An Integrated Approach for Identifying Key Driving Factors of Urban Exceedance Flooding Based on Spatio-Statistical Dual Perspective Using OPGD-MLR Modeling

Guangqi Liu[1], Meifang Ren[1]*, Jiacao Ma[2], Jing Li[1], Weichao Yang[2,3]*

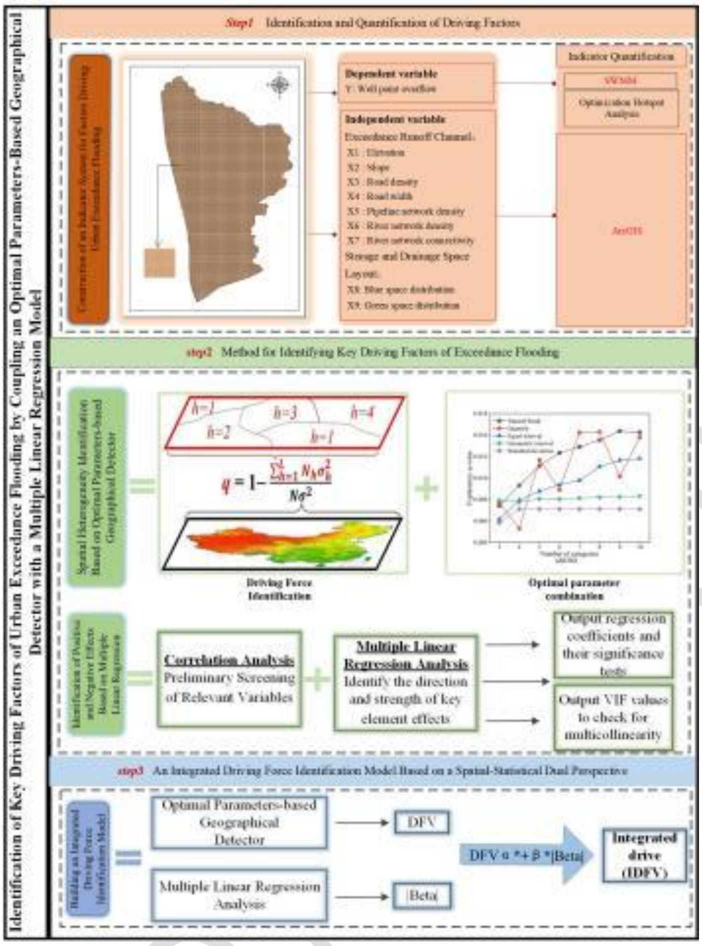
[1]China Academy of Urban Planning & Design, Beijing 100044, China.

[2]Tianjin Key Laboratory of Soft Soil Characteristics & Engineering Environment, Tianjin Chengjian University, Tianjin, 300384, China.

[3]School of Civil Engineering, Tianjin University, Tianjin, 300072, China.

*Correspondence: Meifang Ren (renmeifang@mail.bnu.edu.cn) and Weichao Yang (yangweichao@tju.edu.cn)

Graphical abstract



Abstract

Amplified by global climate change, the frequency and intensity of extreme precipitation events are escalating, leading to increasingly severe urban exceedance flooding. Identifying the key drivers of urban exceedance flooding carries profound theoretical significance and practical implications. Existing approaches face significant challenges in terms of comprehensiveness and precision for key drivers identification of urban exceedance flooding. The study proposes an integrated approach for identifying key driving factors of urban exceedance flooding based on spatio-statistical dual perspective, coupling the Optimal Parameters-based Geographical Detector (OPGD) with multiple linear regression (MLR). Firstly, an urban exceedance flooding potential driving factors system is constructed, categorized into two dimensions like exceedance runoff channel and storage-drainage space layout and totally nine indicators. Subsequently, the OPGD and MLR models are employed to quantify the influence of each driver on flooding intensity from the explanatory power for spatial heterogeneity and the direction-magnitude of statistical effects. Furthermore, by applying a weighted fusion of spatial and statistical association measures, the Integrated Driving Force Value (*IDFV*) index is proposed for each driver, representing its influence on flooding intensity. This methodology effectively integrates the respective strengths of the two models, enhancing the scientific rigor and reliability of the identification results and overcoming the limitations of the traditional Geographical Detector, which cannot indicate the direction of influence. Taking elevation as an illustrative example, the q-statistic under the conventional five-class average scheme is merely 0.035, whereas the optimal-parameter scheme elevates it to 0.192—an absolute increment of 0.158, corresponding to a relative improvement of approximately 456 percent. It is applied to the High-tech Industrial Development Zone, China, demonstrating enhanced accuracy and robustness in identifying the key drivers of exceedance flooding. In summary, this study provides a novel methodology for identifying the critical drivers of urban exceedance flooding, offering practical guidance for urban disaster prevention, mitigation, and spatial planning.

Keywords: Urban Exceedance Flooding; Optimal Parameters-based Geographical Detector; Multiple Linear Regression; Driving force; Urban Disaster Prevention

1 Introduction

Against the backdrop of escalating global climate change and rapid urbanization, flood disasters have emerged as one of the most severe natural hazards threatening human societal development (Karthik. et al., 2025; Selvanarayanan et al., 2024; Sivasubramanian et al., 2025; Sundarapandi et al., 2024). Particularly in highly urbanized areas, the swift increase of impervious surfaces coupled with the encroachment on natural reservoirs like wetlands and lakes has altered surface runoff generation and concentration processes, exacerbating urban waterlogging. As the complexity of urban waterlogging intensifies, the academic focus has shifted from addressing routine rainfall events to extreme flooding occurrences. Data indicate that since 2006, over 100 Chinese cities have experienced inundation annually (Chang et al., 2021; Gong et al., 2017), with the number reaching 258 in 2010 (Lee et al., 2018; Pirani & Modarres, 2020; Suthinkumar et al., 2023). Urban waterlogging disasters have thus become a critical constraint on sustainable socioeconomic development (Gurung, 2016). Notable events especially in China, such as the catastrophic Zhengzhou "7·20" rainstorm in 2021 (Liu & Sun, 2025), the extreme rainfall events in Beijing and Shenzhen in 2023 (Li et al., 2026), further underscore the inadequacy of traditional disaster prevention systems in coping with precipitation exceeding design standards. It demonstrates that urban flooding triggered by exceedance rainfall is a pressing issue urgently requiring resolution (Yang et al., 2017; Yeh et al., 2011). Facing the practical challenge, the accurate identification and quantification of the key driving factors of exceedance flooding, along with a profound understanding of their complex disaster formation mechanisms and interactions, have become a forefront and focal topic for both academia and engineering practice (Zhang et al., 2018). Consequently, conducting in-depth research on identifying the key driving factors of urban exceedance flooding carries significant theoretical value and practical importance.

The development of indicator systems for analyzing the driving factors of urban exceedance flooding demonstrates a distinct evolution from single-element assessment toward systematic integration. Kaźmierczak et al., (2011) established a unified framework for evaluating rainstorm-induced waterlogging losses by incorporating hazard factors, exposure, and vulnerability of the disaster-bearing entities. Hategekimana et al., (2018) and Seyoum et al., (2012) introduced a dynamic risk assessment concept, constructing an evaluation system comprising 18 indicators, including rainfall intensity, surface runoff generation capacity, and drainage system efficiency. Wang et al., (2022) utilized machine learning algorithms to develop an indicator system that integrates multi-source remote sensing and ground monitoring data, significantly enhancing the capability for dynamic assessment of waterlogging risk. Liu et al., (2022) applied a structural equation model to verify the moderating effect of socioeconomic indicators on waterlogging risk, thereby refining the dual-system framework of natural-social factors. In recent years, indicator systems have shown trends toward multidimensional refinement and increased granularity. Bell et al., (2018) innovatively integrated real-time traffic and social media data into the indicator system, achieving minute-level dynamic early warning for waterlogging risk. Bhattacharyya et al., (2022) employed principal component analysis to reduce 32 preliminary indicators to eight core indicators, establishing a more operational and simplified assessment system. Luppichini et al., (2023) introduced the concept of urban resilience, constructing a three-dimensional indicator system encompassing ecological background, engineering infrastructure, and management efficacy, offering a new perspective for waterlogging prevention and control. However, current indicator systems for factors influencing urban exceedance flooding still exhibit deficiencies in indicator selection, standardization, and scale matching. Chen et al., (2022) and Suthinkumar et al., (2023) noted that flood propagation during exceedance rainfall events depends on the synergistic interaction of slope concentration pathways, pipeline overflow nodes, and river channel conveyance bottlenecks, yet existing indicators struggle to quantify the critical failure mechanisms of these dynamic pathways. Liu et al., (2022) mentioned the dynamic migration of waterlogging hotspots in relation to drainage engineering modifications but did not establish indicators linking engineering layouts to runoff pathways. Based on water cycle balance analysis, Chang et al., (2023) and Liao et al., (2020) pointed out that the filling of urban water systems and wetland fragmentation directly weaken regional storage capacity, yet current research lacks quantitative indicators for the structural layout of blue spaces.

Thus, prevailing studies have predominantly focused on conventional indicators—such as rainfall intensity, impervious surface ratio, and drainage network density—exhibiting a discernible limitation in systematically capturing the distinctive disaster formation mechanisms of exceedance flooding. While most existing indicator systems can explain waterlogging under conventional rainfall conditions, they remain inadequate in characterizing the critical mechanisms—such as sudden changes in runoff pathways and the critical failure of storage and drainage spaces—during precipitation events that exceed design standards. Secondly, current indicators lack an in-depth association with spatial layout characteristics. Although some studies mention factors like topographic relief or green space ratio, they fail to further deconstruct the dynamic progression paths of exceedance flooding and their spatial coupling relationship with blue infrastructure. Taking the simulation of the Zhengzhou "7·20" rainstorm as an example (Liu & Sun, 2025), spatial factors such as river longitudinal slope and the distribution of urban depressions significantly influenced flood concentration pathways, yet existing indicators struggle to quantify such features effectively. Furthermore, the scale compatibility of indicator systems proves insufficient. Liu et al., (2013) and Song et al., (2019) noted that the absence of unified quantification standards often leads to scale mismatch issues across different studies. Macro-scale indicators (e.g., for urban agglomerations) frequently overlook localized drivers like micro-topography and pipeline network nodes, while micro-scale analyses fail to reflect the imbalance in regional flood storage capacity. Addressing these gaps, this study innovatively constructs an indicator system from two dimensions including exceedance runoff channelsand storage and drainage space layout, aiming to more systematically characterize the formation mechanisms and spatial heterogeneity of exceedance flooding. It addresses the oversight in traditional indicators regarding the relationship between dynamic pathways and spatial layout, offering a new perspective for the refined simulation and spatially targeted management of exceedance flooding.

