Safety Emergency Science

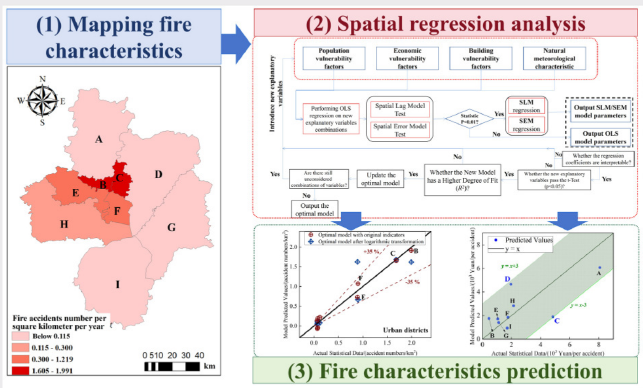
Research Article

City-scale fire risk modeling based on spatial regression methods

Hanchao Ma[1,✉], Xiaoxiao Sun[1], Jinlong Zhao[2]

Cite this article: Ma HC, Sun XX, Zhao JL. *Safety Emergency Science*, 2025, **1**(1): 9590002. https://doi.org/10.26599/SES.2025.9590002

ABSTRACT: Assessing the frequency of fires and the resulting economic losses is crucial for allocating emergency resources and firefighting personnel across different areas. In this study, spatial regression methods were employed to analyze the correlation between fire risk and various influencing factors (economic vulnerability, population vulnerability, building vulnerability, etc.), and an optimal model determination procedure was proposed and validated. The results indicate that the spatial lag model with population density, per capita income, and the degree of personnel concentration around high-risk points of interest (POIs) as indicators was the best model for predicting fire frequency. Moreover, logarithmic transformation of the indicators effectively improved the prediction accuracy in low-population density areas. An ordinary least squares model with the illiteracy rate and average surface water resources as indicators was the best model for predicting direct economic losses from fires, and the distribution of high-risk POIs can qualitatively explain the differences between the predicted results and actual data. The present work not only enriches the research on city-scale fire risk assessment but also reveals that the optimal regression model determination process can provide technical support for the application of spatial regression methods in fire risk assessment.



KEYWORDS: fire frequency; direct economic loss; fire risk modeling; spatial regression method

1 Introduction

With the continuous and rapid development of China's economy, the corresponding fire risk has increased, which seriously threatens the lives of people. In 2020, China recorded 252,000 fire incidents, which resulted in 1,183 fatalities and caused a direct property loss of 4.01 billion yuan (Liu et al., 2022). Consequently, it is essential to conduct studies that analyze regional factors that are highly correlated with fire accident characteristics (frequency, severity) to provide technical support for city-level fire safety management.

In recent years, many scholars have investigated fire risk analysis. Wang (2021) used the Apriori algorithm and GIS technology to analyze the relationships among fire causes, types, and places and reported that urban fires mostly occur in buildings. Moreover, they noted that there are certain differences in the main causes of fires in different types and places. Song (2017) conducted a study on predicting fire accident locations in Hefei city via machine learning and spatial econometric modeling. They discovered that spatial econometric modeling was more effective than machine learning in accurately predicting fire locations. However, the model's performance was better in urban areas, with a noticeable bias in predicting fire locations in counties. On the basis of the Code for Planning of Urban Fire Control (Ministry of Housing and Urban-Rural Development of the People's Republic of China, Ministry of Public Security of the People's Republic of China, 2015), Tan et al. (2023) categorized the collected POI data into six types, including crowded places and flammable and explosive areas, and analyzed the POI data of Panjin city via kernel density analysis and the SAVEE method, which verified the effectiveness of the fire risk assessment on the basis of POI data. Xu (2012) used a vector error correction model to analyze the spatiotemporal variation patterns of fires at the city level in China. They reported that economic development (such as GDP per capita) and climate change (such as environmental humidity) are the macro factors driving fire variations. Bispo et al. (2023) employed various spatial regression models and used a set of socioeconomic explanatory variables to spatially model the distribution of fire incident frequencies in Portugal. The spatial Durbin error model, which considers population density,

[1] School of Safety Science, Tsinghua University, Beijing 100084, China. [2] School of Emergency Management & Safety Engineering, China University of Mining & Technology, Beijing 100083, China.

✉ Corresponding author. E-mail: mhc22@mails.tsinghua.edu.cn

Received: August 19, 2024; Revised: September 4, 2024; Accepted: September 5, 2024

degraded building density, and buying power, was the most accurate; however, the model still had relatively large deviations in areas with low incident frequencies. Zhou (2012) conducted a regression analysis study on the relationships between meteorological factors and the monthly average number of fire accidents by applying a logarithmic transformation to indicator factors. They reported that the average temperature, average relative humidity, average rainfall, and number of fires in the previous month significantly impacted the number of fires in the current month. Liu et al. (2022) selected 10 socioeconomic and governmental data governance indicators to explain urban fire risk through correlation analysis on the basis of multisource statistical data from 105 Chinese cities between 2016 and 2018. They then constructed multiple regression models for each city cluster via ordinary least squares (OLS) regression to test the effect of governmental data governance. The results indicate that the constructed regression models have a good explanatory effect on the fire risk of different cities, with all the models having a coefficient of determination (R^2) exceeding 0.65. Hu et al. (2019) conducted a study on fire frequency and total economic loss in 283 cities in China via fire-related data from 2013 to 2016. They identified seven socioeconomic indicators closely related to urban fire risk and performed regression modeling of fire risk via normalized data and multiple regression models. The aforementioned literature has laid a solid foundation for conducting fire risk assessments. However, most studies have focused on the differences in fire risk between cities, with few studies investigating the dominant parameters of fire risk among different districts and counties within a city. Additionally, the average property loss, as an important indicator for evaluating the severity of fire accidents, has not been widely studied and requires further analysis of the related factors.

To fill the research gap mentioned above, this work selects a typical area in China as the study area, extracts six indicator parameters, and employs spatial regression models to analyze the correlation between these indicators and accident frequency/direct economic loss. Additionally, prediction models for accident frequency and direct economic loss were constructed and analyzed.

