Study on the Influence of Abnormal Rainfall Station Detection in the Basin on the Simulation Accuracy of Hydrological Model

Ke Jia[1,2], Lan Lan[1*], Zhenliang Ma[3], Guoliang Ji[4], YiFei Tian[2]

[1]Guizhou East Frontal Engineering Design Consulting CO, LTD, Wuhan, 430010, China
[2]Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan, 430010, China
[3]Hubei Key Laboratory of Petroleum Geochemistry and Environment, Yangtze University, Wuhan, 430100, China
[4]China Three Gorges Corporation, Wuhan, 430010, China

*Corresponding author: yinshanxu@163.com

Abstract. The presence of abnormal rainfall stations in the basin will affect the quality of hydrological model input, thereby reducing the simulation accuracy of the hydrological model. In order to improve the simulation accuracy of the hydrological model, this paper attempts to first calculate the rainfall process missing index and rainfall process similarity index, and use the DBSCAN clustering algorithm to cluster the calculated index to screen out the missing abnormal stations; secondly, a lumped Xin'anjiang model is constructed to simulate the long flow of 1h, 3h, and 6h before and after the screening of abnormal rainfall stations; finally, the simulation results of the hydrological model are evaluated by the NS of the model simulation results and the Re of the water volume simulation. The results of the closed basin above the Xiangtun Station of Le'an River as the research object show that after using the method proposed in this paper to remove the abnormal rainfall station data, the NS of the model simulation is increased by 0.07, and the relative error of the water volume simulation is increased by 18.08%, which has a certain effect on improving the accuracy of the hydrological model simulation.

Keywords: Rainfall Station, Anomaly Detection, Rainfall Process Index, Xin'anjiang Model, DBSCAN Clustering

1 Introduction

Rain gauges are widely distributed and numerous as the main facilities for monitoring hydrological data. However, when monitoring rainfall data, a large number of rain gauges inevitably have abnormal situations such as missing measurements, missed measurements, and misreporting. As the input data of the hydrological model, the quality and authenticity of rainfall data have a great impact on the output results of the model. The output results of the hydrological model will also affect actual production

W. Nie et al. (eds.), *Proceedings of the 2024 6th International Conference on Hydraulic, Civil and Construction Engineering (HCCE 2024)*, Atlantis Highlights in Engineering 37,
https://doi.org/10.2991/978-94-6463-726-7_16

processes such as flood forecasting, water resources allocation, and reservoir scheduling [1]. In the modeling process, hydrologists need to prepare many years of rainfall data as the input of the model. In the face of a large amount of rainfall data, how to accurately filter out the data of abnormal stations is important for the refined hydrological simulation of the basin. By screening the hydrological stations in the region and providing high-quality data input for the hydrological model under the condition of reducing the station density, can the accuracy of the hydrological simulation be improved? In this regard, it is necessary to further explore the impact of removing regional abnormal rainfall data on the simulation accuracy of the hydrological model through experimental design.

An outlier is an outlier that deviates from the group[2]. The abnormal rainfall data monitored by the rain gauge station in the basin is manifested as the rainfall process or the recorded value does not conform to the rainfall law. At present, anomaly detection technology mainly includes three categories: statistical method, deep learning (DL), and machine learning (ML)[3]. The anomaly detection of rainfall data mostly adopts statistical methods, such as Hampel method, Grubbs criterion, and Raida criterion. Statistical methods calculate the statistics of the data as the confidence interval and regard rainfall data outside the interval as abnormal[4][5][6]. Liu Xiulin et al. used the Grubbs criterion to detect the cumulative rainfall value and screened out the annual rainfall data of suspicious stations[7]. Gao Yueming et al. applied statistical methods such as Raida criterion, Chauville criterion, Grubbs test, and Dixon test to detect whether the rainfall at a certain point is abnormal in the spatial dimension[8]. Tian Jiyang et al. combined the Hampel method and Grubbs criterion with radar-assisted verification to identify anomalies in large-scale rainfall monitoring data[9]. Chaudhary U et al. used statistics such as coefficient of variation, skewness, and kurtosis to identify outliers in rainfall data [10]. Deep learning constructs a multi-layer neural network structure, trains massive data, extracts complex features from it, and then provides accurate prediction results [11]. For hydrological time series data, most people use long short-term memory (LSTM) network models to learn time features, and combine other detection algorithms to build a coupling model [12][13] to identify abnormal data through simulation prediction. Machine learning is to "learn" rules from a large number of samples through a combination of various algorithms, and then use the rules to classify and predict data samples [14]. Among them, the clustering algorithm puts samples with high similarity in the data set in the same cluster, and samples with low similarity in different clusters to detect abnormal samples in the data set.

