

Article

# Analysis and Mapping of Rainfall-Induced Landslide Susceptibility in A Luoi District, Thua Thien Hue Province, Vietnam

Nguyen Thanh Long <sup>1</sup> and Florimond De Smedt <sup>2,\*</sup> 

<sup>1</sup> Department of Economic Geology and Geomatics, Vietnam Institute of Geosciences and Mineral Resources, Hanoi 084-04, Vietnam; ntlong1974@yahoo.com

<sup>2</sup> Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, 1050 Brussels, Belgium

\* Correspondence: fdesmedt@vub.ac.be; Tel.: +32-26293547

Received: 30 October 2018; Accepted: 21 December 2018; Published: 29 December 2018



**Abstract:** Rainfall-induced landslides form an important natural threat in Vietnam. The purpose of this study is to explore regional landslide susceptibility mapping in the mountainous district of A Luoi in Thua Thien Hue Province, where data on the occurrence and causes of landslides are very limited. Three methods are applied to examine landslide susceptibility: statistical index, logistic regression and certainty factor. Nine causative factors are considered: elevation, slope, geological strata, fault density, geomorphic landforms, weathering crust, land use, distance to rivers and annual precipitation. The reliability of the landslide susceptibility maps is evaluated by a receiver operating characteristic curve and the area under the curve is used to quantify and compare the prediction accuracy of the models. The certainty factor model performs best. This model is optimized by maximizing the difference between the true positive rate and the false positive rate. The optimal model correctly identifies 84% of the observed landslides. The results are verified with a validation test, whereby the model is calibrated with 75% randomly selected observed landslides, while the remaining 25% of the observed landslides are used for validation. The validation test correctly identifies 81% of the observed landslides in the training set and 73% of the observed landslides in the validation set.

**Keywords:** landslide; natural hazard; statistical index; logistic regression; certainty factor; Vietnam

## 1. Introduction

Landslides caused by tropical storms are common in the mountainous areas of Vietnam and have a major impact on local living conditions [1]. Therefore, prediction of the landslide susceptibility is of great importance to reduce the loss of life or property. Causative factors for landslides are numerous and varied, including geomorphological and geological as well as human-influenced factors [2]. Rainfall-induced landslides are generally caused by changes in pore water pressure, which lead to a reduction of soil strength and consequently slope instability. These changes are often produced by an increase in the groundwater level [3–6], an increase in the saturation degree of the soil [7–10], or by the formation of a perched water table [11]. Hence, causative factors for predicting landslide susceptibility must be carefully assessed on the basis of relevance, availability and scale. However, this is difficult in Vietnam because systematic studies and inventories of spatial characteristics and land-use have only been initiated recently. Lee and Dan [12] evaluated the landslide susceptibility in the province of Lai Chau in Vietnam, with an emphasis on the influence of tectonic fractures. Bui et al. [13] applied the statistical index and the logistic regression methods to estimate the landslide susceptibility in the province of Hoa Binh and concluded that the distance to roads, slope and lithology are important

causative factors for landslides. Long and De Smedt [14] used the analytical hierarchical process approach for analysing landslide susceptibility in the province of Thua Thien Hue and concluded that slope angle and precipitation are the main causative factors for landslides. Long and De Smedt [15] applied a physically based slope stability model to determine safety factors for landslides in Thua Thien Hue Province using the maximum daily precipitation recorded in a 28-year period as a triggering mechanism for slope instability. Hung et al. [16] presented landslide susceptibility maps for the upper Lo River catchment in northern Vietnam, using the analytical hierarchy process and a weighted linear combination model, and concluded that slope and weathering crust are of great importance for predicting landslides.

The most obvious approach to every study of landslide susceptibility is to compile a landslide inventory and analyse the relationship with possible causative factors to predict landslide prone areas. Various statistical techniques have been proposed for assessing landslide susceptibility. In bivariate statistical analysis, the importance of each causative factor is determined on the basis of the observed landslide density within each class of a factor map, after which all factor maps are combined by a weighting procedure to obtain a landslide susceptibility map. There are several ways in which such analyses can be performed. A very popular technique is the statistical index method [17–29].

In multivariate statistical analyses, all causal factors controlling the landslide events are analysed together to indicate the relative contribution of each of these factors to the degree of hazard [30]. There are also different techniques for this. One is the logistic regression analysis, which is widely used to predict success or failure of a process based on a set of observations. Instead of using a linear relationship between the independent variables and the response, a logistic function is used [28]. Many studies have used logistic regression to assess landslides [13,25,31–41]. Budmir et al. [42] presented an overview of landslide probability mapping using logistic regression.

Certainty factor analysis was introduced as an alternative to statistical approaches to avoid limitations such as independency of the data. Certainty factor analysis expresses belief or the disbelief of a hypothesis, including inconclusiveness due to insufficient, inaccurate or contradicting data. Certainty factor analysis has become the standard approach for uncertainty management in rule-based systems and has been applied successfully for assessing landslide susceptibility [26,39,43–48].

The aim of this work is to test the statistical index, logistic regression and certainty factor analysis for predicting rainfall-induced landslide susceptibility in Vietnam. The results obtained by the different methods are compared and the most accurate technique is identified and validated.

## 2. Materials and Methods

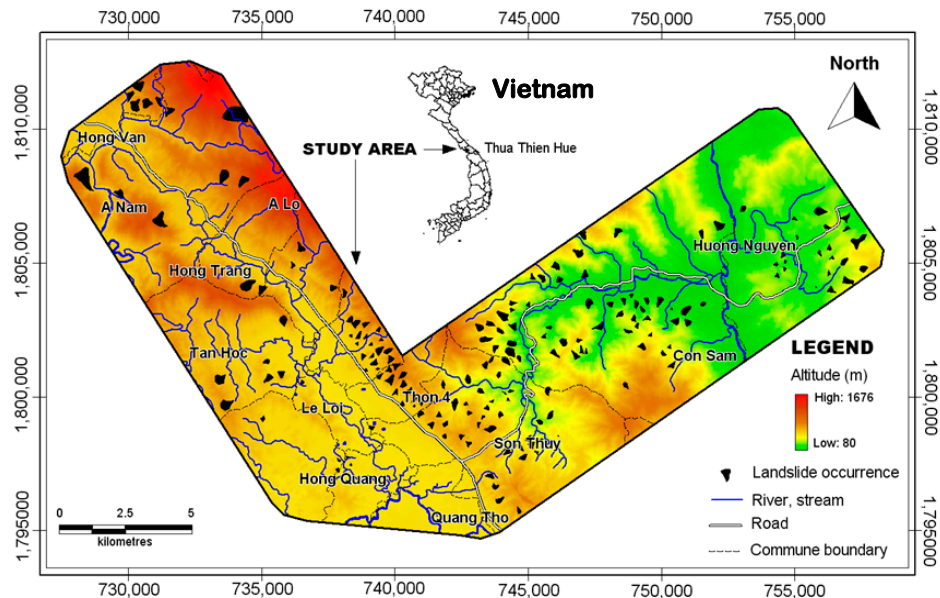
### 2.1. Study Area

The Thua Thien Hue Province is located in the centre of Vietnam. The climate is generally warm and humid because it is located in the tropical monsoon region. The area is frequently influenced by an intertropical convergence zone that typically causes tropical low pressures and typhoons, leading to annual rainfall of about 3500 mm with an average of 200 rainy days per year [49]. The wet season lasts 4 months, from September to December, with 70–80% of the total rainfall.

