \dots. & & \\ & \dots. & & \\ x_{m1} & x_{m2} & \dots & x_{mp} \end{bmatrix}$$

163 Among them, $x_{i'j'}$ is the j'-th index parameter value of the i'-th sample group, i'
164 =1,2,..., m , j'=1,2,..., p ; then:

$$\overline{x_{j'}} = \frac{1}{m} \sum_{i'=1}^{m} x_{i'j'}$$
165
$$s_{j'} = \left[\frac{1}{m-1} \sum_{i'=1}^{m} (x_{i'j'} - \overline{x_{j'}}) \right]^{1/2}$$
$$y_{i'j'} = (x_{i'j'} - \overline{x_{j'}}) / s_{j'}$$

In the formula, $\overline{x_{j'}}$ is the mean value of the j'th index parameter; $s_{j'}$ is the standard deviation of the j'th index parameter; $y_{ij'}$ is the standardized value of the parameter value $x_{ij'}$.

2.4 Model evaluation Indicator Selection

The absolute error (AE) is the absolute value of the difference between the measured value and the true value. Here it refers to the absolute value of the difference between the simulated water depth of the model and the actual measured water depth (Hu et al. 2018). The mathematical expression is as follows:

$$AE = |S_i - O_i|$$

where S_i and O_i are the simulated and observed water depth at the i-th number.

Relative error (*RE*) refers to the value obtained by multiplying the ratio of the absolute error caused by the measurement to the measured (conventional) true value by 100%, expressed as a percentage. Generally speaking, the relative error can better reflect the credibility of the measurement. It is mostly used in the evaluation of urban rain and flood models to indicate the credibility of the simulated value of flood peak discharge (Hu et al. 2018). It is described as follows:

$$RE = \frac{S_i - O_i}{O_i} \times 100\%$$

where S_i and O_i are the simulated and observed water depth at the i-th number.

The mathematical expressions of these metrics can be described as follows (J.E. et al. 1970):

$$NSE = 1 - \frac{\sum_{i=1}^{n}(S_i - O_i)^2}{\sum_{i=1}^{n}(S_0 - \overline{O})^2}$$

where S_0 (m) and O_0 (m^3/s) represent the water depth of the simulated and observed hydrographs, respectively; $\overline{O}$ is the mean value of the observed water depth, and n is the data points number. *NSE* measures the ability of the model to predict variables different from the mean, gives the proportion of the initial variance accounted for by the model, and ranges from 1 (perfect fit) to -∞. Values closer to 1 provide more accurate predictions (Gupta et al. 1999).

The coefficient of determination (R^2) is often used to describe the degree of fit between data (Hu et al. 2018). When R^2 is closer to 1, it means that the reference value of the related equation is higher; on the contrary, when it is closer to 0, it means that the reference value is lower. It is described as follows:

$$R^2 = \frac{\left[\sum_{i=1}^{n}(S_i - \overline{S})(O_i - \overline{O})\right]}{\sum_{i=1}^{n}(S_i - \overline{S})^2 (O_i - \overline{O})^2}$$

where n is the total number of measured data, S_i is the simulated water depth for data point i, O_i is the measured water depth for data point i, and is the averaged value of the measured water depth (Jackson et al. 2019).

3 Results and discussion

3.1 Study area

The geographical coordinates of Zhengzhou City are 112°42'-114°14' E and 34°16'-34°58' N. It belongs to the continental monsoon climate in the northern temperate zone, with hot and rainy summers and autumns, cold and dry springs and

winters, with an average annual rainfall of 542 mm. The basic terrain tends to be low in the northwest and high in the southeast, with serious waterlogging during the flood season (J. Y. Wang et al. 2020). Zhengzhou University is located in the West Fourth Ring of Zhengzhou City. The climate characteristics of the school are consistent with those of Zhengzhou City. On the whole, the building pattern and drainage pipe network of Zhengzhou University are distributed in a gridded rectangle, which is very beneficial to the division of the sub-catchment of the model, as shown in **Figure 2.**

3.2 Database

According to research needs, we collected rainfall data from Henan Hydrological Bureau during 2014-2018; and obtained the water depth of the outlet node of Zhengzhou University through experimental monitoring. The land use data is visually interpreted by Google's high-definition image map and corrected on-site, with an accuracy of 1×1m. The pipe network data comes from the Infrastructure Management Office of Zhengzhou University. The elevation data is obtained through surveying and mapping. Uncertain parameter samples are derived from 26 research results published by peers.

Table 2 shows the rainfall information of 6 flood events. Corresponding to the rainfall data, we selected 6 flood events to calibrate the parameters of the model to compare the reliability of the manual tuning method and the parameter sample clustering method.

Table 3 is the sample data used to calculate the clustering parameters. There are 26 research results published by peers in total. Because this research is aimed at the

application of the SWMM model in China, the previous results of the selected peers are all in the research area of China.

3.3 Sub-catchment area division

According to the requirements of China's current "Outdoor Drainage Design Code", the pipeline network is generalized. The pipeline network is generalized into node elevation, pipeline shape, length, pipe diameter, etc., and combined with pipeline network distribution information, topography, and surface type information, the study area Divided into 29 sub-catchments, 37 nodes, 37 pipes and 1 total water outlet, of which the pipes and water outlets are free flow. Considering factors such as actual topography, land use type, and location of drainage pipe network to divide sub-catchment areas. Ensure that each sub-catchment is an independent urban functional area. As shown in **Figure 3.**

The division of land use types is an important prerequisite for the division of sub-catchments. The land use type of the study area is determined by visual interpretation of the 1×1m high-definition remote sensing method, and field inspections are used to make corrections, so as to ensure the division of sub-catchments. Reliability of the natural attributes of the water area. On the whole, Zhengzhou University’s architectural pattern and drainage pipe network are grid-shaped and rectangular. The drainage pipe network is relatively independent and systematic. This is very beneficial to the model’s sub-catchment division, which ensures the society of sub-catchment divisions. Attribute reliability.