DBSCAN is a density-based clustering algorithm that can adapt to data clustering of arbitrary shape distribution. It does not need to set the number of clusters and can effectively identify the outliers in it [15]. In a certain area, the rainfall data monitored by the rainfall station has a certain correlation [16][17]. Therefore, the DBSCAN clustering algorithm can be used to cluster the rainfall data in the basin to detect outliers. The DBCSAN clustering algorithm is affected by two parameters: the neighborhood radius (Eps) and the minimum number of data points in the neighborhood radius (Minpts) [18][19]. In order to improve the accuracy of clustering results, many scholars have conducted extensive research on the parameter selection of clustering algorithms [20][21][22]. Jahirabadkar et al. determined parameters based on the distribution of data to adapt to data sets with varying density [23]. JH Kim et al. proposed an approximate adaptive density clustering algorithm (AA-DBSCAN) to determine the Eps parameter and improve clustering performance [24]. Khan et al. proposed an ADBSCAN algorithm to automatically determine the appropriate Eps and Minpts values to identify all clusters in the data set [25].

In order to study the impact of rainfall data input on the accuracy of hydrological model simulation, this paper takes the Le'an River Basin of Poyang Lake as the research area, detects the rainfall data with missing anomalies in the rain gauge station, and improves the quality of hydrological model input by removing the abnormal rainfall station data. First, the rainfall process missing index and rainfall process similarity index of the rain gauge station are calculated according to the distance weight; secondly, the missing anomalies and singular points of the station are detected by the DBSCAN clustering algorithm based on the calculated rainfall process index; finally, the impact of different rainfall data inputs on hydrological model simulation is explored through experimental design.

2 Overview and Data of the Study Area

The Le'an River Basin is located in the northeast of Jiangxi Province. It originates from the southern foot of Zhanggong Mountain in Wuyuan County, Jiangxi Province, and flows from east to west into Poyang Lake. The basin area is 8376km2. The study area of this paper is located in the closed basin above Xiangtun Station of Le'an River. The basin area is 3893km2 and includes 84 rainfall stations. The average distance between rainfall stations in the basin is 31.9km, and each station contains 8 years of rainfall monitoring data. The distribution of rainfall stations in the study area is shown in Figure 1.

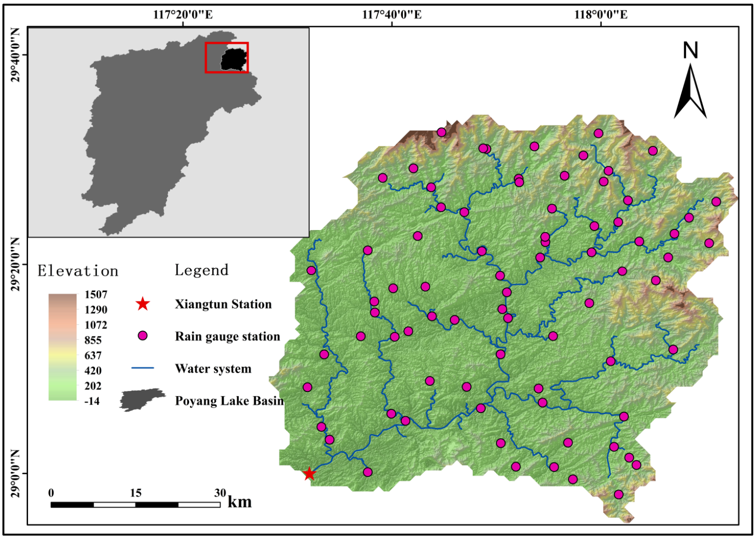

Fig. 1. Distribution map of rainfall stations in the study area

3 Research Methods

This paper detects the anomaly of missing rainfall data, screens rainfall stations from the perspective of missing detection and singular point detection, and then constructs a lumped Xin'anjiang model for the Xiangtun station in the study area. Hydrological simulations are performed for the 1h, 3h, and 6h long time schemes before and after the abnormal rainfall stations are removed, respectively, to explore the impact of different rainfall data inputs on the simulation accuracy of the Xin'anjiang model. The overall flow chart of this study is shown in Figure 2.

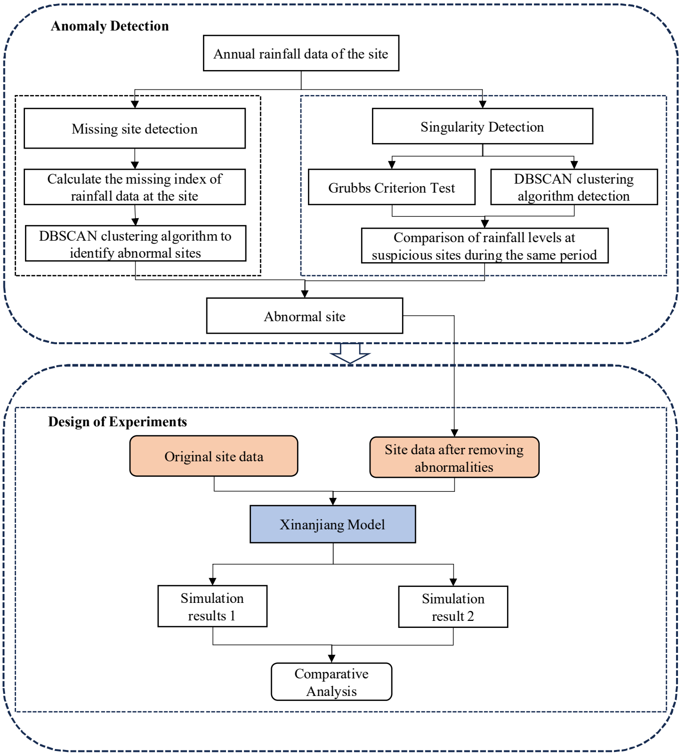

Fig. 2. Overall flow chart of the study

3.1 Rainfall Process Missing Index

The spatial distribution of rainfall stations in the region will affect the similarity of rainfall monitoring data. In order to reflect that the similarity of rainfall data at neighboring stations is higher than that at remote stations, distance weighting (DW) is introduced. In this paper, the Gaussian function is used to calculate the distance weight of rainfall stations. The formula is as follows:

