Article

Landslide Susceptibility Mapping Based on Selected Optimal Combination of Landslide Predisposing Factors in a Large Catchment

Qianqian Wang [1], Dongchuan Wang [1],*, Yong Huang [2], Zhiheng Wang [1], Lihui Zhang [1], Qiaozhen Guo [1], Wei Chen [1], Wengang Chen [1] and Mengqin Sang [1]

[1] School of Geology and Geomatics, Tianjin Chengjian University, No. 26 Jinjing RD., Xiqing District, Tianjin 300384, China; wq1365216@163.com (Q.W.); Wangzhiheng@tcu.edu.cn (Z.W.); zhanglihui@tjuci.edu.cn (L.Z.); gqiaozhen@tcu.edu.cn (Q.G.); towei@163.com (W.C.); cwg844321258@163.om (W.C.); jessicasang@sina.com (M.S.)

[2] Appraisal Center for Environment & Engineering Ministry of Environmental Protection, No. 8 Beiyuan RD., Chaoyang District, Beijing 100012, China; jiajia8211@126.com

* Correspondence: wangdongchuan@tcu.edu.cn; Tel.: +86-22-2377-3032

Abstract: Landslides are usually initiated under complex geological conditions. It is of great significance to find out the optimal combination of predisposing factors and create an accurate landslide susceptibility map based on them. In this paper, the Information Value Model was modified to make the Modified Information Value (MIV) Model, and together with GIS (Geographical Information System) and AUC (Area Under Receiver Operating Characteristic Curve) test, 32 factor combinations were evaluated separately, and factor combination group with members Slope, Lithology, Drainage network, Annual precipitation, Faults, Road and Vegetation was selected as the optimal combination group with an accuracy of 95.0%. Based on this group, a landslide susceptibility zonation map was drawn, where the study area was reclassified into five classes, presenting an accurate description of different levels of landslide susceptibility, with 79.41% and 13.67% of the validating field survey landslides falling in the Very High and High zones, respectively, mainly distributed in the south and southeast of the catchment. It showed that MIV model can tackle the problem of "no data in subclass" well, generate the true information value and show real running trend, which performs well in showing the relationship between predisposing factors and landslide occurrence and can be used for preliminary landslide susceptibility assessment in the study area.

Keywords: predisposing factor combination; Modified Information Value model; ROC curve; Baoxing catchment; China

1. Introduction

Landslides are often triggered under the interaction of many factors. In sensitive regions of geological environment, the original structure of rock was damaged by frequent earthquakes, and loose debris mass gathered there. Geological environment instability is further aggravated in these areas, due to a combination of complicated geological structure, steep terrain, erosion of frequent heavy rain and human activities [1–3]. Owing to the diversity and uncertainty of landslide predisposing factors and high increase in incidence frequency, the potential threat of landslides is increased, which not only endangers the safety of life and property for the local population, but also results in huge damage to regional resources and environment [4,5].

Many scholars in the field of landslides [6,7] found that research on predisposing factors for sensitive area is extremely important. The reason is that dynamic mechanism of landslides and landslide predisposing factors have uncertain characteristics, such as the non-uniqueness of factors and the uncertainty of the master–slave relationship, and they may change with seasons and some triggering events (such as the concentrated precipitation of monsoon climate area) [8]. In the sensitive regions of the geological environment, landslide predisposing factors, such as topography and geomorphology, geological structures, rock properties and hydrological conditions, interact with each other, which made it more difficult to determine the time and scale of landslides [9]. Many studies have been done on predisposing factors for evaluation. Dahl *et al.* [1] carried out research in the Faroe islands in the North Atlantic and found that landslide initiation area was determined by slope, lithology, soil coverage, and the primary area mainly has slopes from 25° to 40°. Scholars found that factors like topography and geomorphology, geological conditions, cutting intensity, vegetation environment, and human intervention are all indispensable in the study of geological predisposing factors, and the inner relationship between landslides and the predisposing factors should be further studied [10–12]. Chinese scholars have done lots of forecasting and evaluating work on landslide and debris flow with the GIS technology and sensitivity index model in southwest China. Lan *et al.* [13] considered precipitation a significant landslide predisposing factor. Landslide trigger threshold by earthquake will be significantly reduced after precipitation. From the perspective of terrain factors, based on remote sensing image interpretation, Zhang *et al.* [14] focused on the relationship between the characteristics of landslide and terrain factors through three factors (*i.e.*, slope, relief, and gully vertical drop) closely related to landslides. To summarize, landslides are usually initiated under complex conditions, in which some important factors should not be ignored, such as the terrain conditions, lithologic distribution, precipitation, seismic activity, *etc.* [15]. The triggering mechanism is from a combined action of geological structure and environment. Different factors should be taken into comprehensive consideration, including topography and geomorphology, geological conditions, cutting intensity, vegetation environment, human factors and so on, excluding those factors which exert little effect. Therefore, in practical analysis and evaluation of landslides, it is of great significance to find out the most influential combination of factors to landslides in specific area.

It is extremely important to objectively analyze the distribution of every factor in the study area and the relationship between predisposing factors and landslide bodies before the geological hazard assessment and prediction [16]. Different methods have been taken in the evaluation of landslide susceptibility. Kayastha [17] summarized the research methods of sensitive factors as five types: (1) direct map plotting method; (2) landslide catalog; (3) the exploratory analysis; (4) statistical methods such as fuzzy logic and artificial neural network; and (5) the conceptual model. Among them, the direct map cartography and landslides catalog methods are the most intuitive and basic ways to identify regional disasters for research. Bathrellos [18] proposed a unique approach of using mainly natural hazards as well as geological–geomorphological–geographical characteristics of the study area for urban planning and sustainable development. As for other research methods, Kritikos and Davies [19] developed an approach based on GIS (Geographic Information System) to make susceptibility assessment of precipitation-induced shallow landslide. They applied fuzzy logic technique to deal with uncertainties and intricate relationships between conditioning factors and landslide.

Information value model is a statistical analysis method which was developed from information theory and is now often applied to spatial prediction of geological hazards and disaster risk assessment [20–23]. However, some problems are hidden in this IV model. When no landslide exists in certain subclass, there would be no significance. Researchers usually assign "0" [20] or "no data" [22] to those pixels, which would make the results much more exaggerated if a large number of those pixels existed, since "0" value in the model means that the ratio of landslide pixels in subclass i is equal to the average ratio of the study area, and if no landslides existed, the results should approach infinitesimal. To overcome this problem, Oliveira *et al.* [24] defined that when Npix (x_i) = 0, I (H, x_i)

was not calculated and was qualitatively determined as the lower Information value considering the data set of predisposing variables, which could avoid the problem of high exaggeration, while the results could not exactly show the information value of this area.

Baoxing Catchment has suffered frequent geological disasters in the past decade. Landsides have caused huge life and property losses and posed great threats to post-disaster reconstruction. Taking it as the study area, based on ARCGIS and SPSS software, this paper aims to find out the most influential factor combination to landslides in the study area and make susceptibility map based on it. Modified Information Value (MIV) Model was established, and together with AUC (Area Under Receiver Operating Characteristic Curve) test, was taken in the selection of the optimal combination group.