3.4 Parameter clustering results

The urban surface has both natural and social attributes, and the two attributes are superimposed to divide the city's functions (Du. 2020). It is divided into three types: industrial and commercial areas (ICA), residential areas (RA), and public areas (PA).

Taking the values of these parameters in **Table 3** as samples, the K-means clustering algorithm is used to calculate the parameter values on different urban functional areas, the number of clusters *K* is set to 3. This also refers to the number of functional districts in the city and the parameters clustering results are shown in **Table 4**.

For other parameters, the calculation results of flood control and drainage in Zhengzhou City (Li et al., 2016) are used as the values of model parameters. Among them, MAX is 76.2mm/h; MIN is 3.6mm/h; DE is 3; MC is 0.014.

The cluster classification results of DI, NI, DP, and NP are shown in **Figure 4**. The three axes are composed of parameter values, sample size, and fixed width. The fixed width has no practical meaning. The light gray separation layer represents the dividing line of the sample parameter clustering, and the category number corresponding to the Roman numeral. The black arrows represent the cluster centers of each category. These four parameters are divided into three ranges, and the different categories are caused by the differences in the underlying surface conditions. The industrial and commercial area in the study area is mainly used for commercial and financial land. It is located in the center of the study area with convenient traffic and on both sides of the main traffic line (street) or at the corner intersection. The buildings are tall and dense, and the

hardened road surface is flat and densely distributed. The depression has the smallest storage capacity and the smallest surface Manning coefficient; residential areas are located outside or around the industrial and commercial districts, residential areas are fragmented, high- and low-differentiated, public facilities are complete, the layout is complete, and the environmental quality is good. At the same time, its building density and floor area ratio are controlled. It has both hardened roads and roofs, as well as a green area rate of about 30%. The surface depressions are moderately stocked and the surface Manning coefficient is moderate; the public areas are mainly comprehensive parks and specialized parks, with certain recreational facilities and service facilities. The green land, which has the comprehensive functions of improving the ecology, beautifying the city, and preventing and reducing disasters, has certain permeability and water storage capacity. Therefore, in the sub-catchment area, the surface depression has the largest storage capacity and the surface Manning coefficient is the largest.

3.5 Model parameter evaluation

The focus of this research is to compare the application effects of manual adjustment method and parameter sample clustering method in SWMM model for flood simulation. The difference is that the manual tuning method consists of two parts: parameter calibration and verification, while the parameter sample clustering method only has parameter clustering. Therefore, we selected 6 flood events of 20140524, 20150707, 20160605, 20170812, 20180801, 20180818 to compare the simulation results. Among them, in the manual debugging method, the three flood events of 20140524, 20150707, and 20160605 were used for calibration, and the three flood

293 events of 20170812, 20180801, and 20180818 were used for verification.

294 As shown in **Figure 5**. Orange represents Precipitation, and the black square
295 dotted line represents the measured water depth change process. The blue circular
296 dotted line represents the simulation result of the manual tuning method, and the red
297 triangle represents the simulation result of the parameter sample clustering method. As
298 a whole, it seems that the parameter sample clustering method is closer to the measured
299 process line.

300 Although it is difficult to simulate the water depth change process at the outlet of
301 the urban drainage network, the parameter sample clustering method and manual
302 adjustment method both show good simulation results (as shown in **Figure 5**). In the
303 simulation of the peak water depth period, the parameter sample clustering method is
304 closer to the measured value than the manual adjustment method, but both are smaller
305 than the measured value. In the retreat stage, the same parameter sample clustering
306 method is closer to the measured value than the manual adjustment method, but both
307 are larger than the measured value. Therefore, the parameter sample clustering method
308 has a better simulation effect than the manual tuning method.

309

310 The above research is a qualitative macro-analysis. **Table 5** quantitatively
311 compares the prediction performance of runoff depth of the SWMM model using
312 manual adjustment method and parameter sample clustering method. Four performance
313 indicators of *AE*, *RE*, *NSE* and R^2 were used to evaluate them respectively. When the
314 manual adjustment method is used, the *AE* and *RE* of the simulation results of the water

depth change process during the calibration period are both less than 0.07m, 15%, and the *NSE* is greater than 0.880; the *AE* and *RE* of the simulation results during the verification period are both less than 0.09m, 20%, and the *NSE* is greater than 0.850. This shows that the traditional model parameter debugging can still guarantee the simulation accuracy of the SWMM model. When using the parameter sample clustering method, the *AE* and *RE* of the simulation results of the water depth change process during the calibration period are both less than 0.06m, 12%, and the *NSE* are both greater than 0.910; the *AE* and *RE* of the simulation results during the verification period are both less than 0.06m, 11%, and the *NSE* is both Greater than 0.950. This shows that the parameter sample clustering method can significantly improve the simulation accuracy of the SWMM model. This conclusion can be better reflected in the difference between the performance evaluation indicators of the two methods. In the 6 flood events, when the SWMM model uses the sample parameter clustering method, the *AE* and *RE* are both decreasing, and the *NSE* is increasing.