Research on the identification of urban flooding driving factors and the analysis of their mechanisms has given rise to a diverse methodological framework, encompassing statistical correlation analysis, GIS-based spatial analysis, field investigations, multivariate statistical modeling, decision tree analysis, and empirical induction. Among all of them, GIS technology and statistical correlation analysis have evolved into predominant methodological pathways, owing to their respective strengths in handling spatially heterogeneous data and quantifying the influence of individual factors. A significant body of work focuses on the integration of multiple methods and the refinement of models to enhance both the accuracy of factor identification and the depth of mechanistic understanding. For instance, Diakakis et al., (2016) systematically evaluated twelve factors including elevation, slope, and drainage networks through an integrated application of GIS technology, logistic regression, and statistical correlation analysis. Their work revealed the significant impact of soil moisture content and cumulative rainfall on exceedance flooding events. Similarly, Lee et al., (2012) employed bivariate statistical analysis in conjunction with GIS tools to systematically elucidate the association mechanisms between natural/anthropogenic factors and the occurrence of flood disasters. To address the limitations of traditional bivariate methods in expressing probability and characterizing multi-factor interactions, Tehrany et al., (2013) combined binary probability analysis with multivariate logistic regression. They constructed a multi-dimensional set of independent variables, including precipitation, digital elevation models, slope, geological conditions, green space distribution, and river density. The predictive accuracy of the resulting flood susceptibility maps was subsequently validated, providing a scientific foundation for local flood management planning. Furthermore, Pradhan et al., (2010) utilized multiple regression models to examine the combined effect of several variables on the spatial distribution patterns of flooding. The integration of decision tree analysis with GIS technology has also matured as a robust approach, effectively synthesizing the advantages of both bivariate and multivariate statistical models to become a key technique for diagnosing the causes of waterlogging (Tehrany et al., 2014). In a comprehensive assessment, Mahmood et al., (2017) integrated seven distinct methodologies—ranging from rainfall characteristic analysis to water quality evaluation—and concluded that both natural conditions and socioeconomic factors jointly drive the increasing risk of flash flood disasters. Jin et al., (2005) applied field surveys to systematically identify potential causes and perform regional zoning for exceedance flooding in Shandong Province, China. Chen et al., (2014) employing Mann-Kendall trend tests, GIS tools, and empirical analysis, investigated the interplay of rainfall characteristics, topographic potential, and solid material source conditions, revealing the foundational driving factors of exceedance flooding formation in mountainous regions.

Although significant progress has been made in identifying the driving factors of urban exceedance flooding, several notable shortcomings remain. First, much of the existing research relies on single methodological approaches or simplistic bivariate analyses, lacking a comprehensive analytical framework capable of simultaneously capturing both the explanatory power for spatial heterogeneity and the positive/negative effects under statistical control of the driving factors. While GIS-based spatial analysis effectively visualizes factor distributions, it struggles to quantify their independent contributions. Conversely, statistical regression models can identify the direction of effects but are sensitive to multicollinearity and often overlook the complex spatial dependencies between variables. This singular analytical perspective frequently leads to biased identifications or insufficient explanatory power regarding the key drivers. Second, most studies fail to effectively integrate the identification of spatial heterogeneity in driving factors with the determination of their directional influence. Methods such as decision trees or bivariate analysis can identify associations between high-risk areas and certain factors but cannot ascertain whether a specific factor exacerbates or mitigates flooding risk after controlling for other variables. Traditional statistical models, while capable of determining the direction of influence, often neglect the local variability and spatial non-stationarity in how these factors exert their effects.

In light of above study limitaion, this study proposes a comprehensive approach for identifying the key driving factors of urban exceedance flooding. The method achieves a multi-dimensional analysis of both natural and anthropogenic driving factors by coupling the Optimal Parameters-based Geographical Detector (OPGD) (Yongze et al., 2020) with the Multiple Linear Regression (MLR) model (Agbiji et al., 2025). The specific research framework is detailed as follows. Firstly, based on the disaster formation mechanisms of urban exceedance flooding, a key driving factor index system is constructed with exceedance flooding intensity as the dependent variable and a total of 9 indicators from two dimensions - exceedance runoff conveyance channels and storage-drainage space layout - as independent variables. Secondly, the OPGD and MLR models are applied independently to quantify the influence of each driver from two complementary perspectives. Finally, The *IDFV* is constructed to evaluate the overall influence strength of each factor. This index is derived from a weighted integration of the q-statistic and the absolute value of the standardized regression coefficient (|*Beta*|), assigning respective weights to balance the contributions of spatial pattern influence and independent causal effect. The approach ensures the assessment concurrently considers a factor's influence on the spatial pattern and its independent statistical effect. It is applied to the High-tech Industrial Development Zone in China. A comparative analysis with results from traditional, single-method approaches highlights the advantages of the coupled model in terms of identification accuracy and mechanistic interpretation.

2 Approach for Identifying Key Driving Factors of Urban Exceedance Flooding

This study introduces a novel and objective methodology for the comprehensive identification of key drivers behind urban exceedance flooding, structured around three principal phases. Initially, establish and quantifyan indicator system for identifying the key driving factors of urban exceedance flooding. Subsequently, the Optimal Parameters-based Geographical Detector model is employed to identify the pivotal drivers, complemented by a Multivariate Linear Regression analysis to delineate the direction and magnitude of their individual effects. The third phase involves a weighted integration of spatial and statistical association measures to quantify the composite driving force exerted by each indicator on flooding intensity. Finally, a comparative analysis of results from individual versus the coupled methodological approach is conducted through a case study, demonstrating the enhanced analytical robustness and scientific reliability of the proposed framework. The overall research structure is delineated in Figure1.

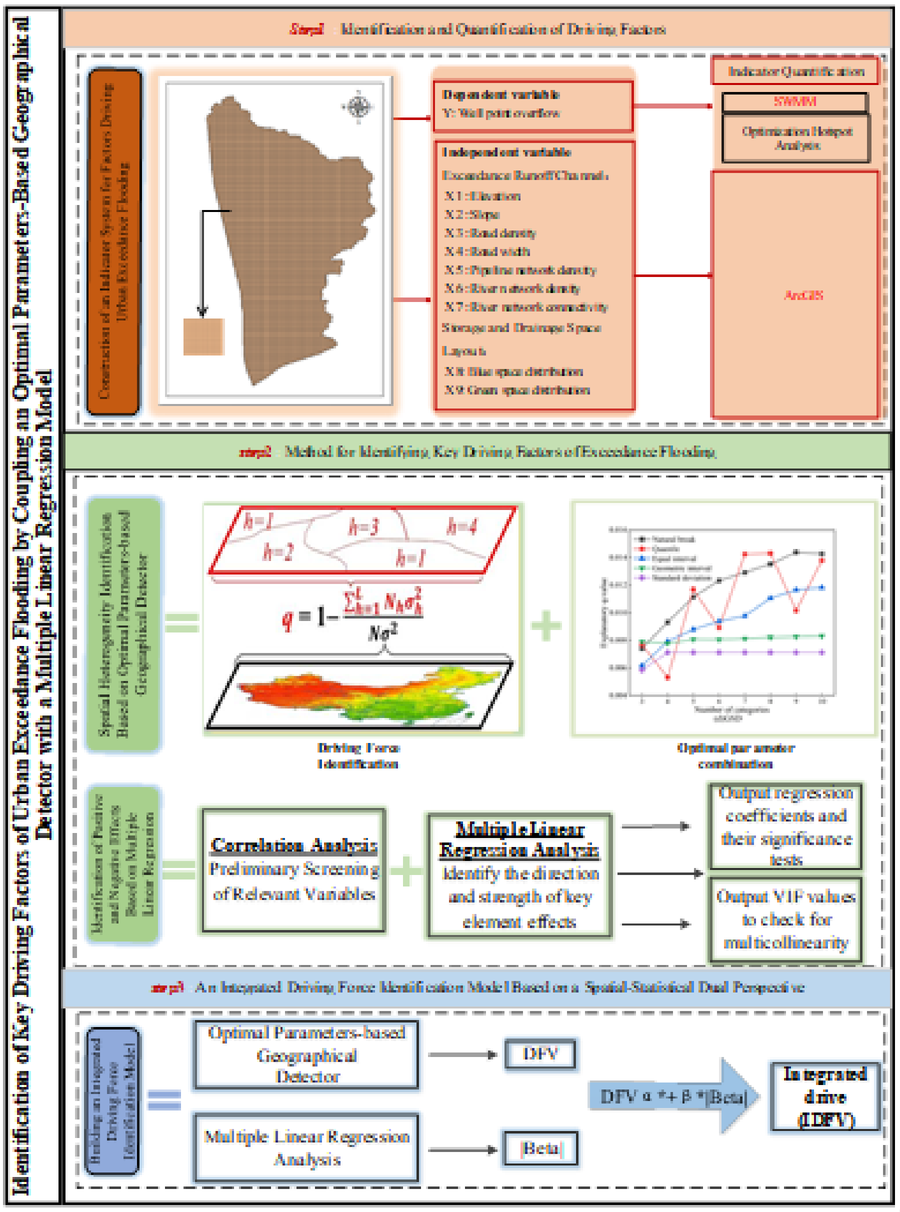


Fig. 1 Framework for identifying key driving factors of exceedance flooding

2.1 Identification and Quantification of Driving Factors

2.1.1 Construction of the Driving Factor System

Grounded in an in-depth analysis of the formation mechanisms of urban exceedance flooding and a systematic review of extant literature, this study establishes an indicator system for identifying key drivers. The system innovatively selects key influencing factors from two dimensions—exceedance runoff channel and storage and drainage space layout—as independent variables, with exceedance flooding intensity serving as the dependent variable. Based on criteria of correlation, representativeness, accuracy, and supporting literature, these variables can be further operationalized into more granular indicators. The framework of this indicator system is presented in Figure 2.

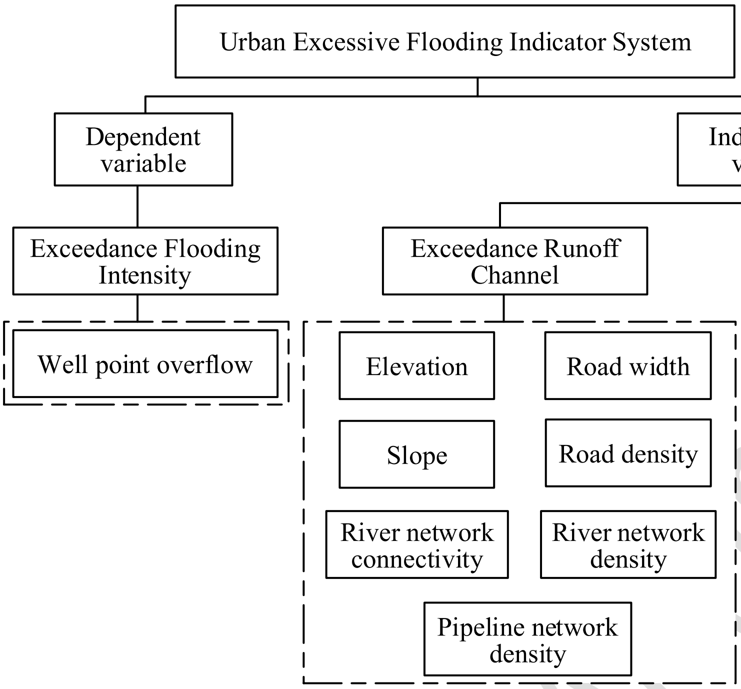


Fig. 2 Indicator system

2.1.1.1 Exceedance Flooding Intensity

The study on identifying key drivers of urban exceedance flooding introduces the concept of exceedance flooding intensity here. It is defined as the spatial aggregation degree of the total overflow volume at each manhole point. The concept quantifies the destructive capacity of floodwaters that result when rainfall intensity surpasses the design standard of an urban waterlogging prevention and control system. Synthesizing drainage system data from the study area, a one-dimensional pipe network model was constructed based on the Storm Water Management Model. The model simulates the manhole overflow process under a 100-year return period precipitation event. The spatial aggregation degree of the total overflow volume at each manhole point is then utilized as the exceedance flooding intensity value for the analytical unit.

2.1.1.2 Driving Factors

Conventional approaches to screening key drivers of urban exceedance flooding have predominantly focused on factors such as natural topography and gray infrastructure like drainage networks. Moving beyond this paradigm, and synthesizing insights from the disaster formation mechanism and existing research, this study proposes two novel, complementary dimensions for the systematic screening of key indicators like exceedance runoff channels (addressing how excess water is conveyed) and storage and drainage space layout (addressing where it can be stored or discharged). By constructing a comprehensive indicator system based on these two dimensions, this study aims to quantitatively detect the degree of influence each driving factor exerts on flooding intensity, thereby enabling the precise identification of core drivers responsible for exceedance flooding. This framework offers a theoretical foundation for refined prevention and control strategies against urban exceedance flooding.

