



# Analyzing spatial non-stationarity effects of driving factors on landslides: a multiscale geographically weighted regression approach based on slope units

Feifan Lu<sup>1</sup> · Guifang Zhang<sup>1,2,3</sup> · Tonghao Wang<sup>1</sup> · Yumeng Ye<sup>1</sup> · Junwei Zhen<sup>4</sup> · Wanli Tu<sup>4</sup>

Received: 12 March 2024 / Accepted: 22 August 2024  
© The Author(s) 2024

## Abstract

Landslides pose a significant threat to the safety of people and their property. Previous landslide susceptibility assessment efforts have often overlooked the impact of spatial variations in the distribution of driving factors on disaster occurrences. Geographically Weighted Regression (GWR) is the most commonly used method for spatially heterogeneous modeling. However, it uses a single bandwidth and cannot explain the spatially varying scaling parameters of each factor. This study focuses on Luhe as the research area and introduces the Multiscale Geographically Weighted Regression (MGWR) model. By considering different spatial scales, the spatial non-stationarity of the relationship between landslides and their driving factors was explored. The final results indicate that the MGWR model outperforms the global regression Ordinary Least Squares (OLS) model and the traditional GWR model. Within the study area, the influence of various factors on landslides is complex, exhibiting a multipolar pattern in space. Elevation emerges as the dominant driving factor in the research area, showing a negative correlation with landslides. By employing the concept of multiscale spatial local regression, one can better analyze the interaction patterns between factors and landslides, providing improved insights for disaster prevention and mitigation.

**Keywords** Landslides · Slope unit · Spatial non-stationarity · Moran's I · Multiscale geographically weighted regression

## Introduction

Landslides, as a frequently occurring global geological hazard, have significant impacts on the natural environment, human society, and economic development (Petley 2012; Han et al. 2019). Relevant studies indicate that between 1900 and 2020, landslide events resulted in a total of 66,438 fatalities and ~ 10.8 billion USD in economic losses, with Asia being the most affected region (Guha-Sapir et al. 2020).

Landslide susceptibility mapping (LSM), which describes the spatial probability of landslide occurrences based on local topographic conditions, is an essential tool for promoting landslide risk management and reducing disaster losses (Huang et al. 2020; Nhu et al. 2020; Kaur et al. 2024). In previous studies, most scholars chose to evaluate susceptibility by focusing on causative factors closely related to landslide disasters in the study area, such as elevation, slope, aspect, lithology, and then calculated the importance of the selected factors in the prediction process (Wang et al. 2020; Liu et al. 2021). The assessment methods can be classified into two categories: (1) qualitative analysis, such as the Analytic Hierarchy Process (Zhang et al. 2016a; Panchal and Shrivastava 2022). These methods often have subjectivity, relying heavily on expert knowledge and experience, making them susceptible to human interference (Zhang et al. 2022). (2) Quantitative methods are typically based on mathematical models, and common quantitative analyses include information value (Wang et al. 2019), logistic regression (LR) (Budimir et al. 2015), support vector machines (SVM)

✉ Guifang Zhang  
zhgfang@mail.sysu.edu.cn

<sup>1</sup> School of Earth Sciences and Engineering, Sun Yat-Sen University, Zhuhai 519082, China

<sup>2</sup> Guangdong Provincial Key Lab of Geological Processes and Mineral Resources, Zhuhai 519082, China

<sup>3</sup> Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China

<sup>4</sup> The Seventh Geological Brigade of Guangdong Provincial Geological Bureau, Huizhou 516000, China

(Huang and Zhao 2018), random forests (RF) (Youssef et al. 2016; Sun et al. 2021), among others. These methods can eliminate human interference and objectively infer the likelihood of landslides. As research advances, ensemble learning in landslide susceptibility assessment is gaining favor among scholars (Liu et al. 2023). Comparatively, assessment methods involving multiple model couplings often yield more accurate landslide susceptibility zones than those obtained from single models (Arabameri et al. 2019; Zhao et al. 2021; Zhou et al. 2021).

However, past research has often focused on improving the accuracy of landslide susceptibility assessment but has rarely considered the issue of spatial heterogeneity of landslide factors. Most studies generally assume a fixed relationship between influencing factors and landslide occurrences within a given region (Gu et al. 2022). Traditional machine learning models, such as SVM and RF, construct classifiers based on training datasets to achieve optimal predictions. While this to some extent diminishes the impact of subjective human influence and enhances predictive accuracy compared to statistical models, they remain inherently global in nature. Global models overlook variations in the contributions of individual influencing factors within the study area, potentially resulting in biases in parameter estimation and suboptimal predictions (Anselin and Griffith 1988). Owing to spatial distribution differences, i.e., the presence of spatial heterogeneity, a certain factor may have varying impacts on disasters in different sub-regions of the study area, exhibiting its spatial non-stationarity relationship with disasters. GWR offers unique advantages in addressing the spatial non-stationarity of geospatial data (Wheeler and Páez 2009). It incorporates the geographic coordinates of the data into the regression parameters, performing local fitting by accounting for changes in explanatory factors corresponding to changes in geographic location. For instance, generally, areas with steeper slopes are more prone to landslides compared to those with gentler slopes. This is because as the slope increases, the shear stress at the foot of the slope gradually increases, and the slope is more susceptible to the influence of tensional fractures. GWR reveals the spatial variation of its relationship with landslides by conducting regressions separately in areas with different slopes. The method of local regression is gaining attention in landslide assessment studies. Erenler and Duzgun (2010) found that incorporating the spatial structure of regression parameters significantly enhances model explanatory power. Feuillet et al. (2014) highlighted that the GWR model can retain crucial correlated factors that might perform poorly in global models. Chalkias et al. (2020) demonstrated the GWR model's effectiveness in explaining spatial relationships between landslides and driving factors in Greece. Additionally, the GWR model has also demonstrated advantages over traditional statistical and machine learning models in predicting

landslide susceptibility (Hong et al. 2017; Boussouf et al. 2023).

Despite the GWR model demonstrates high applicability in exploring the spatial non-stationarity of landslide factors, its use of a single "optimal average" bandwidth for estimating all influencing factors prevents it from explaining the parameter variations of each factor at different spatial scales. As a crucial control parameter in constructing the GWR model, the bandwidth has a far greater impact on the final fitting results than the choice of the kernel function (Fotheringham et al. 2002; Chalkias et al. 2020). In the GWR model, observations near each calibration point typically exhibit similar characteristics, known as spatial autocorrelation (Song et al. 2022). However, the spatial autocorrelation strength varies for each factor. The relationship between each landslide factor and landslides in space may manifest as local, regional, or global. If these differences are consolidated into a single spatial scale, it may lead to misjudgments in the regression results. Multiscale Geographically Weighted Regression (MGWR) (Fotheringham et al. 2017) addresses the geographic multiscale issue in GWR to some extent by seeking unique bandwidths for different factors based on the extent of their spatial heterogeneity. Compared to a single bandwidth, MGWR minimizes model overfitting and reduces bias in the parameter estimation process to the greatest extent (Oshan et al. 2019). Additionally, unlike the stepwise regression model, the GWR model cannot automatically remove irrelevant variables (Chen et al. 2022). This necessitates considering the nature of spatial heterogeneity when evaluating land features, an aspect often overlooked in previous research. The choice of evaluation units significantly influences the assessment of spatial non-stationarity in factors. Some studies opt to interpolate coefficients after obtaining regression results based on raster cells (Zulkaffi et al. 2023), but this approach may lead to illogical local fitting outcomes. As the fundamental units for landslide occurrences, slope units offer a relatively accurate representation of various terrain features within the disaster-prone area compared to grid units (Huang et al. 2021). This allows for a better representation of homogeneity within units and heterogeneity between units, highlighting the spatial distribution differences of factors.

This study analyzes the non-stationarity of the relationship between influencing factors and landslides at different spatial scales using the MGWR model. It should be noted that this non-stationarity refers to spatial relationships and does not yet involve temporal series. Luhe County in Guangdong Province is selected as the study area to compare the advantages of the MGWR model relative to global regression models (OLS) and traditional GWR model. Additionally, we explore the spatial patterns of landslide drivers and their relationship with landslides through visualization. The study aims to provide new methods and perspectives

for analyzing interaction patterns between factors and landslides, as well as for disaster prevention and reduction.

### Study area

Luhe is situated along the southeastern coast of Guangdong Province, spanning the latitude range of 115°24' to 115°49' N and the longitude range of 23°08' to 23°28' E (Fig. 1a). It covers a total area of 986 km<sup>2</sup>, with mountainous terrain accounting for 787 km<sup>2</sup>, i.e., 79.8% of the total area. Luhe exhibits a range in elevation from 12 m above sea level to 1219 m, located on the southeast side of the Lianhua Mountains Range (Fig. 1b). Its topography generally varies from higher in the west to lower in the east, with the steepest slope reaching 77° and an average slope of 16°. The average annual temperature in Luhe is 22 °C, with an annual average rainfall exceeding 1800 mm, most concentrated between May and July, accounting for 40% of the total annual rainfall. Four major faults traverse the region, with the Hetian Fault being an active fault, exhibiting an overall gentle undulating extension and large scale (Fig. 1b). Based on lithological characteristics and physical properties, the engineering rock groups in the study area can be categorized into five types: single-layer soil, double-layer soil, multilayer soil, stratified clastic rock, and massive intrusive rock formations (Fig. 4e). The primary mode of transportation within the area is road networks, with over 90% of towns connected by roads, rendering transportation highly convenient (Fig. 4f).

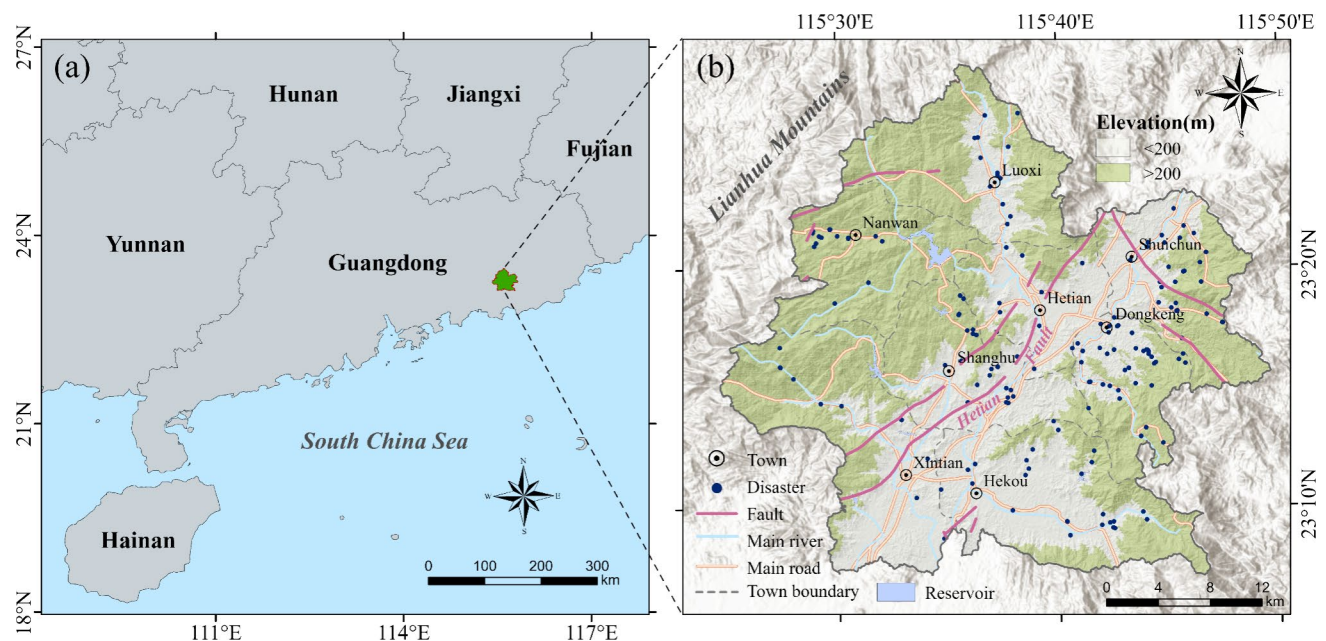
The region, abundant in mountainous terrain and ample rainfall, coupled with active tectonics and extensive land development, has made landslides a significant geological hazard, posing a severe threat to the safety of people and their property.