2. Study Area

In this paper, Baoxing Catchment in Sichuan province, China was taken as the study area, including Baoxing county and Lushan county. Baoxing catchment is located at $102°26'$–$103°14'$ E, $30°02'$–$30°57'$ N, in the western part of Sichuan basin, with a total area of 4319 km^2 (Figure 1). Mountainous terrain dominates this catchment, and the elevation is gradually reduced from 5268 m in the northwest to 557 m in the southeast. It belongs to the joints of the Yangtze Paraplatform and Ganzisongpan Geosynclinal fold system, across three seismic belts, including the Longmenshan Mountain Belt, Xianshuihe River Belt, and Anning River Belt. Intensive fault structures lie in Sichuan Province, China, and frequent seismic activities happened in the past decade. Wenchuan earthquake (12 May 2008) and Lushan earthquake (20 April 2013) struck this area successively. Both earthquakes caused huge losses of life and property losses and posed great threats to post-disaster reconstruction [25]. The water resources, ecology, communication, electric power and road system suffered the influence of different levels [26]. The rock structure loosens and the shear strength of geological structural reduces significantly when several earthquakes occur in the same area [27]. The study area is located in the monsoon climate area, where the mean annual precipitation reaches 1000 mm, with most concentrated in heavy precipitations.
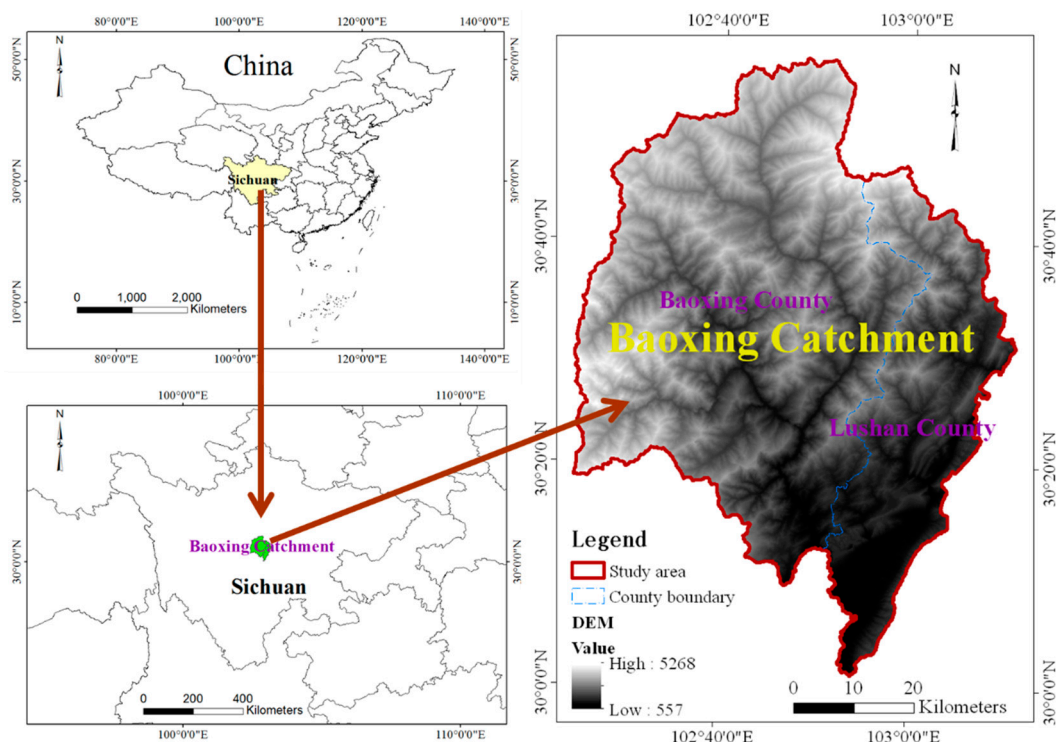

Figure 1. Location of the study area.

3. Data Source and Preprocessing

Remote sensing (RS) image data Landsat TM and ETM+ data were downloaded from the data sharing platform Geospatial Data Cloud (http://www.gscloud.cn/). The atmospheric correction of RS image was processed by FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) model and geometric correction was done according to the 1:250,000 topographic maps and GPS field sampling points. Through data processing and integration, multi-spectral color images were generated with the spatial resolution of 15 m. Drainage network, roadside, landslides and vegetation coverage information were interpreted by the combination of object-oriented classification method and manual vectorization. Furthermore, high spatial resolution remote sensing images from Google Earth were referenced to get a higher accuracy result. Altogether, 1258 landslides were extracted here (Figure 2), 97% of which were small- and medium-sized shallow slides or collapses, mainly triggered by Wenchuan earthquake (12 May 2008) and Lushan earthquake (20 April 2013). All the landslides together affected a total area of 3,529,530 m^2, which corresponds to 0.08% of the study area. The landslide size ranges from 8.9 to 173,459 m^2 and the mean value is 2790 m^2. The landslide density is 0.29 landslide/km^2.
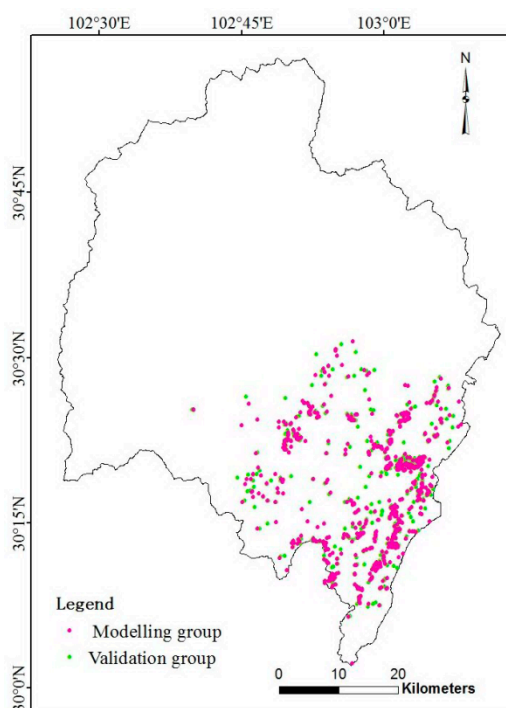
Figure 2. Spatial distribution of landslides.

The geological information used in this article was derived from 1:250,000 geological map of the Geological Survey Institute of Sichuan province. Seismic data were based on information from the national earthquake science data sharing center (http://data.earthquake.cn). The precipitation data were collected from meteorological science data sharing service network of Sichuan province (http://www.climate.sc.cn). Digital Elevation Data (ASTER GDEM 30 m) of Baoxing catchment were used to delineate drainage network, slope and aspect by ARCGIS software. The field survey landslide distribution map was provided by the Geological Survey Institute of Sichuan province. Half of those landslides were used in the calculation of Information Value, and the left half were used to evaluate the accuracy of the prediction in AUC calculation.

4. Methodology

4.1. Selection of Predisposing Factors

There are no universal guidelines regarding the selection of factors for landslide susceptibility [28]. A certain parameter may be an important controlling factor for landslide occurrence in one area but not in another [20]. According to the analysis of previous data and geological environment research literature on the sensitive area of landslides, the critical predisposing factors of landslides include regional geological conditions (fault structure, slope, aspect, formation lithology, *etc.*), climatic factors (precipitation distribution, temperature, *etc.*), geological disaster factors (density, frequency, types, *etc.*), vegetation conditions, river terrain cutting, and human engineering activities (roads and other engineering constructions) [29]. In consideration of the complex conditions in the study area, together with data accessibility, we choose nine factors from four characteristic groups as landslide predisposing factors, covering most of the critical elements leading to landslides in this region. That is, Topography group (including slope, aspect and terrain), Geological lithology group (including lithology and faults), Terrain cutting group (including river and road cutting intensity) and Natural environment group (including vegetation coverage and annual precipitation distribution).

Each predisposing factors was reclassified into several classes (Table 1), which were utilized as different zonal statistic areas in the Modified Information Model to evaluate landslide susceptibility.

Table 1. The reclassification of different predisposing factors.