Figure 6 is the correlation diagram of the simulated water depth change process and actual measured value of the SWMM model using two simulation methods. The abscissa is the actual measured value, and the ordinate is the simulated value. **Figure 6(a)** is the manual adjustment method. We can see that the water depth data The points are clustered near the 45° line, and the R^2 value is 0.924, but there are still some water depth data points showing scattered characteristics; **Figure 6(b)** is the simulation result of the parameter sample clustering method, and we can see that the water depth data points are more concentrated Is concentrated near the 45° line, and the R^2 value is 0.967,

which shows that the SWMM model has higher simulation accuracy when using the parameter sample clustering method to determine the model parameters.

We propose to improve the parameter value of the SWMM model based on the parameter value principle of a single city functional area. The research results show that the parameters determined by the sample parameter clustering method have higher accuracy when the SWMM model simulates urban flooding. In the process of using the SWMM model, the traditional manual parameter adjustment method takes a long time and cannot avoid the phenomenon of "different parameters with the same effect", which has obvious shortcomings. Of course, this does not mean that the manual tuning method is completely inapplicable. On the contrary, the parameter sample clustering method proposed in this study is essentially to obtain samples from a large number of peer-to-peer tuning results. Mining the data clustering characteristics of the same parameter, and interpreting this characteristic as the difference of urban functional areas, so as to better ensure the simulation ability and efficiency of the SWMM model. The research results can also be used in urban areas where there is no measured rainfall-runoff data.

4 Conclusion

This paper applies the SWMM model to urban flood simulation, and proposes a calculation method based on the clustering of parameter samples of a single city functional area, and the following conclusions are obtained:

1. The SWMM model has a good rainfall-flood simulation effect in urban areas. Through manual adjustment of parameters, during the application process, the average

values of *AE*, *RE* and *NSE* during the calibration period are 0.051m, 12.93%, and 0.888 respectively; the average of the verification period The average values of *AE*, *RE*, and *NSE* are 0.066m, 15.95%, and 0.881, respectively. Using the improved SWMM model using the parameter sample clustering method, the average values of *AE*, *RE*, and *NSE* are 0.040m, 9.08%, and 0.949, respectively; the various evaluation indicators are better than the manual adjustment method.

2. Using parameter sample clustering method to calculate the SWMM model parameters is more applicable than manual tuning method. Compared with the measured value, the R^2 value of the former is 0.959, and the R^2 value of the latter is 0.924. It shows that this method can replace the traditional manual parameter adjustment method, and can further carry out flood simulation in urban areas without data. In the process of constructing and using SWMM, the sample parameter clustering method is more efficient than the manual parameter adjustment method, which greatly simplifies the parameter selection process of SWMM. This is of great significance to the promotion and application of the model.

3. After calculating the same parameters through sample clustering, the values of the cluster centers are different. This difference may be caused by the difference in the underlying surface conditions. Among them, PA (road, hard block) has the largest DI, DP, NI, and NP; RA (residential area) has the second highest DI, DP, NI, and NP; ICA has the smallest DI, DP, NI, and NP. Therefore, if SWMM takes different parameter values in different urban functional zones, the model can reflect the spatial variation characteristics of the urban surface.

Authors Contributions

For this research paper with several authors, a short paragraph specifying their individual contributions was provided. Chengshuai Liu developed the original idea and contributed to the research design for the study. Bingyan Ma was responsible for data collecting. Caihong Hu provided guidance and contributed to the research design. Qiang Wu, Yue Sun, and Xiang Li provided some guidance for the writing of the article. All authors have read and approved the final manuscript.

Funding

This work was funded by Key projects of National Natural Science Foundation of China, grant number 51739009, National Natural Science Foundation of China, grant number 51979250.

Declaration of competing interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Availability of data and material

Not applicable. The data in this manuscript is also used in other ongoing research, so data and materials are not applicable.

References

Chen, W. J., Huang, G. R., Zhang, H., & Wang, W. Q. (2018). Urban inundation response to rainstorm patterns with a coupled hydrodynamic model: A case study in Haidian Island, China. Journal of Hydrology, 564, 1022-1035. Retrieved from <Go to ISI>://WOS:000445316200080.

403 doi:10.1016/j.jhydrol.2018.07.069
404 Du Xian. Research on the refined simulation of urban rainstorm and flood based on
405 SWMM model [D]. Zhengzhou University, 2020.
406 de Souza, B. A., Paz, I. d. S. R., Ichiba, A., Willinger, B., Gires, A., Amorim, J. C.
407 C., . . . Schertzer, D. (2018). Multi-hydro hydrological modelling of a complex
408 peri-urban catchment with storage basins comparing C-band and X-band radar
409 rainfall data. *Hydrological Sciences Journal-Journal Des Sciences*
410 *Hydrologiques, 63*(11), 1619-1635. Retrieved from <Go to
411 ISI>://WOS:000450274900003. doi:10.1080/02626667.2018.1520390
412 Gironas, J., Roesner, L. A., Rossman, L. A., & Davis, J. (2010). A new applications
413 manual for the Storm Water Management Model (SWMM). *Environmental*
414 *Modelling & Software, 25*(6), 813-814. Retrieved from <Go to
415 ISI>://WOS:000276030300014. doi:10.1016/j.envsoft.2009.11.009
416 Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of Automatic Calibration
417 for Hydrologic Models: Comparison with Multilevel Expert Calibration.
418 *Journal of Hydrologic Engineering, 4*(2).
419 Hapuarachchi, H. A. P., Wang, Q. J., & Pagano, T. C. (2011). A review of advances in
420 flash flood forecasting. *Hydrological Processes, 25*(18), 2771-2784. Retrieved
421 from <Go to ISI>://WOS:000294201200001. doi:10.1002/hyp.8040
422 Hu, C. H., Liu, C. S., Yao, Y. C., Wu, Q., Ma, B. Y., & Jian, S. Q. (2020). Evaluation
423 of the Impact of Rainfall Inputs on Urban Rainfall Models: A Systematic
424 Review. *Water, 12*(9). Retrieved from <Go to ISI>://WOS:000580334000001.

