


## Article

# Landslide Susceptibility Mapping Using Logistic Regression Analysis along the Jinsha River and Its Tributaries Close to Derong and Deqin County, Southwestern China

Xiaohui Sun, Jianping Chen \*, Yiding Bao, Xudong Han, Jiewei Zhan and Wei Peng 

College of Construction Engineering, Jilin University, Changchun 130026, China;  
sunxh17@mails.jlu.edu.cn (X.S.); baoyd17@mails.jlu.edu.cn (Y.B.); hanxd15@mails.jlu.edu.cn (X.H.);  
zhanjw16@mails.jlu.edu.cn (J.Z.); pengwei17@mails.jlu.edu.cn (W.P.)  
\* Correspondence: chenjp@jlu.edu.cn; Tel.: +86-431-8850-2353

Received: 5 September 2018; Accepted: 4 November 2018; Published: 8 November 2018



**Abstract:** The objective of this study was to identify the areas that are most susceptible to landslide occurrence, and to find the key factors associated with landslides along Jinsha River and its tributaries close to Derong and Deqin County. Thirteen influencing factors, including (a) lithology, (b) slope angle, (c) slope aspect, (d) TWI, (e) curvature, (f) SPI, (g) STI, (h) topographic relief, (i) rainfall, (j) vegetation, (k) NDVI, (l) distance-to-river, (m) and distance-to-fault, were selected as the landslide conditioning factors in landslide susceptibility mapping. These factors were mainly obtained from the field survey, digital elevation model (DEM), and Landsat 4–5 imagery using ArcGIS software. A total of 40 landslides were identified in the study area from field survey and aerial photos' interpretation. First, the frequency ratio (FR) method was used to clarify the relationship between the landslide occurrence and the influencing factors. Then, the principal component analysis (PCA) was used to eliminate multiple collinearities between the 13 influencing factors and to reduce the dimension of the influencing factors. Subsequently, the factors that were reselected using the PCA were introduced into the logistic regression analysis to produce the landslide susceptibility map. Finally, the receiver operating characteristic (ROC) curve was used to evaluate the accuracy of the logistic regression analysis model. The landslide susceptibility map was divided into the following five classes: very low, low, moderate, high, and very high. The results showed that the ratios of the areas of the five susceptibility classes were 23.14%, 22.49%, 18.00%, 19.08%, and 17.28%, respectively. And the prediction accuracy of the model was 83.4%. The results were also compared with the FR method (79.9%) and the AHP method (76.9%), which meant that the susceptibility model was reasonable. Finally, the key factors of the landslide occurrence were determined based on the above results. Consequently, this study could serve as an effective guide for further land use planning and for the implementation of development.

**Keywords:** landslide susceptibility mapping; frequency ratio; principal component analysis; logistic regression analysis; receiver operating characteristic curve

## 1. Introduction

Landslides have become one of the most destructive disasters in mountainous areas [1]. Landslide susceptibility mapping at regional scales is of great significance to risk mitigation and land planning in mountainous areas [2,3]. Landslide susceptibility can be thought of as the tendency of a region to generate landslides [4,5]. Landslide susceptibility only takes into account

the likelihood of the landslide predisposing factors of landslide occurrence, not including the instability process and the return period of landslide occurrence [6,7].

With the development of the geographic information system (GIS), global positioning system (GPS), and remote sensing (RS), many researchers have applied these technologies to landslide susceptibility mapping [8–10]. Over the last decades, many statistical methods have been used in landslide susceptibility mapping, such as the logistic regression analysis [11], frequency ratio (FR) [10], statistical index (SI) [12], certainty factor (CF) [13], discriminant analysis (DA) [14], evidential belief function [15], and index of entropy [16]. In addition to the statistical methods, a lot of machine learning algorithms, such as artificial neural network models (ANN) [10], support vector machines (SVM) [13], maximum entropy (MaxEnt) [16], and naïve Bayes [17] have also been used for landslide susceptibility mapping. But beyond that, some ensemble methods have also been used in landslide susceptibility mapping, such as the ANN–SVM, ANN–MaxEnt, SVM–MaxEnt, ANN–MaxEnt–SVM [18], and ANN–Bayes analyses [19].

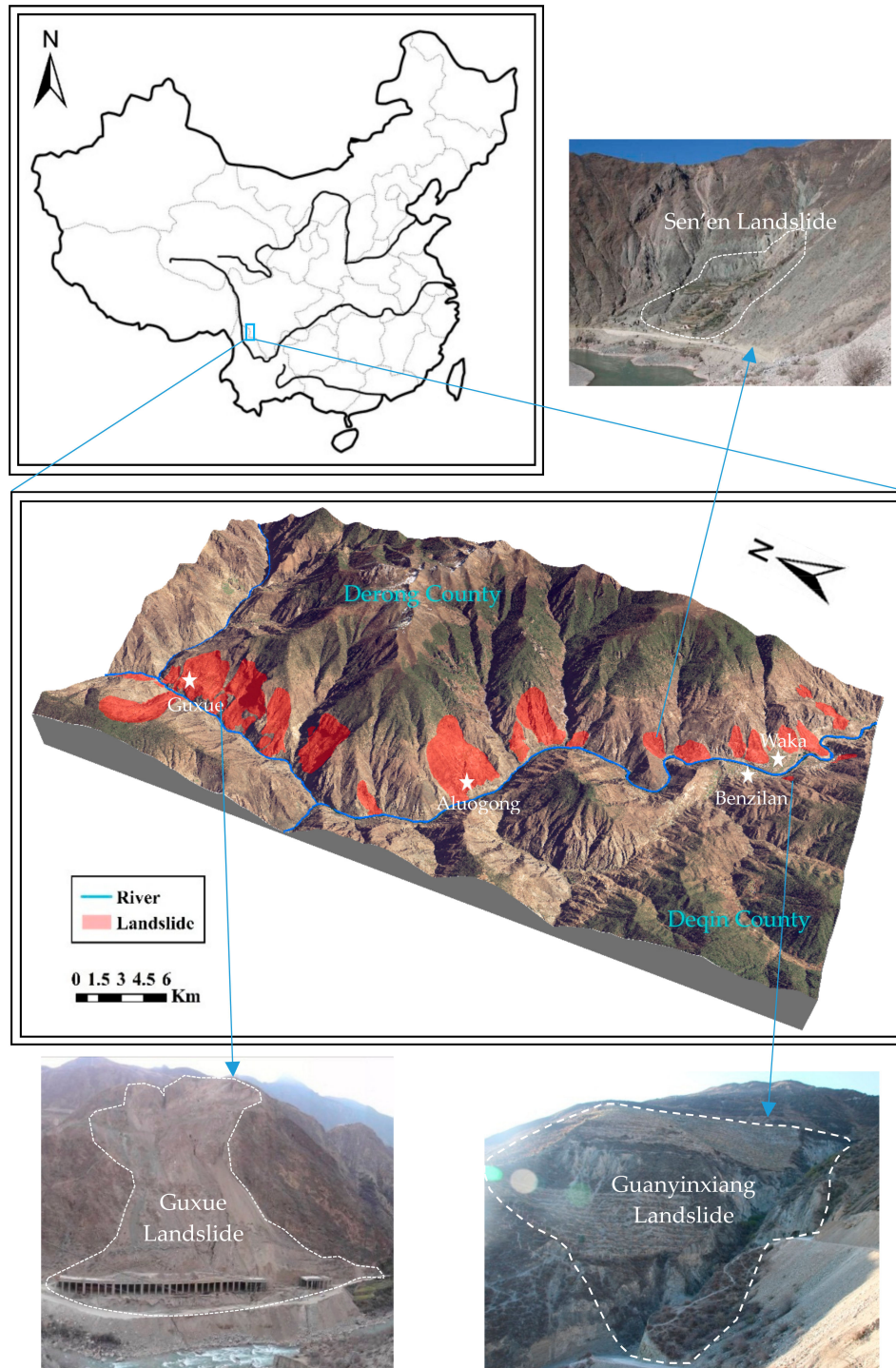
In this study, logistic regression was applied to produce the landslide susceptibility map of the study area, which is a multivariate statistical method and has been widely used in landslide susceptibility analysis. Logistic regression is a statistical method, but it is actually a machine learning method, except that its mathematical expression is known. Compared with other evaluation methods, it has the following features: (1) it is based on statistical methods, and it has low requirements for the quality and quantity of samples; (2) the independent variable can be continuous or discrete, and it does not have to satisfy the normal distribution; (3) this method is very mature, with many mature tests, and the results are easy to test [5,8–11]. Our choice of this method was based on the fact that landslide occurrence is controlled by many linear and nonlinear influencing factors. The aims of this study are to identify the areas that are most susceptible to the occurrence of landslides, and to find the key factors associated with landslides. So, this study is mainly divided into six steps, as follows: (1) According to the remote sensing and aerial photos' interpretation and the field survey, a total of 40 landslides were mapped in the study area (including rock slope deformation, rock planar slide, and rock flexural topple, not including debris flow). (2) Based on the field survey, the mechanism of the landslide, the local geo-environmental conditions, and the previous studies, 13 influencing factors, were selected to produce the landslide susceptibility map. (3) In order to clarify the relationship between the landslide occurrence and the influencing factors, the frequency ratio (FR) model was used to describe their relationship. (4) Principal component analysis (PCA) was used to eliminate the multiple collinearity between the 13 influencing factors. (5) Logistic regression analysis was used to produce the landslide susceptibility map, and was compared with other methods. (6) The receiver operating characteristic (ROC) curve was used for validation.

The purpose of this study is to find an accurate landslide susceptibility map, and to find the key factors associated with landslides, which provides a reasonable tool for the landslide risk mitigation of the study area. The landslide susceptibility map could serve as an effective guide for the further land use planning and for the implementation of development.

## 2. Study Area

The study area lies on the border between the Sichuan province and Yunnan province of China, along the upper reaches of the Jinsha River. On the left bank of the Jinsha River is Derong County, and the right bank is Deqin County. The study area ranges from 99°12' E to 99°21' E longitude and from 28°12' N to 28°26' N latitude, covering a total area of 364.10 km<sup>2</sup> (Figure 1). The elevation of the study area ranges from 1989 to 4888 m, and the maximum elevation difference is 2899 m. The study area belongs to the Tibetan Plateau, which is a rapidly uplifting region. Previous studies have shown that since the Quaternary, neotectonic movement has made the study area uplift at a rate of 5 mm per year [20]. As a result of the rapid uplift, a rapid river incision was caused. Many landslides occurred along the river because of the combined effect of both the rapid uplift and river incision. The annual temperature is 13.8–19.2 °C. As a result of the study area being under the influence of

southwest and southeast monsoons, the climatic characteristics of this area are complex. The study area belongs to the subtropical dry-hot valley climate. Because of the huge elevation difference and the monsoons, the foehn effect here is very significant. The mean annual precipitation of the low elevation area is around 300 mm. However, it can reach more than 1000 mm in the high elevation area.

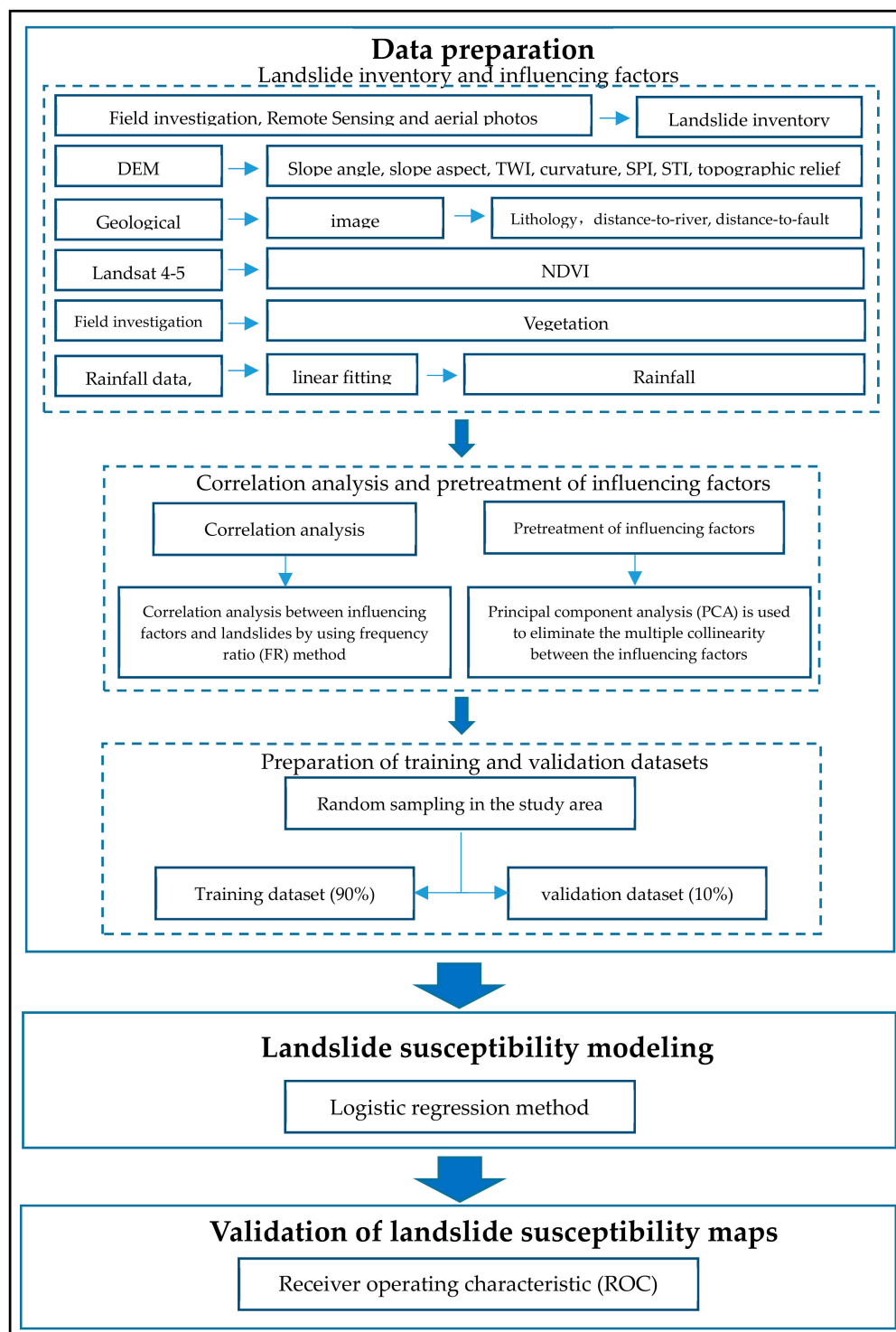


**Figure 1.** Geographical position and landslide inventory of the study area.

### 3. Methodology

The methodology of this study is shown in Figure 2. Based on the flow chart, this study is mainly divided into the following three steps: (a) data preparation, (b) landslide susceptibility modeling

using logistic regression analysis; and (c) evaluation accuracy analysis using a receiver operating characteristic (ROC) curve.