Groups	Factors	Subclasses	Area (hm²)	Factors	Subclasses	Area (hm²)
Topography	Slope	0–5°	7680.06	Aspect	<0° (Flat)	34,934.22
		5–10°	16,243.80		0–22.5 (N1)	39,252.06
		10–15°	28,682.50		22.5–67.5° (NE)	42,176.25
		15–20°	44,728.50		67.5–112.5° (E)	52,289.10
		20–25°	60,272.50		112.5–157.5° (SE)	50,455.53
		25–30°	70,456.10		157.5–202.5° (S)	41,343.39
		30–35°	71,933.70		202.5–247.5° (SW)	39,180.60
		35–40°	58,948.10		247.5–292.5° (W)	41,553.45
		40–45°	37,528.60		292.5–337.5° (NW)	46,666.89
		45–50°	19,707.60		337.5–360° (N2)	42,801.39
		50–55°	8795.88	Terrain (relief amplitude)	<100 m	235,095.12
		55–60°	3632.94		100–200 m	189,105.30
		>60°	2042.82		>200 m	9745.65
Geological lithology	Lithology	Hard	113,977.35	Faults (buffer distance)	0–3000 m	262,886.31
		Middle hard	277,983.36		3000–6000 m	73,262.25
		Middle soft	36,291.87		6000–9000 m	23,076.63
		Soft	2397.60		9000–12000 m	17,117.37
					12000–15000 m	15,132.96
					>15000 m	39,119.94
Terrain cutting	Road (buffer distance)	>1000 m	59,003.19	Drainage network (buffer distance)	<200 m	58,350.06
		1000–2000 m	48,580.38		200–400 m	53,166.96
		2000–3000 m	43,454.52		400–600 m	51,164.73
		3000–4000 m	70,729.56		600–800 m	46,314.72
		>4000 m	208,882.44		800–1000 m	44,561.34
					>1000 m	177,095.07
Natural environment	Precipitation	<800 mm	33,498.17	Vegetation (coverage)	<0.2	275.31
		800–900 mm	131,643.14		0.2–0.3	25,049.16
		900–1000 mm	121,177.49		0.3–0.4	80,806.50
		1000–1200 mm	84,867.92		0.4–0.5	189,018.27
		>1200 mm	59,463.28		0.5–0.6	122,846.40
					>0.6	12,652.02

4.2. Modified Information Value (MIV) Model

4.2.1. Information Value (IV) Model

The information value (IV) for each subclass of the factors is calculated with the following equation [20,30],

$$I(H, x_i) = \ln \frac{Npix(x_i) / Npix(N_i)}{\sum Npix(x_i) / \sum Npix(N_i)} \tag{1}$$

In the study area, $I(H, x_i)$ is the information value of subclass i of a predisposing factor; $Npix(x_i)$ is the number of landslide pixels in subclass i; $Npix(N_i)$ is the total pixel number of subclass i; $\sum Npix(x_i)$ is the total landslide pixel number of in the study area; and $\sum Npix(N_i)$ is the total pixel number of the study area.

Therefore, the total information value I_{total} for each pixel is calculated by summing up all the information values of each factor layer using Equation (2).

$$I_{total} = \sum_{i=1}^{n} I(H, x_i) \tag{2}$$

where n is the number of layers of the predisposing factors.

4.2.2. The Establishment of the Modified Information Value (MIV) Model

To avoid the problem ("no data" for a certain subclass) mentioned in the introduction, and to quantitatively express the results, Equation (1) in IV model was modified in this paper to establish the MIV model as the following equation,

$$I(H, x_i) = \log_2 \left(\frac{Npix(x_i) / Npix(N_i)}{\sum Npix(x_i) / \sum Npix(N_i)} + 1 \right) \tag{3}$$

where, when no landslides existed in a certain subclass, $I(H, x_i)$ equals "0", which would be the smallest value; when $I(H, x_i)$ is "1", it means that the ratio of landslide pixels in subclass i equals to the average of the study area; and when the value is larger than "1", more landslides pixels in the subclass lie than those in the study area, and the larger the value, the higher landslides ratio in the subclass, *vice versa*. Taking factor Road (buffering distance) as an example, statistic maps of the two models are shown in Figure 3. When the distance was larger than 4000 m, no landslide was found in field survey, therefore, different fitting trends appear at the last subclass. Furthermore, obviously MIV model makes the true information value and the real running trend.
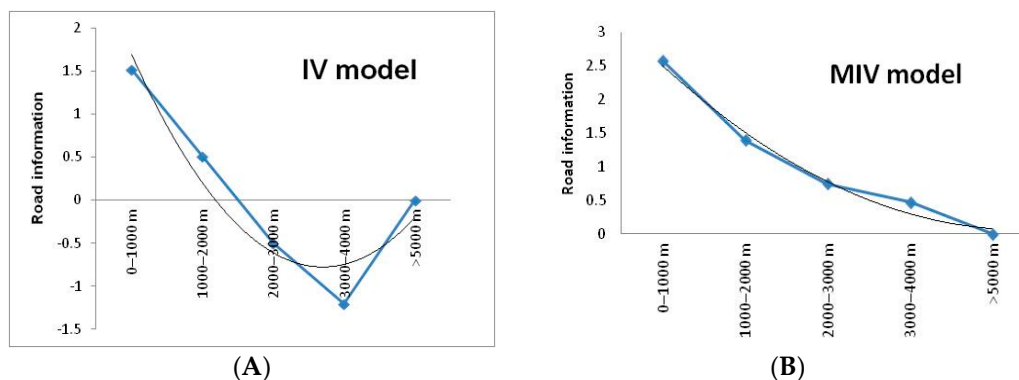
Figure 3. Result comparison of the two models. (**A**) IV model; (**B**) MIV model.

I_{total} in MIV model is still calculated with Equation (2). Landslides were affected by many factors, and each factor's contribution was different. Thus, the weight value for each factor cannot be judged

simply by artificial decision. Half of the field survey landslides were used to calculate the information value of each subclass of every predisposing factor with Equation (3), which indicates the degree of contribution of the subclass to landslides. Summing up all the information layers of all factors, layer of I_{total} was achieved to show the spatial distribution of I_{total} values, where the larger the value is, the higher the landslide susceptibility.

4.3. The Selection of Optimal Factor Combination

4.3.1. Factor Combination

Due to differences of geological environments in different regions, the inducing factors of landslides would not be completely consistent and the contribution of every factor would be different, so there would be an optimal factor combination, under the function of which spatial probability of landslide occurrence would be the highest. One of the aims of this paper is to find the optimal factor combination in the study area.

Enumeration method was used in this paper to make different combinations of the nine predisposing factors. Four typical predisposing factors, each from a different group, are selected through the following analysis as the basic factors and 32 test factor combinations are set up by adding 1, 2, 3, 4, or 5 of the left factors each time. For each combination, raster layer of the total information value on every pixel was generated in ARCGIS with MIV model.

4.3.2. Selection of Optimal Combination Based on ROC Curve Test

The receiver operating characteristic curve, referred to as "ROC curve", is an effective method of evaluating the performance of dichotomy problems, which divide the objectives into two classes, positive and negative, like in diagnostic tests [31]. An ROC curve is constructed using a true positive rate and false positive rate pair for each possible threshold value of the test. Each point on the curve is created by plotting the unique true positive rate (TPR) and false positive rate (FPR) associated with each unique test value. The area under the ROC curve (AUC) is a common metric that can be used to compare different tests (indicator variables). An AUC close to 0.5 corresponds to a poor diagnostic test. The larger the AUC, the more accurate the test is. Landslide susceptibility maps are often validated using AUC [32,33]. This method has been widely used as a measure of performance of a predictive rule [20,34,35]. With the implementation of ROC analysis we can assess the prediction accuracy of a model [36].

In this paper, together with enumeration method, AUC of ROC curve was applied to evaluate the 32 factor combinations separately in SPSS software. The combination with the maximum AUC value (and larger than 0.9) would be chosen as the optimal combination of predisposing factors to the landslides in the area. The value also indicates the combination's "diagnosis" ability for landslides, for higher score indicates the higher consistency of landslide sustainability with the actual spatial distribution of landslides.

