Frontiers in Environmental Research
ISSN: 2960-026X
DOI: https://doi.org/10.61784/fer3029

CAN OPEN WEB DATA ASSESS URBAN FLOOD RISK? EVIDENCE FROM ZHENGZHOU

Da Mao[1,2*], Yan Xie[1], ZhiYu Yuan[1]
[1]*School of Horticulture and Landscape Architecture, Henan Institute of Science and Technology, Xinxiang 453003, Henan, China.*
[2]*Xinxiang Urban-Rural and Landscape Digital Technology Engineering Research Center, Xinxiang 453003, Henan, China.*
Corresponding Author: Da Mao, Email: maoda@foxmail.com

Abstract: Under extreme rainfall scenarios, urban stormwater pipe networks are prone to saturation and failure, rendering flood risk assessment under such conditions of critical importance. This study established a flood risk assessment framework using GIS technology and open web data to evaluate 270 catchment subunits in Zhengzhou's main urban area. Results reveal an overall stepped spatial pattern of flood risks, characterized by lower risks in the southwest and higher risks in the northeast, with contiguous high-risk clusters identified in the city center. Spatial autocorrelation analysis confirms strong spatial dependence of flood risks: High-High clusters concentrate in the urban core and northeastern districts, while Low-Low clusters are distributed across the higher-elevation southwestern region. The study further conducts bivariate correlation analysis between zonal flood risks and influencing factors, culminating in the proposal of targeted mitigation strategies.
Keywords: Waterlogging; Flood; Risk; Open data; Evaluate

1 INTRODUCTION

Urban waterlogging refers to the phenomenon of flood caused by excessive rainfall intensity or prolonged duration, which exceeds the drainage capacity of pipe networks. In recent years, extreme heavy rainfall events have occurred with increasing frequency in various regions[1-2]. Once urban flooding disasters triggered by torrential rains occur, disaster-bearing bodies such as pedestrians, roads, and buildings exposed to floods are highly vulnerable to threats, leading to casualties, waterlogging-induced traffic disruptions, water and power outages, economic losses, and even potential collapse of urban emergency command systems and basic support systems such as communication, electricity, and medical services . Urban flood not only affects the work and life of urban residents but also easily causes severe losses of life and property.How to assess the risk of urban flood disasters and ensure the safety of urban drainage and flood prevention has become an urgent issue for urban development[3].

Risk assessment is an effective means to alleviate urban flood disasters and has attracted the attention of many studies[4-5]. Previous studies generally suggest that the formation of flooding disasters is primarily caused by two factors: natural and anthropogenic. Natural factors mainly include urban topography and climatic characteristics, while anthropogenic factors include: the increase in impervious surfaces during urbanization, which reduces natural buffer zones, slows down rainwater infiltration, and leads to surface water accumulation during storms[6].

The high spatial heterogeneity of the built environment and the complexity of disaster-causing processes pose considerable challenges to urban pluvial flooding research, manifesting as insufficient model representativeness, low computational efficiency, and scarcity of basic and validation data[7]. Under extreme conditions, intensified surface water accumulation renders underground stormwater pipe networks nearly ineffective. Meanwhile, for large cities, the traditional method of establishing catchment areas involves enormous workloads, and data on underground pipelines are difficult to obtain. Many scholars from various countries have conducted research and attempts on the assessment of urban flood risks[8-11]. However, due to different ways of obtaining data, many indicators lack universality[12]. In this context, more and more research has begun to use open data based on the Internet[13-15]. This study proposes using open web data for flood risk assessment under extreme conditions to derive evaluation conclusions more efficiently and concisely.

2 DATA AND METHODS

2.1 Study Area

Zhengzhou, the capital of Henan Province, serves as a key central city in China's central region. On July 20, 2021, under the impact of extreme weather, Zhengzhou was struck by severe rainstorms and flooding. The hourly rainfall reached a staggering 201.9 mm that day, with a daily cumulative rainfall of 624.1 mm. Urban waterlogging has imposed significant impacts on Zhengzhou's strategic status and its economic and social fabric. The selection of Zhengzhou's main urban area as the study area is highly representative.

