

Article

GIS-Based Random Forest Weight for Rainfall-Induced Landslide Susceptibility Assessment at a Humid Region in Southern China

Peng Wang ^{1,5}, Xiaoyan Bai ^{2,*}, Xiaoqing Wu ³, Haijun Yu ⁴, Yanru Hao ⁵ and Bill X. Hu ⁵

- ¹ Department of Ecology, Jinan University, Guangzhou 510632, China; wangp@jnu.edu.cn
- ² Department of Environmental Engineering, School of Environmental Science and Engineering, Guangdong University of Technology, Guangzhou 510006, China
- ³ South China Institute of Environment Sciences, Ministry of Environment Protection of PRC, Guangzhou 510535, China; wxqingAnn@163.com
- ⁴ State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, Research Center on Flood and Drought Disaster Reduction of the Ministry of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing 100038, China; leivi@gmx.com
- ⁵ Institute of Groundwater and Earth Sciences, Jinan University, Guangzhou 510632, China; thyru@jnu.edu.cn (Y.H.); bill.x.hu@gmail.com (B.X.H.)
- * Correspondence: xybai7829@163.com

Received: 10 May 2018; Accepted: 28 July 2018; Published: 1 August 2018



Abstract: Landslide susceptibility assessment is presently considered an effective tool for landslide warning and forecasting. Under the assessment procedure, a credible index weight can greatly increase the rationality of the assessment result. Using the Beijiang River Basin, China, as a case study, this paper proposes a new weight-determining method based on random forest (RF) and used the weighted linear combination (WLC) to evaluate the landslide susceptibility. The RF weight and eight indices were used to construct the assessment model. As a comparison, the entropy weight (EW) and weight determined by analytic hierarchy process (AHP) were also used, respectively, to demonstrate the rationality of the proposed weight-determining method. The results show that: (1) the average error rates of training and testing based on RF are 18.12% and 15.83%, respectively, suggesting that the RF model can be considered rational and credible; (2) RF ranks the indices elevation (EL), slope (SL), maximum one-day precipitation (M1DP) and distance to fault (DF) as the Top 4 most important of the eight indices, occupying 73.24% of the total, while the indices runoff coefficient (RC), normalized difference vegetation index (NDVI), shear resistance capacity (SRC) and available water capacity (AWC) are less consequential, with an index importance degree of only 26.76% of the total; and (3) the verification of landslide susceptibility indicates that the accuracy rate based on the RF weight reaches 75.41% but are only 59.02% and 72.13% for the other two weights (EW and AHP), respectively. This paper shows the potential to provide a new weight-determining method for landslide susceptibility assessment. Evaluation results are expected to provide a reference for landslide management, prevention and reduction in the studied basin.

Keywords: landslide susceptibility; index weight; random forest; weighted linear combination; geographical information system; the Beijiang River Basin

1. Introduction

The occurrence frequency of natural disaster has increased in recent decades on the background of global warming [1–9]. Rainfall-induced landslides are considered one of the most common natural disasters resulting in significant economic damage and devastating loss of life [10,11].



Large-scale landslide occurrences are estimated to have led to at least 60,000 deaths with losses of more than US \$9.7 billion worldwide from 1900 to 2016 [12]. Changing climatic patterns and increased anthropogenic activities (e.g., deforestation, land reclamation, slope excavation and reservoir construction) in mountainous regions have contributed to a global increase in the occurrence of landslide events [13–16]. Defining optimum preventive and palliative measures for appropriate landslide defense and management is essential within this context as landslide-induced losses may be reduced by nearly 90% at an estimated cost of 10.3% of the potential losses [17].