4.4. Landslide Sustainability Zoning Based on the Optimal Combination

According to the former analysis, the larger the I_{total} value is, the higher the landslide susceptibility. Using the natural break method in ARCGIS, layer of the total information values (I_{total}) for the optimal combination was reclassified into five classes (very high, high, moderate, low, and very low) to define different levels of landslide susceptibility in the study area. Furthermore, then layer of field survey landslides was overlaid with the landslide susceptibility layer to show the accuracy of sustainability assessment.

5. Results

5.1. Spatial Distribution Characteristics of Information Value

Information Values were calculated in ARCGIS for every landslide impact factor and layers of IV for different factors were acquired (Figure 4) and IV for each subclass for those factors is shown in Figure 5. Larger IV suggests that correlation degree of factors with landslides is higher, and *vice versa*. Information value is not only a quantitative indicator for the influence of a factor on the landslides, but also the reflection of the landslide distribution in each factor. Based on Figures 4 and 5 spatial distribution analyses of IV were carried on in each predisposing factor.

5.1.1. In Topography Group

Different Topography conditions may lead to different possibilities of landslides. Here, in this paper, Slope, Aspect and Terrain (Relief amplitude) were taken as predisposing factors in Topography group to show the changes of Information Values.

In general, the higher the slope is, the more unfavorable it is to rock and soil consolidation, and the possibility of landslides would be greater [37]. Lower slope IVs are mainly distributed in areas where it is relatively flat, while higher IVs are concentrated in the central and southern mountains. The relationship between IV and slope presented a state of fluctuation, but the overall trend is it rises with the increase of slope. When the slope is lower than 35°, slope IV was relatively low, which showed that the probability of landslide is lower in slopes under 35°, especially when it is in subclass 5°–10°, the contribution to landslide is the lowest degree (with IV 0.45). When the slope is higher than 35°, slope IV increased dramatically with a drop for subclass 45°–50°, and the highest IV (3.22) was found in the subclass >60°.
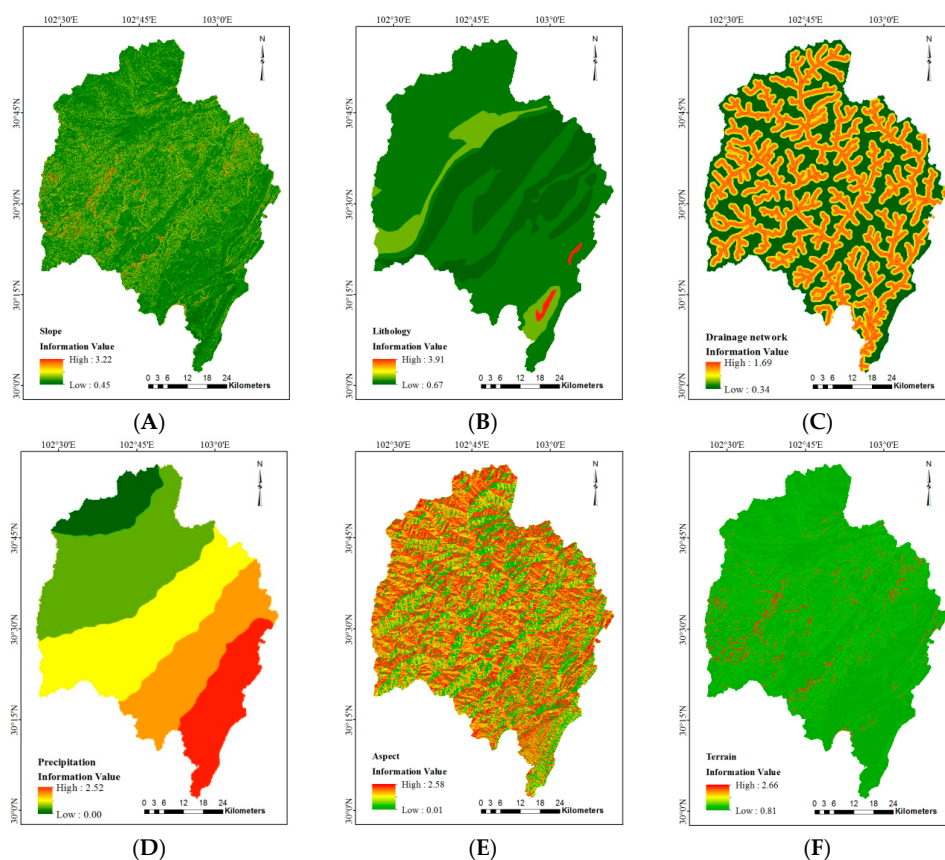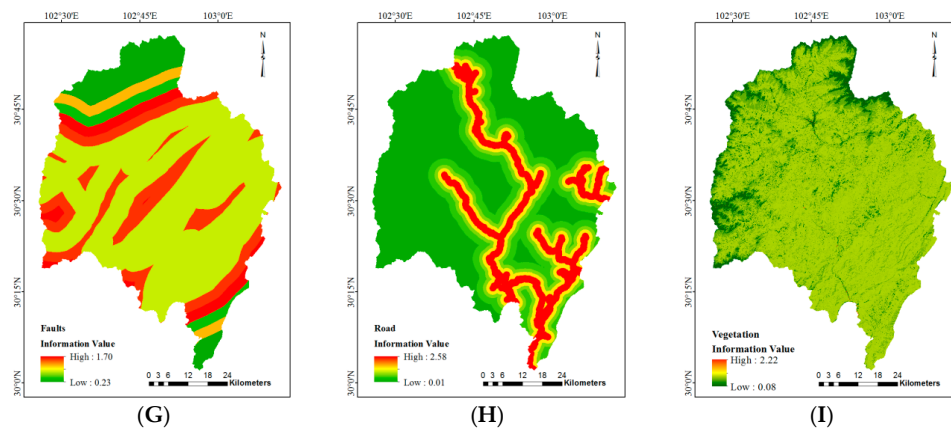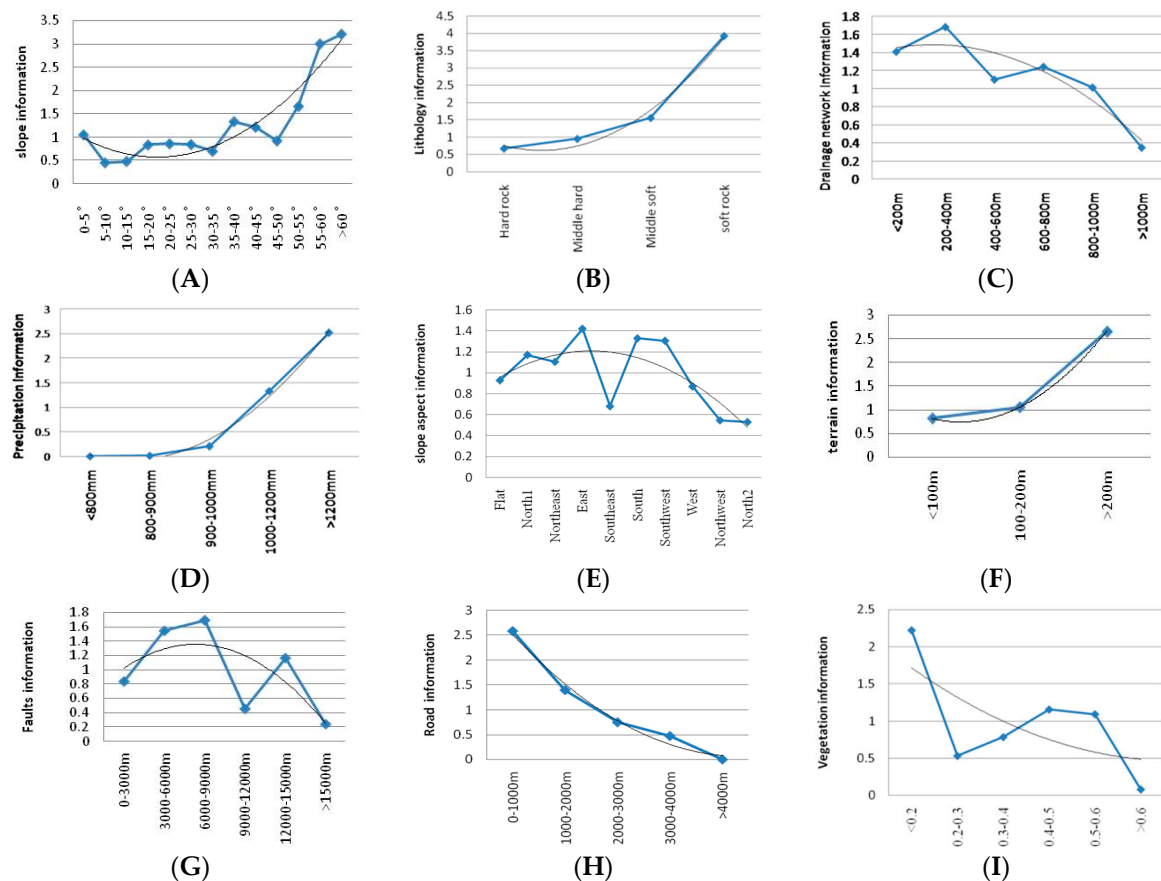