This study selected a continuous area within the central urban district specified in the public *Zhengzhou Sponge City Special Plan (2017-2030)* for assessment. Geographic information data were established through ArcGIS digitization and georeferencing. The assessment area, projected using WGS1984-UTM-Zone50N, encompasses 838.56 km² divided into 270 subunits (Figure 1). The subunits range from 0.77 km² to 13.91 km² in area. These subunits are assigned distinct annual runoff control rate targets in the Zhengzhou Sponge City Special Plan (2017-2030), which well align with practical requirements for urban water management. The numbers in the right figure are the partition serial numbers given by the author.

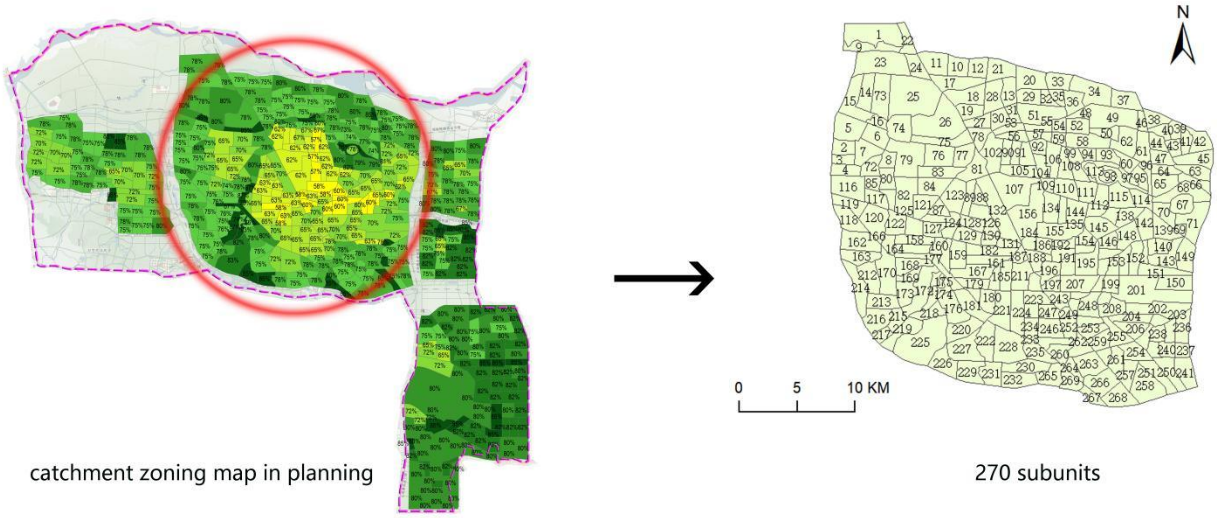


Figure 1 Study Area and Subunits Diagram

2.2 Open Web Data

2.2.1 DEM data

DEM (Digital Elevation Model) is a digital simulation of terrain through limited topographic elevation data, which realizes the digital expression of terrain surface morphology. It is currently widely applied in geographic assessment and analysis. The DEM data used in this study was downloaded from the Geographic Information Data Cloud, with a precision of 30 meters. By overlaying the evaluation units with DEM rasters in ArcGIS, three indices—minimum elevation, average elevation, and average slope—of each evaluation unit were extracted using tools such as Spatial Analyst (Figure 2).