2 Methods and data description

2.1 Spatial regression model

The application of spatial regression models in fire risk assessment has gained increasing attention in recent years (Bispo et al., 2023; Song, 2017). The classical ordinary least squares (OLS) linear regression model is typically represented by Eq. (1):

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{y}$ is the spatial sequence (y_1, ..., y_n) with a sample size of n; $\boldsymbol{X}$ is an $n \times k$ matrix consisting of k columns of explanatory variables; $\boldsymbol{\beta}$ represents the corresponding regression coefficients; and ϵ denotes the disturbance term.

However, for spatial data, there may be internal correlations between observations that are geographically close to each other. Consequently, using models such as OLS, which assumes independence among observations, may lead to significant errors (Bivand and Piras, 2015). To measure the spatial distance relationships between regions, a spatial weight matrix was introduced:

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix} \tag{2}$$

where w_{ij} represents the spatial weight value between region i and region j (i = 1, ..., n, j = 1, ..., n), and the inverse of the distance between regions is used as the spatial weight (Getis and Aldstadt, 2004), that is, $w_{ij} = 1/d_{ij}$, where d_{ij} is the straight-line distance between the geographic centroids of region i and region j.

On the basis of the spatial weight matrix, various methods can be employed to incorporate spatial correlation into the OLS model, thereby constructing a spatial econometric model. Common models include the spatial lag model (Eq. (3)) and the spatial error model (Angulo et al., 2017) (Eq. (4)), which are widely used to predict urban growth trends (Gao et al., 2020), regional environmental pollution (Hou and Zhu, 2022), and the spatial distribution of disasters such as floods and fires (Li et al., 2023; Song, 2017):

$$\boldsymbol{y} = \lambda \boldsymbol{W y} + \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{3}$$

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u} \tag{4}$$

$$\boldsymbol{u} = \rho \boldsymbol{W u} + \boldsymbol{\varepsilon} \tag{5}$$

where λ is the spatial effect coefficient in the spatial lag model (SLM); u represents the disturbance term with spatial dependence in the spatial error model (SEM); and ρ is the spatial effect coefficient for the disturbance term u.

2.2 Data description

This work selects a city in China as the assessment area, which includes 4 districts, 4 counties, and 1 county-level city, covering a total area of 11,445 km^2 with a permanent population of 9.634 million. Fire accident data were obtained from the China Statistical Yearbook (National Bureau of Statistics, 2025), which details the number of accidents and corresponding total direct economic losses in each administrative region per year. Considering the city's rapid development and randomness of fire, average fire accident data from 2017 to 2021 were extracted for the study, including (1) the number of fire accidents per square kilometer per year and (2) the average direct economic loss per accident.

The factors affecting fire accident frequency and average direct economic losses can be categorized into four main groups (Xu, 2012; Zhang and Huang, 2013): (1) population vulnerability, which is primarily represented by population density, the proportion of the population aged below 15 and over 65, and the proportion of illiteracy among those aged 15 and above; (2) economic vulnerability, which is represented mainly by per capita disposable income at the district and county levels (Song, 2017; Zhang and Huang, 2013); (3) natural meteorological characteristics, which is represented primarily by average surface water resources (Zhou, 2012); and (4) building vulnerability, which is represented mainly by POI kernel density and the degree of personnel concentration around high-risk POIs (Tan et al., 2023). The degree of personnel concentration around high-risk POIs is calculated primarily as follows. Referring to the "Code for planning of urban fire control" (GB51080-2015) and other relevant literature (Tan et al., 2023), the collected POI data are categorized into four main types of high-risk locations: crowded places, flammable and explosive areas, vulnerable areas, and important sites, as detailed in Table 1. The POI indicator weight for different high-risk locations is based on the work of Tan et al. (2023). There is a significant correlation between the average number of people around high-risk POIs and fire risk (Song, 2017). Therefore, the value of personnel concentration around high-risk POIs is obtained by multiplying the population per unit area in the high-risk POI region by the POI indicator weight. The entire calculation is carried out via ArcGIS software. The specific process is as follows: (1) the Fishnet function is used to divide the studied area into grids with a resolution of 100 m × 100 m, and then the Raster function is used to count the population density data (from LandScan) within each grid. (2) For each POI, its type and its corresponding grid location are identified, the population density data within the grid are obtained, which are multiplied by the POI type weight to obtain the corresponding personnel concentration around this high-risk POI. (3) On the basis of the boundaries of the administrative areas within the city, all the POIs within different administrative areas are counted. For each area, the average value of the personnel concentration around different POIs is calculated to characterize the personnel concentration around high-risk POIs in the region.

Notably, in addition to the above parameters, many other factors can be used to characterize the frequency and loss of fires. However, owing to the limited accessibility of the data and the high degree of autocorrelation among some factors, only the aforementioned parameters have been selected for analyzing the fire risk characteristics within a city.

The sources and specific information for all the relevant data mentioned above are detailed in Table 2. After collinearity testing, the fitting coefficient between population density and POI kernel density was greater than 0.85, as shown in Fig. 1. Therefore, POI kernel density was not used for spatial regression. The fitting coefficients among the other variables were all less than 0.8. Therefore, the main explanatory variables used were population density, per capita disposable income, the proportion of the population aged 15 and over 65, the proportion of illiteracy among those aged 15 and above, surface water resources, and the degree of personnel concentration around high-risk POIs.

2.3 Optimization for regression modeling

The use of a spatial regression model to predict fire risk is a complex process. Verifying the accuracy and rationality of the selected independent variables and regression models is necessary. To determine the optimal spatial regression model efficiently, an optimized process is proposed. The specific process is depicted in Fig. 2:

(1) On the basis of the analysis in Section 3.1, the explanatory variables with significant correlations were extracted from the set of influencing factors, and ordinary least squares (OLS) regression was performed between these variables and the frequency of fire accidents.

