



## Article

# Quantitative Assessment of Landslide Risk Based on Susceptibility Mapping Using Random Forest and GeoDetector

Yue Wang<sup>1,2</sup>, Haijia Wen<sup>3,4,5,\*</sup> , Deliang Sun<sup>6</sup> and Yuechen Li<sup>1,2</sup>

<sup>1</sup> Chongqing Engineering Research Center for Application of Remote Sensing Big Data, School of Geographical Sciences, Southwest University, Chongqing 400715, China; wwyue1998@aliyun.com (Y.W.); liyuechen@swu.edu.cn (Y.L.)

<sup>2</sup> Chongqing Jinpo Mountain National Field Scientific Observation and Research Station for Karst Ecosystem, Chongqing 400715, China

<sup>3</sup> Key Laboratory of New Technology for Construction of Cities in Mountain Area, Ministry of Education, Chongqing 400045, China

<sup>4</sup> National Joint Engineering Research Center of Geohazards Prevention in the Reservoir Areas, Chongqing 400044, China

<sup>5</sup> School of Civil Engineering, Chongqing University, Chongqing 400045, China

<sup>6</sup> Key Laboratory of GIS Application in Chongqing University, Chongqing 401331, China; sundeliang2002@aliyun.com

\* Correspondence: jhw@cqu.edu.cn; Tel.: +86-132-5132-1327

**Abstract:** This study aims to evaluate risk and discover the distribution law for landslides, so as to enrich landslide prevention theory and method. It first selected Fengjie County in the Three Gorges Reservoir Area as the study area. The work involved developing a landslide risk map using hazard and vulnerability maps utilizing landslide dataset from 2001 to 2016. The landslide dataset was built from historical records, satellite images and extensive field surveys. Firstly, under four primary conditioning factors (i.e., topographic factors, geological factors, meteorological and hydrological factors and vegetation factors), 19 dominant factors were selected from 25 secondary conditioning factors based on the GeoDetector to form an evaluation factor library for the LSM. Subsequently, the random forest model (RF) was used to analyze landslide susceptibility. Then, the landslide hazard map was generated based on the landslide susceptibility mapping (LSM) for the study region. Thereafter, landslide vulnerability assessment was conducted using key elements (economic, material, community) and the weights were provided based on expert judgment. Finally, when risk equals vulnerability multiplied by hazard, the region was categorized as very low, low, medium, high and very high risk level. The results showed that most landslides distribute on both sides of the reservoir bank and the primary and secondary tributaries in the study area, which showed a spatial distribution pattern of more north than south. Elevation, lithology and groundwater type are the main factors affecting landslides. Fengjie County landslide risk level is mostly low (accounting for 73.71% of the study area), but a small part is high and very high risk level (accounting for 2.5%). The overall risk level shows the spatial distribution characteristics of high risk in the central and eastern urban areas and low risk in the southern and northern high-altitude areas. Secondly, it is necessary to strictly control the key risk areas, and carry out prevention and control zoning management according to local conditions. The study is conducted for a specific region but can be extended to other areas around the investigated area. The developed landslide risk map can be considered by relevant government officials for the smooth implementation of management at the regional scale.

**Keywords:** landslide; susceptibility assessment; random forest; GeoDetector; risk assessment



**Citation:** Wang, Y.; Wen, H.; Sun, D.; Li, Y. Quantitative Assessment of Landslide Risk Based on Susceptibility Mapping Using Random Forest and GeoDetector. *Remote Sens.* **2021**, *13*, 2625. <https://doi.org/10.3390/rs13132625>

Academic Editor: William Frodella

Received: 29 May 2021

Accepted: 29 June 2021

Published: 4 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Landslides are one of the most severe and common geological hazards in the world, being significantly widespread, catastrophic and destructive, prone to chain disasters, and mainly occurring in mountainous areas [1,2]. The periodic water level rise and fall

in the Three Gorges Reservoir Area makes an unstable state in the slope on both sides of the reservoir, which aggravates the existing landslide recurrence or potential landslide instability, so there are countless hazards. Landslides can directly result in a threat to life, livelihood, casualties, agricultural livestock and forest growth throughout the world, ranging from minor social disruption to serious economic losses [3,4]. Landslide research has attracted worldwide attention, mainly due to the continuous improvement in people's awareness of the socio-economic impact of landslide, and the increasing pressure of urbanization on mountain environments [5]. During the periods between 2004–2010, 2620 landslide events were recorded worldwide, causing a total of 32,322 fatalities [6]. Only in China, more than 25,000 people have died from landslides over the past 60 years, and up to \$50 million a year of economic losses were caused by landslides [7]. This grim situation makes measures to prevent and forecast landslide disasters extremely urgent. To minimize losses and damages, studies need to be strengthened, starting from landslide data collection for landslide risk assessment.

In 1984, Varnes [8] first proposed the concept of landslide risk, which refers to the possible loss of population and economic activity caused by landslide disasters over a certain period of time. Moreover, landslide risk assessment refers to the interaction between the disaster-causing body and the disaster-bearing body in order to evaluate and estimate the number of casualties or property losses that may be caused by landslides [9]. The disaster-causing body can cause danger but does not consider the hazard object, reflecting a natural attribute of landslides. The disaster-bearing body has the disaster-bearing function, which refers to the personnel, property, etc., which suffer the landslide disaster, manifesting its social consequences. Therefore, landslide risk assessment results can intuitively show landslide risk distribution in the study area, and guide disaster prevention and mitigation work according to local conditions.

Risk assessment methods can be divided into qualitative, quantitative and qualitative-semi-quantitative methods. Qualitative methods are generally carried out depending on the experience of expert engineering geologists and geomorphologists, which may be subjective. Qualitative-semi-quantitative methods are a combination of the above two methods. Biçer et al. [9] assessed landslide risk by a semi-quantitative approach in a landslide-prone area located in the Eastern Mediterranean region of Turkey and produced a landslide risk index map. However, quantitative methods are performed by using statistical and/or mathematical modelling techniques. Based on the value estimation of the hazard bearing body at different times and working conditions, Bonachea et al. [10] established the quantitative analysis model of landslide disaster risk assessment and the results showed that this quantitative method is feasible. Risk assessment for single landslide has become relatively mature, while landslide risk assessment at the regional scale has not been a frequent topic in the literature. Xu et al. [11] took the Ganba landslide in Xuanen County of Hubei Province in China as a case study of landslide risk assessment. Zhang et al. [12] researched the risk of a barrier dam induced by the Caijiaba landslide, finding that the riverway would be blocked by the debris, forming a weir dam. However, the scope of single landslide risk research is limited, whereas research on a regional scale can grasp the basic situation of regional landslide risk more macroscopically, so as to provide a theoretical basis for management departments.

As the basic premises and core work of landslide risk assessment [13], landslide susceptibility, hazard and vulnerability assessments have been conducted by many scholars during the recent decades [14–16]. As a foundation for landslide prevention and spatial planning, landslide susceptibility mapping (LSM) depicts the future possibility of landslides in a region [17,18]. The adoption of modeling methodologies plays a key role in the effectiveness of LSM [19]. There are many methods of landslide susceptibility assessment, and the early stage is mainly statistical analysis. Due to the complex nonlinear characteristics of landslide development, many problems with landslide susceptibility assessment research have not been systematically solved. In the context of the rapid development of data mining technology, most researchers have begun to use machine learning

algorithms to study landslide susceptibility, including random forest (RF) [14,19], logistic regression (LR) [20,21], artificial neural networks (ANN) [15,22], support vector machine (SVM) [23,24] and other models. Further research shows that, compared with other machine learning algorithms, the random forest, tree-based ensemble algorithm, can achieve better results, because of its robust performance and high accuracy. It only needs a small amount of adjustment before model training [25]. In addition to the different algorithms used, redundancy and noise factors will also increase the uncertainty of the model and reduce the prediction ability. Therefore, screening of dominant and effective factors is conducive to improving the accuracy of risk assessment. Usually, factor selection methods can be classified into three categories: statistical methods, machine learning methods, and other methods. The commonly used statistical methods include factor analysis [26], correlation coefficient and rough set [27]. Machine learning methods including RF [28], and LR [29] has been employed in the literature. Although these methods can filter out relatively important influence factors and increase the reliability of LSM to a certain extent, they do not consider the pattern characteristics of spatial data, and the improvement of accuracy is limited. However, different to the above methods, GeoDetector [30] takes into account the spatial pattern characteristics between factors and landslide data, and the selected factors are more representative, which improves the accuracy of LSM.

In this paper, considering the above concepts, a quantitative assessment of landslide risk based on susceptibility mapping using random forest and GeoDetector was performed in Fengjie County in the Three Gorges Reservoir Area. Firstly, based on summarizing the theoretical methods of landslide risk assessment, taking 1522 historical landslides of Fengjie County as a sample, the development and distribution characteristics of landslides were deeply analyzed, and the condition factors optimized by GeoDetector. Then, the LSM was generated by RF model. Lastly, landslide risk assessment was studied combined with hazard and vulnerability assessment, and the landslide risk assessment model of Fengjie County was constructed. The risk assessment levels of different regions were obtained, and the spatial distribution characteristics of landslide risk were revealed, which provided a scientific basis for the prevention and control of landslide disasters in Fengjie County. The highlights of this paper include: (1) GeoDetector was adopted for factor screening; (2) Machine learning method was applied to LSM; (3) Landslide risk assessment at regional scale was conducted.

## 2. Materials and Methods

### 2.1. Study Area

Fengjie County is located in the east of Chongqing, the center of the Three Gorges Reservoir Area ( $109^{\circ}1'17''$  E~ $109^{\circ}45'58''$  E,  $30^{\circ}29'19''$  N~ $31^{\circ}22'33''$  N, Figure 1). Situated in the east of the Sichuan Basin, as a mountainous region the county is the junction of the Dabashan arc fold fault zone and east Sichuan arc concave fold zone, with a sophisticated structural stress field (Figure 2). The region is mainly mountainous, with the highest elevation at 2123 m. Topographic characteristics are high in the north, low in the south, and vary widely. Its climate is subtropical monsoon, with frequent rainfall and annual average precipitation of 1132 mm, predominantly occurring from May to September, and an average annual temperature of approximately  $16.5^{\circ}\text{C}$ . There are many water systems in the county, with a drainage area of more than  $50\text{ km}^2$ . The Yangtze River runs through the central part, with an average flow of  $13,700\text{ m}^3/\text{s}$  over many years.

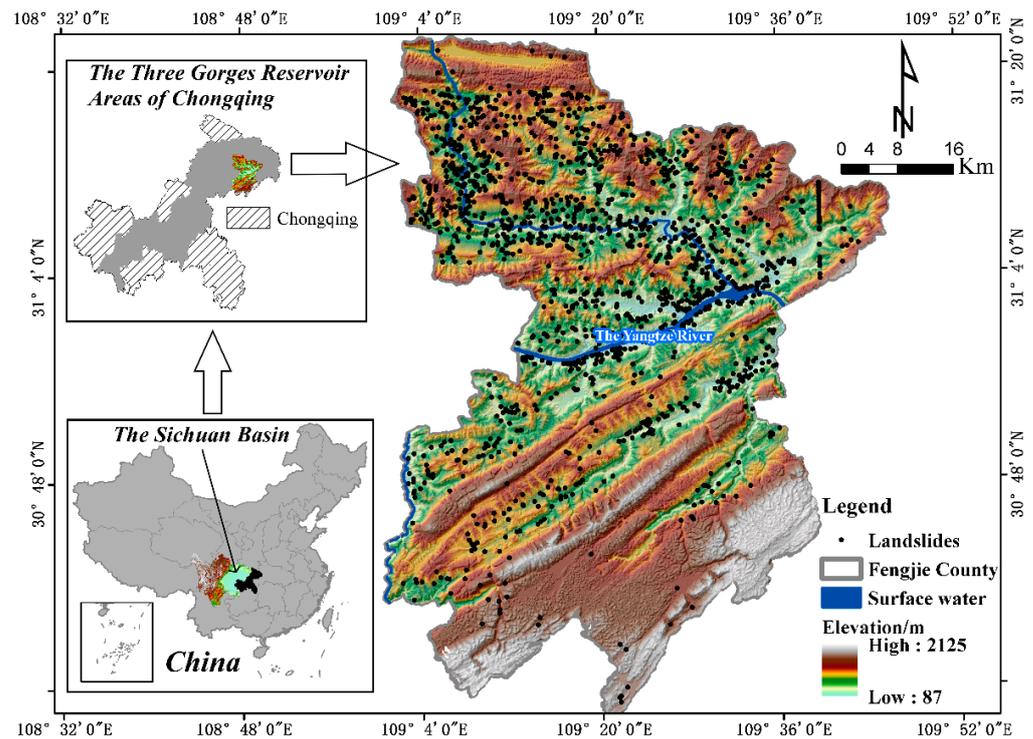


Figure 1. Location and landslide distribution of Fengjie County.

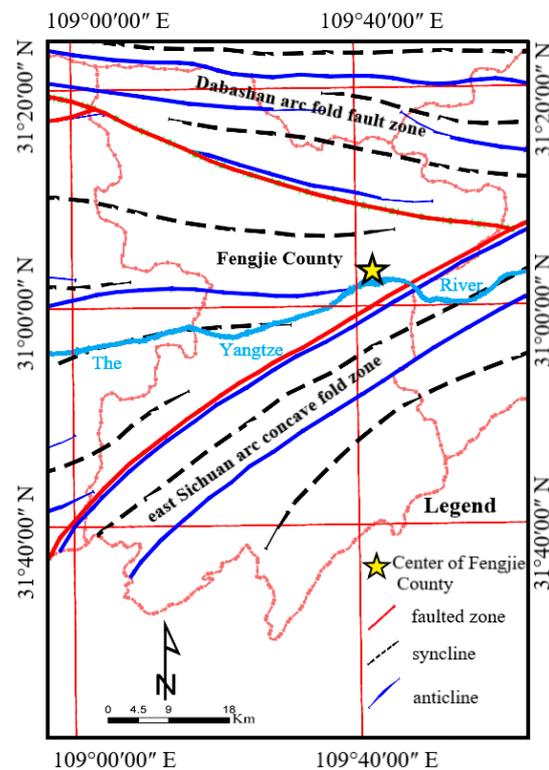
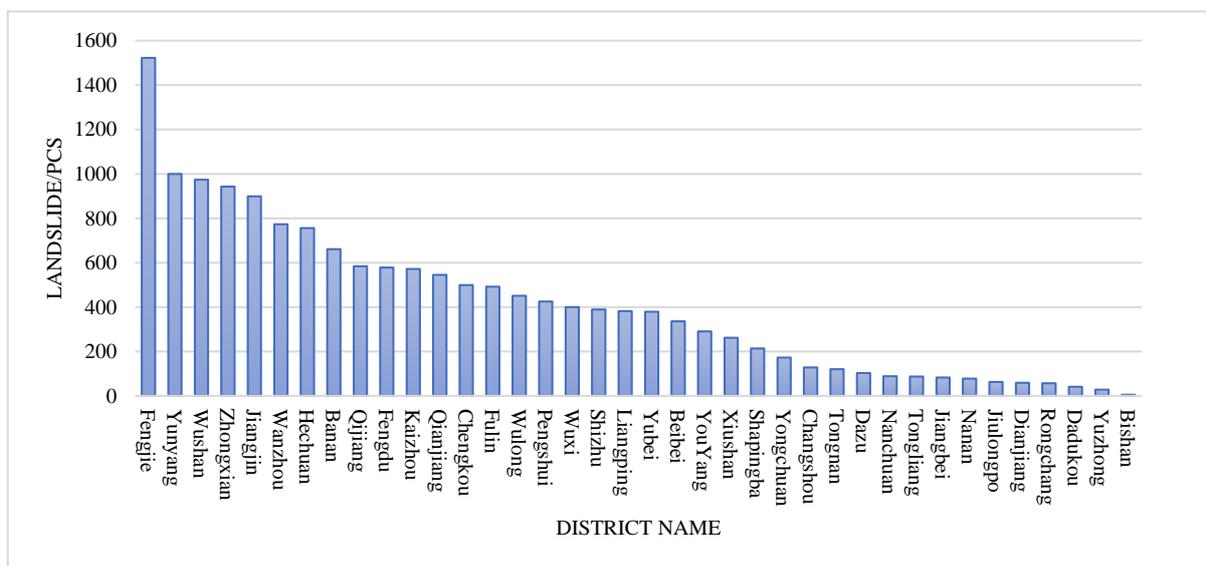


Figure 2. Geological Structure Outline Map of Fengjie County.