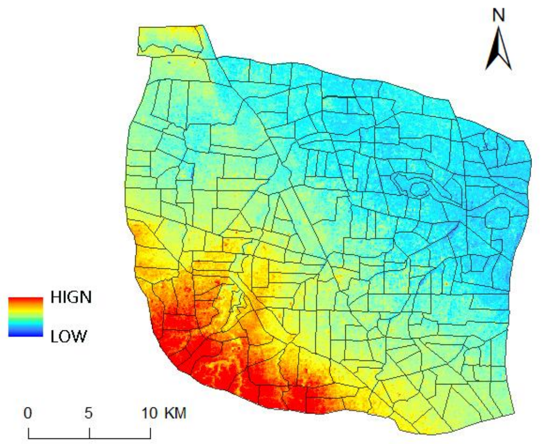


Figure 2 DEM Data and Overlay Map of Subunits

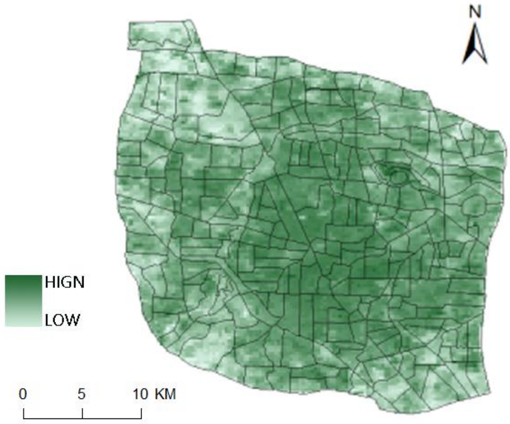


Figure 3 NDVI Data and Overlay Map of Subunits

2.2.2 NDVI data

NDVI (Normalized Difference Vegetation Index) is a digital characterization of surface vegetation coverage achieved through normalized operations of the difference and sum of reflectance values from remote sensing bands. Its core value lies in transforming vegetation physiological status into quantifiable numerical signals, which has been widely applied in ecological monitoring, agricultural yield estimation, and climate change research. The NDVI data used in this study was derived from the Geospatial Data Cloud in August 2023, with a spatial resolution of 250 meters. Using the ArcGIS platform, the vector boundaries of the study area were spatially overlaid with NDVI raster data (Figure 3), and the average NDVI values within each unit were extracted.

2.2.3 Population density data

The population density data in this study was primarily obtained from the LandScan global population density spatial distribution dataset, with a spatial resolution of 1 kilometer and a temporal coverage of 2020. Although population distribution is dynamic, this dataset effectively characterizes the spatial distribution and density of human populations. The study achieved a more refined simulation of population spatial distribution by allocating the total population of statistical units to raster grids and performing overlay extraction (Figure 4).

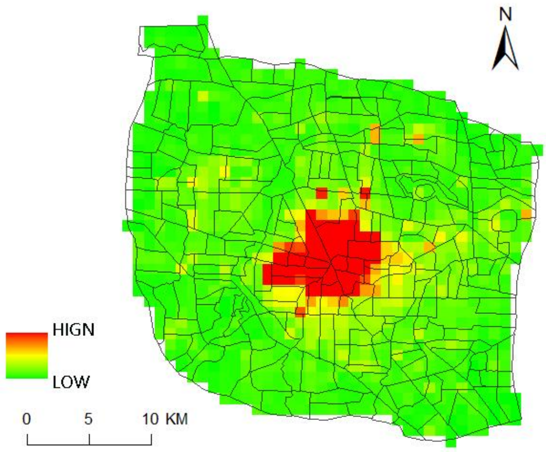


Figure 4 Population Density Data and Overlay Map of Subunits

2.2.4 Building density data

Building density is calculated by dividing the total building footprint area by the unit area. As a key indicator for measuring urban spatial development intensity, building scale, and land-use efficiency, it also serves as a critical metric for assessing urban impervious surfaces and flood risk. The building area data used in this study was derived from the vector building datasets of 77 major Chinese cities published online in 2022. After projecting these datasets, we overlaid them with the evaluation units (Figure 5) to calculate the building density of each unit.