**Figure 2.** Flow chart of this study.

### 3.1. Methods

In the analysis of landslide susceptibility, the influencing factors are usually used as the independent variables, and whether the landslides occurred is shown as a binary ("1" represents a landslide occurred, "0" represents a landslide has not occurred). Because of the existence of



non-continuous variables in the influencing factors (such as lithology), the multivariate linear regression method will no longer be applied to the derivation of the relationship between such independent variables and dependent variables. However, the logistic regression (LR) method can solve this problem.

The logistic regression method is a commonly used method for the statistical analysis of dichotomous dependent variables (the dependent variable only takes two values). This method can describe the relationship between a binary dependent variable and a series of independent variables. The independent variable can be continuous or discrete, and it does not have to satisfy the normal distribution. Logistic regression can describe the complex nonlinear relationship between the natural phenomena using simple linear regression, and can be used to predict the probability of an event's occurrence. The odds ratio estimated using logistic regression can also be used to test the strength of the correlation between the independent variables and the dependent variables. So, this method has been widely used in the analysis of landslide susceptibility.

The main idea of the logistic regression model is to determine the likelihood of future landslide occurrence after each factor is converted to a logical variable. Logistic regression use the maximum likelihood method to look for the "best fit". The simplified logical regression method can be described by the following equation [10,21]:

$$P = \frac{1}{1 + e^{-z}}, \quad (1)$$

where  $P$  is the probability of landslide occurrence, and the value of  $P$  ranges between 0 and 1;  $z$  is the liner combination:

$$z = \beta_0 + \beta_1 Y_1 + \beta_2 Y_2 \dots + \beta_G Y_G, \quad (2)$$

where  $\beta_0$  is a constant;  $\beta_1, \beta_2, \dots$ , and  $\beta_G$  are the regression coefficients; and  $Y_1, Y_2, \dots$ , and  $Y_G$  are the influencing factors. In the analysis of landslide susceptibility, the pixels with landslides have a value of 1, and the pixels without landslides have a value of 0. By using the logistic regression and the observed data, the probability of the landslide occurrence can be calculated.

### 3.2. Landslides and Influencing Factors

The study area of this paper is very large, and the terrain is very steep. Many areas are inaccessible to human beings. It is very important to conduct a detailed geological survey of the landslides in the study area, but, because of time and manpower constraints, it is impossible to conduct a detailed field survey of all of the landslides in the study area. Therefore, it is necessary to carry out the investigation and research of landslides by means of aerial photos and remote sensing interpretation technology. Landslides have certain recognition features in remote sensing images and aerial photos, such as shape, color, shadow, and differences with surrounding topography and landform [22,23]. In previous remote sensing interpretation work, if there was an area where there were obvious landslide characteristics, we would have tentatively defined it as a landslide. Subsequently, the field survey in the study area was used to determine the accuracy of the interpretation of the landslides, and to supplement those that could not be interpreted.

There are many factors that affect the occurrence of landslides [4,24–32]. In order to understand the main factors affecting landslide sensitivity, Hamid Reza Pourghasemi and Mauro Rossi reviewed 220 scientific papers published between 2005 and 2012 in different ISI (Internation Scientific Indexing) journals [15], and they counted the application frequency of the influencing factors, which were used in landslide susceptibility mapping. The statistical results show that the 20 factors with the highest application frequency were the slope degree, lithology, slope aspect, land cover/land use, distance from river, elevation, distance from faults, plan curvature, profile curvature, distance from road, soil type, topographic wetness index (TWI), rainfall, normalized difference vegetation index (NDVI), slope-length, steam power index (SPI), drainage density, geomorphology, soil thickness, and fault density. According to the statistical results and the geological environment characteristics of the study area, 13 influencing factors were selected

for landslide susceptibility mapping in this study, and they can be divided into the following three categories: lithology, geomorphological, and environment. These influencing factors maps were converted to a pixel size of  $10 \times 10$  m, and the digital elevation model (DEM) was also converted to the same pixel size. All of those maps are conducted using ArcGIS 10.2 software. The classification of the continuous influencing factors was based on previous studies [24,25].

### 3.2.1. Landslide Inventory

By means of the aerial photo and remote sensing interpretation and extensive field survey, a total of 40 landslides were identified and mapped along the Jinsha River and Dingqu River. Fourteen of these landslides are located on the right bank of the Jinsha River and eight are on its left bank. Five landslides are located on the right of the Dingqu River, and three are located on its left (Figure 1). As shown in Figure 1, the landslides in the study area are mainly distributed on the left bank of the Jinsha River and Dingqu River. The reason for this is that landslide occurrence is affected by topography, stratigraphic lithology, geological structure, meteorological conditions, hydrological conditions, human engineering activities, and so on.

### 3.2.2. Lithology Factor

The influence of formation lithology on the occurrence of landslides is obvious. The type, degree of hardness, structural characteristics, and so on have great influence on the physical and mechanical properties, weathering resistance, deformation, and failure modes of the slopes. According to the geological map of scale 1:200,000, the exposed strata in the study area are from the Devonian, Carboniferous, Permian, Triassic, and Quaternary. The Quaternary strata include the landslide accumulation ( $Q_h^{del}$ ) and the bench gravel and sand layer ( $Q_p^3$ ). The Triassic strata include the Jiabila formation ( $T_{3j}$ ) and Qugasi formation ( $T_{2q}^1$ ,  $T_{2q}^2$ , and  $T_{2q}^3$ ). The lithology of the Jiabila formation are mainly composed of siltstone, volcanic rock, slate, sandstone, and limestone. The lithology of the Qugasi formation are mainly composed of volcanic rock, slate, sandstone, and limestone. The Permian strata include the Gangdadao formation ( $P_2$  and  $P_{2g}$ ) and the Ranlang formation ( $P_{1b}$ ,  $P_{1a}$ , and  $P_{1r}$ ). The lithology of the Gangdadao formation and the Ranlang formation are mainly composed of volcanic rock, slate, sandstone, and limestone. The carboniferous strata include the Dingpo formation ( $C_3$ ) and Zhapu formation ( $C_2$ ). The lithology of the Dingpo formation and the Zhapu formation are mainly composed of basalt, andesite, rhyolite, and volcanic breccia. The Devonian strata include the Qiongcuo formation ( $D_{2q}$ ) and the Gerong formation ( $D_{1g}$ ). The lithology of the Qiongcuo formation and Gerong formation are mainly composed of volcanic rock, slate, sandstone, and limestone. The geological map was used to extract a lithology map.

### 3.2.3. Geomorphological Factors

The statistical results of Hamid Reza Pourghasemi and Mauro Rossi show that geomorphological factors have a significant influence on the occurrence of landslides. According to the statistical results, seven factors are selected in this category, including the slope angle, slope aspect, topographic wetness index (TWI), curvature, stream power index (SPI), sediment transport index (STI), and topographic relief.

**Slope angle:** Slope angle is one of the most important factors for landslide susceptibility mapping [33]. Within a certain slope angle, because of the increases of the slope angle, the gravity stress and shear stress of the slope generally increase and the probability of slope failure is increased [34]. Based on the digital elevation model (DEM), with a resolution of 10 m of the study area, the slope angle map can be extracted using the ArcGIS 10.2.

**Slope aspect:** As another important factor for landslide susceptibility mapping, the slope aspect affects the rainfall direction, and the amount and the effluence of the solar radiation of the slope. So, it makes the moisture and vegetation unevenly distributed in the slope [15,24,25,35].

Therefore, the slope aspect has a different influence on the slope stability. The slope aspect map was produced using the DEM.

**TWI:** The TWI reflects the amount of flow accumulation at any point in the study area [36]. To some extent, the TWI represents the distribution of the soil moisture [37]. The TWI can be calculated using the following equations [38–40]:

$$TWI = \ln(A_S \div \tan\beta), \quad (3)$$

where  $A_S$  is the upslope contributing area and  $\beta$  is the slope angle.

**Curvature:** Curvature describes the morphological characteristics of the slope shape, which reflects the formation of surface erosion and surface runoff. The slope shape provides spaces for slope sliding [41]. The curvature map was extracted using the DEM.

**SPI:** The SPI reflects the erosion capacity of water flow in the study area [42]. The SPI can be calculated using the following equations [43]:

$$SPI = A_S \times \tan\beta, \quad (4)$$

where  $A_S$  is the upslope contributing area and  $\beta$  is the slope angle.

**STI:** The STI is a dimensionless parameter, and it is calculated by combining the length and steepness. It describes the process of erosion and deposition of the study area [44]. The STI can be calculated using the following equations [40]:

$$STI = \left(\frac{A_s}{22.13}\right)^{0.6} \times \left(\frac{\sin\beta}{0.0896}\right)^{1.3}, \quad (5)$$

where  $A_S$  is the upslope contributing area and  $\beta$  is the slope angle.

**Topographic relief:** Topographic relief can reflect the change of the rolling of the slope surface and can reveal the law of topography change of an entirety area.

### 3.2.4. Environmental Factors

The environmental factors include the average annual rainfall, vegetation, normalized difference vegetation index (NDVI), distance-to-river, and distance-to-fault.

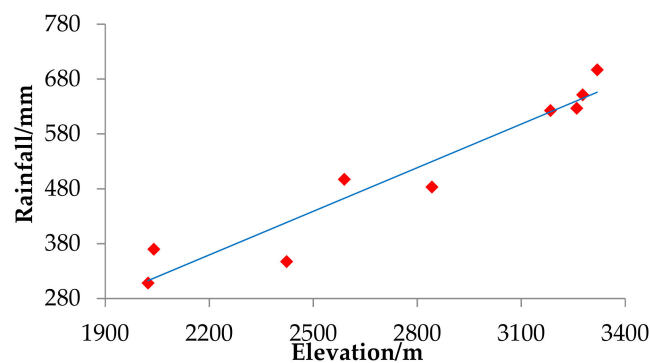
**Average annual rainfall:** Rainfall is one of the most important factors that trigger landslides. Rainfall will cause the erosion on the slope surface. The water infiltration will increase the gravity of the rock and reduce the shear strength of the joints, thus inducing landslide hazards. Because of the foehn effect, the precipitation of the study area follows an obvious vertical distribution. Previous studies have shown that the precipitation increases with increasing elevation, and it is proportional to the elevation [45]. There are many precipitation stations distributed in the Yunnan and Sichuan province (China Meteorological Data Service Center) [24,46], but most of these are distributed in the county, rather than along Jinsha River and Dingqu River. Based on the distance to the Jinsha River and Dingqu River, the climate zone, and other factors, nine precipitation stations were selected to establish the relationship between the average annual rainfall and elevation, as listed in Table 1. In Figure 3, the red points represent the rainfall data collected from the nine precipitation stations. The fitting equation is as follows:

$$P_a = 0.265H - 223.4, \quad (6)$$

where  $H$  is the elevation of the precipitation stations and  $P_a$  is the average annual rainfall. Caochen [12] suggested that the precipitation gradient is 24.4 mm/100 m of the Xulong reservoir, which is similar to this area. The precipitation gradient of this study is 26.5 mm/100 m. So, 26.5 mm/100 m is a reasonable precipitation gradient of the study area.

**Table 1.** Annual precipitation measured by precipitation station at different elevation.

| Precipitation Station | Longitude | Latitude | Elevation/m | Average Annual Precipitation/mm | Data Resources/Year |
|-----------------------|-----------|----------|-------------|---------------------------------|---------------------|
| Derong                | 99°10.2′  | 28°25.8′ | 2422.9      | 347.1                           | 1981–2010           |
| Batang                | 99°03.6′  | 30°00.0′ | 2589.2      | 497.0                           | 1981–2010           |
| Xiangcheng            | 99°28.8′  | 28°33.6′ | 2842.0      | 483.1                           | 1981–2010           |
| Xianggelila           | 99°25.2′  | 27°30.0′ | 3276.7      | 651.1                           | 1981–2010           |
| Deqin                 | 98°33.0′  | 28°17.4′ | 3319.0      | 696.7                           | 1981–2010           |
| Dege                  | 98°35.0′  | 31°48.0′ | 3184.0      | 622.4                           | 1981–2010           |
| Baiyu                 | 98°50.0′  | 31°13.0′ | 3260.0      | 626.6                           | 1981–2010           |
| Benilan               | 99°17.0′  | 28°17.0′ | 2023.0      | 308.0                           | 1965–1998           |
| Shangqiaotou          | 99°24.0′  | 28°10.0′ | 2040.0      | 369.7                           | 1961–2004           |

**Figure 3.** The relationship between elevation and average annual precipitation based on the nine precipitation stations.

**Vegetation:** For the field survey, the vegetation of the study area can be divided into the following five types: (a) in the elevation range, 1989 to 2500 m is the bare soil; (b) in the elevation range, 2500 to 3300 m is the brush-forbs; (c) in the elevation range, 3300 to 4200 m is the woods; (d) in the elevation range, 4200 to 4500 m is the grassland; and (e) in the elevation range, more than 4500 m is the snow.

