RESEARCH ARTICLE

Study on Influence Factors of Suspended Solids Concentration Change in East Lake of Wuhan before and after Rainfall Events Based on GWR Model

GAN Wen-xia[1], PAN Jun-jie[1], XIAO An-na[2], YANG Ru-qin[3]

[1] *School of Civil Engineering and Architecture，Wuhan Institute of Technology，Wuhan 430205，China；*

[2] *Hubei Communication Investment Intelligent Detection Co.，Ltd，Wuhan 430050，China；*

[3] *Wuhan Natural Resources and Planning Information Center，Wuhan 430014，China*

ABSTRACT

The research combines remote sensing technology and geostatistical spatial analysis technology to quantitatively analyze the impact of different factors such as social, economic and infrastructure factors on the concentration of suspended solids in urban lakes. Through remote sensing data, the suspended solids concentration of East Lake in Wuhan before and after the rainfall from January 20 to January 28, 2021 was quantitatively retrieved, and the spatial characteristics of the suspended solids concentration change before and after the event and the impact of different factors on the suspended solids concentration change were analyzed by using geographical weighted regression. The analysis results show that the building area, the number of intersections and the road area are positively correlated with the concentration of suspended solids. The area of forest and green space is negatively correlated with the concentration of suspended solids in the water edge area. This research can provide theoretical support for urban environmental management and precise pollution prevention and control.

Keywords: Traffic pollution；Water quality；Concentration of suspended solids；Remote sensing；Geostatistical spatial analysis

1. Introduction

As a city with hundreds of lakes, Wuhan's lake water environment directly affects the lives of surrounding residents, the harmonious development of urban ecology and economy. In recent years, with the implementation of the comprehensive water environment management plan in Wuhan, the goal of "sewage entering the pipeline network and rainwater entering the lakes and ponds" has been basically achieved. The main source of water pollution in urban lakes is rainwater runoff[1]. Analyzing the factors and characteristics affecting lake water quality before and after rainfall events has important theoretical significance and practical value.

Scholars have done a lot of research on the influencing factors of water environment, among which there are more studies on the impact of land use factors on rainwater runoff [2, 3]. The research results show that differences in land use can lead to a regular distribution of water quality concentration. With the vigorous development of road construction, traffic activities are becoming increasingly frequent, and pollution caused by road runoff washing pollutants into receiving water bodies has also attracted people's attention [4,5]. However, existing studies have mainly analyzed the overall state of water quality, treating the research object as a whole to analyze its water quality or pollution sources [6]. However, when the water surface area is large, its water quality distribution and its impact characteristics have spatial heterogeneity, and a holistic analysis is difficult to reveal the spatial distribution heterogeneity of the impact factors. There is a lack of detailed analysis of the water quality impact characteristics brought about by typical rainfall events. The Geoweighted Regression (GWR) model in geostatistics can explore the intensity and direction of the influence of different independent variables on the dependent variable, and reflect the spatial heterogeneity characteristics of influencing factors[7]. It is widely used in different industries, such as environmental spatial heterogeneity analysis of ecological indicators, air quality, and carbon emission indices. Zhao Jingya et al. [8] used population, road network, and transportation data from Los Angeles, USA, and analyzed the impact of multiple influencing factors on nitrogen dioxide emissions using a geographically weighted regression model; Chen Xiaohui et al.[9] used geographically weighted regression to analyze the impact of different built-up areas on the ecological index in Fuzhou City, and quantitatively analyzed the spatial heterogeneity of the ecological index; Wang et al.[10] used the GWR model to analyze the spatial heterogeneity impact of road network traffic characteristics on urban air quality; Yuan Changwei et al. [11] constructed a geographically weighted regression model for carbon emissions in various provinces of China based on the three major influencing factors of transportation industry scale, technology, and structure, and thus achieved the division of carbon emission reduction pressure indices in different regions.

Based on the comprehensive background and current situation, this article takes Donghu Lake in Wuhan City as an example and uses geographic weighted regression method to quantitatively analyze the changes in suspended solids concentration in water before and after rainfall events and the impact characteristics between different influencing factors. As a typical large-scale urban lake, East Lake in Wuhan has a close geographical interaction with the city and its population. Conducting analysis and research on rainfall events in the study area of East Lake can help reveal the spatial heterogeneity of different influencing factors on water quality and provide a basis for regional water environment governance.