doi:10.3390/w12092484

Hu, C. H., Wu, Q., Li, H., Jian, S. Q., Li, N., & Lou, Z. Z. (2018). Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation. *Water, 10*(11), 16. Retrieved from <Go to ISI>://WOS:000451736300050. doi:10.3390/w10111543

Hu Yuwei, Li Hao, Hao Shuangyan, Aoki, Chen Kang, Liu Yupeng. The world pays attention to China's rainstorm rescue [N]. Global Times, 2021-07-23 (016).

J.E., N., & J.V., S. (1970). River flow forecasting through conceptual models part I: A discussion of principles. *Elsevier, 10*(3).

Jackson, E. K., Roberts, W., Nelsen, B., Williams, G. P., Nelson, E. J., & Ames, D. P. (2019). Introductory overview: Error metrics for hydrologic modelling - A review of common practices and an open source library to facilitate use and adoption. *Environmental Modelling & Software, 119*, 32-48. Retrieved from <Go to ISI>://WOS:000478965700003. doi:10.1016/j.envsoft.2019.05.001

Kastridis, A., Kirkenidis, C., & Sapountzis, M. (2020). An integrated approach of flash flood analysis in ungauged Mediterranean watersheds using post-flood surveys and unmanned aerial vehicles. *Hydrological Processes, 34*(25), 4920-4939. Retrieved from <Go to ISI>://WOS:000575128500001. doi:10.1002/hyp.13913

Li Shihao. Flood Risk Analysis and Waterlogging Simulation Research in Zhengzhou City[D]. Zhengzhou University, 2016.

Li, Q., Liang, Q. H., & Xia, X. L. (2020). A novel 1D-2D coupled model for

hydrodynamic simulation of flows in drainage networks. *Advances in Water Resources, 137*, 14. Retrieved from <Go to ISI>://WOS:000519527900013. doi:10.1016/j.advwatres.2020.103519

Liwanag, F., Mostrales, D. S., Ignacio, M. T. T., & Orejudos, J. N. (2018). Flood Modeling Using GIS and PCSWMM. *Engineering Journal-Thailand, 22*(3), 279-289. Retrieved from <Go to ISI>://WOS:000437090700021. doi:10.4186/ej.2018.22.3.279

Luan, Q. H., Fu, X. R., Song, C. P., Wang, H. C., Liu, J. H., & Wang, Y. (2017). Runoff Effect Evaluation of LID through SWMM in Typical Mountainous, Low-Lying Urban Areas: A Case Study in China. *Water, 9*(6), 21. Retrieved from <Go to ISI>://WOS:000404559900077. doi:10.3390/w9060439

Ma, B. Y., Wu, Z. N., Wang, H. L., & Guo, Y. (2020). Study on the Classification of Urban Waterlogging Rainstorms and Rainfall Thresholds in Cities Lacking Actual Data. *Water, 12*(12). Retrieved from <Go to ISI>://WOS:000602841700001. doi:10.3390/w12123328

Sarauskiene, D., Akstinas, V., Nazarenko, S., Kriauciuniene, J., & Jurgelenaite, A. (2020). Impact of physico-geographical factors and climate variability on flow intermittency in the rivers of water surplus zone. *Hydrological Processes, 34*(24), 4727-4739. Retrieved from <Go to ISI>://WOS:000574046800001. doi:10.1002/hyp.13912

van der Sterren, M., Rahman, A., & Ryan, G. (2014). Modeling of a lot scale rainwater tank system in XP-SWMM: A case study in Western Sydney,

Australia. *Journal of Environmental Management, 141*, 177-189. Retrieved from <Go to ISI>://WOS:000338599800020. doi:10.1016/j.jenvman.2014.02.013

Wang, C., & Wu, D. H. (2021). A K-Means Clustering-Based Multiple Importance Sampling Algorithm for Integral Global Optimization. *Journal of the Operations Research Society of China*(prepublish).

Wang, J. Y., Hu, C. H., Ma, B. Y., & Mu, X. L. (2020). Rapid Urbanization Impact on the Hydrological Processes in Zhengzhou, China. *Water, 12*(7). Retrieved from <Go to ISI>://WOS:000558701300001. doi:10.3390/w12071870

Wu, X., Wang, Z., Guo, S., Liao, W., Zeng, Z., & Chen, X. (2017). Scenario-based projections of future urban inundation within a coupled hydrodynamic model framework: A case study in Dongguan City, China. *Journal of Hydrology, 547*, 428-442. Retrieved from <Go to ISI>://WOS:000398871100032. doi:10.1016/j.jhydrol.2017.02.020

Wu, Z. I., Ma, B. Y., Wang, H. L., Hu, C. H., Lv, H., & Zhang, X. Y. (2021). Identification of Sensitive Parameters of Urban Flood Model Based on Artificial Neural Network. *Water Resources Management, 35*(7), 2115-2128. Retrieved from <Go to ISI>://WOS:000646644200001. doi:10.1007/s11269-021-02825-3