(1) Exceedance Runoff Channel

Exceedance runoff channels refer to specific pathways designed to convey and safely discharge exceedance stormwater runoff when rainfall intensity surpasses the design standard of the conventional regional drainage system. Based on their formation mechanisms, these channels can be categorized into two types: firstly, actively engineered conveyance channels, including metrics such as pipe network density, road width, and road density; secondly, passively formed natural surface convergence paths, encompassing topography-related factors like elevation, slope, river network density, and river network connectivity. Together, these two categories constitute a multi-layered defense system against rainfall events exceeding design standards, aiming to systematically divert excess rainwater that surpasses the capacity of source control facilities and the drainage network.

(2) Storage and Drainage Space Layout

It refers to the core management framework for urban responses to exceedance rainfall runoff, whose systemic function manifests in the synergy between storage and drainage. The drainage function relies on flood discharge channels to safely convey and discharge exceedance stormwater, while the storage function depends on flood storage and regulation facilities to detain, retain, and regulate runoff. In terms of spatial carriers, urban storage and drainage spaces can be categorized into two types: firstly, natural storage spaces, including rivers and lakes, which are quantitatively characterized in this study through blue space distribution indicators; secondly, artificial storage spaces, such as green spaces and agricultural/forestry land, which can be quantified using green space distribution indicators.

2.1.2 Quantification of Driving Factors

To identify the key drivers of urban exceedance flooding and evaluate their spatial distribution characteristics and influence mechanisms, this study selects manhole overflow volume as the dependent variable to represent exceedance flooding intensity. A set of nine indicators is chosen as independent variables to construct the indicator system like elevation, slope, pipe network density, road density, road width, river network density, river network connectivity, green space distribution, and blue space distribution. Leveraging the robust spatial data processing and analytical capabilities of the ArcGIS platform, each indicator is quantified and expressed spatially. The specific quantification procedures are detailed as follows:

(1) Exceedance Flooding Intensity

To quantify urban exceedance flooding intensity, this study employs a SWMM model to simulate manhole overflow volumes under a 100-year return period extreme rainfall scenario. The spatial distribution of this intensity is then derived using the Optimized Hot Spot Analysis tool within the ArcGIS platform. Optimized Hot Spot Analysis is a spatial clustering process that reveals the aggregated patterns in the distribution of geospatial phenomena. Its principle is based on the local *Getis-Ord Gi** algorithm for spatial cluster analysis. This section details the procedure for applying this method to calculate the exceedance flooding intensity across the study area, as follows:

Initially, spatial cluster analysis is conducted on the manhole overflow volumes within the study area using a grid as the aggregation unit, deriving the *Gi** statistic value for each grid cell. These values are then standardized according to the following formula to obtain the normalized *GiZ* value for each grid. The *GiZ* value reflects not only the magnitude of the data but also its spatial clustering pattern, which is defined herein as the aggregation degree. Therefore, this study utilizes *GiZ* to represent the magnitude of exceedance flooding intensity. A higher *GiZ* value indicates a more concentrated distribution of high-intensity exceedance flooding areas, while a lower value suggests a more dispersed pattern. The principle formulas for *Gi** and *GiZ* are shown in Eq. (1) to (4):

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij} x_j - \overline{X} \sum_{j=1}^{n} w_{ij}}{S \sqrt{\frac{1}{n-1}\left[n \sum_{j=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_{ij} \right)^2 \right]}} \left(i, j = 1, 2, \ldots, n \right) \tag{1}$$

$$\overline{X} = \frac{1}{n} \sum_{j=1}^{n} x_j \tag{2}$$

$$S = \sqrt{\frac{1}{n} \sum_{j=1}^{n} x_j^2 - \left(\overline{X} \right)^2} \tag{3}$$

$$GiZ_i = \frac{GiZ_i - \mu_i}{\sigma_i} \tag{4}$$

Where, *Gi** represents the overflow volume statistic of exceedance flooding in grid *I*; *n* is the total number of all grids; w_{ij} is the spatial weight value between grid *i* and grid *j*, which is obtained through ArcGIS; x_j is the overflow volume of exceedance flooding in grid *j*; GiZ_i is the standardized value of *Gi**;μ_i is the mean of *Gi**; σ_i is the standard deviation of *Gi**.

(2) Elevation (DEM): It refers to the height of a specific point on the ground surface. As a primary factor governing runoff generation from precipitation, ground elevation is represented through a digital elevation model derived from existing topographic data of the study area. Utilizing available satellite-based DEM data, a raster map of elevation for the study area was generated by extracting the region of interest using the Extract by Mask tool in GIS software. The elevation metric was subsequently calculated by aggregating pixel values within the designated analytical units via the Zonal Statistics tool.

(3) Slope (Slope): It metric represents the gradient or steepness of a terrain surface unit. A higher slope value is positively correlated with an increased urban waterlogging risk index. Leveraging the Slope tool within GIS software, the Digital Elevation Model raster for the study area was converted into a corresponding slope raster. Subsequently, the average slope value for each designated analytical unit was calculated using the Zonal Statistics tool within the Spatial Analyst module.

(4) Pipe Network Density (PND): It is defined as the total length of drainage pipes per unit area. The calculation involved a three-step GIS procedure. First, the Intersect tool was employed to overlay the drainage network layer with the fishnet grid, generating a drainage network layer confined to the study area. Second, the pipe length within each grid cell was computed using the Calculate Geometry function. Finally, the PND for each cell was derived by dividing the pipe length by the grid cell area via the Field Calculator.

(5) Road Density (RD): It quantifies the total length of roads per unit area. The calculation followed a three-step GIS procedure. First, the Intersect tool was used to overlay the road network layer with the fishnet grid, thereby creating a road network layer specific to the study area. Next, the Calculate Geometry function was applied to obtain the total road length within each grid cell. Finally, the road density for each cell was computed by dividing the road length by the grid cell area using the Field

Calculator.

(6) Road Width (RW): It refers to the cross-sectional width of a road. The calculation procedure was as follows. First, road centerlines were digitized based on the road network layer. Second, a width attribute field was added and populated with values using GIS measurement tools. Subsequently, the Spatial Join tool was used to associate the road features with the fishnet grid cells. Within the field mapping settings, the merge rule for the road width field was set to "Mean." The process ultimately outputs the average road width value for each grid cell.

(7) River Network Density (RND): It is defined as the total length of rivers per unit area. The calculation followed a three-step GIS procedure. First, the Intersect tool was applied to overlay the hydrographic network layer with the fishnet grid, generating a hydrographic network layer specific to the study area. Second, the total river length within each grid cell was computed using the Calculate Geometry function. Finally, the RND for each cell was derived by dividing the river length by the grid cell area via the Field Calculator.

(8) River Network Connectivity (RNC): It quantifies the number of hydrographic nodes per unit area. The calculation procedure comprised the following steps. First, all nodes were extracted from the vector river network using the Feature Vertices to Points tool in GIS software. Subsequently, the Spatial Join function was employed to count the number of nodes contained within each grid cell of the fishnet. A field named "Connectivity Density" was added to the fishnet attribute table. Finally, the RNC value for each grid cell was computed by dividing the node count by the grid cell area.

(9) Green Space Distribution (GSD): It is defined as the area of green space per unit area. The calculation followed a standard GIS workflow. First, the Intersect tool was applied to overlay the green space layer with the fishnet grid, generating a green space layer specific to the study area. Subsequently, the area of green space within each grid cell was computed using the Calculate

Geometry function. Finally, the GSD value for each cell was derived by dividing the green space area by the grid cell area via the Field Calculator.

(10) Blue Space Distribution (BSD): is defined as the area of blue space per unit area. The calculation followed a standard GIS workflow. First, the Intersect tool was applied to overlay the blue space layer with the fishnet grid, generating a blue space layer specific to the study area. Subsequently, the area of blue space within each grid cell was computed using the Calculate Geometry function. Finally, the BSD value for each cell was derived by dividing the blue space area by the grid cell area via the Field Calculator.

2.2 Optimal Parameters-based Geographical Detector

The Geographical Detector was originally developed by Wang's research team and first applied in the field of public health to investigate the association mechanisms between health risk factors and the spatial distribution of diseases. Owing to its independence from classical statistical assumptions, such as homoscedasticity and normal distribution, and its effectiveness in handling categorical variables, the model has been widely adopted in recent years for spatial driver analysis across disciplines such as economics, meteorology, geology, and urban geography. However, the traditional Geographical Detector requires the discretization of continuous variables, a process where both the classification method and the number of classes often rely on subjective, arbitrary user definitions. This subjectivity can potentially constrain the model's explanatory power. To address this limitation, the Optimal Parameters-based Geographical Detector (OPGD) was developed. This enhanced model introduces an algorithmic optimization procedure. It systematically evaluates the explanatory power (q-value) of each continuous variable under various combinations of discretization methods (e.g., equal interval, quantile, natural breaks, geometric interval, and standard deviation) and different class numbers (ranging from 3 to 10). A higher q-value indicates a stronger explanatory capacity of that driving factor for the spatial heterogeneity of exceedance flooding intensity. The model used the Geographical Detector Version 2018 (http://geodetector.cn/) to select the discretization scheme that maximizes the q-value as the optimal parameter set. This approach significantly enhances the objectivity of the discretization process, thereby ensuring the model's analytical power for spatial patterns of exceedance flooding is optimized. This advancement facilitates a more robust and precise identification of the key drivers underlying the spatial patterns of flood disasters.

This study takes the spatial distribution of historical exceedance flooding events as the phenomenon under analysis and proposes nine driving factors influencing exceedance flooding intensity as detection variables. The factor detector is selected to quantify the extent to which each driving factor explains the spatial heterogeneity of exceedance flooding intensity. Its basic principle is shown in Figure 3, and the quantification formulas are shown in Eq. (5) to (7):

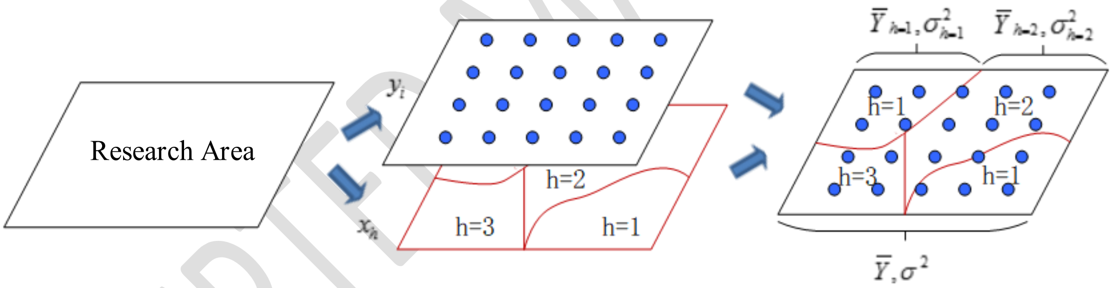


Fig. 3 Schematic Diagram of the Working Principle of a Geographic Detector

$$q = 1 - \frac{1}{N\sigma^2}\sum_{h=1}^{L} N_h\sigma_h^2 \qquad (5)$$

$$\sigma_h^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(N_{h,i} - \overline{N_h})^2 \qquad (6)$$

$$\sigma^2 = \frac{1}{N - 1}\sum_{j=1}^{N}(N_j - \overline{N})^2 \qquad (7)$$

Where, q represents the explanatory power of a single factor X on the spatial distribution of exceedance flooding intensity Y, which is defined here as the driving force. The value of q ranges from 0 to 1, where q=0 indicates that there is no relationship between X and Y, and q=1 indicates that Y is completely determined by X. N is the total number of analysis units in the study area; σ^2 is the global variance of exceedance flooding intensity Y in the study area; N_h is the number of analysis units in region h; σ_h^2 is the variance of exceedance flooding intensity Y in region h; L is the number of factor partitions; $N_{h,I}$ is the value of exceedance flooding intensity Y in the i-th analysis unit of region h; $\overline{N_h}$ is the average value of exceedance flooding intensity Y in region h; N_j is the exceedance flooding intensity value at the j-th grid in the entire study area, and represents the global average of exceedance flooding Y in the entire study area.