The A Luoi district in the west along the border with Laos is very mountainous. The mountains are strongly incised and steep, with attitudes ranging from 500 to 1700 m. The landscape is dominated mainly by shrubs or bare soil, as a result from bombings and spraying of defoliants during the Second Indochina War, and by small remains of broad evergreen tropical forest mixed with afforested land, mainly acacia plantations. The area is sparsely populated, mostly by ethnic minorities clustered in small villages in the valleys. These people practice crop cultivation, for example wetland rice, cassava, and maize, and harvest forest products such as rattan and bamboo.

In the A Luoi district, tropical storms usually occur several times a year, which often causes landslides, especially in the mountainous areas with steep slopes. In recent years, the risk of landslides has increased due to man-made activities as agriculture and deforestation [50]. Vãn et al. [51] identified

181 landslides with a total area of about 7 km<sup>2</sup>, based on the interpretation of aerial photographs and field research. A study area of 263 km<sup>2</sup> was selected for assessing the landslide susceptibility covering the main mountain ranges and all observed landslides (Figure 1). The observed landslide density in this area is about 2.7%. Unfortunately, no details were found or noted about the width, depth, types or causes of the landslides. Such data are often missing or incomplete, especially in remote and rural regions such as this study area.



**Figure 1.** Location of the study area in the district of A Luoi, Thua Thien Hue Province, Vietnam.

## 2.2. Data Sources

Nine digital causative factor maps with a pixel size of 30 m by 30 m were prepared for landslide susceptibility analysis [14,50]:

- Elevation: a digital elevation map (DEM) was obtained by digitizing the topographic map on scale 1:50,000 of the Ministry of Natural Resources and Environment (Figure 1), and an elevation class map was derived by classifying the height in intervals of 250 m (Table 1).
- Slope: a digital slope angle map was derived from the DEM and a slope class map by separating the slope angles into six classes: (1) flat-gentle (<5°), (2) fair (5–15°), (3) moderate (15–25°), (4) fairly moderate (25–35°), (5) steep (35–45°), and (6) very steep (>45°).
- Geology: a digital geological class map was derived from the Thua Thien Hue geological map, on a scale 1:50,000, compiled by Văn et al. [51], indicating 13 geological formations (Table 1). A description of the lithology of these formations is given by Long [50].
- Fault density: faults were digitized from the geological map and the fault density was derived as total length of faults per 1 km<sup>2</sup>; a categorical fault density map was obtained by classifying the fault density in intervals of 500 m/km<sup>2</sup> (Table 1).
- Geomorphology: eight geomorphological units identified by Văn et al. [51] were transformed in a digital map (Table 1).
- Weathering crust: a digital categorical map was derived from fieldwork in Thua Thien Hue Province carried out by Văn et al. [51], indicating Quaternary deposits and four types of weathering crusts: Sialite, Sialferrite, Ferrosialite, and mixtures of Silixite.
- Land use: a digital map was derived from a Landsat TM5 image of 20 February 1999 (Path/row: 125/48); four land uses were identified and verified in the field by Văn et al. [51], resulting in four land use classes: agriculture, forest, shrub and bare hills, and build-up land.

- Drainage distance: a digital map of the shortest distance to a watercourse was derived from the topographic map and a drainage distance class map was obtained by subdividing the values into classes <50 m, 50–200 m and >200 m (adapted from the literature, e.g., reference [20]).
- Precipitation: average annual precipitation was selected as rainfall causative factor for landslide analysis, because precise information about the intensity of individual storms is not available in the study area; the precipitation map was derived by spatial interpolation (inverse distance weighting) of the average annual precipitation observed from 1976 to 2003 in three climate stations in the A Luoi district [52]; the values range from about 2900 mm/year to 3500 mm/year and because this is a rather small range the precipitation class map was derived by dividing the values into just three classes: <3100 mm/year, 3100–3300 mm/year and >3300 mm/year.

### 2.3. Methods for Landslide Susceptibility Analysis

The statistical index (SI) method is based on a bivariate statistical comparison of a landslide inventory map with a categorical causative factor map [17,29]. Weight values for each class of the causative factor are determined as the natural logarithm of the landslide density in that class divided by the landslide density in the entire map [17]

$$w_{ij} = \ln(f_{ij}/f), \quad (1)$$

where  $w_{ij}$  is the weight of class  $j$  of parameter  $i$ ,  $f_{ij}$  the landslide density within class  $j$  of parameter  $i$ , and  $f$  the landslide density within the entire map. Statistically,  $f_{ij}$  is the conditional probability of a landslide event occurring in class  $j$  of parameter  $i$  and  $f$  is the prior probability of a landslide event occurring in the entire study area. Thus, each causative factor map is overlaid with the landslide map and the landslide frequency ratio  $f_{ij}/f$  and weight value in each class of the factor map is determined. The weight value is calculated only for classes with landslide occurrences and a zero value is assigned otherwise [17,20], which implies that the related parameter class has no impact on the landslide susceptibility. By overlying all causative factor class maps and adding the weights, a landslide susceptibility map is obtained that expresses the relative likelihood for landslide occurrence.

In logistic regression (LR), the quantitative relationship between the occurrence of landslides and its dependence on a set of causative factors is expressed as a logistic function [31,42]

$$p = [1 + \exp(-a_0 - \sum_i a_i x_i)]^{-1}, \quad (2)$$

where  $p$  is the probability of a landslide event,  $x_i$  are causative factors, and  $a_i$  are regression coefficients. The coefficients are estimated by non-linear regression, imposing  $p = 1$  in known landslide areas and  $p = 0$  elsewhere. Then the probability for landslides in each mapping unit is predicted with Equation (2) to obtain a landslide susceptibility map. One of the advantages of LR over other methods is that the probabilities always fall between 0 and 1. Equation (2) implies that all causative factors are numerical variables. Thus, in the case of categorical maps, the classes can be substituted by their corresponding landslide frequency ratios  $f_{ij}/f$ . In this study, the SAS 9.1 software (SAS Institute Inc., Cary, NC, USA) [53] is used to process the data and estimate the regression coefficients.

The certainty factor approach is an expert system similar to probabilistic reasoning but is less formally. In certainty factor (CF) analysis, the relationship between the occurrence of landslides and a categorical causative factor map is calculated as [43]

$$CF_{ij} = \begin{cases} (f_{ij} - f) / [f_{ij}(1 - f)] & f_{ij} \geq f \\ (f_{ij} - f) / [f(1 - f_{ij})] & f_{ij} < f \end{cases}, \quad (3)$$

where  $CF_{ij}$  is the certainty factor of class  $j$  of parameter  $i$ . The values range between  $-1$  and  $+1$ , whereby  $-1$  means definitely false and  $+1$  means definitely true. Positive values indicate an increasing

certainty in causality, while negative values correspond to the opposite. A value equal to or close to zero means that it is difficult to give any indication about causality. A combination of two CF values,  $x$  and  $y$ , is a CF value,  $z$ , calculated as follows [40,43]

$$z = \begin{cases} x + y - xy & x, y \geq 0 \\ (x + y) / [1 - \min(|x|, |y|)] & xy < 0 \\ x + y + xy & x, y < 0 \end{cases} \quad (4)$$

The combination rule expressed by Equation (4) enables to combine the CF values of all causative factor class maps to obtain a landslide susceptibility map.