2. Research area and data

Donghu Lake is located in the eastern part of the urban area of Wuhan City, Hubei Province. It can be further divided into sub lakes such as Guoguo Lake, Guozheng Lake, Tangling Lake, Miao Lake, Tuan Lake, Houhu Lake, Yujia Lake, etc. The total area of the lake is 31.75 square kilometers at a normal high water level of 19.78 meters. It is the largest urban lake in China (**Figure 1**), an important water source in Wuhan City, and a well-known ecological tourism scenic area. The western water area is closely connected to Wuchang District, surrounded by Hongshan District and Qingshan District to the east, south, and north.

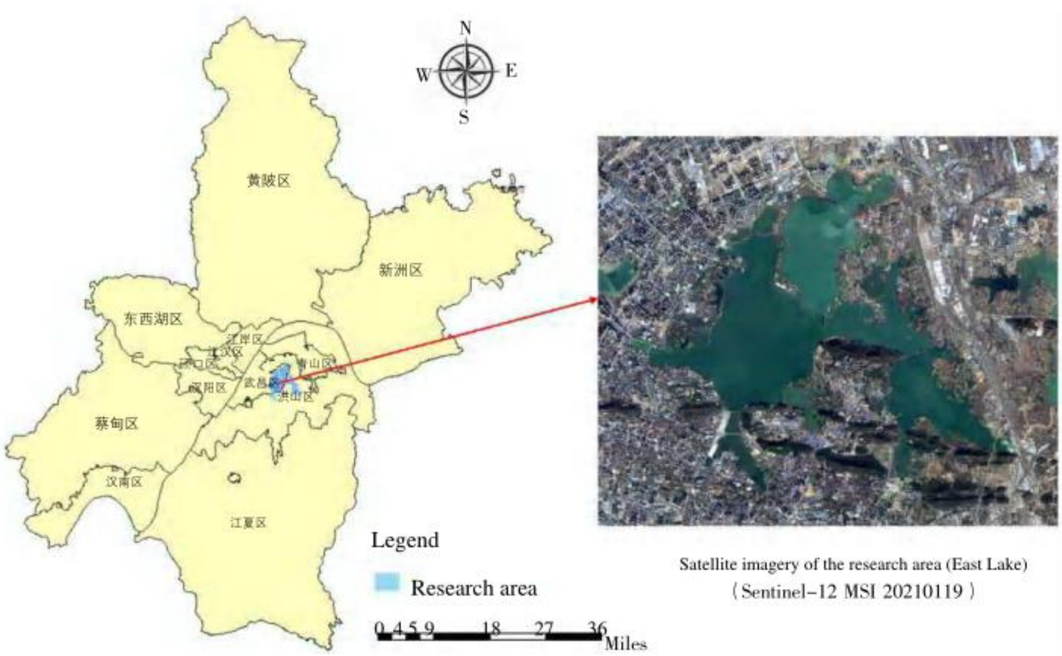


Figure 1. Geographical Location Map of the Study Area.

2.1. Remote sensing images and measured water quality data

This study selected Sentinel-2 MSI data from satellite multispectral remote sensing images as the data source for remote sensing water quality inversion. Image data from January 19, 2021 and January 29, 2021 before and after rainfall were selected for suspended solids concentration inversion. The suspended solids inversion model of East Lake used the band combination inversion model of East Lake water quality parameters constructed by Lu Dongshuo [13].

2.2. Impact factor data

Starting from the actual social, economic, and infrastructure conditions of the research area, combined with the exploration of rainwater runoff and its influencing factors in existing research, and the analysis of the process of rainfall affecting lake water quality through rainwater erosion, a preliminary set of influencing factors for urban lake water quality was constructed as shown in Table 1. The data sources include publicly available online data and interpreted data [14]. In the study, factor screening methods will also be used to further select the main influencing factors.

Table 1. Factors affecting the preliminary selection of urban lake water quality.