Figure 1 shows the geographical location of Fengjie County and the distribution of historical landslides. As shown, most of the landslides are distributed among the banks of the reservoir and both sides of the rivers, showing the spatial distribution characteristics more for the north and less for the south. Signs indicate that the human engineering activities of the Three Gorges Reservoir Area, such as town construction, resettlement, water

storage, roads, bridges, and power generation of the reservoir areas, as well as continuous precipitation, have led to a substantial impact on induced landslides. In addition, in terms of time distribution characteristics, landslides mainly occur from May to October. This is because the slope body is more prone to landslides under the influence of heavy rainfall in summer after being soaked in winter. Therefore, the high occurrence period of landslides is consistent with the annual flood season. Specifically, the landslide is relatively stable in the dry season or normal conditions, but in severe convective weather or rainstorm, the stability of the slope body will decrease.

According to the statistical results of landslide numbers in the counties of Chongqing (Figure 3) from 2001 to 2016, compared with other counties, Fengjie County has the largest number of landslides (1522). Especially, it has nearly 1.5 times as many historical landslides as Yunyang County. The results show that the Fengjie County landslide disaster situation is grim, and landslide disaster risk assessment is of great significance.

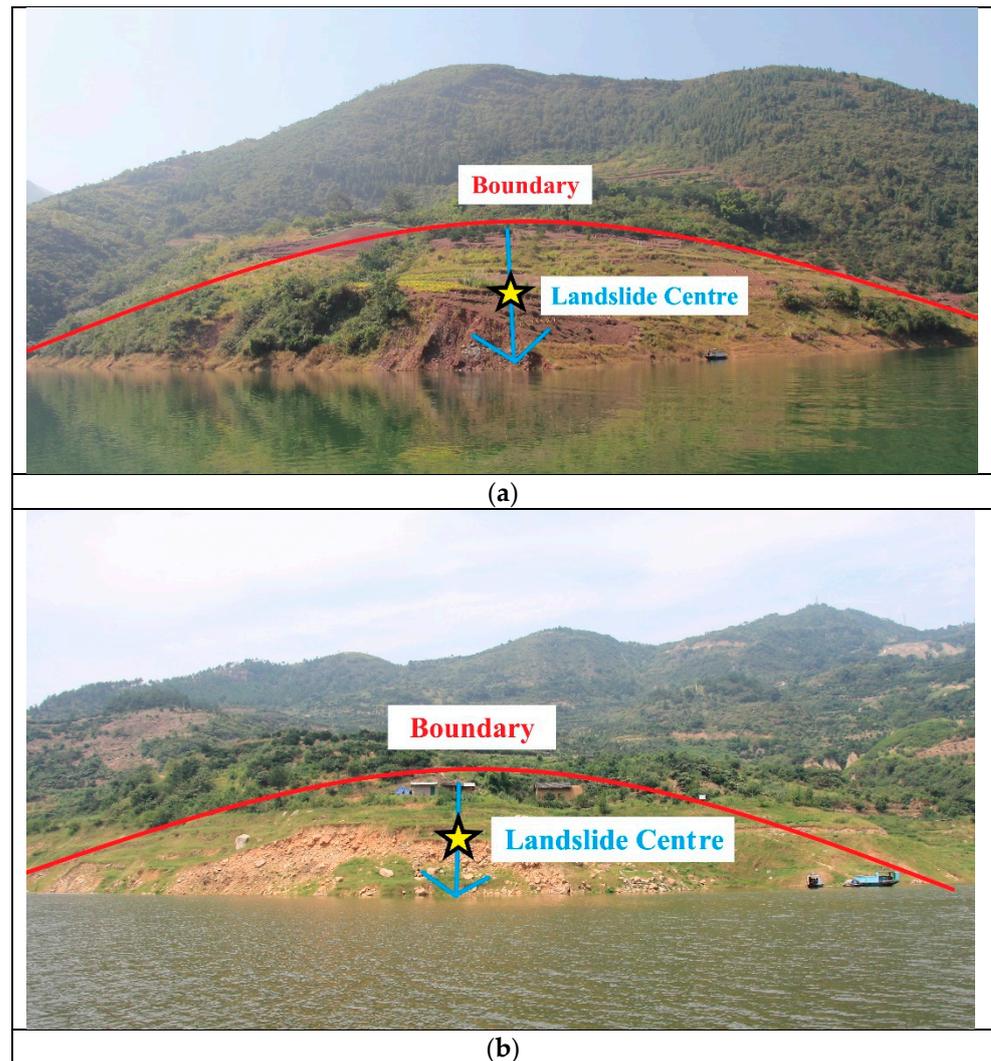


**Figure 3.** Landslide numbers for the counties of Chongqing from 2001 to 2016.

## 2.2. Data

The data on 1522 historical landslides in Fengjie County from 2001 to 2016 are from the Chongqing Geological Monitoring Station. The attribute table contains information on landslide name, occurrence location, and time. Since 2007, the geological disaster management department of Chongqing formed a geological disaster garrison (about 500 members) permanently stationed in geological disaster-prone areas. Once the landslide emergency is found, they will be responsible for the evacuation of personnel and recording basic information on the landslide. Given the actual situation of the landslides in the study area, the historical landslide data were divided into two types. Most of the landslides in the study area were shallow/soil (81.68%) and only 18.32% were deep/rock landslides. Therefore, we will generate a comprehensive susceptibility map for generic landslides in the results.

During the field survey, the central latitude and longitude coordinates of the landslide are recorded as the location, as a point to use as input data in the susceptibility model. Therefore, we selected two typical landslides in Fengjie County, namely, the Xiawazhaping Landslide and the Zhujiatian Landslide (Figure 4a,b), to show the location and center of the landslide.



**Figure 4.** Field survey of landslides in Fengjie County: (a) the Xiawazhaping Landslide; (b) the Zhujiatian Landslide.

POI (point of interest) data in 2016 were crawled by a python program. These data points include buildings such as hospitals, primary and secondary schools, business centers, parks and squares, taking into account various types of commercial and educational activity that can represent human engineering. Other data sources, types and accuracy are shown in Table 1.

**Table 1.** Data and data sources.

Data Name	Data Sources	Type	Scale
Historical landslide	Chongqing Geological monitoring station	Dataset	
Elevation	Aster satellite	Grid	30 m
Geological data	National Geological Data Center	Grid	1:200,000
Land cover	Chongqing Municipal Bureau of land and resources	Vector	1:100,000
Administrative division	Chongqing Municipal Bureau of land and resources	Vector	1:100,000
River network	Chongqing Water Resources Bureau	Vector	1:100,000
Annual rainfall	Chongqing Meteorological Administration	Dataset	90 m
Road	Chongqing Transportation Commission	Vector	1:100,000
Satellite image	Geospatial Data Cloud platform	Grid	30 m
POI of Chongqing	Web Crawler	Dataset	
GDP (Gross Domestic Product)	Resource and Environment Science and Data Center	Grid	1 km × 1 km
Population	Resource and Environment Science and Data Center	Grid	1 km × 1 km

### 2.2.1. Data on Landslide Susceptibility Assessment

Landslide development is not only controlled by the geological conditions of the slope but also interfered with by hydrological and climatic conditions [16]. Some scholars have found that there are up to 596 factors in landslide susceptibility research [31]. According to the principles of significance, representativeness, scientificity and operability, the primary factors were selected. Based on the collected data and related literature, combined with the spatial laws and regional characteristics of landslide disaster distribution in Fengjie County, as shown in Table 2, four types of factor libraries were established in this study and 25 disaster-causing factors were selected:

**Table 2.** Classification of conditioning factors of landslide susceptibility.

Type	Factor	Classification
Topographic Factors	plane curvature	1. <-1.4; 2. -1.4~-0.38; 3. -0.38~0.4; 4. 0.4~1.5; 5. >1.5
	elevation/m	1. <343; 2. 343~538; 3. 538~712; 4. 712~872; 5. 872~1025; 6. 1025~1185; 7. 1185~1357; 8. 1357~1554; 9. 1554~1783; 10. >1783
	elevation coefficient of variation	1. <0.008; 2. 0.008~0.02; 3. 0.02~0.035; 4. 0.035~0.055; 5. 0.055~0.085; 6. 0.085~0.153; 7. >0.153
	slope/°	1. <6; 2. 6~13; 3. 13~19; 4. 19~24; 5. 24~30; 6. 30~35; 7. 35~42; 8. 42~50; 9. >50
	aspect	1. Flat; 2. North; 3. Northeast; 4. East; 5. Southeast; 6. South; 7. Southwest; 8. West; 9. Northwest
	slope variability	1. <4; 2. 4~7; 3. 7~10; 4. 10~13; 5. 13~17; 6. 17~20; 7. 20~25; 8. 25~31; 9. >31
	curvature	1. <-2; 2. -2~-0.8; 3. -0.8~0.7; 4. 0.7~2.8; 5. >2.8
	profile curvature	1. <-2; 2. -2~-0.6; 3. -0.6~0.4; 4. 0.4~1.8; 5. >1.8
	slope shape	1. Convex slope; 2. Concave slope; 3. Straight slope
	RDLS/m	1. <15; 2. 15~29; 3. 29~43; 4. 43~58; 5. 58~78; 6. 78~112; 7. >112
	slope position	1. Valleys; 2. Flats slope; 3. Ridge; 4. Middle slope; 5. Lower slope; 6. Upper slope
	micro-landform	1. Canyons, Deeply incised streams; 2. Open slopes; 3. Midslope ridges, Small hills in plains; 4. Plains; 5. Upland drainages, Headwaters; 6. Mountain tops, High narrow ridges; 7. Local ridges hills in valleys; 8. Midslope drainages, shallow valleys; 9. Upper slopes, Plateau; 10. U-shape valleys
	TRI	1. <1.07; 2. 1.07~1.2; 3. 1.2~1.4; 4. 1.4~1.8; 5. >1.8
incision density	1. <0; 2. 0~2; 3. 2~3; 4. 3~4; 5. 4~5; 6. >5	
Incision depth/m	1. <433; 2. 433~616; 3. 616~700; 4. 933~1126; 5. 1126~1369; 6. 1369~1835; 7. >1835	
TWI	1. <5; 2. 5~7; 3. 7~10; 4. 10~15; 5. >15	
Geological Factors	lithology	1. T3xj; 2. T3b1; 3. T2b2; 4. T1j; 5. T1d; 6. S1-2; 7. P2; 8. P1; 9. J3p/J3s; 10. J2s/J2xs; 11. J1-2z/J1z; 12. D2/D3
	CRDS	1. Reverse slope; 2. Tangential slope; 3. Outward slope; 4. Oblique slope; 5. Flat; 6. Dip-slope I; 7. Dip-slope II
	distance from the fault/m	1. <500; 2. 500~1000; 3. 1000~1500; 4. 1500~2000; 5. 2000~2500; 6. 2500~3000; 7. >3000
Meteorological and Hydrological Factors	distance from rivers/m	1. <100; 2. 100~200; 3. 200~300; 4. 300~400; 5. 400~500; 6. 500~600; 7. >600
	SPI	1. <15; 2. 15~30; 3. 30~45; 4. 45~60; 5. 60~100; 6. 100~1000; 7. >1000
	groundwater type	1. Carbonate fissure cave water; 2. Fracture water of clastic rock interbedded karst cave; 3. Crushed rock fissure water; 4. Sandstone fissure/Gravel fissure/shale pore fissure water; 5. Sandy pebble micro confined water; 6. Dolomite fissure karst water; 7. Mud dolomite fissure water; 8. Weathering fissure water; 9. Without water
	STI	1. <133; 2. 133~1071; 3. 1071~3483; 4. 3483~8708; 5. >8708
Vegetation Factors	NDVI	1. <0.1; 2. 0.1~0.2; 3. 0.2~0.3; 4. 0.3~0.4; 5. 0.4~0.5; 6. 0.5~0.6; 7. >0.6
	land cover	1. Meadow; 2. Farmland; 3. Water area; 4. Forest; 5. Garden plot; 6. Others; 7. Residential land; 8. Transportation

(1) Topographic Factors: plane curvature, elevation, elevation coefficient of variation [32], slope, aspect, slope variability, curvature, profile curvature, slope shape, relief degree of the land surface (RDLS), slope position, micro-landform, terrain roughness index (TRI), incision density, incision depth and topographic wetness index (TWI). These factors were all calculated with a digital elevation model (DEM). DEM data from Aster satellite are at 30 m spatial resolution. Topographic factors are closely related to landslide occurrence and they are the main factors which control the spatial distribution of landslide disasters. Mark et al. [33] studied the corresponding relationship between frequency of shallow landslide and the terrain. The results show that landslide disaster has a good correlation with steep terrain.

(2) Geological Factors: lithology, combination reclassification of stratum dip direction and slope aspect (CRDS), distance from the fault. As important internal causes of landslides, different geological factors show large differences in physical and mechanical parameters and directly affect the slope stability. In general, the occurrence and formation of landslide occurs under certain geological environmental conditions, such as free interface of slope, sliding soil and rock mass, cutting slope and groundwater active tectonic surface.

(3) Meteorological and Hydrological Factors: distance from rivers, stream power index (SPI), groundwater type, sediment transport index (STI). Factors such as surface water and groundwater are important in affecting slope stability. Many landslides are related to the role of water, or water is their trigger factor. Water softens or mudifies the rock mass of the slope, which greatly reduces the shear strength of rock mass. The erosion of surface water and the dissolution of groundwater also directly damage the slope. In this study, most of the historical landslides are distributed along rivers, because Fengjie County is located in the center of the Three Gorges Reservoir, and the periodic rise and fall of the water level is one of the main causes of landslides.

(4) Vegetation Factors: Normalized Difference Vegetation Index (NDVI) and land cover. NDVI, as an important parameter of ecological environment quality, directly affects the degree of soil erosion. Land cover shows the degree of human disturbance and destruction of rock and soil. Forest is beneficial to solid slope and reduces the occurrence of landslide, while farmland and residential land will destroy the stability of slope and cause slope damage.

The lithology and fault factor can be obtained by vectorizing the 1:10,000 geological map. Groundwater type was generated after vectorization based on a 1: 200,000 hydro-geological map. NDVI data was generated by Landsat 8 OLI processing. Distance from the fault and rivers is needed to establish multi-level buffer zones for faults and rivers. In summary, 30 m × 30 m was selected as the basic unit of susceptibility assessment and for establishing the factor geospatial database (Figure 5).