(2) Using the results of the OLS regression, a Lagrange

Table 1 Classification of POI data

High-risk location type	POI data type	Indicator weight (Tan et al. 2023)
Crowded places	Shopping services, catering services, sports and leisure services, transportation hubs (airports, train stations, bus stations, subway and bus stations, etc.)	0.40
Flammable and explosive areas	Industrial parks and mineral, metallurgy and chemical, fireworks and firecracker factories, etc.	0.55
Population vulnerable areas	Educational, scientific, and cultural services (training institutions, kindergartens, etc.), medical institutions (emergency centers, clinics, disease prevention agencies, etc.), and nursing homes	0.40
Important sites	Government agencies and social organizations (at the district and county level), research institutions, tourist attractions (museums, historical sites, scenic areas, etc.), educational training (science museums, research institutions, libraries, etc.), and media and culture (radio, television, art galleries, exhibition halls, etc.)	0.30

Table 2 Statistical data utilized in this paper, its basis of application, and the data sources

Data type	Data name	Basis of application	Data source	Relationship with the severity/frequency of fires
Fire accident data	Number of fire accidents in various districts and counties from 2017 to 2021	—	China Statistical Yearbook 2018–2022 (National Bureau of Statistics, 2025)	—
	Direct economic losses from fire accidents in various districts and counties from 2017 to 2021	—	China Statistical Yearbook 2018–2022 (National Bureau of Statistics, 2025)	—
Population vulnerability	Population density	(Zhou, 2012)	Landscan (Oak Ridge National Laboratory, 2025)	Positive correlation
	Proportion of the population aged below 15 and over 65	(Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2015)	China Population Census Yearbook 2020 (Office of the Leading Group of the State Council for the Seventh National Population Census, 2025)	Positive correlation
	Proportion of illiteracy among those aged 15 and above	(Bispo et al., 2023)	China Population Census Yearbook 2020 (Office of the Leading Group of the State Council for the Seventh National Population Census, 2025)	Positive correlation
Economic vulnerability	Per capita disposable income	(Xu, 2012)	China Statistical Yearbook 2018–2022 (National Bureau of Statistics, 2025)	Related to the area development degree (Xu, 2012)
Natural meteorological characteristic	Average surface water resources	(Xu, 2012)	China Statistical Yearbook 2018–2022 (National Bureau of Statistics, 2025)	Negative correlation
Building vulnerability	POI data	(Song, 2017)	Amap API	Positive correlation

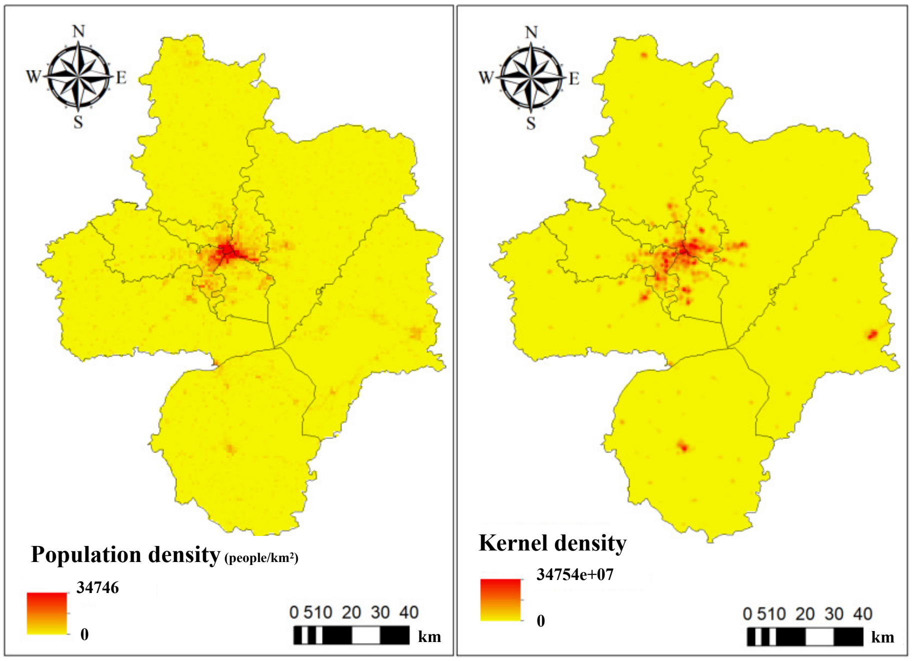


Fig. 1 High correlation between population density and POI kernel density.

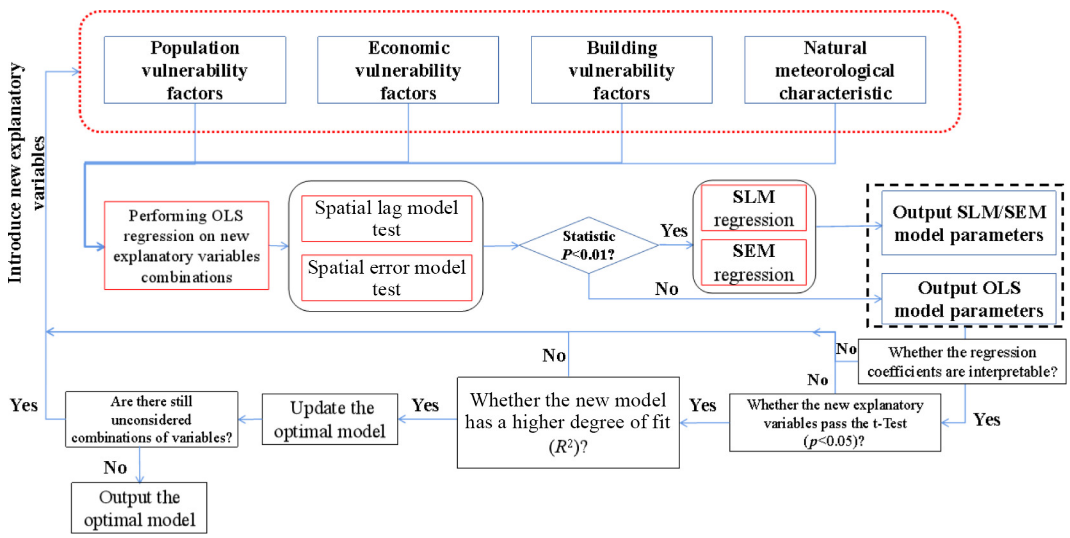


Fig. 2 Flowchart for determining the optimal regression model.

multiplier (LM) test was conducted to analyze whether the indicators in the spatial lag model (SLM) test and the spatial error model (SEM) test rejected the "no spatial autocorrelation" hypothesis. If the hypothesis is rejected, the distribution of fire accident frequency has spatial autocorrelation, and either an SLM or SEM regression is performed to output the corresponding model parameters. Otherwise, the OLS regression parameters are output.