Factors	Primary selection factor	Abbreviation	Unit	Variable number
Land use factors	Building area	BR	m^2	X_1
	Forest area	FR	km^2	X_2
	Green area	GR	km^2	X_3
	Road area	RR	m^2	X_4
Road traffic factors	Road width	RW	m	X_5
	Road length	RL	km	X_6
	Number of intersections	RC	Pieces/km2	X_7
	Vehicle speed	RS	km/h	X_8

3. Research methods

In this study, the spatial autocorrelation of suspended particulate matter concentration in Donghu Lake was first statistically analyzed. Then, correlation analysis and collinearity test were conducted, and the influencing factors were screened based on the characteristics of the event. Finally, a geographically weighted regression model was established between the suspended particulate matter concentration and different factors in 2.2. Based on the modeling results, the degree and spatial characteristics of the impact of different factors on the changes in suspended particulate matter concentration before and after rainfall were analyzed. The specific methods are as follows.

3.1. Spatial autocorrelation analysis

Spatial autocorrelation can reflect the degree of interdependence of spatial variables in various regions[15]. This article uses the classic Moran index in geostatistical research to explore the global spatial autocorrelation of water quality, as the basis for determining spatial heterogeneity [16]. The calculation formula is [17]:

$$I = \frac{N\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\left(z_i - \bar{z}\right)\left(z_j - \bar{z}\right)}{\left(\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\right)\sum_{i=1}^{N}\left(z_i - \bar{z}\right)^2} \tag{1}$$

In the formula: I is the autocorrelation measurement indicator; N is the number of data points for suspended solids concentration in the study area; $\bar{z}$ is the average concentration of suspended solids; z_i and z_j represent the spatial positions of suspended solids concentration data, while w_{ij} represents the distance between z_i and z_j.

3.2. Geographically weighted regression model

The Geographically Weighted Regression model embeds the spatial location of data into regression parameters, and models the different spatial relationships between variables and dependent variables at different geographic locations [18]. The model is shown in formula (2):

$$y_i = \beta_0(\mu_i, v_i) + \sum_{k=1}^{p} \beta_k(\mu_i, v_i) x_{ik} + \varepsilon_i, \varepsilon_i \sim N(0, \varepsilon_i^2) \tag{2}$$

In this study, (μ_i, v_i) is the coordinate of the i-th suspended solids concentration inversion data (longitude and latitude projection coordinates, unit: m), $\beta_0(\mu_i, v_i)$ is the constant term of the i-th suspended solids concentration inversion data, $\beta_k(\mu_i, v_i)$ is the estimated coefficient of the k-th independent variable in the i-th suspended solids concentration inversion data, p represents the number of independent variables significantly affecting water quality distribution, and ε_i is the random error of the i-th suspended solids concentration inversion data.

3.3. Correlation analysis and collinearity test

When solving a geographically weighted regression model, if there is local multicollinearity between factors, the results of the model will become unstable. Therefore, it is necessary to conduct correlation analysis and multicollinearity testing between alternative influencing factors to ensure the independence of explanatory variables and improve the accuracy of the model. This article uses tolerance and variance inflation factor (VIF) indicators to jointly measure multicollinearity among influencing factors [19]. The variance inflation factor is the reciprocal of tolerance, and its calculation formula is:

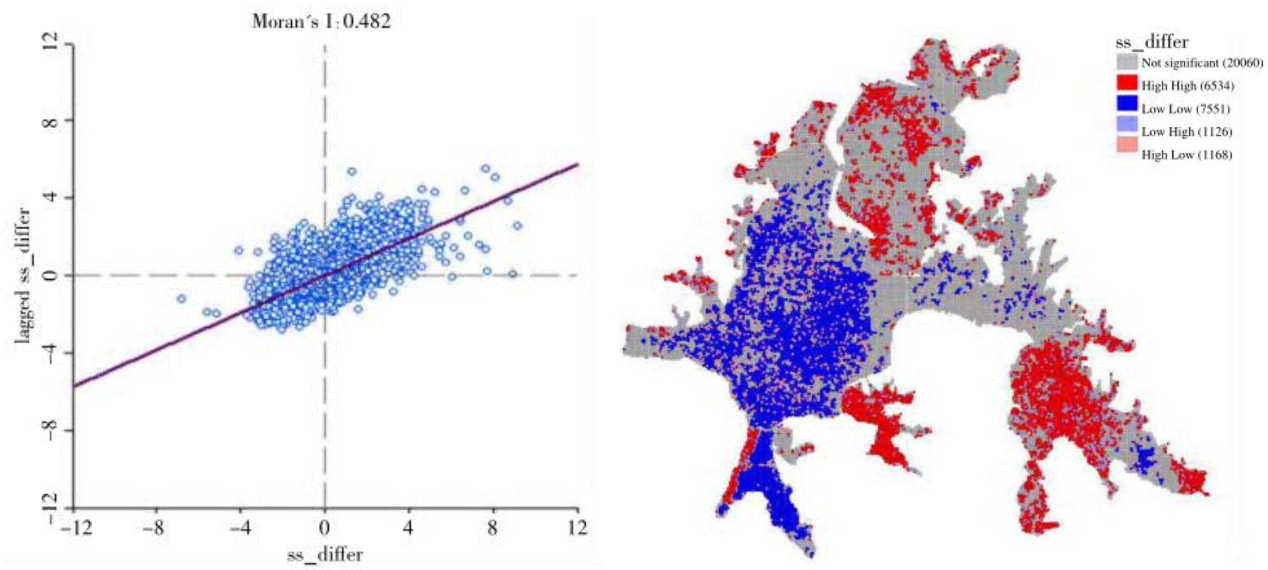
$$VIF = \frac{1}{1 - R_i^2} \tag{3}$$

In the formula, R_i represents the correlation coefficient between this influencing factor and other influencing factors, and i represents the numbering of different factors.

4. Results and discussion

4.1. Spatial autocorrelation analysis of road water quality before and after typical rainfall events

Figure 2 shows the local Moran index scatter plot and LISA clustering map of the spatial autocorrelation analysis of the changes in suspended solids concentration before and after rainfall, respectively. The Moran's index for the difference in suspended solids concentration before and after rainfall is 0.482, indicating a significant positive spatial autocorrelation (P<0.01), indicating that it tends to cluster in space. From the LISA clustering map, it can be observed that its spatial clustering characteristics are obvious. The northern and eastern parts of Donghu Lake show high high clustering, while the western part shows low low clustering, showing significant spatial heterogeneity.



(a) Local Moran exponent scatter plot (b) LISA clustering map

Figure 2. Spatial autocorrelation analysis of suspended solids changes before and after rainfall.

4.2. Factor correlation analysis and collinearity test in geographically weighted regression models

The correlation coefficients between the initially selected influencing factors listed in Table 1 are shown in **Table 2**. From Table 2, it can be found that among the eight selected factors, X_4 (road area), X_5 (road width), X_6 (road length), and X_8 (vehicle speed) have significant correlations, indicating a close relationship between these initial influencing factors, especially between road width and road speed, with a correlation coefficient of 0.992, indicating a high correlation between lane width, vehicle speed, and road capacity.

Table 2. Correlation coefficients of primary influencing factors.

Impact factors	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1							
X_2	-.585**	1						
X_3	-.372**	-.414**	1					
X_4	.109**	.256**	-.546**	1				
X_5	.680**	-.172**	-.689**	.681**	1			
X_6	.423**	.127**	-.740**	.844**	.912**	1		
X_7	-.907**	.462**	.595**	-.265**	-.794**	-.572**	1	
X_8	.656**	-.152**	-.680**	.705**	.992**	.920**	-.778**	1

Table 3 shows the collinearity diagnosis of the primary influencing factors. Among the eight primary factors, the VIFs of the primary factors for Building area (BR) and Road area (RR) are less than 10, with values of 8.946 and 5.005, respectively, indicating that these two factors do not have multicollinearity with other primary factors. The VIF of Forest area (FR), Green area (GR), Road length (RL), and Number of intersections (RC) are greater than 10 and less than 100, indicating a strong multicollinearity between these four factors and other primary factors. It is necessary to combine the results of correlation analysis to determine whether they should be retained.

Table 3. Diagnosis of collinearity of primary influencing factors.

Primary selection factor	Tolerance	VIF
Building area (BR)	.112	8.946
Forest area (FR)	.034	29.584
Green area (GR)	.023	44.062
Road area (RR)	.200	5.005
Road width (RW)	.009	111.304
Road length (RL)	.049	20.546
Number of intersections (RC)	.018	55.545
Road Speed (RS)	.009	107.368

Based on the correlation coefficients and multicollinearity test results of the preliminary selected influencing factors, this study chose to retain the influencing factors that have a high VIF but a correlation less than 0.7. Finally, five factors, namely Building area (BR), Forest area (FR), Green area (GR), Road area (RR), and Number of intersections (RC), were selected as the influencing factors for geographically weighted regression.