Wu, Z. N., Ma, B. Y., Wang, H. L., & Hu, C. H. (2020). Study on the Improved Method of Urban Subcatchments Division Based on Aspect and Slope- Taking SWMM Model as Example. *Hydrology, 7*(2), 17. Retrieved from <Go to

ISI>://WOS:000551260500009. doi:10.3390/hydrology7020026

Xiang, C., Liu, J., Wang, H., & Shao, W. (2017). Monitoring and Simulation of Urban Hydrological Unit: A Case Study of Tsinghua University, Beijing. [城市水文单元雨洪监测与模拟研究-以清华大学校园为例]. *Journal of China Hydrology, 37*(4), 12-17. Retrieved from <Go to ISI>://CSCD:6058698.

Xu, H. S., Ma, C., Xu, K., Lian, J. J., & Long, Y. (2020). Staged optimization of urban drainage systems considering climate change and hydrological model uncertainty. *Journal of Hydrology, 587*, 11. Retrieved from <Go to ISI>://WOS:000568819100081. doi:10.1016/j.jhydrol.2020.124959

Zaghloul, N. A. (1983). SENSITIVITY ANALYSIS OF THE SWMM RUNOFF-TRANSPORT PARAMETERS AND THE EFFECTS OF CATCHMENT DISCRETIZATION. *Advances in Water Resources, 6*(4), 214-223. Retrieved from <Go to ISI>://WOS:A1983SS39100004. doi:10.1016/0309-1708(83)90059-3

Zhang, N., Li, N., Gan, H., & Wang, J. (2017). Improvement of underground drainage computation method for urban flood simulation model. *Journal of Hydraulic Engineering, 48*(5), 526-534. Retrieved from <Go to ISI>://CSCD:6001685.

Du Xian. (2020). Research on refined simulation of urban rainstorm and flood based on SWMM model. (Master), Zhengzhou University, Available from Cnki

Fu Xiaoran, Wang Dong, Luan Qinghua, Liu Jiahong, & Wang Haichao. (2020). Construction and simulation of large-scale pipeless flow data urban SWMM——II. Model parameter verification and storm runoff simulation

513 analysis. Advances in Water Science, 31 (01), 51-60.
514 Liu Kexin, Bao Weimin, Que Jiajun, Li Jiajia, & Shu Huilian. (2015). Application of
515 K-means clustering method based on principal component analysis in flood
516 forecasting. Journal of Wuhan University (Engineering Edition), 48(04), 447
517 -450+458.
518 Liu Xingpo. (2009). Urban rainfall runoff model parameter calibration method based
519 on runoff coefficient. Water Supply and Drainage, 45(11), 213-217.
520 Luan Qinghua, Fu Xiaoran, Wang Haichao, Liu Jiahong, & Gao Xuerui. (2019).
521 Construction and simulation of SWMM for large-scale urban area without
522 pipe flow data——I. Key technology for digital refinement of complex
523 underlying urban area. Advances in Water Science, 30 (05), 653-660.
524 Mei Chao. (2019). Urban hydrology and hydrodynamic coupling model and its
525 application research. (Ph.D.), China Institute of Water Resources and
526 Hydropower Research, Available from Cnki
527 Mei Chao, Liu Jiahong, Wang Hao, Li Zejin, Xia Lin, & Wang Ying. (2017). SWMM
528 principle analysis and application prospects. Water Resources and
529 Hydropower Technology, 48(05), 33-42.
530 Qian Kun, Bao Weimin, Li Sisong, & Si Wei. (2012). Application of K-means cluster
531 analysis method in flood forecasting. Hydropower Energy Science, 30(05),
532 41-44.
533 Xu Yuanhao, Wu Qiang, Li Changqing, Chen Youqian, Zhang Li, Ran Guang, & Hu
534 Caihong. (2020). Simulation and prediction of flood process in the middle
535 reaches of the Yellow River based on long short-term memory (LSTM) neural
536 network. Journal of Beijing Normal University (Natural Science Edition),
537 56(03), 387-393.