**Table 1.** Causative factors for landslide, classes in each factor, and associated area ratios, landslide frequency ratios, SI weights and CF values (significant values are indicated in bold).

Factor	Class	Area (%)	$f_{ij}/f$	$W_{ij}$	$CF_{ij}$
Elevation	<250 m	14.6	0.53	−0.63	−0.47
	250–500 m	21.1	1.46	0.38	0.33
	500–750 m	48.9	0.81	−0.21	−0.20
	>750 m	15.4	1.42	0.35	0.30
Slope	<5°	21.6	<b>0.10</b>	<b>−2.30</b>	<b>−0.90</b>
	5–15°	19.5	0.57	−0.55	−0.43
	15–25°	23.4	1.42	0.35	0.30
	25–35°	21.1	1.52	0.42	0.35
	35–45°	11.2	1.46	0.38	0.32
	>45°	3.3	1.55	0.44	0.37
Geology	Dai Loc Complex (igneous)	8.8	<b>2.16</b>	<b>0.77</b>	<b>0.55</b>
	Lower A Lin (sedimentary)	11.6	0.97	−0.03	−0.03
	Lower A Vuong (metamorphic)	5.8	<b>2.19</b>	<b>0.78</b>	<b>0.56</b>
	Lower Ben Giang - Que Son (igneous)	13.9	1.00	0.00	0.00
	Lower Long Dai (metamorphic)	8.4	1.12	0.11	0.11
	Lower Nui Vu (metamorphic)	10.5	<b>0.47</b>	<b>−0.76</b>	<b>−0.54</b>
	Middle A Vuong (metamorphic)	0.8	0	0	−1.00
	Middle Long Dai (metamorphic)	17.6	1.26	0.23	0.21
	Middle-upper Pleistocene	2.1	0	0	−1.00
	Upper A Lin (sedimentary)	8.0	<b>0.25</b>	<b>−1.37</b>	<b>−0.75</b>
	Upper Ben Giang - Que Son (igneous)	1.3	0	0	−1.00
	Upper Long Dai (metamorphic)	9.3	0.52	−0.67	−0.49
Upper Nui Vu (metamorphic)	1.9	0	0	−1.00	
Fault density	<500 m/km <sup>2</sup>	5.1	1.20	0.17	0.16
	500–1000 m/km <sup>2</sup>	60.0	0.96	−0.04	−0.04
	1000–1500 m/km <sup>2</sup>	28.0	0.74	−0.30	−0.26
	>1500 m/km <sup>2</sup>	6.8	<b>2.28</b>	<b>0.83</b>	<b>0.58</b>
Geomorphology	Alluvium deposits	6.3	<b>0.11</b>	<b>−2.27</b>	<b>−0.90</b>
	Erosional channels and riverbeds	10.7	0	0	−1.00
	Early Quaternary valley pediment	22.5	0.54	−0.62	−0.47
	Wash slope	5.5	0.53	−0.64	−0.48
	Erosional-denudational slope	7.1	<b>2.01</b>	<b>0.71</b>	<b>0.52</b>
	Quick gravity slope (debris flow)	11.8	1.23	0.21	0.19
	Slow gravity slope (earth flow)	28.5	1.82	0.60	0.46
	Planation surface	7.6	<b>0.47</b>	<b>−0.73</b>	<b>−0.53</b>
Weathering crust	Quaternary deposit	2.1	0	0	−1.00
	Ferrosialite	20.9	0.67	−0.39	−0.33
	Mixtures of Silixite	28.5	0.90	−0.10	−0.10
	Sialferrite	35.1	1.22	0.20	0.18
	Sialite	13.3	1.32	0.27	0.24
Land use	Agriculture	4.2	0	0	−1.00
	Forest	27.2	1.20	0.18	0.17
	Shrub and bare hill	67.1	1.01	0.01	0.01
	Built-up area	1.5	0	0	−1.00
Distance to river	≤50 m	8.7	<b>0.21</b>	<b>−1.59</b>	<b>−0.80</b>
	50–200 m	22.7	0.64	−0.44	−0.36
	>200 m	68.6	1.22	0.20	0.18
Precipitation	<3100 mm/y	32.8	0.92	−0.08	−0.08
	3100–3300 mm/y	39.0	1.23	0.20	0.19
	>3300 mm/y	28.1	0.78	−0.25	−0.23

The areas of all factor classes, expressed as a percentage of the total study area, are shown in Table 1.

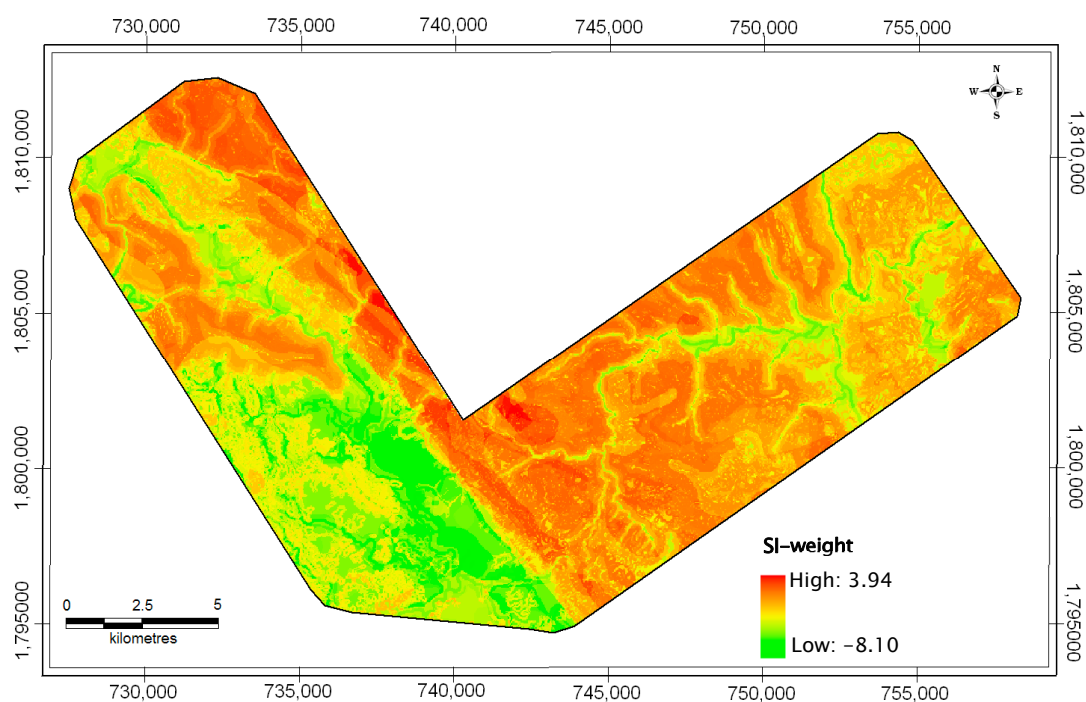
#### 2.4. Model Verification

The prediction accuracy of the models is evaluated by means of a receiver operating characteristic (ROC) curve, obtained by plotting the true positive rate (TPR) against the false positive rate (FPR) e.g., [13,23,25,26,40]. The first is the fraction of observed landslides in the zone with landslide susceptibility values (i.e., SI weight, LR probability or CF value) greater than a certain threshold and the latter is the fraction of areas free from landslides in the zone with scores larger than the threshold. The threshold varies from the minimum to the maximum landslide susceptibility value, so that TPR and FPR values vary between zero and one. The overall quality of a model is determined by the area under the ROC curve (AUC) and the model with the highest AUC value is considered the best e.g., references [13,23,26,40]. An AUC value of one indicates a perfect model, while a model that randomly predicts occurrences of landslides gives a value of 0.5. In practice, the AUC values are somewhere between these two extremes.