4.3. Analysis of geographical weighted regression model results

The geographical weighted regression R^2 and the distribution map of influencing factor coefficients are shown in **Figure 3**. The overall fitting accuracy of the model shows a decreasing circle layer relationship from outward to inward diffusion, that is, the fitting effect of the lake shore is good, but the fitting effect of the lake center is poor. The main reason is that the water quality on the lakeshore is highly influenced by influencing factors, and the water quality distribution in the lake center is relatively uniform due to the fluidity of the water, resulting in no significant relationship in the geographically weighted regression model.

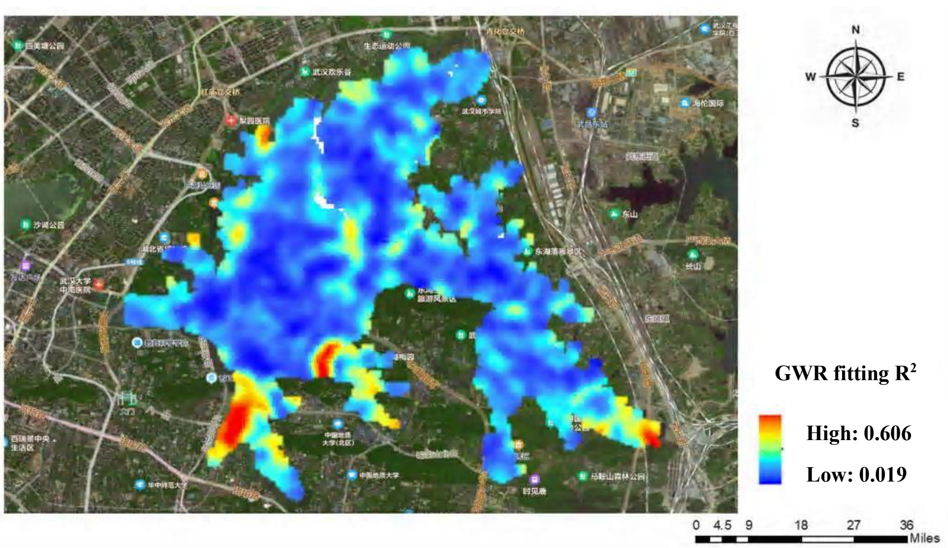


Figure 3. Geographical Weighted Regression R^2 Distribution Map.

The distribution diagram of the coefficients of the geographically weighted regression influencing factors is shown in **Figure 4**. Overall, the regression coefficients of each influencing factor exhibit significant non-stationary characteristics in geographical space. The impact capacity and effect of each influencing factor on water quality are different, further proving the spatial heterogeneity of the influencing factors.