The occurrence of landslide is regarded as a comprehensive result of many determinants such as precipitation, topography, morphology, lithology and land-use type [18]. The exact location of such geological disaster implies all varieties of information of hazard inducing environment factors [19]. Therefore, landslides are not reciprocally irrelevant events; there are some correlations between hazard inducing environment factors and location in a certain region. The occurrence probability can be expected if such correlations are properly revealed and estimated. Landslide susceptibility assessment, one of most important measures analyzing the correlations, becomes a vital parameter for landslide early warning systems and is a necessary component of natural and urban planning for government policies worldwide [20–26]. Benefitting from development of computer technique, the convenience in application and compatibility of geographical information systems (GIS), numerous assessment methods have been applied to evaluate the landslide susceptibility. These methods can be generally categorized into two groups. The first is a deterministic or engineering approach based on mathematical models of the physical mechanisms that control slope failure, e.g., TRIGRS [27,28]. The significant limitation of this kind method is the requirement for material data (mechanical properties, water saturation, etc.) that are difficult to obtain over large areas [29]. The second general approach is statistical and thus does not posit mechanisms that control slope failure, but assumes rather that occurrences of past landslides can be related arbitrarily to measurable characteristics of the landscape [30–32]. In turn, these characteristics can be used to predict future landslide occurrence and then many common algorithms were applied including weighted linear combination (WLC), multiple regression model [33–35], artificial neural network model [36,37], and support vector machine [38,39]. All these statistical methods could properly present the probability distribution at spatial scale and show a prefect effect in practice. Among these methods, WLC, first introduced by Voogd [40], has been intensively applied, benefitting from high precision, easy comprehension, simple use and convenience when combining with GIS [41–43]. However, the determination of a suitable index weight is a significant step when applying the WLC method because a group of suitable weights helps to better and more sensitively assess the susceptibility level. Generally, subjective weight (SW) and objective weight (OW) are two main weight-determining methods used in the evaluation system [44]. SW is typically determined by the decision maker's intentions and strongly affected by expert knowledge and biases, resulting in high subjectivity [45,46]. For example, analytic hierarchy process, a method of quantitative and qualitative analysis, is able to determine a comprehensive weight by expert score; however, such weight may not be proper if the experts lack enough experience or neglect some implicit information. A suitable index weight in landslide susceptibility assessment should objectively reflect each index's real contribution/importance and should not be affected by the decision maker's intentions when considering the objective existence of a landslide event. In this case, OW is regarded as a more suitable weight than SW. Deficiencies are still featured in currently common OW methods, which include entropy theory [47,48], technique for order preference by similarity to ideal solution method (TOPSIS) [49,50], gray relational analysis (GRA) [51] and the criteria importance though intercriteria correlation method (CRITIC) [52,53]. These traditional methods could fetch objective information of sample data using self-contained mathematical theory and analysis; however, they depend on sample data excessively and are easy to get disturbed by data fluctuation, resulting in many deficiencies including complicated calculations, poor relevance and even overlooking practical situations [54].

Random forest (RF) is a machine-learning algorithm proposed by Breiman in 2001 that provides estimates regarding hierarchy of variables in classification and evaluation and features the capability of estimating index importance to total susceptibility level [55]. The method has been applied to fields including genomic ranking [56], neuroscience prediction [57], T-cell epitope classification [58], soil parent material mapping [59], vegetable oil analysis [60], and flood hazard risk assessment [61]. Theoretical and empirical studies have demonstrated that RF may perform classification work effectively and quantitatively give objective estimates of what variables are important in the classification. The quantitative estimate of variable importance is consistent to the idea of index weight, implying that the OW could, in theory, be computed using the importance of the variables predicted by RF. However, no study has focused on determining OW utilizing RF in the field of landslide susceptibility assessment. Therefore, this study aims to apply this novel OW (i.e., weight determined by RF) in the field of landslide susceptibility assessment.

Primary objectives of this study were to: (1) adopt the Beijiang River Basin where located in humid region in Southern China as a case study and construct a landslide susceptibility assessment model utilizing the WLC; and (2) demonstrate that RF can estimate an objective and suitable index weight at basin scale. The study was intended to provide a scientific reference for index weight calculation, landslide prediction, warning, and management, as well as for soil and water conservation planning in the studied basin.