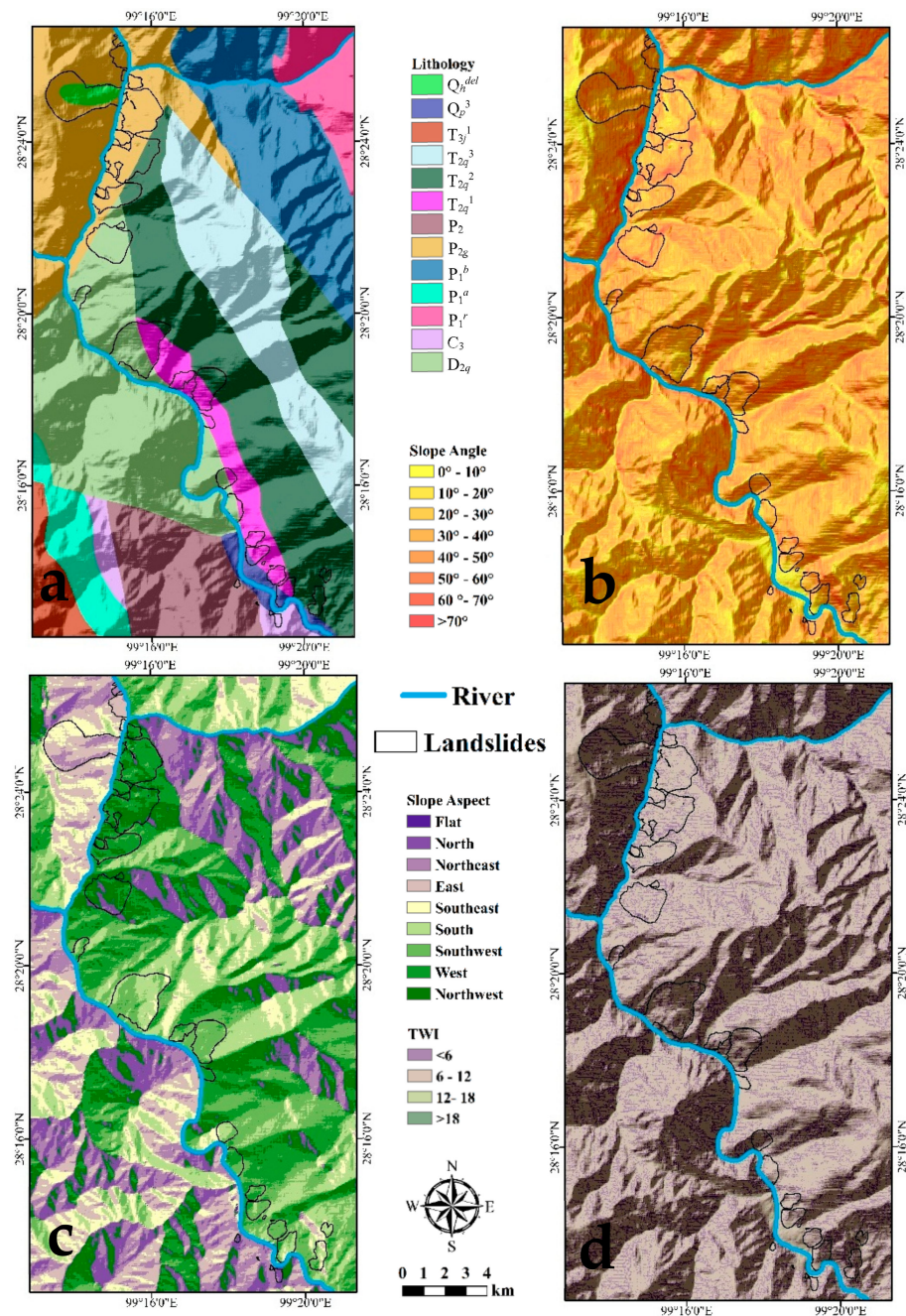
**NDVI:** The NDVI can be used to reflect the vegetation coverage of the study area. If the normalized vegetation index is less than zero, it means that the ground is covered with water or snow. If the normalized vegetation index is equal to zero, it means that there is bare land or rock. If the normalized vegetation index is greater than zero, it indicates that there is vegetation cover, and the greater the value, the higher the vegetation coverage. The Landsat 4–5 image was used to extract the NDVI map.

**Distance-to-river:** The slope on both sides of the river is usually eroded by rivers. In normal conditions, at the closer distance to the river, the stronger the erosion and the higher probability of the occurrence of landslides [47]. The distance-to-river map was calculated in 300 m intervals.

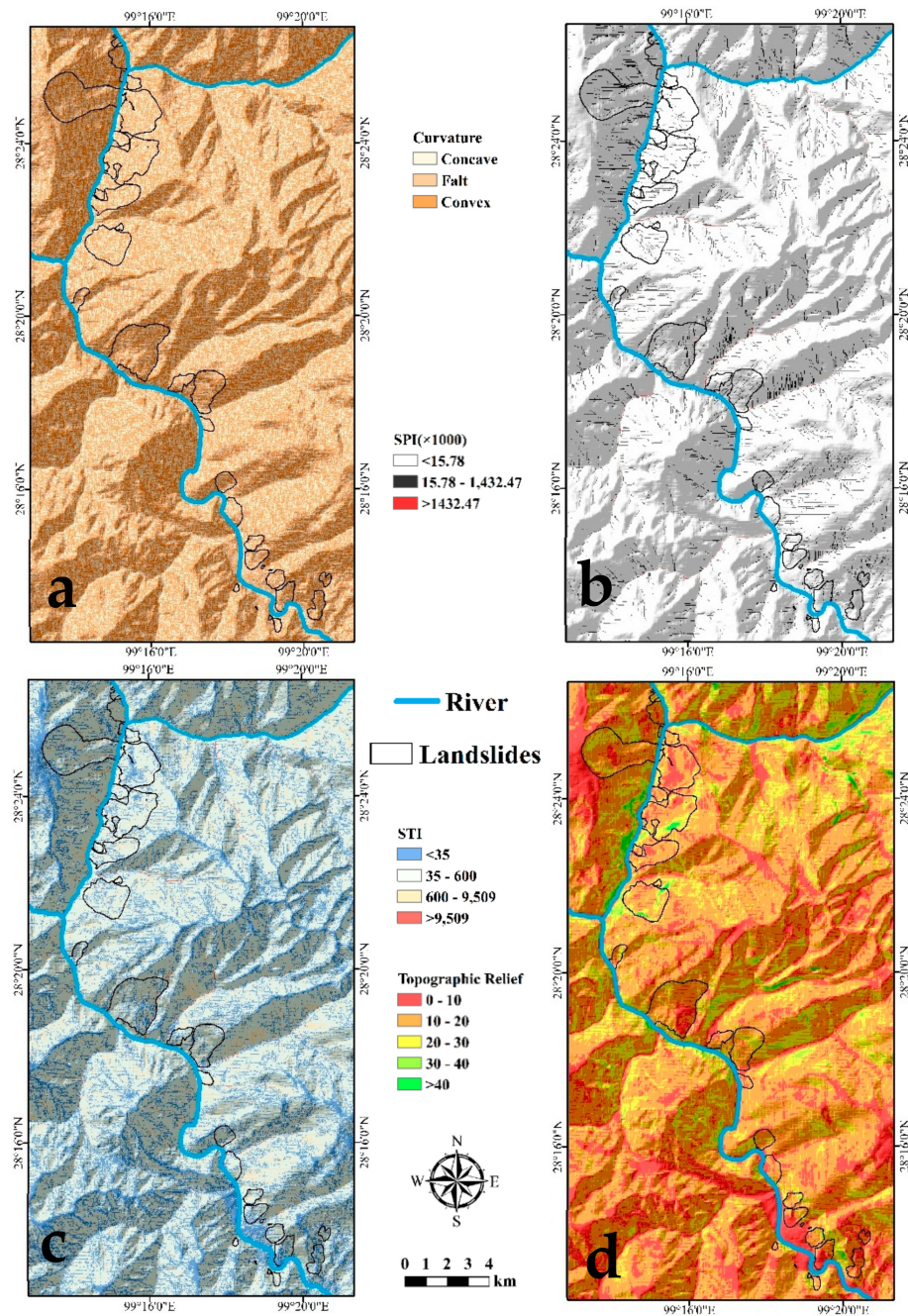
**Distance-to-fault:** In the faulted zone, the rock is relatively broken and the joint fracture is developed, which makes the slope of these areas less stable and more prone to landslide occurrence [48]. The distance-to-fault map was calculated in 300 m intervals.

All of the influencing factor maps are shown in Figures 4–7.



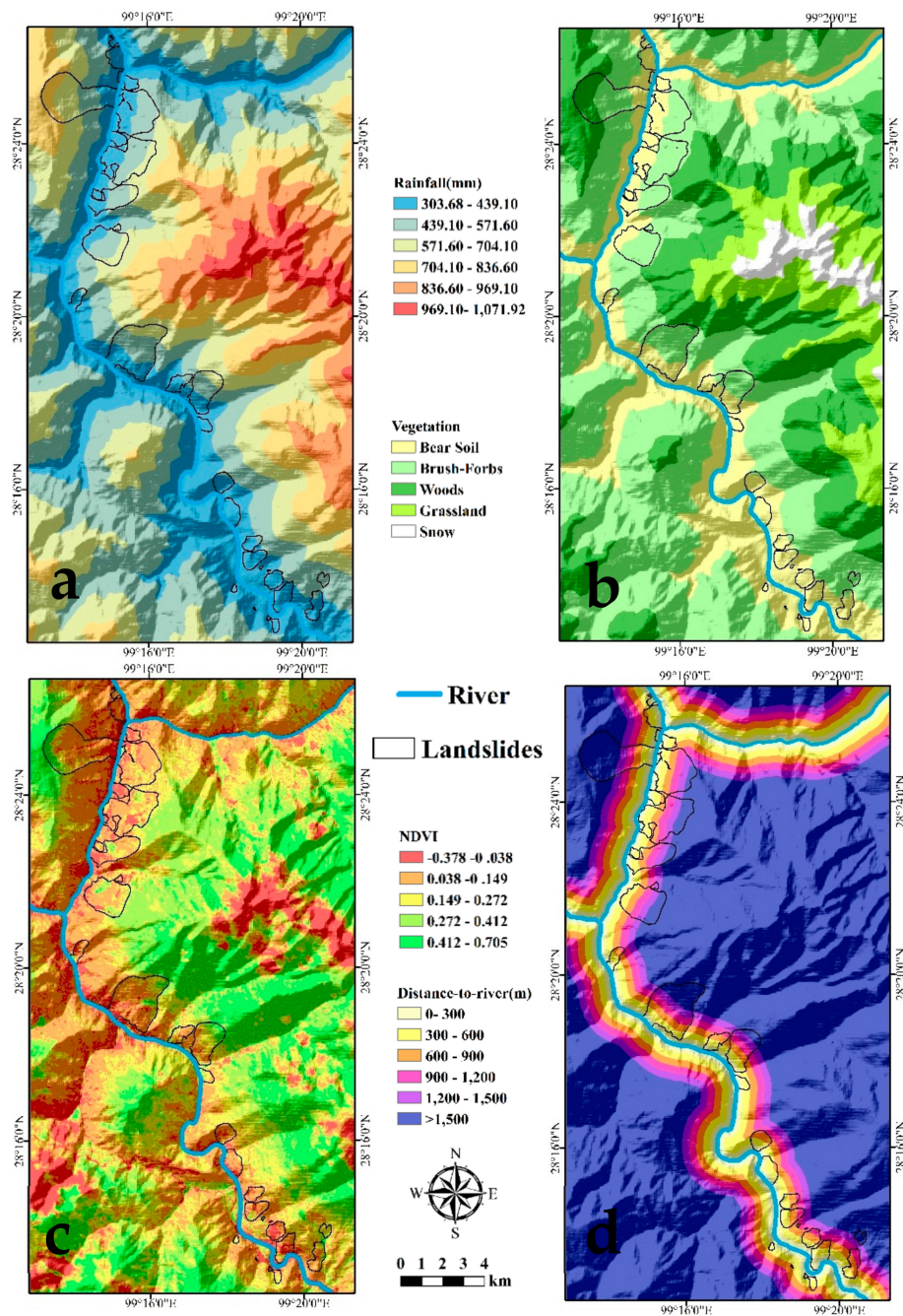


**Figure 4.** Influencing factors maps of the study area: (a) lithology; (b) slope angle; (c) slope aspect; and (d) topographic wetness index (TWI).

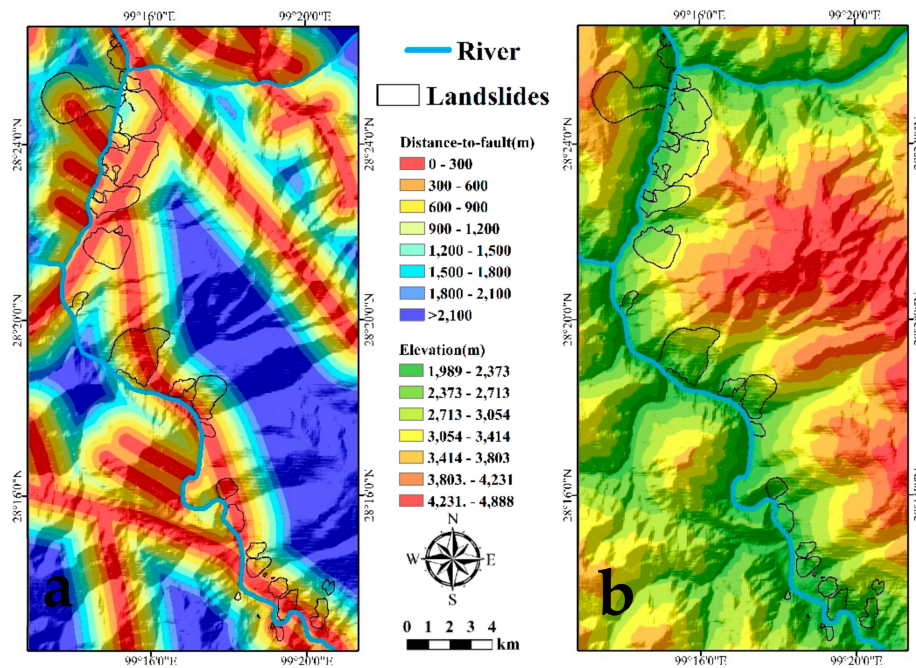


**Figure 5.** Influencing factor maps of the study area: (a) curvature; (b) steam power index (SPI); (c) sediment transport index (STI); and (d) topographic relief.





**Figure 6.** Influencing factor maps of the study area: (a) rainfall; (b) vegetation; (c) normalized difference vegetation index (NDVI); and (d) distance-to-river.



**Figure 7.** Influencing factors maps of the study area: (a) distance-to-fault and (b) elevation.

### 3.3. Evaluation of Influencing Factors

#### 3.3.1. Probabilistic Relationship Analysis between Landslides and the Influencing Factors

Bivariate statistical methods are commonly used to compute the probabilistic relationship between the dependent and independent variables. In this paper, the frequency ratio (FR) method will be used to ensure the relationship between the influencing factors and the occurrence of the landslides. In the FR method, the quantitative relationship between the landslide occurrence and the different conditioning parameters can be identified and expressed as an FR value. The FR value calculation process is very concise, and can be realized as follows [49]:

$$FR = \frac{a/A}{b/B}, \quad (7)$$

where  $a$  is the number of pixels with landslides for each conditioning factor,  $A$  is the total number of pixels with landslides in study area,  $b$  is the number of pixels for each conditioning factor, and  $B$  is the total number of pixels in the study area. If the values are greater than 1, it means there is a greater correlation, whereas values less than 1 represent a minor correlation [12].