2.3 Multiple Linear Regression

Multiple Linear Regression (MLR) is a modeling method based on statistical principles. It quantifies the independent influence of various factors on a target variable by constructing a linear relationship between the dependent variable and multiple independent variables. The method employs the ordinary least squares (OLS) technique for parameter estimation, optimizing the model's fit by minimizing the sum of squared residuals. In this study, based on MLR, the driving factors influencing exceedance flooding are treated as independent variables, while the *GiZ-score* from Optimized Hot Spot Analysis is used to characterize exceedance flooding intensity as the dependent variable. Utilizing the linear regression analysis in IBM SPSS Statistics 23 (https://www.ibm.com/cn-zh/products/spss), the study investigates the correlation between exceedance flooding intensity and its influencing factors within the study area. This process elucidates the mathematical relationship between the dependent and independent variables, tests for multicollinearity among the driving factors, and establishes a quantitative relationship between exceedance flooding intensity and the drivers. By comparing the magnitude (absolute value) and sign (positive/negative) of the standardized regression coefficients, the relative influence strength and the direction (positive or negative effect) of different independent variables on the dependent variable are effectively identified. The principle formula is shown in Eq. (8):

$$Y = B_0 + B_1X_1 + B_2X_2 + \cdots\cdots + B_nX_n + \varepsilon \qquad (8)$$

Where, Y is the exceedance flooding intensity; X_1, X_2, ..., X_n are the driving factors affecting exceedance flooding intensity; B_1, B_2,..., B_n are the regression coefficients of the driving factors;B_0 is the constant term; ε is the random error after removing the effects of the n independent variables on Y, also called the residual.

2.4 Integrated Driving Force Identification Model Based on Spatio-Statistical Dual Perspective

To overcome the respective limitations of the Optimal Parameters-based Geographical Detector (OPGD) and Multiple Linear Regression (MLR) in identifying key drivers, this study proposes a novel identification method based on the coupling of these two models. The integrated approach aims to combine their strengths, enabling a multidimensional analysis of the key drivers behind urban exceedance flooding. The coupled model leverages the OPGD to objectively determine the optimal parameter set for discretization schemes, thereby quantifying the independent explanatory power of each driving factor on the spatial heterogeneity of flooding. Concurrently, it incorporates MLR to precisely identify the direction and magnitude of influence each factor exerts on flooding intensity while controlling for the effects of other variables, thus addressing the Geographical Detector's inability to characterize causal direction. Consequently, the integrated application of both methods allows for a macro-level understanding of which factors dominate the spatial differentiation, alongside a micro-level clarification of the specific effect size and direction of each factor. Building on this foundation, a composite driving force metric is constructed by assigning distinct, scientifically-grounded weights to the q-value (from OPGD) and the absolute value of the standardized regression coefficient (from MLR), respectively. This metric is used to evaluate the overall effect strength of each driving factor. The formula for calculating the Integrated Driving Force Value (IDFV) is shown in Eq. (9):

$$IDFV = \alpha \times DFV + \beta \times |Beta| \tag{9}$$

Where, DFV is the q value;$|Beta|$ is the absolute value of the standardized regression coefficient;α and β are weighting coefficients(α+β=1).The weight ratio can be adjusted according to research needs.For regions with significant spatial heterogeneity,the weight of the q value can be appropriately increased.In this paper, α is chosen as 0.6 and γ as 0.4.

3 Case Study

3.1 Research Area

The study area is situated within Southwest China (Figure 4). This region is characterized by significant topographic relief. Local low-lying areas are prone to water accumulation. Compounded by accelerated urbanization, the increase in impervious surface area has significantly raised the surface runoff coefficient. Concurrently, aging drainage infrastructure in some areas, coupled with insufficient pipeline capacity, makes the region susceptible to exceedance flooding during heavy or prolonged rainfall events. In light of these conditions, this study selects the High-tech Zone as the research area to identify the key drivers of urban exceedance flooding, quantify the influence degree of various factors, and thereby provide scientific guidance for enhancing urban resilience and mitigating disaster losses. A detailed description of the data used is provided in Table 1.

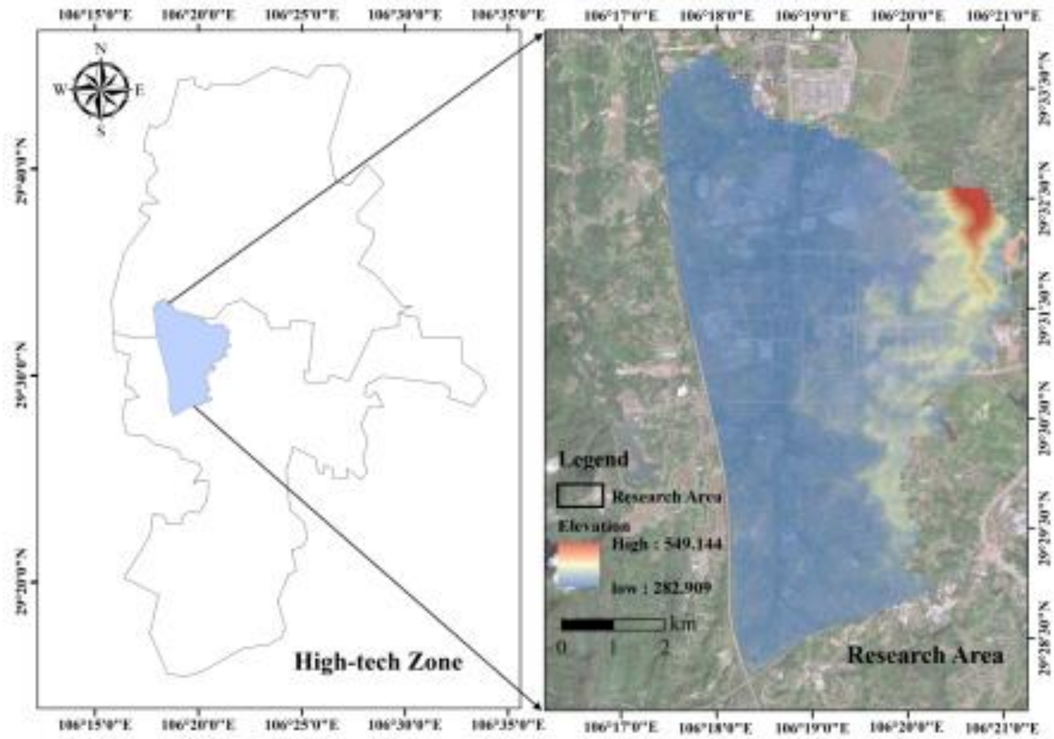


Fig. 4 Map of the Study Area's Geographical Location

Table 1 Research data

Data Category		**Factor**	**Data Description**
Exceedance flooding intensity		Well point overflow	Model simulation result
Key influencing factors	Exceedance runoff channel	Elevation	The spatial resolution is 6m × 6m
		Slope	Generated from DEM data
		Pipeline network density	Generated from pipeline network data

Data Category	Factor	Data Description
	Road density	Generated from road data
	Road width	Generated from road data
	River network density	Generated from hydrological data
	River network connectivity	Generated from hydrological data
Storage and drainage space layout	Green space distribution	Generated from land use data
	Blue space distribution	Generated from land use data

3.2 Quantitative Results of Driving Factors

3.2.1 Spatial Distribution of exceedance flooding intensity

3.2.1.1 Exceedance flooding total overflow volume Simulation and validation

Given the absence of field-monitored waterlogging data under observed rainfall conditions in the study area, direct validation of the two-dimensional model's accuracy was not feasible. Accordingly, overflow volume was employed as a proxy metric to represent exceedance flooding intensity, and the runoff coefficient was adopted as the calibration target for parameterizing the rainfall-runoff model. This approach effectively addresses the challenge of model parameter calibration in regions lacking empirical observational data. Specifically, the composite runoff coefficient used in the design of stormwater drainage networks served as the a priori benchmark, against which the simulated runoff coefficient from the model was compared to calibrate the model parameters. Three rounds of iterative adjustments were performed on the following parameters: impervious roughness, pervious roughness, impervious depression storage, pervious depression storage, maximum infiltration rate, minimum infiltration rate, decay constant, and antecedent dry days (Table 2). Following calibration, the model was validated against rainfall events with 10-year and 20-year return periods (Table 3), and the results fell within acceptable ranges. This procedure thereby effectively resolved the issue of model calibration and validation in the absence of locally observed flow data.

Table 2 Parameter Calibration Table

Parameter	**Initial value**	**Iteration 1**	**Iteration 2**	**Iteration 3**
Maximum infiltration rate (%)	3	25.4	26.3	30.1
Minimum infiltration rate (%)	0.5	2.3	2.6	3.2
Decay constant	4	4.5	4.6	5
Antecedent dry days (day)	7	8	8.5	9
Impervious roughness	0.018	0.0183	0.01831	0.01832
Pervious roughness	0.1	0.15	0.151	0.152
Impervious depression storage (mm)	0.05	1.27	1.3	2.3
Pervious depression storage (mm)	0.05	2.54	2.82	3.31
Simulation runoff coefficient	0.880347	0.70372	0.692901	0.64608

Table 3 Verification Table

Scenario	**Simulation Runoff Coefficient**	**Design Runoff Coefficient**
Verified for 10 years	0.694477	0.639831416 ~ 0.739831416
Verified for 20 years	0.728742	

3.2.1.2 Spatial Distribution results of exceedance flooding intensity

To accurately characterize the spatial influence features of urban exceedance flooding intensity, this study quantified its spatial distribution using the Optimized Hot Spot Analysis method within ArcGIS10.8 (https://desktop.arcgis.com/zh-cn/desktop/index.html). The results are presented in Figure 5. The research indicates that the spatial distribution of exceedance flooding intensity within the study area exhibits significant areal clustering characteristics. High-value clusters are predominantly concentrated in the central-northern region, while the southern and most peripheral areas generally exhibit lower intensity. Overall, high-intensity zones are relatively concentrated, whereas low-intensity zones are more dispersed. This spatial differentiation pattern can be summarized as high in the core, low at the edges; severe in the north, mild in the south.