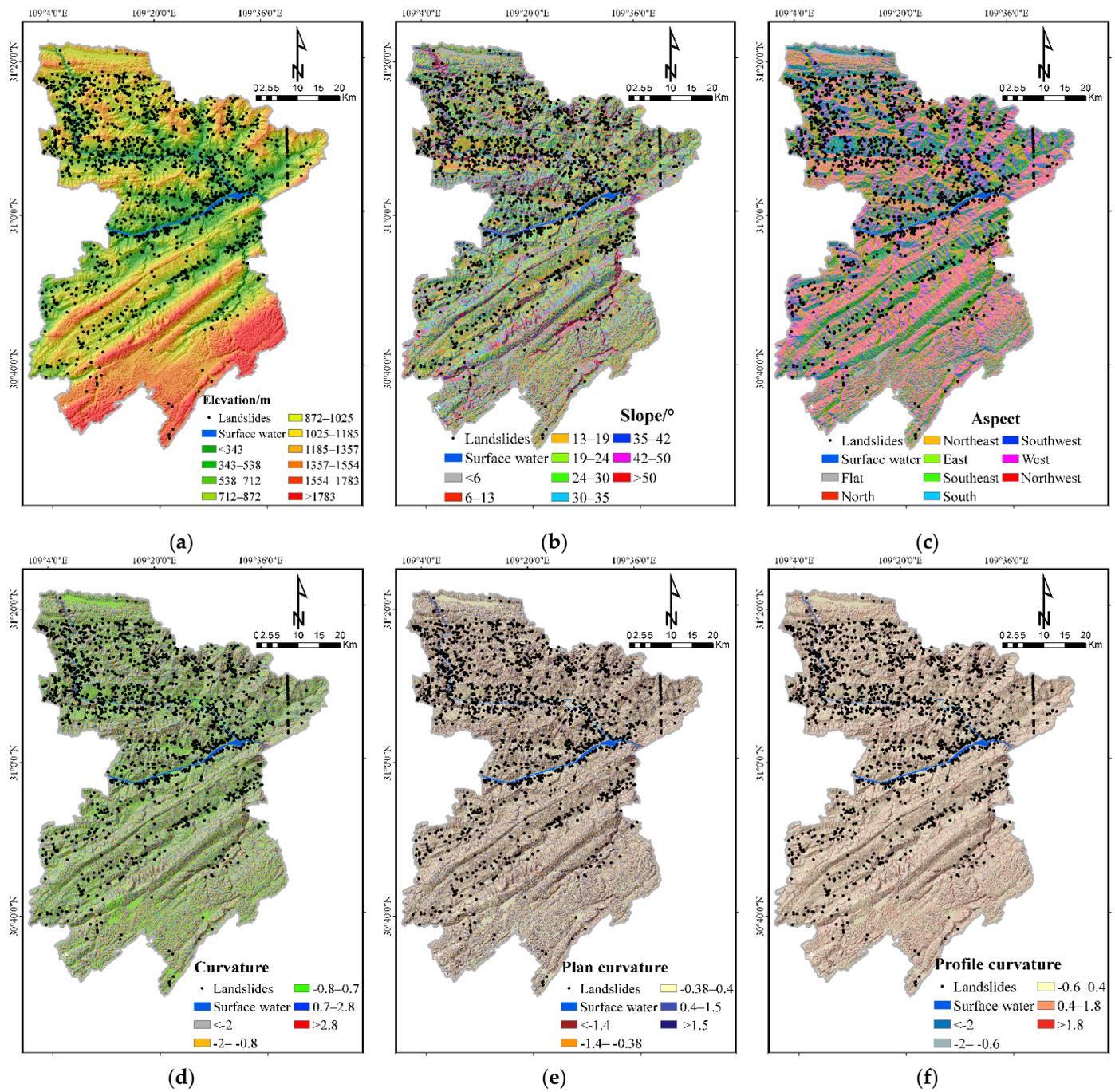


Figure 5. Cont.

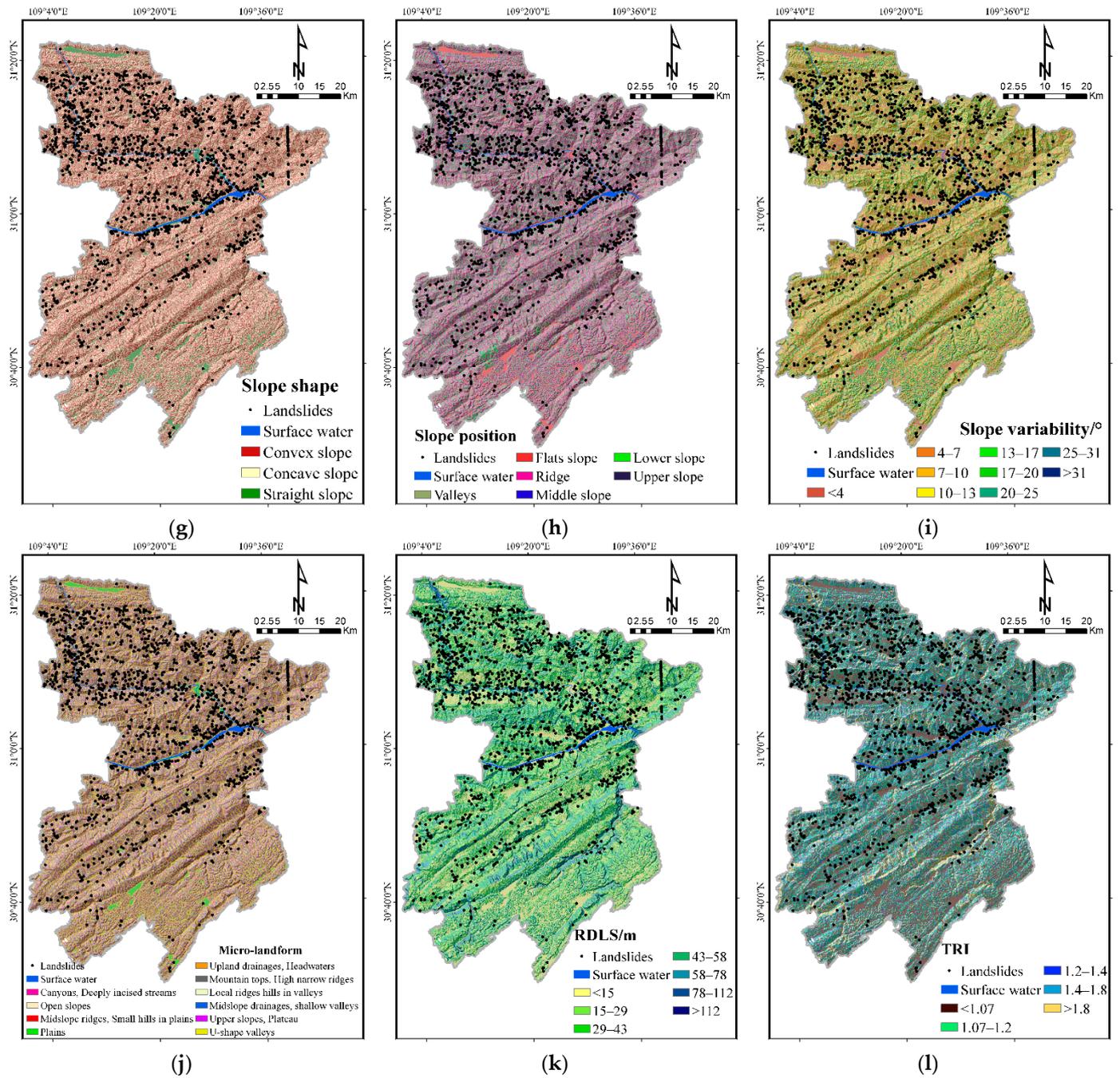


Figure 5. Cont.

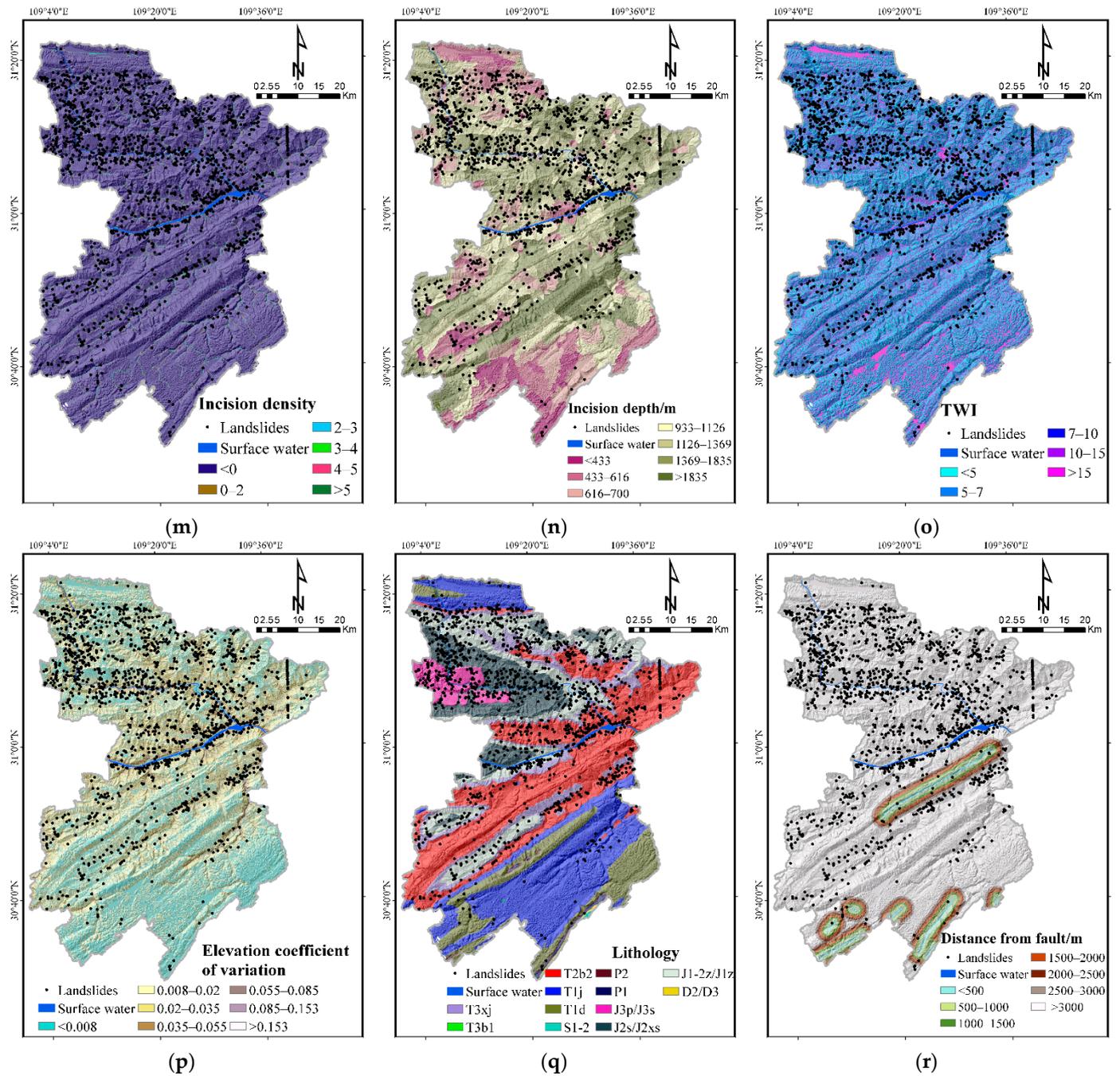


Figure 5. Cont.

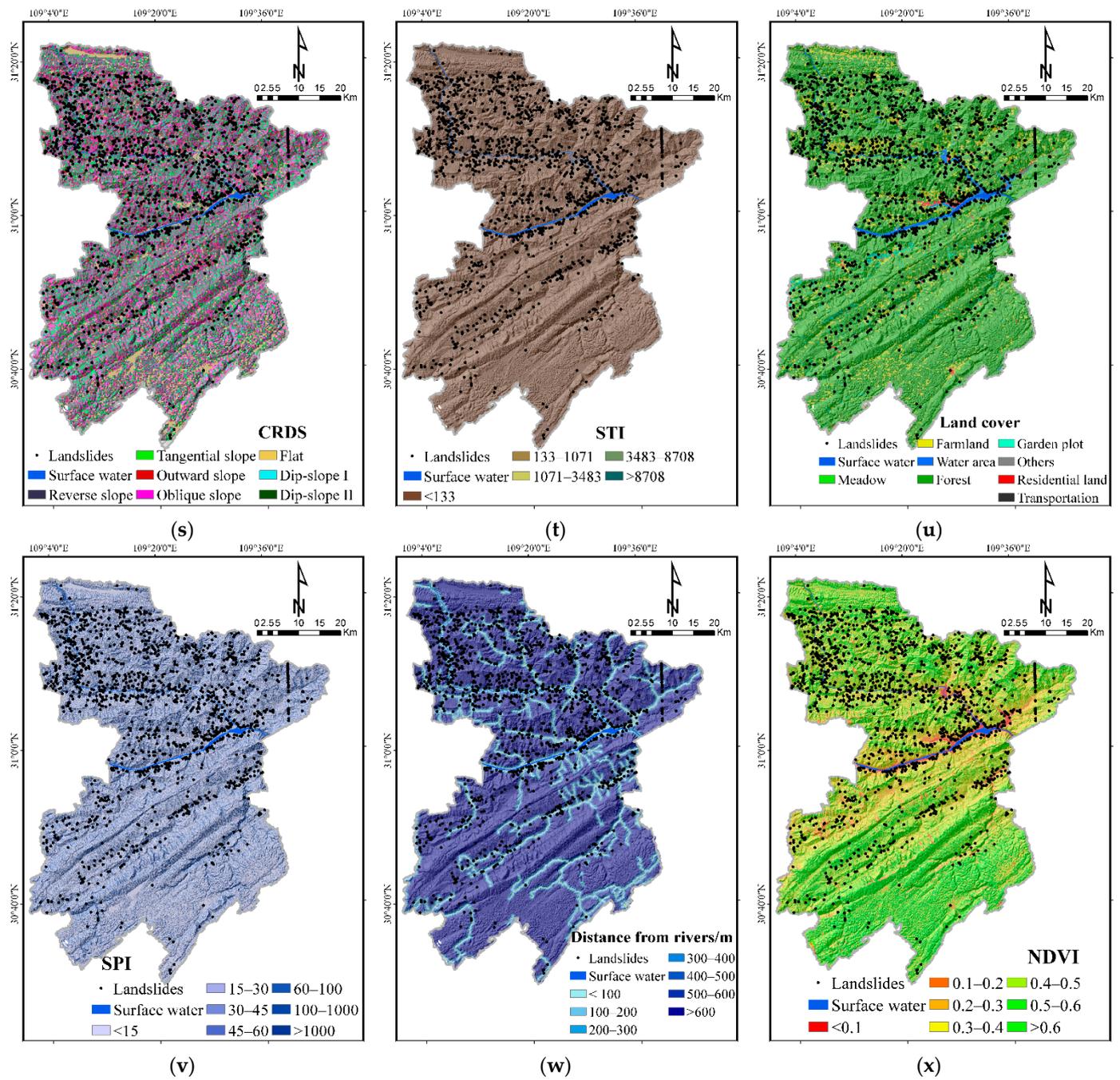
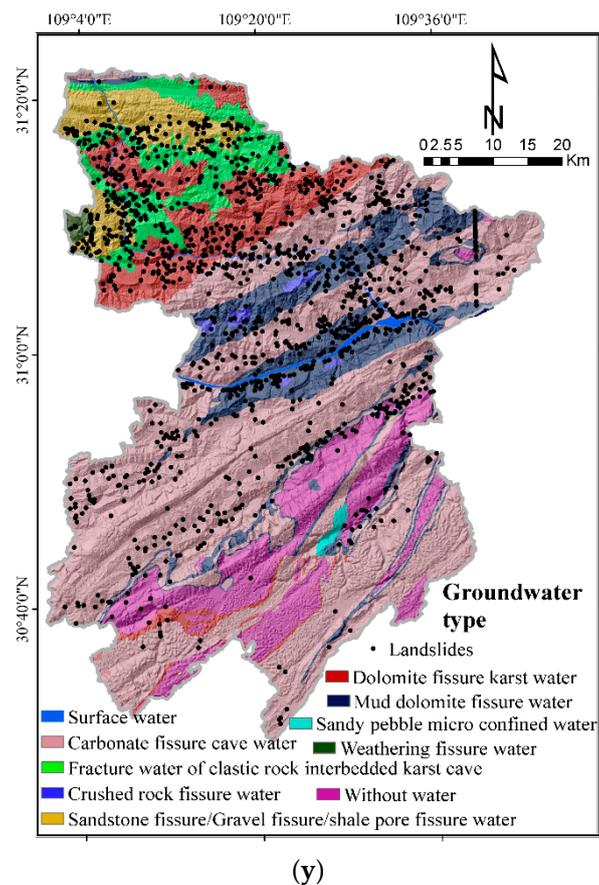