Figure 4. *Cont.*

Figure 4. Layers of Information Values for different factors. (**A**) Slope; (**B**) Lithology; (**C**) Drainage network (buffering distance); (**D**) Annual precipitation; (**E**) Aspect; (**F**) Terrain (relief amplitude); (**G**) Faults (buffering distance); (**H**) Road (buffering distance); (**I**) Vegetation coverage.

Figure 5. Statistics of IV (Information value) for each subclass in those factors. (**A**) Slope; (**B**) Lithology; (**C**) Drainage network (buffering distance); (**D**) Annual precipitation; (**E**) Aspect; (**F**) Terrain (relief amplitude); (**G**) Faults (buffering distance); (**H**) Road (buffering distance); (**I**) Vegetation coverage.

As for layer of Aspect, it was found that the higher IVs mainly distributed in subclass E (East), S (South) and SW (Southwest) and it was lower in SE (Southeast), NW (Northwest) and N (North). Analysis from the remote sensing image interpretation showed that most landslides concentrated in east and south directions. The situation was related to the natural geographical

environment differences. Sunny hillsides receive plenty of sunlight, leading to high evaporation and poor vegetation coverage; therefore, the conditions of water and hot combination is poorer than on shady hillsides, and the degree of rock weathering is relatively more severe, therefore more landslides developed [8]. While on shady hillsides, less moisture evaporation took place, a good combination water and heat conditions benefit the growth of vegetation, which can also prevent soil erosion, then the degree of rock consolidation will be fine, so the possibility of landslide is lower.

Relief amplitude is a sign of gentle relief condition in a region. The greater the Relief amplitude, the more complex the topography is. Overall, the trend line of IVs was increasing in the layer of Terrain. Terrain IV is the lowest in less than 100 m (0.81), then it increased with the fluctuation degree and reached the highest point at the subclass of >200 m (2.66), though the area of this subclass is only 2.24% of the whole study area. Higher IVs were mainly distributed in the central and southern mountains with high relief amplitude, and the area of low IV was located in the northern and southern plains.

5.1.2. In Geological Lithology Group

Geological lithology plays a key role in controlling spatial distribution of regional geological disasters. These areas suffer from many landslides because of the tectonic activity and its geological setting [38]. In this paper, two predisposing factors Lithology and Faults (buffering distance) were considered in Geological lithology group.

Lithology is one of the most basic and important factors in the research of geological hazards, and influences the stability of regional geological structure [39]. Distribution of lithology was complex, hard and soft rocks were staggered in the research area. The fitting curve of Lithology IVs rose from hard rock to soft group in a monotone increasing state with a high fitting degree. The minimum IV 0.67 was in the hard rock subclass, and the peak IV 3.91 appeared in the soft rock subclass. The information values of middle soft and soft rock were both larger than 1, which showed that they exerted greater effect on the development possibility of landslides.

Faults are one important factor of the control component for the stability of the regional geological. Fault activities tend to induce landslide and mud-rock flows in zonal distribution. Fitting curve of fault distance information is presented a trend of rise first, and then fall. High IVs mainly distributed in the range of 3000–9000 m from Faults, and landslide area was the largest in this region. Furthermore, in 6000–9000 m, it reached the highest value, 1.70, and exerted the greatest influence on the landslide. Altogether, there are three low points, in subclasses 0–3000 m, 9000–12,000 m, and >15,000 m. Fault IV reached at a minimum value of 0.23 when the distance was larger than 15,000 m. In the subclass 9000–12,000 m, Fault IV is 0.45, but the information value increased to 1.18 in the next subclass 12,000–15,000 m.

5.1.3. In Terrain Cutting Group

In the terrain cutting group, two predisposing factors were chosen, namely Drainage network and Road, which usually exert strong cutting action on terrain. Multi-ring buffering distance (every 200 m for Drainage network and 1000 m for Road) were used to generate subclasses.

In the natural environment, drainage network would have cutting effect on the surrounding geography, and, at the same time, influence the stability of rock and soil along the coast [40]. In the study area, IVs of drainage network presented circular distribution along the network, and decreased with the distance. The closer distance to the network, the higher the information value is, and the farther, the lower it is. In mountain areas, sections along drainage network are often regions with very frequent geological disasters, which were closely related to drainage network in the history. Landslides often develop around water system [27]. From Figures 4 and 5 it was found that drainage network information fitting curve was in a declining trend with the increase of the distance, and the reduction was relatively uniform and moderate. Within a distance of 400 m, IVs of drainage network were greater than 1, so the effect on landslide was obvious. Furthermore, IV were the highest in

200–400 m distance from drainage network, 1.68. When the distance was larger than 600 m, influence degree on the landslides decreased continuously, and when it was larger than 1000 m, the IV reached the minimum, 0.34.

Road were selected as manmade predisposing factor on landslides. The road information curve had a high fitting degree, monotonically decreasing with the distance of road buffer. The nearer to the road, the higher the Road IV is. Road constructed would produce disturbance on the slope ground, leaving the rock structural plane in an instability state. The maximum IV (2.58) was within 1000 m buffer distance, and the landslide area in this range, accounting for 68.2% of the total landslide area. When the distance was larger than 4000 m, no landslide was found in field survey, the influence degree dropped to the minimize value, 0. In mountainous area, road may artificially cut towards the steep slopes which might cause high landslide occurrence in the near distances from the road [41].

5.1.4. In Natural Environment Group

Annual precipitation distribution and Vegetation coverage are taken as the representations of the group of Natural Environment, which usually play important roles in landslide occurrence [42,43].

Precipitation is a dynamic source of rock sliding and at the same time provides sliding surface for slipmass, therefore, it is one of the essential factors in the monsoon climate region for landslides [44,45]. In the research area, annual precipitation increases gradually from north to south. Precipitation IV fitting curve showed a trend of monotone increasing with the increase of annual precipitation. The lowest IV, approximate to zero, lay in the north end of the study area with annual average precipitation less than 800 mm, while, the peak IV, 2.52, appeared in the southernmost of the study area where annual average precipitation is larger than 1200 mm. Where annual average precipitation is greater than 1000 mm, area and the number of landslides became larger. The monotonic trend of precipitation IV with precipitation further proved that precipitation is the one of the main induce factors of landslides.