$$DW = ae^{-\frac{(distance-b)^2}{2c}} \tag{1}$$

Where: distance represents the straight-line distance between sites; a represents the maximum weight value; b represents the minimum distance between sites; and c is the parameter value for adjusting the maximum distance between sites to correspond to the minimum weight.

The rainfall process missing index (RPMI) is used to describe the magnitude of rainfall data missing at the central site relative to other sites in the same time period. The site to be calculated is recorded as the central site. The formula is as follows:

$$RPMI_k = \sum_{k=1}^{N} [len(S_k^*) \times DW_k] \tag{2}$$

Where: $len(S_k^*)$ represents the number of rainfall data that the central station and the k th station did not monitor in the same time period; DW_k represents the distance weight between the central station and the k th station.

3.2 Rainfall Process Similarity Index

The rainfall process similarity index (RPSI) of the site is used to describe the similarity of the rainfall processes monitored between rainfall sites. The calculated rainfall process similarity index is normalized, and the formula is as follows:

$$RPSI_k = \sum_{k=1}^{N} (Ed_k * DW_k) \tag{3}$$

$$Ed_k = \sqrt{\sum_{t=1}^{T} (x_{k1}^t - x_{k2}^t)^2} \quad k1、k2 \in [1, N], t \in [1, T] \tag{4}$$

$$RPSI_{norm} = \frac{RPSI_k - min(RPSI)}{max(RPSI) - min(RPSI)} \quad k \in [1, ..., N] \tag{5}$$

Where: Ed_k is the Euclidean distance between the rainfall data of the central station and the rainfall data of the k th station; N is the number of rainfall stations in the region; T is the number of moments of rainfall data at the rainfall station.

3.3 DBSCAN Clustering

The DBSCAN clustering algorithm is a density-based spatial clustering algorithm that uses the spatial density of the data set as the basis for clustering, so there is no need to set the number of clusters. The DBSCAN clustering algorithm can cluster arbitrarily distributed data and effectively identify outliers. Figure 3 is a schematic diagram of the clustering process of the DBSCAN clustering algorithm.

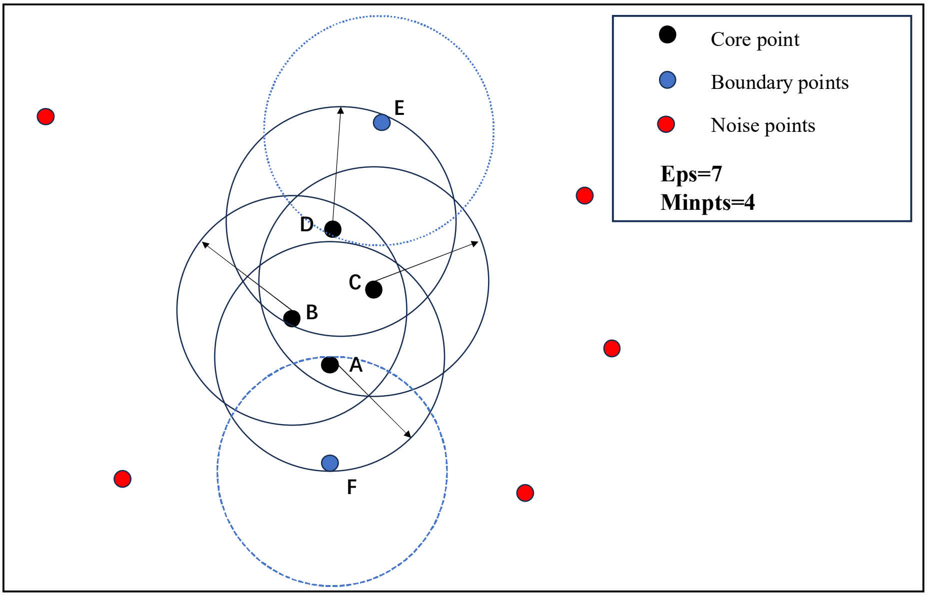

Fig. 3. Clustering diagram of DBSCAN clustering algorithm

In Figure 3, the neighborhood radius Eps of the DBSCAN clustering algorithm is 7, and the minimum number of data points in the neighborhood Minpts is 4. The search starts from the core point A, and the Eps range includes three data points B, C, and F. Since point F does not meet the Minpts requirement within the Eps range, point F is a boundary point. The core point propagates from point A to point B, and the data points included in the Eps range of core point B meet the Minpts requirement. Similarly, the core point propagates from point A to point D. When the propagation reaches point E, since the Minpts requirement is not met within its Eps range, point E is also a boundary point, and clustering ends. The red data points in the figure exceed the Eps range of each core point and cannot be propagated by other core points, which are outliers. Through the above process, data points A~F are finally clustered into one category, and the remaining data points are identified as outliers.