### Materials and methodology

This study first constructs slope units to establish the spatial adjacency relationships of driving factors, further examining the multicollinearity among factors and the spatial autocorrelation of each factor. Subsequently, we introduced variables into the MGWR model, obtained regression results for the factor coefficients, and compared the model's performance with OLS and GWR models. Finally, the spatial non-stationarity in the effects of factors on disasters were further discussed through visualization. Figure 2 shows the study framework for this paper.

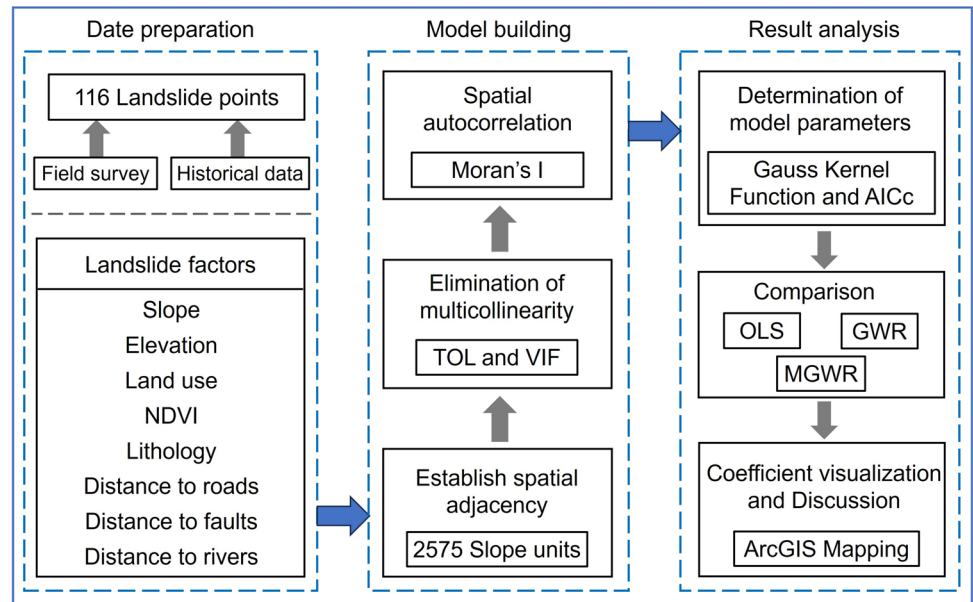
### Data preparation

The inventory of landslides in the study area comprises a combination of engineering geological surveys, remote sensing interpretation, and the collection of disaster points registered in the natural resources bureau database of Luhe (Fig. 1b). A total of 166 landslide sites were identified, of which 124 had a landslide volume of less than or equal to 10 × 10<sup>4</sup> m<sup>3</sup>. Figure 3 illustrates examples of four landslides within Luhe, predominantly characterized as soil slides with

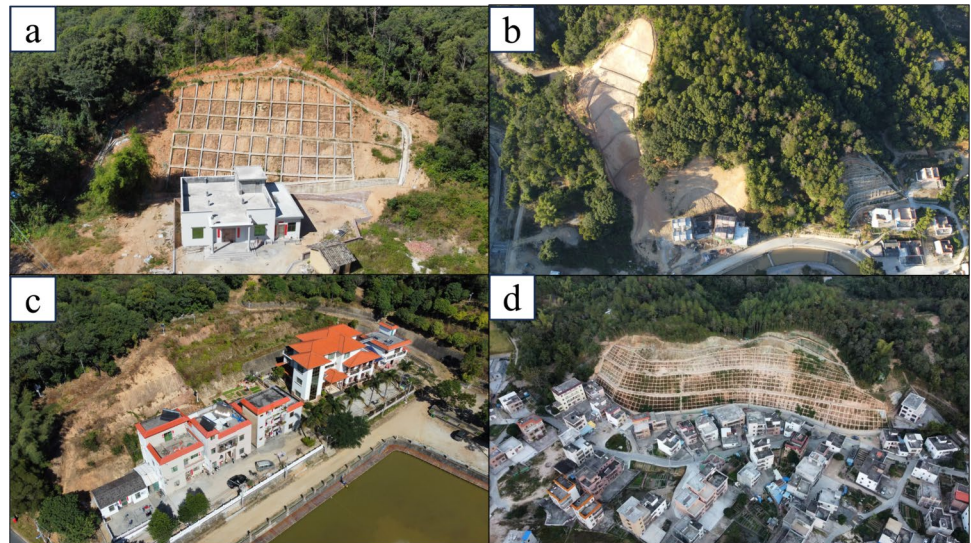


**Fig. 1** Basic information of the study area: (a) Location of the study area; (b) Map of topography, fault and disaster distribution in the study area

**Fig. 2** The overall flowchart of the study methods



**Fig. 3** Examples of landslides from field investigation and their locations



a planar morphology dominated by tongue-like features. In terms of spatial distribution, these landslides are mainly located in the low hills and ridges of the towns of Nanwan, Dongken, Luoxi, and Shanghu.

The Digital Elevation Model (DEM) data were employed for subsequent processing of landslide driving factors, such as slope. The land use data includes ten major land cover types: arable land, woodland, grassland, shrubland, wetland, water bodies, tundra, artificial surfaces, bare land, glaciers, and permanent snow, derived from GlobeLand30 (<http://loess.geodata.cn>). After comparison, it was found that the 2020 land use in Luhe was very similar to that in 2000, which to some extent mitigates the issue of data timeliness concerning land use at the time of landslide occurrence. Ultimately, we decided to use the 2020 land use data, which

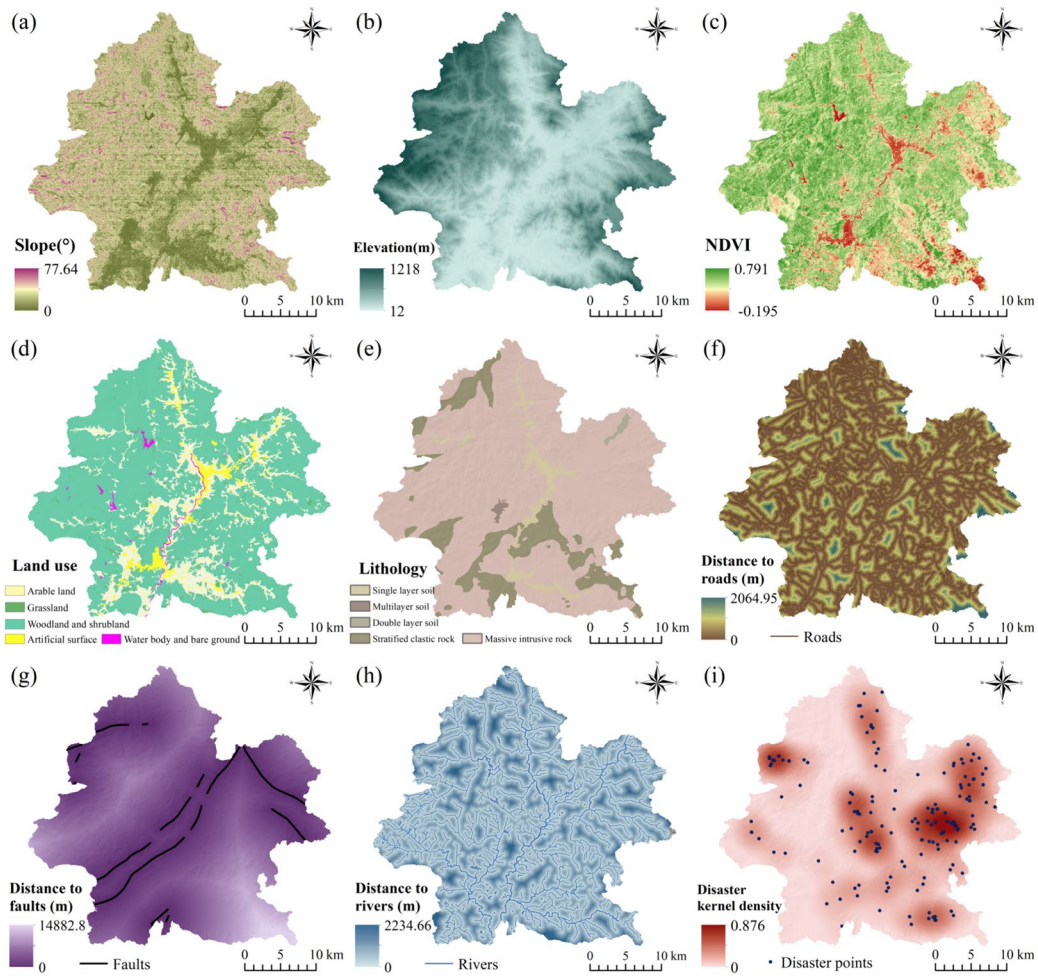
is closer to the completion time of the landslide survey and has higher data quality, as our landslide driving factor. Additionally, vector data, including roads, river networks, and faults, were utilized for distance calculations in ArcGIS. The characteristics of each dataset are summarized in Table 1. To facilitate analysis, all vector data were converted to raster format with a spatial resolution set at 10 m.

### Landslide driving factors

The occurrence of landslides is influenced by a variety of natural and human factors within a specific region. These controlling factors can be broadly divided into two categories: static or conditioning factors (e.g., topographic, geological conditions, and land cover) and dynamic or triggering

**Table 1** Data used for driving factors analysis in this study

| Data name            | Data format    | Scale or resolution | Data source                                  |
|----------------------|----------------|---------------------|--|
| DEM                  | GeoTIFF        | 12.5m               | ALOS GDEM                                    |
| Disaster points      | .shp (point)   | 1:50000             | Field work and remote sensing interpretation |
| Remote sensing image | GeoTIFF        | 30m                 | LIT product of LANDSAT-8                     |
| Road                 | .shp (line)    | 1:50,000            | Calibration data based on Amap               |
| Water body           | .shp (polygon) | 1:50,000            | Calibration data based on Amap               |
| Fault                | .shp (line)    | 1:50,000            | Zijin regional geological map                |
| Land use             | GeoTIFF        | 30m                 | GlobeLand30                                  |



**Fig. 4** Layers for driving factors analysis: (a) slope; (b) elevation; (c) NDVI; (d) land use; (e) lithology; (f) distance to roads; (g) distance to faults; (h) distance to rivers; (i) disaster kernel density

factors (e.g., rainfall and earthquakes) (Popescu 2002; Ng et al. 2021). Among these, rainfall is often considered a triggering factor, particularly short-term intense rainfall. When considering the influence of triggering factors, some studies choose to use matrix integration methods to combine landslide susceptibility maps with rainfall thresholds to obtain new susceptibility zoning (Segoni et al. 2018; Pradhan et al.