#### 3.3.2. Principal Component Analysis

In general, there is no independent test on the selected influencing factors before logistic regression. However, the adjustment of the logistic regression model is sensitive to the linear correlation of the influencing factors [50]. The linear correlation of the influencing factors will increase the variance of the logistic regression coefficients. Some studies use an independence test to verify the mutual independence of each influencing factor, such as the variance inflation factor (VIF) [51] and conditional independence test [52], and can then exclude the influencing factors, which are highly correlated. However, compared with these methods, the principal component analysis (PCA) can not only eliminate the multicollinearity problem among the influencing factors, but can also be used to evaluate by how much the different influencing factors affect the landslide susceptibility of the study area. This is crucial to the subsequent search for the key factor of landslide occurrence. So, in this paper, PCA is used to reduce the dimension of the preselected influencing factors and change the factors, which are reselected, so as to make them independent of each other. Then, the reselected factors will



be used in the logistic regression to eliminate the influence of the linear correlation between the factors on the predicted results.

The principal component analysis uses the linear correlation between the preselected factors, replacing the preselected factors with a small number “principal components”. Those “principal components” can represent most of the information of the preselected factors [53]. The algebraic essentials of PCA are as follows: Let  $Y(t, x)$  be a preselected data at point  $x$  ( $x = 1, \dots, p$ ) and time  $t$  ( $t = 1, \dots, m$ ). The matrices  $\{Y(t, x): x = 1, \dots, p\}$  mean all of the values of  $Y(t, x)$  at point  $t$  from 1 to  $m$ , and the matrices center on their time averages. Those matrices can be replaced as the  $p \times 1$  column vectors,  $Y(t) = [Y(t, 1), \dots, Y(t, p)]^T$ , and “T” means the transformation operation. The vectors will form a series of points around the origin of a  $p$ -dimensional Euclidian space,  $E_p$ . So, PCA can transform the preselected factors system to a new factors system, using the linear transformation. PCA makes the greatest variance using any projection of the data lies as the first principal component, and the second greatest variance as the second principal component, and so on. Thus, by retaining the characteristics of the data set that contribute most to its variance, PCA can be used to reduce the dimensional of the data set.

The steps of PCA are as follows:

- (1) Use the following equation to normalize the preselected influencing factors:

$$M = \frac{H - H_{\min}}{H_{\max} - H_{\min}}, \quad (8)$$

where  $M$  is the preselected influencing factor’s normalized value,  $H$  is the value of each preselected influencing factor’s pixels,  $H_{\max}$  and  $H_{\min}$  are the maximum and minimum values of each preselected influencing factor, respectively.

- (2) In ArcGIS 10.2 software, a  $20 \times 20$  m fishnet was built to sample 13 preselected factors.
- (3) Using the Kaiser–Meyer–Olkin (KMO) test and the Bartlett’s test of the sample data, the applicability of PCA can be verified.
- (4) PCA was carried out for the sample data, and a correlation matrix eigenvalue greater than 0.9 was selected as the principal component.
- (5) According to the principal component, a new influencing factors system will be built.

### 3.4. Data for the Logistic Regression Analysis

In order to establish the logistic regression model, which needed pixels with or without the presence of landslides [54], we created datasets containing 200,000 pixels with landslides and an equal number of non-landslides pixels, which were randomly chosen from the study area. Both of the landslides pixels and non-landslides pixels are divided into two sets. One set of the pixels that was used as the training dataset for the regression analysis included 90% (180,000 pixels) of the pixels. And the other set was used as the validation dataset, which included 10% (20,000 pixels) of the pixels. So, the final datasets consisted of 400,000 pixels. All of the pixel’s information are put into a table. One column of the table contained the status information of the landslides. A value of 1 was assigned to the pixels with landslides, and a value of 0 was assigned for the pixels without landslides. In the other column, the one for the influencing factors, contained the influencing factors value information. Finally, the datasets will be used in the logistic regression analysis, and the value of  $\beta_0, \dots, \beta_G$  can be achieved, which can be used to calculate the value of  $z$ .

### 3.5. Model Development

First, before the PCA, the datasets should be tested using the Kaiser–Meyer–Olkin test and the Bartlett’s test. The Kaiser–Meyer–Olkin test is used to test the correlation between the influencing factors, and its values range from 0 to 1. The Bartlett’s test is used to test whether the influencing factors

are independent from each other. When the KMO value is greater than 0.6 and the Bartlett's value is less than 0.01, it is suitable for the PCA. Second, the goodness of fit of the logistic regression model was evaluated using the Cox and Snell pseudo  $R^2$  test and the Nagelkerke pseudo  $R^2$  test [55]. The value of the Cox and Snell pseudo  $R^2$  test is usually less than 1. The value of the Nagelkerke pseudo  $R^2$  test ranges from 0 to 1 [55,56]. If an  $R^2$  value is more than 0.2, it means that it is a good fit [57]. Finally, the landslide susceptibility map produced by the logistic regression model will be divided into the following five classes using the natural breaks method: very high, high, moderate, low, and very low.

### 3.6. Model Validation

It is necessary to validate the accuracy of the model. In this study, the receiver operating characteristic curve (ROC) analysis was used to evaluate the prediction power of the model. The ROC curve is drawn with the false positive rate (sensitivity) as the X-axis, and the true positive rate (1-specificity) as the Y-axis. It has a chance diagonal (the connection between the origin and the point (1, 1)), and the ROC curve area (AUC) of the opportunity diagonal is 0.5. The farther away the opportunity diagonal, the larger the AUC value, and the more accurate the prediction. For any prediction experiment, the value of AUC is between 0.5 and 1. The ROC curve area is commonly used as a standard to evaluate the goodness of the susceptibility model [25,51,58–61].

## 4. Results

### 4.1. Evaluation of Influencing Factors

In order to clarify the relationship between the landslide occurrence and the influencing factors, the FR model was used to describe the relationship between them. Table 2 shows the results of the application of the FR model. From this table, it is seen that the percentages of the landslide area of the  $Q_h^{del}$ ,  $T_{2q}^2$ ,  $T_{2q}^1$ ,  $P_{2g}$ , and  $D_{2q}$  were 7.32%, 14.00%, 17.40%, 35.67%, and 22.82%, respectively, which means 97.21% of the landslide area was distributed among the five lithologies. The lithologies with an FR value greater than 1 are  $Q_h^{del}$ ,  $Q_3^P$ ,  $T_{2q}^1$ ,  $P_{2g}$ , and  $D_{2q}$ , and the highest value of the FR was  $Q_h^{del}$  (13.04), following by  $T_{2q}^1$  (4.72),  $P_{2g}$  (2.48),  $Q_3^P$  (1.37), and  $D_{2q}$  (1.26). This means that the relationship between the lithology and the occurrence of the landslide from small to large is  $T_{2q}^1$ ,  $P_{2g}$ ,  $Q_3^P$ , and  $D_{2q}$ . The slope angle was between 0 and 73° in the study area. For the slope angle factor, classes 20–30, 30–40, and 60–70 had a positive FR value (1.11, 1.05, and 2.06, respectively). For the slope angle classes <20, 40–60, and >70, the FR values were negative. This means that the slope angle classes of 20–40 and 60–70 were prone to landslide occurrence. As for the slope aspect, 75.83% of the landslide areas were found in the slope aspect of the S, SW, W, and NW. The areas facing the SW, W, and NW have higher FR values, which means that they have higher probabilities of landslide occurrence. The TWI of the study area was divided into the following four classes: <6, 6–12, 12–18, and >18. Over 70 % of the landslide areas were found in the class of 6–12. The FR values of the four classes were 0.93, 1.01, 1.42, and 0.60, respectively. It can be seen that the TWI classes of 6–12 and 12–18 are prone to landslide occurrence. As for the curvature, the percentages of the landslide area of the classes of concave, flat, and convex were 39.71%, 29.00%, and 39.30%, respectively, but the FR values of the three classes were not high.

**Table 2.** Distribution of the training pixels. NDVI—normalized difference vegetation index; FR—frequency ratio; TWI—topographic wetness index; SPI—steam power index; STI—sediment transport index; NDVI—normalized difference vegetation index.

| Factors               | Class         | Landslide Not Occurred |        | Landslide Occurred |        | Total Count | FR    |
|-----------------------|---------------|------------------------|--------|--------------------|--------|-------------|-------|
|                       |               | Count                  | Ratio  | Count              | Ratio  |             |       |
| Lithology             | $Q_h^{del}$   | 1067                   | 0.03%  | 19,379             | 7.32%  | 20,446      | 13.04 |
|                       | $Q_3^p$       | 44,741                 | 1.33%  | 4964               | 1.88%  | 49,705      | 1.37  |
|                       | $T_{3j}^1$    | 82,644                 | 2.45%  | 0                  | 0.00%  | 82,644      | 0.00  |
|                       | $T_{2q}^3$    | 433,458                | 12.84% | 0                  | 0.00%  | 433,458     | 0.00  |
|                       | $T_{2q}^2$    | 646,006                | 19.13% | 37,075             | 14.00% | 683,081     | 0.75  |
|                       | $T_{2q}^1$    | 88,043                 | 2.61%  | 46,056             | 17.40% | 134,099     | 4.72  |
|                       | $P_2$         | 265,722                | 7.87%  | 639                | 0.24%  | 266,361     | 0.03  |
|                       | $P_{2g}$      | 430,106                | 12.74% | 94,434             | 35.67% | 524,540     | 2.48  |
|                       | $P_1^b$       | 461,126                | 13.66% | 0                  | 0.00%  | 461,126     | 0.00  |
|                       | $P_1^a$       | 127,972                | 3.79%  | 0                  | 0.00%  | 127,972     | 0.00  |
|                       | $P_1^r$       | 141,015                | 4.18%  | 0                  | 0.00%  | 141,015     | 0.00  |
|                       | $C_3$         | 55,805                 | 1.65%  | 1763               | 0.67%  | 57,568      | 0.42  |
|                       | $D_{2q}$      | 598,585                | 17.73% | 60,422             | 22.82% | 659,007     | 1.26  |
| Slope Angle           | 0–10          | 99,022                 | 2.93%  | 3832               | 1.45%  | 102,854     | 0.51  |
|                       | 10–20         | 390,709                | 11.57% | 27,926             | 10.55% | 418,635     | 0.92  |
|                       | 20–30         | 1,105,152              | 32.73% | 96,973             | 36.63% | 1,202,125   | 1.11  |
|                       | 30–40         | 1,284,364              | 38.04% | 105,990            | 40.04% | 1,390,354   | 1.05  |
|                       | 40–50         | 430,872                | 12.76% | 25,518             | 9.64%  | 456,390     | 0.77  |
|                       | 50–60         | 60,726                 | 1.80%  | 3549               | 1.34%  | 64,275      | 0.76  |
|                       | 60–70         | 5331                   | 0.16%  | 939                | 0.35%  | 6270        | 2.06  |
|                       | >70           | 114                    | 0.00%  | 5                  | 0.00%  | 119         | 0.58  |
| Slope Aspect          | Flat          | 1938                   | 0.06%  | 6                  | 0.00%  | 1944        | 0.04  |
|                       | N             | 433,950                | 12.85% | 7207               | 2.72%  | 441,157     | 0.22  |
|                       | NE            | 364,642                | 10.80% | 13,411             | 5.07%  | 378,053     | 0.49  |
|                       | E             | 387,480                | 11.48% | 28,779             | 10.87% | 416,259     | 0.95  |
|                       | SE            | 340,425                | 10.08% | 14,563             | 5.50%  | 354,988     | 0.56  |
|                       | S             | 435,604                | 12.90% | 31,770             | 12.00% | 467,374     | 0.93  |
|                       | SW            | 466,658                | 13.82% | 60,966             | 23.03% | 527,624     | 1.59  |
|                       | W             | 509,216                | 15.08% | 67,252             | 25.40% | 576,468     | 1.60  |
| TWI                   | NW            | 436,377                | 12.92% | 40,778             | 15.40% | 477,155     | 1.18  |
|                       | <6            | 812,439                | 24.06% | 58,599             | 22.14% | 871,038     | 0.93  |
|                       | 6–12          | 2,486,604              | 73.65% | 197,698            | 74.68% | 2,684,302   | 1.01  |
|                       | 12–18         | 70,336                 | 2.08%  | 8122               | 3.07%  | 78,458      | 1.42  |
|                       | >18           | 6911                   | 0.20%  | 313                | 0.12%  | 7224        | 0.60  |
| Curvature             | Concave       | 1,341,951              | 39.75% | 105,133            | 39.71% | 1,447,084   | 1.00  |
|                       | Flat          | 689,221                | 20.41% | 55,595             | 21.00% | 744,816     | 1.03  |
|                       | Convex        | 1,345,078              | 39.84% | 104,044            | 39.30% | 1,449,122   | 0.99  |
| SPI ( $\times 10^4$ ) | <15.78        | 1,029,734              | 30.50% | 2,447,736          | 93.58% | 3,477,470   | 0.98  |
|                       | 15.78–1432.47 | 138,638                | 4.11%  | 16,422             | 6.20%  | 155,060     | 1.46  |
|                       | >1432.47      | 7898                   | 0.23%  | 574                | 0.22%  | 8472        | 0.93  |
| STI                   | <35           | 887,114                | 26.27% | 60,897             | 23.00% | 948,011     | 0.88  |
|                       | 35–600        | 2,320,987              | 68.74% | 185,085            | 69.91% | 2,506,072   | 1.02  |
|                       | 600–9509      | 162,241                | 4.81%  | 18,227             | 6.89%  | 180,468     | 1.39  |
|                       | >9509         | 5948                   | 0.18%  | 523                | 0.20%  | 6471        | 1.11  |
| Topographic Relief    | 0–10          | 707,752                | 20.96% | 52,211             | 19.72% | 759,963     | 0.94  |
|                       | 10–20         | 2,052,327              | 60.79% | 170,331            | 64.34% | 2,222,658   | 1.05  |
|                       | 20–30         | 552,583                | 16.37% | 37,517             | 14.17% | 590,100     | 0.87  |
|                       | 30–40         | 54,636                 | 1.62%  | 2928               | 1.11%  | 57,564      | 0.70  |
|                       | >40           | 8992                   | 0.27%  | 1745               | 0.66%  | 10,737      | 2.24  |

Table 2. Cont.