Figures

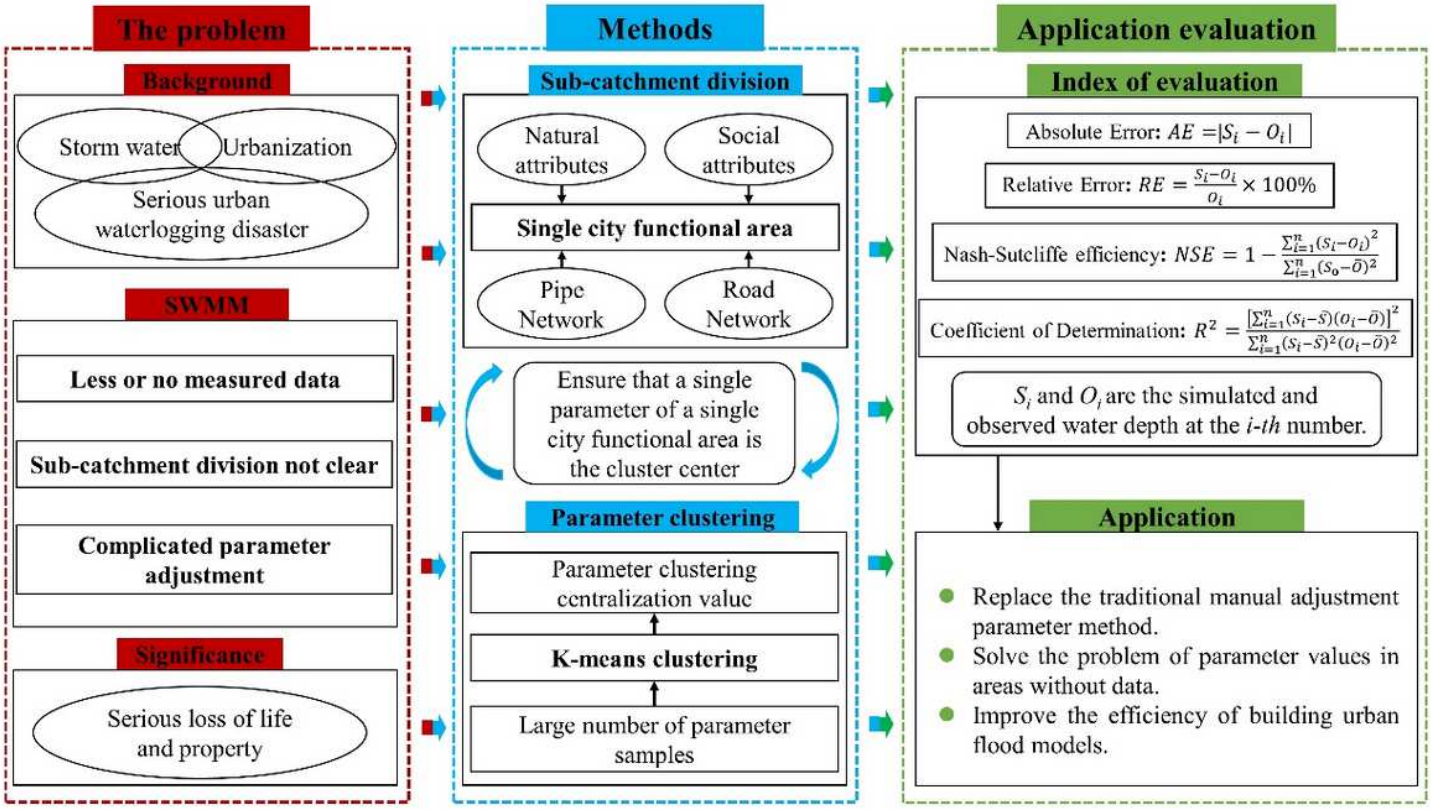

Figure 1

Research frame diagram.

Figure 2

Study area.