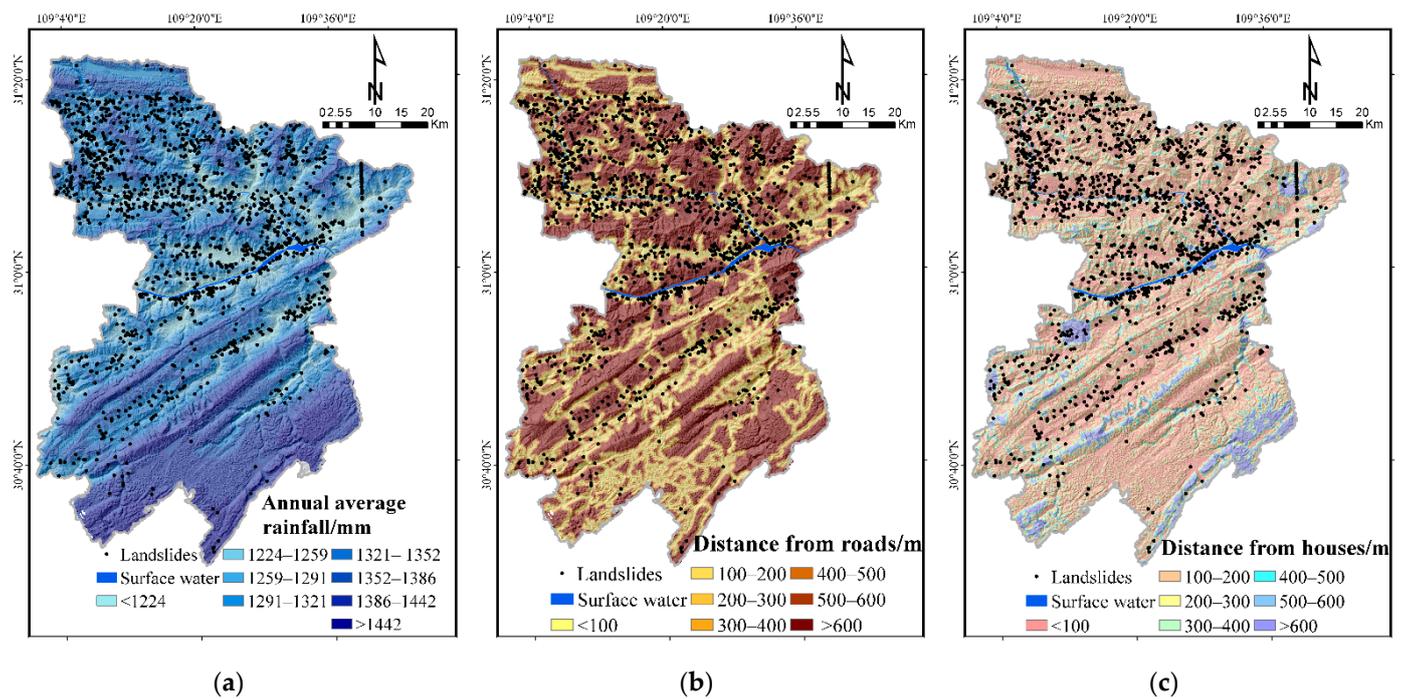
Figure 5. Cont.



**Figure 5.** Conditioning Factors on Layer of Landslide Susceptibility: (a) Elevation; (b) Slope; (c) Aspect; (d) Curvature; (e) Plan curvature; (f) Profile curvature; (g) Slope shape; (h) Slope position; (i) Slope variability; (j) Micro-landform; (k) RDLS; (l) TRI; (m) Incision density; (n) Incision depth; (o) TWI; (p) Elevation coefficient of variation; (q) Lithology; (r) Distance from fault; (s) CRDS; (t) STI; (u) Land cover; (v) SPI; (w) Distance from rivers; (x) NDVI; (y) Groundwater type.

### 2.2.2. Data for Landslide Hazard Assessment

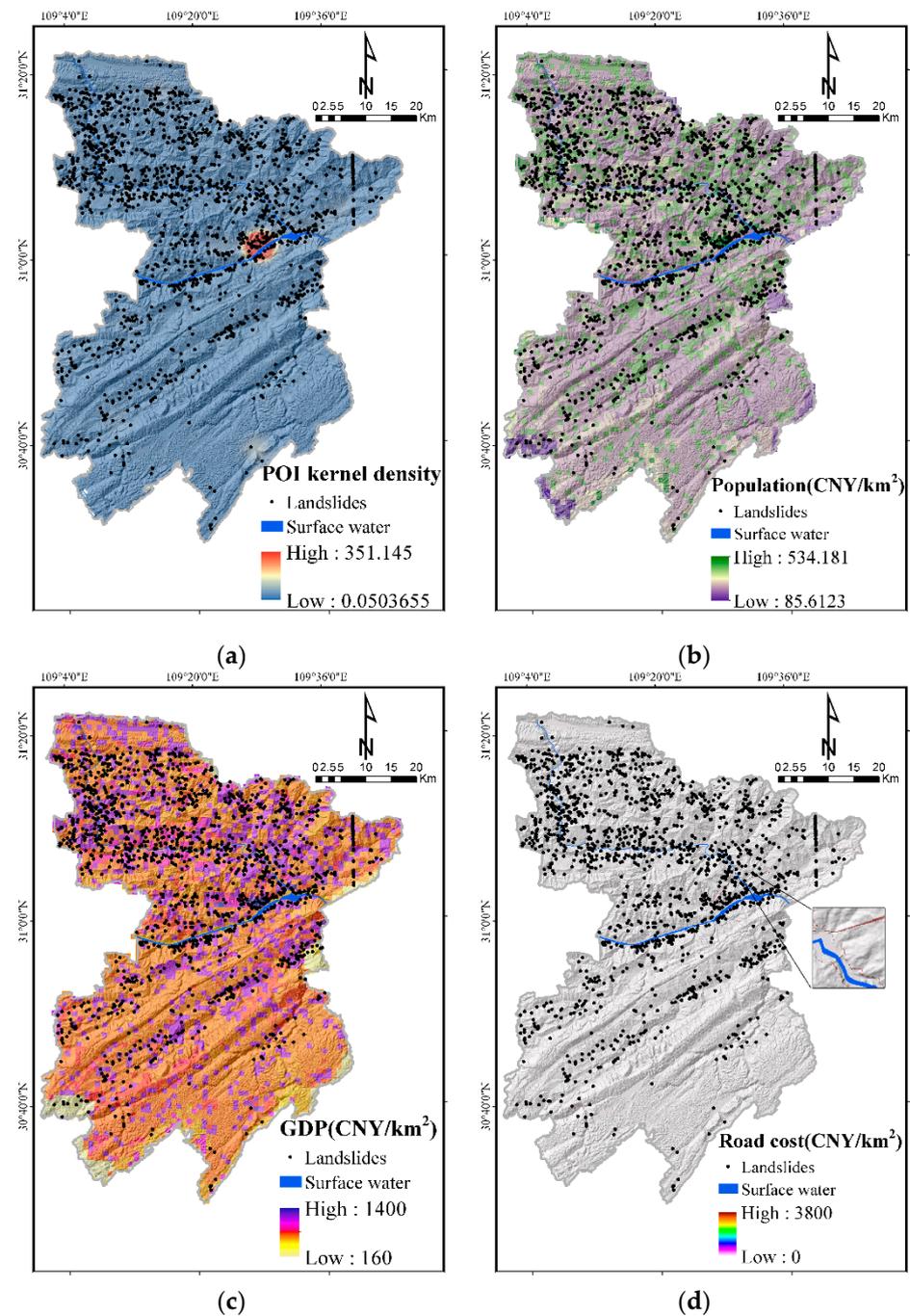
Based on the collected data and related literature [34–36], combined with the spatial distribution and regional characteristics of Fengjie County landslides, the landslide risk evaluation library was constructed by selecting three secondary triggering factors under the two kinds of index, annual average rainfall and human engineering activities, including distance from roads and houses. The annual average rainfall is obtained by spatial interpolation based on the data of meteorological monitoring stations in the county. The distance from roads and houses are obtained by multiple buffers ranging from less than 100 m to over 600 m. Landslide hazard assessment factors are shown in Figure 6.



**Figure 6.** Triggering factors Layer of Landslide Hazard: (a) Annual average rainfall; (b) Distance from roads; (c) Distance from houses.

### 2.2.3. Data of Landslide Vulnerability Assessment

Evidence from various studies indicates that material, community, and economic factors need to be considered in vulnerability assessment [13,37,38]. Based on disaster-affected body data collection, remote sensing interpretation and field investigation, we selected four important factors in constructing the landslide vulnerability evaluation library. These are widely used in previous studies and best reflect vulnerability, including POI kernel density, road cost (CNY/km<sup>2</sup>), population (CNY/km<sup>2</sup>) and GDP (CNY/km<sup>2</sup>). Landslide vulnerability assessment factors are shown in Figure 7. POI kernel density is related to material and is based on location services. If each POI site is regarded as a functional unit, then the higher the POI density, the more landslide vulnerability is likely to increase. POI kernel density analysis was made with ArcGIS software. Generally, with an increase in road cost, population and GDP, landslide vulnerability is likely to increase [38,39]. According to the principle that different roads correspond to different prices, road cost was created in ArcGIS software. Furthermore, population and GDP came from the Resource and Environment Science and Data Center and were also made with ArcGIS software.



**Figure 7.** Influence Factors on Layer of Landslide vulnerability: (a) POI kernel density; (b) Population; (c) GDP; (d) Road cost.

To reduce the data dispersion, all the factors after reclassification should be normalized. The classification index values of these factors were transformed linearly, thereby reducing the values to [0,1] intervals. The normalization formula is denoted as follows:

$$X^* = (X - X_{min}) / (X_{max} - X_{min}) \quad (1)$$

where  $X^*$  is the normalized data;  $X$  is the original data;  $X_{min}$  is the minimum value after each factor is assigned; and  $X_{max}$  is the maximum value after each factor is assigned.

### 2.3. Methodology

Referring to the landslide risk assessment framework of Van et al. (2006) [40], this work is divided into four steps, using the technical route shown in Figure 8: (1) Landslide susceptibility assessment. According to field investigation and related data, combined with 1522 historical landslides in Fengjie County and the ArcGIS platform, 25 landslide influencing factors were selected to construct the landslide susceptibility database. Then, the dominant factors were screened by GeoDetector, and the landslide susceptibility was evaluated by the random forest method. (2) Landslide hazard assessment. Based on the relevant data, interpretation and field investigation, the hazard assessment index of the study area is established, and the landslide hazard is further evaluated combined with the results of susceptibility assessment. (3) Landslide vulnerability assessment. Spatial analysis and quantification of selected vulnerability factors are carried out to evaluate landslide vulnerability. (4) Quantitative risk assessment of landslides. Based on the above assessments, the quantitative risk assessment for the Fengjie County landslides was carried out based on the GIS platform.

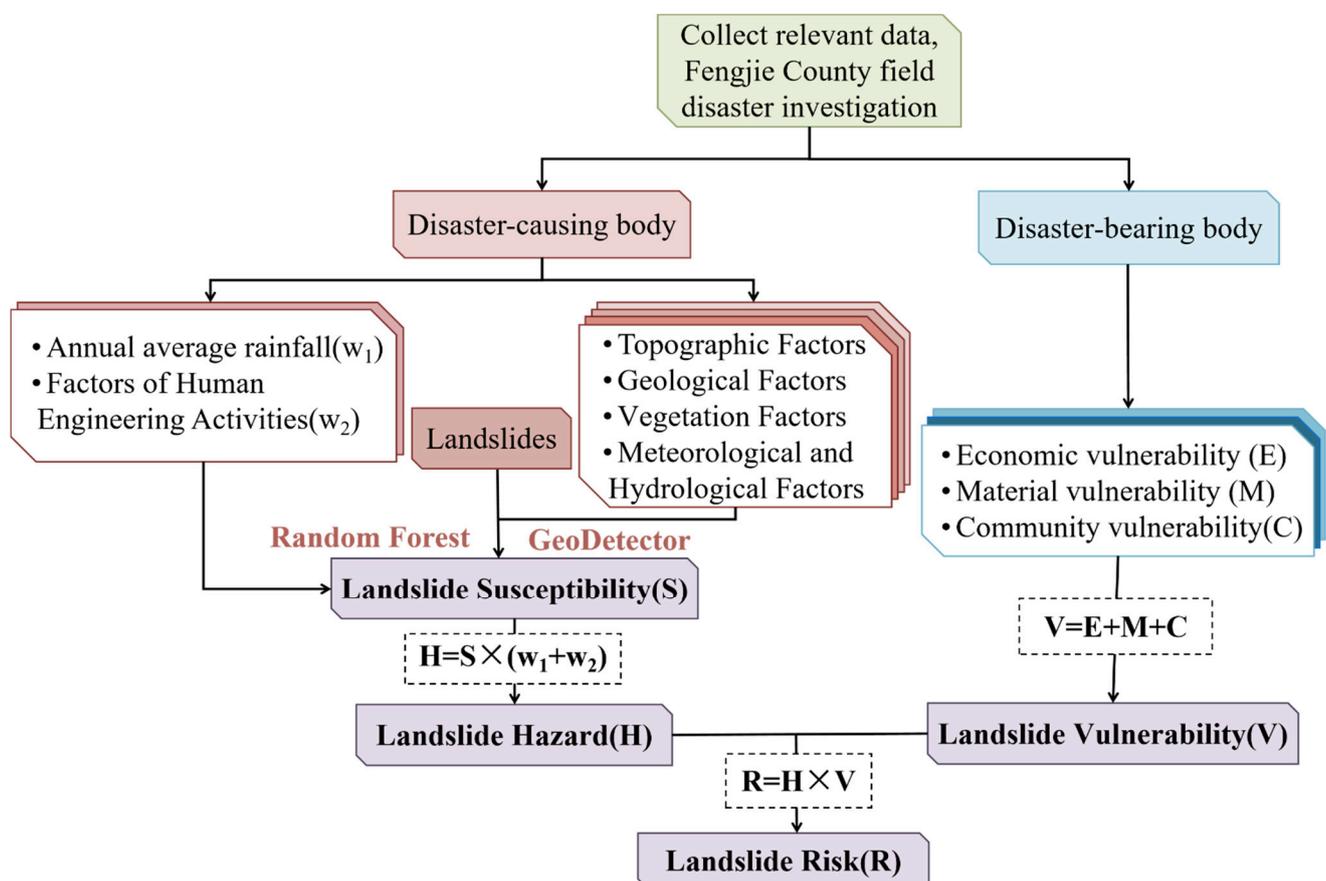


Figure 8. The methodological framework of the study.

#### 2.3.1. Landslide Susceptibility Assessment Method

##### 1. Random Forest Model (RF)

First proposed by Breiman (2001) [41], Random Forest (RF) is an ensemble method of separately trained binary decision trees. Compared with the traditional landslide division methods, the RF method introduces two random samplings (samples and features). The decision trees improve the accuracy and stability of the model more than a single decision tree, by using a randomly generated method to select samples and features. Then, the judgment results of multiple decision trees are voted on to arrive at the final output.