| Factors           | Class          | Landslide Not Occurred |        | Landslide Occurred |        | Total Count | FR   |
|-------------------|----------------|------------------------|--------|--------------------|--------|-------------|------|
|                   |                | Count                  | Ratio  | Count              | Ratio  |             |      |
| Rainfall          | 303.68–439.10  | 684,222                | 20.27% | 120,189            | 45.40% | 804,411     | 2.05 |
|                   | 439.10–571.60  | 940,470                | 27.86% | 93,156             | 35.19% | 1,033,626   | 1.24 |
|                   | 571.60–704.10  | 751,015                | 22.24% | 40,886             | 15.44% | 791,901     | 0.71 |
|                   | 704.10–836.60  | 526,648                | 15.60% | 10,501             | 3.97%  | 537,149     | 0.27 |
|                   | 836.60–969.10  | 364,558                | 10.80% | 0                  | 0.00%  | 364,558     | 0.00 |
|                   | 969.10–1071.92 | 109,377                | 3.24%  | 0                  | 0.00%  | 109,377     | 0.00 |
| Vegetation        | Bare Soil      | 684,222                | 20.27% | 120,189            | 45.40% | 804,411     | 2.05 |
|                   | Brush-forbs    | 1,410,763              | 41.78% | 126,123            | 47.64% | 1,536,886   | 1.13 |
|                   | Woods          | 831,805                | 24.64% | 18,420             | 6.96%  | 850,225     | 0.30 |
|                   | Grassland      | 340,123                | 10.07% | 0                  | 0.00%  | 340,123     | 0.00 |
|                   | Snow           | 109,377                | 3.24%  | 0                  | 0.00%  | 109,377     | 0.00 |
| NDVI              | −0.378–0.038   | 369,510                | 10.94% | 45,345             | 17.13% | 414,855     | 1.50 |
|                   | 0.038–0.149    | 901,201                | 26.69% | 141,964            | 53.63% | 1,043,165   | 1.87 |
|                   | 0.149–0.272    | 799,708                | 23.69% | 51,241             | 19.36% | 850,949     | 0.83 |
|                   | 0.272–0.412    | 734,348                | 21.75% | 18,177             | 6.87%  | 752,525     | 0.33 |
|                   | 0.412–0.705    | 571,523                | 16.93% | 8005               | 3.02%  | 579,528     | 0.19 |
| Distance-to-river | 0–300          | 260,359                | 7.71%  | 30,297             | 11.44% | 290,656     | 1.43 |
|                   | 300–600        | 222,237                | 6.58%  | 52,314             | 19.76% | 274,551     | 2.62 |
|                   | 600–900        | 210,492                | 6.23%  | 46,048             | 17.39% | 256,540     | 2.47 |
|                   | 900–1200       | 204,869                | 6.07%  | 39,290             | 14.84% | 244,159     | 2.21 |
|                   | 1200–1500      | 198,806                | 5.89%  | 31,041             | 11.73% | 229,847     | 1.86 |
|                   | >1500          | 2,279,527              | 67.52% | 65,742             | 24.83% | 2,345,269   | 0.39 |
| Distance-to-fault | 0–300          | 522,093                | 15.46% | 79,029             | 29.85% | 601,122     | 1.81 |
|                   | 300–600        | 510,160                | 15.11% | 75,870             | 28.66% | 586,030     | 1.78 |
|                   | 600–900        | 297,156                | 8.80%  | 38,661             | 14.60% | 335,817     | 1.58 |
|                   | 900–1200       | 533,958                | 15.81% | 46,486             | 17.56% | 580,444     | 1.10 |
|                   | 1200–1500      | 348,652                | 10.33% | 18,600             | 7.03%  | 367,252     | 0.70 |
|                   | 1500–1800      | 286,178                | 8.48%  | 5486               | 2.07%  | 291,664     | 0.26 |
|                   | 1800–2100      | 207,871                | 6.16%  | 600                | 0.23%  | 208,471     | 0.04 |
|                   | >2100          | 670,222                | 19.85% | 0                  | 0.00%  | 670,222     | 0.00 |

As for the SPI, the highest FR value was found to be related to the class of 15.78–1432.47 (1.46), followed by the classes of <15.78 (0.98) and >1432.47 (0.93). Over 90% of the landslide area was found in the class of <15.78. The FR values for each class of the STI were 0.88, 1.02, 1.39, and 1.11, respectively. Over 69% of the landslide area was found in the class of 35–600. The FR values of the topographic relief classes of 0–10, 10–20, 20–30, 30–40, and >40 were 0.94, 1.05, 0.87, 0.70, and 2.24, respectively, which means that the class of >40 was favorable for landslide hazards. Over 60% of the landslide area was found in the class of 20–30. As for the rainfall, 45.40% of the landslide area was found in the class of 303.68–439.10, 35.19% of the landslide area was found in the class of 439.10–571.60, and 15.44% of the landslide area was found in the class of 571.60–704.10. The classes of 303.68–439.10 and 439.10–571.60 had positive FR values (2.05 and 1.24, respectively). For the rainfall class of >571.60, the FR value was negative. As for the vegetation, 45.40% of the landslide areas were found in the bare soil zones, and 47.64% of the landslide areas were found in brush-forbs zones. The FR value of the vegetation classes of bare soil, brush-forbs, woods, grassland, and snow were 2.05, 1.13, 0.30, 0.00, and 0.00, respectively. The NDVI was between −0.378 and 0.705. More than 90% of the landslide area had an NDVI value below 0.272. The classes of −0.378–0.038 and 0.038–0.149 had positive FR values (1.50 and 1.87, respectively). The landslides were mainly distributed within 0–1500 m of the river, and 0–1200 m of the faults.

#### 4.2. Result of the PCA

The adjustment of the logistic regression model is sensitive to the linear correlation of the influencing factors. In this study, the PCA was used to eliminate the linear correlation between the influencing factors. First, all of the influencing factors were normalized. Next, a 20 × 20 m fishnet



was built in the study area, so as to sample 13 preselected factors. A total of 12,748,442 sampling points were obtained. Then, the Kaiser–Meyer–Olkin (KMO) test and the Bartlett’s test of the sample data were carried out. The test results are shown in Table 3. It can be seen from Table 3, that the KMO test value is 0.640 and the  $p$ -value is  $<0.05$ , which shows that there was a certain correlation between the influencing factors, and it was suitable for the PCA.

**Table 3.** Results of the Kaiser–Meyer–Olkin (KMO) test and the Bartlett’s test.

| KMO test        | 0.640         |
|-----------------|---------------|
| Bartlett’s test | 8,177,019.716 |
| $p$ -value      | 0.000         |

The correlation matrix of the influencing factors is shown in Table 4. As can be seen from Table 4, the correlation coefficient between the slope angle and the topographic relief was 0.90, the correlation coefficient between the SPI and STI was 0.96, and the correlation coefficients between the rainfall and vegetation was 0.95. These results show that there was a high correlation between some of the influencing factors. In other words, there were extra elements between the preselected influencing factors.

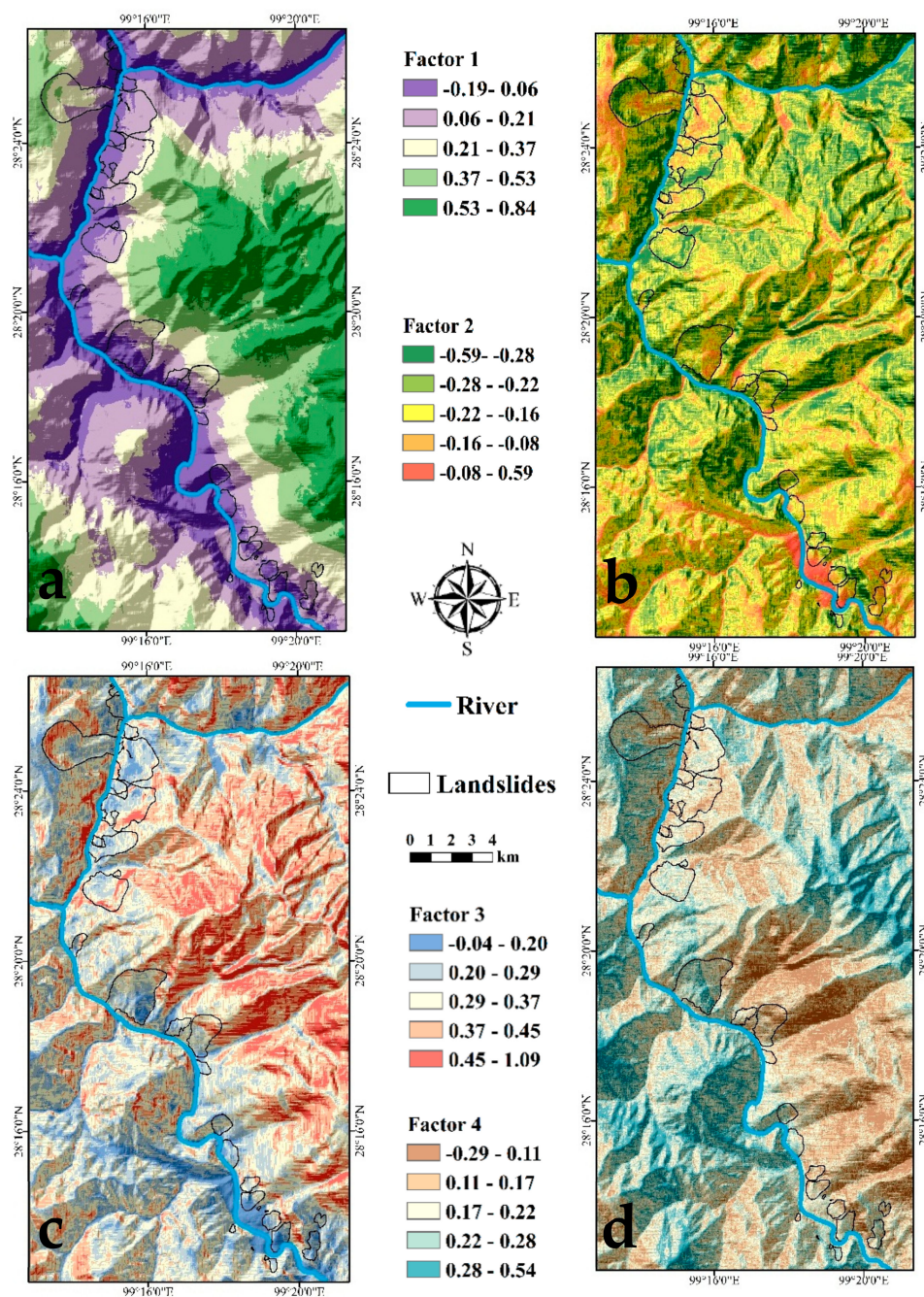
According to the eigenvalue of the correlation matrix, six principal components with eigenvalues greater than 0.9 were selected. In general, the higher the eigenvalue, the greater the difference reflected by the principal component, and the more the actual information of the preselected influencing factors can be retained. From Table 5, it can be seen that the sum of the variance contribution rates of the six principal components was 82.36%, which means that they extracted 82.36% of the information from the original data.

In the process of dimension reduction, PCA is used. The newly generated factors are liner combinations of the preselected influencing factors. According to the component score coefficient matrix, shown in Table 6, six new factors can be obtained, namely, Factor 1, Factor 2, Factor 3, Factor 4, Factor 5, and Factor 6. In the matrix, the higher the coefficient of the preselected influencing factor was, the higher the correlation between the new factors and the preselected influencing factors was. The new factor maps are shown in Figures 8 and 9.

**Table 4.** The correlation matrix of the influencing factors.

| Factors | F1    | F2    | F3    | F4    | F5    | F6    | F7    | F8    | F9    | F10   | F11   | F12   | F13   |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| F1      | 1.00  | 0.12  | −0.08 | −0.03 | 0.00  | 0.01  | 0.02  | 0.12  | −0.40 | −0.38 | −0.27 | −0.34 | −0.41 |
| F2      | 0.12  | 1.00  | −0.01 | −0.25 | 0.01  | −0.02 | 0.00  | 0.90  | −0.06 | −0.07 | −0.04 | −0.13 | −0.02 |
| F3      | −0.08 | −0.01 | 1.00  | 0.00  | 0.00  | 0.00  | 0.00  | −0.01 | 0.09  | 0.07  | 0.08  | 0.02  | 0.09  |
| F4      | −0.03 | −0.25 | 0.00  | 1.00  | −0.28 | 0.19  | 0.31  | −0.26 | −0.08 | −0.07 | −0.04 | −0.04 | −0.01 |
| F5      | 0.00  | 0.01  | 0.00  | −0.28 | 1.00  | −0.02 | −0.05 | 0.01  | 0.03  | 0.02  | 0.01  | 0.01  | 0.00  |
| F6      | 0.01  | −0.02 | 0.00  | 0.19  | −0.02 | 1.00  | 0.96  | −0.02 | −0.04 | −0.04 | −0.04 | −0.04 | −0.01 |
| F7      | 0.02  | 0.00  | 0.00  | 0.31  | −0.05 | 0.96  | 1.00  | 0.00  | −0.06 | −0.05 | −0.04 | −0.05 | −0.01 |
| F8      | 0.12  | 0.90  | −0.01 | −0.26 | 0.01  | −0.02 | 0.00  | 1.00  | −0.07 | −0.09 | −0.05 | −0.15 | −0.02 |
| F9      | −0.40 | −0.06 | 0.09  | −0.08 | 0.03  | −0.04 | −0.06 | −0.07 | 1.00  | 0.95  | 0.48  | 0.80  | 0.38  |
| F10     | −0.38 | −0.07 | 0.07  | −0.07 | 0.02  | −0.04 | −0.05 | −0.09 | 0.95  | 1.00  | 0.41  | 0.77  | 0.34  |
| F11     | −0.27 | −0.04 | 0.08  | −0.04 | 0.01  | −0.04 | −0.04 | −0.05 | 0.48  | 0.41  | 1.00  | 0.35  | 0.34  |
| F12     | −0.34 | −0.13 | 0.02  | −0.04 | 0.01  | −0.04 | −0.05 | −0.15 | 0.80  | 0.77  | 0.35  | 1.00  | 0.36  |
| F13     | −0.41 | −0.02 | 0.09  | −0.01 | 0.00  | −0.01 | −0.01 | −0.02 | 0.38  | 0.34  | 0.34  | 0.36  | 1.00  |

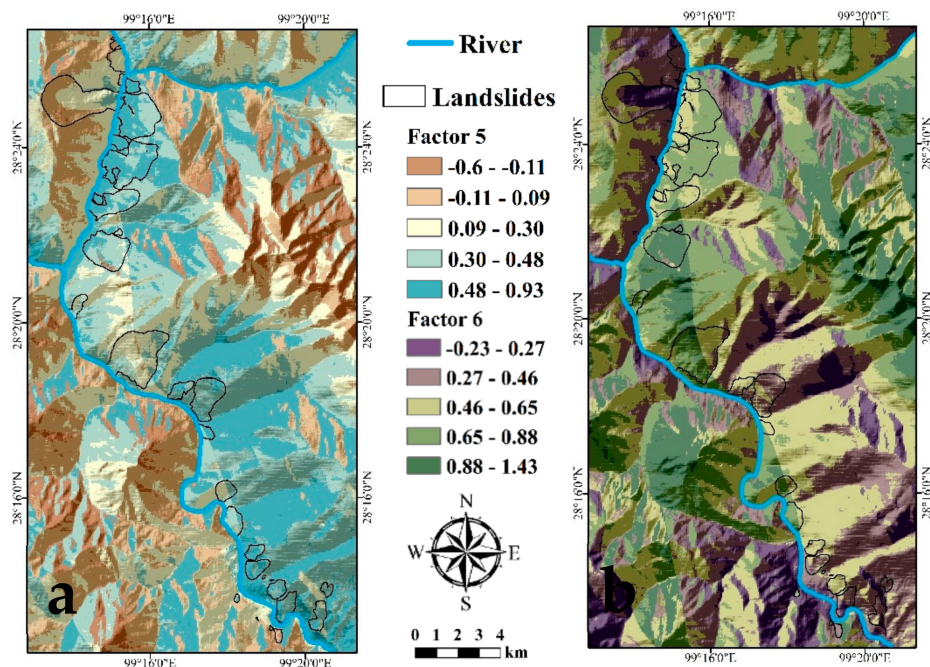
Notes: F1—lithology; F2—slope angle; F3—slope aspect; F4—TWI; F5—curvature; F6—SPI; F7—STI; F8—topographic relief; F9—rainfall; F10—vegetation; F11—NDVI; F12—distance-to-river; F13—distance-to-fault.