2019). Some scholars also incorporate rolling rainfall data (e.g., maximum rolling 12-h rainfall) into prediction models, deriving the spatiotemporal probability of landslides by fitting the nonlinear relationship between landslide spatial susceptibility and rolling rainfall (Ng et al. 2021; Xiao et al. 2023; Ren et al. 2024). However, this type of research requires highly accurate rainfall data, typically necessitating

the collection of precise rainfall amounts for periods before and after landslides, as well as the number of landslides triggered by these rainfall events. Due to the difficulty in obtaining rainfall data and concerns about the quality of the rainfall data within the study area, this study decided not to include rainfall and other triggering factors in the model, focusing solely on static factors. Referring to previous research outcomes and considering the characteristics of the study area, 8 landslide driving factors were ultimately selected: slope, elevation, Normalized Difference Vegetation Index (NDVI), land use, lithology, distance to roads, distance to faults, and distance to rivers. Simultaneously, we used the landslide kernel density values (number/km<sup>2</sup>) as the dependent variable (Fig. 4i), replacing the conventional binary variable (0 and 1). By employing this approach to cover the dependent variable across the entire study area and integrating it with subsequently established slope units, not only allows the model to perform regression over the entire region, showcasing its local regression characteristics, but also addresses the potential issue of imprecise local coefficients that may arise from interpolation. Data processing and visualization for all variables were conducted using ERSI ArcGIS 10.8 and ENVI 5.6 software, and the data layers for each variable are depicted in Fig. 4.

Overall, for all the driving factors investigated in this study, each variable exhibits different units and ranges. To facilitate comparison and enhance the fitting accuracy of the models, the following formula was employed for standardization:

$$x_{sij} = \frac{x_{ij} - \text{mean}(X_i)}{\text{std}(X_i)} \quad (1)$$

where  $x_{sij}$  is the standardized value for variable  $i$  at point  $j$ ;  $x_{ij}$  is the raw value of  $x_{sij}$ ;  $X_i$  is the variable  $i$  ( $X_i = x_{i1}, x_{i2}, \dots, x_{in}$ ); and  $\text{mean}(X_i)$  and  $\text{std}(X_i)$  are the mean and standard deviation of variable  $X_i$ , respectively.

In regression analysis, the common solution for handling categorical variables such as lithology is to create layers of binary values (dummy variables) for each class of an independent parameter (Ohlmacher and Davis 2003; Zhang et al. 2011). However, when there are too many parameters, this method can result in an excessively long regression equation, potentially leading to numerical issues or multicollinearity among the variables (Ayalew and Yamagishi 2005). If linear fitting is performed by directly assigning numerical codes to each lithology category, the subjective nature of the coding order can lead to completely opposite regression coefficients when the order is reversed. This not only fails to reflect the specific relationship between each category and landslides but also leads to differences in the interpretation of the final results. To address the issues mentioned above, we applied the landslide density method

to handle land use and lithology. The density values were assigned as new values for each category, which were then normalized according to Eq. (1). To enhance the interpretability of the results, we grouped similar land use categories together, combining woodland and shrubland, as well as water bodies and bare land. The calculation principle of landslide density is shown in Eq. (2), and the results are presented in Table 2.

$$FR = \frac{N_i/N}{S_i/S} \quad (2)$$

where  $N_i$  is the number of landslides in the  $i$ -th class of influencing factor,  $N$  is the number of the landslide in the study area,  $S_i$  is the area of the influencing factor in the  $i$ -th class, and  $S$  is the area of the study area.

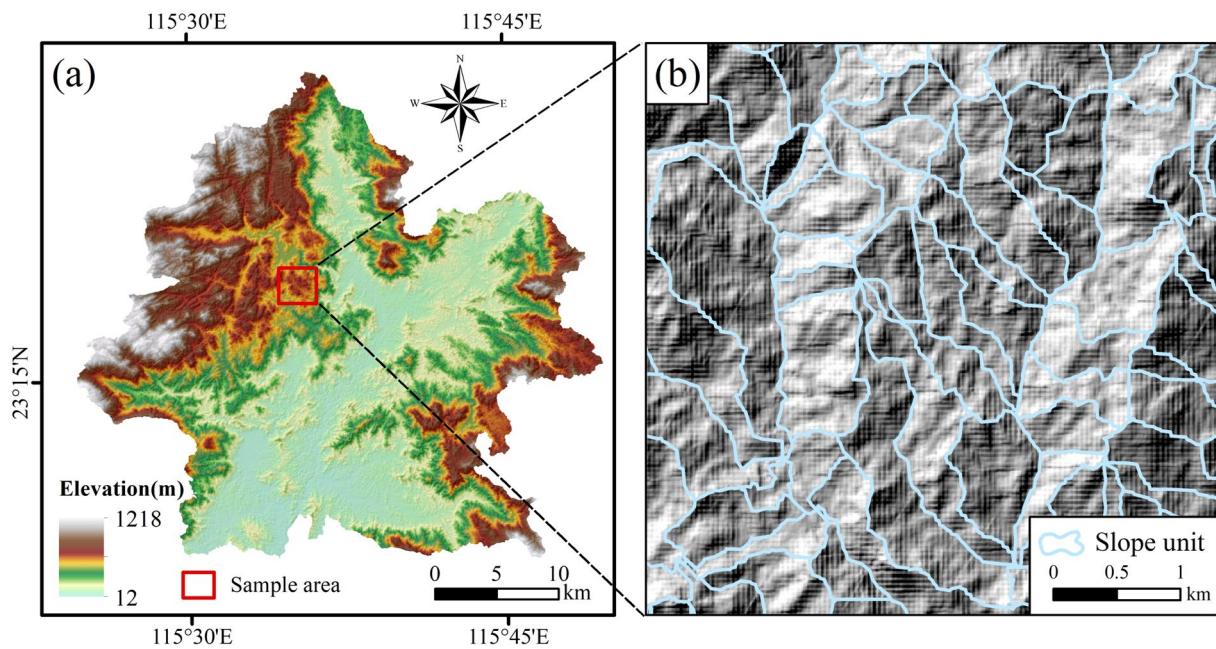
### Model evaluation subunit

Slope units were adopted as the subunits for research. Utilizing the commonly used hydrological analysis method, ridge and valley lines for the study area were obtained through both forward and reverse hydrological analysis operations, and subsequently overlaid. After manual adjustments, the study area was divided into a total of 2575 subunits (Fig. 5).

The adoption of slope units serves not only to enhance the accuracy of landslide susceptibility assessment but, more importantly, to establish spatial adjacency relationships between features. Spatial adjacency represents the spatial distance relationship between two units, and the GWR model regards each subunit within a uniform region, basing its modeling on spatial adjacency (Li et al. 2020). In geological hazard assessment, the boundaries of slope units are the ideal choice for representing this spatial adjacency. In

**Table 2** The calculation results of landslide relative density for categorical variables

| Variable                   | Area (km <sup>2</sup> ) | Landslides (number) | FR value |
|----------------------------|-------------------------|---------------------|----------|
| <b>Land use</b>            |                         |                     |          |
| Artificial surfaces        | 32.563                  | 15                  | 2.735    |
| Arable land                | 171.797                 | 71                  | 2.454    |
| Water bodies and bare land | 5.826                   | 2                   | 2.039    |
| Grassland                  | 19.183                  | 4                   | 1.238    |
| Woodland and shrubland     | 756.320                 | 74                  | 0.581    |
| <b>Lithology</b>           |                         |                     |          |
| Double layer soil          | 4.452                   | 2                   | 2.670    |
| Single layer soil          | 41.760                  | 8                   | 1.139    |
| massive intrusive rock     | 797.290                 | 142                 | 1.059    |
| Stratified clastic rock    | 139.760                 | 14                  | 0.595    |
| Multilayer soil            | 3.412                   | 0                   | 0.000    |



**Fig. 5** Division of slope units in the study area: (a) The entire study area; (b) Schematic diagram of slope unit

comparison to the commonly used administrative boundaries (Ma et al. 2020) and grid boundaries (Sabokbar et al. 2014) in GWR models, slope units better express the homogeneity and heterogeneity of features within subregions, thereby improving the accuracy of GWR model fitting.

### Elimination of multicollinearity

The prerequisite for building a regression model is the independence of all explanatory variables (Zhang et al. 2016b). If there is a strong linear correlation between two or more specific variables, it can lead to unstable parameter estimates in the regression model, causing bias in explaining the meaning and impact of other variables, known as the problem of multicollinearity. To address this issue, two indicators, tolerance (TOL), and Variance Inflation Factor (VIF), are employed to consider the potential multicollinearity in our study. If  $TOL < 0.2$  and  $VIF > 5$ , it indicates the presence of multicollinearity between explanatory variables (Ren et al. 2024).  $VIF > 10$  suggests a relatively severe multicollinearity problem, and the variable should be considered for removal from the model (Akinwande et al. 2015). The formulas for calculating TOL and VIF are as follows:

$$TOL = 1 - R_j^2 \tag{3}$$

$$VIF = \frac{1}{TOL} \tag{4}$$

where  $R_j^2$  is the coefficient of determination indicating the regression value of the  $j$ -th landslide conditioning factor against all other conditioning factors of landslide.

### Spatial autocorrelation

Before applying the MGWR model to analyze the selected causative factors, it is essential to conduct a spatial autocorrelation test of the geospatial parameters (Yu et al. 2020b). This test helps determine whether these factors exhibit varying spatial effects across the entire study area, which, in turn, guides the decision on whether to use a spatial regression model (GWR and MGWR) or a global regression model (OLS) for fitting. The most commonly used spatial variability test is known as the Moran's I test, which reflects the global spatial autocorrelation trends of the explanatory variables. Its calculation formula is as follows (Moran 1950):

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{5}$$

where  $n$  is the number of spatial units;  $\omega_{ij}$  is the weight between location  $i$  and  $j$ ;  $y_i, y_j$  represents the selected attribute value at locations  $i$  and  $j$  respectively; and  $\bar{y}$  is the average value of all observations.

The range of Moran's I varies between -1 and 1. Higher positive values indicate that the attributes of the features are positively correlated in space, which means that nearby observations have similar attribute values, while

observations at longer distances have dissimilar attribute values, leading to spatial clustering. Conversely, negative values signify spatial dispersion, and values close to 0 suggest a random spatial distribution.

### Multiscale geographically weighted regression

Traditional linear regression typically employs the Ordinary Least Squares (OLS) model to quantify the relationship between explanatory variables and the dependent variable. However, as a global homogeneity model, OLS treats the explanatory variables and the dependent variable as having a fixed relationship during fitting, failing to reveal the spatial non-stationarity of factor data in space (Feng et al. 2021). The equation for OLS is as follows:

$$Y = \beta_0 + \sum_{k=1}^m \beta_k X_k + \varepsilon \quad (6)$$

where  $\beta_0$  is a constant parameter;  $\beta_k$  is the respective regression coefficient for explanatory variable  $X_k$ , which is constant over space; and  $\varepsilon$  is the residual error.