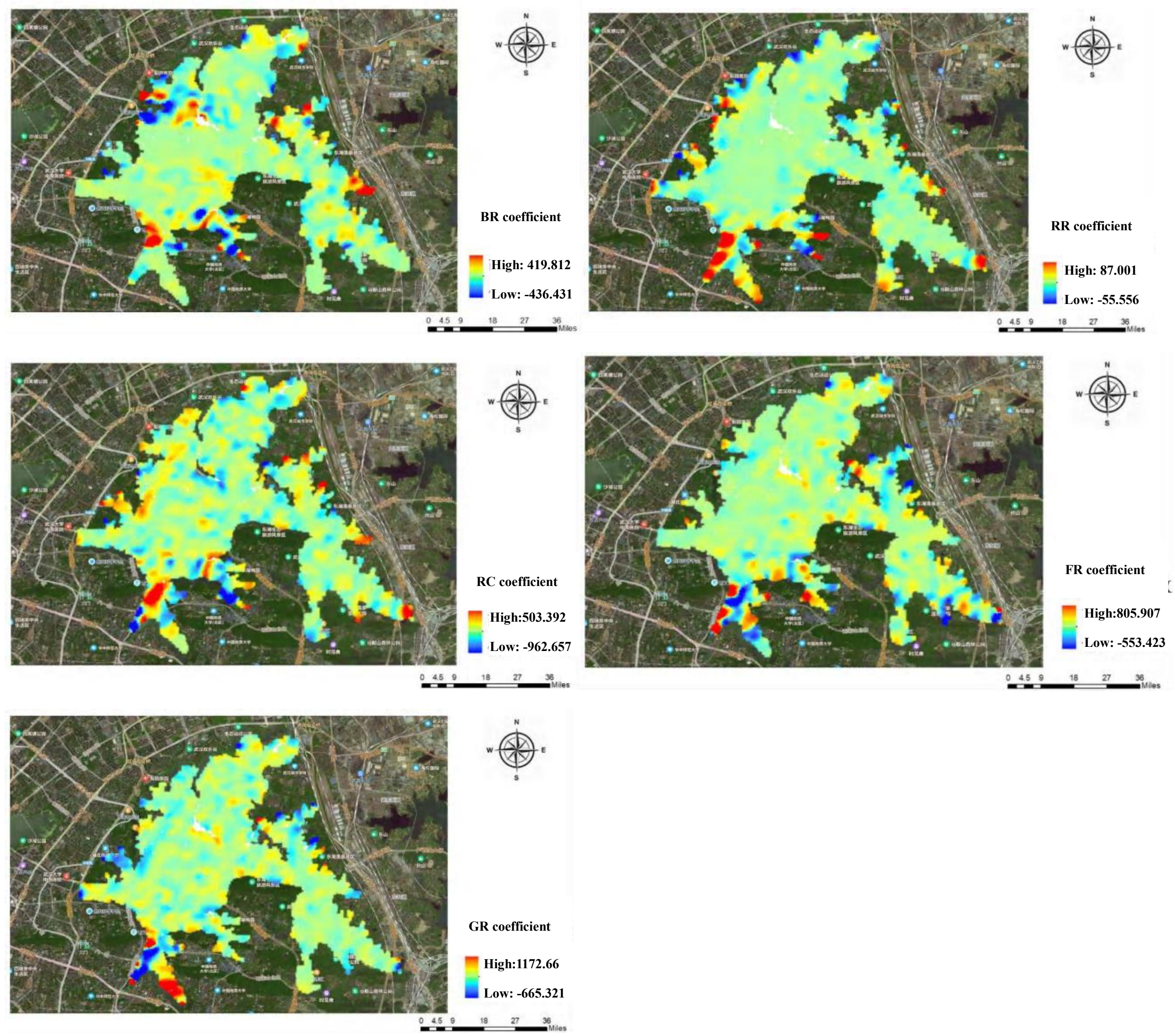


Figure 4. Distribution of Geographical Weighted Regression Independent Variable Coefficients.

(1) The road area, number of intersections, and suspended solids concentration are generally positively correlated in the water edge area, that is, the larger the road area and the more intersections there are, the higher the suspended solids concentration in that area. The main reason is that the larger the road area, the greater the traffic flow, and the more residual pollutants on the road.

(2) The area of buildings is positively correlated with the concentration of suspended solids in the edge area of water bodies, meaning that the larger the area occupied by buildings, the higher the concentration of suspended solids in that area. The main reason is that high-density areas of buildings will inevitably generate more roof rainwater runoff discharge, leading to intensified pollution of solid suspended solids in water bodies.

(3) The area of forests and green spaces is negatively correlated with the concentration of suspended solids in the edge areas of water bodies. The main reason is that forests and green spaces have the ability to intercept, absorb, and purify rainwater, reducing the pollution caused by rainwater runoff discharge to water bodies. The areas surrounding the positively correlated areas are densely populated with buildings and roads, with a large area of road hardening. The reduction effect of trees and green spaces on rainwater runoff is not significant.

5. Conclusion

This article takes the changes in suspended solids in urban lakes before and after rainfall events as the research object, and uses geographic weighted regression method to analyze its influencing factors. The results indicate that:

(1) The GWR model considering spatial heterogeneity has a good fit between the changes in suspended solids concentration before and after rainfall and the influencing factors.

(2) The coefficients of each influencing factor exhibit significant non stationarity in geographic space, further proving the spatial heterogeneity of the influencing factors.

(3) The area of buildings is positively correlated with the concentration of suspended solids in the edge area of water bodies, that is, the larger the area occupied by buildings, the higher the concentration of suspended solids in that area; The area of forests and green spaces is negatively correlated with the concentration of suspended solids in the edge areas of water bodies; The number of road intersections, road area, and suspended solids concentration are positively correlated in the water edge area.