**Figure 8.** Influencing factor maps selected using principal component analysis (PCA): (a) Factor 1; (2) Factor 2; (3) Factor 3; and (b) Factor 4.

**Table 5.** Total variance explained.

| Components | Initial Eigenvalues |               |              | Extraction Sums of Squared Loadings |               |              |
|------------|---------------------|---------------|--------------|-------------------------------------|---------------|--------------|
|            | Total               | % of Variance | Cumulative % | Total                               | % of Variance | Cumulative % |
| 1          | 3.506               | 26.969        | 26.969       | 3.506                               | 26.969        | 26.969       |
| 2          | 2.190               | 16.843        | 43.812       | 2.190                               | 16.843        | 43.812       |
| 3          | 1.873               | 14.409        | 58.220       | 1.873                               | 14.409        | 58.220       |
| 4          | 1.146               | 8.813         | 67.034       | 1.146                               | 8.813         | 67.034       |
| 5          | 1.038               | 7.985         | 75.019       | 1.038                               | 7.985         | 75.019       |
| 6          | 0.910               | 6.997         | 82.016       | 0.910                               | 6.997         | 82.016       |
| 7          | 0.724               | 5.568         | 87.584       | -                                   | -             | -            |
| 8          | 0.624               | 4.797         | 92.382       | -                                   | -             | -            |
| 9          | 0.570               | 4.384         | 96.766       | -                                   | -             | -            |
| 10         | 0.252               | 1.939         | 98.705       | -                                   | -             | -            |
| 11         | 0.096               | 0.737         | 99.441       | -                                   | -             | -            |
| 12         | 0.040               | 0.310         | 99.751       | -                                   | -             | -            |
| 13         | 0.032               | 0.249         | 100.000      | -                                   | -             | -            |

**Table 6.** Component score coefficient matrix.

| Factors | 1      | 2      | 3      | 4      | 5      | 6      |
|---------|--------|--------|--------|--------|--------|--------|
| F1      | −0.577 | −0.076 | −0.019 | 0.112  | −0.299 | 0.447  |
| F2      | −0.207 | −0.588 | 0.717  | −0.177 | −0.034 | −0.003 |
| F3      | 0.126  | 0.005  | 0.032  | −0.115 | 0.798  | 0.573  |
| F4      | −0.054 | 0.611  | −0.091 | −0.486 | −0.061 | −0.004 |
| F5      | 0.032  | −0.195 | −0.001 | 0.850  | 0.198  | −0.112 |
| F6      | −0.102 | 0.719  | 0.619  | 0.223  | 0.012  | 0.009  |
| F7      | −0.117 | 0.746  | 0.626  | 0.139  | 0.004  | 0.007  |
| F8      | −0.220 | −0.590 | 0.714  | −0.174 | −0.020 | −0.012 |
| F9      | 0.925  | −0.052 | 0.140  | 0.035  | −0.176 | 0.199  |
| F10     | 0.895  | −0.038 | 0.124  | 0.047  | −0.210 | 0.230  |
| F11     | 0.592  | −0.040 | 0.097  | −0.069 | 0.132  | −0.107 |
| F12     | 0.842  | 0.016  | 0.057  | 0.045  | −0.251 | 0.165  |
| F13     | −0.577 | −0.076 | −0.019 | 0.112  | −0.299 | 0.447  |

Notes: F1—lithology; F2—slope angle; F3—slope aspect; F4—TWI; F5—curvature; F6—SPI; F7—STI; F8—topographic relief; F9—rainfall; F10—vegetation; F11—NDVI; F12—distance-to-river; F13—distance-to-fault.



### 4.3. Landslide Probability

The six factors that were obtained using the PCA will be introduced into the logistic regression analysis. We checked the significance of each factor. We will retain only the significant factor that had a  $p$ -value less than 0.05. In other words, we will exclude the factors with a  $p$ -value more than 0.05 from the model. The first regression analysis results (Table 7) show that the  $p$ -value of Factor 4 was greater than 0.05, so it was excluded from the logistic regression analysis model.

**Table 7.** The first logistic regression analysis results.

| Factors    | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 |
|------------|----------|----------|----------|----------|----------|----------|
| $P$ -value | 0.000    | 0.042    | 0.000    | 0.784    | 0.000    | 0.000    |

The Cox and Snell pseudo  $R^2$  test and the Nagelkerke pseudo  $R^2$  test were used to evaluate the goodness of the fit of the logistic regression model. For the final logistic regression model, the Cox and Snell pseudo  $R^2$  value was 0.233, and the Nagelkerke pseudo  $R^2$  value was 0.310 (Table 8). Both the Cox and Snell  $R^2$  value and the Nagelkerke pseudo  $R^2$  value were greater than 0.200, which indicates that the fitting result was good.

**Table 8.** Results of the Cox and Snell pseudo  $R^2$  test and the Nagelkerke pseudo  $R^2$  test.

| Pseudo $R^2$ test               | value |
|---------------------------------|-------|
| Cox and Snell pseudo $R^2$ test | 0.233 |
| Nagelkerke pseudo $R^2$         | 0.310 |

In the final logistic regression model, Factor 1, Factor 2, Factor 3, Factor 5, and Factor 6, were introduced into the logistic regression analysis model. In this study, the odds ratio was used to assess the relationship of the factors and the landslide susceptibility. If the odds ratio value of the factors is greater than 1, it means that the factors are related to landslide susceptibility. If the odds value of the factors is equal to 1, it means that the factors are neutral with landslide susceptibility. If the odds ratio value is less than 1, it means that the factors are negated with landslide susceptibility. From Table 9, it can be seen that Factor 2 and Factor 5 were related with landslide susceptibility, while Factor 1, Factor 3, and Factor 6 were negated with landslide susceptibility.

**Table 9.** The factors estimated coefficients.

| Factors  | $B_G$  | Standard Error of Estimate | Wald $\chi^2$ Value | $p$ -Value | Odds Ratio |
|----------|--------|----------------------------|---------------------|------------|------------|
| Factor 1 | −5.370 | 0.036                      | 21,795.771          | 0.000      | 0.005      |
| Factor 2 | 0.478  | 0.168                      | 8.081               | 0.004      | 1.613      |
| Factor 3 | −0.859 | 0.131                      | 42.868              | 0.000      | 0.424      |
| Factor 5 | 2.324  | 0.019                      | 14,953.978          | 0.000      | 10.215     |
| Factor 6 | −0.538 | 0.017                      | 991.685             | 0.000      | 0.584      |
| Constant | 0.925  | 0.016                      | 3183.937            | 0.000      | 2.522      |

Using Equation (1), we calculated the predicted probability of landslides for the entirety of the study area. The result was a raster map and the value of each pixel of the map represents the estimated probability of landslide occurrence. The map was divided into the following five classes: very high, high, moderate, low, and very low (Figure 10). Table 10 shows that the areas of the five classes are 62.93, 98.48, 65.55, 81.89, and 84.25 km<sup>2</sup>, respectively.

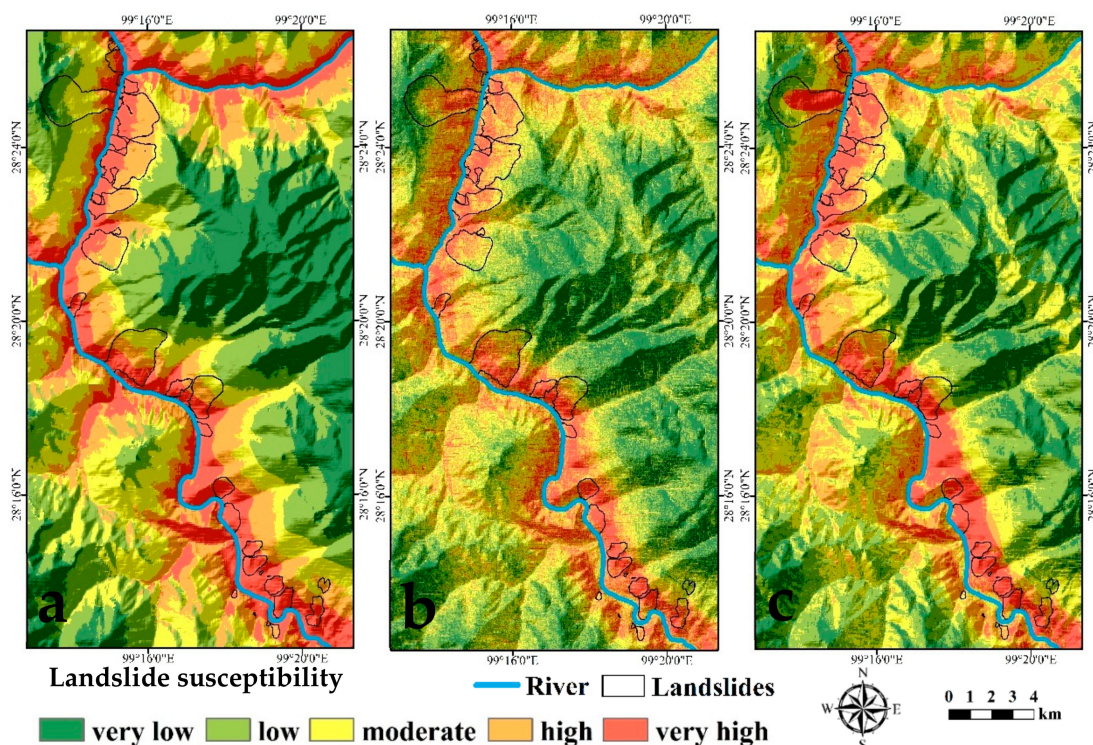
In order to establish a more accurate model of landslide susceptibility, the FR method and the analytic hierarchy process (AHP) were also used in this study. As the application of the FR method and the AHP in landslide susceptibility modeling is quite known, the theory was not introduced in this study. This study only lists the evaluation results of the FR method and the AHP



method. The landslide susceptibility maps of the FR method and the AHP method were also divided into five classes using the natural breaks method (Figure 10). Table 10 shows that the areas of the five susceptibility classed of the AHP method (very high, high, moderate, low, and very low) were 40.14, 70.10, 88.99, 103.14, and 61.73 km<sup>2</sup>, respectively. For the FR method, they were 35.08, 74.00, 84.82, 101.67, and 68.53 km<sup>2</sup>, respectively.

**Table 10.** Statistical results of the landslide susceptibility mapping. PCA-LR—principal component analysis logistic regression; FR—frequency ratio; AHP—analytic hierarchy process.

| Models | Susceptibility | Landslide Occurred |        |                         | Total Study Area |        |                         | Prediction Accuracy |
|--------|----------------|--------------------|--------|-------------------------|------------------|--------|-------------------------|---------------------|
|        |                | Count              | Ratio  | Area (km <sup>2</sup> ) | Count            | Ratio  | Area (km <sup>2</sup> ) |                     |
| PCA-LR | Very Low       | 8021               | 3.02%  | 0.80                    | 842549           | 23.14% | 84.25                   | 83.4%               |
|        | Low            | 1625               | 6.12%  | 1.63                    | 818895           | 22.49% | 81.89                   |                     |
|        | Moderate       | 33,901             | 12.76% | 3.39                    | 655499           | 18.00% | 65.55                   |                     |
|        | High           | 88,080             | 33.15% | 8.81                    | 694770           | 19.08% | 69.48                   |                     |
|        | Very High      | 1,184,800          | 44.59% | 11.85                   | 629309           | 17.28% | 62.93                   |                     |
| AHP    | Very Low       | 2441               | 0.92%  | 0.24                    | 617269           | 16.95% | 61.73                   | 76.9%               |
|        | Low            | 16,843             | 6.34%  | 1.68                    | 1031436          | 28.33% | 103.14                  |                     |
|        | Moderate       | 29,421             | 11.07% | 2.94                    | 889896           | 24.44% | 88.99                   |                     |
|        | High           | 76,814             | 28.91% | 7.68                    | 701028           | 19.25% | 70.10                   |                     |
|        | Very High      | 139,213            | 52.39% | 13.92                   | 401393           | 11.02% | 40.14                   |                     |
| FR     | Very Low       | 4774               | 1.80%  | 0.48                    | 685253           | 18.82% | 68.53                   | 79.9%               |
|        | Low            | 18,598             | 7.00%  | 1.86                    | 1016745          | 27.92% | 101.67                  |                     |
|        | Moderate       | 44,106             | 16.60% | 4.41                    | 848232           | 23.30% | 84.82                   |                     |
|        | High           | 101,138            | 38.06% | 10.11                   | 740007           | 20.32% | 74.00                   |                     |
|        | Very High      | 96,116             | 36.17% | 9.61                    | 350785           | 9.63%  | 35.08                   |                     |

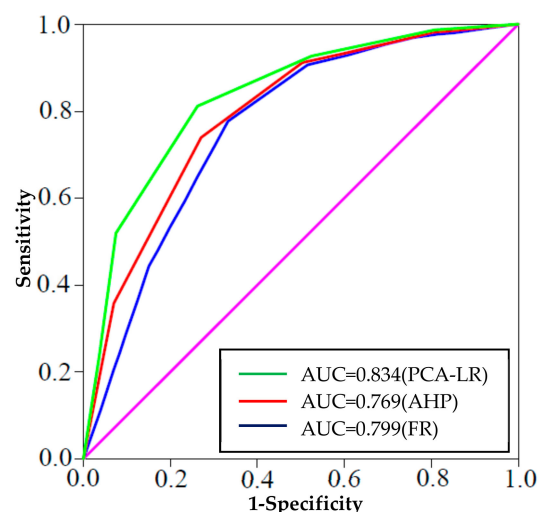


**Figure 10.** Landslide susceptibility map: (a) PCA-logistic regression (LR) method; (b) FR method; and (c) analytic hierarchy process (AHP) method.