3.4 Singularity Detection

The Grubbs criterion determines whether an outlier is abnormal by measuring the degree to which the outlier is far from the sample mean[8]. In order to eliminate the shielding effect of outliers on the same side, the mean in the original formula is replaced by the median[26]. The annual cumulative rainfall of rain stations in the region is tested using the Grubbs criterion to screen out rain stations with abnormal annual cumulative rainfall. The formula is as follows:

$$G_1 = (X_{median} - x_1) / \delta \tag{6}$$

$$G_n = (x_n - X_{median}) / \delta \tag{7}$$

$$\delta = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{8}$$

Where: x_1, x_n are the maximum and minimum values of the sample sequence X after sorting, respectively; X_{median} is the median of the sample sequence X; δ is the standard deviation of the sample sequence X.

The process of Grubbs criterion for distinguishing abnormalities is as follows, where G_0 is the statistical critical coefficient obtained by checking the Grubbs critical value table. If $G_1 \geq G_n$ A and $G_1 > G_0$, then x_1 is judged as an abnormal value and eliminated; if $G_n \geq G_1$ and $G_n > G_0$, then x_n is judged as an abnormal value and eliminated; if $G_0 > G_1$ and $G_0 > G_n$, then there is no abnormal value. If there is an abnormal value, it will be eliminated and the above process will be repeated with the remaining samples until no abnormal value is eliminated.

The suspicious stations screened by the Grubbs criterion and the suspicious stations detected by the DBSCAN clustering algorithm for the rainfall process similarity index are compared with the rainfall values monitored by the adjacent non-suspicious stations in the same time period. When the rainfall value of a certain period of the suspicious station differs from the rainfall value of the same period of the adjacent station by more than one rainfall level [9], and the monitored rainfall value is 0, the monitored rainfall value of the suspicious station in this period is judged to be a singular point. Table 1 is a rainfall level classification table.

Table 1. Rainfall classification table

grade	rainfall (mm)				
	1h	3h	6h	12h	24h
light rain	0.1~1.5	0.1~2.9	0.1~3.9	0.1~4.9	0.1~9.9
moderate rain	1.6~6.9	3~9.9	4~12.9	5~14.9	10~24.9
heavy rain	7~14.9	10~19.9	13~24.9	15~29.9	25~49.9
rainstorm	15~39.9	20~49.9	25~59.9	30~69.9	50~99.9
heavy rainstorm	40~49.9	50~69.9	60~119.9	70~139.9	100~249.9
extremely heavy rainstorm	>=50	>=70	>=120	>=140	>=250

Note: Refer to the meteorological department's rainfall classification standards.

3.5 Model Calibration and Testing

This paper uses the Nash efficiency coefficient (NS) and the relative error (Re) of water volume simulation as evaluation indicators of model simulation effect [27]. The NS coefficient reflects the overall effect of model simulation. When NS is closer to 1, the model simulation is better. A positive value of Re indicates that the simulated value is higher than the true value, and a negative value indicates that the simulated value is lower than the true value. When Re is closer to 0, the model simulation result is better.

4 Results Analysis

4.1 Calculation Results of Distance Weights between Sites

The distance weight is calculated according to formula (1), where parameter a is 1, parameter b is 0, and parameter c is calculated by inverse calculation based on the weight value corresponding to the maximum station distance of 0.01. Figure 4 shows the distance weight calculation result between stations in the study area. The closer the color is to orange, the closer the distance between stations is, the closer the calculated distance weight is to 1, and the higher the similarity of the monitored rainfall data is; the closer the color is to blue, the farther the distance between stations is, the closer the calculated distance weight is to 0, and the lower the similarity of the monitored rainfall data is.