The key point of RF is to combine  $n$  independent decisions ( $u(X, \theta_k; k = 1, 2, \dots, n)$ ) to build a model. Each decision tree in the model judges or predicts the samples. Different classification models  $u_1(X), u_2(X), \dots, u_k(X)$  are obtained after sample training. Then, these classification models can be used to build RF models:

$$U(X) = \arg \max_Z \sum_{i=1}^k I(u_i(X) = Z) \quad (2)$$

where  $U(X)$  represents an RF model,  $u_i(X)$  denotes a single decision tree model,  $Z$  means output variable, and  $I(\cdot)$  is an explicit function.

Figure 9 shows the steps of the RF algorithm.

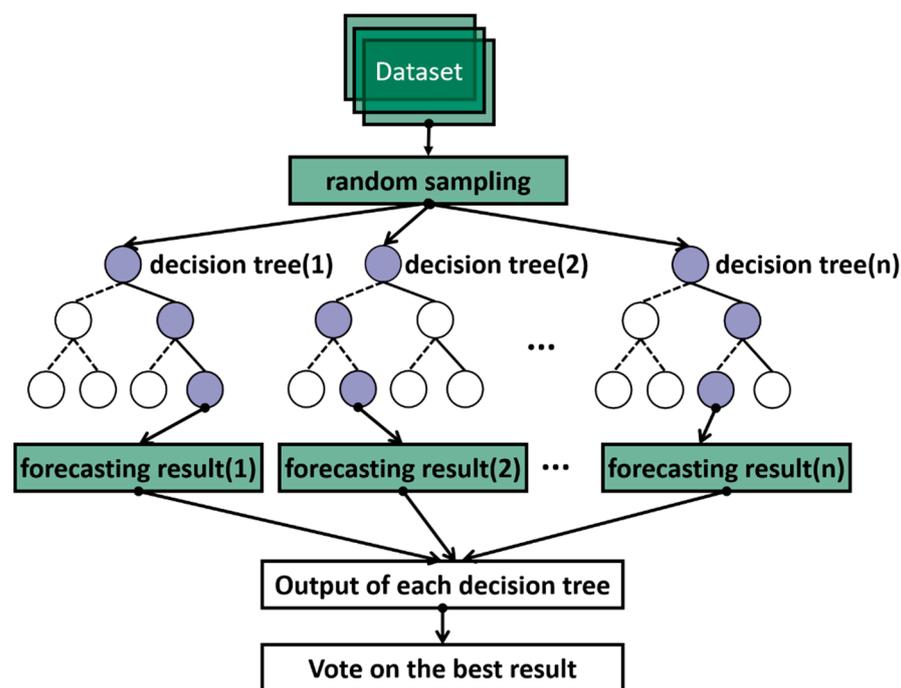


Figure 9. The schematic diagram of the RF algorithm.

In order to build decision trees, we use the Classification and Regression Tree (CART) algorithm to split the nodes in this study. CART follows the minimum principle of Gini. At node  $t$ , CART randomly extracts an object which is assigned to class  $i$  according to probability  $p(i|t)$ . The estimated probability that the object belongs to class  $j$  is  $p(j|t)$ . Under this rule, the estimated probability of misclassification is as follows:

$$Gini = \sum_{i \neq j}^J (p(i|t)p(j|t)) \quad (3)$$

## 2. GeoDetector

Geodetector is a new method proposed by Wang et al. [30] to detect spatial differences and reveal driving factors. Unlike other statistical methods, it gives a clear physical meaning and may overcome the limitations of statistical methods in dealing with variables [42]. The general assumption of the application of Geodetector to landslide research can be expressed as follows: if the condition factors control or contribute to the occurrence of the landslide, the spatial distribution characteristics of the landslide and the condition factor should be similar. Geodetector includes factor detector, risk detector, ecological detector and interaction detector. In this study, the factor detector is mainly used to calculate the explanatory  $q$  value of the conditional factor  $X$  when the landslide occurs, and the spatial

correspondence between  $X$  and the dependent variable  $Y$  is measured by the explanatory degree of the factor  $X$ , expressed as:

$$q = 1 - \frac{\sum_{m=1}^S N_m \sigma_m^2}{N \sigma^2} = 1 - \frac{WSS}{TSS} \quad (4)$$

$$WSS = \sum_{m=1}^S N_m \sigma_m^2 \quad (5)$$

$$TSS = N \sigma^2 \quad (6)$$

where  $m = 1, \dots, S$  is the stratification of variable  $Y$  or factor  $X$ ,  $N_m$  is the number of units in the entire area, and  $\sigma_m^2$  and  $\sigma^2$  are the variance of the  $Y$  value of the layer  $m$  and the entire area, respectively.  $WSS$  is the sum of variance within the layer, and  $TSS$  is the total variance of all the regions. The range of  $q$  is  $[0,1]$ , and the larger the value of  $q$ , the stronger the spatial heterogeneity of  $Y$ .

### 3. Evaluation of LSM Model

It is important to evaluate the model, which can reflect the model performance, and different aspects can be assessed. The precision (positive predictive value), sensitivity (true positive rate), specificity (true negative rate), and accuracy are usually considered effective indicators of fitting and predictive accuracies. Therefore, this paper applied these indicators to evaluate the performances of the RF model in the present research (Table 3).

**Table 3.** Explanation of statistical-index-based evaluations.

No.	Metric	Equation	Definition
1	Precision	$Precision = \frac{TP}{TP+FP}$	The fraction of relevant instances in the retrieved instances.
2	Sensitivity (SST)	$SST = \frac{TP}{TP+FN}$	The percentage of landslide cells that are correctly classified.
3	Specificity (SPF)	$SPF = \frac{TN}{TN+FP}$	The percentage of non-landslide cells that are correctly classified.
4	Accuracy (ACC)	$ACC = \frac{TP+TN}{M}$	The proportion of landslide and non-landslide cells are correctly classified.
5	Recall	$Recall = \frac{TP}{TP+FN}$	It indicates how many positive examples in the sample are predicted correctly.

$TP$  is the number of correctly predicted landslide cells.  $FP$  is the sum of cells of non-landslides that are classified as a landslide.  $FN$  is the sum of cells of landslides that are classified as non-landslide.  $TN$  is the number of correctly predicted non-landslide cells.  $M$  is the sum of landslides and non-landslides.

Otherwise, the Receiver Operating Characteristic (ROC) curve is also a method to measure the effectiveness of a model. The area under the receiver operating characteristic (AUC) value is used as the basis for determination [43]. This value ranges from 0.5 (very poor performance) to 1.0 (perfect performance). When the AUC value is greater than 0.7, the closer it is to 1, the more accurate the model's prediction. The value of AUC can be computed by the trapezoidal rule of integral calculus, as shown in Equation (7).

$$AUC = \sum_{p=1}^n (X_{p+1} - X_p) \times (S_{p+1} - S_p - S_p/2) \quad (7)$$

where  $X_p$  is specificity and  $S_p$  is sensitivity.

#### 2.3.2. Landslide Hazard Assessment Method

The susceptibility assessment is only aimed at the analysis and evaluation of static factors, without considering the dynamic factors that affect the occurrence of landslides. Therefore, based on the landslide susceptibility assessment, this paper incorporates the external dynamic factors mainly based on average rainfall and human engineering activities

over many years to realize the hazard assessment for landslide disaster in Fengjie County. The calculation formula is:

$$H = S \times (w_1 + w_2 + \dots + w_n) \quad (8)$$

$H$  is the landslide hazard index;  $S$  is the regional landslide susceptibility value; and  $w_1, w_2, \dots, w_n$  are the normalized risk assessment factors.

### 2.3.3. Landslide Vulnerability Assessment Method

The vulnerability assessment model for landslides mainly considers the disaster-bearing body. It refers to objects that suffer from landslide disasters, such as human beings, property, resources or the ecological environment [37]. Within the hazard range, evaluating the damage and the degree of damage that the hazard-bearing body may produce when suffering from a landslide disaster is known as vulnerability evaluation. After considering the characteristics of the study area and the difficulty of data acquisition, three types of vulnerability assessment indicators, material, social and economic, are selected. Among these, material vulnerability refers to POI density and road cost in the county. Social vulnerability considers population density in the county, and economic vulnerability mainly considers GDP. The calculation formula is:

$$V = E + M + C \quad (9)$$

In the formula,  $V$  represents the vulnerability of the disaster-bearing body;  $M$ ,  $C$ , and  $E$  represent material, community, and economic vulnerability, respectively. Considering that the importance of these three parts is indistinguishable, the weights account for 1/3 each.

### 2.3.4. Landslide Risk Assessment Method

Risk refers to the expected value of loss of human life, property, and economic and social activities due to a certain natural disaster in a certain area and period. A landslide is a natural phenomenon, but if it threatens human society, then it is a disaster. In 1984, Varnes [8], a well-known landslide expert in the United States, proposed a basic definition of geological hazard risk, which was universally recognized. Landslide risk is the study of the possibility of losses caused by landslide damage, including the possibility of disasters and the magnitude of losses. Based on the above concept of 'risk' and the definition by scholars such as Varnes (1984), Einstein (1988) [44], Fell (1994) [45], the product of hazard and vulnerability is generally used as the value of landslide risk [31,46]:

$$\text{Risk} = \text{Hazard} \times \text{Vulnerability} \quad (10)$$

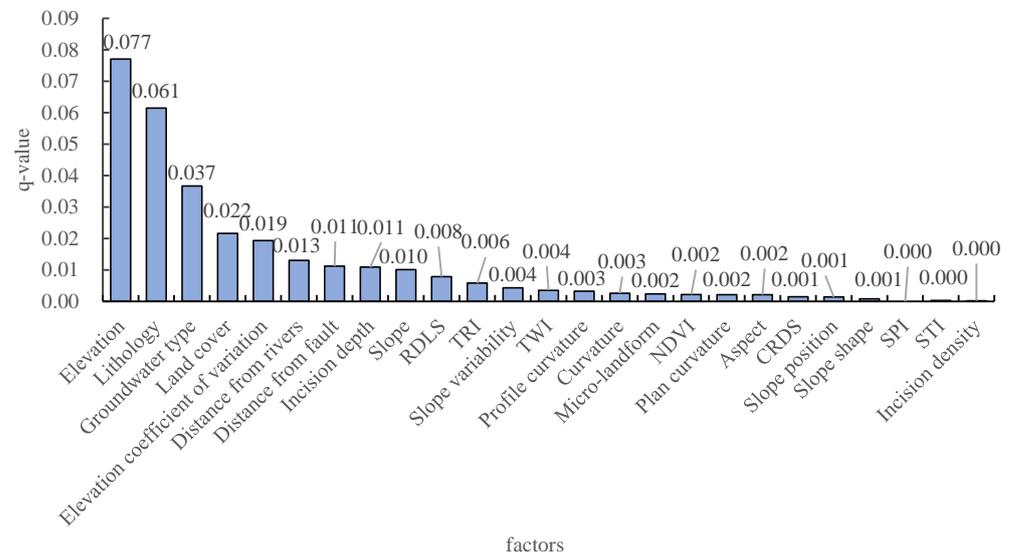
Hazard reflects the natural attribute of landslide, and vulnerability reflects the social attribute of landslide. Through the calculation of this formula, the randomness and uncertainty of landslide occurrence and development are included, reflecting the close relationship between nature and human society.

## 3. Results

### 3.1. Results of Landslide Susceptibility

The factor detector based on Geodetector screened the influencing factors and obtained the detection results (Figure 10).  $q$  value explains the contribution rate of the factor, namely the degree of influence degree of the factor on the landslide. The results show that elevation, lithology, groundwater type, land cover, incision depth, elevation coefficient of variation, distance from rivers, distance from the fault, slope, RDLS, TWI, TRI, slope variability, plane curvature, curvature, micro-landform, NDVI, profile curvature and aspect are relatively important. Among these, elevation has the strongest explanatory power for the occurrence of landslides, while the  $q$  value of the CRDS, slope position, slope shape, SPI, STI and incision density are less than 0.001, which does not have explanatory power, indicating

that the relationship with the occurrence of landslides in the study area is very limited. Therefore, this paper conducted a landslide comprehensive susceptibility evaluation based on the above 19 factors ( $q \geq 0.002$ ).



**Figure 10.** Factor detector results.

Based on 1522 historical landslides in the study area, the 500 m buffer zone and the river area are excluded as non-landslide areas. Since the number of training samples will directly affect the training accuracy, this paper constructs the modeling data set according to a ration of landslide (1522):non-landslide (15,220), i.e., 1:10. Five-fold cross-validation is used to reduce the impact of a single sampling method. The basic principle is that all data sets (1522 landslides and 15,220 non-landslides) are randomly and averagely divided into five disjoint subsets, one subset used for each test, and the remaining subsets used for model training. As shown in Table 4, the average accuracy of RF model training and test samples are 0.976 and 0.913, respectively. Among them, sample 3 has the highest test accuracy (0.919). Therefore, the model constructed with this sample is used for the simulation of global landslide comprehensive susceptibility.

**Table 4.** The accuracy of five-fold cross-validation.

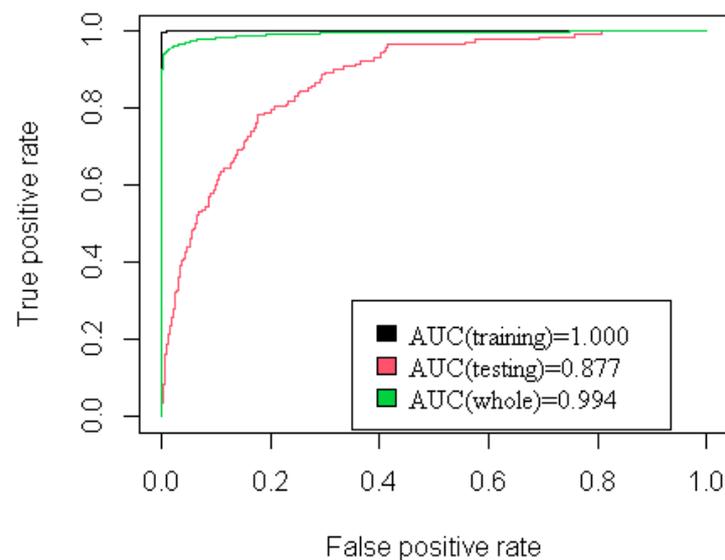
Subset	Accuracy	
	Training	Testing
1	0.977	0.908
2	0.977	0.917
3	0.975	0.919
4	0.975	0.904
5	0.976	0.918
Average	0.976	0.913

For the binary classification problem (landslide 1, non-landslide 0), the confusion matrix is often used to analyze the prediction accuracy. The confusion matrix of all data sets of the random forest model is given in Table 5, and is classified by using the library 'Information Value' to select a better threshold, instead of the traditional threshold of 0.5. If the predicted value is greater than the threshold, landslide will occur, and vice versa. It can be seen that the overall accuracy of the RF model is 0.991, the prediction accuracy of landslide and non-landslide are 0.997 and 0.939, and the sensitivity and specificity are 0.930 and 0.997, respectively. The results show that the RF model has good prediction performance.

**Table 5.** Confusion matrix of landslide comprehensive susceptibility.

RF		True Condition		Summation
		Landslide	Non-Landslide	
Prediction Condition	Landslide	1416 (TP)	40 (FP)	Precision: 0.997
	Non-landslide	106 (FN)	15,180 (TN)	Precision: 0.939
Summation		Recall: 0.930	Recall: 0.997	Accuracy: 0.991

Additionally, the results of landslide comprehensive susceptibility constructed by the RF model can also be tested by the receiver operating characteristic (ROC). Area under the ROC curve can quantitatively test the accuracy of the model prediction. In this study, R language was used to perform ROC curve analysis in R Studio software. The AUC values of training, testing, and all samples were 1.000, 0.877 and 0.994 (Figure 11), respectively. Especially, the AUC value of testing is greater than 0.7, which means the model has high accuracy and reliability.