## 5. Discussion

### 5.1. Validation

The validation is very important for landslide susceptibility mapping. Without validation, the landslide susceptibility model will have no meaning. In order to verify the quality of the prediction and the stability of the model, the ROC curve has been used to estimate the model's accuracy, which is used as a quantitative measurement. The ROC curves of the model built in this study are shown in Figure 11. From Figure 11, it can be seen that the AUCs of the PCA-LR model, AHP model, and FR model were 0.834, 0.769, and 0.799, respectively. Many studies have introduced the traditional academic point system into the accuracy ranking, and they have suggested that the accuracy rate between 0.90 and 1.00 is excellent, the accuracy rate between 0.80 and 0.90 is good, the accuracy rate between 0.70 and 0.80 is fair, the accuracy rate between 0.60 and 0.70 is poor, and the accuracy rate between 0.50 and 0.60 is failing [62,63]. Thus, the accuracy rate of the PCA-LR model fell within the “good” classification category, and the accuracy rate of the FR model and the AHP model fell within the “fair” classification category. We also compared our results with other studies in similar areas. The prediction accuracy of the landslide susceptibility model of the Xulong reservoir, which is similar to this area, based on the combination of the information content method and the hierarchical analysis method established by Caochen [24] is 85.74%. This result is basically equal to the result of the PCA-LR model established in this paper, and the prediction accuracy of this model is the highest. Therefore, the subsequent discussion in this paper is based on the PCA-LR model.



**Figure 11.** Receiver operating characteristic (ROC) curve of the model.

### 5.2. Key Factors for Landslide Occurrence

The landslide susceptibility mapping should not only produce the landslide susceptibility map, but also identify the main factors of landslide occurrence, and evaluate the contribution and influence of these factors. In order to establish the landslide susceptibility model, we adopted an FR method to analyze the correlation between landslide occurrence and preselected factors, using PCA to eliminate the multicollinearity between the preselected factors. Finally, the 13 preselected factors were reduced to six factors, and the landslide susceptibility model was established using logistic regression. For the logistic regression model, the odds ratios ( $\text{Exp}(\beta_G)$ ) can be used to measure the correlation between the factors and the landslide occurrence. The component score coefficient of the PCA shows the extent of the correlation between the principal components and the preselected factors. The FR method can reflect the correlation between each class of each preselected factor and landslide occurrence. Based on the above discussion, we can find the combination of the most favorable factors for landslide occurrence.

Because the odds ratios of Factor 2 (1.613) and Factor 5 (10.215) (Table 9) are greater than 1, this indicates that Factor 2 and Factor 5 play a major role in landslide occurrence in the study area. From Table 6, it can be seen that the slope angle (−0.588), TWI (0.611), SPI (0.719), STI (0.746), and topographic relief (−0.590) are the preselected factors with the highest correlation for Factor 2, and the lithology (0.299), slope aspect (−0.588), vegetation (0.132), rainfall (−0.210), NDVI (−0.251), distance-to-river (0.798), and distance-to-fault (−0.299) are the preselected factors with the highest correlation for Factor 5. This means that these preselected factors have a stronger effect on landslide occurrence than the other factors.

**Lithology:** From Table 2, it can be seen that the lithologies with an FR value greater than 1 are  $Q_h^{del}$ ,  $Q_3^P$ ,  $T_{2q}^1$ ,  $P_{2g}$ , and  $D_{2q}$ . The lithology of these strata is mainly limestone, volcanic rock, slate, and green schist. In the area where the landslides densely occurred in the study area, the thickness of the rock mass is thin or medium–thin. The rock mass is cut by joints and fractures, and these discontinuities are highly developed. Thus, the local gravity deformation of the rock mass is serious. It is common to see the bending phenomena and tearing deformation phenomena in the rock mass. Therefore, these factors provide favorable conditions for landslide occurrence.

**Slope angle and distance-to-river:** According to the FR value, the slope angle classes that were the most prone to landslide were 20–40 and 60–70. As for the distance-to-river, the class that was most prone to landslide was the class of 0–1500. The study area is located in a rapidly uplifting region [64]. According to previous research, it was shown that the annual uplift rate of the study area was  $5.8 \pm 1.0$  mm from 1970 to 2012 [65]. The rapid uplift give rise to a rapid river incision. Under the combined action of the bedrock uplift and the river incision, the slope along the river become steeper [24,25,34,66]. In this case, landslides can make the slopes adjust to the rapid river incision quickly [66]. So, a large number of landslides have occurred along the Jinsha River and Dingqu River.

**Slope aspect:** The FR value of the slope aspect classes show that the areas facing the SW, W, and NW have higher probabilities of landslide occurrence. The slope aspect usually affects the slope structure of the rock mass. The altitude of rock dipping toward to the inner slope was more prone to bending, and the dip bedded rock slope is more prone to landslide occurrence than the escarpment slope is.

**TWI, SPI, and STI:** From Table 2, it can be seen the TWI classes of 6–12 and 12–18, the SPI class of 1.58–1432.47, and the STI classes of 35–600, 600–9509, and >9509 had a positive effect on the landslide occurrence. These factors reflect the hydrologic condition of the study area. The smaller the TWI value, the lower the moisture. The higher TWI value symbolizes a higher order water channel. In this study, the TWI classes of 6–12 and 12–18 represent a lower order drainage, which is vulnerable to instability. As for SPI and STI, a high value is indicative of water contributions from upslope and high water flow velocities, and the effect of topography on erosion, which are directly linked to landslide occurrence [67].

**Topographic relief:** The FR value of the topographic relief shows that the class of >40 was favorable for landslide hazards.

**Rainfall, vegetation, and NDVI:** As for the rainfall, vegetation, and NDVI, the classes most prone to landslide were the class of 303.68–439.10 (rainfall), classes bare soil and brush-forbs (vegetation), and the class −0.378–0.149 (NDVI). It can be seen that the area with high rainfall has fewer landslides. The reason for this was that the annual rainfall in the study was generally low (around 300 mm in the low elevation and 1000 mm in the high elevation), and the precipitation commonly occurred as snowfall in the high elevation area (elevation more than 4800 m) [12]. Therefore, it is difficult to have effective rainfall in a short period, leading to the occurrence of landslides. Second, because of the close relationship between the vertical distribution of the precipitation and the distribution of the vegetation, the distribution of the vegetation also followed a vertical distribution law. The low elevation areas have little vegetation, because of the low rainfall (the low and high elevation areas has a low NDVI value and the moderate elevation area have a high NDVI value). Areas without vegetation cover are more prone to cause the landslides.

**Distance-to-fault:** The landslides mainly occurred within 0–1200 m of the faults. In the faulted zone, the rock is relatively broken and the joint fracture is developed, which makes the slope of these areas less stable and more prone to landslide occurrence [48].

In summary, the factors that are most favorable to landslide occurrence are as follows: (a) lithology:  $Q_h^{del}$ ,  $Q_3^p$ ,  $T_{2q}^1$ ,  $P_{2g}$ , and  $D_{2q}$ ; (b) slope angle: 20–40 and 60–70; (c) slope aspect: SW, W, and NW; (d) TWI: 6–18; (e) SPI: 1.58–1432.47; (f) STI: >35; (g) topographic relief: >40; (h) rainfall: 303.68–571.60; (i) vegetation: bare soil and brush-forbs; (j) distance-to-river: 0–1500 m; (k) distance-to-fault: 0–1200 m.

### 5.3. Landslide Susceptibility Mapping

The landslide susceptibility map (PCA-LR model) is shown in Figure 6, and the statistical results of the landslide susceptibility mapping are shown in the Table 10. It can be seen from Figure 5 and Table 10 that the very low and the low susceptibility areas had an area of 84.25 km<sup>2</sup> and 81.89 km<sup>2</sup>, accounting for 23.14% and 22.49% of the total study area, respectively. This region is mainly distributed in the high and moderate elevation area. The strata of this area are  $T_{3j}^1$ ,  $T_{2q}^3$ ,  $P_2$ ,  $P_1^b$ ,  $P_1^a$ , and  $P_1^r$ . The vegetation is mainly woods, grassland, and snow. The NDVI value of this area is high, which means that this area has a high vegetation coverage. The rainfall is high relative to the whole study area, but the rainfall commonly occurred as snowfall in the high elevation area. This area is far away from the rivers and has little erosion from the rivers. The moderate, high, and very high susceptibility areas had an area of 65.55 km<sup>2</sup>, 69.48 km<sup>2</sup>, and 62.93 km<sup>2</sup>, accounting for 18.00%, 19.08%, and 17.28%, respectively. The vegetation of this area is mainly brush-forbs and bare soil. The NDVI value of this area is low, which indicates that this area has low vegetation coverage. The strata of this area are  $Q_h^{del}$ ,  $Q_3^p$ ,  $T_{2q}^1$ ,  $P_{2g}$ , and  $D_{2q}$ . This area is close to the rivers and faults. In the faulted zone, the rock is relatively broken and the joint fracture is developed, which makes the slope of these areas less stable, and landslides are more likely to occur. The study area belongs to the rapidly uplifting region, and the interaction between the bedrock uplift and river incision made the landslides occur widely along rivers.

As for the landslide occurrence, the very low, low, moderate, high, and very high susceptibility areas had an area of 0.80 km<sup>2</sup>, 1.63 km<sup>2</sup>, 3.39 km<sup>2</sup>, 8.81 km<sup>2</sup>, and 11.85 km<sup>2</sup>, accounting for 3.02%, 6.12%, 12.76%, 33.15%, and 44.59% of the entire landslide area, respectively. The moderate, high, and very high susceptibility area make up 90.86% of the total landslide area. According to the field survey, the landslide mainly occurred within the high and very high susceptibility ranges. Hence, the landslide susceptibility map that was produced in this study is reasonable.

The landslide susceptibility map shows that the very high, high, and moderate susceptibility areas are mainly distributed in Guxue town, Taentong village, Yongduo village, Rancun village, Deze village, Aluogong village, Jiaxue village, Senen village, Benzilan town, Waka town, and so on, all of which are located near both sides of the Jinsha River and Dingqu River. These villages are densely populated, with a high density of buildings and cultivated land, and some villages have developed industries. Moreover, these villages are located in the high susceptibility areas of the landslide occurrence, so these villages suffer a higher degree of landslide hazards. Therefore, there should be a focus on disaster reduction and prevention in these villages. The low and very low susceptibility areas are mainly distributed in the regions far away from the Jinsha River and Dingqu River. Human activities in this area are relatively weak, and even if landslide occurs, the damage is relatively small.

In general, the areas with very high, high, and moderate susceptibility to landslide occurrence are mainly distributed in the areas with intensive human activities, so disaster prevention and reduction should be emphasized. Human activities are sparse in the areas with a low and very low susceptibility to landslide occurrence, and the potential threat caused by landslide disasters is small or harmless, but the prevention of disaster risk reduction should also be done.



## 6. Conclusions

According to the field survey, the mechanism of the landslide, the local geo-environmental conditions, and the previous studies, 13 influencing factors, including (a) lithology, (b) slope angle, (c) slope aspect, (d) TWI, (e) curvature, (f) SPI, (g) STI, (h) topographic relief, (i) rainfall, (j) vegetation, (k) NDVI, (l) distance-to-river, and (m) distance-to-fault, were selected to produce the landslide susceptibility map in this study. In order to clarify the relationship between the landslides and the influencing factors, the FR model was used to describe their relationship, because the adjustment of the logistic regression model is sensitive to the linear correlation of the influencing factors. In this paper, the principal component analysis (PCA) is used to reduce the dimension of the preselected influencing factors and to change the factors, which are then reselected, so as to make them independent of each other. According to the eigenvalue of the correlation matrix, six principal components with eigenvalues greater than 0.9 were selected. The sum of variance contribution rates of the six principal components was 82.36%, which means that they extracted 82.36% of the information of the original data. As for the logistic regression analysis, the  $p$ -value was used to check the significance of the six factors obtained using the PCA. The factors with a  $p$ -value more than 0.05 were excluded from the LR model. Because the  $P$ -value of Factor 4 is 0.784, it was excluded from the model. The odds ratio was used to assess the relationship of the six factors and landslide susceptibility. It can be seen that Factor 2 and Factor 5 were related with landslide susceptibility, while Factor 1, Factor 3, and Factor 6 were negated with landslide susceptibility. The slope angle, TWI, SPI, STI, and topographic relief are the preselected factors with the highest correlation with Factor 2, and the lithology, slope aspect, vegetation, rainfall, distance-to-river, and distance-to-fault are the preselected factors with the highest correlation with Factor 5. These factors have been identified as key factors in the occurrence of landslides. The Cox and Snell pseudo  $R^2$  test and the Nagelkerke pseudo  $R^2$  test were used to evaluate the goodness of the fit of the logistic regression model. Both the Cox and Snell  $R^2$  value and Nagelkerke pseudo  $R^2$  value were greater than 0.200, which indicates that the fitting result was good. The landslide susceptibility map that was produced by the logistic regression model was divided into the following five classes using the natural breaks method: very low, low, moderate, high, and very high. The ratios of the areas of the susceptibility classes were 23.14%, 22.49%, 18.00%, 19.08%, and 17.28%, respectively. The total proportion of the landslide pixels of the moderate, high, and very high susceptibility area was 90.86%. The validation result shows that the prediction accuracy of the model was 84.9%, which means that the landslide susceptibility map was reliable and reasonable. Consequently, this study could serve as an effective guide for further land use planning and for the implementation of development.

**Author Contributions:** X.S. contributed to the data analysis and manuscript writing. J.C. proposed the main structure of this study. Y.B., X.H., J.Z., and W.P. provided useful advice and revised the manuscript. All of the authors read and approved the final manuscript.

**Funding:** This research was funded by the National Natural Science Fund of China, grant number 41330636, and the Graduate Innovation Fund of Jilin University, grant number 2017137.