2. Methodology

Taking the Beijiang River Basin as a study case, we first selected 11 indices closely related to landslide and determined 181 rainfall-induced landslide spots. We then divided these spots into training dataset and validation dataset. The RF algorithm was executed to compute the weight of indices and the results should pass the five-fold cross validation. Afterwards, the landslide susceptibility was assessed by combining the RF weight and weighted linear combination method. As a comparison, the entropy weight (EW) and weight determined by analytic hierarchy process (AHP) were also used to further demonstrate the rationality of RF weight.

2.1. Weighted Linear Combination

Weighted linear combination (WLC), the best known and most commonly used multi criteria-GIS method [40], was applied to calculate landslide susceptibility in this study. The WLC method is a simple but effective method where susceptibility indices affecting a landslide may be combined by applying weights [62]. Assuming there are m indices and weights in the assessment system, the calculation formula of WLC is as follows:

$$y = \sum_{j=1}^{m} w_j x_j \tag{1}$$

where *y* is the comprehensive landslide susceptibility value; w_j is the weight of the *j*th index with a range of 0 to 1 and meets the condition of $\sum_{j=1}^{m} w_j = 1$ (j = 1, 2, ..., m); and x_j is the normalized value of the *j*th susceptibility index that may be calculated in the following formulas:

$$x_j = \frac{x - x_{min}}{x_{max} - x_{min}} \text{ or } x_j = \frac{x_{max} - x}{x_{max} - x_{min}}$$
(2)

where *x* is the raw value of the susceptibility index, and x_{min} and x_{max} are the minimum and maximum values, respectively. The former formula is available for the positive indices, as, the larger the value is, the greater the occurrence probability of a landslide. The latter formula is available for the negative indices, as, the larger the value is, the smaller the occurrence probability of a landslide.

2.2. Weight Definition

2.2.1. Random Forest Weight

Suitable weights greatly improve the accuracy and quality of landslide susceptibility assessment. This study utilized a random forest (RF) to determine the index weight. An RF is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \Theta_k), k = 1, ...\}$, where $\{\Theta_k\}$ are independent, identically distributed random vectors, and each decision tree (DT) casts a unit vote for the most popular class at input *x* [55]. Multiple samples are drawn in a RF utilizing the resampling bootstrap method, and classification and regression trees (CARTs) are built corresponding to each bootstrap sample (Figure 1).



Figure 1. Random forest operating principle.

Classification and regression trees (CARTs) (Figure 2), consist of root node (t_1), internal node (t_i , i = 1, 2, 3 and 4) and leaf node (N_T). The root node is split into two internal nodes according to a certain split standard when the tree begins to grow. The internal node then becomes root node and is split again and the splitting process repeats constantly until the terminal leaf node generates. If there are *M* input variables (i.e., susceptibility indices in this study), a number $m \ll M$ is specified so that, at each node, *m* variables are selected at random out of *M*, and the best split of these *m* is applied to split the node. The value of *m* remains constant during the forest's growth. The minimum Gini value is the split standard of the node, with the corresponding variable considered as the optimal variable. The Gini value is calculated as follows:

$$Gini(t) = 1 - \sum_{j=1}^{k} [p(j|t)]^2$$
(3)

where p(j|t) is the probability of class *j* at node *t*. Each time a node split is made on variable *i*, the Gini impurity criterion for the two descendent nodes is less than that of the parent node, which provides Mean Gini Decrease (MGD) after each split. Combining the MGD for each individual variable over all the trees in the forest rapidly provides an importance parameter named Gini importance that is typically consistent with the permutation importance measure [60,61]. Thus, this study proposes the random forest weight (RFW) as:

$$w_i = \frac{D_i}{\sum_{i=1}^M D_i} \tag{4}$$

where w_i and D_i are the *i*th variable weight and MGD value, respectively. The RFW equation is therefore based on MGD without involving subjective factors. The RFW equation measures the importance of the variables and is available for providing reasonable weights for the WLC.



Figure 2. Structure chart of classification and regression trees.