For vegetation coverage, The Vegetation IV fitting curve shows a slow downward trend, with an inflexion point at subclass 0.2–0.3. When vegetation coverage is lower than 0.2, Vegetation IV reached the highest value, 2.22. The maximum Vegetation impact on the landslide occurred in this area, with more landslides in less subclass area. Furthermore, landslides increase land instability, and it is likely to slide again if natural reforestation is slow [46]. When the coverage is from 0.4 to 0.6, vegetation IV was slightly larger than 1. In the left three subclasses, vegetation IVs were all lower than 1, especially when the coverage is higher than 0.6, vegetation IV reached the minimum value, 0.08. High vegetation coverage can usually reduce the frequency and severity of geological hazards. Vegetation IV curve was not in a monotone drop trend with the rising of vegetation coverage in this area, which shows vegetation coverage is only the auxiliary elements affecting landslide development in the study area.

5.2. Optimal Combination Selection Based on ROC Curve Test

Among those predisposing factors, there would be an optimal combination to express landslide susceptibility. Based on the statistics of IV in Figure 5, Slope (from Topography group), Lithology (from Geological lithology group), Drainage network (from Terrain cutting group) and Precipitation (from Natural environment group) were selected as the basic combination group, and then 32 combination groups were made by adding 0,1, 2, 3, 4, or 5 of the left factors each time (Table 2).

AUC of ROC curve was applied to evaluate the 32 factor combinations separately in SPSS software. AUC results of all the 32 combination groups are shown in Table 2.

Table 2. AUC (Area Under Receiver Operating Characteristic Curve) values of all the combination groups.

| Group ID	Factor Combination	AUC Value	Group ID	Factor Combination	AUC Value
1	A, B, C, D	0.878	17	A, B, C, D, E, F, G	0.877
2	A, B, C, D, E	0.872	18	A, B, C, D, E, F, H	0.938
3	A, B, C, D, F	0.875	19	A, B, C, D, E, F, I	0.878
4	A, B, C, D, G	0.888	20	A, B, C, D, E, G, H	0.943
5	A, B, C, D, H	0.943	21	A, B, C, D, E, G, I	0.889
6	A, B, C, D, I	0.890	22	A, B, C, D, E, H, I	0.940
7	A, B, C, D, E, F	0.867	23	A, B, C, D, F, G, H	0.946
8	A, B, C, D, E, G	0.881	24	A, B, C, D, F, G, I	0.893
9	A, B, C, D, E, H	0.939	25	A, B, C, D, G, H, I	0.950
10	A, B, C, D, E, I	0.882	26	A, B, C, D, F, H, I	0.943
11	A, B, C, D, F, G	0.883	27	A, B, C, D, E, F, G, H	0.943
12	A, B, C, D, F, H	0.870	28	A, B, C, D, E, F, G, I	0.886
13	A, B, C, D, F, I	0.886	29	A, B, C, D, E, G, H, I	0.944
14	A, B, C, D, G, H	0.947	30	A, B, C, D, E, F, H, I	0.940
15	A, B, C, D, G, I	0.897	31	A, B, C, D, F, G, H, I	0.947
16	A, B, C, D, H, I	0.944	32	A, B, C, D, E, F, G, H, I	0.944

(A) Slope; (B) Lithology; (C) Drainage network (buffering distance); (D) Annual precipitation; (E) Aspect; (F) Terrain (relief amplitude); (G) Faults (buffering distance); (H) Road (buffering distance); (I) Vegetation coverage. Among groups with five factor members, Group 5 had the highest AUC value, 0.943, where Road was added to the basic group, showing that Road was more important than the left four factors in landslide occurrence. When group member reached six, Group 14, with Road and Faults added to the basic group, became the best combination with AUC value, 0.947, which was also higher than that of Group 5. Furthermore, when three factors were added to the basic group, the highest AUC value belonged to Group 25, reaching 0.950, the highest of all 32 groups. Here, Road and Faults were still group members. Vegetation was added to Group 14. When the members were continually added, AUC value of Group 31 was the highest of the left groups, but it was still lower than Group 25. This means that starting from Group 25, when more factors are added, the AUC value decreases. That is to say, more factors do not always mean better results. When Aspect or Terrain or both were added to Group 25, the AUC value is lower. Therefore Group 25, made up of Slope, Lithology, Drainage network, Annual precipitation, Faults, Road and Vegetation, was selected as the optimal combination group to express the susceptibility of landslide occurrence in the study area.

6. Discussions

For every landslide predisposing factor, information values are different between subclasses, among which there is one subclass that contributes the most to landslide development. For all factors, each would function in landslide development, but would not be the only dominating factor. Therefore, landslide is the result of joint action by different kinds of factors.

Theoretically speaking, the group with the highest AUC value is the optimal combination of landslide predisposing factors, while according to the predictive capacity among models, there is no significant differences between similar AUC values, like 0.946, 0.947 and 0.950, with respect to ROC curves. From this point of view, it can be seen that all the groups with AUC value higher than 0.940 are with H (Road) included in the group besides the basic four factors, which shows that H is an important predisposing factor in considering landslide susceptibility; while, when either or both E (Aspect) and F (Terrain) are added to the group, AUC values are slightly lower than before, indicating that E and F could be excluded in the study area. What is more, a group with the same high AUC value but less predisposing factor members would be commonly accepted as a more suitable one for the selection of optimal combination, especially when data acquisition were difficult. Therefore, Group 25 can be defined as the optimal combination in the study area and when either or both G (faults) and I (vegetation coverage) were difficult to obtain, Group 14, Group 16 or Group 5 can suffice.

The whole Baoxing Catchment was reclassified into five classes (very high, high, moderate, low, and very low) to define different levels of landslide susceptibility based on I_{total} values of the optimal combination group (Group 25 in Table 2). The landslide susceptibility map was produced, and was overlaid with the layer of field survey landslides (Figure 6). Statistics results were listed in Table 3 to show the accuracy of susceptibility assessment.
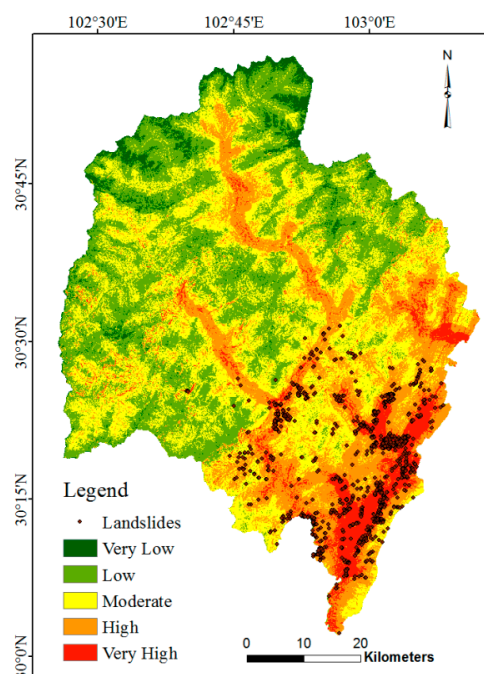
Figure 6. Landslide susceptibility zonation map overlaid with field survey landslides.

Table 3. Statistics fitting results of landslides in each landslide susceptibility zone.