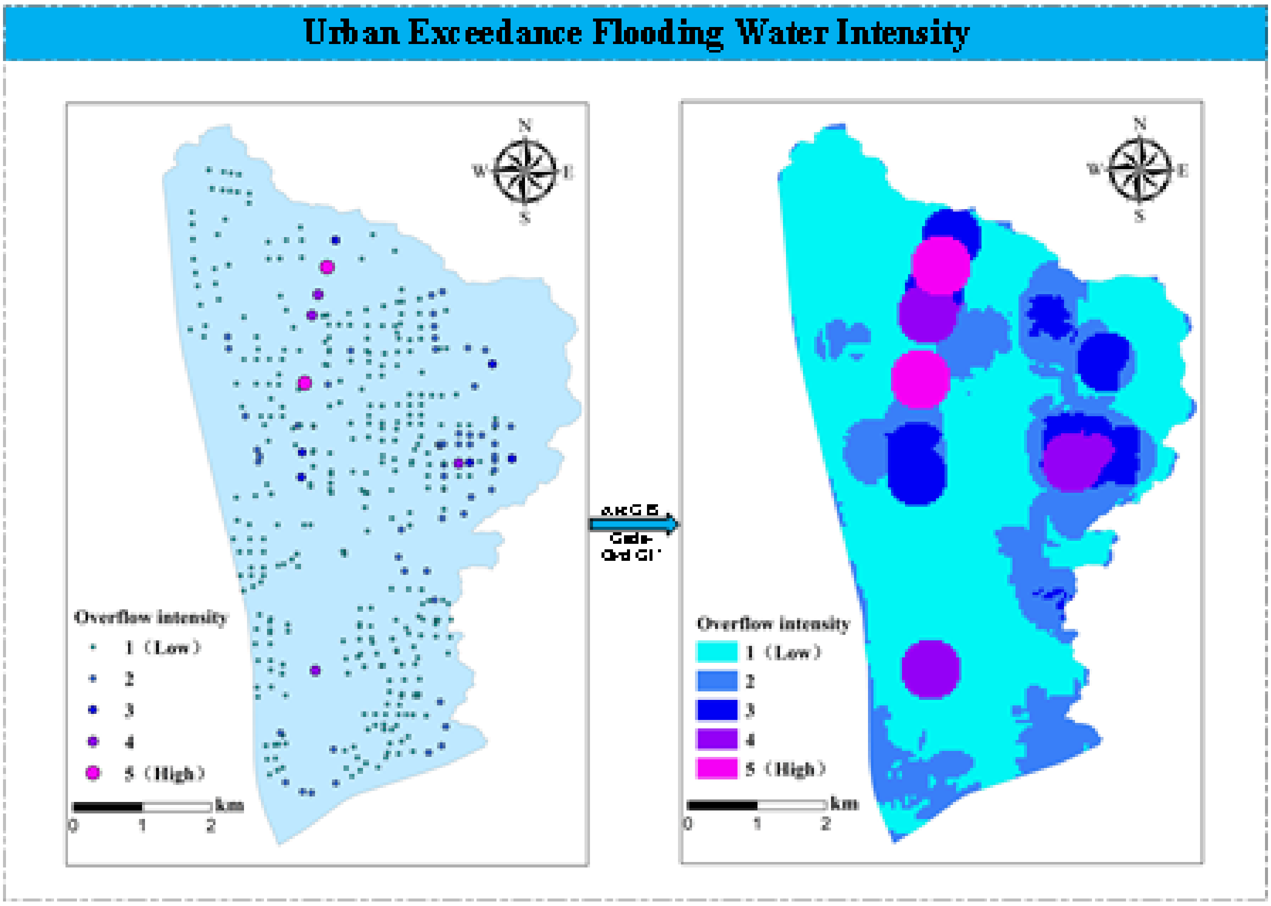


Fig.5 Spatial Distribution of Exceedance Flooding Intensity

3.2.2 Spatial Distribution of Exceedance Runoff Channel

To identify the influence of exceedance runoff channels on exceedance flooding intensity in the study area, this study conducted a spatial quantitative analysis of elevation, slope, pipe network density, road density, road width, river network density, and river network connectivity using ArcGIS. The results are shown in Figure 6. The analysis reveals significant spatial heterogeneity in the geographic factors. The topography exhibits a general stepped pattern, higher in the southeast and lower in the northwest. Roads and drainage networks are highly concentrated in the central and northwestern parts, forming high-density urbanized areas, which stand in stark contrast to the low-density eastern and southern regions. The overall connectivity of the water system is favorable, and the high-density urbanized areas are precisely located in the middle and lower reaches of the river network.This indicates that the urban core areas spatially couple high-intensity road and pipe networks and are situated in critical locations for catchment concentration and flow conveyance. They not only receive runoff inputs from upstream catchment areas but also generate high-intensity runoff from their own extensive impervious surfaces, which directly impacts the hydrological processes of downstream river sections. The formation and progression of regional flooding are significantly influenced by this synergistic effect of natural topography–urban infrastructure–river network structure.

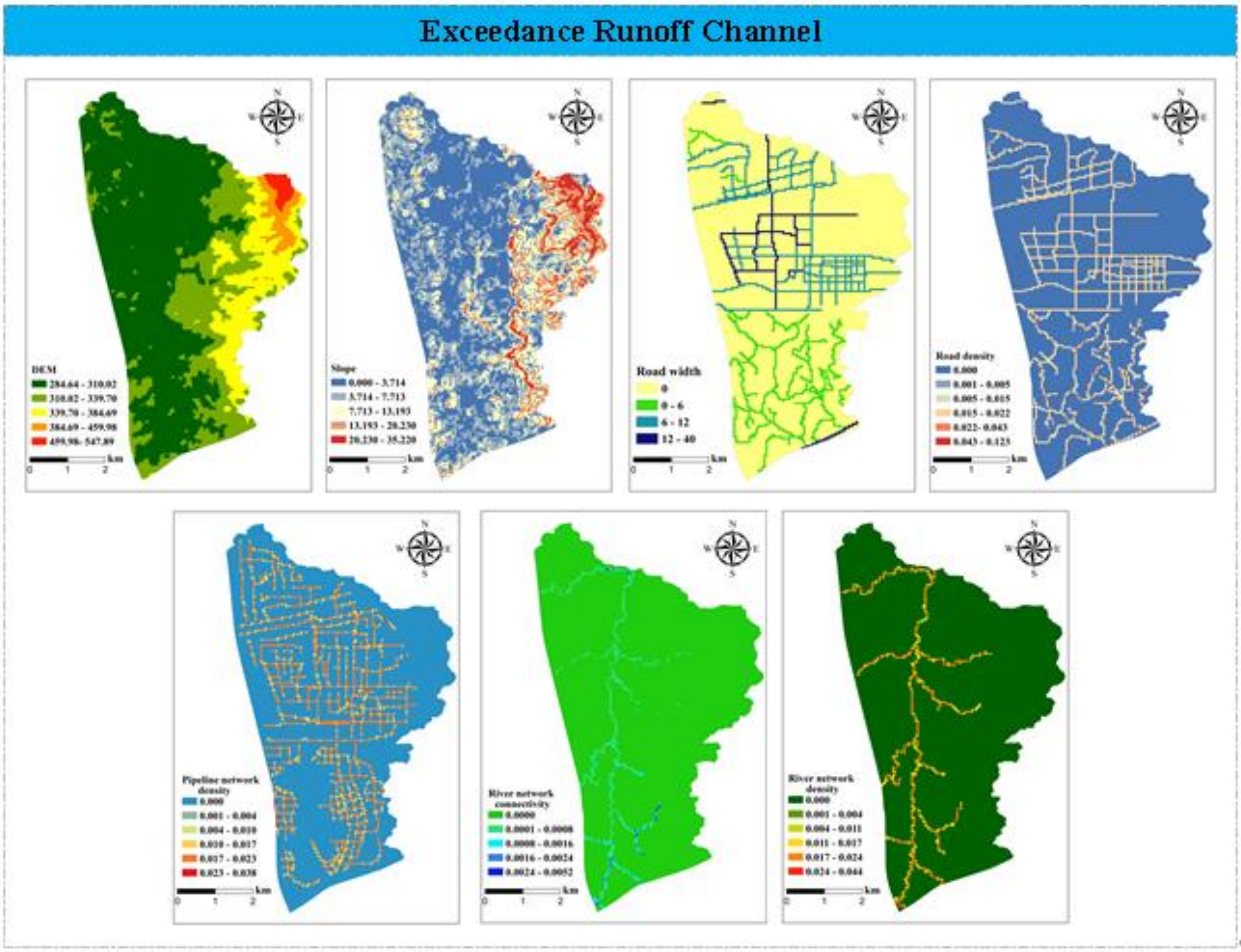


Fig.6 Spatial Distribution of Exceedance Runoff Channel Indicators

3.2.3 Spatial Distribution of storage and drainage space layout

The quantification results for the storage and drainage space layout indicators are presented in Figure 7. The figure clearly illustrates the spatial heterogeneity patterns of blue and green spaces within the study area. The distribution of blue space exhibits a discontinuous, patchy pattern, with its high-value areas concentrated in zones containing water bodies, wetlands, and lakes. In contrast, the distribution of green space is relatively extensive, with its high-density zones corresponding to areas of concentrated green spaces, parks, and regions with high vegetation coverage. A certain degree of spatial complementarity exists between the two. Together, they constitute the natural storage and infiltration underlying surface of the study area.

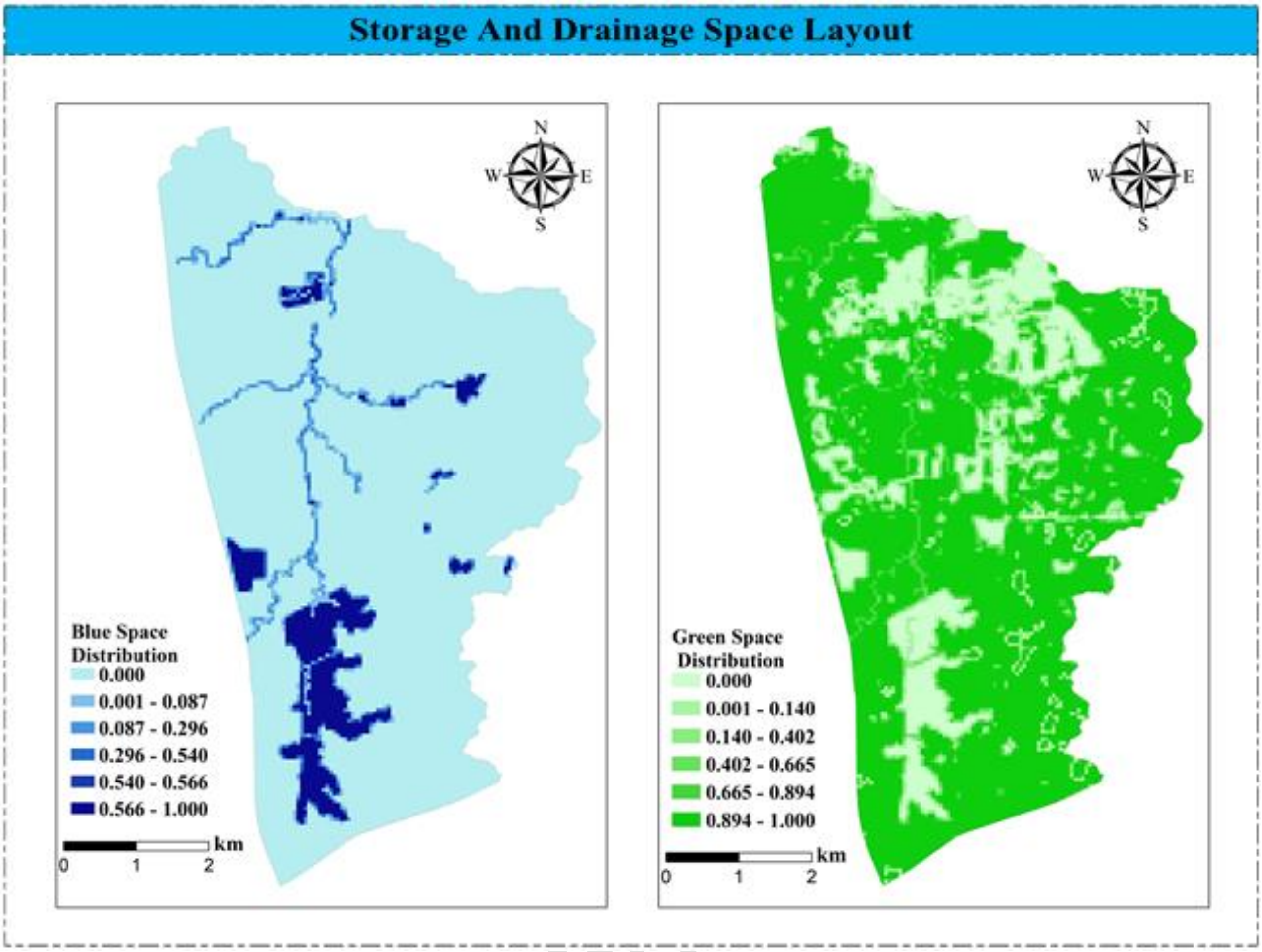


Fig.7 Spatial Distribution of Storage and Drainage Space Layout Indicators

3.3 Driving Force Identification Results of Exceedance Flooding Driving Factors

3.3.1 Spatial Relationship Analysis Results

The parameter combination yielding the highest *q-value* is selected as the optimal discretization scheme. Taking the discretization results of elevation, pipe network density, river network density, and green space distribution as examples when considering overflow volume as the dependent variable: for the elevation factor, the q-value is maximized when using the quantile interval method with 10 classes; therefore, this method is selected to classify elevation into 10 categories. Similarly, the geometric interval method is chosen to classify pipe network density into 8 classes, and the same method is applied to river network density with 8 classes. The natural breaks method is selected to classify green space distribution into 9 classes. The optimal discretization schemes for other continuous driving factors are detailed in Table 4.