2.2.2. Entropy Weight

Entropy was utilized as a commonly applied method of OW to compare with the RFW. The concept of entropy, as a parameter measuring the degree of disorder or randomness, originates from thermodynamics and represents heat energy that cannot be utilized to generate work [63]. Entropy was first applied to the information theory in 1948 by Shannon, which became the measurement of ordering of one system [64]. Entropy weight (EW) is based on the information entropy theory and reflects useful information content offered by each variable [44,47].

A judgement matrix Y with *m* evaluation objects and *n* variables is constructed for the calculation of EW as:

$$Y = (y_{ij})_{m \times n} \ (i = 1, 2, \dots, n; j = 1, 2, \dots, m)$$
(5)

The influence of variable dimension and numerical range is eliminated when Y is normalized to a standard matrix $X = (x_{ij})_{m \times n}$ (i = 1, 2, ..., n; j = 1, 2, ..., m) by Equation (2). According to information theory, the variable's entropy value H_i is calculated as:

$$H_i = -\frac{1}{\ln m} \sum_{j=1}^m f_{ij} \ln f_{ij} \ (i = 1, 2, \dots, n; j = 1, 2, \dots, m)$$
(6)

where $f_{ij} = \frac{x_{ij}}{\sum_{j=1}^{m} x_{ij}}$ and $0 \le H_i \le 1$. The EW may then be computed as:

$$w_{i} = \frac{1 - H_{i}}{n - \sum_{i=1}^{n} H_{i}}$$
(7)

where the EW should meet the condition $\sum_{i=1}^{n} w_i = 1$. A smaller entropy value obviously relates to a larger EW, indicating the variable is more crucial.

2.2.3. Analytic Hierarchy Process

The weight determined by Analytic hierarchy process (AHP) was also utilized as a comparison. AHP is regarded as an ideal SW method featuring efficient and flexible framework based on psychology and mathematics. Its multi-criteria decision-making technique provides a systematic approach for assessing and integrating the effects of various factors, involving several levels of dependent or independent qualitative and quantitative information [65,66].

By analyzing the relations among indices, this method builds a hierarchical organization, including goal, criterion and sub-criterion levels, to objectively form a multi-level analysis model. The goal level is a problem's objective, and the criterion level includes factors which have influence on the objective decision. The sub-criterion level contains indices subordinated to those belonging to the criterion level. Judgment matrices are established, and a weight vector is determined according to these matrices.

3. Case Study

3.1. Study Area

The Beijiang River is the second largest river in the Pearl River system located in the Guangdong Province, China [67], and is approximately 582 km long with a drainage area of 46,649 km² (Figure 3). The Beijiang River Basin predominantly constitutes two cities and sixteen counties and features a subtropical monsoon climate with a multi-year average precipitation of 1800 mm. The flood season of the basin is from April to September, and the dry season is from October to March of the next year [68]. Approximately 70-80% of the annual rainfall is concentrated in the flood season with the rainfall fastigium from May to July [69]. Main soil type of the basin is red soil, typical to hilly topographical areas of south China, and converted to a soft soil once infiltrated by rain. Geological structure of the basin includes 68 lithology types and a substantial number of small faults in the middle and upper reaches, composing a complicated and adverse geological environment with potential for geological instability [70]. The characteristics of high-intensity rainfall, poor agrotype, complicated landform, and complicated and adverse geological environment in the basin can lead to a high probability of landslide hazard. Examples of occurrences in the area include a serious landslide in Qingxin County resulting from continuously heavy rain in March 2012, causing 7 deaths and 1 injury; four people of Nanxiong County were buried by a rainstorm-triggered landslide in May 2013; a landslide in Huaiji County caused by continuous heavy rain then killed 2 villagers and injured 3 children in May 2014; etc. The accidents suggest the Beijiang River Basin is facing great challenges in prevention and reduction of landslides. Taken together, the studied area is considered as a typical case for landslide susceptibility assessment.



Figure 3. Location and topographical condition map of the study area.