(3) On the basis of the obtained output parameters, the first step is to reference the positive/negative correlations between various factors and fire frequency in Table 2 to ensure that the model parameters are consistent, indicating that the indicators have interpretability. A t test is subsequently conducted to check the significance of the explanatory variables involved in the output model parameters, and the fitness of the new model is evaluated to determine whether it exceeds the current optimal model. If all the above conditions are met, the new model is updated as the new optimal model.

(4) Check whether there are still factor combinations that have not been used for modeling. If such combinations exist, steps (1)–(3) are repeated until all the combinations have been tested.

Additionally, to avoid instability in the regression coefficients due to multicollinearity after too many variables are selected, no more than three indicators are introduced for regression modeling.

Figure 3 presents a typical case of the optimization process for regression modeling within the flowchart. After one cycle, the SLM model under the new combination of indicators has a greater degree of fit.

3 Results and discussion

3.1 Spatial distribution characteristics of fire risk and related factors

Figure 4(a) presents the frequency of fire accidents across different

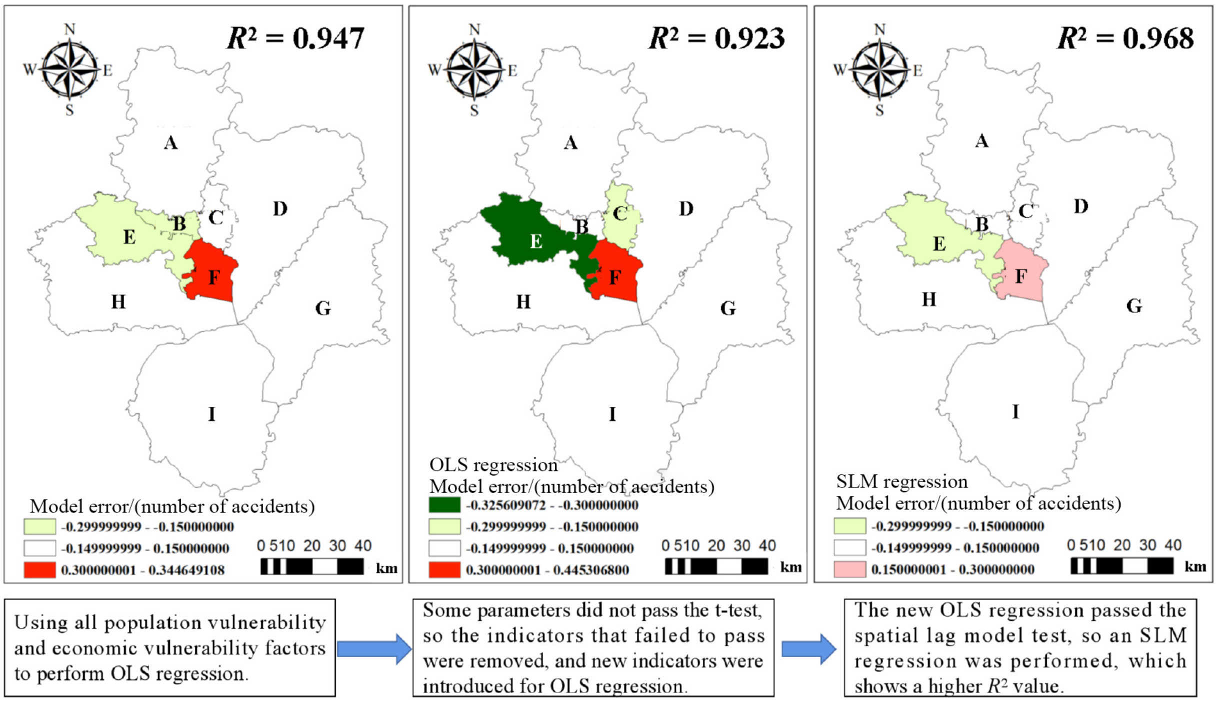


Fig. 3 Typical case of the optimization process, where the model error unit in the figure is the number of accidents per square kilometer.

areas, showing that high-frequency areas are predominantly concentrated within the urban area, with a decreasing trend in frequency as the distance from the city center increases. The results of Moran's I index test indicate an I value of 0.255 and a p value of 0.011, suggesting significant spatial autocorrelation. Figures 4(b) and 4(c) depict the distributions of population density and per capita disposable income, respectively. The distribution characteristics of these two factors clearly align closely with the accident frequency. Figures 4(e) and 4(f) show the distributions of the proportion of vulnerable populations and the proportion of illiterate individuals. Accurately determining the correlation between these two factors and fire frequency is difficult, but overall, both are negatively correlated with the frequency of accidents, which may be attributed to the negative correlation between per capita disposable income and these two factors. In summary, considering the observations and test results, there is a significant association between accident frequency and various factors, with spatial autocorrelation. Therefore, a spatial autocorrelation model can be employed to quantify the internal relationship between these indicators and frequency.

Figure 4(d) presents the average direct economic loss per accident across different regions. The results of Moran's I test indicate an I value of –0.106 and a p value of 0.895, suggesting the absence of significant spatial autocorrelation. Additionally, no substantial correlation was observed between the distribution of population density/disposable income and economic losses. However, there is an overall positive correlation between the distributions of vulnerable population ratios and illiteracy rates and losses, although this relationship exhibits considerable deviation. For example, while the main reason for the high economic losses in region **A** can be attributed to the existence of a high illiteracy rate and the highest rate of vulnerable populations in the region, it is difficult to draw similar deductions for some other regions, such as region **C**.