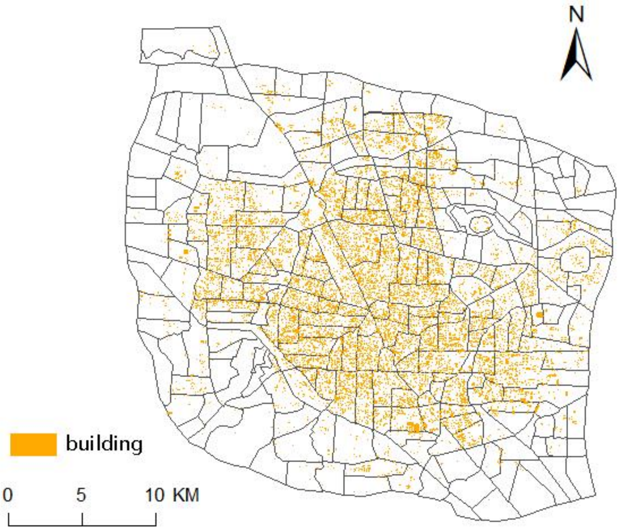


Figure 5 Building Density Data and Overlay Map of Subunits

2.3 Methods

2.3.1 Evaluation System

The unit-based flood risk assessment for Zhengzhou's urban core under extreme conditions is a concept with relatively clear objectives but vague details, necessitating the introduction of multiple evaluation indicators. The indicators were selected based on their maximum correlation with surface flood risk after urban stormwater pipe networks become saturated under extreme conditions. Referencing relevant research, an evaluation system was established (Table 1).

Table 1 Evaluation System of unit-based Flood Risk

Code	Indicator Name	Calculation Method	Effect
A1	Minimum Elevation	Extract the minimum value from the intersection of the unit area and DEM raster.	Negative
A2	Average Elevation	Extract the average value from the intersection of the unit area and DEM raster.	Negative
A3	Average Slope	Extract the average value from the intersection of the unit area and slope raster.	Negative
A4	Building Density	Calculate by dividing the building area by the area of the unit.	Positive
A5	Average NDVI	Extract the average value from the intersection of the unit area and NDVI raster.	Negative
A6	Population Density	Extract the average value from the intersection of the unit area and population density raster.	Positive

The scientific hypothesis of this assessment is that under extreme conditions, urban stormwater pipe networks are saturated and unable to function effectively in the short term, leading to urban waterlogging risks. The minimum elevation and average elevation within units are negative indicators, meaning that lower elevation values correspond to higher flood risks. The average slope is also a negative indicator because gentle slopes under extreme conditions slow surface runoff velocity, prolonging rainwater retention time in catchment units. If drainage facilities (e.g., pipe networks, drainage ditches) have insufficient design capacity for flood control, rainwater easily accumulates in low-lying areas, forming standing water.

Building density is a positive indicator: as building roofs are impervious surfaces, higher building density amplifies risks. The average NDVI is a negative indicator—lower average NDVI values signify higher waterlogging risks. Although green spaces lose much of their infiltration capacity under extreme conditions, vegetation itself still retains and dissipates rainwater. Finally, population density is a positive indicator: as vulnerable recipients of waterlogging disasters, higher population density correlates directly with greater risks.

2.3.2 Weight Determination

The flood risk assessment for Zhengzhou's urban core area under extreme conditions is a concept with relatively clear objectives but vague extensions, thus multivariate statistical analysis methods can be adopted for evaluation. The established evaluation system exhibits strong complexity, fuzziness, and dynamics in data information. To avoid weight bias caused by subjective analysis methods, the entropy weight model was selected for evaluation to objectively reflect data characteristics and patterns. Specifically, in this evaluation, smaller differences among units within the same indicator imply lower indicator weights. The specific calculation process is as follows:

First, the range method was adopted to standardize the statistical values of each indicator. Considering the requirement for logarithmic calculations in subsequent formulas, which necessitates avoiding zero values, the linear interpolation method was used to standardize the data within the range of 0.1 to 1. The formula is as follows:

$$X'_{ij} = \frac{X_{ij} - \min(X_j)}{\max(X_j) - \min(X_j)} \times 0.9 + 0.1 \quad \text{(Positive)} \tag{1}$$

$$X'_{ij} = \frac{\max(X_j) - X_{ij}}{\max(X_j) - \min(X_j)} \times 0.9 + 0.1 \quad \text{(Negative)} \tag{2}$$