3.2. Data and Pre-Processing

The index selection varies among study areas according to the specific characteristics of each location [71]. One index can have significant impacts on the landslide susceptibility in a specific area but may have a limited influence in another area. First, we selected 11 indices representing the conditions of rainfall, topography, geology, and human activity; afterwards, we estimated these indices by the RF methods and then abandoned three indices (i.e., slope aspect, topographic wetness index and distance from stream) featuring smaller rate of Gini importance (less than 2%) for the purpose of convenient calculation and redundancy elimination [72,73]. The remaining eight indices are as follows:

The maximum one-day precipitation: Intensive rainfall acted as a trigger factor causing most landslide events in the study basin [74–76]. Short duration precipitation exerts greater influence in the studied area on landslide formation and development than average yearly or monthly rainfall [77,78]. The maximum one-day precipitation (M1DP) was selected finally among the maximum 6 h, 12 h, one-day and three-day precipitation because we found most of the historical landslides occurred in the study area after a consecutive one-day rainstorm [79]. Precipitation data (1961–2005) were provided by 48 rainfall observation stations scattered across the Beijiang River Basin and were accessed from the Hydrology Bureau of Guangdong Province (http://www.gdsw.gov.cn/wcm/gdsw/index.html). Kriging interpolation was then employed to generate the layer based on the rainfall observation stations.

Elevation (EL, m): Most landslides occur in mountainous areas with a large drop with elevation reflecting characteristics of a discontinuous terrain [80–82]. Digital elevation model (DEM—30 m) was utilized to represent the elevation index. The range of DEM is 48–1871 m, with an average elevation of 365.25 m in the study basin. Mountainous areas are typically located in the northern basin, whereas the southern basin features lower elevations. The DEM dataset was provided by Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences (http://www.gscloud.cn).

Slope angle (SL, degree): Slope angle is frequently applied as an index reflecting the degree of topographic change in landslide susceptibility studies as landslides are directly related to slope angle [78–82]. SL was generated by DEM using the "Slope" tool of Arc.GIS9.3 and it meets (Degree of slope = θ , tan $\theta = rise/run$). Areas with steep slopes feature high occurrence probabilities for landslides. The range of SL in the Beijiang River Basin is from 0° to 71.5°, with an average slope of 11.8°, and steep slopes mainly located in the central basin.

Normalized difference vegetation index (NDVI): The condition of vegetation cover is represented by this index. A large NDVI value indicates the area is comprised of luxuriant vegetation, providing a well-developed root system to maintain and stabilize soils. Areas with high vegetation cover are then generally safer than are bare areas. Average NDVI value in the Beijiang River Basin is 0.50, suggesting that vegetation cover is at a moderate level. This index was calculated for each Landsat 5 TM image data. Landsat data in 2005 acquired from the USGS Global Visualization Viewer was terrain-, radiometrically-, and geographically-corrected, and formatted to fit in an 8-bit number (ranging from 0–255). NDVI is expressed by NDVI = (band 4 – band 3)/(band 4 + band 3) where band 4 and 3 represent near-infrared band and infrared band, respectively, with a spatial resolution of 30 m \times 30 m. Use the Raster Calculator tool in the Spatial Analyst toolbar to perform the calculation.

Distance to fault (DF, m): The geological fault areas are highly susceptible to landslides because the surrounding rock strength decreases due to tectonic breaks [83]. DF is utilized in this study to reflect the degree of landslide susceptibility, thus the closer to the fault, the more dangerous exists [78]. Fault data (1:250,000) were obtained from the National Geological Archives of China (http://www.ngac.org.cn).

Shear resistance capacity (SRC, MPa): Lithology is an important index for the susceptibility assessment [80]. Lithological variations often result in strength and permeability differences in rocks and soils, significantly affecting the occurrence of landslides. Thus, this research used SRC to quantify the lithology. A large SRC value indicates that lithology can withstand a large collapsing force. A total of 68 lithology types exist in the study basin and each type was assigned a SRC comprehensive value according to the *Design code for engineered slopes in water resources and hydropower projects of China*

(*SL* 386-2007). The design code is a national normative criterion based on a significant number of tests and experiments in different areas of China, thus a SRC value is recommended for use that directly corresponds to a certain lithology type. Lithology data (1:250,000) were obtained from the National Geological Archives of China (http://www.ngac.org.cn).