A good model has a high TPR value and a low FPR value. The optimal threshold of a model can thus be obtained by maximising TPR and minimising FPR. There are different ways to achieve this. One possibility used in this study is to maximise the difference  $TPR - FPR$  (i.e., Youden's index [44]), which corresponds to the point on the ROC curve furthest from a random guess ROC curve, given by  $TPR = FPR$ , i.e., the diagonal line of the ROC graph. Finally, the reliability of the optimal model for predicting the landslide susceptibility in the study area is verified by means of a validation test, where the observed landslides are randomly divided into a 75% training set used for model calibration, while the remaining 25% are used for validation of the model prediction efficiency e.g., references [39,47].

### 3. Results

Landslide frequency ratios and the resulting SI weights obtained with Equation (1) for each class of the nine causative factors are shown in Table 1. These weights are added by overlapping the factor maps resulting in a landslide susceptibility map depicted in Figure 2.

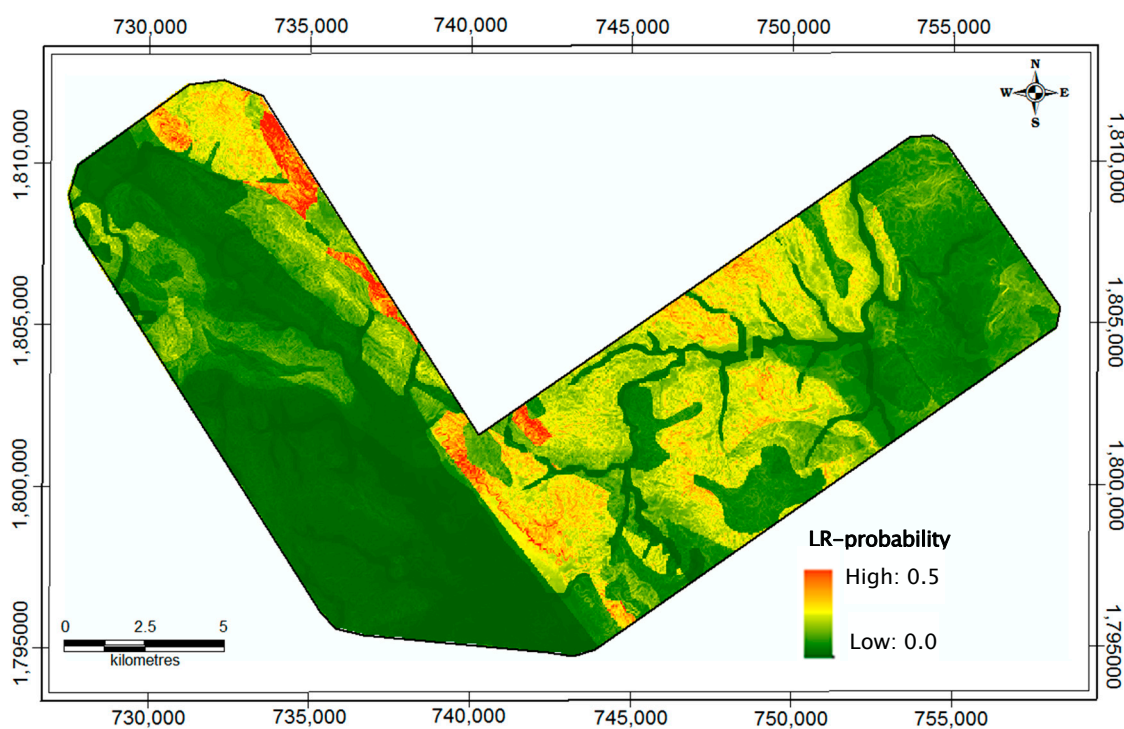


**Figure 2.** Landslide susceptibility map based on the statistical index model.

For application of the LR model, numerical maps are used for elevation, slope, fault density, drainage distance and precipitation, and the classes of geology, geomorphology, land use and weathering crust are substituted by their corresponding landslide frequency ratio. Thus, all factor data become numerical. The estimated regression coefficients and their standard error, Student’s *t*-statistic and corresponding *p*-value are shown in Table 2. With these coefficients and Equation (2), a landslide susceptibility map is derived, as shown in Figure 3.

**Table 2.** Estimated coefficients of the logistic regression model, with standard error and corresponding Student’s *t*-statistic and *p*-value.

Parameter	Coefficient	Standard Error	<i>t</i> -Score	<i>p</i> -Value
Intercept	−6.88	0.366	−18.8	<10 <sup>−4</sup>
Elevation	4.29 × 10 <sup>−4</sup>	0.44 × 10 <sup>−4</sup>	9.75	<10 <sup>−4</sup>
Slope	0.018	0.001	18.0	<10 <sup>−4</sup>
Geology	0.660	0.021	31.3	<10 <sup>−4</sup>
Fault density	7.90 × 10 <sup>−5</sup>	3.30 × 10 <sup>−5</sup>	2.39	0.008
Geomorphology	0.774	0.021	37.8	<10 <sup>−4</sup>
Land use	−0.134	0.061	−2.21	0.014
Weathering crust	−0.141	0.057	−2.47	0.007
Distance to river	3.06 × 10 <sup>−4</sup>	0.30 × 10 <sup>−4</sup>	10.2	<10 <sup>−4</sup>
Precipitation	−1.87 × 10 <sup>−3</sup>	0.03 × 10 <sup>−3</sup>	−69.3	<10 <sup>−4</sup>



**Figure 3.** Landslide susceptibility map based on the logistic regression model.

The CF values determined with Equation (3) for all classes of each causative factor are given in the last column of Table 1. The combination of these CF values using Equation (4) results in the landslide susceptibility map shown in Figure 4.