Class ID	Zone	Section Area (hm^2)	Landslide Area (hm^2)	Landslide Density	Landslide Ratio (%)
1	Very Low	11,926.62	0.03	0.01%	0.01
2	Low	156,442.41	0.81	0.02%	0.32
3	Moderate	148,201.92	16.45	0.11%	6.58
4	High	82,849.23	34.21	0.41%	13.67
5	Very High	30,855.06	198.68	2.33%	79.41

According to the AUC values in Table 2, Group 25 owned the Highest AUC, 0.950, which also means that the accuracy of the susceptibility map generated by Group 25 was 95.0%. From Figure 6, it can be visually judged that most of the landslides fall in High and Very High zones with high fitting degree. The High and Very High zones mainly distributed in the south and southeast of the catchment, and Low and Very Low zones mainly in the north and west. Table 3 shows that in Very Low and Low zones, there were almost no landslides, reaching only 0.01% and 0.32% of all the landslides area, respectively, while in the High and Very High zones, the ratio reached altogether 93.08%, especially in the Very High zone, the ratio is 79.41%, with low zone area (30,855.06 hm^2) but Very High landslide area, the landslide density is 2.33%.

In this paper, we have made an analysis between landslides and their predisposing factors, without considering landslide typology. The four basic factors in those groups were selected based on experience. Moreover, in the MIV model proposed in this paper, the typical negative information values in IV modal are constrained between 0 and 1 and that could possible affect the predictive capacity of the model. All those points may bring some uncertainty to the results of this paper. While,

through the validation of ROC model, it can be judged that the results are highly reliable. To precisely judge the effects of these points, we will make it one of our key issues in our future research.

7. Conclusions

Landslides are one of the most hazardous geological disasters in nature and are usually initiated under complex environmental conditions. To find the optimal combination of predisposing factors and create a landslide susceptibility map based on them is of great significance and interest to planning agencies for preliminary hazard studies, especially when a regulatory planning policy is to be implemented. In the present paper, the Information Value model was modified to create the MIV model, based on which raster layers of total information value on every pixel was generated in ARCGIS for all 32 combinations separately, and together with AUC test in ROC curve, the optimal combination group of predisposing factors was selected and a landslide susceptibility map was drawn based on the group. Results show that, (1) MIV model can tackle the problem of "no data in subclass" well and generate the true information value and real running trend, which performs well in showing the relationship between predisposing factors and landslide occurrence and can be perfectly used for preliminary landslide susceptibility assessment in the study area; (2) factor combination group with members of Slope, Lithology, Drainage network, Annual precipitation, Faults, Road and Vegetation, was selected as the optimal combination group to express the susceptibility of landslide occurrence in Baoxing catchment, with an accuracy of 95.0%; (3) in the landslide susceptibility zonation map drawn based on the optimal group, the whole Baoxing Catchment was reclassified into five classes (very high, high, moderate, low, and very low), which presented an accurate description of different levels of landslide susceptibility with 79.41% and 13.67% of the validating field survey landslides falling in the very High and High zones, respectively, both mainly distributed in the south and southeast of the catchment.

Author Contributions: Dongchuan Wang conceived and designed the research. Dongchuan Wang and Lihui Zhang generally supervised the research group. Qianqian Wang and Mengqin Sang drafted the article. Qianqian Wang and Zhiheng Wang, Wei Chen and Wengang Chen collected and processed the data; Wengang Chen, Qianqian Wang, Qiaozhen Guo and Yong Huang interpreted and analyzed the data. All authors have revised the article critically and approved the final manuscript.

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

References

1. Dahl, M.P.J.; Mortensen, L.E.; Veihe, A.; Jensen, N.H. A simple qualitative approach for mapping regional landslide susceptibility in the Faroe Islands. *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 159–170. [CrossRef]
2. Lepore, C.; Kamal, S.A.; Shanahan, P.; Bras, R.L. Rainfall-induced landslide susceptibility zonation of Puerto Rico. *Environ. Earth Sci.* **2012**, *66*, 1667–1681. [CrossRef]
3. Mezughi, T.H.; Akhir, J.M.; Rafek, A.G.; Abdullah, I. Landslide susceptibility assessment using frequency ratio model applied to an area along the E-W Highway (Gerik-Jeli). *Am. J. Environ. Sci.* **2011**, *7*, 43–50. [CrossRef]
4. Gong, J.H.; Wang, D.C.; Li, Y.; Zhang, L.H.; Yue, Y.J.; Zhou, J.P.; Song, Y.Q. Earthquake-induced geological hazards detection under hierarchical stripping classification framework in the Beichuan area. *Landslides* **2010**, *7*, 181–189. [CrossRef]

5.   Bathrellos, G.D.; Gaki-Papanastassiou, K.; Skilodimou, H.D.; Skianis, G.A.; Chousianitis, K.G. Assessment of rural community and agricultural development using geomorphological-geological factors and GIS in the trikala prefecture (central Greece). *Stoch. Environ. Res. Risk Assess.* **2013**, *27*, 573–588. [CrossRef]

6.   Khezri, S. Landslide susceptibility in the Zab Basin, northwest of Iran. *Procedia Soc. Behav. Sci.* **2011**, *19*, 726–731. [CrossRef]

7.   Niu, Q.F.; Cheng, W.M.; Lan, H.X.; Liu, Y.; Xie, Y.W. Susceptibility assessment of secondary geological disaster based on information value methodology for Yushu earthquake region. *J. Mt. Sci.* **2011**, *29*, 243–249. (In Chinese).

8.   Sujatha, E.R.; Rajamanickam, V. Landslide susceptibility mapping of Tevankarai Ar sub-watershed, Kodaikkanal taluk, India, using weighted similar choice fuzzy model. *Nat. Hazards* **2011**, *59*, 401–425. [CrossRef]

9.   Bathrellos, G.D.; Kalivas, D.P.; Skilodimou, H.D. GIS-based landslide susceptibility mapping models applied to natural and urban planning in Trikala, central Greece. *Estud. Geol.* **2009**, *65*, 49–65. [CrossRef]

10. Dubey, C.S.; Chaudhry, M.; Sharma, B.K.; Pandey, A.C.; Singh, B. Visualization of 3-D digital elevation model for landslide assessment and prediction in mountainous Terrain: A case study of Chandmari landslide, Sikkim, eastern Himalayas. *Geosci. J.* **2005**, *9*, 363–373. [CrossRef]

11. Romeo, R.W.; Floris, M.; Veneri, F. Area-scale landslide hazard and risk assessment. *Environ. Geol.* **2006**, *51*, 1–13. [CrossRef]

12. Hart, A.B.; Hearn, G.H. Landslide assessment for land use planning and infrastructure management in the Paphos District of Cyprus. *Bull. Eng. Geol. Environ.* **2013**, *72*, 173–188. [CrossRef]

13. Lan, H.X.; Wu, F.; Zhou, C.; Wang, L.J. Spatial hazard analysis and prediction on rainfall-induced landslide using GIS. *Chin. Sci. Bull.* **2003**, *48*, 703–708. [CrossRef]

14. Zhang, Z.Y.; Fan, J.R.; Zhang, J.Q.; Su, F.H. Topographic factors analysis of Lushan earthquake-hit area. *J. Mt. Sci.* **2013**, *31*, 624–631. (In Chinese).