Available water storage capacity (AWC): Topsoil plays a key role in the formation of landslide [84]. Soil-type data used in this study include information related to AWC, an index reflecting the maximum water amount that is held per unit of earth column. A classification value could be consulted directly from the *Harmonized World Soil Database* (2009) [85] as it provides a standard between classification value and AWC value (Table 1), thus AWC was used to represent and quantify the soil type. A large AWC value indicates soil absorbs more water with the absorption likely to weaken and break the soil structure, increasing the probability of landslides. Seven AWC measurement values were assigned to each soil type according to the *Harmonized World Soil Database* (2009) (Table 1). Soil-type data (1 km × 1 km) were obtained from the Food and Agriculture Organization of the United Nations (http://www.fao.org/home/en/).

Table 1. Available water capacity (AWC) value and classification.

Class	1	2	3	4	5	6	7
AWC (mm/m)	150	125	100	75	50	15	0

Runoff coefficient (RC): Land-cover types (LCT) are often affected by human activities, including bare land, open forest land and rural residential areas, and present high landslide potential [86,87]. A runoff coefficient (RC), measuring the runoff quantity that is converted by rainfall [61], was applied to quantify the LCT. A large RC value indicates that more rainwater is converted into surface runoff and less water infiltrates into the underground environment, significantly reducing probability of soil structure breakdown. Twenty-four land cover types exist in the study basin and were assigned corresponding RC values (Table 2) according to the *Code for Design of Building Water Supply and Drainage of China (GB 50015-2003)* and the *Code for Design of Outdoor Wastewater Engineering of China (GB 50014-2006)*. The two design codes were similar to SL 386-2007 as well as national normative criteria, thus the recommended value could be applied directly. Land-cover type data for 2005 were employed and provided by the Resources and Environment Science Data Center of the Chinese Academy of Sciences (http://www.resdc.cn/Default.aspx).

Table 2. Land-cover type and the corresponding runoff coefficient (RC).

Land-Cover Type	Runoff Coefficient	Land-Cover Type	Runoff Coefficient
Paddy field	0.98	Water body	1
Non-irrigated farmland	0.6	intertidal zone	0.4
Open forest land	0.15	Mudflat	0.5
Shrubbery	0.18	Urban land	0.9
Closed forest land	0.22	Rural residential area	0.8
High coverage grassland	0.2	Construction land	0.85
Moderate coverage grassland	0.25	Sand	0.1
Low coverage grassland	0.3	Bare land	0.7

Positive factors among the eight indices are M1DP, EL and SL, while NDVI, DF, SRC, AWC and RC constitute the negative factors. Figure 4 presents spatial distribution characteristics of the indices with all indices converted into grid format with a cell size of $30 \text{ m} \times 30 \text{ m}$ using the GIS technique and the Beijiang River Basin consisting of approximately 52 million grids. Data-processing tools included the open source software R, Arc. GIS 9.3 and MS Excel.



Figure 4. Spatial distribution characteristics of the eight susceptibility indices. Note: M1DP—maximum one-day precipitation; EL—elevation; SL—slope angle; NDVI—normalized difference vegetation index; DF—distance to fault; SRC—shear resistance capacity; AWC—available water storage capacity; RC—runoff coefficient.

3.3. Landslide Susceptibility Assessment Model

Training and validation datasets must be created prior to employing RF. Historical landslide spots were utilized as the dataset for the ability to accurately reflect the characteristic and spatial distribution of landslides. Historical landslide inventory (1995-2005) was available from comprehensive field surveys, including field evaluation, air photo/satellite image interpretations, the China Geological Environment Information Network landslide database (http://www.hbgec.org/), and news report records. Only the rainfall-induced landslide spots occurred after extreme rainfalls were considered; spots caused by artificial actions, including slope excavation, mine excavation and reservoir construction were not considered in this study. Altogether 181 landslide spots (Figure 5) distributed over the basin were finally utilized for the dataset. The five-fold cross validation and the final validation accuracy of susceptibility map are the two important criteria for dividing the sample for training and validation. Among the 181 landslide spots, a random sample of two-thirds (120) was applied to create a training dataset with the remaining (61) employed as validation data for the final susceptibility map. The 120 spots were classified as first category and marked with "1" while 120 non-landslide spots were classified as second category and marked with "0". The non-landslide spots were of the same sample size with intense human activities and no recorded landslides and were drawn randomly and uniformly to contribute to the training dataset. Samples were then created by extracting normalized values of the eight indices based on the 240 spots using the tool "Sample" of Arc.GIS 9.3. The total