**Acknowledgments:** Thanks to anonymous reviewers for their valuable feedback on the manuscript.

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

## References

1. Aleotti, P.; Chowdhury, R. Landslide hazard assessment: Summary review and new perspectives. *Bull. Eng. Geol. Environ.* **1999**, *58*, 21–44. [[CrossRef](#)]
2. Saha, A.K.; Gupta, R.P.; Sarkar, I.; Arora, M.K.; Csaplovics, E. An approach for GIS-based statistical landslide susceptibility zonation—With a case study in the himalayas. *Landslides* **2005**, *2*, 61–69. [[CrossRef](#)]
3. Fell, R.; Corominas, J.; Bonnard, C.; Cascini, L.; Leroi, E.; Savage, W.Z. Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng. Geol.* **2008**, *102*, 85–98. [[CrossRef](#)]

4. Guzzetti, F.; Reichenbach, P.; Ardizzone, F.; Cardinali, M.; Galli, M. Estimating the quality of landslide susceptibility models. *Geomorphology* **2006**, *81*, 166–184. [[CrossRef](#)]
5. Raja, N.B.; Çiçek, I.; Türkoğlu, N.; Aydın, O.; Kawasaki, A. Correction to: Landslide susceptibility mapping of the sera river basin using logistic regression model. *Nat. Hazards* **2018**, *91*, 1423–1423. [[CrossRef](#)]
6. Brabb, E.E.; Pampeyan, E.H.; Bonilla, M.G. Landslide Susceptibility in San Mateo County, California. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 1972.
7. Corominas, J.; Westen, C.V.; Frattini, P.; Cascini, L.; Malet, J.P.; Fotopoulou, S. Recommendations for the quantitative analysis of landslide risk. *Bull. Eng. Geol. Environ.* **2014**, *73*, 209–263. [[CrossRef](#)]
8. Bai, S.B.; Wang, J.; Lü, G.N.; Zhou, P.G.; Hou, S.S.; Xu, S.N. Gis-based logistic regression for landslide susceptibility mapping of the zhongxian segment in the three gorges area, China. *Geomorphology* **2010**, *115*, 23–31. [[CrossRef](#)]
9. Yesilnacar, E.; Topal, T. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Eng. Geol.* **2005**, *79*, 251–266. [[CrossRef](#)]
10. Yılmaz, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey). *Comput. Geosci.* **2009**, *35*, 1125–1138. [[CrossRef](#)]
11. Chen, W.; Pourghasemi, H.R.; Zhao, Z. A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. *Geocarto Int.* **2017**, *32*, 367–385. [[CrossRef](#)]
12. Cao, C.; Xu, P.; Wang, Y.; Chen, J.; Zheng, L.; Niu, C. Flash flood hazard susceptibility mapping using frequency ratio and statistical index methods in coalmine subsidence areas. *Sustainability* **2016**, *8*, 948. [[CrossRef](#)]
13. Chen, W.; Li, W.; Chai, H.; Hou, E.; Li, X.; Ding, X. GIS-based landslide susceptibility mapping using analytical hierarchy process (AHP) and certainty factor (CF) models for the Baozhong region of Baoji city, China. *Environ. Earth Sci.* **2016**, *75*, 1–14. [[CrossRef](#)]
14. He, S.; Pan, P.; Dai, L.; Wang, H.; Liu, J. Application of kernel-based fisher discriminant analysis to map landslide susceptibility in the Qinggan river delta, Three Gorges, China. *Geomorphology* **2012**, *171*–172, 30–41. [[CrossRef](#)]
15. Pourghasemi, H.R.; Rossi, M. Landslide susceptibility modeling in a landslide prone area in Mazandarn Province, north of Iran: A comparison between GLM, GAM, MARS, and M-AHP methods. *Theor. Appl. Climatol.* **2017**, *130*, 1–25. [[CrossRef](#)]
16. Lombardo, L.; Bachofer, F.; Cama, M.; Märker, M.; Rotigliano, E. Exploiting maximum entropy method and aster data for assessing debris flow and debris slide susceptibility for the Giampilieri catchment (north-eastern Sicily, Italy). *Earth Surf. Process. Landf.* **2016**, *41*, 1776–1789. [[CrossRef](#)]
17. Pham, B.T.; Bui, D.T.; Pourghasemi, H.R.; Indra, P.; Dholakia, M.B. Landslide susceptibility assessment in the Uttarakhand area (India) using GIS: A comparison study of prediction capability of naive bayes, multilayer perceptron neural networks, and functional trees methods. *Theor. Appl. Climatol.* **2015**, *122*, 1–19. [[CrossRef](#)]
18. Chen, W.; Pourghasemi, H.R.; Kornejady, A.; Zhang, N. Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geocarto Int.* **2017**, *305*, 314–327. [[CrossRef](#)]
19. Lee, S.; Ryu, J.H.; Won, J.S.; Park, H.J. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. *Eng. Geol.* **2004**, *71*, 289–302. [[CrossRef](#)]
20. Wang, E.; Burchfiel, B.C. Late Cenozoic to Holocene deformation in southwestern Sichuan and Adjacent Yunnan, China, and its role in formation of the southeastern part of the Tibetan Plateau. *Geol. Soc. Am. Bull.* **2000**, *112*, 413–423. [[CrossRef](#)]
21. Can, T.; Nefeslioglu, H.A.; Gokceoglu, C.; Sonmez, H.; Duman, T.Y. Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. *Geomorphology* **2005**, *72*, 250–227. [[CrossRef](#)]
22. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.T. Landslide inventory maps: New tools for an old problem. *Earth-Sci. Rev.* **2012**, *112*, 42–66. [[CrossRef](#)]
23. Yang, X.; Chen, L. Using multi-temporal remote sensor imagery to detect earthquake-triggered landslides. *Int. J. Appl. Earth Obs.* **2010**, *12*, 487–495. [[CrossRef](#)]



24. Cao, C.; Wang, Q.; Chen, J.; Ruan, Y.; Zheng, L.; Song, S.; Niu, C. Landslide susceptibility mapping in vertical distribution law of precipitation area: Case of the Xulong hydropower station reservoir, Southwestern China. *Water* **2016**, *8*, 270. [\[CrossRef\]](#)
25. Wang, F.; Xu, P.; Wang, C.; Wang, N.; Jiang, N. Application of a GIS-based slope unit method for landslide susceptibility mapping along the Longzi river, southeastern Tibetan plateau, China. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 172. [\[CrossRef\]](#)
26. Li, J.; Wang, C.; Wang, G.; Liu, W. Analysis of landslide influential factors and coupling intensity based on third theory of quantification. *Chin. J. Rock Mech. Eng.* **2010**, *29*, 1206–1213.
27. Li, J.-X.; Wang, C.M.; Wang, G.C. Landslide risk assessment based on combination weighting-unascertained measure theory. *Rock Soil Mech.* **2013**, *34*, 468–474.
28. Marjanović, M.; Kovačević, M.; Bajat, B.; Voženílek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **2011**, *123*, 225–234. [\[CrossRef\]](#)
29. Kanungo, D.P.; Arora, M.K.; Sarkar, S.; Gupta, R.P. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Eng. Geol.* **2006**, *85*, 347–366. [\[CrossRef\]](#)
30. Yalcin, A. GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations. *Catena* **2008**, *72*, 1–12. [\[CrossRef\]](#)
31. Akgun, A.; Sezer, E.A.; Nefeslioglu, H.A.; Gokceoglu, C.; Pradhan, B. An easy-to-use Matlab program (Mamland) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Comput. Geosci.* **2012**, *38*, 23–34. [\[CrossRef\]](#)
32. Mohammady, M.; Pourghasemi, H.R.; Pradhan, B. Landslide susceptibility mapping at Golestan province, Iran: A comparison between frequency ratio, dempster–shafer, and weights-of-evidence models. *J. Asian Earth Sci.* **2012**, *61*, 221–236. [\[CrossRef\]](#)
33. Lee, S.; Min, K. Statistical analysis of landslide susceptibility at Yongin, Korea. *Environ. Geol.* **2001**, *40*, 1095–1113. [\[CrossRef\]](#)
34. Simons, M. *The Morphological Analysis of Landforms: A New Review of the Work of Walther Penck (1888–1923)*; JSTOR: New York, NY, USA, 1962.
35. Conforti, M.; Pascale, S.; Robustelli, G.; Sdao, F. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo river catchment (Northern Calabria, Italy). *Catena* **2014**, *113*, 236–250. [\[CrossRef\]](#)
36. Hungr, O.; Leroueil, S.; Picarelli, L. The varnes classification of landslide types, an update. *Landslides* **2014**, *11*, 167–194. [\[CrossRef\]](#)
37. Ozdemir, A. Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). *J. Hydrol.* **2011**, *405*, 123–136. [\[CrossRef\]](#)
38. Gokceoglu, C.; Sonmez, H.; Nefeslioglu, H.A.; Duman, T.Y.; Can, T. The 17 March 2005 Kuzulu Landslide (Sivas, Turkey) and landslide-susceptibility map of its near vicinity. *Eng. Geol.* **2005**, *81*, 65–83. [\[CrossRef\]](#)
39. Beven, K.; Kirkby, M.J. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **1979**, *24*, 43–69. [\[CrossRef\]](#)
40. Moore, I.D.; Burch, G.J. Physical Basis of the Length-slope Factor in the Universal Soil Loss Equation. *Soil Sci. Soc. Am. J.* **1986**, *50*, 1294–1298. [\[CrossRef\]](#)
41. Oh, H.J.; Pradhan, B. Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. *Comput. Geosci.* **2011**, *37*, 1264–1276. [\[CrossRef\]](#)
42. Jebur, M.N.; Pradhan, B.; Tehrany, M.S. Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (lidar) data at catchment scale. *Remote Sens. Environ.* **2014**, *152*, 150–165. [\[CrossRef\]](#)
43. Moore, I.D.; Grayson, R.B.; Ladson, A.R. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Process.* **1991**, *5*, 3–30. [\[CrossRef\]](#)
44. Pradhan, A.M.S.; Kang, H.S.; Lee, S.; Kim, Y.T. Spatial model integration for shallow landslide susceptibility and its runout using a GIS-based approach in Yongin, Korea. *Geocarto Int.* **2016**, *32*, 420–441. [\[CrossRef\]](#)
45. Cheng, Q.; Ko, C.; Yuan, Y.; Ge, Y.; Zhang, S. GIS modeling for predicting river runoff volume in ungauged drainages in the greater Toronto area, Canada. *Comput. Geosci.* **2006**, *32*, 1108–1119. [\[CrossRef\]](#)

46. Huang, Y.; Chen, S.; Cao, Q.; Hong, Y.; Wu, B.; Huang, M.; Qiao, L.; Zhang, Z.; Li, Z.; Li, W.; et al. Evaluation of version-7 TRMM multi-satellite precipitation analysis product during the Beijing extreme heavy rainfall event of 21 July 2012. *Water* **2013**, *6*, 32–44. [\[CrossRef\]](#)
47. Park, S.; Choi, C.; Kim, B.; Kim, J. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environ. Earth Sci.* **2013**, *68*, 1443–1464. [\[CrossRef\]](#)
48. Hong, H.; Pradhan, B.; Xu, C.; Bui, D.T. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena* **2015**, *133*, 266–281. [\[CrossRef\]](#)
49. Laxton, J. Geographic information systems for geoscientists—Modelling with GIS—Bonhamcarter, G.F. *Int. J. Geogr. Inf. Syst.* **1996**, *10*, 355–356. [\[CrossRef\]](#)
50. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*, 3rd ed.; John Wiley: Hoboken, NJ, USA, 2013.
51. Djeddaoui, F.; Chadli, M.; Gloaguen, R.; Djeddaoui, F.; Chadli, M.; Gloaguen, R. Desertification susceptibility mapping using logistic regression analysis in the Djelfa area, Algeria. *Remote Sens.* **2017**, *9*, 1031. [\[CrossRef\]](#)
52. Agterberg, F.P.; Cheng, Q. Conditional independence test for weights-of-evidence modeling. *Nat. Resour. Res.* **2002**, *11*, 249–255. [\[CrossRef\]](#)
53. Preisendorfer, R.W.; Mobley, C.D. Principal component analysis in meteorology and oceanography. *Dev. Atmos. Sci.* **1988**, *17*, 55–72.
54. Dai, F.C.; Lee, C.F. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* **2002**, *42*, 213–228. [\[CrossRef\]](#)
55. Bewick, V.; Cheek, L.; Ball, J. Statistics review 14: Logistic regression. *Crit. Care* **2005**, *9*, 112–118. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Regmi, N.R.; Giardino, J.R.; McDonald, E.V.; Vitek, J.D. A comparison of logistic regression-based models of susceptibility to landslides in Western Colorado, USA. *Landslides* **2014**, *11*, 247–262. [\[CrossRef\]](#)
57. Clark, W.; Hosking, P. *Statistical Methods for Geographers*; John Wiley & Sons: New York, NY, USA, 1986.
58. Mathew, J.; Jha, V.K.; Rawat, G.S. Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. *Landslides* **2009**, *6*, 17–26. [\[CrossRef\]](#)
59. Othman, A.A.; Gloaguen, R.; Andreani, L.; Rahnama, M. Landslide susceptibility mapping in Mawat area, Kurdistan Region, NE Iraq: A comparison of different statistical models. *Nat. Hazards Earth Syst. Sci. Discuss.* **2015**, *3*, 1789–1833. [\[CrossRef\]](#)
60. Devkota, K.C.; Regmi, A.D.; Pourghasemi, H.R.; Yoshida, K.; Pradhan, B.; Ryu, I.C.; Dhital, M.R.; Althuwaynee, O.F. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Nat. Hazards* **2013**, *65*, 135–165. [\[CrossRef\]](#)
61. Lee, S. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int. J. Remote Sens.* **2005**, *26*, 1477–1491. [\[CrossRef\]](#)
62. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. Model-building strategies and methods for logistic regression. In *Applied Logistic Regression*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2000; pp. 89–151.
63. Alatorre, L.C.; Sánchez-Andrés, R.; Cirujano, S.; Beguería, S.; Sánchez-Carrillo, S. Identification of mangrove areas by remote sensing: The ROC curve technique applied to the northwestern Mexico coastal zone using Landsat imagery. *Remote Sens.* **2011**, *3*, 1568–1583. [\[CrossRef\]](#)
64. Chen, Y.; Booth, D.C. *The Wenchuan Earthquake of 2008: Anatomy of a Disaster*; Springer Science & Business Media: New York, NY, USA, 2011.
65. Hao, M.; Wang, Q.; Shen, Z.; Cui, D.; Ji, L.; Li, Y.; Qin, S. Present day crustal vertical movement inferred from precise leveling data in eastern margin of Tibetan plateau. *Tectonophysics* **2014**, *632*, 281–292. [\[CrossRef\]](#)

66. Burbank, D.W.; Leland, J.; Fielding, E.; Anderson, R.S.; Brozovic, N.; Reid, M.R.; Duncan, C. Bedrock incision, rock uplift and threshold hillslopes in the northwestern Himalayas. *Nature* **1996**, *379*, 505–510. [[CrossRef](#)]
67. Pradhan, A.M.S.; Kim, Y.T. Relative effect method of landslide susceptibility zonation in weathered granite soil: A case study in Deokjeok-ri Creek, South Korea. *Nat. Hazards* **2014**, *72*, 1189–1217. [[CrossRef](#)]



© 2018 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 (<http://creativecommons.org/licenses/by/4.0/>).