To further analyze the possible reasons for the distribution characteristics of economic losses, kernel density spatial distribution maps of different types of POIs were created, as shown in Fig. 5. All types of POIs have higher kernel density values in the city center. Moreover, compared with other types of POIs, the kernel density of flammable and explosive POIs is relatively lower at the urban boundary intersections and higher within the **C** district. This could be one of the reasons for the greater average loss of accidents in the **C** district.

3.2 Fire frequency analysis based on spatial regression models

Spatial modeling of the fire accident frequency distribution from 2017 to 2021 was conducted on the basis of the influencing factors listed in Table 1. By comparing and cycling through all possible combinations via the optimal procedure described in Section 2.3, the indicator models that meet the output conditions throughout the entire process are organized and ranked, as shown in Table 3. As the number of indicators increases, the overall degree of fit R^2 shows an upward trend, and the optimal explanatory variable combination is (1) population density and (2) per capita disposable income. The best model is the SLM model, and the specific model parameters are presented in Table 3. Notably, although the model with three indicators was tested for its ability to predict accident frequency, no better regression R^2 was observed than that of the double indicator model. Therefore, only two-parameter fitting is used in the end.

On the basis of the data in Table 3, the following observations can be made: (1) In Table 3, the average number of people around high-risk POIs and population density are the indicators most closely associated with fire frequency, whereas water resources (environmental humidity) have a weaker correlation with accident frequency. This suggests that in the study area, the impact of the environment on accident frequency is much lower than that on human behavior. (2) When per capita disposable income is considered a single variable, there is a positive correlation between frequency and income. This is primarily due to the rapid development of the economy, which has led to an increase in urban population density and consequently an increase in

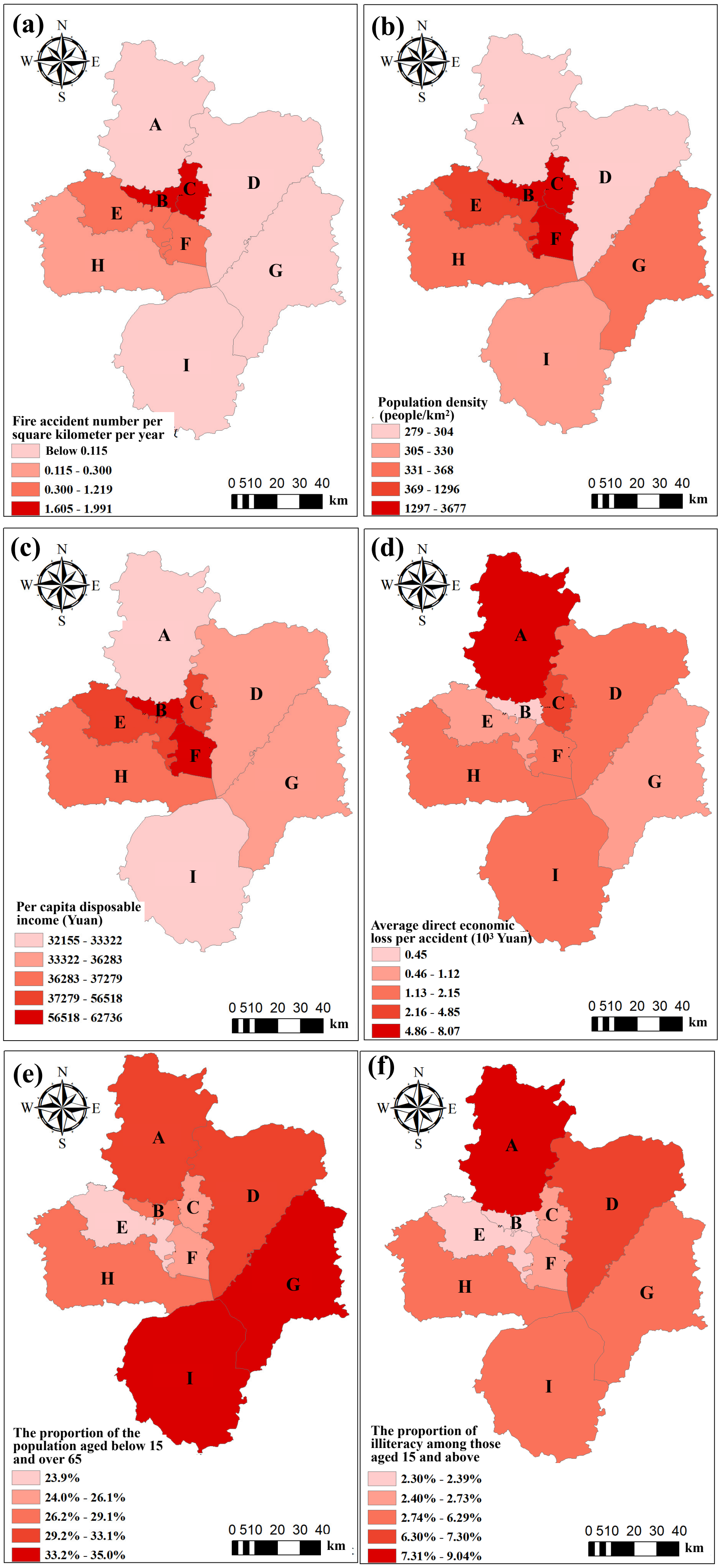


Fig. 4 Spatial distributions of different indicators.