240 samples, including eight normalized values (EL, SL, M1DP, DF, RC, NDVI, SRC and AWC) and a category value (0 or 1), constitute a complete training dataset.



Figure 5. Location of landslide spots and training samples.

The total 240 samples were input into the RF package of the software R to train the training data. The number of classification trees and variables attempted at each split was set to 2500 and 3, respectively, following multiple attempts. Effects of calculation occasionality were then reduced utilizing the five-fold cross validation, a common model-checking algorithm [59]. The stable and

reliable performance of the model can be checked by this validation technique. After training, the Gini decrease value of each index can be obtained, and the RFW can be calculated by Equation (4).

Normalized grid layers and the weight were then input into Equation (1) utilizing the raster calculator of GIS to calculate landslide susceptibility value and generate a susceptibility map. The landslide susceptibility map was classified into five susceptibility levels—very high, high, moderate, low and very low—by the quantile method as contained by an equal number of features. The flow chart of the assessment is shown in Figure 6.



Figure 6. The flow chart of the landslide susceptibility assessment in the Beijiang River Basin.

4. Results

4.1. Five-Fold Cross Validation

The 240 samples, the training data of five-fold cross validation in this case, were randomly divided into five sub-samples. A single sub-sample was retained as the model validation data, whereas the other four sub-samples were used to train the model. Each sub-sample was only validated once during the process of five-fold cross validation and then we can obtain five sets of results [61].

Table 3 demonstrates that the error rate of training and testing ranges 14.06–20.83% and 10.42–20.83%, respectively. The average error rates are 18.12% and 15.83%, respectively, indicating that average accuracy reaches 81.88% and 84.17%, respectively. Generally, the verification accuracies of both training and testing present stable and reliable performance, suggesting that the model can be considered rational and credible [88] and the weight calculated by RF can be used for the next step.

Fold	1	2	3	4	5	Average
Training	18.23	17.19	20.31	20.83	14.06	18.12
Testing	16.67	12.50	10.42	18.75	20.83	15.83

Table 3. Error rate (%) of five-fold cross validation.

4.2. Random Forest Weight Analysis

Five sets of Gini decrease values were obtained after the five-fold cross validation and an average RFW was calculated. Table 4 confirms that indices EL, SL, M1DP and DF are the Top 4 most-important of the eight indices, occupying 73.24% of the weight and suggesting these specific indices contribute overwhelmingly to total landslide susceptibility. EL is the most important index comprising approximately 35.00% of the total. High elevation indicates a mountainous region location as the Beijing River Basin is naturally characterized by hilly terrain, significantly increasing the probability of a landslide event. Figure 5 illustrates that most landslide spots are located in the mountainous regions of the central and northern basin, verifying that the high impact index EL, as identified by RF, bears significance in landslide susceptibility. SL is similar to EL and is considered to be the second-most-important index by RF. A large drop provides significant potential energy to cause earth-body sliding. Figures 4c and 5 also illustrate that most landslide spots are located in areas with a large SL value (substantial drop), verifying that SL also plays a vital role in landslide susceptibility. M1DP is regarded as the third-ranked index, with a percentage of 12.09%. Figure 4a demonstrates that M1DP in the south basin, especially in the southeast, is greater than in the north, suggesting the spatial variation of M1DP is quite notable, and is the primary explanation for the RF model ranking the index in third place. Certain landslide spots are in locations with relatively slight rainfall, yet the rainfall amount may be sufficient enough (minimum value still reaches 85 mm) to trigger a landslide. Many faults exist in the central basin where most landslide spots are concentrated; thus, the RF model ranks DF as the fourth-ranked index. Indices RC, NDVI, SRC and AWC are less consequential, with only a 26.76% index weight of the total.