Then, the standardized indicator data are normalized to obtain a matrix:

$$Z_{ij} = (X'_{ij} / \sum_{i=1}^{m} X'_{ij})_{m\times n} \tag{3}$$

Calculate the entropy value of the j-th indicator:

$$E_j = -\frac{1}{\ln m} \sum_{i=1}^{m} Z_{ij} \ln Z_{ij} ; \tag{4}$$

Thus, the weight of the j-th indicator can be derived as:

$$W_j = (1 - E_j) / \sum_{j=1}^{n} (1 - E_j) \tag{5}$$

2.3.3 Model Calculation

Multiply the standardized indicator values of each subunit by the indicator weights determined via the entropy weight method to derive the flood risk assessment model for each catchment subunit in Zhengzhou's urban core under extreme conditions. This model is expressed as the sum of the products of individual risk indicators and their respective weights within each subunit. The formula is as follows:

$$R_i = \sum_{j=1}^{n} W_j X_{ij} \tag{6}$$

3 RESULTS AND ANALYSIS

3.1 Results

Results from the entropy weight method show that the differences in indicator weights are not significant (Table 2), indicating minimal variation in the within-indicator data across the 270 subunits. Specifically, Minimum Elevation and Average Elevation exhibit the highest weights, highlighting the most pronounced topographic differences.

Table 2 Indicator Weights Derived from Entropy Weight Method

Minimum Elevation	Average Elevation	Average Slope	Building Density	Average NDVI	Population Density
0.1669	0.1669	0.1668	0.1664	0.1667	0.1663

By substituting the indicator weights and standardized indicator values into the calculation model, the risk values for individual indicators and total flood risk values of the 270 subunits were derived. The average risk value across subunits is 0.4994, with the maximum and minimum values being 0.7066 and 0.1909, respectively—showcasing remarkable differences.

Jenks natural breaks classification divides data into different categories by minimizing the sum of squared within-class variances, enabling the identification of classification intervals to most appropriately group similar values. By inputting the total risk values of each subunit into ArcGIS and applying the Jenks natural breaks classification method for legend categorization, a flood risk assessment distribution map for each subunit was generated (Figure 6). The legend values represent flood risk levels, where higher numbers indicate greater risks.

The flood risks in Zhengzhou's urban core under extreme conditions exhibit an overall stepped spatial distribution, lower in the southwest and higher in the northeast. However, contiguous high-risk areas appear in the urban central region, mainly concentrated west of Zhongzhou Avenue and north of Zhengbian Road-East Street, as well as in the Longhu and Longzi Lake areas. Contiguous high-risk zones between Huayuan Road and Lianhuo Expressway in the northern region are also notably present.

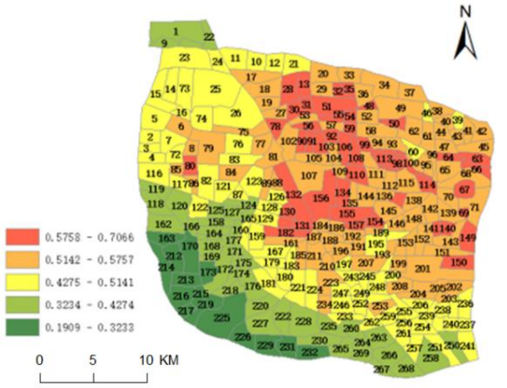
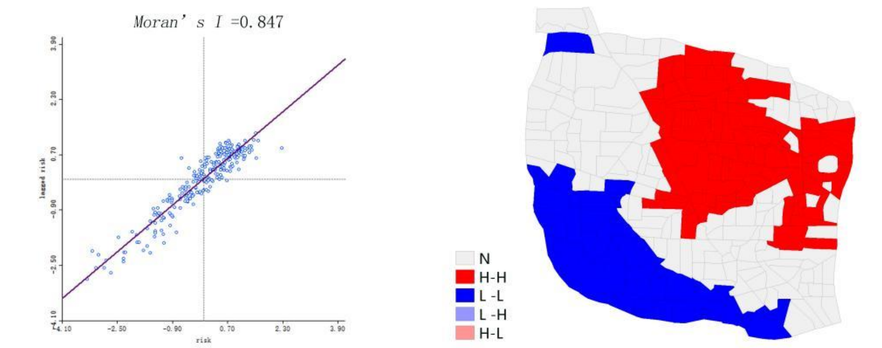