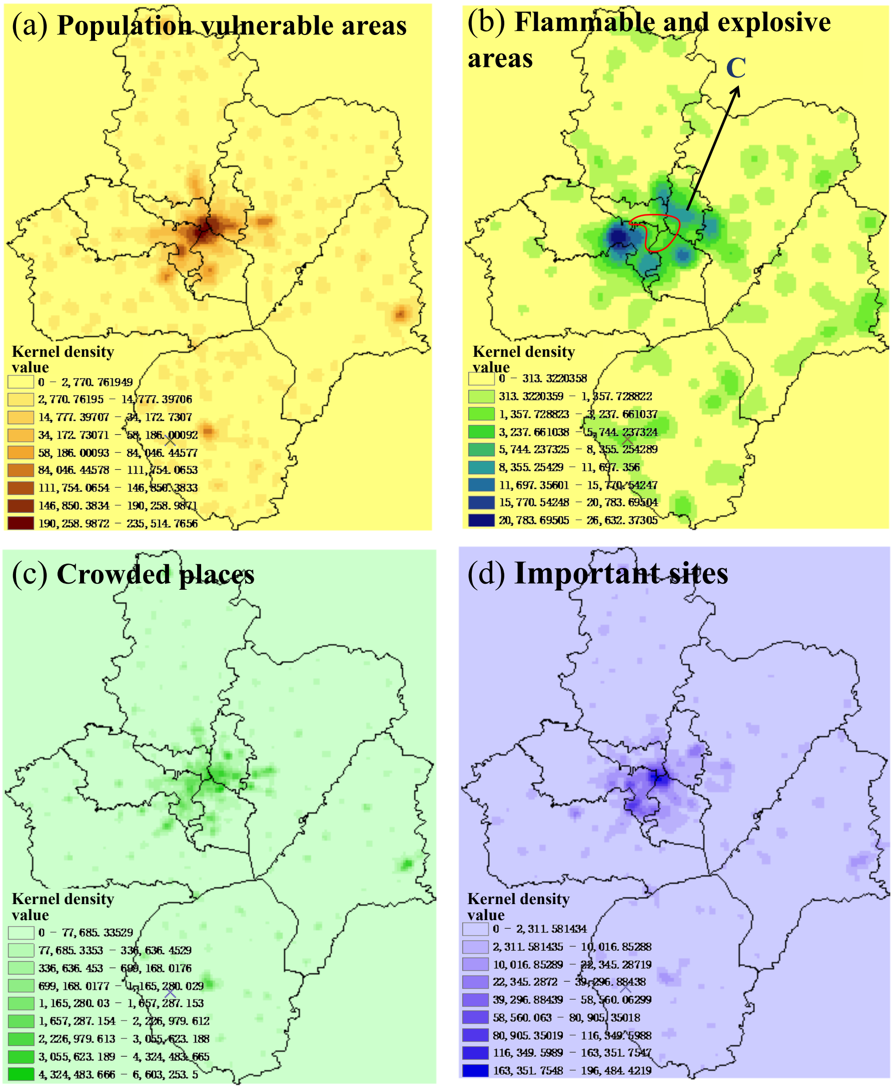


Fig. 5 Kernel density spatial distribution maps of different types of POIs.

Table 3 Regression model parameters with different combinations of indicators

Original indicator regression results				Regression results for indicators after logarithmic transformation	
Single indicator model	R^2	Double indicator model	R^2	Single indicator model	R^2
Population density	**0.945**	Population density & per capita disposable income	**0.968**	Population density	0.961
Degree of personnel concentration around high-risk POIs	**0.915**			Degree of personnel concentration around high-risk POIs	0.975
Proportion of the population aged below 15 and over 65	0.878			Proportion of the population aged below 15 and over 65	0.635
Proportion of illiteracy among those aged 15 and above	0.833			Proportion of illiteracy among those aged 15 and above	0.904
Average surface water resources	0.033 5			Average surface water resources	0.010
Per capita disposable income	0.710			Per capita disposable income	0.905

accident frequency. In the dual-indicator model of population density and per capita income, per capita disposable income is negatively correlated with fire frequency. This may suggest that the overall speed of improvement in the degree of fire prevention can meet the rapid development requirements of a city (Xu, 2012).

To further clarify the correlation between the proposed model and accident frequency, comparison plots of the predicted values versus the actual values are plotted in Figs. 6 and 7. These figures indicate that the proposed model can accurately predict accident frequencies within the urban area but fails to effectively estimate the accident values for the counties. This is primarily due to the significant differences in accident frequencies between the counties and the urban area, which the model regression cannot effectively address. Similar phenomena have been observed in previous studies (Bispo et al., 2023; Song, 2017). Considering the lower weight of the counties due to their lower accident frequency values, this paper further applies logarithmic transformation to all indicators and conducts optimization following the process in Fig. 2. The results indicate that the model considering the distribution of high-risk POIs is the optimal model, with the model parameters and errors detailed in Table 4 and Figs. 6 and 7. This model significantly outperforms the spatial model without logarithmic transformation in predicting the counties, suggesting that logarithmically transformed indicators in spatial autoregressive modeling have superior capabilities in predicting disaster frequencies at the city scale. Moreover, when assessing the accident frequency distribution at the municipal scale, it is necessary to consider the distribution characteristics of high-risk POIs.

3.3 Analysis of direct fire economic loss characteristics via spatial regression models

The direct economic losses caused by fire accidents are closely related to accident development patterns and the vulnerability of disaster-bearing entities. Therefore, to quantitatively analyze the main factors affecting direct economic losses, spatial regression modeling was conducted using the indicators proposed in Section 2.1. The specific modeling process is consistent with Fig. 4, and the regression results are shown in Table 5. The following results can be obtained: (1) The illiteracy rate and climate humidity are the main factors affecting the direct economic loss per accident, and the two can be considered the best explanatory variable combination. (2) The optimal model is the OLS model, indicating that the economic loss situation does not exhibit a spatial autocorrelation effect in this city. (3) The degree of personnel concentration around high-risk POIs has a relatively low correlation with direct economic losses, suggesting that the current POI risk indicator cannot reflect the accident severity across different regions. This can be attributed to the fact that the current POI indicator value has not yet considered the scale and operational characteristics of specific risk points. (4) The proportion of the population aged below 15 and over 65, population density, and per capita disposable income also have relatively low correlations with direct economic losses. In contrast, as discussed in Section 3.2, these parameters are significantly correlated with the fire frequency. These findings suggest that these indicators may have a greater influence on the frequency of accidents than on their consequences.