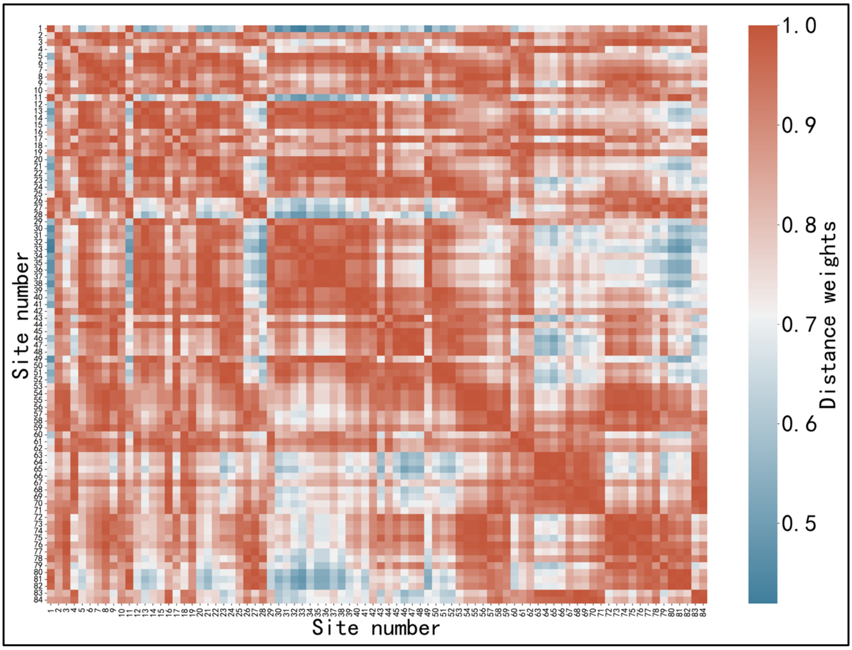

Fig. 4. Distance weight results between sites in the study area

4.2 Rainfall Data Index Based on Distance Weight Calculation

Taking the 3-h period scheme as an example, Figure 5 shows the calculation results of the rainfall process missing index of each station in the study area in each month in 2016, and Figure 6 shows the calculation results of the rainfall process similarity index of the stations in the study area from 2016 to 2023.

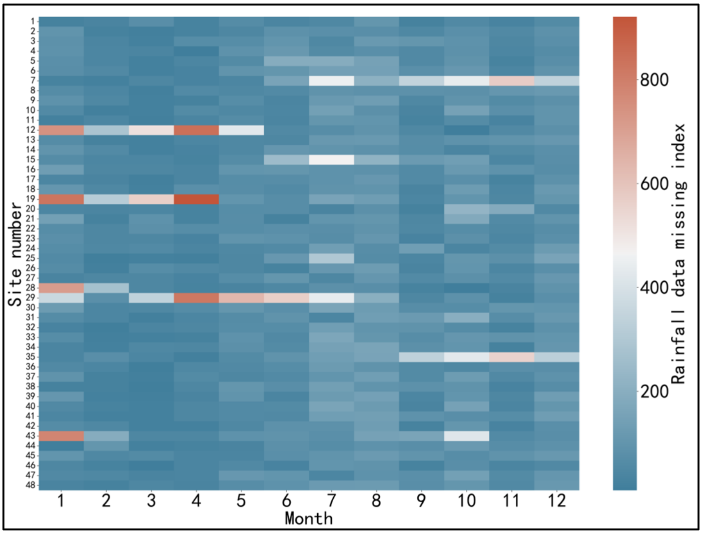

Fig. 5. Monthly rainfall missing index of each station in the study area in 2016

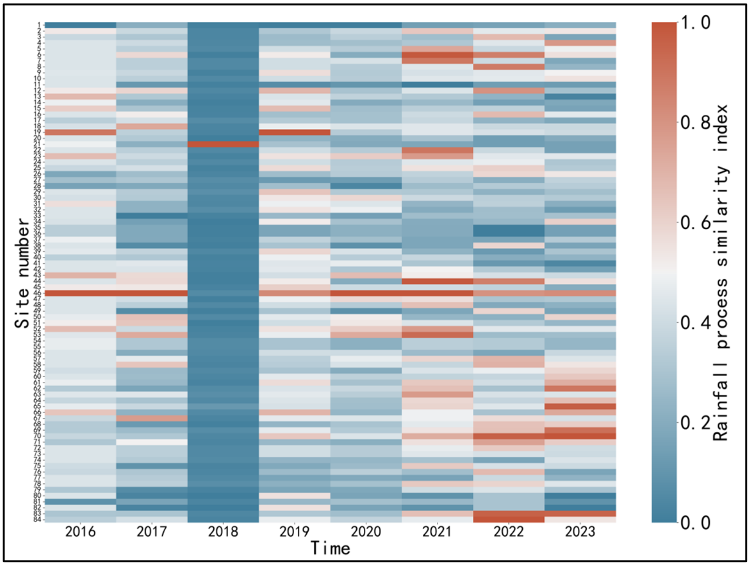

Fig. 6. Similarity index of rainfall processes at each station in the study area from 2016 to 2023

The DBSCAN clustering algorithm is used to cluster the rainfall process missing index of each station in each month of each year, and abnormal stations are detected. The results of the missing station detection are manually checked, and the recall rate of the detection results is shown in Table 2. The recall rate is the percentage of samples predicted to be positive among the actual positive samples. From the results in Table 2, it can be shown that the recall rate of the missing station detection method in this paper reaches more than 75%, which can effectively detect the missing stations.

Table 2. Detection rate of missing stations in the 3-hour long-term scheme in the study area

years	2016	2017	2018	2019	2020	2021	2022	2023
Recall	100%	81%	78%	86%	75%	88%	92%	100%

4.3 Analysis of the Simulation Results of the Xinanjiang Model Based on Abnormal Site Detection

In order to compare the impact of abnormal rainfall station detection in the basin on the simulation accuracy of the hydrological model, a lumped Xin'anjiang model was established in the closed basin above Xiangtun Station in the Le'an River Basin. Hydrological simulation was performed on rainfall data with a time period of 1h, 3h, and 6h, and the average surface rainfall in the basin before and after the abnormal rainfall station was detected was calculated using Thiessen polygons. This study used the rainfall data for 6 years from 2016 to 2021 as the rate period of the model, and 2022 and 2023 as the test period of the model. The cooperative search algorithm (CSA) was used to search for the globally optimal parameters of the Xin'anjiang model parameters. According to the detection method of this paper, the rainfall data of 1h, 3h, and 6h at the stations in the study area from 2016 to 2023 were detected year by year. Table 3 is a statistical table of the detection results of abnormal rainfall stations in different time length schemes.