Figure 6 Flood Risk Assessment Distribution Map of Subunits

3.2 Spatial Autocorrelation Analysis

It is necessary to employ spatial autocorrelation to analyze the spatial characteristics of flood risks in Zhengzhou's urban core, using *Moran's I* index to determine whether spatial clustering exists within the region. The *Moran's I* statistic was computed using GeoDa 1.20 to quantify spatial clustering patterns of flood risk.



a. Flood Risk Univariate *Moran' s I* Scatter Plot b. Flood Risk Univariate LISA Cluster Map

Figure 7 Flood Risk *Moran' I* Scatter Plot and LISA Cluster Map

Moran's I value reflects the spatial clustering of an indicator within a region: a value greater than 0 indicates positive correlation, meaning high-indicator subunits are surrounded by high-indicator neighbors, and low-indicator subunits are surrounded by low-indicator neighbors. The univariate *Moran's I* value for flood risk across subunits was first calculated as 0.847, indicating significant spatial autocorrelation (Figure 7, a). Meanwhile, a local LISA cluster map generated by Geo-Da software (Figure 7, b) shows that High-High clusters are primarily concentrated in the city center and northeastern areas, while Low-Low clusters are mainly distributed in the higher-terrain southwestern regions.

The GeoDa software was also used to calculate local *Moran's I* values between bivariate variables and generate LISA cluster maps. This study sequentially calculated the raw values of flood risk against the minimum elevation, average elevation, average slope, building density, NDVI values, and population density of each subunit, generating corresponding bivariate scatter plots and LISA cluster maps(Figure 8) in turn.