GWR extends OLS by weighting spatial dependencies to highlight its characteristics as a local regression model (Su et al. 2017). It includes the geographic coordinates of features in the model, allowing its parameters to vary with spatial location, thereby revealing the spatial variability in the relationship between explanatory variables. The equation for the GWR model is as follows (Lu et al. 2020):

$$Y_i = \beta_0(\mu_i, \nu_i) + \sum_{k=1}^m \beta_k(\mu_i, \nu_i) X_{ik} + \varepsilon_i \quad (7)$$

where  $X_{ik}$  and  $Y_i$  represent the values of the independent and dependent variables at location  $i$ , respectively;  $(\mu_i, \nu_i)$  represents the coordinates of regression analysis point  $i$ , which, in this study, are the projected coordinates at the centroid of each constructed evaluation subunit,  $\beta_0(\mu_i, \nu_i)$  represents the intercept and  $\varepsilon_i$  is the random error;  $\beta_k(\mu_i, \nu_i)$  represents the local regression analysis coefficient.

However, GWR employs a single bandwidth that restricts the local relationships of factors to vary at the same spatial scale. Due to the spatial heterogeneity of each factor, considering the relationship between each factor and landslides at different spatial scales is imperative. Fotheringham et al. (2017) introduced MGWR by employing the back-fitting algorithm (BFA) for model calibration. By setting different bandwidths for each factor, MGWR allows the relationships between explanatory variables and the response variable to vary at different spatial scales. The fundamental formula for MGWR is as follows:

$$Y_i = \beta_0(\mu_i, \nu_i) + \sum_{k=1}^m \beta_{b_{wk}}(\mu_i, \nu_i) X_{ik} + \varepsilon_i \quad (8)$$

where  $b_{wk}$  in  $\beta_{b_{wk}}$  indicates the bandwidth used for calibration of the  $k$ -th conditional relationship.

Similar to the GWR model, the establishment of spatial weights is a crucial step in the MGWR model, which can be divided into two parts: determining the kernel function and optimizing the bandwidth. Based on the fundamental principle of calculating weights in MGWR, which is that the farther the distance, the larger the weight, and vice versa, resulting in a monotonically decreasing function with values ranging between 0 and 1, called the kernel function (Fotheringham et al. 2002; Cho et al. 2010). In order to compare the differences between the MGWR and GWR models, this study uniformly selects the fixed Gaussian kernel function, with its relevant formula as follows (Hong et al. 2017):

$$\omega_{ij} = \exp\left(-\left(d_{ij}/b\right)^2\right) \quad (9)$$

where  $\omega_{ij}$  is the weight for unit  $j$  in the neighborhood of unit  $i$ , and  $d_{ij}$  is the distance between the center point of the unit  $i$  and  $j$  as the measurement of spatial proximity degree.  $b$  is the bandwidth of the Gauss kernel function.

The bandwidth is a crucial control parameter for the GWR model's weight calculation, and its magnitude directly determines the rate at which weights decrease with distance. A bandwidth that is too small can lead to local overfitting, while one that is too large may result in parameter estimates becoming too global (Lu et al. 2020). The bandwidths for GWR and MGWR models were selected uniformly using the corrected Akaike information criteria (AICc). Furthermore, we utilized the MGWR 2.2 software developed by Oshan et al. (2019) to implement the MGWR method, employing the Golden Section method to search for bandwidths. The relevant formula for AICc is as follows (Yu et al. 2020a):

$$AICc = 2n \ln\left(\frac{RSS}{n}\right) + n \ln(2\pi) + n \left[ \frac{n + tr(S)}{n - 2 - tr(S)} \right] \quad (10)$$

where  $n$  denotes the number of observations,  $RSS$  is the residual sum of squares of the model, and  $tr(S)$  denotes the trace of the hat matrix  $S$  as well as the number of effective parameters of the model.

### Model evaluation metrics

In this study, Sigma-square, R-squared, Adjusted R-squared, and the corrected Akaike information criteria (AICc) were chosen to assess the predictive capabilities and complexity of the three models. Sigma-square denotes the variance of model residuals, with lower values indicating higher predictive accuracy. R-squared and Adjusted R-squared reflect the goodness of fit of the model, with higher values indicating a better fit. AICc provides a comprehensive measure of



the goodness of fit and model complexity for GWR model results. For the same dataset and dependent variable in the study, a lower AICc value indicates a better model (Fotheringham et al. 2002), and when the AICc values differ by more than 4, it can be considered as an improvement (Park and Kim 2015).

## Result and discussion

### Global regression and multicollinearity test

In order to mitigate the issue of global collinearity among factors and reduce the risk of model overfitting, we initially employed the OLS method to conduct multiple linear regression analyses for each factor. The TOL and VIF values for each factor were also calculated, as shown in Table 3. All VIF values for the factors were below 3, indicating only mild multicollinearity among them. It is noteworthy that OLS, being a global regression model, provides a single regression coefficient for each factor. The regression results indicate that elevation is the most significant impact factor on landslides, displaying a strong negative correlation. For categorical variables, a class with a high landslide density, which corresponds to a parameter having a higher positive coefficient, was considered to play a greater role in causing landslides (Ayalew and Yamagishi 2005). For example, in the land use category, artificial surfaces have a higher landslide density, while woodland and shrubland have lower landslide densities. If the regression coefficients are greater than 0, it can be inferred that landslides are more likely to occur in artificial surfaces. However, this relationship is global and does not provide insights into their impact on landslide occurrences in specific local regions. The regression coefficient for distance to rivers is 0.007, showing a slight positive correlation at the global level, but it did not pass the significance test at the 95% confidence level ( $p$ -value > 0.05). Considering that subsequent analyses will involve local regression for each factor, distance to rivers has not been excluded in the following stages of the study.

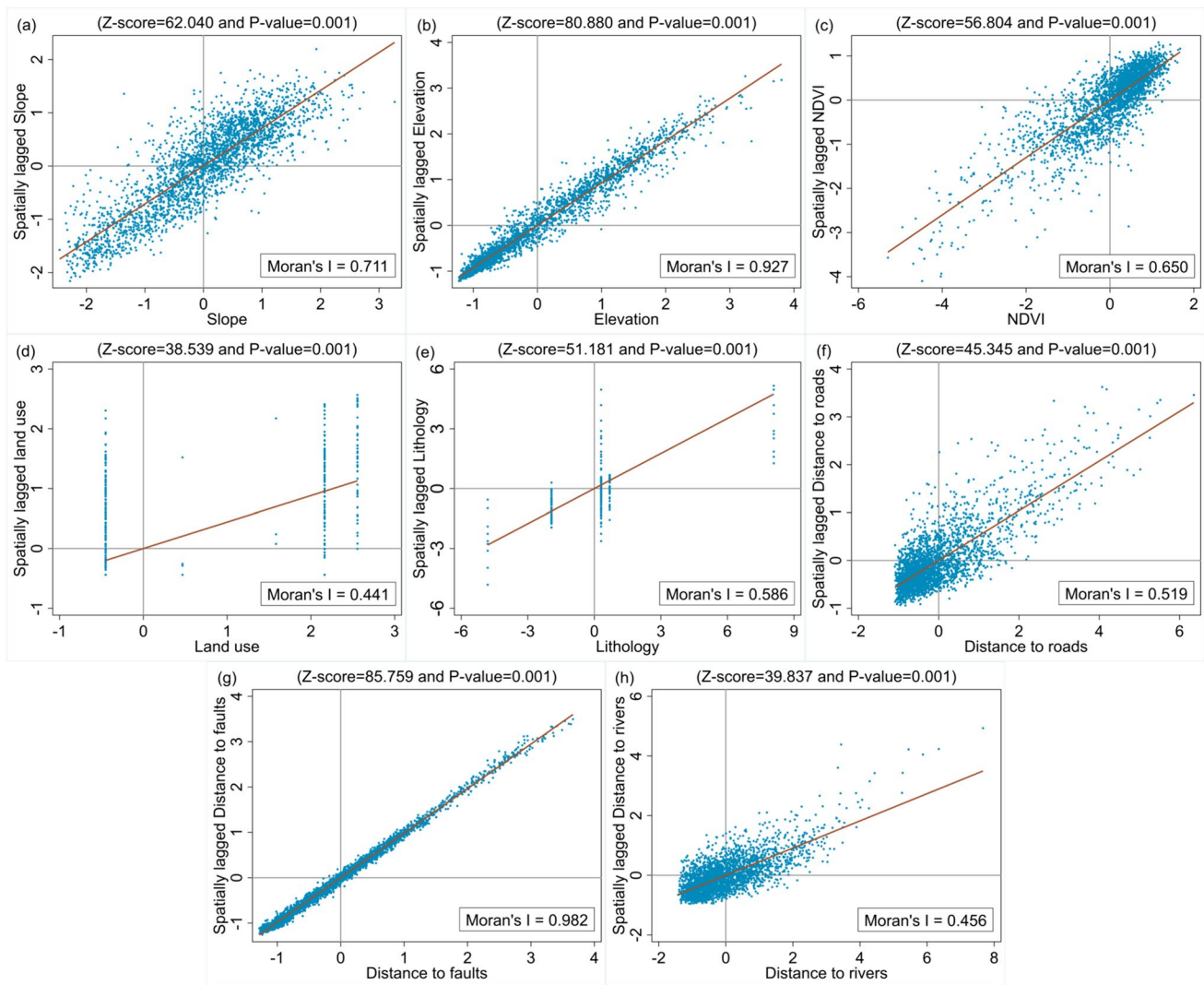
### Spatial autocorrelation test

The spatial autocorrelation of geospatial factors reflects the variations in their own spatial distribution. When explanatory factors exhibit this spatial heterogeneity, it indicates that applying the global regression OLS model may lead to bias. Moreover, the heterogeneity in attribute values of the driving factors across spatial locations forms the foundation for comprehending the spatially varying relationships between these factors and landslides. Based on the previously established slope units, an adjacency spatial weight matrix was created, and the Moran's I test was conducted for each explanatory variable. Moran scatter plots for various factors were plotted with spatial lag distances on the y-axis and standardized factor values on the x-axis, as shown in Fig. 6.

The Z-score is typically used as a significance indicator for the Moran's I statistic to test the null hypothesis (Moran 1950). The first and third quadrants in the Moran scatter plot represent the clustering characteristics of neighboring areas, indicating that high values tend to cluster with high values and low values with low values. All factors exhibited spatial clustering among similar values, with most falling within the first and third quadrants (Fig. 6). And they all passed the significance test at the 95% level, with Moran's I values ranging from 0 to 1, indicating an obviously positive spatial correlation. The spatial heterogeneity of factors further validates the rationality of utilizing local regression models (GWR and MGWR) in the study area. Furthermore, the examination results reveal that each factor's degree of spatial aggregation varies. Among them, elevation and distance to faults have Moran's I values of 0.9 or above, indicating the strongest spatial autocorrelation. On the other hand, land use, distance to rivers and distance to roads have relatively smaller Moran's I values. While they exhibit a positive correlated distribution, they are more spatially dispersed compared to other factors. These results suggest that the relationship between each factor and landslides may be local or tend towards global, indicating that different factors are associated with landslides at different spatial scales. This implies that the adoption of multi-scale GWR models with

**Table 3** Results of OLS model parameter estimation and multicollinearity diagnosis

| Variable           | Coefficient | Std Error | t-value | p-value | TOL   | VIF   |
|--------------------|-------------|-----------|---------|---------|-------|-------|
| Slope              | 0.084       | 0.031     | 2.696   | 0.007   | 0.368 | 2.720 |
| Elevation          | -0.273      | 0.025     | -10.986 | 0.000   | 0.570 | 1.754 |
| NDVI               | -0.007      | 0.028     | -0.268  | 0.009   | 0.461 | 2.167 |
| Land use           | 0.019       | 0.025     | -0.733  | 0.004   | 0.549 | 1.823 |
| Lithology          | 0.120       | 0.019     | 6.302   | 0.000   | 0.977 | 1.024 |
| Distance to roads  | -0.125      | 0.020     | -6.298  | 0.000   | 0.898 | 1.114 |
| Distance to faults | -0.037      | 0.020     | -1.862  | 0.003   | 0.912 | 1.097 |
| Distance to rivers | 0.007       | 0.020     | 0.368   | 0.713   | 0.850 | 1.176 |



**Fig. 6** Moran scatterplot of explanatory factors: (a) slope; (b) elevation; (c) NDVI; (d) land use; (e) lithology; (f) distance to roads; (g) distance to faults; (h) distance to rivers

different bandwidths may better consider the impact scale of various factors on landslides.