Table 3. Statistics of abnormal site detection results for the long-term schemes in the study area for 1h, 3h, and 6h periods

plan	2016	2017	2018	2019	2020	2021	2022	2023
1-h	10	27	10	14	13	18	21	12
3-h	15	23	10	12	6	19	17	12
6-h	10	30	14	9	5	16	13	13

The model calibration results before and after the abnormal rainfall station detection in the basin are shown in Table 4. It can be seen from Table 4 that after the abnormal rainfall station detection, the NS of the regular simulation process of the Xin'anjiang model rate at the Xiangtun station in different time periods of long schemes are all above 0.896, the Re of the water volume simulation is within -29.1%, and the NS of the simulation process in the test period is all above 0.929, and the Re of the water volume simulation is within -28.9%; the NS of the regular simulation process of the

Xin'anjiang model rate at the Xiangtun station in different time periods of long schemes without abnormal rainfall station detection is all above 0.869, the Re of the water volume simulation is within -39.8%, the NS of the simulation process in the test period is all above 0.916, and the Re of the water volume simulation is within -33.7%.

Table 4. Calibration results of Xin'anjiang model at Xiangtun Station

plan		Rate Regular (2016-2021)		Inspection period (2022-2023)	
		NS	Re(%)	NS	Re(%)
1-h	Detection	0.917	-29.1	0.931	-28.9
	Not detected	0.916	-29.3	0.939	-23.0
3-h	Detection	0.918	-27.4	0.937	-28.7
	Not detected	0.869	-35.5	0.916	-29.1
6-h	Detection	0.896	-25.5	0.929	-26.5
	Not detected	0.892	-39.8	0.926	-33.7

Figures 7 to 9 show the comparison of the total annual runoff simulated by the 1h, 3h, and 6h time-period schemes at Xiangtun Station. From the result comparison chart, it can be seen that the average annual runoff simulated by the different time-period schemes that detect abnormal rainfall sites is closer to the measured total amount, and the average annual water volume Re simulated by the 1h, 3h, and 6h time-period schemes that have been detected by abnormal rainfall sites are -29.5%, -28.3%, and -26%, respectively, while the average annual water volume Re simulated by the 1h, 3h, and 6h time-period schemes that have not been detected by abnormal rainfall sites are -28.5%, -34.7%, and -39.1%, respectively. According to the above analysis results, the flow process simulated by the Xin'anjiang model at Xiangtun Station after detecting abnormal sites is closer to the measured value than the simulated flow process without detecting abnormal sites, indicating that the accuracy of the Xin'anjiang model simulation can be improved by removing the data of abnormal rainfall sites in the basin.

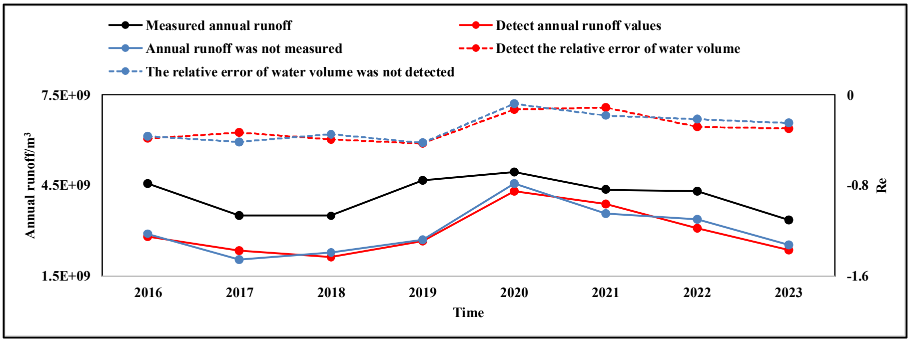

Fig. 7. Comparison of the total annual runoff simulated by the 1-hour long scheme at Xiangtun Station

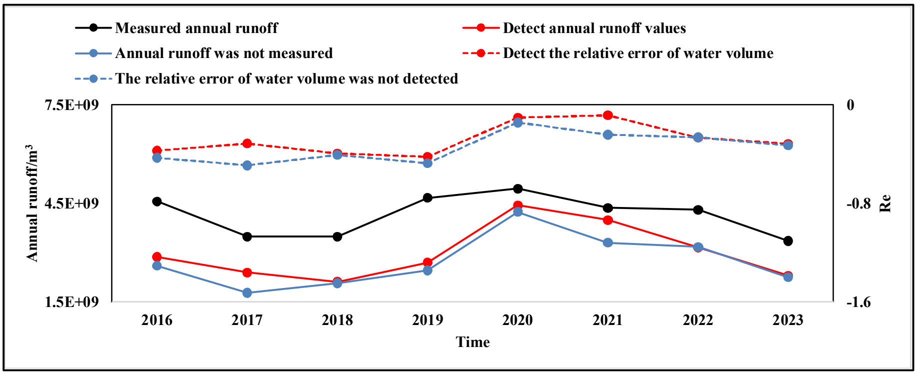

Fig. 8. Comparison of the total annual runoff simulated by the 3-hour long-term scheme at Xiangtun Station