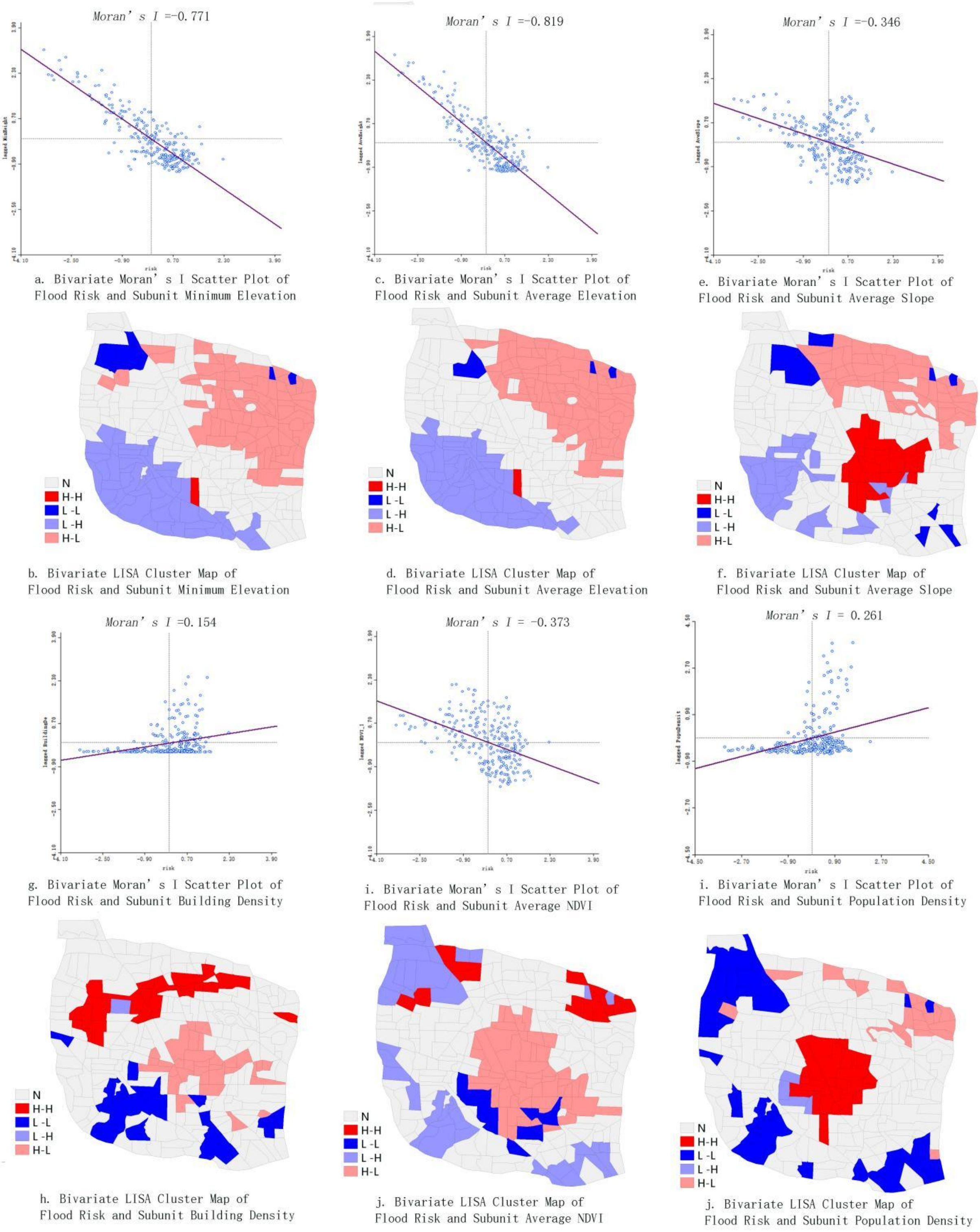


Figure 8 Bivariate *Moran's I* Scatter Plots and LISA Cluster Maps of Flood Risk and Various Parameters

Bivariate spatial autocorrelation analysis shows that subunit average elevation has the highest correlation with flood risk, with a *Moran's I* value of -0.819, indicating a negative correlation. In terms of cluster distribution, High-Low and Low-High clusters are contiguous in the northeastern and southwestern urban areas. Both average slope and population density also exhibit negative correlations with flood risk, but unlike elevation indicators, High-High clusters are remarkably concentrated in the city center. The building density indicator has the smallest correlation with flood risk but shows a positive correlation (*Moran's I* = 0.154), still demonstrating a significant relationship. In spatial clustering, High-High clusters are aggregated in the northwestern region, High-Low clusters in the city center, and Low-Low clusters in the southern region. NDVI also has a significant negative correlation with flood risk (*Moran's I* = -0.373), with High-Low clusters mainly in the city center and Low-High clusters on the urban periphery, except for High-High clusters appearing only in the northern region.

4 CONCLUSION

This study presents a comprehensive flood risk assessment for Zhengzhou's urban core under extreme conditions, incorporating open web data with entropy weight modeling, spatial autocorrelation analysis, and GIS-based spatial distribution mapping. Key findings include:
(1) Spatial Distribution Pattern: Flood risks exhibit a stepped gradient from southwest to northeast, with high-risk clusters concentrated in central urban areas (west of Zhongzhou Avenue, north of Zhengbian Road-East Street) and northeastern districts (Longhu, Longzi Lake areas). Contiguous high-risk zones between Huayuan Road and Lianhuo Expressway further validate topographic influence.
(2) Indicator Significance: Topographic indicators (minimum/average elevation) dominate entropy weights, highlighting terrain as the primary risk determinant. Negative correlations between elevation/slope and flood risk confirm that lower-lying areas face higher vulnerability.
(3) Spatial Autocorrelation: Univariate *Moran's I* (0.847) reveals strong global spatial clustering, while bivariate analysis identifies elevation as the most influential factor *(Moran's I* = -0.819). Building density shows weak positive correlation (*Moran's I* = 0.154), indicating urban construction's indirect impact.
(4) Policy Implications: High-risk clusters in central business districts and northeastern developments necessitate targeted drainage upgrades, while NDVI-negative correlation underscores green infrastructure's flood mitigation potential.