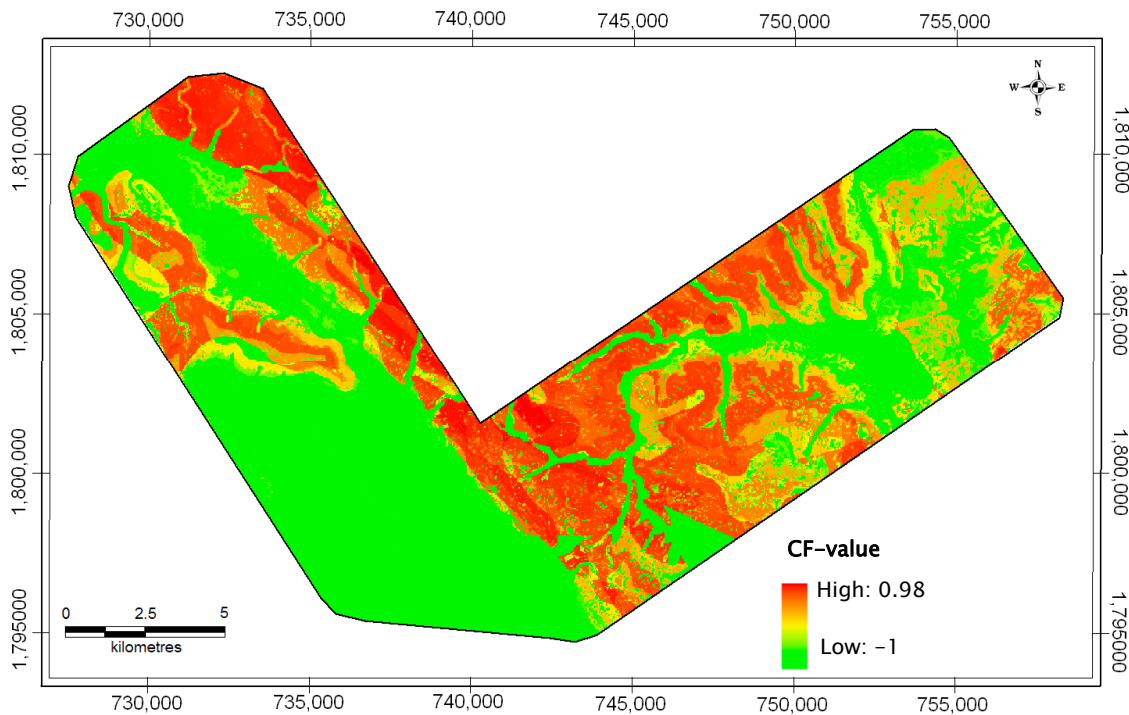


Figure 4. Landslide susceptibility map based on the certainty factor model.

ROC curves of all models are presented in Figure 5 together with their AUC values, from which it can be concluded that the CF model perform best, although differences with the other models are small. The optimum point on the ROC curve of the CF model corresponds to an optimal CF threshold value of 0.15, resulting in TPR and FPR values of 0.84 and 0.42, respectively.

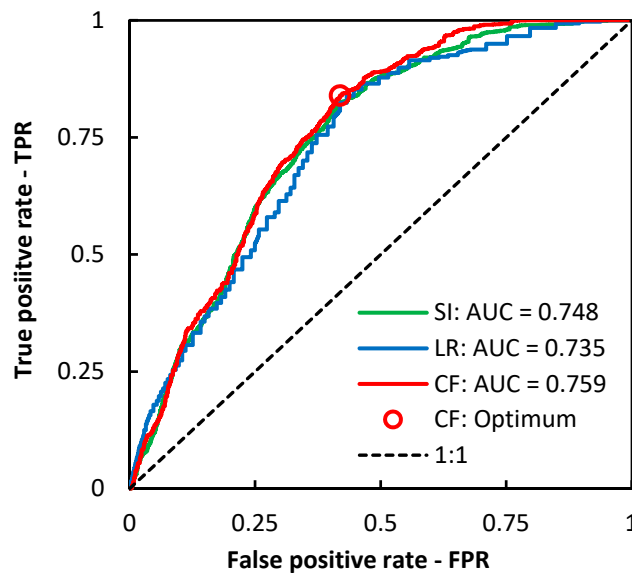


Figure 5. Receiver operating characteristic (ROC) curves and area under the curve (AUC) values showing the accuracy of the models for predicting landslide susceptibility (SI: statistical index model, LR: logistic regression model, CF: certainty factor model) and the optimum threshold point for the CF model.

For the validation test performed with the CF model and the 75% training set of observed landslides, results are obtained that closely resemble those in Table 1 and Figure 5 (even slightly better because AUC = 0.780) and the resulting landslide susceptibility map is almost identical as shown in



Figure 4. However, the optimal CF threshold value is 0.20, which is slightly larger than before, and the corresponding TPR and FPR values are 0.81 and 0.37 respectively, which are slightly smaller than before. The 75% training set consists of 136 landslides of which the model correctly identifies 110 landslides (81%), while for the remaining 25% validation set consisting of 45 landslides the model correctly identifies 33 landslides (73%).

## 4. Discussion

### 4.1. Statistical Index Model

The importance of each causative factor class can be inferred from its landslide frequency ratio  $f_{ij}/f$ . Factor classes that contribute strongly to landslides can be identified by frequency ratios larger than 2, which means that the landslide susceptibility is more than double the overall average. There are only four classes that strongly promote landslides: geology classes Dai Loc Complex (magmatic rock) and Lower A Vuong Formation (schist), both fairly soft rock types, and the highest fault density class and the geomorphology erosional-denudational slope class, both for obvious reasons. Factor classes that strongly avert landslides have frequency ratios of less than 0.5, which means that the landslide susceptibility is less than half of the overall average. There are six classes that fall into this category: the smallest slope class, geomorphology classes Alluvial deposits and Planation surface and the smallest drainage distance class (river banks), which all relate directly to nearly flat areas, and the geological classes Lower Nui Vu Formation and Upper A Lin formation, which also occur predominantly in flat areas.

Factor classes with a frequency ratio around one have little influence on the occurrence of landslides because the observed landslide density in these classes is close to the overall average landslide density in the study area. Some classes have a zero-frequency ratio, meaning that no landslides have been observed in these areas. Most of these areas are very small, except for the geomorphology class Erosional channels and riverbeds. Zero-frequency ratios pose a problem for the SI model because weights cannot be determined by taking the logarithm. The standard remedy for this is to set the weight equal to zero, meaning one cannot decide whether the class affects the occurrence of landslides. This might be a correct interpretation in case of small areas where no landslides have occurred, but not for large areas without landslides. In the present case, a zero value for the frequency ratio for the geomorphology class Erosional channel and riverbeds probably indicates that landslides in such areas are very unlikely and therefore it is incorrect to set the weight to zero.

In the SI model, landslide frequency ratios of the causative factor classes are transformed into weights and accumulated to represent landslide susceptibility. The resulting landslide susceptibility values range from around minus eight to four (Figure 2). The high values indicate areas with a higher susceptibility for landslide occurrence, which are indicated in Figure 2 by red and orange colors. These areas form a large part of the study area and cover most of the observed landslides, except for some landslides in the southwest around Tan Hoc (Figure 1). Because the weights are accumulated without distinction or grading, each causative factor has the same impact and relationships or correlations between factors are ignored. This can lead to overestimation or underestimation of the landslide susceptibility. For example, a combination of high elevation and steep slope leads to high landslide susceptibility, although these causative factors might be strongly correlated and indicate the same tendency for occurrence of landslides. Therefore, extreme SI values are probably exaggerated. For future studies, this can be avoided by using principal components analysis or similar techniques to reduce redundant information and to produce a smaller set of uncorrelated variables.

The AUC value of 0.748 for the SI model is approximately half way through the feasible range of 0.5–1, indicating that the prediction accuracy is reasonable. However, values reported in the literature for SI models are usually slightly larger, for example in the range 0.79–0.86 [23–28], indicating that the data in the current study are insufficient to provide better prediction accuracy. The causative landslide factors used in this study were based on consideration of relevance, but especially on availability on

a regional scale. That is why the selection was relative and subjective and should be improved in future research.