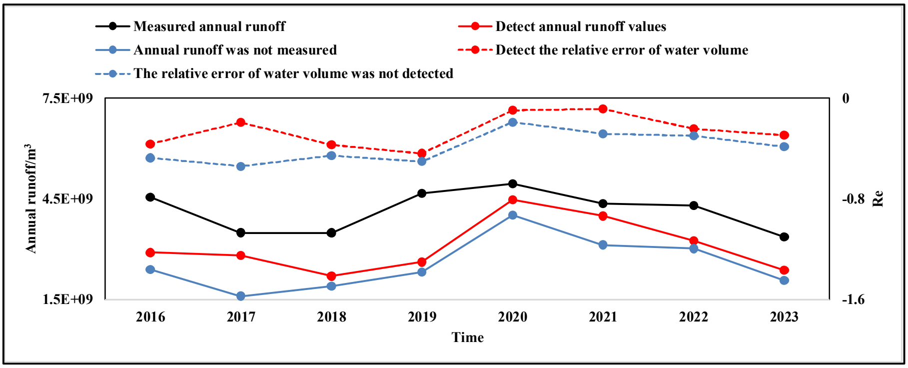

Fig. 9. Comparison of the total annual runoff simulated by the 6-hour long-term scheme at Xiangtun Station

5 Conclusion

Based on the method of regional abnormal rainfall site detection proposed in this paper, the lumped Xin'anjiang model is established for the closed basin above Xiangtun Station, and the 1h, 3h, and 6h long-term schemes are established respectively. Among them, the model rate NS of the model detected by the abnormal rainfall site is above 0.896 on a regular basis, and the NS of the validation period is above 0.929; the model rate NS of the model not detected by the abnormal rainfall site is above 0.869 on a regular basis, and the NS of the validation period is above 0.916. After the abnormal rainfall site detection, the average annual water volume Re of the 1h long-term scheme is -29.5%, the average annual water volume Re of the 3h long-term scheme is -28.3%, and the average annual water volume Re of the 6h long-term scheme is -26%; without the abnormal rainfall site detection, the average annual water volume Re of the 1h long-term scheme is -28.5%, the average annual water volume Re of the 3h long-term scheme is -34.7%, and the average annual water volume Re of the 6h long-term scheme is -39.1%. It shows that the detection method in this paper can improve the accuracy of hydrological model simulation after eliminating abnormal rainfall stations, and can provide ideas and methods for refined hydrological data simulation in the basin.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2023YFC3206005) and Scientific Research Projects of China Three Gorges Group Co. (0704218)

References

1. Wang Sen , Wang Xiayu, Jia Wenhao, et al. Research Progress and Prospect of Reservoir Intelligent Operation[J]. Pearl River, 2023, 44(10):25-34. https://doi.org/10.3969/j.issn.1001-9235.2023.10.003.
2. Ane BLAZQUEZ-GARCIA, CONDE A, MORI U, et al. A review on outlier/anomaly detection in time series data[J]. ACM Computing Surveys (CSUR), 2021, 54(3):1-33. https://doi.org/10.48550/arXiv.2002.04236.
3. BRAEI M, WAGNER S. Anomaly Detection in Univariate Time-series: A Survey on the State-of-the-Art[J]. arXiv preprint arXiv:2004.00433,2020. https://doi.org/ 10.48550/arXiv.2004.00433.
4. Qi Minfang. Big Data Technology and Its Application on the Analysis of Power Plant Units[D]. North China Electric Power University, 2016. https://doi.org/ 10.7666/d.Y3114001.
5. Yang Maolin, Lu Yansheng. Pruning based Outlier Mining from Large Dataset[J]. Computer Science, 2012,39(10):152-156. https://doi.org/ 10.3969/j.issn.1002-137X.2012.10.033.
6. SENGONUL E, SAMET R, ABU AL-HAIJA Q, et al. An analysis of artificial intelligence techniques in surveillance video anomaly detection: A comprehensive survey[J]. Applied Sciences, 2023, 13(8):4956. https://doi.org/10.3390/app13084956.
7. Lin Xiulin, Zhang Hangnan, Fang Yuanhao, et al. Study on data quality of hourly rainfall of telemetry rainfall stations in lower reaches of Jinsha River[J]. Pearl River, 2019, 50(03):131-135+171. https://doi.org/ CNKI:SUN:RIVE.0.2019-03-023.
8. Gao Yueming, Lin Qinghua, Liu Shupiao, et al. Spatial Dimension Analysis and Judgement of Abnormal Rainfalls[J]. Pearl River, 2022, 43(12):97-103+154.
9. Tian Jiyang, Liu Hanying, Liu Ronghua, et al. Anomaly recognition method of massive rainfall data[J]. Journal of China Institute of Water Resources and Hydropower Research, 2022, 20(5):438-448. https://doi.org/ 10.13244/j.cnki.jiwhr.20210239
10. CHAUDHARY U, LIMBU M N, KAPHLE G C, et al. Statistical Analysis of Rainfall Distribution in Biratnagar, Nepal: A Case Study[J]. Pragya Darshan, 2023, 5(1):52-57. https://doi.org/10.3126/pdmdj.v5i1.52307.
11. MIAU S, HUNG W H. River flooding forecasting and anomaly detection based on deep learning[J]. Ieee Access, 2020, 8:198384-198402. https://doi.org/ 10.1109/ACCESS.2020.3034875.
12. CHEN Z, NI X, LI H, et al. FedLGAN: a method for anomaly detection and repair of hydrological telemetry data based on federated learning[J]. PeerJ Computer Science, 2023, 9:e1664. https://doi.org/ 10.7717/peerj-cs.1664.