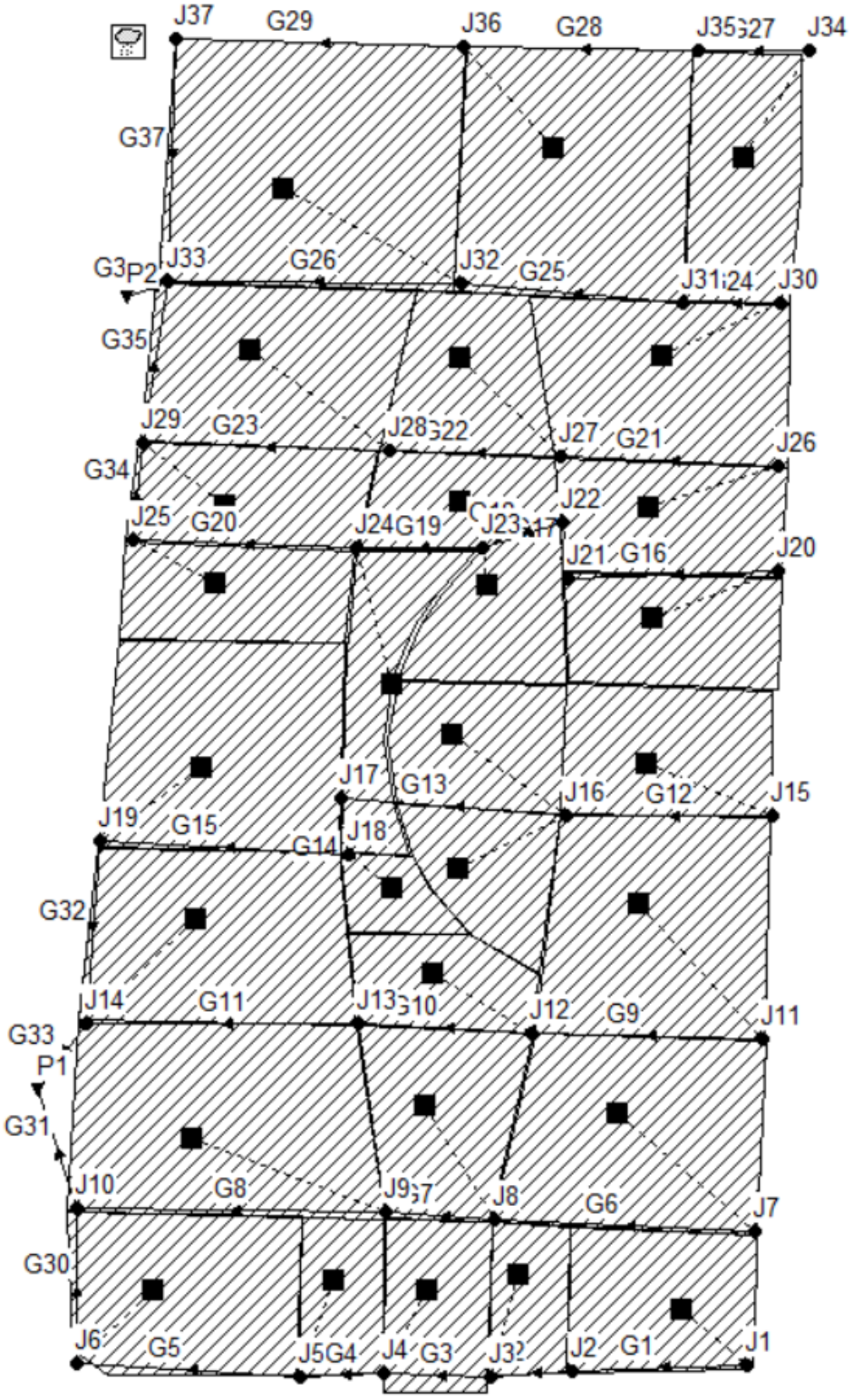

Figure 3

Generalized map of the sub-catchment in the study area.

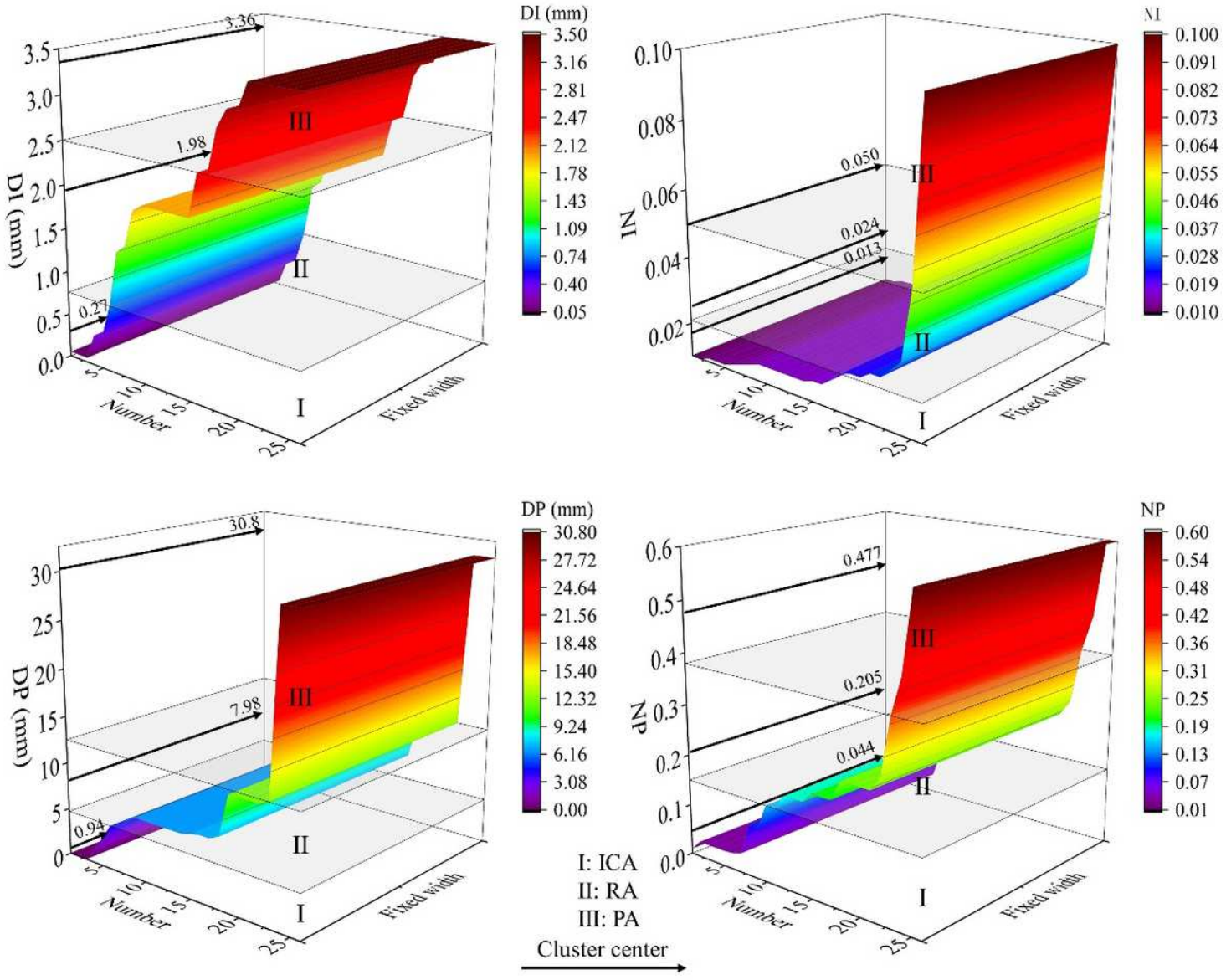

Figure 4

Parameter clustering results.