**Figure 11.** ROC curve and AUC value.

The RF model can learn, after training the sample data. It can be applied to the geospatial database of the whole study area, and then the probability value of each grid pixel landslide (0~1) can be obtained. Then, according to the expert experience method [47], the susceptibility results are divided into five grades: very low, low, medium, high, and very high (Figure 12). Very low and low susceptibility level areas indicate that landslide disasters are not easy to occur under basic topographic and geological conditions. The medium level area indicates that landslide disasters are more likely to occur under basic topographic and geological conditions. High level areas indicate that landslide disasters are prone to occur. Very high susceptibility indicates that landslide disasters are easy to occur.

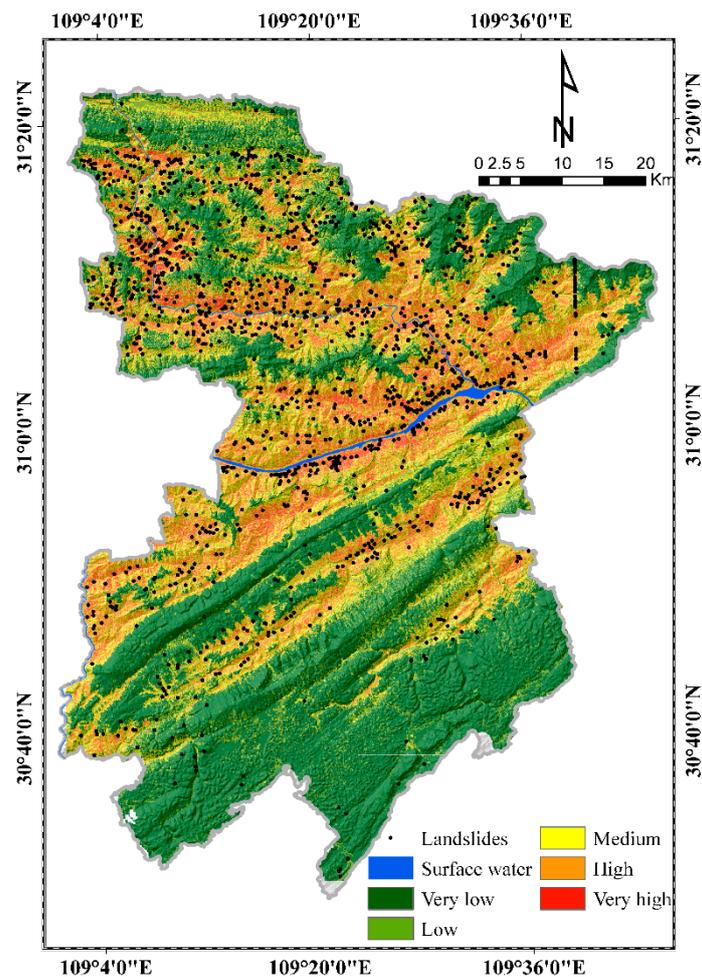


Figure 12. Landslide comprehensive susceptibility map.

To quantitatively analyze the defined susceptibility mapping result, the grid number, area proportion, landslide number and landslide density of each susceptibility grade are counted, as shown in Table 6. It can be seen that areas of high and very high susceptibility account for 2% of the total area of Fengjie County, and the number of landslides accounts for 89.94% of the total. Areas of low and very low susceptibility are more than half of the county area, accounting for 53%, and the number of landslides is only 3.61% of the total. The area of medium-prone area accounts for 19.28%, and the proportion of landslides in this area is 6.44%. Overall, with an increase in susceptibility grade, the smaller the area ratio and the larger the proportion of landslides. There is a significant positive correlation between the number of historical landslides and the susceptibility level, and the area and the proportion of landslides in the susceptibility area are also at a reasonable level. The landslide comprehensive susceptibility mapping based on the RF method is consistent with the actual situation.

Table 6. Statistical results for landslide comprehensive susceptibility in different classes.

Landslide Probability	Susceptibility Class	Grid Number	Area Proportion	Landslide	Landslide Proportion
<0.16	Very low	1,775,732	39.41%	16	1.05%
0.16–0.23	Low	635,784	14.11%	39	2.56%
0.23–0.31	Medium	868,521	19.28%	98	6.44%
0.31–0.41	High	922,251	20.47%	269	17.67%
>0.41	Very high	303,324	6.73%	1100	72.27%

### 3.2. Results of Landslide Hazard

Based on the ArcGIS10.4 platform, the above landslide comprehensive susceptibility result and hazard factors are superimposed and calculated according to Formula (8). The results were divided into five grades according to the natural breakpoint method as shown in Figure 13, namely, very low, low, medium, high and very high, and the landslide hazard map of Fengjie County was obtained, because the natural breakpoint method is a statistical method for classification based on numerical statistical distribution and can ensure the categories have relative consistency. In principle, there are some natural turning points and feature points in any statistical sequence that can be used to divide the object of study into different groups, which means the difference between the same category of data is the smallest, and the difference between different categories of data is the largest [48]. Different hazard levels represent the possibility of landslides over a short time. The higher the grade, the greater the risk of landslide. The hazard level of Fengjie County shows an obvious spatial distribution, and the very low and low hazard areas are mostly distributed in the high-altitude mountainous areas in the south and southeast. High and very high hazard areas are concentrated along rivers and central towns. The medium hazard area is distributed in the low mountains outside of the high hazard area.

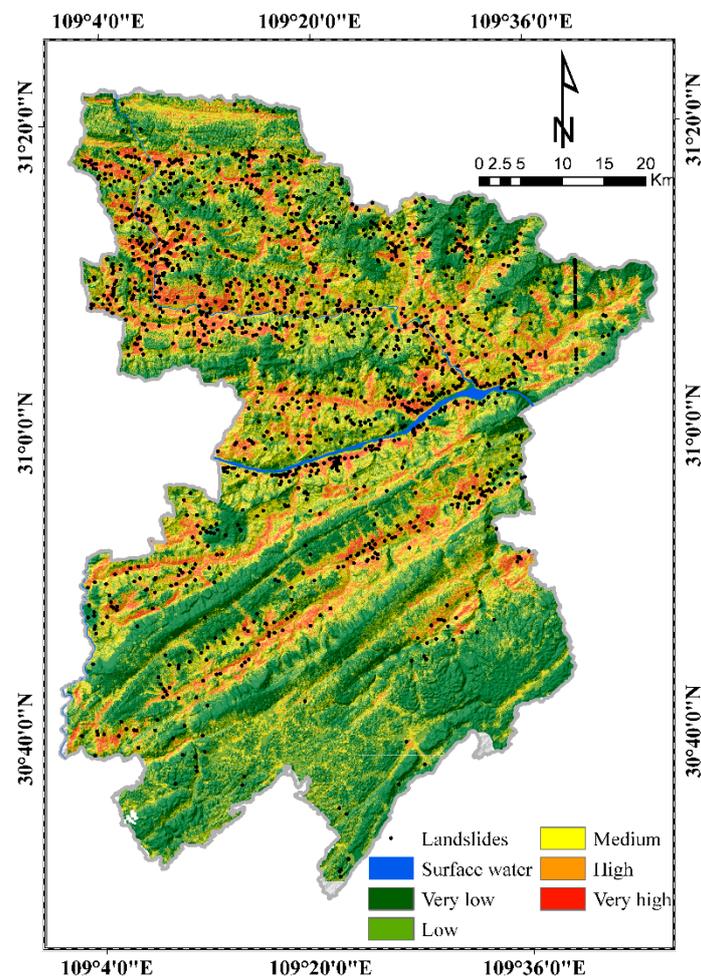


Figure 13. Landslide hazard map.

To further analyze the hazard grade division, the ArcGIS10.4 software tool is used to count the number of grids, the percentage of grids, the number of landslides and other data in each division, as shown in Table 7.

**Table 7.** Statistical result of landslide hazard in different classes.

Hazard Class	Grid Number	Area Proportion	Landslide	Landslide Proportion
Very low	1,380,163	30.63%	19	1.25%
Low	1,321,706	29.33%	76	4.99%
Medium	963,268	21.38%	211	13.86%
High	585,668	13.00%	395	25.95%
Very high	242,624	5.38%	821	53.94%

According to the statistical results from the above table, the number of grid units in the very low hazard area compared to the very high hazard area decreases in turn, and the grid percentage decreases from 30.63% to 5.38%. The areas with low hazards and below account for 59.97% of the county area, indicating that more than half of the areas in the county have a small probability of landslide under the influence of natural and human activities. The high and very high hazard area accounts for 18.38% of the total area, but it contains 79.89% of the landslides, indicating that the distribution range of landslides in the county is relatively concentrated, which is consistent with the actual situation. The density of landslides increased by about 250 times (from 0.015 to 3.760) in the process of evaluation grading from very low to very high. There was a significant positive correlation between the density of landslides and the hazard degree.

### 3.3. Results of Landslide Vulnerability

Based on the grid calculator of ArcGIS 10.4, the four vulnerability factors of POI kernel density, population, GDP and road cost are superimposed and calculated to obtain the vulnerability results for the study area. The study area is still divided into very low, low, medium, high and very high vulnerability areas by using the natural breakpoint method, and the map for Fengjie County landslide disaster vulnerability is obtained by mapping and synthesis (Figure 14).

The grid cell number, area and area ratio of each vulnerability partition in the study area were statistically analyzed, as shown in Table 8. The area of very low and low landslide vulnerability areas in the study area is 4001.49 km<sup>2</sup>, accounting for about 99.48% of the total. Most of these areas are uninhabited or far from the main roads. The medium, high and very high vulnerability areas accounted for only 0.52% of the total area, and this huge difference in number is because in Fengjie County, due to its mountainous terrain hindering economic development, the overall GDP level is low, the population density is small, and housing and transportation facilities are still relatively lacking. The very high vulnerability areas are almost all concentrated in the central urban area of Fengjie County, where the urban and regional population density is relatively large, and infrastructure such as housing and factories are built. Once these areas slide, people's lives and property will be seriously damaged.

**Table 8.** Statistical result of landslide vulnerability in different classes.

Vulnerability Class	Grid Number	Area Proportion	Area (km <sup>2</sup> )
Very low	3,147,329	70.42%	2832.60
Low	1,298,767	29.06%	1168.89
Medium	12,899	0.29%	11.61
High	7481	0.17%	6.73
Very high	2580	0.06%	2.32

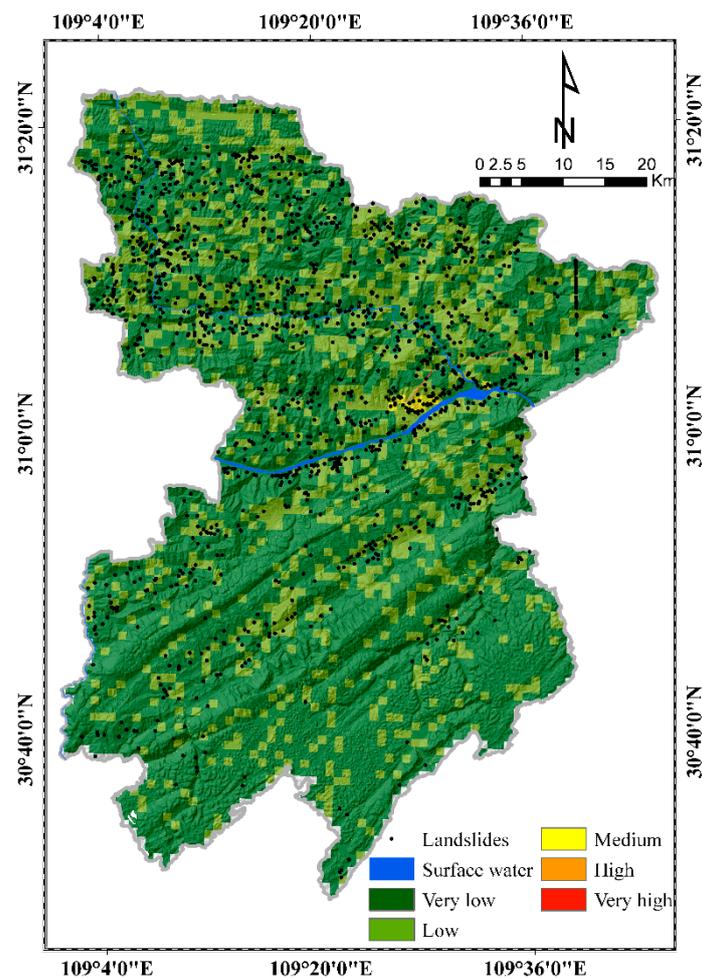


Figure 14. Landslide vulnerability map.

### 3.4. Results of Landslide Risk

Based on the results of landslide hazard and vulnerability assessment, the grid superposition calculation was carried out. Formula (10) was used to calculate the landslide risk value of Fengjie County, and this was divided into five risk areas by the natural breakpoint method: very low, low, medium, high and very high. A regional landslide risk map is obtained by cartographic generalization (Figure 15).

According to the landslide risk map (Figure 15), natural factors such as topography, environmental conditions, meteorological hydrology and social factors such as population and economic factors give the landslide risk map of Fengjie County a certain regularity. Table 9 shows the statistical results of the landslide risk map. The very low and low-risk area is 2949.17 km<sup>2</sup>, accounting for 73.71% of the study area, indicating that the risk level of most areas is below middle risk, and most of these areas are distributed at high altitude, above the mountains, belong to the unpopulated area, where human activities are weak, even if there is a low risk landslide. The medium-risk area accounts for about 23.79%, which is mostly distributed along the valleys and rivers. Villages are scattered in these areas, and the villagers suffer great landslide risks. Very high and high areas accounted for 2.5% of the study area, concentrated in the central city of Fengjie County, and these areas are located along the Three Gorges Reservoir Area, densely populated, and where building density is larger. Under the influence of extreme weather such as reservoir water fluctuation and heavy rainfall, landslides are prone to occur, which is more likely to cause major casualties and property losses. Therefore, the risk degree in these regions is also high.

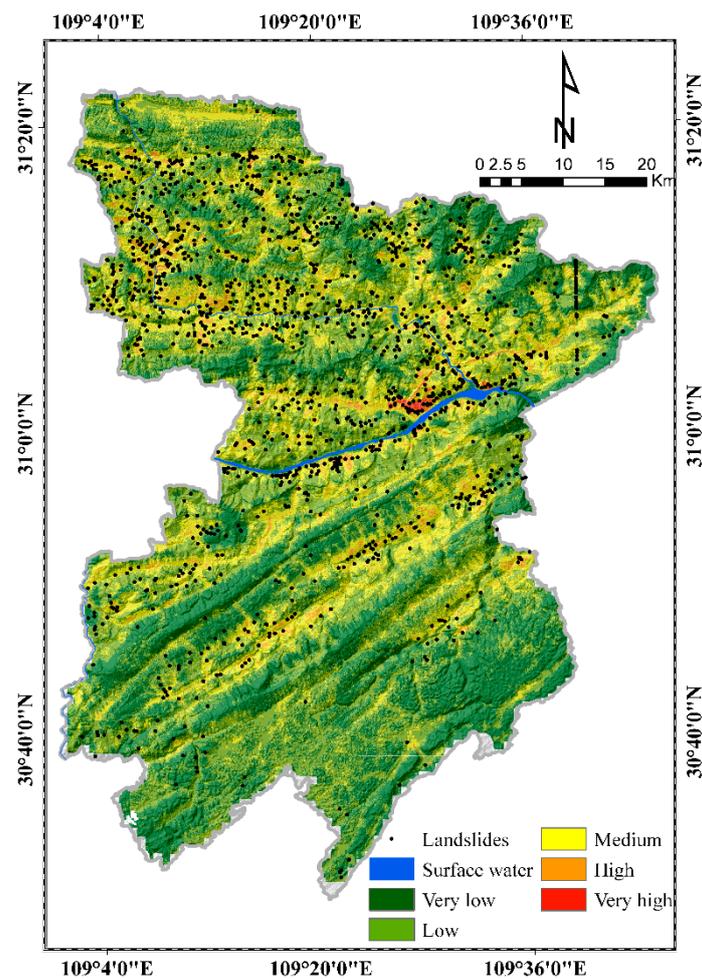


Figure 15. Landslide risk map.