Figure 8 shows the difference between the predicted direct fire economic loss per accident and the actual statistical data. The prediction model is able to reflect the overall trend of average fire losses by using climate and social vulnerability indicators, but there is still a significant error. **C** District and **D** County are the two areas with the largest prediction deviations, which may be related to factors such as the ratio of aging buildings and the specific risk characteristics of the POIs. Specifically, **C** District is the only old industrial urban area in this city, with a relatively dense concentration of large-scale high-risk POIs and more severe urban aging issues, leading to actual losses that are higher than the model's predictions. **D** County, on the other hand, has a relatively

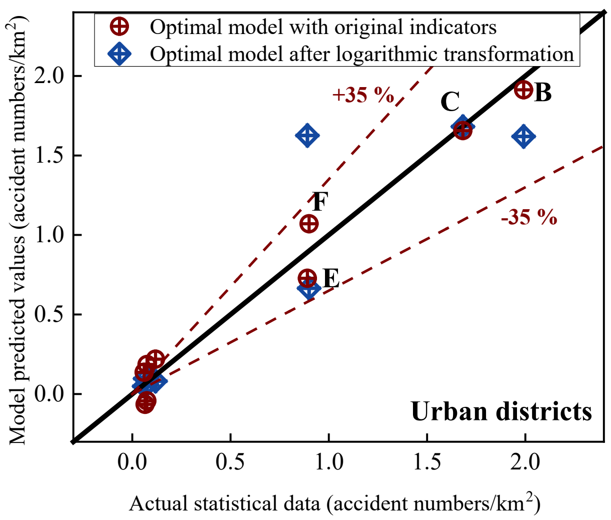


Fig. 6 Comparison between the statistical fire frequency data and model-predicted values (urban districts).

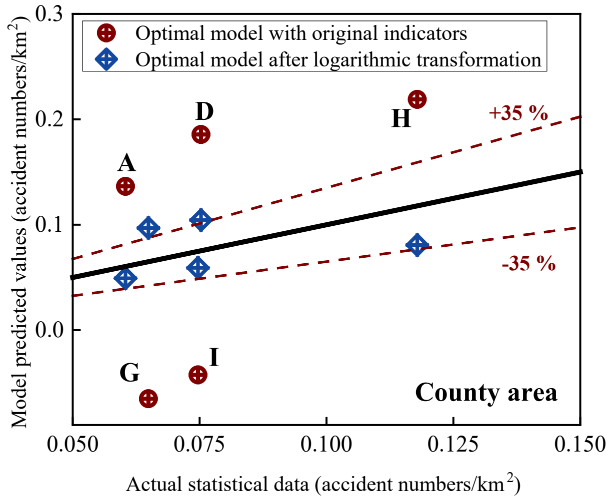


Fig. 7 Comparison between the statistical fire frequency data and model-predicted values (county area).

Table 4 Optimal model parameters (both are SLM models)

Optimal model by using original indicators				
Explanatory variable name	Model estimated coefficient	Standard error	*z* value	*p* value
Population density	0.000 506 5	0.000 076 8	6.60	< 0.001
Per capita disposable income	−0.000 024 8	9.79×10^{-6}	−2.53	0.011
Disturbance term	0.458 289 2	0.298 245		
Spatial effect coefficient term	0.008 282 0	0.002 321 1	3.57	< 0.001
Optimal model by using indicators after logarithmic transformation				
Explanatory variable name	Model estimated coefficient	Standard error	*z* value	*p* value
Degree of personnel concentration around high-risk POIs	1.178 178	0.256 536 9	4.59	< 0.001
Disturbance term	−1.388 115	0.603 252 1		
Spatial effect coefficient term	−0.025 341 8	0.009 232 3	−2.74	0.006

Table 5 Regression model parameters with different combinations of indicators for direct fire economic loss per accident

Single indicator model	R^2	Double indicator model	R^2
Proportion of illiteracy among those aged 15 and above	0.2620	Proportion of illiteracy among those aged 15 and above & average surface water resources (OLS model)	0.4992
Average surface water resources	0.3090		
Proportion of the population aged below 15 and over 65	0.0090		
Population density	0.0300		
Degree of personnel concentration around high-risk POIs	0.1106		
Per capita disposable income	0.1101		

Table 6 Optimal model parameters for direct fire economic loss per accident (OLS model)

Explanatory variable name	Model estimated coefficient	Standard error	z value	p value
Proportion of illiteracy among those aged 15 and above	42.73453	28.3126	1.51	0.143
Average surface water resources	−1,349.7	800.8074	−1.69	0.143
Disturbance term	5.98051	3.788877		

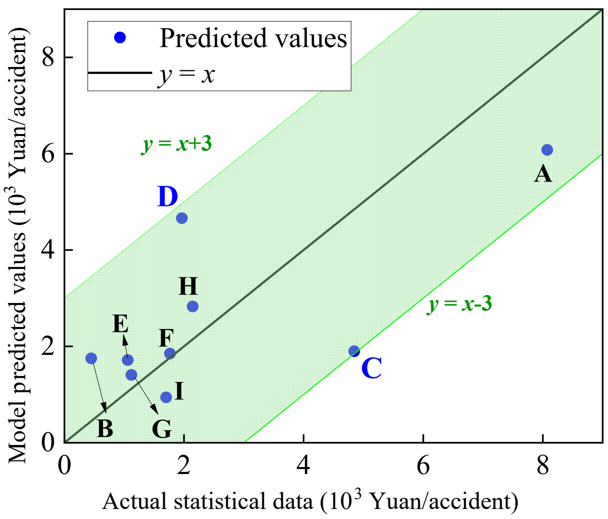


Fig. 8 Comparison between the direct fire economic loss statistical data and model-predicted values.

recent development history, with strategic emerging industries such as intelligent manufacturing, photovoltaics, and new energy as its main industries in recent years, and thus has a lower fire risk than traditional chemical industrial parks. Owing to limited relevant statistical data and the difficulty of quantitatively assessing fire risk across different industries, it is currently challenging to obtain better regression results, and further research needs to be conducted in the future.