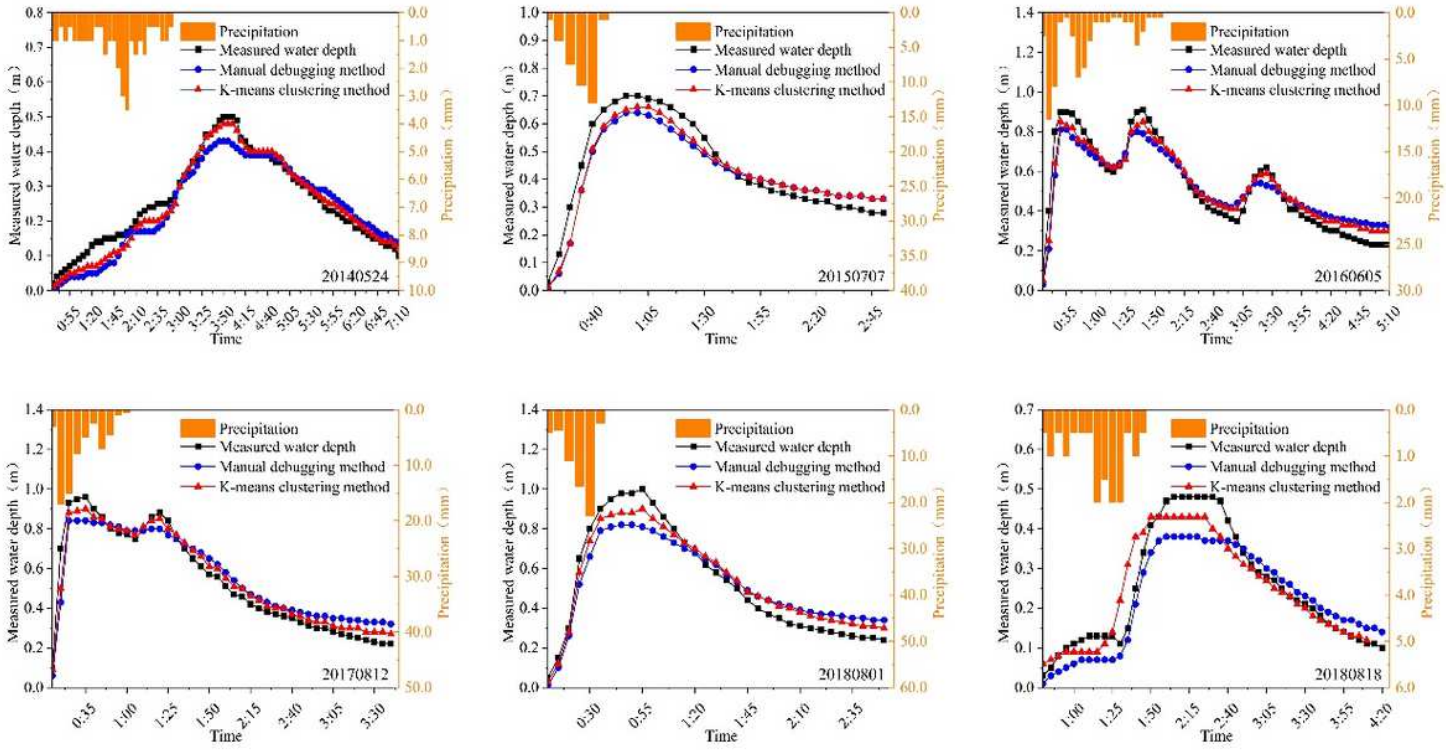

Figure 5

Observed and estimated hydro-graphs of the Manual debugging method and K-means clustering method in SWMM model at the validation stage in 6 flood events.

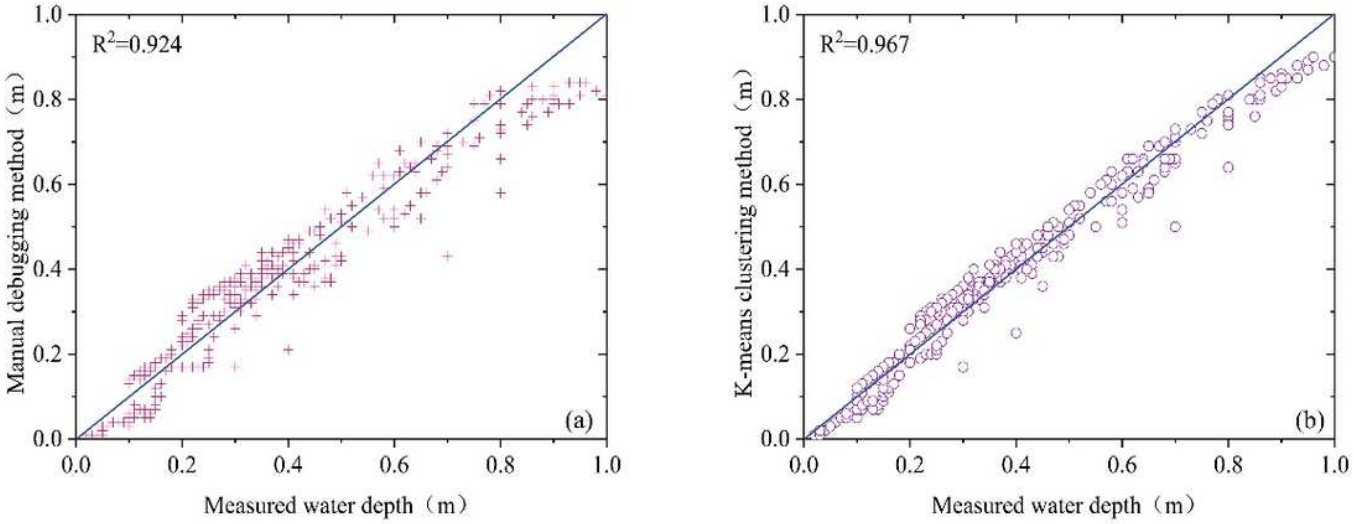

Figure 6

Correlation coefficient graph.