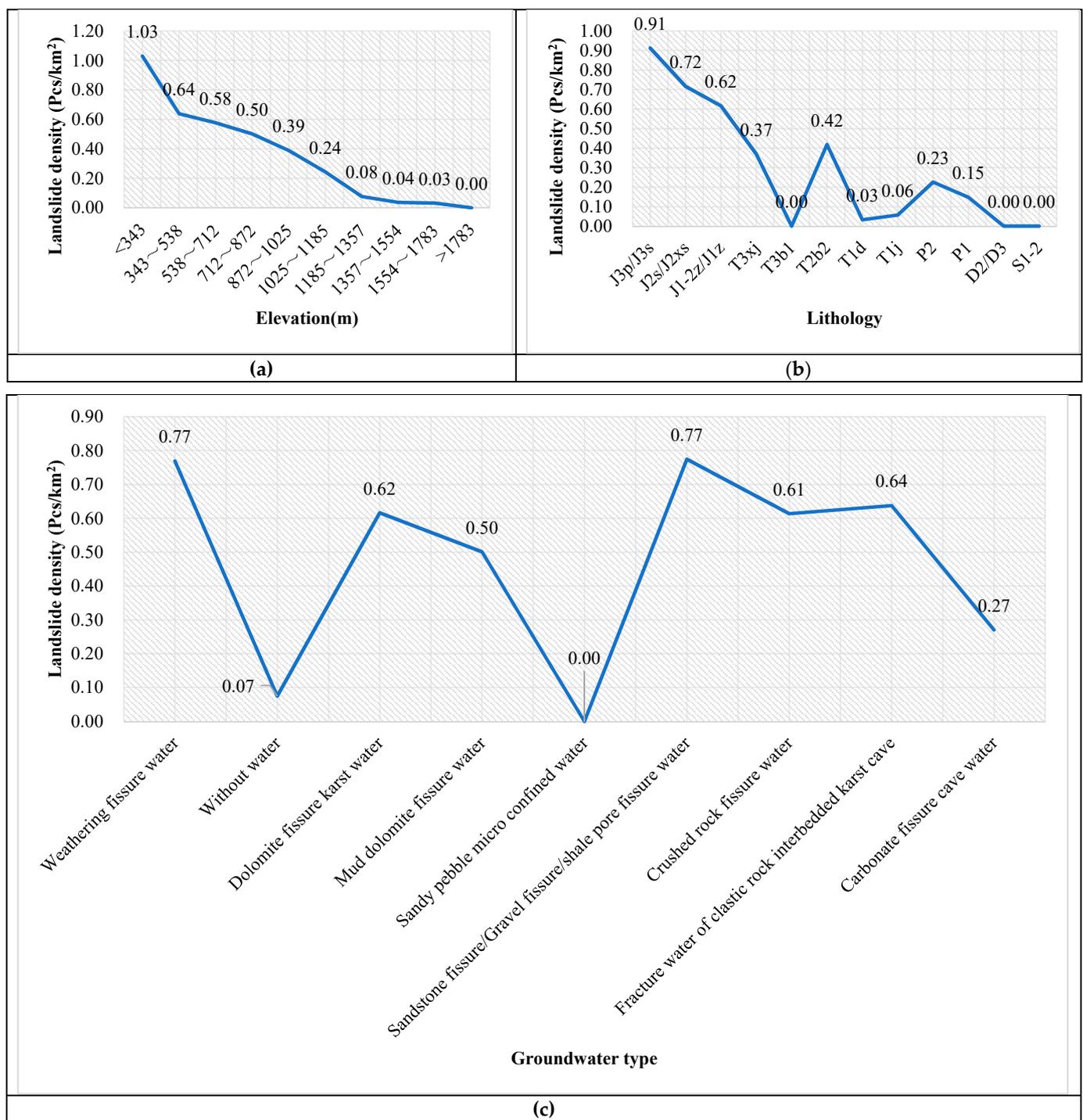
Table 9. Statistical result of landslide risk in different classes.

Risk Class	Grid Number	Area Proportion	Area (km <sup>2</sup> )
Very low	1,341,361	30.17%	1207.22
Low	1,935,495	43.54%	1741.95
Medium	1,057,730	23.79%	951.96
High	94,178	2.12%	84.76
Very high	16,714	0.38%	15.04

## 4. Discussion

### 4.1. Importance of Contributing Factors

Effective and contribution factors play an important role in landslide research. Analyzing the contribution rate and influence of each factor on landslide occurrence and identifying the dominant factors can provide important guidance for landslide disaster prediction and prevention. Therefore, based on the GeoDetector, we have given the  $q$ -value statistical results of 25 landslide factors (Figure 10). To better analyze the relationship between factors and landslides, a statistical chart of historical landslide density ranking with the top three factors of  $q$  value was drawn: elevation, lithology, and groundwater type (Figure 16).



**Figure 16.** Typical factors in landslide density statistics: (a) Elevation; (b) Lithology; (c) Groundwater type.

It can be seen from Figure 16a that the landslide density is negatively correlated with elevation: the landslide density is higher at lower elevation. Fengjie County is located in a typical mountain environment, with high difference and low altitude. The low altitude area has flat terrain, fertile soil, and is close to water sources, which is convenient for human beings to carry out economic and life activities. Human engineering activities are frequent. Therefore, landslides occur frequently. The area with higher altitude has steep terrain, inconvenient transportation and less human activity, so there are fewer landslides.

As an important internal cause of landslides, different lithologic characteristics contribute to great differences in physical and mechanical parameters, which directly affect

slope stability. Fengjie County has many types of lithology, mainly including Jurassic (J3p, J2s, J1, etc.) and Triassic (T1d, T1j, T2b2, etc.). According to the statistical results in Figure 16b, since the lithological geology of Jurassic soft–hard interphase strata has unique characteristics, the landslide-intensive areas are mostly distributed in this region. The soft–hard interphase structure formed by sandstone and mudstone is unstable, which is a common type of sliding bed structure in China. It is widely distributed in the counties of the Three Gorges Reservoir Area and even in the eastern part of Sichuan.

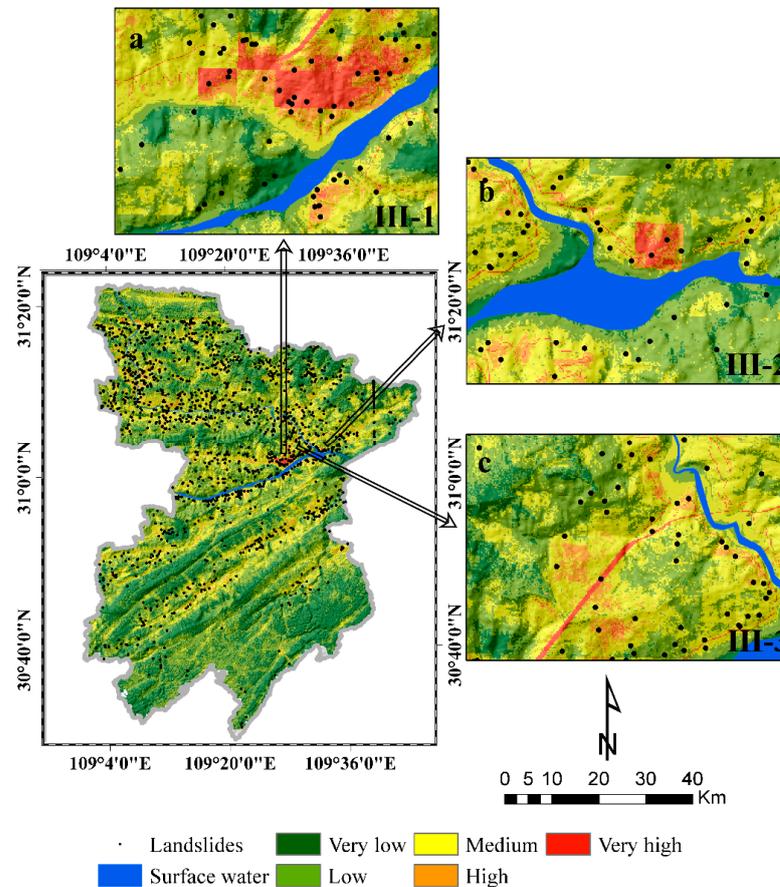
Figure 16c shows that that different types of groundwater have a significant impact on the stability and deformation of landslides. Among them, the landslide density of weathering fissure water, dolomite fissure karst water, sandstone fissure/gravel fissure/shale pore fissure water and other groundwater types is larger. Because of the special geological environment of these groundwater, sandstone, sandy conglomerates, carbonate and shale are characterized by weathering disintegration and interaction of soft and hard rock layers. Groundwater has developed underground, and mudstone with weak resistance to rainwater erosion and weathering is easy to collapse and form cavities. Sandstone is more prone to instability and failure due to the cutting effect of the structural plane, resulting in collapse and landslide, which seriously affects the stability and durability of the slope.

#### 4.2. Risk Prevention Zoning

According to the results of the risk map, the landslide prevention and control area is divided into very low and low-risk area as the general prevention and control area, medium risk area as the sub-key prevention and control area, and high and very high-risk level area as the key prevention and control area. The results are shown in Table 10. The general prevention and control areas are mainly located in scenic spots, nature reserves and mountains, accounting for 73.71% of the total area. The landslide density is only 0.07/km<sup>2</sup>, and the risk level is low. Sub-key control areas are mostly located in the valley in the transition landslide zone unstable slope area, accounting for 23.79% of the area, landslide density 0.82/km<sup>2</sup>; compared with the general control area, this is a nearly twelve-fold increase. As shown in Figure 17, the key prevention and control areas are mainly divided into three sub-regions. The key prevention and control sub-region (III-1) is located in the landslide group at the eastern Yonggan Town and the western Zhuyi Town (Figure 17a). The region is located in the central urban area of Fengjie County, with a high density of buildings and population. The county town is built along the river, near high mountains, steep slopes and fewer rocks. Secondly, the probability of slope instability will greatly increase due to the excavation of mountain slopes by human engineering activities, resulting in high landslide risk. The key prevention and control sub-region (III-2) includes the Chenjiabao landslide, Guanmiaotuo landslide and Linjiawan landslide (Figure 17b). These three landslides are concentrated in the vicinity of schools and shops. Once the slope slides again, the loss will be extremely serious. In Figure 17c, the landslide risk of the key prevention and control sub-area (III-3) in the central urban section of Hurong Expressway is also very high. Fengjie County is a typical mountainous terrain. The construction of a mountainous expressway will inevitably fill and excavate a large number of slopes along the line, and destroy the slope.

**Table 10.** Division of landslide disaster prevention in Fengjie County.

Division Name and Code	Subregion Name and Code	Area (km <sup>2</sup> )	Area Proportion	Landslide	Landslide Proportion	Landslide Situation
General Control Area (I)	Scenic spots, nature reserves and other general prevention and control areas of high mountains and extremely high mountains	2949.17	73.71%	208	0.07%	It belongs to the very low and low-risk area of landslide disaster. The landslide density is small and the risk level is low.
Sub-focus Areas (II)	Subkey control area of landslide and unstable slope in the transition zone of a river valley	951.96	23.79%	785	0.82%	It belongs to the risk area of the landslide; the landslide density is large and the risk level is high.
Key Control Areas (III)	Key prevention and control subregion of landslide group at the junction of eastern Yongan town and western Zhuyi town (III-1)	5.35	0.13%	29	5.42%	It belongs to the very high-risk area of landslide, with high landslide density and high-risk level.
	Chenjiabao landslide, Guanmiaotuo landslide and Linjiawan landslide key prevention and control sub-region (III-2)	0.98	0.02%	3	3.06%	It belongs to very high-risk area, with big landslide density and high-risk level.
	Key Prevention and Control Sub-districts in the Central Urban Section of Hurong Expressway (III-3)	1.45	0.04%	6	4.13%	It belongs to very high-risk area, with large landslide density, and high-risk level.



**Figure 17.** Key prevention area for landslide hazard. (III-1) in the eastern Yongan Town and the western Zhuyi Town, (III-2) in the Chenjiabao landslide, Guanmiaotuo landslide and Linjiawan landslide, (III-3) in the central urban section of Hurong Expressway.

### 4.3. Contributions and Shortcomings

In this study, we adopted the natural breakpoint method to classify the results. In fact, there are different methods to classify a map: e.g., quantile, standard deviation, geometric interval, etc. Since the elements are grouped into each class in the same number by the quantile classification method, the maps obtained are often misleading. Similar elements may be placed in adjacent classes, or elements with large differences in values may be placed in the same class. The standard deviation classification method is used to display the difference between the attribute value and the average value of the elements. The disadvantage is that it is vulnerable to the influence of two extreme values. A geometric interval classification scheme is used to create a classification interval according to the group distance, with geometric series. As one of the most commonly used classification methods, the natural breakpoint method can maximize the difference between classes. The elements will be divided into multiple categories and their boundaries will be set at positions where the data values are relatively different, so as to achieve the best classification results.

In landslide risk assessment, the distance from roads and houses are closely related to the landslide occurrence. The construction of massive roads and houses is a process whereby humans transform the natural environment, which includes transportation, erosion, and accumulation of surface soil. Excessive digging, application of external loads and vegetation destruction lead to steep slopes and loose soil. Finally, precipitation and earthquakes can trigger landslides. According to the statistics, with an increase in distance from the road, landslide density gradually decreases, and there is a significant negative correlation between the two. The area within 200 m from the road is the high landslide occurrence area. Landslide density within 400 m away from the houses is significantly negatively correlated with the house distance. The highest density is 100 m away from the houses because of the great damage to the soil caused by various human development activities, which increases the probability of landslide.

Although this study provides a relative contribution to landslide risk assessment, there are some limitations. Firstly, the Geodetector has a premise hypothesis regarding landslide influencing factors, that is, there should be strong spatial heterogeneity among factors, while the heterogeneity of factors such as land cover and CRDS is usually small in adjacent units. In this case, the model may not fully fit the spatial heterogeneity of this type of factor [49].

Secondly, the evaluation unit used in this study is a grid unit; such a unit may contain multiple landslides or a landslide may be shared by several adjacent grids, that is, a grid may not represent a specific landslide. After using the Geodetector, the grid unit can be improved to some extent, but the problem of the landslide evaluation unit has not been fundamentally solved. Some studies have compared different mapping units to test the spatial scale effect of mapping, but there is no general method to obtain the optimal mapping unit [50,51]. Therefore, to further improve the accuracy of landslide related models, more reasonable evaluation units will be explored in subsequent studies.

Thirdly, the outcomes of landslide susceptibility mapping could be subject to uncertainties, despite the fact that the RF model has good prediction performance and results [14]. In this study, factor selection, hyper-parameter optimization in the model, data used, sample fraction, etc., may be the main sources of uncertainty. We will attempt to explore other methods to minimize uncertainties and improve landslide predictions. In vulnerability evaluation, the four most important factors are selected in this study, but the factors considered are not comprehensive, which may become one of the sources of uncertainty. In addition, it should be noted that GDP and population create the problem of data matching at different precisions, which will affect the accuracy of the results. Hence, the reliability of the selected data must be improved.