Table 4 Optimal Discretization Results Table of Geographic Detector

Driving Factor	**Classification method**	**Number of categories**	***q***
Elevation	Quantile	10	0.19211248
Slope	Geometric interval	10	0.000425437
Pipeline network density	Geometric interval	8	0.002504265
Road density	Natural break	10	0.008003684
Road width	Natural break	4	0.02496086
River network density	Geometric interval	8	0.009592402
River network connectivity	Natural break	10	0.009176541
Green space distribution	Natural break	9	0.01437928
Blue space distribution	Natural break	10	0.021590991

The analysis of key drivers for exceedance flooding in the study area, based on the model results, indicates significant differences in the explanatory power (*q-value*) of each factor regarding the spatial heterogeneity of flooding. According to the magnitude of the *q-values*, the drivers can be categorized into three tiers of influence:The core driver is elevation, whose *q-value* far exceeds that of other factors. This indicates that natural topographic conditions are the fundamental element dominating the spatial pattern of flooding, establishing the spatial baseline of flooding risk by controlling the basic direction of surface flow concentration and defining convergence areas.Important influencing factors, forming a second tier, include road width, blue space distribution, and green space distribution. This suggests that, apart from natural topography, the capacity of artificially planned flood discharge channels and the layout of storage spaces provided by blue-green ecological infrastructure play a significant regulatory role in flooding intensity. These represent key actionable layers for mitigating waterlogging risks.Factors including river network density, river network connectivity, road density, pipe network density, and slope all have q-values below 0.01, classifying them as secondary influencing factors. This implies that, within the study area, the direct explanatory power of any single one of these factors on flooding intensity is relatively limited. Their effects are more dependent on spatial coupling and interaction with the core and important factors, or may only demonstrate more pronounced influence on a larger regional scale.

3.3.2 Statistical Relationship Analysis Results

Based on the Multiple Linear Regression (MLR) model, this study conducted correlation and multicollinearity analyses on the driving factors, screened the optimal factors, and constructed the best-fitting regression model. The results are presented in Table 5. The analysis reveals no significant correlations among the nine driving factors, and the model demonstrates a good fit. The study indicates that elevation, road density, river network connectivity, green space distribution, and blue space distribution are negatively correlated with overflow volume, while the remaining driving factors show positive correlations. Furthermore, the significance test (*p-values*) for pipe network density and road density exceeded 0.05, whereas the p-values for all other driving factors were less than 0.05.

Based on the magnitude of the standardized regression coefficients, the influence of the various driving factors on exceedance flooding intensity can be categorized into three tiers:The core driver is road width, whose coefficient value is significantly higher than that of other factors. It indicates that road width is the most influential factor. Wider roads act as crucial discharge channels during extreme rainfall events, effectively concentrating and rapidly removing surface runoff—a finding that has been repeatedly validated in urban waterlogging research.Important influencing factors, constituting the second tier, include river network density, elevation, river network connectivity, blue-green space distribution, and road density. Apart from elevation, these factors reflect the significant regulatory effect that the capacity of artificially planned flood discharge channels and the storage space layout of blue-green ecological infrastructure have on flooding intensity. The standardized coefficient for elevation is 0.105, which is at a relatively high level. Combined with the physical principle that low-lying areas are prone to becoming flood concentration zones, this further confirms that topography is a fundamental factor controlling the spatial distribution of flooding.The standardized regression coefficients for both pipe network density and slope are below 0.01, classifying them as secondary influencing factors. This suggests that, within the study area, the explanatory power of these individual factors on flooding intensity is relatively limited. Their dominant role is superseded by surface discharge channels and storage space factors.

Table 5 Multiple Linear Regression Analysis Results

Driving Factor	***Beta***	***Significance***	***Collinearity***
Elevation	-0.105	0.000	1.064
Slope	0.001	0.000	1.001
Pipeline network density	0.008	0.859	1.026
Road density	-0.058	0.311	2.037

Driving Factor	*Beta*	*Significance*	*Collinearity*
Road width	0.160	0.000	2.058
River network density	0.123	0.000	2.351
River network connectivity	-0.080	0.000	2.332
Green space distribution	-0.075	0.000	1.272
Blue space distribution	-0.015	0.000	1.345

The analysis based on the sign (positive/negative) of the standardized regression coefficients reveals a dichotomous pattern in the direction of influence that driving factors exert on urban waterlogging risk, as illustrated in Figure 8. The study demonstrates significant differences in the direction and strength of influence among different categories of drivers. Factors including elevation, road density, river network connectivity, and blue-green space distribution exhibit significant negative effects, meaning an increase in their values contributes to the mitigation of flooding intensity. Taking elevation and river network connectivity as examples: areas at higher elevations are typically located in the upper reaches of catchments, allowing surface runoff to drain naturally downstream via gravity, thereby reducing the risk of local ponding and manhole surcharge. Conversely, low-lying areas are prone to water accumulation; if drainage capacity is insufficient, the risk of overflow increases significantly. Higher river network connectivity enhances the storage and conveyance capacity of the water system, enabling rapid distribution and accommodation of localized intense runoff through the networked channels, thus suppressing flooding risks caused by poor drainage. It reflects the inherent mitigating effect of natural topographic conditions and ecological infrastructure on waterlogging risk.In contrast, slope, pipe network density, road width, and river network density show significant positive effects, where an increase in their values exacerbates flooding risk. Considering pipe network density and road width: high-density drainage networks form an efficient convergence network. While they accelerate the concentration of stormwater from upstream areas, they can create conveyance bottlenecks if the downstream system's capacity is inadequate, leading to a sharp increase in pressure in lower sections and a higher probability of manhole overflow. Roads, as primary impervious surfaces, contribute to increased risk in two ways when widened: firstly, by expanding the runoff-generating area, which increases total runoff volume and the load on the drainage system; secondly, by facilitating rapid lateral concentration of runoff across the wider surface. The results in a large volume of runoff impacting inlet structures within a short time, generating instantaneous peak loads at network entry points. If the inlet capacity is insufficient, this can easily trigger surface ponding and overflow.

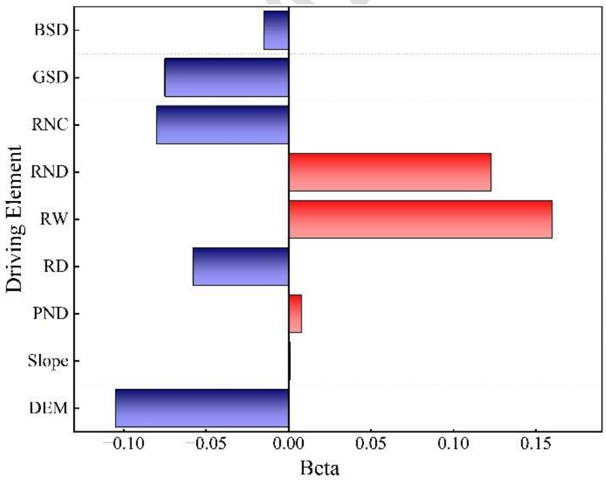


Fig.8 Driver Factors Positive and Negative Effects

3.3.3 Integrated Driving Force Value Identification Results

By integrating the Optimal Parameters-based Geographical Detector (OPGD) and multiple linear regression (MLR), we quantify the influence of different factors on exceedance flooding intensity and identify the most critical driving factors, as detailed in Table 6.

Table 6 Integrated Driving Force Value Results

Driving Factor	***q***	**|*Beta*|**	***IDFV***
Elevation	0.19211248	0.105	0.157267
Slope	0.000425437	0.001	0.000655
Pipeline network density	0.002504265	0.008	0.004703
Road density	0.008003684	0.058	0.028002
Road width	0.02496086	0.160	0.078977
River network density	0.009592402	0.123	0.054955
River network connectivity	0.009176541	0.080	0.037506
Green space distribution	0.01437928	0.075	0.038628
Blue space distribution	0.021590991	0.015	0.018955

The OPGD excels in identifying key drivers from the perspective of "spatial distribution" through its q-statistic, whereas the standardized regression coefficients of MLR are proficient in quantifying the "statistical magnitude" and direction of each factor's influence. The synthesized results, combined with the findings presented in Figure 9, allow for a clear identification of the key drivers of exceedance flooding in the study area and their hierarchical influence.Elevation and road width, with Integrated Driving Force Values (*IDFV*) of 0.157267 and 0.078977, respectively, significantly surpass other factors, collectively constituting the core echelon of drivers for flooding risk. This confirms the fundamental control exerted by the natural topographic context on runoff convergence patterns, as well as the pivotal role of high-intensity impervious surfaces as rapid runoff generation and conveyance pathways. These two factors are the primary forces governing the spatial heterogeneity and intensity distribution of flooding within the study region.River network density, green space distribution, and river network connectivity, with *IDFV* values ranging between 0.037 and 0.055, form a secondary tier of influential factors. The result underscores the significant regulatory and buffering functions that the structural characteristics of the natural water system and the spatial configuration of ecological land exert on the flooding process, representing an intermediate yet indispensable layer of factors in systemic flood management.In contrast, blue space distribution, road density, pipe network density, and slope, all with *IDFV* values below 0.03, are classified as tertiary factors. This indicates that, within the study area, the static distribution of water bodies, the localized density of road and drainage networks, and micro-topographic variations possess relatively limited independent and overall explanatory power for flooding intensity. Their influence is likely more contingent upon, or subordinate to the core and secondary factors described above.

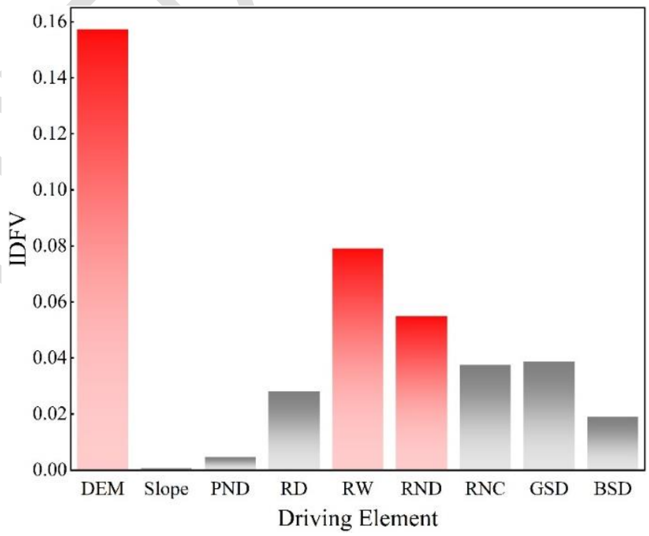


Fig.9 Results for the Coupling Method of Driving Factors

4 Discussion

4.1 Rationality Analysis

Identifying the key drivers of exceedance flooding provides crucial guidance for urban waterlogging prevention and control. It enables the prioritized allocation of limited resources to extremely high-risk areas, thereby mitigating flood risks at their source. Consequently, the scientific and rational identification of these key factors within a study area is of paramount importance, as the choice of identification method directly impacts the validity of high-risk factor determination. To enhance the reliability of identification results, the rationale for employing an integrated driving force identification method that couples the Optimal Parameters-based Geographical Detector with multiple linear regression models is elaborated below.