5 DISSCUSION

The study bridges methodological gaps by integrating entropy weight objectivity with spatial autocorrelation's locational insight. The -0.819 *Moran's* I between elevation and flood risk aligns with hydrological theory, but the unexpected positive correlation in building density (0.154) warrants further exploration — potentially reflecting impervious surface expansion overriding topographic effects. High-risk zones coincide with both low-lying areas and urban heat islands, possibly exacerbated by climate change-induced extreme precipitation.
Limitations include:
(1) Data Timeliness: NDVI and population density data represent snapshots, failing to capture recent urban expansion;
(2) Climate Scenario Simplification: Extreme conditions assume historical rainfall peaks, ignoring projected climate variability;
(3) Microtopography Omission: Subunit-scale elevation data may obscure localized depressions critical for flood accumulation.
Future research should incorporate real-time rainfall-runoff modeling, fine-scale LiDAR topography, and climate change projections to refine risk zonation. The framework also offers transferability to other rapidly urbanizing regions with complex topographies.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This research was funded by the following projects: Key Science and Technology Research and Development Program of Henan Province, China (222102320233), Key Science and Technology Research and Development Program of Henan Province, China (232102320180), Cultivation Plan of Young Backbone Teachers in Colleges and Universities of Henan Province, China (2021GGJS118).

REFERENCES

[1] Koc K, Isik Z. Assessment of Urban Flood Risk Factors Using Monte Carlo Analytical Hierarchy Process. Natural Hazards Review, 2021, 22(4): 14.
[2] Idowu D , Zhou WY. Global Megacities and Frequent Floods: Correlation between Urban Expansion Patterns and Urban Flood Hazards. Sustainability, 2023, 15(3). DOI: 10.3390/su15032514.
[3] Liu J. Research and Countermeasure on Key Issues in the Construction of Shenzhen Waterlogging Disaster Assessment System Oriented to Risk Management. Urban Geotechnical Investigation & Surveying, 2021(01): 22-26.
[4] Bin LL, Xu K, Pan H, et al. Urban flood risk assessment characterizing the relationship among hazard, exposure, and vulnerability. Environmental Science and Pollution Research, 2023, 30(36): 86463-86477.
[5] O'Donnell EC, Thorne CR. Drivers of future urban flood risk. Philosophical Transactions of The Royal Society a-Mathematical Physical and Engineering Sciences, 2020, 378(2168). DOI:10.1098/RSTA.2019.0216
[6] Jia B, Li Y. Study on Roughness Height and Resistance Characteristies of SubmergedVegetation Buffer Zone in Open Channel Flow Condition. Water Resources and Power, 2021, 39(6): 60-63,115.
[7] Huang H, Wang X, Liu L. A review on urban pluvial floods: Characteristics.mechanisms, data, and research methods. Progress in Geography, 2020, 40(6): 1048-1059.

[8] Caradot N, Granger D, Chapgier J, et al. Urban flood risk assessment using sewer flooding databases. Urban flood risk assessment using sewer flooding databases, 2011, 64(4): 832-840.
[9] Bakhtiari V, Piadeh F, Behzadian K, et al. A critical review for the application of cutting-edge digital visualisation technologies for effective urban flood risk management. Sustainable Cities and Society, 2023, 99: 14.
[10] Seemuangngam A, Lin HL. The impact of urbanization on urban flood risk of Nakhon Ratchasima, Thailand. Applied Geography, 2024, 162. DOI: 10.1016/j.apgeog.2023.103152.
[11] Peixoto JPJ, Costa DG, Portugal P. Flood-Resilient Smart Cities: A Data-Driven Risk Assessment Approach Based on Geographical Risks and Emergency Response Infrastructure. Smart Cities, 2024, 7(1): 662-679.
[12] Zhu SA, Jiang YX, Zhang J, et al. Evaluating the effect of urban flooding on spatial accessibility to emergency shelters based on social sensing data. Transactions in GIS, 2024, 28(1): 23-39.
[13] Niemi TJ, Warsta L, Taka M, et al. Applicability of open rainfall data to event-scale urban rainfall-runoff modelling. Journal of Hydrology, 2017, 547: 143-155.
[14] Lin T, Liu XF, Song JC, et al. Evaluating the effect of urban flooding on spatial accessibility to emergency shelters based on social sensing data. Habitat International, 2018, 71: 88-96.
[15] Ren C , Cai M, Li XW, et al. Developing a rapid method for 3-dimensional urban morphology extraction using open-source data. Sustainable Cities and Society, 2020, 53. DOI: 10.1016/j.scs.2019.101962.