### Model results and comparison

Based on the constructed slope units, GWR and MGWR regression analyses were conducted separately for the selected eight factors using a fixed Gaussian kernel function and the AICc method. For the 2575 subunits, both models generated the corresponding factor regression coefficients. Tables 4 and 5 present the parameter estimation results for GWR and MGWR, respectively.

The traditional GWR model employs the same bandwidth for all driving factors (the optimal bandwidth adopted for the GWR model is 819.18m in this study). In contrast, the MGWR model searches for specific bandwidths for variables

at different spatial scales (Table 6), resulting in differences in the final regression parameters between the two models. Specifically, there are significant differences in the ranges of regression coefficients for each factor obtained by the two models (Tables 4 and 5). The GWR model yields a positive mean NDVI regression coefficient, whereas the MGWR result is precisely the opposite. The MGWR model suggests that, overall, the probability of landslide occurrence increases as NDVI decreases. Furthermore, in the GWR model, the regression coefficient for elevation has both positive and negative values, but in the MGWR model, the coefficient is consistently negative. The MGWR model posits that elevation maintains a consistently negative correlation with landslides throughout the study area. The underlying reason for these differences is the varying degree of spatial clustering or dispersion of each factor, indicating different

**Table 4** Parameter estimates of GWR model

| Variable           | Coefficient |       |       |        |        |       |
|--------------------|-------------|-------|-------|--------|--------|-------|
|                    | Min         | Max   | Range | Median | Mean   | STD   |
| Slope              | -0.330      | 0.368 | 0.698 | 0.034  | 0.033  | 0.139 |
| Elevation          | -1.395      | 0.178 | 1.573 | -0.373 | -0.369 | 0.316 |
| NDVI               | -0.536      | 0.399 | 0.935 | -0.018 | 0.023  | 0.178 |
| Land use           | -0.265      | 0.550 | 0.814 | -0.004 | 0.024  | 0.153 |
| Lithology          | -0.081      | 0.399 | 0.481 | 0.068  | 0.076  | 0.100 |
| Distance to roads  | -0.472      | 0.306 | 0.778 | -0.037 | -0.082 | 0.150 |
| Distance to faults | -1.051      | 0.634 | 1.685 | -0.032 | -0.062 | 0.294 |
| Distance to rivers | -0.167      | 0.278 | 0.445 | 0.007  | 0.041  | 0.104 |

**Table 5** Parameter estimates of MGWR model

| Variable           | Coefficient |        |       |        |        |       |
|--------------------|-------------|--------|-------|--------|--------|-------|
|                    | Min         | Max    | Range | Median | Mean   | STD   |
| Slope              | -0.605      | 0.493  | 1.098 | 0.047  | 0.038  | 0.118 |
| Elevation          | -0.687      | -0.029 | 0.659 | -0.277 | -0.296 | 0.106 |
| NDVI               | -0.619      | 0.266  | 0.884 | -0.051 | -0.063 | 0.162 |
| Land use           | -0.325      | 0.405  | 0.730 | 0.010  | 0.009  | 0.087 |
| Lithology          | -0.947      | 1.323  | 2.269 | 0.097  | 0.125  | 0.207 |
| Distance to roads  | -0.705      | 0.152  | 0.857 | -0.037 | -0.063 | 0.105 |
| Distance to faults | -0.813      | 0.535  | 1.348 | -0.035 | -0.058 | 0.175 |
| Distance to rivers | -0.163      | 0.295  | 0.458 | 0.023  | 0.036  | 0.072 |

**Table 6** Model bandwidth for MGWR

| Variable  | Bandwidth (m) | Variable           | Bandwidth (m) |
|-----------|---------------|--------------------|---------------|
| Slope     | 620.43        | Lithology          | 866.50        |
| Elevation | 501.74        | Distance to roads  | 920.43        |
| NDVI      | 698.98        | Distance to faults | 479.14        |
| Land use  | 1100.66       | Distance to rivers | 1053.23       |

**Table 7** Model diagnostics of OLS, GWR and MGWR

| Diagnostic criteria | OLS      | GWR      | MGWR     |
|---------------------|----------|----------|----------|
| Sigma-square        | 0.909    | 0.356    | 0.137    |
| R-square            | 0.094    | 0.659    | 0.869    |
| Adjusted R-square   | 0.091    | 0.644    | 0.863    |
| AICc                | 7072.151 | 4700.969 | 2314.983 |

spatial autocorrelation strengths. MGWR takes into account the different spatial scales inherent in each factor, leading to deviations in its regression results from the GWR model. Specifically, the MGWR model searches for the smallest bandwidth for the distance to faults factor, with a value of 479.14m, indicating the strongest spatial heterogeneity of the distance to faults. This result corresponds to the earlier calculation, where the distance to faults factor had the highest Moran's I value (Fig. 5). In the study area, the relationship between the distance to faults and landslides exists at a relatively small spatial scale. In contrast, the bandwidth values for the land use, distance to rivers, and distance to roads factors are relatively larger, indicating less spatial heterogeneity, and their relationship with landslides tends to be more global in spatial scale.

Additional diagnostic results for the OLS, GWR, and MGWR models are presented in Table 7. It can be observed that the overall fitting performance of the GWR and MGWR

models, utilizing local regression, surpasses that of the global regression OLS model. They exhibit higher R-square values and lower AICc values. The MGWR model, considering different spatial scales, demonstrates the best performance. Compared to the GWR model, the MGWR model exhibits an increase of 0.210 in R-square and a decrease of 2385.986 in AICc. This indicates that the MGWR model, by considering the spatial scale diversity of driving factors, reduces noise and collinearity issues during the regression process, enhancing the model's robustness. Furthermore, the MGWR model, by seeking different optimal bandwidths and efficiently calculating spatial weight values through kernel functions, more effectively reveals the spatial influence scale of each variable.

The spatial distribution of standardized residuals from the OLS, GWR, and MGWR models (Fig. 7) reveals that the regression residuals of the MGWR model are generally closer to 0 and exhibit more randomness in space.

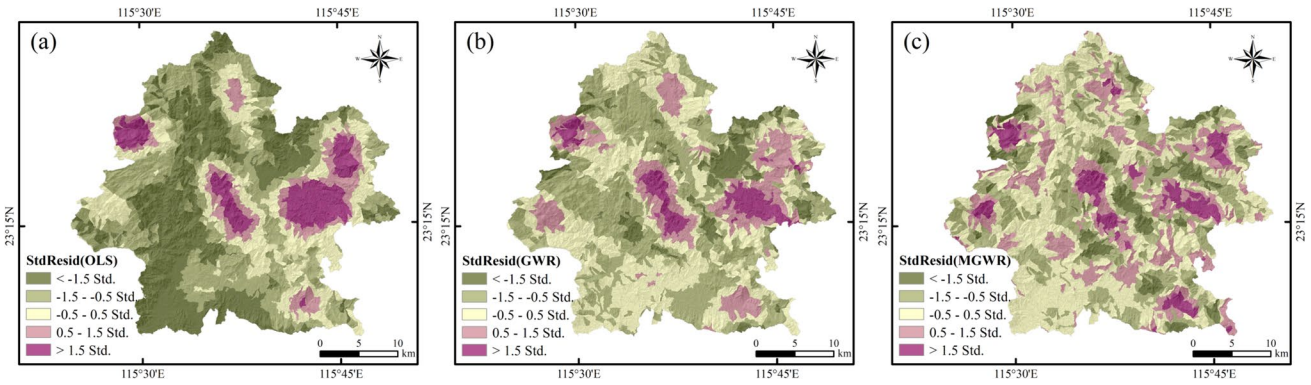


Fig. 7 Spatial distribution of standardized residuals: (a) OLS model; (b) GWR model; (c) MGWR model

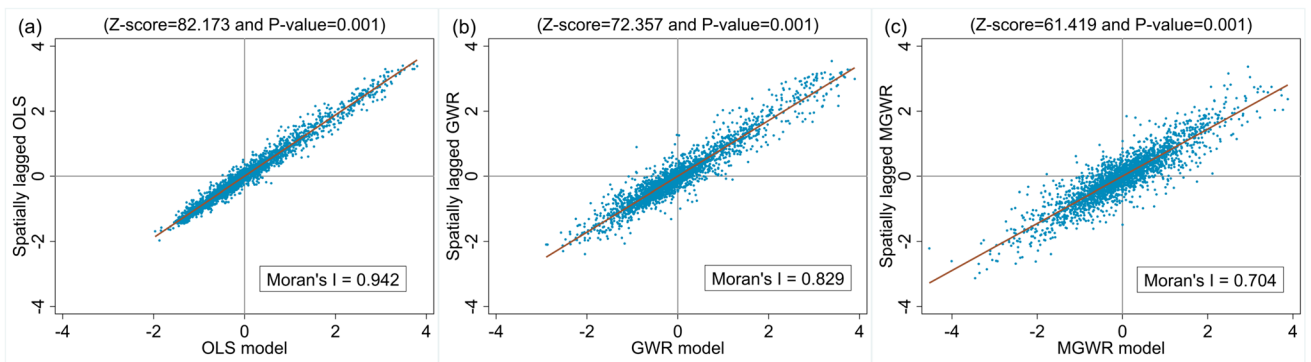


Fig. 8 Moran scatterplot of standardized residuals: (a) OLS model; (b) GWR model; (c) MGWR model

Further analysis involves plotting Moran scatter plots for the standardized residuals of the three models (Fig. 8). It is evident that, compared to the global regression OLS model, both GWR and MGWR exhibit smaller Moran's I value and greater spatial dispersion, indicating that local regression methods considering the spatial non-stationarity of factors are more suitable for a broader range of study areas. It is worth emphasizing that the MGWR model, which further considers the influence of spatial scale, has the smallest Moran's I values. This suggests that the MGWR model can comprehensively consider the spatial variations of various factors affecting landslides, contributing to a more accurate understanding of the spatial patterns of landslide occurrences. In summary, the MGWR model exhibits the best fitting performance and the most uniform distribution of regression residuals. Therefore, we choose the regression results of the MGWR model to explore the spatial variations in the impact of different factors on landslides in subsequent studies (Table 5).