13. Luo Chaolin, Zhang Bo, Meng Qinkui, et al. Reservoir Flood Forecasting Based on Long-Short-Term Memory Neural Network[J]. Pearl River, 2022,4 3(12):128-134. https://doi.org/ 10.3969/j.issn.1001-9235.2022.12.018.
14. Tian Yifei, Yang Haicong, JI Guoliang, et al. Research on Similar Flood Analysis and Intelligent Recommendation Modeling[J]. Journal of Water Resources Research, 2023, 12(5), 472-485. https://doi.org/ 10.12677/JWRR.2023.125052.
15. Song Dongfei, Xu Hua. Research and parallelization of DBSCAN algorithm[J]. Computer Engineering and Applications, 2018, 54(24): 52-56+122. https://doi.org/ 10.3778/j.issn.1002-8331.1808-0423.
16. Zhao Xiaoyan. Analysis of Spatiotemporal Variation Characteristics of Precipitation in the Upper Yangtze River[J]. Advances in Geosciences, 2022, 12(11):19. https://doi.org/ 10.12677/AG.2022.1211146.
17. Li Lulu, Xiao Tiangui, Yang Mingxin, et al. Spatial and Temporal Distribution Characteristics and Analysis of Precipitation Anomalies in Southwest China[J]. Climate Change Research Letters, 2019,8(6):12. https://doi.org/ 10.12677/CCRL.2019.86076.
18. ALI T, ASGHAR S, SAJID N A. Critical analysis of DBSCAN variations[C]//2010 international conference on information and emerging technologies. IEEE, 2010: 1-6. https://doi.org/ DOI: 10.1109/ICIET.2010.5625720.
19. Soman, K.P, Diwakar, Shyam, Ajay V, et al. Basic tutorial on data mining[M]. China Machine Press, 2009.
20. Li Wenjie, Yan Shiqiang, Jiang Ying, et al. Research on Method of Self-Adaptive Determination of DBSCAN Algorithm Parameters[J]. Computer Engineering and Applications, 2019, 55(05):1-7+148. https://doi.org/ CNKI:SUN:JSGG.0.2019-05-002.
21. Zhou Hongfang, Wang Peng. Research on Adaptive Parameters Determination in DBSCAN Algorithm[J]. Journal of Xi'an University of Technology, 2012, 28(03):289-292. https://doi.org/ 10.3969/j.issn.1006-4710.2012.03.007.
22. Xia Luning, Jin Jiwu. SA-DBSCAN:A self-adaptive density-based clustering algorithm[J]. Journal of University of Chinese Academy of Sciences, 2009, 26(04):530-538. https://doi.org/ 10.7523/j.issn.2095-6134.2009.4.015.
23. JAHIRABADKAR S, KULKARNI P. Algorithm to determine ε-distance parameter in density based clustering[J]. Expert Systems with Applications, 2014, 41(6): 2939-2946. https://doi.org/10.1016/j.eswa.2013.10.025.
24. KIM J H, CHOI J H, YOO K H , et al. AA-DBSCAN: an approximate adaptive DBSCAN for finding clusters with varying densities[J]. Springer US, 2019, 75(1):142-169. https://doi.org/10.1007/s11227-018-2380-z
25. KHAN M M R, SIDDIQUE M A B, ARIF R B, et al. ADBSCAN: Adaptive density-based spatial clustering of applications with noise for identifying clusters with varying densities[C]. 2018 4th international conference on electrical engineering and information & communication technology (iCEEiCT). IEEE, 2018: 107-111. https://doi.org/ 10.1109/CEEICT.2018.8628138.
26. Yang Jinyan, Jiang Zengjie ,Chen Wei, et al. Application of Robust Statistics and Grubbs Criterion in Proficiency Testing Results Analysis[J]. Acta Metrologica Sinica, 2018, 39(06):862-867. https://doi.org/ 10.3969/j.issn.1000-1158.2018.06.21.
27. Du Yong, Fu Yupeng, Wei Yongjiang, et al. Study on River System Flood Forecasting Scheme for Longtan Hydropower Station Interval in Hongshui River[J]. Pearl River, 2024, 45(01):122-130. https://doi.org/ 10.3969/j.issn.1001-9235.2024.01.012.