There are still some shortcomings in the experimental process of this study, and the selection of influencing factors can be expanded. For example, starting from natural, social, built environment, transportation system, etc., a water quality impact index system for urban lakes can be constructed. Measure the influencing factors as comprehensively as possible to improve the accuracy of the constructed model, achieve more ideal fitting results, and make the research more scientific and accurate.

Conflict of interest

Author declare no conflict of interest.

References

1. Gu Yu, Wang Xuan, Fang Zheng. Investigation and Analysis of Pollution Characteristics of Rainfall Runoff in Typical Cities [J]. China Rural Water Resources and Hydropower, 2020 (06): 46-50,57
2. Xie Jifeng, Hu Zhixin, Xu Ting, et al. Analysis of water quality characteristics of rainfall runoff on different underlying surfaces in Hefei City [J]. Chinese Environmental Science, 2012, 32 (06): 1018-1025
3. Wang Zhu, Liu Yang, Wang Fang, et al. Spatial and temporal distribution characteristics and attribution analysis of water quality in Qinhuangdao [J]. China Rural Water Resources and Hydropower, 2020 (06): 63-71
4. Gong Miaomiao, Cai Chenghao, Miao Hanqian, et al. Characteristics and source analysis of rainfall runoff pollution on roads in different functional areas of Lin'an District [J]. Environmental Monitoring Management and Technology, 2019, 31 (04): 18-22
5. Luo Hui, Zhang Zeyu, Hu Xiaobo, et al. Research progress on heavy metal characteristics in road rainwater runoff [J]. Applied Chemistry, 2019, 48 (10): 2456-2461
6. Xie Weimin, Zhang Fang, Zhang Jingdong, et al. Research on the Changes and Treatment Methods of Urban Rainwater Runoff Pollutants [J]. Environmental Science and Technology, 2005 (06): 30-31, 49110-111
7. Pratt B, Chang H. Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales[J]. Journal of Hazardous Materials, 2012, 209-210: 48-58.
8. Zhao Jingya, Xu Chengcheng, Liu Pan. NO2 emission prediction model based on geographically weighted regression [J]. Journal of Safety and Environment, 2019, 19 (03): 964-970
9. Chen Xiaohui, Zeng Xiaoying, Zhao Chaochao, et al. Analysis of Road Network Ecological Effects Based on Remote Sensing Ecological Index: A Case Study of Fuzhou City [J]. Journal of Ecology, 2021, 41 (12): 4732-4745
10. Wang Q , Feng H , Feng H , et al. Analysis of the Impact of Road Traffic on the Urban Air Quality[J]. 2021.
11. Yuan C W , Rui X L , Wu D Y J P. Carbon Emission Reduction Pressure Index for China'sProvincial Transport Sector Based on Geographical Weighted Regression Model [J]. China Journal of Highway and Transport, 2016, 029(006): 262-270.
12. Hubei Provincial Local Chronicles Compilation Committee. Hubei Provincial Chronicles Geography [M]. Wuhan: Hubei People's Publishing House, 1997

13. Lu Dongshuo, Li Jian, AFlippi. Analysis of suspended solids concentration in Donghu Lake based on long-term Landsat images [J]. Journal of Wuhan University (Engineering Edition), 2019, 52 (10): 854-861
14. Geographic Monitoring Cloud Platform, Wuhan Land Use Data [EB/OL] [2023-02-17] http://www.dsac.cn/dataproduct/Index/301801.
15. Getis A，Ord J K. The Analysis of Spatial Association by Use of Distance Statistics[J]. Geographical Analysis，1992.
16. Zhang Songlin, Zhang Kun. Comparative study of global spatial autocorrelation Moran index and G coefficient [J]. Journal of Sun Yat sen University (Natural Science Edition), 2007 (04): 93-97
17. Sawada B M. The General Cross Product Statistics [J] two thousand and nine
18. Wan Gang, Cao Xuefeng. Historical Evolution and Reflection on Geospatial Information Grid [J]. Journal of Surveying and Mapping, 2016, 45 (S1): 15-22
19. Bai Xuemei, Zhao Songshan. Exploration of Several Issues in the Comprehensive Evaluation Method of Principal Component Analysis [J]. Statistical Research, 1995 (06): 47-51