#### 4.2. Logistic Regression Model

In the LR model, causative factors are weighted by means of the regression coefficients (Table 2), which enables to give less weight to correlated factors and to rank the causative factor according to their impact on landslides. A disadvantage is that there are no classes within a causative factor that can be graded individually and so the relative importance within a factor is fixed. For example, an elevation of 500 m will be ten times more susceptible for landslides than an elevation of 50 m, and so on.

The relative contribution of each causative factor to the logistic function can be obtained by comparing the significance of the corresponding regression coefficient. The p-values obtained for the estimated regression coefficients show that all coefficients are significant (Table 2). The relative importance of the causative factors in the LR model can be evaluated by comparing their Student's *t*-statistic value. As such, precipitation, geomorphology and geology are important causative factors, whereas slope, elevation and drainage distance are of intermediate importance, and weathering crust, fault density and land use are less important. In similar studies conducted in Vietnam, Bui et al. [13] concluded that slope and lithology are important factors for landslides and land-use, rainfall, distance to faults and distance to rivers are of medium importance, while Hung et al. [16] concluded that slope and weathering crust are of great importance. It is clear that the type of data and the way in which it is collected and analyzed has a strong influence on the results. That is why it remains difficult to decide in advance which factors are more important than others. Presumably, all factors that may affect a particular case should be considered and the ranking can only be achieved after the analysis is completed.

Some regression coefficients are negative, which needs to be explained or interpreted in a correct way. The negative coefficients for land use and weathering crust are abnormal because the values of these factors are landslide frequency ratios, which should contribute in a positive way to landslide susceptibility. The only explanation is that these factors are of little importance in the regression equation, as indicated by their low Student's *t*-statistic. The negative regression coefficient for precipitation can be explained by the fact that fewer landslides have been observed in areas with more rainfall, as indicated by the low landslide frequency ratio of the highest precipitation class (Table 1).

The resulting landslide susceptibility, obtained as probabilities estimated with Equation (2), range from zero to 0.5 (Figure 3). The high values, represented by red and orange colors in the landslide susceptibility map, constitute a much smaller portion of the study area than obtained with the SI method (Figure 2) and cover much less of the observed landslides. However, since the overall observed density of the observed landslides is very small, about 0.027, all probabilities above that value are significant, which also includes the areas colored in green in Figure 3. Therefore, most of the observed landslides are also well predicted by the LR landslide susceptibility map, except for landslides observed in the southwest around Tan Hoc. However, the ROC curve of LR is mostly below the curve of SI and the AUC value obtained for LR is lower than for SI (Figure 5), which indicates that the landslide susceptibility map obtained with the SI model predicts landslides better and is to be preferred over the map of the LR model. In similar studies, the LR model was preferred over the SI model [13,22], while in other studies it was the other way around [25]. The AUC value of 0.735 obtained for the LR model in the current study is on the low side compared to values reported in the literature, for example 0.74–0.89 [25,39,40], which indicates a shortcoming of the current data.

#### 4.3. Certainty Factor Model

The CF values derived from the landslide frequency ratios yield results that at first sight are comparable to those of the SI model. Causative factor classes that strongly promote landslides have CF values larger than 0.5, more specifically the highest fault density class, the geomorphology

erosional-denudational slope class and the geomorphology classes Dai Loc Complex and Lower A Vuong Formation, while classes that avert landslides have CF values lower than  $-0.5$ , in particular the smallest slope class, geomorphology classes Alluvial deposits and Planation surface, the smallest drainage distance class and geological classes Lower Nui Vu Formation and Upper A Lin formation (Table 1). However, a difference with the SI model is that these values are not accumulated, but instead are combined using Equation (4), which can reduce the problem of overestimation or underestimation due to correlated factors. Another advantage is that the classes with no occurrence of landslides are not treated as missing data, but get a CF value of  $-1$ , which means strongly preventing landslide occurrences. This is the case for the geology classes Middle A Vuong, Middle-upper Pleistocene, Upper Ben Giang–Que Son and Upper Nui Vu, geomorphology class Erosional channels and riverbeds, weathering class Quaternary deposits, and land use classes Agriculture and Built-up area. This interpretation is more realistic, especially for a class like Erosional channels and riverbeds that covers more than 10% of the study area.

The ROC curve of the CF model is mostly above the curves of the SI and the CF models and the AUC value obtained for CF is the largest (Figure 5), indicating that the landslide susceptibility map obtained with the CF model best matches the observed landslides and is to be preferred over the other maps. However, other studies have reported otherwise, for example the SI model with a larger AUC value than the CF model [26] or the LR model with a larger AUC value than the CF model [39], so there can be no general conclusion. Also, AUC values reported in the literature for other CF models are generally larger, for example 0.78 [26] and 0.89 [39].

#### 4.4. Optimal Model

The prediction accuracy of the models evaluated by means of the ROC curve and the corresponding AUC value indicates that differences between the model results are rather small. Nevertheless, the CF model can be considered the best because its AUC value is slightly larger than for the other models, but also because of the clear handling of inconclusive data as causative factor classes where no landslides have been observed. The AUC value of 0.759 obtained for the CF model is approximately in the middle of the value one for a perfect model and 0.5 for a random guess model, which implies that the predictive accuracy of the CF model is reasonable, but clearly could use some improvement. The reason for this is more than likely the lesser quality and incompleteness of the data and the lack of detailed information about the origin and causes of the landslides. Important factors that were not considered in this study could significantly affect the landslide susceptibility of a given site, such as geotechnical properties of the involved soils (e.g., strength parameters and hydraulic conductivity) or water pressure and saturation degree regimes initially existing in the slope.

The optimum point on the ROC curve of the CF model is obtained for a CF cut-off value of 0.15. The area with CF values larger than or equal to 0.15 corresponds to the orange and red zones indicated in Figure 4, covering about 43% of the study area. Long and De Smedt [14] found 37% of the study area highly susceptible to landslides using a hierarchical process approach, while Long and De Smedt [15] predicted with a physically based model that 29% of the area consists of very unstable slopes. All these results are more or less comparable, which indicates that irrespective of the research technique used, a large part of the area is susceptible to landslides. The optimal CF model correctly identifies 152 of the 181 observed landslides, i.e., 84%. Long and De Smedt [14] used the hierarchical process approach and predicted only 55% of the observed landslides, but Long and De Smedt [15] correctly predicted 87% of the landslides with a physically based slope stability model. The former negative result is probably due to the fact that the hierarchical process approach method is based on a subjective judgment on the relevance of the causative factors.

The 16% unidentified landslides are mainly located in the southwest of the study area around Tan Hoc and Hong Quang (Figure 1). These are areas with an average topographic height and with fair to moderate slopes (5–25%). Specific differences with other areas involve the Ferrosialite weathering crust and the higher precipitation rate. However, this does not explain the occurrence of landslides

observed in this region. Therefore, the causative factors of these landslides remain unanswered. Presumably, specific local circumstances that promote landslides were not considered in the current research, or perhaps not recognized as a result of errors or misinterpretations of field data and derived factor maps.