## 5. Conclusions

Taking Fengjie County in the Three Gorges Reservoir Area as the research region, this paper studied quantitative risk assessment of landslides based on susceptibility mapping using random forest and GeoDetector. The main conclusions are drawn as follows:

(1) 19 dominant factors, such as elevation, lithology, groundwater type, etc., were selected, by using Geodetector, as susceptibility assessment factors, and the landslide comprehensive susceptibility assessment model was established based on the RF model. Secondly, the annual average rainfall and human engineering activities were determined as risk assessment indexes, and the hazard assessment model of Fengjie County was established based on GIS software and the weighted superposition method. At the same time, the vulnerability evaluation factors include POI kernel density, population, GDP and road cost. The landslide vulnerability assessment model was established based on GIS grid technology and weight superposition method. On this basis, the distribution of Fengjie County landslide and quantitative risk assessment research was developed.

(2) Most landslides in Fengjie County are distributed on both sides of the reservoir bank and the primary and secondary tributaries, showing a spatial distribution pattern which is more north than south. In terms of quantity, landslide risk is mostly in the low-risk level, a small part of the high risk or very high-risk level. Very low and low-risk areas accounted for the largest proportion, 73.7%, of the study area. The middle, high and very high-risk areas accounted for 23.79% and 2.5%, respectively. From the perspective of spatial pattern, the overall risk level shows the high spatial distribution characteristics in the central and eastern urban areas, and low in the southern and northern high-altitude areas. Because very low and low-risk areas are mostly distributed above the mountains, human activities are weak, even if there is a low risk landslide. The middle-risk areas are mostly located near scattered villages along valleys and rivers, so the villagers suffer more landslide risk. High and very high-risk areas are located along the Three Gorges Reservoir, concentrated in the central city, densely populated and full of buildings. Under the influence of extreme weather such as reservoir water lifting and heavy rainfall, once landslides occur they cause serious casualties and property losses. The results of risk zoning are in line with the actual situation of Fengjie County, which can provide a basis for disaster prevention and mitigation and land space planning in the study area.

(3) The importance of the results for different landslide conditioning factors are in line with basic geological laws and the regional characteristics. Elevation, lithology, and groundwater type are the main factors. Secondly, the general, sub-key and key prevention and control areas were divided, and the landslide prevention and control management targeted. The general control area accounted for 73.71%, and the landslide density was  $0.07/\text{km}^2$ , which is widely distributed and has low risk. Sub-key control areas are mostly located in the valley to the transition zone landslide, an unstable slope area, accounting for 23.79% of the area, and landslide density of  $0.82/\text{km}^2$ ; compared with the general control area, this is an increase of nearly 12 times. The key prevention and control areas are divided into three sub-regions, which are mostly located around the central towns, landslides and highways. The landslide density can reach 3–5 places/ $\text{km}^2$ . It is necessary to strictly prevent and control landslide disasters. Starting from two aspects of prevention and control, according to the difference of risk levels in various regions, zoning management is carried out according to local conditions. This study is helpful for all levels of management departments to make timely and accurate disaster prevention and mitigation decisions, and to provide decision-making information for regional urban planning, land resources development, land use development and social and economic sustainable development.

**Author Contributions:** Conceptualization, D.S. and H.W.; methodology, Y.W.; formal analysis, Y.W.; resources, H.W., D.S. and Y.L.; writing—original draft preparation, Y.W.; supervision, D.S., H.W. and Y.L.; validation, Y.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Key research and development program of the Ministry of science and technology, grant number 2018YFC1505501 and the Fundamental Research Funds for the Central Universities, grant number 2021CDJKYJH036.

**Data Availability Statement:** The data presented in this study can be available on request from the corresponding author.

**Acknowledgments:** The authors would like to thank Wen (at Chongqing University) for providing the scripts used for data processing with the Random Forest technique, and also thank Sun and Li for patient guidance. The authors are also grateful to the editor and anonymous reviewers for their positive comments on the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hong, H.; Pourghasemi, H.R.; Pourtaghi, Z.S. Landslide susceptibility assessment in Lianhua County (China): A comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology* **2016**, *259*, 105–118. [[CrossRef](#)]
2. Huang, Y.; Zhao, L. Review on landslide susceptibility mapping using support vector machines. *Catena* **2018**, *165*, 520–529. [[CrossRef](#)]
3. Petley, D. Global patterns of loss of life from landslides. *Geology* **2012**, *40*, 927–930. [[CrossRef](#)]
4. Acharya, S.; Pathak, D. Landslide hazard assessment around MCT zone in Marsyangdi River basin, west Nepal. *J. Nepal Geol. Soc.* **2017**, *53*, 93–98. [[CrossRef](#)]
5. Aleotti, P.; Chowdhury, R. Landslide hazard assessment: Summary review and new perspectives. *Bull. Eng. Geol.* **1999**, *58*, 21–44. [[CrossRef](#)]
6. Salvati, P.; Bianchi, C.; Rossi, M.; Guzzetti, F. Societal landslide and flood risk in Italy. *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 465–483. [[CrossRef](#)]
7. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Bui, D.T.; Duan, Z.; Ma, J. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* **2017**, *151*, 147–160. [[CrossRef](#)]
8. Varnes, D.J. *Landslide Hazard Zonation: A Review of Principles and Practice*; Unesco: Paris, France, 1984.
9. Biçer, Ç.T.; Ercanoglu, M. A semi-quantitative landslide risk assessment of central Kahramanmaraş City in the Eastern Mediterranean region of Turkey. *Arab. J. Geosci.* **2020**, *13*. [[CrossRef](#)]
10. Bonachea, J.; Remondo, J.; de Teran, J.R.; Gonzalez-Diez, A.; Cendrero, A. Landslide risk models for decision making. *Risk Anal.* **2009**, *29*, 1629–1643. [[CrossRef](#)]
11. Yong, X.; Zhipeng, L.; Chunying, G. Risk Study on quantitative risk analysis of Rainfall Landslide—A Case Study of Ganba Landslide in Xuanen County. *Geol. Miner. Resour. South China* **2018**, *34*, 294–301. [[CrossRef](#)]
12. Nianchang, Z.; Bolong, L.I.U.; Junning, X.I.E. Numerical Investigation on the Barrier Dam Risk Caused by Landslide—A Case Study on Caijiaba Landslide. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *658*. [[CrossRef](#)]
13. Michellier, C.; Pigeon, P.; Paillet, A.; Trefon, T.; Dewitte, O.; Kervyn, F. The Challenging Place of Natural Hazards in Disaster Risk Reduction Conceptual Models: Insights from Central Africa and the European Alps. *Int. J. Disaster Risk Sci.* **2020**, *11*, 316–332. [[CrossRef](#)]
14. Adnan, M.S.G.; Rahman, M.S.; Ahmed, N.; Ahmed, B.; Rabbi, M.F.; Rahman, R.M. Improving Spatial Agreement in Machine Learning-Based Landslide Susceptibility Mapping. *Remote Sens.* **2020**, *12*, 3347. [[CrossRef](#)]
15. Can, A.; Dagdelenler, G.; Ercanoglu, M.; Sonmez, H. Landslide susceptibility mapping at Ovacık-Karabük (Turkey) using different artificial neural network models: Comparison of training algorithms. *Bull. Eng. Geol. Environ.* **2017**, *78*, 89–102. [[CrossRef](#)]
16. Du, C.; Yi, Q.; Zhou, B.; Qin, S.; Zeng, H. Evaluation of landslide susceptibility in Yunyang County of Three Gorges Reservoir Area Based on GIS and weighted information. *J. China Three Gorges Univ. (Nat. Sci.)* **2017**, *39*, 48–53.
17. Steger, S.; Brenning, A.; Bell, R.; Glade, T. The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. *Landslides* **2017**, *14*, 1767–1781. [[CrossRef](#)]
18. Ahmad, H.; Ningsheng, C.; Rahman, M.; Islam, M.M.; Pourghasemi, H.R.; Hussain, S.F.; Habumugisha, J.M.; Liu, E.; Zheng, H.; Ni, H.; et al. Geohazards Susceptibility Assessment along the Upper Indus Basin Using Four Machine Learning and Statistical Models. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 315. [[CrossRef](#)]
19. Wang, Y.; Sun, D.; Wen, H.; Zhang, H.; Zhang, F. 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* **2020**, *17*, 4206. [[CrossRef](#)]
20. Lombardo, L.; Mai, P.M. Presenting logistic regression-based landslide susceptibility results. *Eng. Geol.* **2018**, *244*, 14–24. [[CrossRef](#)]
21. Zhao, Y.; Wang, R.; Jiang, Y.; Liu, H.; Wei, Z. GIS-based logistic regression for rainfall-induced landslide susceptibility mapping under different grid sizes in Yueqing, Southeastern China. *Eng. Geol.* **2019**, *259*. [[CrossRef](#)]

22. Tian, Y.; Xu, C.; Hong, H.; Zhou, Q.; Wang, D. Mapping earthquake-triggered landslide susceptibility by use of artificial neural network (ANN) models: An example of the 2013 Minxian (China) Mw 5.9 event. *Geomat. Nat. Hazards Risk* **2018**, *10*, 1–25. [[CrossRef](#)]
23. Zhou, C.; Yin, K.; Cao, Y.; Ahmed, B.; Li, Y.; Catani, F.; Pourghasemi, H.R. Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China. *Comput. Geosci.* **2018**, *112*, 23–37. [[CrossRef](#)]
24. Huang, F.; Yao, C.; Liu, W.; Li, Y.; Liu, X. Landslide susceptibility assessment in the Nantian area of China: A comparison of frequency ratio model and support vector machine. *Geomat. Nat. Hazards Risk* **2018**, *9*, 919–938. [[CrossRef](#)]
25. Sun, D.; Xu, J.; Wen, H.; Wang, Y. An Optimized Random Forest Model and Its Generalization Ability in Landslide Susceptibility Mapping: Application in Two Areas of Three Gorges Reservoir, China. *J. Earth Sci.* **2020**, *31*, 1068–1086. [[CrossRef](#)]
26. Lin, J.-W.; Hsieh, M.-H.; Li, Y.-J. Factor analysis for the statistical modeling of earthquake-induced landslides. *Front. Struct. Civ. Eng.* **2019**, *14*, 123–126. [[CrossRef](#)]
27. Chang, S.-H.; Wan, S. Discrete rough set analysis of two different soil-behavior-induced landslides in National Shei-Pa Park, Taiwan. *Geosci. Front.* **2015**, *6*, 807–816. [[CrossRef](#)]
28. Sun, D.; Xu, J.; Wen, H.; Wang, D. Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest. *Eng. Geol.* **2021**, *281*. [[CrossRef](#)]
29. Soma, A.S.; Kubota, T.; Mizuno, H. Optimization of causative factors using logistic regression and artificial neural network models for landslide susceptibility assessment in Ujung Loe Watershed, South Sulawesi Indonesia. *J. Mt. Sci.* **2019**, *16*, 383–401. [[CrossRef](#)]
30. Wang, J.; Xu, C. Geodetector: Principle and prospective. *Acta Geogr. Sin.* **2017**, *72*, 116–134. [[CrossRef](#)]
31. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [[CrossRef](#)]
32. Yang, X.; Wang, P.; Li, X.; Xie, C.; Zhou, B.; Huang, X. Application of topographic slope and elevation variation coefficient in identifying the motuo active fault zone. *Seismol. Geol.* **2019**, *41*, 419–435. [[CrossRef](#)]
33. Mark, R.K.; Ellen, S.D. Statistical and Simulation Models for Mapping Debris-Flow Hazard. In *Geographical Information Systems in Assessing Natural Hazards*; Carrara, A., Guzzetti, F., Eds.; Springer: Dordrecht, The Netherlands, 1995; Volume 5, pp. 93–106. [[CrossRef](#)]
34. Ram, P.; Gupta, V. Landslide hazard, vulnerability, and risk assessment (HVRA), Mussoorie township, lesser himalaya, India. *Environ. Dev. Sustain.* **2021**. [[CrossRef](#)]
35. Dikshit, A.; Sarkar, R.; Pradhan, B.; Acharya, S.; Alamri, A.M. Spatial Landslide Risk Assessment at Phuentsholing, Bhutan. *Geosciences* **2020**, *10*, 131. [[CrossRef](#)]
36. Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2015**, *13*, 361–378. [[CrossRef](#)]
37. Park, Y.; Pradhan, A.M.S.; Kim, U.; Kim, Y.-T.; Kim, S. Development and Application of Urban Landslide Vulnerability Assessment Methodology Reflecting Social and Economic Variables. *Adv. Meteorol.* **2016**, *2016*, 1–13. [[CrossRef](#)]
38. Khadka, A.; Katel, P.; Rai, P.; Bahadur Budha, P. Vulnerability Assessment of Peoples Exposed to Landslides in Panchase of Nepal using Analytical Hierarchy Process. *Int. J. Environ.* **2020**, *9*, 81–103. [[CrossRef](#)]
39. Fauzan, M.E.; Damayanti, A.; Saraswati, R. Vulnerability Assessment of Landslide Areas in Ci Manuk Upstream Watershed, Garut District, West Java Province. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2020**, *10*, 219–226. [[CrossRef](#)]
40. van Westen, C.J.; van Asch, T.W.J.; Soeters, R. Landslide hazard and risk zonation—why is it still so difficult? *Bull. Eng. Geol. Environ.* **2005**, *65*, 167–184. [[CrossRef](#)]
41. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
42. Luo, W.; Liu, C.-C. Innovative landslide susceptibility mapping supported by geomorphon and geographical detector methods. *Landslides* **2017**, *15*, 465–474. [[CrossRef](#)]
43. Wang, Y.; Wu, X.; Chen, Z.; Ren, F.; Feng, L.; Du, Q. Optimizing the Predictive Ability of Machine Learning Methods for Landslide Susceptibility Mapping Using SMOTE for Lishui City in Zhejiang Province, China. *Int. J. Environ. Res. Public Health* **2019**, *16*, 368. [[CrossRef](#)] [[PubMed](#)]
44. EINSTEIN, N. Special lecture: Landslide risk assessment procedure. In *Proceedings of the 5th International Symposium on Landslides*, Lausanne, Switzerland, 10–15 July 1988; pp. 1075–1090.
45. FELL, R. Landslide risk assessment and acceptable risk. *Can. Geotech. J.* **1994**, *31*, 261–272. [[CrossRef](#)]
46. Hearn, G.J. Landslide and erosion hazard mapping at Ok Tedi copper mine, Papua New Guinea. *Q. J. Eng. Geol.* **1995**, *28*, 47–60. [[CrossRef](#)]
47. Liu, J.; Li, S.; Chen, T. Landslide Susceptibility Assessment Based on Optimized Random Forest Model. *Geomat. Inf. Sci. Wuhan Univ.* **2018**, *43*, 1085–1091. [[CrossRef](#)]
48. Feng, H. Rainfall-Triggered Landslide Development Regularity Analysis and Hazard Assessment in Chun'an, West Zhejiang. Ph.D. Thesis, China University of Geosciences (Wuhan), Wuhan, China, 1 May 2016.
49. Yang, Y.; Yang, J.; Xu, C.; Xu, C.; Song, C. Local-scale landslide susceptibility mapping using the B-GeoSVC model. *Landslides* **2019**, *16*, 1301–1312. [[CrossRef](#)]

- 
50. Domènech, G.; Alvioli, M.; Corominas, J. Preparing first-time slope failures hazard maps: From pixel-based to slope unit-based. *Landslides* **2019**, *17*, 249–265. [[CrossRef](#)]
  51. Tanyas, H.; Rossi, M.; Alvioli, M.; van Westen, C.J.; Marchesini, I. A global slope unit-based method for the near real-time prediction of earthquake-induced landslides. *Geomorphology* **2019**, *327*, 126–146. [[CrossRef](#)]