Table 8 Proportion statistics of dominant driving factors

| Dominant factors   | Slope unit (number) | Area (km <sup>2</sup> ) | Areal percentage (%) |
|--------------------|---------------------|-------------------------|----------------------|
| Elevation          | 1793                | 685.98                  | 69.52                |
| Lithology          | 313                 | 121.46                  | 12.31                |
| Distance to faults | 278                 | 104.38                  | 10.58                |
| NDVI               | 146                 | 55.03                   | 5.58                 |
| Slope              | 24                  | 11.32                   | 1.15                 |
| Distance to roads  | 21                  | 8.50                    | 0.86                 |
| SUM                | 2575                | 986.67                  | 100                  |

### Spatial non-stationarity of driving factors

Another advantage of spatial local regression models is their ability to visually express quantitative parameter estimates in the form of maps, illustrating the spatial variation in the intensity of the impact of driving factors on landslides.

The positive and negative signs of regression coefficients correspond to the positive or negative correlation between factors and landslides, while the magnitude of the absolute value represents the strength of the influence (Chen et al. 2022). Conducting a census of regression coefficients of the MGWR model across all slope units, the regression outcomes for each factor are stratified using 0 as the threshold. The first 30% of positive values are classified as high positive, while the remaining 70% are labeled as sub positive. Likewise, the first 30% of the minimum negative values are considered high negative, and the subsequent values are categorized as sub negative (Fig. 9). For elevation, which only has negative regression coefficients, it is also classified into four levels using the quantile method.

The slope is typically confirmed as the most significant factor influencing landslide occurrence in many studies (Mind’je et al. 2019; Polykretis et al. 2021). It affects the likelihood of slope movement by influencing potential energy changes, crack development, and water infiltration. The MGWR model yielded relatively high positive values in Nanwan and Shuichun (Fig. 9a). In this region, slope is positively correlated with landslides, indicating that an increase in slope promotes landslide occurrences, and slope remains an important influencing factor. Conversely, highly negatively correlated areas are concentrated in the southern

part of Dongkeng. Overall, the impact of slope on landslides is highly complex, exhibiting a multipolar pattern in space.

The MGWR model shows that the regression coefficients for elevation are all negative (Table 5), and all three models suggest that landslides are more likely to occur in the low-altitude areas of the study region (Tables 3 and 4). According to statistics, the average elevation of 166 landslide points is only 200.87 m, which is below the county average of 314 m in Luhe. As shown in Fig. 1b, landslides tend to be distributed in low mountainous areas adjacent to regions of human activity, a phenomenon that is similar to findings in most landslide-related studies in southeastern coastal areas of China (Zhang et al. 2016a; Liang et al. 2022; Yu et al. 2023). This suggests that human factors, such as engineering activities and land use changes, may be the causes of local disasters. It is worth noting that global regression models may be affected by the concentration bias in landslide data. If the majority of landslide points are concentrated in low-altitude areas, the global regression model will naturally yield stronger negative regression results. However, the local regression model addresses this issue to some extent by analyzing within a smaller area. When the local regression model yields a negative value in a low-altitude region, it can be interpreted as indicating that even though the region is at a low altitude, a further decrease in altitude within that area still contributes to the occurrence of landslides.

**Fig. 9** The spatial variation of local parameter estimates of each factor: (a) slope; (b) elevation; (c) NDVI; (d) land use; (e) lithology; (f) distance to roads; (g) distance to faults; (h) distance to rivers

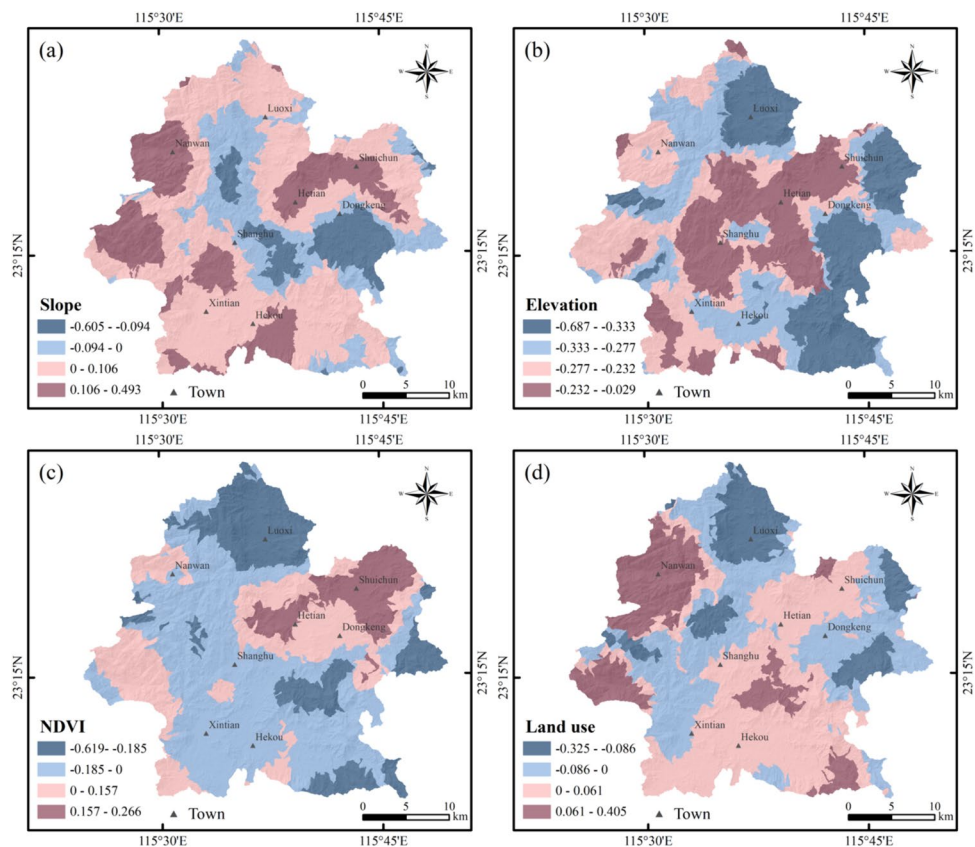
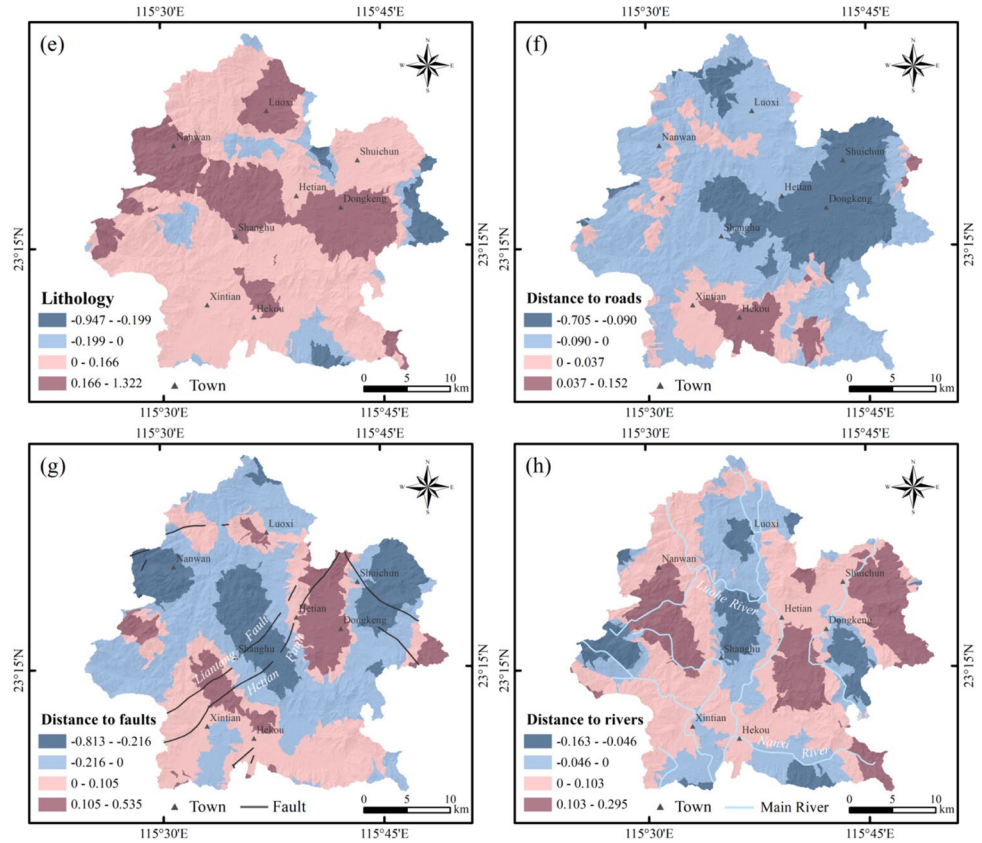


Fig. 9 (continued)



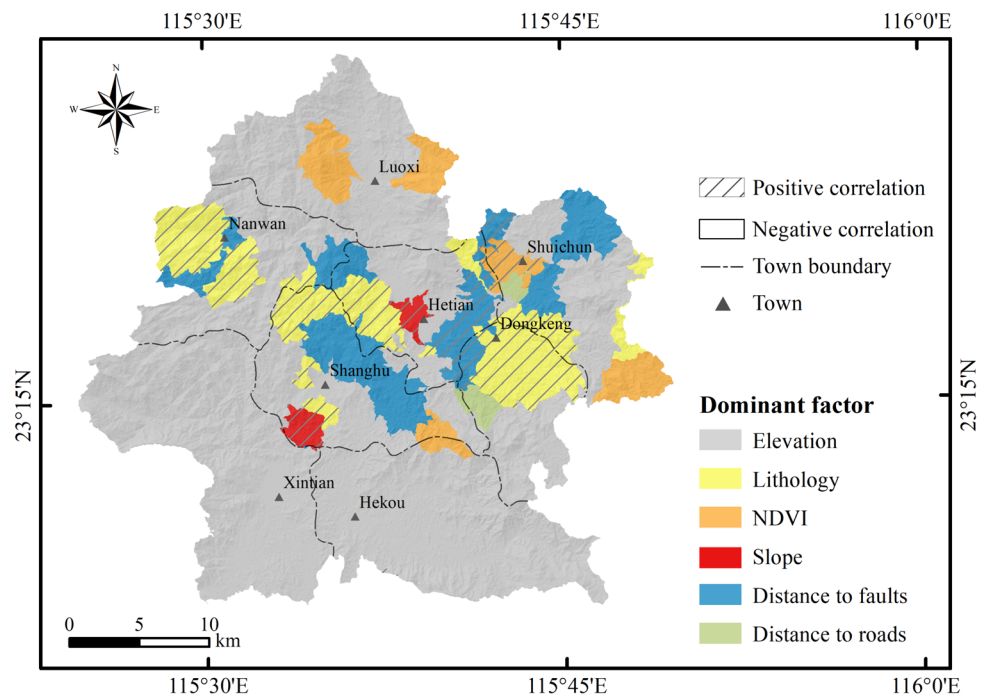
The regression coefficients for NDVI are mostly negative, indicating that overall, an increase in NDVI has a suppressing effect on landslide occurrences. This aligns with the findings of many scholars (Qin et al. 2021; Zhang et al. 2023). However, in the northeastern part of the study area near Shunchun, relatively high positive regression coefficients are observed, indicating that an increase in NDVI has a positive impact on landslides in that region. Zhu Liang et al. (2023) reached similar conclusions in their study of landslides in Tibet, suggesting that while high vegetation cover can enhance slope stability, its relationship with landslides can be unstable. Preliminary analysis suggests that in this area, the rock formation consists of massive intrusive rock, thus excessive vegetation cover may lead to root splitting in the bedrock, which outweighs its stabilizing effect and results in instability.