The validation test leads to comparable results. The optimal model correctly predicts 81% of the observed landslides in the training set and 73% of the observed landslides in the validation set, which is slightly less than previously obtained. The difference is clearly due to fact that fewer observed landslides were used for the model calibration, resulting in a slightly larger CF threshold value and a slightly smaller area with CF values larger than the threshold, which accounts for only 37% of the study area. However, the differences between model results for the training and validation data are not so important as to undermine the reliability of the model.

## 5. Conclusions

In this study, the statistical index, logistic regression and certainty factor models have been applied to derive landslide susceptibility maps for the mountainous district of A Luoi in Thua Thien Hue Province, Vietnam. The prediction accuracy of the models evaluated by means of a receiver operating characteristic indicates that the CF model can be considered the best, although differences between the model results are small. Comparison with the observed landslides indicates that 84% of the observed landslides are correctly predicted, and that 43% of the study area is very susceptible to landslides. A validation test leads to comparable results. The landslide susceptibility mapping provides an estimated spatial distribution of the sensitivity to landslides and can help to understand the relationship with causative factors and allows us to delineate zones where mitigating measures can be implemented. In particular, this study shows that the Dai Loc complex and Lower A Vuong Formation are clearly more favorable for landslides and the Lower Nu Vui and Upper A Lin Formations are clearly unfavorable; also, large faults density and erosional-denudational slopes appear to promote landslides, while alluvial deposits, planation surfaces and small drainage distances have an opposite effect.

**Author Contributions:** N.T.L. performed the data analysis and wrote the draft version of the paper; F.D.S. supervised the work, and corrected and finalized the paper.

**Funding:** This research received no external funding.

**Acknowledgments:** Special thanks are due to the researchers and managers of the Vietnam Institute of Geosciences and Mineral Resources for providing data used in this study. We also like to thank the anonymous reviewers for their valuable remarks and comments that greatly improved the quality of the paper.

**Conflicts of Interest:** The authors declare that there are no conflicts of interest.

## References

1. Nadim, F.; Kjekstad, O.; Peduzzi, P.; Herold, C.; Jaedicke, C. Global landslide and avalanche hotspots. *Landslides* **2006**, *3*, 159–174. [[CrossRef](#)]
2. Gorsevski, P.V.; Gessler, P.E.; Boll, J.; Elliot, W.J.; Foltz, R.B. Spatially and temporally distributed modeling of landslide susceptibility. *Geomorphology* **2006**, *80*, 178–198. [[CrossRef](#)]
3. Corominas, J.; Moya, J.; Ledesma, A.; Lloret, A.; Gili, J.A. Prediction of ground displacements and velocities from groundwater level changes at the Vallcebre landslide (Eastern Pyrenees, Spain). *Landslides* **2005**, *2*, 83–96. [[CrossRef](#)]
4. Rosso, R.; Rulli, K.M.; Vannucchi, G. A physically based model for the hydrologic control on shallow landsliding. *Water Resour. Res.* **2006**, *42*, W06410. [[CrossRef](#)]
5. Ledesma, A.; Corominas, J.; Gonzàles, A.; Ferrari, A. Modelling slow moving landslides controlled by rainfall. In Proceedings of the 1st Italian Workshop on Landslides, Napoli, Italy, 8–10 June 2009; pp. 196–205.
6. Conte, E.; Donato, A.; Pugliese, L.; Troncone, A. Analysis of the Maierato landslide (Calabria, Southern Italy). *Landslides* **2018**, *15*, 1935–1950. [[CrossRef](#)]
7. Iverson, R.M. Landslide triggering by rain infiltration. *Water Resour. Res.* **2000**, *36*, 1897–1910. [[CrossRef](#)]

8. Godt, J.W.; Baum, R.L.; Chleborad, A.F. Rainfall characteristics for shallow landsliding in Seattle, Washington, USA. *Earth Surf. Proc. Land.* **2006**, *31*, 97–110. [[CrossRef](#)]
9. Salciarini, D.; Godt, J.W.; Savage, W.Z.; Baum, R.L.; Conversini, P. Modeling landslide recurrence in Seattle, Washington, USA. *Eng. Geol.* **2008**, *102*, 227–237. [[CrossRef](#)]
10. Conte, E.; Troncone, A. Simplified approach for the analysis of rainfall-induced landslides. *J. Geotech. Geoenviron.* **2012**, *138*, 398–406. [[CrossRef](#)]
11. Shackelford, C.D.; Chang, C.K.; Chiu, T.F. The capillary barrier effect in unsaturated flow through soil barriers. In Proceedings of the First International Congress on Environmental Geotechnics, Edmonton, AB, Canada, 10–15 July 1994; pp. 789–793.
12. Lee, S.; Dan, N.T. Probabilistic landslide susceptibility mapping in the Lai Chau province of Vietnam: Focus on the relationship between tectonic fractures and landslides. *Environ. Geol.* **2005**, *48*, 778–787. [[CrossRef](#)]
13. Bui, D.; Lofman, O.; Revhaug, I.; Dick, O. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Nat. Hazards* **2011**, *59*, 1413–1444. [[CrossRef](#)]
14. Long, N.T.; De Smedt, F. Application of an analytical hierarchical process approach for landslide susceptibility mapping in A Luoi district, Thua Thien Hue Province, Vietnam. *Environ. Earth Sci.* **2012**, *66*, 1739–1752. [[CrossRef](#)]
15. Long, N.T.; De Smedt, F. Landslide susceptibility analysis based on a physical slope stability model: A case study from Thua Thien-Hue Province, Vietnam. *Landslides* **2014**, *11*, 897–907. [[CrossRef](#)]
16. Hung, L.Q.; Van, N.T.H.; Duc, D.M.; Ha, L.T.C.; Van Son, P.; Binh, L.T. Landslide susceptibility mapping by combining the analytical hierarchy process and weighted linear combination methods: A case study in the upper Lo River catchment (Vietnam). *Landslides* **2016**, *13*, 1285–1301. [[CrossRef](#)]
17. van Westen, C. Statistical landslide hazard analysis. In *ILWIS 2.1 for Windows Application Guide*; ITC Publication: Enschede, The Netherlands, 1997; pp. 73–84.
18. Cevik, E.; Topal, T. GIS-based landslide susceptibility mapping for a problematic segment of the natural gas pipeline, Hendek (Turkey). *Environ. Geol.* **2003**, *44*, 949–962. [[CrossRef](#)]
19. Süzen, M.L.; Doyuran, V. A comparison of the GIS based landslide susceptibility assessment methods: Multivariate versus bivariate. *Environ. Geol.* **2004**, *45*, 665–679. [[CrossRef](#)]
20. Yalcin, A. GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey). *Catena* **2007**, *72*, 1–12. [[CrossRef](#)]
21. Regmi, A.D.; Devkota, K.C.; Yoshida, K.; Pradhan, B.; Pourghasemi, H.R.; Kumamoto, T.; Akgun, A. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arab. J. Geosci.* **2014**, *7*, 725–742. [[CrossRef](#)]
22. Pourghasemi, H.R.; Moradi, H.R.; Fatemi Aghda, S.M. Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Nat. Hazards* **2013**, *69*, 749–779. [[CrossRef](#)]
23. Chen, W.; Li, W.; Hou, E.; Bai, H.; Chai, H.; Wang, D.; Cui, X.; Wang, Q. Application of frequency ratio, statistical index, and index of entropy models and their comparison in landslide susceptibility mapping for the Baozhong Region of Baoji, China. *Arab. J. Geosci.* **2015**, *8*, 1829–1841. [[CrossRef](#)]
24. Chen, W.; Chai, H.; Sun, X.; Wang, Q.; Ding, X.; Hong, H. A GIS-based comparative study of frequency ratio, statistical index and weights-of-evidence models in landslide susceptibility mapping. *Arab. J. Geosci.* **2016**, *9*, 204. [[CrossRef](#)]
25. Chen, T.; Niu, R.; Jia, X. A comparison of information value and logistic regression models in landslide susceptibility mapping by using GIS. *Environ. Earth Sci.* **2016**, *75*, 867. [[CrossRef](#)]
26. Wu, Y.; Li, W.; Wang, Q.; Liu, Q.; Yang, D.; Xing, M.; Pei, Y.; Yan, S. Landslide susceptibility assessment using frequency ratio, statistical index and certainty factor models for the Gangu County, China. *Arab. J. Geosci.* **2016**, *9*, 84. [[CrossRef](#)]
27. Wang, Q.; Li, W.; Wu, Y.; Pei, Y.; Xie, P. Application of statistical index and index of entropy methods to landslide susceptibility assessment in Gongliu (Xinjiang, China). *Environ. Earth Sci.* **2016**, *75*, 599. [[CrossRef](#)]
28. Razavizadeh, S.; Solaimani, K.; Massironi, M.; Kavian, A. Mapping landslide susceptibility with frequency ratio, statistical index, and weights of evidence models: A case study in northern Iran. *Environ. Earth Sci.* **2017**, *76*, 499. [[CrossRef](#)]
29. van Westen, C.J. *Application of Geographic Information System to Landslide Hazard Zonation*; ITC-Publication No. 15; Geo-Information Science and Earth Observation (ITC): Enschede, The Netherlands, 1993.