15. Gall, M.; Borden, K.A.; Emrich, C.T.; Cutter, S.L. The unsustainable trend of natural hazard losses in the United States. *Sustainability* **2011**, *3*, 2157–2181. [CrossRef]

16. Lin, S.K. Vulnerability, Risks, and Complexity: Impacts of Global Change on Human Habitats. *Sustainability* **2013**, *5*, 160–162. [CrossRef]

17. Kayastha, P. Landslide susceptibility mapping and factor effect analysis using frequency ratio in a catchment scale: A case study from Garuwa sub-basin, East Nepal. *Arab. J. Geosci.* **2015**, *59*, 401–425. [CrossRef]

18. Bathrellos, G.D.; Gaki-Papanastassiou, K.; Skilodimou, H.D.; Papanastassiou, D.; Chousianitis, K.G. Potential suitability for urban planning and industry development using natural hazard maps and geological-geomorphological parameters. *Environ. Earth Sci.* **2012**, *66*, 537–548. [CrossRef]

19. Kritikos, T.; Davies, T. Assessment of rainfall-generated shallow landslide/debris-flow susceptibility and runout using a GIS-based approach: Application to western southern ALPS of New Zealand. *Landslides* **2014**. [CrossRef]

20. Chen, W.; Li, W.P.; Hou, E.K.; Zhao, Z.; Deng, N.D.; Bai, H.Y.; Wang, D.Z. Landslide susceptibility mapping based on GIS and information value model for the Chencang District of Baoji, China. *Arab. J. Geosci.* **2014**, *7*, 4499–4511. [CrossRef]

21. Melo, R.; Vieira, G.; Caselli, A.; Ramos, M. Susceptibility modelling of hummocky terrain distribution using the information value method (Deception Island, Antarctic Peninsula). *Geomorphology* **2012**, *155–156*, 88–95. [CrossRef]

22. Xu, W.B.; Yu, W.J.; Jing, S.C.; Zhang, G.P.; Huang, J.X. Debris flow susceptibility assessment by GIS and information value model in a large-scale region, Sichuan Province (China). *Nat. Hazards* **2013**, *65*, 1379–1392. [CrossRef]

23. Sharma, L.P.; Patel, N.; Ghose, M.K.; Debnath, P. Development and application of Shannon's entropy integrated information value model for landslide susceptibility assessment and zonation in Sikkim Himalayas in India. *Nat. Hazards* **2015**, *75*, 1555–1576. [CrossRef]

24. Oliveira, S.C.; Zêzere, J.L.; Catalão, J.; Nico, G. The contribution of PSInSAR interferometry to landslide hazard in weak rocks dominated areas. *Landslides* **2015**, *12*, 703–719. [CrossRef]

25. Huang, Y.; Zheng, L.; Wang, D.C.; Chen, W.G.; Wang, Q. Spatial Distribution of Fragmentation by Diversion-Typed Hydroelectric Plant Exploitation in East Baoxing Catchment from 1999 to 2013. *Sustainability* **2015**, *7*, 3515–3527. [CrossRef]

26. Li, W.L.; Huang, R.Q.; Tang, C.; Xu, Q.; Westen, C.V. Co-seismic Landslide Inventory and Susceptibility Mapping in the 2008 Wenchuan Earthquake Disaster Area, China. *J. Mt. Sci.* **2013**, *10*, 339–354. [CrossRef]

27. Lan, H.X.; Zhou, H.C.; Gao, X.; Cheng, W.M.; Wang, Z.H.; Yang, Z.H.; Li, L.P.; Wu, Y.M. Secondary geological hazard assessment and hazard mitigation countermeasures in Lushan, Ya'an Earthquake, Sichuan Province. *Prog. Geogr.* **2013**, *32*, 499–504. (In Chinese).

28. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda Yahiko Mountains, Central Japan. *Geomorphology* **2005**, *65*, 15–31. [CrossRef]

29. Papadopoulou-Vrynioti, K.; Bathrellos, G.D.; Skilodimou, H.D.; Kaviris, G.; Makropoulos, K. Karst collapse susceptibility mapping considering peak ground acceleration in a rapidly growing urban area. *Eng. Geol. Eng. Geol.* **2013**, *158*, 77–88. [CrossRef]

30. Yin, K.L.; Yan, T.Z. Statistical prediction model for slope instability of metamorphosed rocks. In Proceedings of the 5th International Symposium on Landslides, Lausanne, Switzerland, 10–15 July 1988; Bonnard, C., Ed.; Balkema: Rotterdam, The Netherland, 1988; Volume 2, pp. 1269–1272.

31. Park, S.H.; Goo, J.M.; Jo, C.H. Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists. *Korean J. Radiol.* **2004**, *5*, 11–18. [CrossRef] [PubMed]

32. Fratinni, P.; Crosta, G.; Carrara, A. Techniques for evaluating the performance of landslide susceptibility models. *Eng. Geol.* **2010**, *111*, 62–72. [CrossRef]

33. Pereira, S.; Zêzere, J.L.; Bateira, C. Technical Note: Assessing predictive capacity and conditional independence of landslide predisposing factors for shallow landslide susceptibility models. *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 979–988. [CrossRef]

34. Xu, C.; Dai, F.C.; Xu, X.; Lee, Y.H. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* **2012**, *145–146*, 70–80. [CrossRef]

35. Wu, X.L.; Niu, R.Q.; Ren, F.; Peng, L. Landslide susceptibility mapping using rough sets and back-propagation neural networks in the Three Gorges, China. *Environ. Earth Sci.* **2013**, *70*, 1307–1318. [CrossRef]

36. Chalkias, C.; Ferentinou, M.; Polykretis, C. GIS supported landslide susceptibility modeling at regional scale: An expert-based fuzzy weighting method. *ISPRS Int. J. Geo-Inf.* **2014**, *3*, 523–539. [CrossRef]

37. Lee, S.; Pradhan, B. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* **2007**, *4*, 33–41. [CrossRef]

38. Rozos, D.; Bathrellos, G.D.; Skillodimou, H.D. Comparison of the implementation of rock engineering system and analytic hierarchy process methods, upon landslide susceptibility mapping, using GIS: A case study from the eastern Achaia county of Peloponnesus, Greece. *Environ. Earth Sci.* **2011**, *63*, 49–63. [CrossRef]

39. Xu, C.; Dai, F.C.; Yao, X.; Chen, J.; Tu, X.B.; Cao, Y.B.; Xiao, J.Z. GIS based certainty factor analysis of Landslide triggering factors in Wenchuan earthquake. *Chin. J. Rock Mech. Eng.* **2010**, *29*, 2972–2981. (In Chinese).

40. Miao, C.Y.; Ni, J.R.; Borthwick, A.G.L. Recent changes in water discharge and sediment load of the Yellow River basin, China. *Prog. Phys. Geogr.* **2010**, *34*, 541–561. [CrossRef]

41. Akbar, T.A.; Ha, S.R. Landslide hazard zoning along Himalayan Kaghan Valley of Pakistan—By integration of GPS, GIS, and remote sensing technology. *Landslides* **2011**, *8*, 527–540. [CrossRef]

42. Mercer, J.; Kurvits, T.; Kelman, I.; Mavrogenis, S. Ecosystem-based adaptation for food security in the aims SIDS: Integrating external and local knowledge. *Sustainability* **2014**, *6*, 5566–5597. [CrossRef]

43. Miao, C.Y.; Yang, L.; Chen, X.H. The vegetation cover dynamics (1982–2006) in different erosion regions of the Yellow River basin, China. *Land Degrad. Dev.* **2012**, *23*, 62–71. [CrossRef]

44. Cui, P.; Chen, X.Q.; Zhang, J.Q.; Yang, Z.J.; You, Y.; Fan, J.R.; Su, F.H.; Kong, Y.D.; Zhu, X.H. Activities and Tendency of Mountain Hazards Induced by the Ms7.0 Lushan Earthquake, 20 April 2013. *J. Mt. Sci.* **2013**, *31*, 257–265. (In Chinese).

45. Miao, C.Y.; Ashouri, H.; Hsu, K.; Sorooshian, S.; Duan, Q.Y. Evaluation of the PERSIANN-CDR daily rainfall estimates in capturing the behavior of extreme precipitation events over China. *J. Hydrometeorol.* **2015**, *16*, 1387–1396. [CrossRef]

46. Chen, Y.C.; Wu, C.F.; Lin, S.H. Mechanisms of forest restoration in landslide treatment areas. *Sustainability* **2014**, *6*, 6766–6780. [CrossRef]