High positive regression coefficients for land use are observed near Nanwan and show a positive correlation in most regions of the central and southern parts of the study area. Analyzing this in conjunction with landslide density (Table 2), it indicates that artificial surfaces and arable land, which have high landslide density values, contribute to the occurrence of landslides. Land use indirectly reflects human activities, and this result further confirms that human-induced disturbances to the surface alter the distribution of ground stress, significantly impacting slope

stability. The regression coefficients for lithology are mostly greater than 0, particularly in areas like Nanwan, Hetian, and Dongkeng, where special attention should be paid to the development of double layer soil. Within the study area, a total of 142 landslide occurrences are distributed in massive intrusive rock (Table 2). This may be due to the presence of well-developed gullies, significant terrain relief, and the development of weathered fractures in shallow rock layers.

The distance to roads, faults, and rivers is often considered in landslide susceptibility assessments. Zhang et al. (2019) discovered in their research in the Pearl River Delta that areas with higher water density and closer proximity to faults are more prone to landslides. Wang et al. (2023) also demonstrated in their study of the Yunnan region that landslides are more likely to be distributed near roads, faults, and water bodies. In our study, the MGWR model showed negative regression results for the distance to roads and faults, also indicating that landslides are more likely to occur when they are closer to them. High negative coefficients for the distance to roads are concentrated in Dongkeng and Shuicun (Fig. 9h), indicating a significant damaging effect on slopes due to road construction in this area. The areas with high negative coefficients for the distance to faults and rivers are concentrated around the main distribution belts of the Luohe River system, Liantang

**Fig. 10** Dominant driving factors of the study area



fault, and Hetian fault in the central part of the study area. This indicates that faults and rivers have a positive effect on landslides in this region.

**Spatial pattern analysis of the dominant factors**

The driving factor with the largest absolute regression coefficient in each slope unit is considered the dominant factor in that subunit. Among the 2575 slope units, six factors: elevation, slope, NDVI, lithology, distance to faults and distance to roads, have occupied dominant positions (Fig. 10). Overall, there exists a significant spatial correlation between landslide occurrences and their driving factors. With changes in spatial location, different dominant factors emerge within each region, and the positive or negative influence of these dominant factors also changes accordingly.

The proportion of correlation for each factor is presented in Table 8. It is evident that elevation is the primary controlling factor for landslide disasters in Luhe. Within the study area, relevant departments should pay special attention to the impact of human activities on slopes in low mountainous areas. The secondary dominant factor is lithology, which still indicates that double layer soil, single layer soil, and massive intrusive rock promote the occurrence of landslides. Distance to faults demonstrates dominance near the Liantang and Hetian fault zones (Fig. 9g) (Fig. 10). This area may require stricter monitoring and early warning systems, as well as more robust building structures. The slope always shows a positive correlation with the occurrence of landslides when it is dominant. Slope cutting and stabilization

work in high-slope areas within the region should not be neglected. Regions dominated by NDVI are relatively scattered, but most still show a negative correlation. Vegetation restoration can be carried out in these areas to stabilize the soil and reduce surface water runoff. However, in the eastern part of Shuichun, attention should also be paid to the development of unstable slopes with high vegetation cover. The proportion of areas dominated by distance to roads is the smallest. In landslide prevention and control efforts, particular attention should be given to the influence of roads on landslides in this area. It is essential to plan the layout of roads rationally and strengthen slope protection and drainage facilities along roadsides. However, it must be pointed out that this result has a certain degree of specificity because MGWR determines the range of local regression through bandwidth, and there may be some fitting errors in the edge areas due to the lack of data.

**Conclusion**

To address the issue of spatial non-stationarity in previous landslide susceptibility assessments, this study establishes slope units and introduces the MGWR model to investigate the spatial variations in the relationship between landslides and their driving factors in Luhe, Guangdong. The results of the Moran's I test indicate that the Moran's I values for all factors in the study area are above 0.4, demonstrating significant spatial heterogeneity and confirming the necessity of using spatial local regression models. Compared to the global

regression OLS model and the single-bandwidth GWR model, the MGWR model, considering different spatial scales, exhibits the highest R squared (0.869) and the lowest AICc values (2314.983), demonstrating the best fitting performance and overall model performance. Additionally, the residual distribution of the MGWR model is more spatially dispersed, indicating its applicability to a wider range of areas. The analysis of the spatial pattern for the driving factors indicates that there is significant spatial non-stationarity in the relationship between landslides and their driving factors. Elevation is the primary dominant factor contributing to landslides in the study area, and it primarily exhibits a negative correlation with landslides when holding dominant positions.

By employing the Multiscale Geographically Weighted Regression (MGWR) model with multiple spatial scales, one can better analyze the interaction patterns between factors and landslides, providing enhanced insights for disaster prevention and mitigation. However, the study still has certain limitations. Firstly, due to the incomplete data collection, the study might overlook more critical landslide driving factors, leading to a somewhat one-sided analysis of the dominant factors in the study area. Furthermore, the uniformity of slope units constructed using hydrological analysis is relatively poor, leading to spatial estimation errors in the construction of the spatial weight matrix and influencing the final fit of the model. Finally, MGWR, essentially a linear regression model, may not fully account for the nonlinear relationships between landslides and factors. Moreover, the inherent complexity of the MGWR model prompts us to explore its applicability in different regions in future studies.

**Acknowledgements** This work was supported by the Project supported by Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. 311022004), the National Comprehensive Risk Survey of Natural Disasters: 1:50000 Geological Disaster Risk Survey and Evaluation (Luhe and Haifeng) (No. 441501-2021-01489), the Guangdong Basic and Applied Basic Research Foundation (2019A1515010733), and the Science and Technology Program of Guangzhou, China, under Grant (No. 201707010209).

**Funding** Project supported by Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), 311022004, Guifang Zhang, National Comprehensive Risk Survey of Natural Disasters: 1:50000 Geological Disaster Risk Survey and Evaluation (Luhe and Haifeng), 441501-2021-01489, Guifang Zhang, Guangdong Basic and Applied Basic Research Foundation, 2019A1515010733, Guifang Zhang, Science and Technology Program of Guangzhou, China, under Grant, 201707010209, Guifang Zhang

**Data availability** Data will be made available on request.

## Declarations

**Competing Interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Akinwande MO, Dikko HG, Samson A (2015) Variance inflation factor: as a condition for the inclusion of suppressor variable (s) in regression analysis. *Open J Stat* 5:754. <https://doi.org/10.4236/ojs.2015.57075>
- Anselin L, Griffith DA (1988) Do spatial effects really matter in regression analysis? *Pap Reg Sci* 65:11–34
- Arabameri A, Pradhan B, Rezaei K, Lee C-W (2019) Assessment of landslide susceptibility using statistical-and artificial intelligence-based FR–RF integrated model and multiresolution DEMs. *Remote Sens* 11:999. <https://doi.org/10.3390/rs11090999>
- Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65:15–31. <https://doi.org/10.1016/j.geomorph.2004.06.010>
- Boussouf S, Fernandez T, Hart AB (2023) Landslide susceptibility mapping using maximum entropy (MaxEnt) and geographically weighted logistic regression (GWLRL) models in the Rio Aguas catchment (Almeria, SE Spain). *Nat Hazards* 117:207–235. <https://doi.org/10.1007/s11069-023-05857-7>
- Budimir M, Atkinson P, Lewis H (2015) A systematic review of landslide probability mapping using logistic regression. *Landslides* 12:419–436. <https://doi.org/10.1007/s10346-014-0550-5>
- Chalkias C, Polykretis C, Karymbalis E, Soldati M, Ghinoi A, Ferentinou M (2020) Exploring spatial non-stationarity in the relationships between landslide susceptibility and conditioning factors: a local modeling approach using geographically weighted regression. *Bull Eng Geol Env* 79:2799–2814. <https://doi.org/10.1007/s10064-020-01733-x>
- Chen L, Zhang H, Zhang X, Liu P, Zhang W, Ma X (2022) Vegetation changes in coal mining areas: naturally or anthropogenically driven? *Catena*, 208. <https://doi.org/10.1016/j.catena.2021.105712>
- Cho S-H, Lambert DM, Chen Z (2010) Geographically weighted regression bandwidth selection and spatial autocorrelation: an empirical example using Chinese agriculture data. *Appl Econ Lett* 17:767–772. <https://doi.org/10.1080/13504850802314452>
- Erener A, Duzgun HSB (2010) Improvement of statistical landslide susceptibility mapping by using spatial and global regression methods in the case of More and Romsdal (Norway). *Landslides* 7:55–68. <https://doi.org/10.1007/s10346-009-0188-x>
- Feng L, Wang Y, Zhang Z, Du Q (2021) Geographically and temporally weighted neural network for winter wheat yield prediction. *Remote Sens Environ* 262. <https://doi.org/10.1016/j.rse.2021.112514>
- Feuillet T, Coquin J, Mercier D, Cossart E, Decaulne A, Jonsson HP, Saemundsson P (2014) Focusing on the spatial non-stationarity of landslide predisposing factors in northern Iceland: Do paraglacial factors vary over space? *Prog Phys Geograp-Earth Environ* 38:354–377. <https://doi.org/10.1177/0309133314528944>
- Fotheringham AS, Yang W, Kang W (2017) Multiscale Geographically Weighted Regression (MGWR). *Ann Am Assoc Geogr* 107:1247–1265. <https://doi.org/10.1080/24694452.2017.1352480>
- Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons
- Gu T, Li J, Wang M, Duan P (2022) Landslide susceptibility assessment in Zhenxiang County of China based on geographically weighted logistic regression model. *Geocarto Int* 37:4952–4973. <https://doi.org/10.1080/10106049.2021.1903571>
- Guha-Sapir D, Below R, Hoyois P (2020) EM-DAT: the CRED/OFDA international disaster database. Centre for Research on the Epidemiology of Disasters (CRED), Université Catholique de Louvain, Brussels