30. Brown, C.E. *Applied Multiple Statistics in Geohydrology and Related Sciences*; Springer-Verlag: Berlin, Germany, 1998.
31. Atkinson, P.; Massari, R. Generalised linear modelling of susceptibility for landsliding in the Central Apennines, Italy. *Comput. Geosci.* **1998**, *24*, 373–385. [[CrossRef](#)]
32. Lee, S.; Min, K. Statistical analysis of landslide susceptibility at Yongin, Korea. *Environ. Geol.* **2001**, *40*, 1095–1113. [[CrossRef](#)]
33. Dai, F.C.; Lee, C.F. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* **2002**, *42*, 213–228. [[CrossRef](#)]
34. Ohlmacher, C.G.; Davis, C.J. Using multiple regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Eng. Geol.* **2003**, *69*, 331–343. [[CrossRef](#)]
35. 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](#)]
36. Lee, S.; Sambath, L. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environ. Geol.* **2006**, *50*, 847–856. [[CrossRef](#)]
37. Yilmaz, 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](#)]
38. Sujatha, E.R.; Rajamanickam, G.V.; Kumaravel, P. Landslide susceptibility analysis using probabilistic certainty factor approach: A case study on Tevankarai stream watershed, India. *J. Earth Syst. Sci.* **2012**, *121*, 1337–1350. [[CrossRef](#)]
39. 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](#)]
40. Nourani, V.; Pradhan, B.; Ghaffari, H.; Sharifi, S.S. Landslide susceptibility mapping at Zonouz Plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. *Nat. Hazards* **2014**, *71*, 523–547. [[CrossRef](#)]
41. Sangchini, E.K.; Emami, S.N.; Tahmasebipour, N.; Pourghasemi, H.R.; Naghibi, S.A.; Arami, S.A.; Pradhan, B. Assessment and comparison of combined bivariate and AHP models with logistic regression for landslide susceptibility mapping in the Chaharmahal-e-Bakhtiari Province, Iran. *Arab. J. Geosci.* **2016**, *9*, 201. [[CrossRef](#)]
42. Budimir, M.E.A.; Atkinson, I.P.M.; Lewis, I.H.G. A systematic review of landslide probability mapping using logistic regression. *Landslides* **2015**, *12*, 419–436. [[CrossRef](#)]
43. Chung, C.J.F.; Fabbri, A.G. The representation of geoscience information for data integration. *Nonrenew. Resour.* **1993**, *2*, 122–139. [[CrossRef](#)]
44. Youden, W.J. Index for rating diagnostic tests. *Cancer* **1950**, *3*, 32–35. [[CrossRef](#)]
45. Binaghi, E.; Luzi, L.; Madella, P.; Rampini, A. Slope instability zonation: A comparison between certainty factor and fuzzy Dempster-Shafer approaches. *Nat. Hazards* **1998**, *17*, 77–97. [[CrossRef](#)]
46. Lan, H.X.; Zhou, C.H.; Wang, L.J.; Zhang, H.Y.; Li, R.H. Landslide hazard spatial analysis and prediction using GIS in the Xiaojiang watershed, Yunnan, China. *Eng. Geol.* **2004**, *76*, 109–128. [[CrossRef](#)]
47. Wang, Q.; Li, W.; Chen, W.; Bai, H. GIS-based assessment of landslide susceptibility using certainty factor and index of entropy models for the Qianyang County of Baoji city, China. *J. Earth Syst. Sci.* **2015**, *124*, 1399–1415. [[CrossRef](#)]
48. Raja, N.B.; Çiçek, I.; Türkoğlu, N.; Aydın, O.; Kawasaki, A. Landslide susceptibility mapping of the Sera River Basin using logistic regression model. *Nat. Hazards* **2017**, *85*, 1323–1346. [[CrossRef](#)]
49. Mai, D.T.; Thanh, N.T. Precipitation extremes over Vietnam. In Proceedings of the 5th International Scientific Conference on the Global Energy and Water Cycle, Orange County, CA, USA, 20–24 June 2005; p. 4.
50. Long, N.T. Landslide Susceptibility Mapping of the Mountainous Area in A Luoi district, Thua Thien Hue Province, Vietnam. Ph.D. Thesis, Vrije Universiteit Brussel, Brussels, Belgium, 2008.
51. Văn, T.T.; Tuy, P.K.; Giáp, N.X.; Ké, T.D.; Thái, T.N.; Giang, N.T.; Thọ, H.M.; Tuất, L.T.; San, D.N.; Hùng, L.Q.; et al. *Assessment of Geological Hazards in Coastal Central Provinces from Quang Binh to Phu Yen, Present Status, Causes, Forecasts and Proposed Measures to Prevent and Mitigate Consequences (In Vietnamese)*; Project n° 3451; Research Institute of Geology and Mineral Resources: Hanoi, Vietnam, 2003; 215p.

52. Vietnam Hydrometeorological Service. *Summary Report on Hydrometeorological Observation in Period 1976–2003 in Vietnam*; Vietnamese Ministry of Environment and Natural Resources: Hanoi, Vietnam, 2003; 375p.
53. SAS Institute Inc. SAS Software Version 9.1. 2007. Available online: <http://www.sas.com/> (accessed on 11 February 2008).



© 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/>).