To rigorously identify the key driving factors of urban exceedance flooding, this study initially employs both the Optimal Parameters-based Geographical Detector (OPGD) and the multiple linear regression (MLR) model, conducting factor detection from the two dimensions of explanatory power for spatial heterogeneity and statistical effects, respectively. Subsequently, the q-statistic and the standardized regression coefficient (*Beta*) for each driving factor are calculated based on the respective methods. Finally, a comprehensive driving force index, the Integrated Driving Force Value (*IDFV*), is constructed through a weighted fusion of the q and |*Beta*| values, assigning appropriate weights to each, in order to systematically evaluate the overall driving strength of each factor.The *IDFV*, integrating both q and |*Beta*|, essentially combines the dual logics of spatial heterogeneity analysis and numerical statistics. This design aligns with the complexity of urban flooding driving mechanisms—which encompass both geographical spatial differentiation and the interaction of multiple factors. The rankings of the driving factors exhibit some variation based on the different identification methods, as shown in Figure 10.

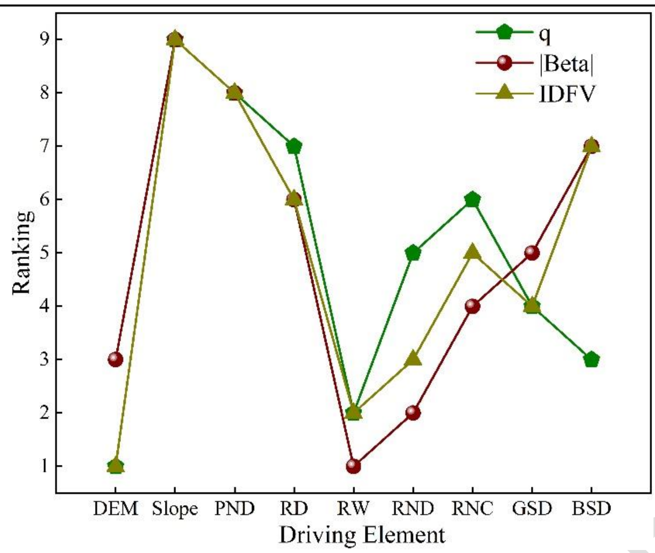


Fig.10 Ranking Fluctuation Chart of Three Identified Driving Factors

The chart reveals that the ranking curves for elevation, slope, and pipe network density demonstrate a high degree of consistency. Elevation consistently ranks at the top (positions 1–3) across all three methods, indicating that topographic factors are dominant drivers. The ranking results for slope and pipe network density in the individual methods align with their *IDFV* rankings, validating the reliability of the integrated driving force identification method in capturing stable, strong-signal factors without distorting mechanisms widely acknowledged in the literature.For road width and road density, the *IDFV* ranking represents a synthesis, neither neglecting their spatial driving effects nor unduly amplifying their linear contributions, reflecting the comprehensiveness and objectivity of this assessment method. The rankings for river network density, river network connectivity, and blue-green space distribution show significant divergence among the different methods. However, the *IDFV* rankings for these factors consistently fall between their respective rankings based solely on q-values and |*Beta*| values. It precisely illustrates the advantage of the *IDFV* method: by fusing the two metrics, it achieves a more robust assessment of factor influence, producing ranking results that possess both geographical explanatory power and statistical reliability. In summary, regarding the identification of key driving factors, the three methods exhibit characteristics of partial convergence and partial complementarity. It demonstrates that the integrated driving force identification method constructed in this study is logically sound, yields robust and insightful results, and provides a valuable analytical framework for the scientific identification of priority targets for urban waterlogging management.

4.2 Advantages Analysis

4.2.1 Advantages Analysis of Optimal Parameters-based Geographical Detector

This study employs the Optimal Parameters-based Geographical Detector (OPGD) model to identify the key drivers of the spatial distribution of urban exceedance flooding. The core of this model lies in its parameter optimization mechanism, which determines the spatial discretization scheme that yields the strongest explanatory power for the dependent variable, thereby objectively revealing the spatial heterogeneity and influence of driving factors. For each continuous driving factor, its explanatory power (*q-value*) is calculated under different classification methods (e.g., equal interval, quantile interval, natural breaks, geometric interval, and standard deviation interval) and with different numbers of classes (ranging from 3 to 10). Taking DEM,PND, RND, and GSD as illustrative examples, the corresponding results are presented in Figure 11.

Traditional studies have predominantly relied upon the five-class Natural Breaks (Jenks) method as the default discretization scheme. By contrast, the present study derives the optimal solution for each indicator using the Optimal Parameters-based Geographical Detector (OPGD). Employing the average explanatory power across five classification methods under a five-class partition as a benchmark, we subsequently calculate the improvement rate in explanatory power of the optimal solution relative to this benchmark. The corresponding results are reported in Table 7.This analysis demonstrates that the principal advantage of the OPGD model does not lie in amplifying the effect of each variable, but rather in correcting the systematic underestimation induced by the interplay between classification method and discretization parameters. Taking elevation as an illustrative example: the q-statistic under the conventional five-class average scheme is merely 0.035, whereas the OPGD-optimized scheme elevates it to 0.192 (Δq = 0.158, representing a relative improvement of 456%). This indicates that the spatial stratification potential of elevation with respect to the response variable in the study area is severely suppressed under non-optimal discretization configurations. For factors exhibiting moderate explanatory power, the OPGD model provides a more stable upper-bound estimate of their explanatory capacity. For instance, the Δq improvements for river network connectivity, road density, and blue-green space distribution range between 0.004 and 0.006. This implies that, under a fixed classification scheme, the contributions of these factors would be underestimated by approximately 20% to 90% in relative terms.

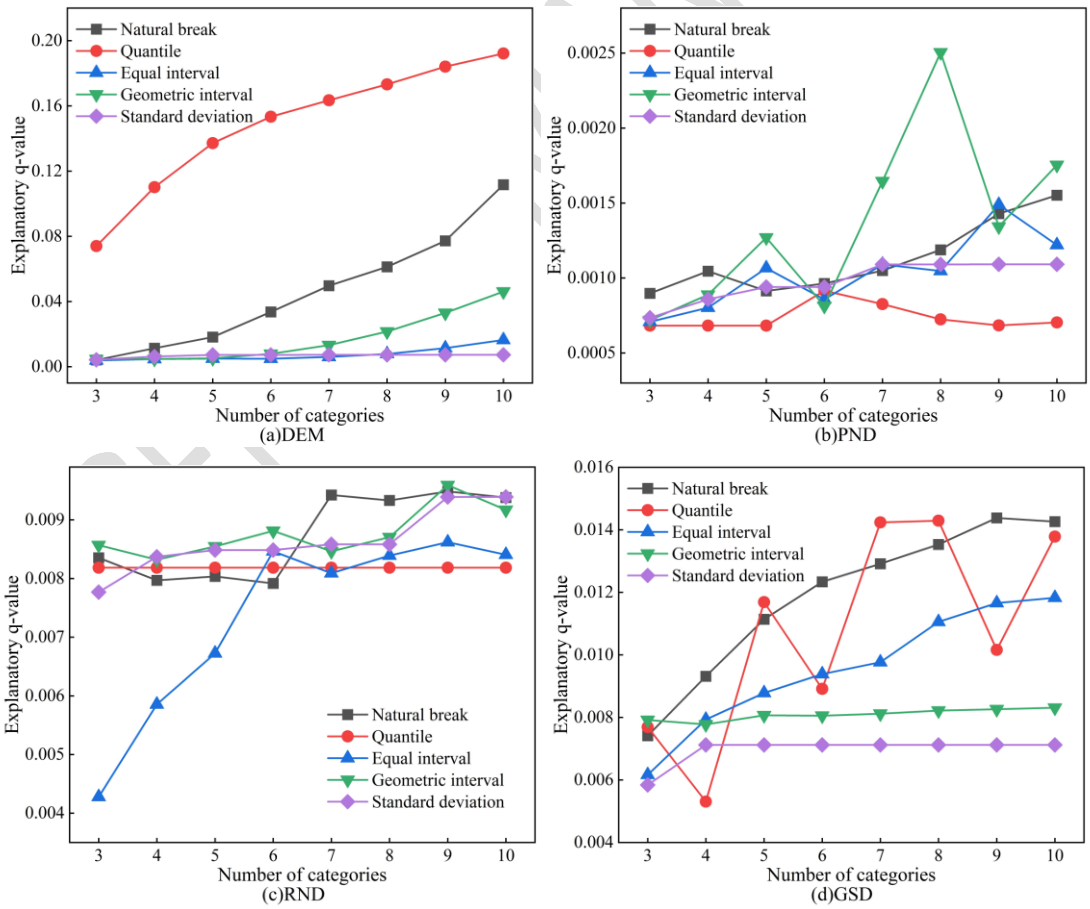


Fig.11 Optimal Discretization of Continuous Variables

693 **Table 7** Accuracy improvement rate of OPGD model driving factors

Driving Factor	***$GEO_{average\ q}$***	***$OPGD_q$***	***Accuracy (%)***
Elevation	0.034533222	0.19211248	456%
Slope	0.000278292	0.000425437	53%
Pipeline network density	0.000974599	0.002504265	157%
Road density	0.004189659	0.008003684	91%
Road width	0.018351234	0.02496086	36%
River network density	0.007995836	0.009592402	20%
River network connectivity	0.004742554	0.009176541	93%
Green space distribution	0.009358543	0.01437928	54%
Blue space distribution	0.015160273	0.021590991	42%

694 *4.2.2 Advantages Analysis of Integrated Driving Force Identification approach*

695 A comparative analysis of results derived from different identification methods is presented in
696 Figure 12, revealing pronounced discrepancies in the performance of driving factors across models.
697 Consider the case of blue space distribution. Within the Optimal Parameters-based Geographical
698 Detector (OPGD) framework, this factor exhibits a relatively high q-statistic, indicating substantial
699 explanatory power regarding the spatial heterogeneity of waterlogging. This arises because the
700 Geographical Detector fundamentally evaluates the degree of spatial overlap between factor

classifications and waterlogging gradations. Blue spaces tend to concentrate in low-lying zones or along river corridors, rendering their spatial configuration highly coincident with historically high-risk waterlogging areas, and this association is therefore readily captured by the OPGD model. In contrast, the standardized regression coefficient for blue space distribution within the Multiple Linear Regression (MLR) framework is comparatively modest, suggesting that, after controlling for other covariates, its statistical effect on waterlogging intensity is not statistically significant. This discrepancy stems from the MLR model's assumption of a global linear relationship between predictors and the dependent variable, whereas the mitigating effect of blue spaces on waterlogging is inherently nonlinear in nature. Consequently, the spatial stratification effect of blue spaces outweighs their global linear contribution. These findings collectively suggest that reliance on a single analytical model may yield results with certain inherent limitations.

The coupled identification results based on the Integrated Driving Force Value (IDFV) indicate that elevation, road width, and river network density rank among the top three factors, establishing themselves as the core drivers of exceedance flooding formation in the study area. This finding is corroborated by the results from individual methods: whether assessed by the q-statistic representing explanatory power for spatial heterogeneity or by the standardized regression coefficient measuring statistical effect, these three factors consistently exhibit prominence. This demonstrates that they not only dominate the spatial distribution of waterlogging risk but also exert a significant direct influence on its intensity. In contrast, slope demonstrates relatively low influence across all three evaluation frameworks. Analysis of topographic conditions reveals that the study area exhibits a stepped configuration—higher in the southeast and lower in the northwest—with most regions characterized by gentle gradients and minimal slope variation. This implies that the spatial stratification effect of slope is weak, and its linear regression coefficient fails to attain statistical significance, thereby preventing slope from emerging as a dominant controlling factor.