- Han Z, Su B, Li Y, Ma Y, Wang W, Chen G (2019) Comprehensive analysis of landslide stability and related countermeasures: a case study of the Lanmuxi landslide in China. *Sci Rep* 9:12407. <https://doi.org/10.1038/s41598-019-48934-3>
- Hong H, Pradhan B, Sameen MI, Chen W, Xu C (2017) Spatial prediction of rotational landslide using geographically weighted regression, logistic regression, and support vector machine models in Xing Guo area (China). *Geomat Nat Haz Risk* 8:1997–2022. <https://doi.org/10.1080/19475705.2017.1403974>
- Huang F, Tao S, Chang Z, Huang J, Fan X, Jiang S-H, Li W (2021) Efficient and automatic extraction of slope units based on multi-scale segmentation method for landslide assessments. *Landslides* 18:3715–3731. <https://doi.org/10.1007/s10346-021-01756-9>
- Huang F, Zhang J, Zhou C, Wang Y, Huang J, Zhu L (2020) A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. *Landslides* 17:217–229. <https://doi.org/10.1007/s10346-019-01274-9>
- Huang Y, Zhao L (2018) Review on landslide susceptibility mapping using support vector machines. *CATENA* 165:520–529. <https://doi.org/10.1016/j.catena.2018.03.003>
- Kaur R, Gupta V, Chaudhary B (2024) Landslide susceptibility mapping and sensitivity analysis using various machine learning models: a case study of Beas valley, Indian Himalaya. *Bull Eng Geol Env* 83:228. <https://doi.org/10.1007/s10064-024-03712-y>
- Li Y, Liu X, Han Z, Dou J (2020) Spatial proximity-based geographically weighted regression model for landslide susceptibility assessment: a case study of Qingchuan area China. *Appl Sci* 10:1107. <https://doi.org/10.3390/app10031107>
- Liang X, Segoni S, Yin K, Du J, Chai B, Tofani V, Casagli N (2022) Characteristics of landslides and debris flows triggered by extreme rainfall in Daoshi Town during the 2019 Typhoon Lekima, Zhejiang Province, China. *Landslides* 19:1735–1749. <https://doi.org/10.1007/s10346-022-01889-5>
- Liang Z, Peng W, Liu W, Huang H, Huang J, Lou K, Liu G, Jiang K (2023) Exploration and comparison of the effect of conventional and advanced modeling algorithms on landslide susceptibility prediction: a case study from Yadong Country Tibet. *Appl Sci* 13:7276. <https://doi.org/10.3390/app13127276>
- Liu R, Li L, Pirasteh S, Lai Z, Yang X, Shahabi H (2021) The performance quality of LR, SVM, and RF for earthquake-induced landslides susceptibility mapping incorporating remote sensing imagery. *Arab J Geosci* 14:1–15. <https://doi.org/10.1007/s12517-021-06573-x>
- Liu R, Ding Y, Sun D, Wen H, Gu Q, Shi S, Liao M (2023) Insights into spatial differential characteristics of landslide susceptibility from sub-region to whole-region cased by northeast Chongqing, China. *Geomat Nat Haz Risk* 14:2190858. <https://doi.org/10.1080/19475705.2023.2190858>
- Lu B, Ge Y, Qin K, Zheng J (2020) A review on geographically weighted regression. *Geom Inform Sci Wuhan Univ* 45:1356–1366. <https://doi.org/10.13203/j.whugis20190346>
- Ma X, Ji Y, Yuan Y, Van Oort N, Jin Y, Hoogendoorn S (2020) A comparison in travel patterns and determinants of user demand between docked and dockless bike-sharing systems using multi-sourced data. *Trans Res Part a: Policy Pract* 139:148–173. <https://doi.org/10.1016/j.tra.2020.06.022>
- Mind'je R, Li L, Nsengiyumva JB, Mupenzi C, Nyesheja EM, Kayumba PM, Gasirabo A, Hakorimana E (2019) Landslide susceptibility and influencing factors analysis in Rwanda. *Environ Dev Sustain* 22:7985–8012. <https://doi.org/10.1007/s10668-019-00557-4>
- Moran PA (1950) Notes on continuous stochastic phenomena. *Biometrika* 37:17–23. <https://doi.org/10.2307/2332142>
- Ng CWW, Yang B, Liu ZQ, Kwan JSH, Chen L (2021) Spatiotemporal modelling of rainfall-induced landslides using machine learning. *Landslides* 18:2499–2514. <https://doi.org/10.1007/s10346-021-01662-0>
- Nhu V-H, Shirzadi A, Shahabi H, Singh SK, Al-Ansari N, Clague JJ, Jaafari A, Chen W, Miraki S, Dou J (2020) Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms. *Int J Environ Res Public Health* 17:2749. <https://doi.org/10.3390/ijerph17082749>
- Ohlmacher GC, Davis JC (2003) Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Eng Geol* 69:331–343. [https://doi.org/10.1016/s0013-7952\(03\)00069-3](https://doi.org/10.1016/s0013-7952(03)00069-3)
- Oshan TM, Ziqi L, Wei K, Wolf LJ, Fotheringham AS (2019) MGWR: a python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information* 8:269. <https://doi.org/10.3390/ijgi8060269>
- Panchal S, Shrivastava AK (2022) Landslide hazard assessment using analytic hierarchy process (AHP): a case study of National Highway 5 in India. *Ain Shams Eng J* 13:101626. <https://doi.org/10.1016/j.asej.2021.10.021>
- Park S, Kim J (2015) A comparative analysis of landslide susceptibility assessment by using global and spatial regression methods in Inje Area, Korea. *J Korean Soc Surv Geod Photogramm Cartogr* 33:579–587. <https://doi.org/10.7848/ksgep.2015.33.6.579>
- Petley D (2012) Global patterns of loss of life from landslides. *Geology* 40:927–930. <https://doi.org/10.1130/G33217.1>
- Polykretis C, Grillakis MG, Argyriou AV, Papadopoulos N, Alexakis DD (2021) Integrating multivariate (GeoDetector) and bivariate (IV) statistics for hybrid landslide susceptibility modeling: a case of the vicinity of Pinios artificial lake, Ilia, Greece *Land* 10:973. <https://doi.org/10.3390/land10090973>
- Popescu ME (2002) Landslide causal factors and landslide remedial options. 3rd international conference on landslides, slope stability and safety of infra-structures. Citeseer, pp 61–81
- Pradhan AMS, Lee S-R, Kim Y-T (2019) A shallow slide prediction model combining rainfall threshold warnings and shallow slide susceptibility in Busan, Korea. *Landslides* 16:647–659. <https://doi.org/10.1007/s10346-018-1112-z>
- Qin Y, Yang G, Lu K, Sun Q, Xie J, Wu Y (2021) Performance evaluation of five GIS-based models for landslide susceptibility prediction and mapping: a case study of Kaiyang County China. *Sustainability* 13:6441. <https://doi.org/10.3390/su13116441>
- Ren T, Gao L, Gong W (2024) An ensemble of dynamic rainfall index and machine learning method for spatiotemporal landslide susceptibility modeling. *Landslides* 21:257–273. <https://doi.org/10.1007/s10346-023-02152-1>
- Sabokbar HF, Roodposhti MS, Tazik E (2014) Landslide susceptibility mapping using geographically-weighted principal component analysis. *Geomorphology* 226:15–24. <https://doi.org/10.1016/j.geomorph.2014.07.026>
- Segoni S, Tofani V, Rosi A, Catani F, Casagli N (2018) Combination of rainfall thresholds and susceptibility maps for dynamic landslide hazard assessment at regional scale. *Frontiers in Earth Science* 6. <https://doi.org/10.3389/feart.2018.00085>
- Song X, Mi N, Mi W, Li L (2022) Spatial non-stationary characteristics between grass yield and its influencing factors in the Ningxia temperate grasslands based on a mixed geographically weighted regression model. *J Geog Sci* 32:1076–1102. <https://doi.org/10.1007/s11442-022-1986-5>
- Su S, Gong Y, Tan B, Pi J, Weng M, Cai Z (2017) Area social deprivation and public health: Analyzing the spatial non-stationary associations using geographically weighed regression. *Soc Indic Res* 133:819–832. <https://doi.org/10.1007/s11205-016-1390-6>
- Sun D, Shi S, Wen H, Xu J, Zhou X, Wu J (2021) A hybrid optimization method of factor screening predicated on geodetector and random forest for landslide susceptibility mapping. *Geomorphology* 379:107623. <https://doi.org/10.1016/j.geomorph.2021.107623>

- Wang Q, Guo Y, Li W, He J, Wu Z (2019) Predictive modeling of landslide hazards in Wen County, northwestern China based on information value, weights-of-evidence, and certainty factor. *Geomat Nat Haz Risk* 10:820–835. <https://doi.org/10.1080/19475705.2018.1549111>
- Wang Y, Sun D, Wen H, Zhang H, Zhang F (2020) Comparison of random forest model and frequency ratio model for landslide susceptibility mapping (LSM) in Yunyang County (Chongqing, China). *Int J Environ Res Public Health* 17:4206. <https://doi.org/10.3390/ijerph17124206>
- Wang H, Xu J, Tan S, Zhou J (2023) Landslide susceptibility evaluation based on a coupled informative-logistic regression model—Shuangbai County as an example. *Sustainability* 15:12449. <https://doi.org/10.3390/su151612449>
- Wheeler DC, Páez A (2009) Geographically weighted regression. *Handbook of applied spatial analysis: software tools, methods and applications*. Springer, pp 461–486
- Xiao T, Zhang LM, Cheung RWM, Lacasse S (2023) Predicting spatio-temporal man-made slope failures induced by rainfall in Hong Kong using machine learning techniques. *Géotechnique* 73:749–765. <https://doi.org/10.1680/jgeot.21.00160>
- Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM (2016) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* 13:839–856. <https://doi.org/10.1007/s10346-015-0667-1>
- Yu H, Fotheringham AS, Li Z, Oshan T, Wolf LJ (2020) On the measurement of bias in geographically weighted regression models. *Spat Stat* 38. <https://doi.org/10.1016/j.spasta.2020.100453>
- Yu H, Gong H, Chen B, Liu K, Gao M (2020) Analysis of the influence of groundwater on land subsidence in Beijing based on the geographical weighted regression (GWR) model. *Sci Total Environ* 738. <https://doi.org/10.1016/j.scitotenv.2020.139405>
- Yu B, Chen W, Feng W, Liu K, Ye L (2023) A case study of shallow landslides triggered by rainfall in Sanming, Fujian Province, China. *Environ Earth Sci* 82. <https://doi.org/10.1007/s12665-023-11118-4>
- Zhang C, Tang Y, Xu X, Kiely G (2011) Towards spatial geochemical modelling: use of geographically weighted regression for mapping soil organic carbon contents in Ireland. *Appl Geochem* 26:1239–1248. <https://doi.org/10.1016/j.apgeochem.2011.04.014>
- Zhang G, Cai Y, Zheng Z, Zhen J, Liu Y, Huang K (2016a) Integration of the statistical index method and the analytic hierarchy process technique for the assessment of landslide susceptibility in Huizhou, China. *Catena* 142:233–244. <https://doi.org/10.1016/j.catena.2016.03.028>
- Zhang M, Cao X, Peng L, Niu R (2016b) Landslide susceptibility mapping based on global and local logistic regression models in Three Gorges Reservoir area, China. *Environ Earth Sci* 75:1–11. <https://doi.org/10.1007/s12665-016-5764-5>
- Zhang H, Zhang G, Jia Q (2019) Integration of analytical hierarchy process and landslide susceptibility index based landslide susceptibility assessment of the Pearl river delta area, China. *IEEE J Select Topics Appl Earth Observ Remote Sens* 12:4239–4251. <https://doi.org/10.1109/JSTARS.2019.2938554>
- Zhang W, Liu S, Wang L, Samui P, Chwała M, He Y (2022) Landslide susceptibility research combining qualitative analysis and quantitative evaluation: a case study of Yunyang County in Chongqing China. *Forests* 13:1055. <https://doi.org/10.3390/f13071055>
- Zhang S, Tan S, Liu L, Ding D, Sun Y, Li J (2023) Slope rock and soil mass movement geological hazards susceptibility evaluation using information quantity, deterministic coefficient, and logistic regression models and their comparison at Xuanwei China. *Sustainability* 15:10466. <https://doi.org/10.3390/su151310466>
- Zhao Z, Liu ZY, Xu C (2021) Slope unit-based landslide susceptibility mapping using certainty factor, support vector machine, random forest, CF-SVM and CF-RF models. *Front Earth Sci* 9:589630. <https://doi.org/10.3389/feart.2021.589630>
- Zhou X, Wen H, Zhang Y, Xu J, Zhang W (2021) Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. *Geosci Front* 12:101211. <https://doi.org/10.1016/j.gsf.2021.101211>
- Zulkafli SA, Abd Majid N, Rainis R (2023) Spatial analysis on the variances of landslide factors using geographically weighted logistic regression in Penang Island, Malaysia. *Sustainability* 15. <https://doi.org/10.3390/su15010852>