4 Conclusions

This paper takes a city in China as an example to statistically analyze the correlations between various influencing factors (economic vulnerability, population vulnerability, natural meteorological characteristics, etc.) and the frequency of fire accidents, as well as the direct economic losses from fire accidents. Spatial regression methods are employed to model the accident frequency and direct economic losses, clarifying the quantitative relationships between fire characteristics and various indicators. The main conclusions of this work are as follows.

(1) In the study area, a significant spatial autocorrelation effect on the distribution of accident frequency was observed, but the spatial distribution of accident losses did not exhibit obvious autocorrelation. Compared with the critical weight matrix, the distance weight matrix better reflects the disaster correlation between different spaces within a city. Additionally, a single vulnerability indicator cannot fully explain the accident distribution characteristics.

(2) An optimal model determination procedure was proposed and used to obtain spatial regression models of fire characteristics that have comprehensive interpretability and predictive accuracy. For the study area of this work, population density, per capita income, and the degree of personnel concentration around high-risk POIs were the most effective indicators for constructing a spatial regression model of accident frequency. The SLM model, which considers spatial effects, can effectively predict the overall accident frequency. Moreover, logarithmically transformed indicators can effectively predict accident frequency in areas with low population density. Therefore, in the follow-up work of city-scale fire risk assessment, the data weight of low-frequency areas can be increased through logarithmic transformation to improve the overall prediction accuracy.

(3) For the study area of this work, the illiteracy rate and average surface water resources were the most effective indicators for constructing a spatial regression model of direct economic losses. The distribution of high-risk POIs can qualitatively explain the deviation between the regression model and actual values.

The present work not only enriches the research on city-scale fire risk assessment but also shows that the optimal regression model determination process can provide technical support for the application of spatial regression methods in fire risk assessment. Importantly, while the distribution characteristics of points of interest (POIs) can qualitatively explain the direct economic loss situation in some areas, it is currently challenging to quantify the specific weights of different types of risk points. Therefore, in future work, it is necessary to conduct quantitative risk assessments of different types of buildings to establish a more effective prediction model for disaster losses at the city scale.

Acknowledgements

This research work was sponsored by the National Natural Science Foundation of China (Grant No. 72091512).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

Angulo, A., Burridge, P., Mur, J. 2017. Testing for a structural break in the weight matrix of the spatial error or spatial lag model. *Spatial Economic Analysis*, **12**: 161–181.

Bispo, R., Vieira, F. G., Bachir, N., Espadinha-Cruz, P., Lopes, J. P., Penha, A., Marques, F. J., Grilo, A. 2023. Spatial modelling and mapping of urban fire occurrence in Portugal. *Fire Safety Journal*, **138**: 103802.

Bivand, R., Piras, G. 2015. Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, **63**(18): 1–36.

Gao, C., Feng, Y., Tong, X., Lei, Z., Chen, S., Zhai, S. 2020. Modeling urban growth using spatially heterogeneous cellular automata models: Comparison of spatial lag, spatial error and GWR. *Computers, Environment and Urban Systems*, **81**: 101459.

Getis, A., Aldstadt, J. 2004. Constructing the spatial weights matrix using a local statistic. *Geographical Analysis*, **36**: 90–104.

Hou, H., Zhu, Y. 2022. Analysis of spillover effects of regional environmental pollution: An interprovincial study in China based on spatiotemporal lag model. *Environmental Science and Pollution Research International*, **29**: 836–853.

Hu, J., Shu, X., Xie, S., Tang, S., Wu, J., Deng, B. 2019. Socioeconomic determinants of urban fire risk: A city-wide analysis of 283 Chinese cities from 2013 to 2016. *Fire Safety Journal*, **110**: 102890.

Li, H., Zhang, C., Chen, M., Shen, D., Niu, Y. 2023. Data-driven surrogate modeling: Introducing spatial lag to consider spatial autocorrelation of flooding within urban drainage systems. *Environmental Modelling & Software*, **161**: 105623.

Liu, Z. G., Li, X. Y., Jomaas, G. 2022. Effects of governmental data governance on urban fire risk: A city-wide analysis in China. *International Journal of Disaster Risk Reduction*, **78**: 103138.

Ministry of Housing and Urban-Rural Development of the People's Republic of China, Ministry of Public Security of the People's Republic of China. 2015. Code for planning of urban fire control: GB 51080 – 2015. Beijing: China Architecture & Building Press.

National Bureau of Statistics. 2025. China Statistical Yearbook. Available at https://www.stats.gov.cn/sj/ndsj/

Oak Ridge National Laboratory. 2025. About LandScan. Available at https://landscan.ornl.gov/about

Office of the Leading Group of the State Council for the Seventh National Population Census. 2025. China Population Census Yearbook. Available at https://www.stats.gov.cn/sj/pcsj/rkpc/d7c/

Song, C. 2017. The spatiotemporal dynamic modeling analysis of fire risk for the location planning of urban fire stations. Ph.D. Thesis. University of Science and Technology of China.

Tan, L. L., Qu, N., Han, L., Sui, Y. F. 2023. GIS-SAVEE model-based assessment of fire risks in industrial cities. *Journal of Institute of Disaster Prevention*, **25**(2): 97–103. (in Chinese)

Wang, Z. 2021. Study on urban fire based on data mining and GIS. Master Thesis. China University of Mining and Technology.

Xu, B. 2012. Macroscopical influences of economic development and climate change on urban fire spatial-temporal variations in China. Ph.D. Thesis. Nanjing University.

Zhang, N., Huang, H. 2013. Social vulnerability for public safety: A case study of Beijing, China. *Chinese Science Bulletin*, **58**: 2387–2394.

Zhou, Y. 2012. Research on regressive model of fire and meteorological factors. Master Thesis. Hefei University of Technology.

Hanchao Ma received his B.Sc. degree in safety engineering from School of Emergency Management and Safety Engineering, China University of Mining & Technology, Beijing, Beijing, China. He is currently pursuing the Ph.D. degree in safety science and engineering from School of Safety Science, Tsinghua University, Beijing, China. His research interests include fire dynamics, disaster prediction and emergency management.