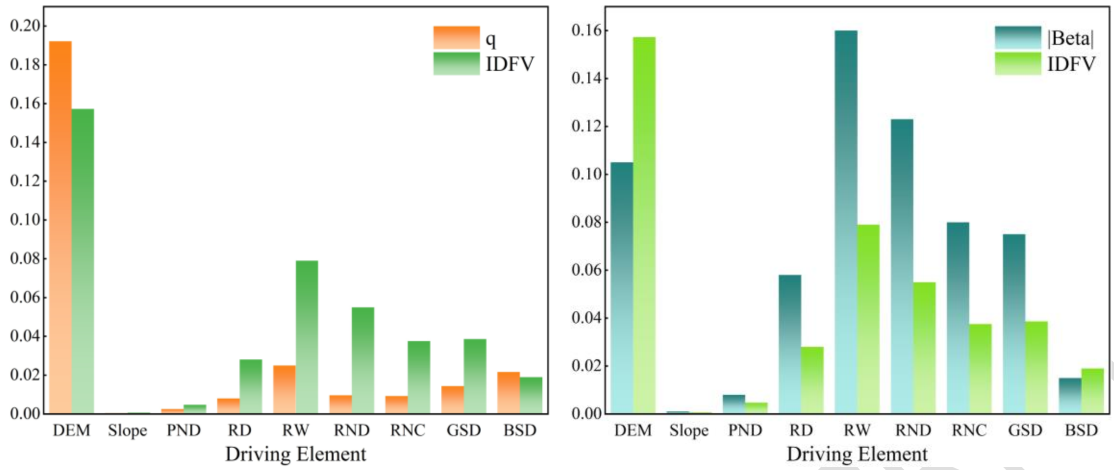


Fig.12 Comparison Chart of Results for Three Recognition Models

4.3 Limitation and Future Work

Due to the absence of field-monitored waterlogging data under observed rainfall events in the study area, validation of the two-dimensional model's accuracy was not feasible. Consequently, overflow volume was adopted as a proxy metric to represent exceedance flooding intensity. While this approach captures certain aspects of waterlogging scenarios, it is acknowledged that actual surface inundation depth would constitute a more physically representative metric. Future research endeavors will involve the installation of relevant monitoring equipment to acquire empirical observational data, thereby enabling two-dimensional surface waterlogging simulations and achieving enhanced computational precision. The Optimal Parameters-based Geographical Detector (OPGD) model reduces subjectivity by optimizing the number of classes and the classification method. However, the discretization of continuous variables still relies on preset parameters, and the results remain sensitive to the chosen discretization scheme. Future research could incorporate adaptive discretization algorithms, such as hierarchical optimization based on machine learning. Secondly, the model exhibits limited capacity for identifying complex non-linear relationships. Meanwhile, Multiple Linear Regression (MLR) assumes linear relationships between variables, which hampers its ability to precisely capture intricate non-linear dynamics or threshold effects between drivers and flooding intensity. Future work could integrate models like Random Forest or Gradient Boosting Trees to elucidate these complex, non-linear interaction mechanisms.Furthermore, due to challenges in data acquisition, this study analyzed only nine indicators. Expanding the dataset with survey data or the integration of multi-source data is a potential avenue for enhancement.

5 Conclusion

This study integrates the Optimal Parameters-based Geographical Detector (OPGD) model with the multiple linear regression (MLR) method, proposing a novel and integrated approach for identifying key factors of urban exceedance flooding. The main research findings are as follows: (1) An integrated spatio-statistical methodology was successfully developed to enhance the identification of key drivers behind urban exceedance flooding. (2) A systematic driver framework comprising two dimensions and nine indicators was established, focusing on exceedance runoff pathways and storage-drainage spatial layout. (3) The practical application of this methodology in the High-tech Industrial Development Zone demonstrated its improved accuracy and robustness in identifying key flooding drivers. (4) The results demonstrate that elevation and road width are the strong-signal factors for the urban exceedance flooding, which the *IDFV* are 0.157 and 0.079, respectively. (5) Expanding the datasets and building a more comprehensive indicator system for urban exceedance flooding will be our future work, providing a better assessment of driving forces. (6) The results of this study provide a new scientific basis for future urban exceedance flooding prevention and control, good for guiding the layout of Sponge City facilities and reducing urban waterlogging risk at the source.

6 Acknowledgement

This work was supported by the National Key Research and Development Program of the Ministry of Science and Technology of China (Project No. 2025YFC3811302), National Natural Science Foundation of China (42201078), Natural Science Foundation of Tianjin City (24JCYBJC00680). The authors acknowledge the assistance of anonymous reviewers.

7 References

Agbiji, N. M., Agunwamba, J. C., & Eshiet, K. I. I. I. (2025). Modelling the impact of climate change on runoff in a sub-regional basin. *Geosciences*, *15*(8), 289.

Bell, J. E., Brown, C. L., Conlon, K., Herring, S., Kunkel, K. E., Lawrimore, J., Luber, G., Schreck, C., Smith, A., & Uejio, C. (2018). Changes in extreme events and the potential impacts on human health. *Journal of the Air & Waste Management Association*, *68*(4), 265-287.

Bhattacharyya, S., Sreekesh, S., & King, A. (2022). Characteristics of extreme rainfall in different gridded datasets over india during 1983–2015. *Atmospheric Research*, *267*, 105930.

Chang, C., Chen, Y., & Huang, J. J. (2023). A comparison study on the role of urbanization in altering the short-duration and long-duration intense rainfall. *Science of the Total Environment*, *857*, 159290.

Chang, C., Li, Y., Chen, Y., Huang, J. J., & Zhang, Y. (2021). Advanced statistical analyses of urbanization impacts on heavy rainfall in the beijing metropolitan area. *Urban Climate*, *40*, 100987.

Chen, F., Wu, M., Dong, M., & Yu, B. (2022). Comparison of the impacts of topography and urbanization on an extreme rainfall event in the hangzhou bay region. *Journal of Geophysical Research-Atmospheres*, *127*(15).

Chen, Z. (2014). Research on the causes and prevention technologies of flash flood disasters in small watersheds. *South China University of Technology*.

Diakakis, M., Deligiannakis, G., Pallikarakis, A., & Skordoulis, M. (2016). Factors controlling the spatial distribution of flash flooding in the complex environment of a metropolitan urban area. The case of athens 2013 flash flood event. *International Journal of Disaster Risk Reduction*, *18*, 171-180.

Gong, Y., Li, X., Zhai, D., Yin, D., Song, R., Li, J., Fang, X., & Yuan, D. (2017). Influence of rainfall, model parameters and routing methods on stormwater modelling. *Water Resources Management*, *32*(2), 735-750.

Gurung, P. (2016). Integration of gauge and radar rainfall to enable best simulation of hydrological parameters. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 1-10.

Jin, H. (2005). Research on the investigation and prevention of flash flood disasters in shandong province based on GIS. *Hohai University*.

Lee, J., Kim, S., & Jun, H. (2018). A study of the influence of the spatial distribution of rain gauge networks on areal average rainfall calculation. *Water*, *10*(11), 1635.

Lee, M. J., Kang, J. E., & Jeon, S. (2012). Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using GIS.

Li, T., Yin, S., Xiao, Y., Wang, M., Chen, L., Zou, W., & Peleg, N. (2026). Quantifying rainfall variability and potential hazards of extreme events in beijing through stochastic simulations. *Journal of Hydrology-Regional Studies*, *63*, 103068.

Liao, D., Zhu, H., & Jiang, P. (2020). Study of urban heat island index methods for urban agglomerations (hilly terrain) in chongqing. *Theoretical and Applied Climatology*, *143*(1-2), 279-289.

Liu, B., Chen, J., Chen, X., Lian, Y., & Wu, L. (2013). Uncertainty in determining extreme precipitation thresholds. *Journal of Hydrology*, *503*, 233-245.

Liu, H., Zou, L., Xia, J., Chen, T., & Wang, F. (2022). Impact assessment of climate change and urbanization on the nonstationarity of extreme precipitation: a case study in an urban agglomeration in the middle reaches of the yangtze river. *Sustainable Cities and Society*, *85*, 104038.

Liu, S., & Sun, Q. (2025). Modeling and performance analysis of emergency intelligence service process based on stochastic petri nets: a case study of the "7·20" zhengzhou rainstorm in china. *Reliability Engineering & System Safety*, 112119.

Liu, Y., Zhang, Y., Zhang, J., Li, Y., Zhai, D., Luo, F., Zhang, Y., & Zhou, G. (2022). Characteristics of summer precipitation in chongqing based on hourly rain gauges data. *Advances in Meteorology*, *2022*, 1-13.

Luppichini, M., Bini, M., Giannecchini, R., & Zanchetta, G. (2023). High-resolution spatial analysis of temperature influence on the rainfall regime and extreme precipitation events in north-central italy. *Science of the Total Environment*, *880*, 163368.

Mahmood, M. I., Elagib, N. A., Horn, F., & Saad, S. A. G. (2017). Lessons learned from khartoum flash flood impacts: an integrated assessment. *Science of the Total Environment*, *601-602*, 1031-1045.

Pirani, F. J., & Modarres, R. (2020). Geostatistical and deterministic methods for rainfall interpolation in the zayandeh rud basin, iran. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *65*(16), 2678-2692.

Pradhan, B., H. J., & Buchroithner, M. (2010). Weights-of-evidence model applied to landslide susceptibility mapping in a tropical hilly area. *Geomatics Natural Hazards & Risk*, *1*(3), 199-223.

Karthik., Surendran., G, S. K., & Srinivasulu, S. (2025). Flood prediction in chennai based on extended elman spiking neural network using a robust chaotic artificial hummingbird optimizer. Global NEST Journal, 1-11.

Sivasubramanian, S., Kumar, G. V. S., Thanarajan, T., & Rajendran, S. (2025). Al-biruni earth radius optimization for enhanced environmental data analysis in remote sensing imagery. *Agrociencia*, 1-18.

Song, X., Zhang, J., Zou, X., Zhang, C., Aghakouchak, A., & Kong, F. (2019). Changes in precipitation extremes in the beijing metropolitan area during 1960–2012. *Atmospheric Research*, *222*, 134-153.

Sundarapandi, A. M. S., Navaneethakrishnan, S. R., A, H., & Rajendran, S. (2024). A light weighted dense and tree structured simple recurrent unit (LDTSRU) for flood prediction using meteorological variables. *Global NEST Journal*.

Suthinkumar, P. S., Varikoden, H., & Babu, C. A. (2023). Changes in extreme rainfall events in the recent decades and their linkage with atmospheric moisture transport. *Global and Planetary Change*, *221*, 104047.

Selvanarayanan, R., Thanarajan, T., & Rajendran, S. (2024). Integrated early flood prediction using sentinel-2 imagery with VANET-MARL-based deep neural RNN. *Global NEST Journal*, 1-9.

Tehrany, M. S., Lee, M. J., Pradhan, B., Jebur, M. N., & Lee, S. (2014). Flood susceptibility mapping using integrated bivariate and multivariate statistical models. *Environmental Earth Sciences*, *72*(10), 4001-4015.

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2013). Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology*, *504*, 69-79.

Yang, P., Ren, G., & Yan, P. (2017). Evidence for a strong association of short-duration intense rainfall with urbanization in the beijing urban area. *Journal of Climate*, *30*(15), 5851-5870.

Yeh, H. C., Chen, Y. C., & Wei, C. (2011). A new approach to selecting a regionalized design hyetograph by principal component analysis and analytic hierarchy process. *Paddy and Water Environment*, *11*(1-4), 73-85.

Yongze, S., Jinfeng, W., Ge, Y., & Chengdong, X. (2020). An optimal parameters-based geo graphical detector model enhances geographic characteristics of ex planatory variables for spatial heterogeneity analysis: cases with dif ferent types of spatial data. *GIScience & Remote Sensing*.

Zhang, W., Villarini, G., Vecchi, G. A., & Smith, J. A. (2018). Urbanization exacerbated the rainfall and flooding caused by hurricane harvey in houston. *Nature*, *563*(7731), 384-388.