Table 4. Index weights determined by RF, entropy and analytic hierarchy process.

Index	EL	SL	M1DP	DF	RC	NDVI	SRC	AWC
RF	0.3500	0.1564	0.1209	0.1051	0.0853	0.0820	0.0760	0.0243
EW	0.0513	0.0346	0.0945	0.0132	0.1459	0.1925	0.1365	0.3315
AHP	0.1806	0.2344	0.1705	0.0796	0.1030	0.0797	0.0541	0.0981

4.3. Spatial Distribution of Landslide Susceptibility

Landslide susceptibility map based on RF weight was generated finally. Figure 7 illustrates the high- and very-high-susceptibility areas are principally located in the central and northern basin in mountainous terrain areas; the low- and very-low-susceptibility areas are distributed in the southern and northeast basin in flat areas; and the moderate-susceptibility zones are typically located in transition areas between high and low susceptibility zones. Zone proportions of each class, from very low to very high, are 19.75%, 20.19%, 20.62%, 20.22% and 19.22%, respectively. Dangerous zones, including the high- and very-high susceptibility zones, occupy approximately 39.44%.

Sixty-one historical landslide spots, approximately one-third of the 181 landslide spots, were utilized to validate reliability of the assessment results. Table 5 demonstrates that 46 historical landslide spots (75.41%) exist in the dangerous zones, 9 spots (14.75%) in the moderate-susceptibility areas and only 5 spots (9.84%) exist in the low- and very-low-susceptibility zones. Fifteen spots (24.59%) remain in the non-dangerous zones (including the moderate-, low- and very-low-susceptibility zones) with data errors, including the historical landslide spot data and index data errors, offering a potential explanation. Some flaws may exist in the dataset of historical landslide spots, for example,

in comprehensive field surveys, locations of landslides may not always be exact, and may even be mistakenly recorded, decreasing precision of validation. For the index data errors, the index NDVI was interpreted and generated based on remote-sensing imagery; however, considerable uncertainty and subjectivity exists in the process, significantly decreasing the index precision [14]; the proportional scale of index DF is only 1:250,000, omitting smaller faults and may also decrease precision. The percentage of spots located in the Dangerous zones attained 75.41%, satisfying accuracy requirements for landslide susceptibility assessments overall.



Figure 7. Landslide susceptibility map based on RFW in the Beijiang River Basin.

Weight	Amount	Very Low	Low	Moderate	High	Very High	Dangerous ¹
RF	Num.	1	5	9	22	24	46
	Per. (%)	1.64	8.20	14.75	36.07	39.34	75.41
AHP	Num.	1	7	9	20	24	44
	Per. (%)	1.64	11.48	14.75	32.79	39.34	72.13
EW	Num.	2	14	9	18	18	36
	Per. (%)	3.28	22.95	14.75	29.51	29.51	59.02

Table 5. Number of validation spots in different susceptibility areas based on different index weights.

¹ Dangerous areas include the high- and very-high susceptibility areas.

To further verify the rationality of the susceptibility map based on RFW, we also collected 16 other landslide samples occurred after May 2005. These samples are from news report and have been verified by the open remote sensing images (Sentinel-2(ESA) and Baidu Map). As shown in Table 6, altogether 12 landslides (75%) locate in the dangerous zones while only 4 spots in the non-dangerous zones, which indicates the map still presents high applicability, even though the indices are mainly based on 2005 (e.g., NDVI and RC).

Date	County	Longitude and Latitude	Susceptibility Level	Location Attribute
20 June 2005	Fogang	113.706518, 24.035404	High	Dangerous
15 July 2006	Lechang	113.293654, 25.370019	Moderate	Non-dangerous
15 June 2008	Renhua	113.755737, 25.087949	Highest	Dangerous
6 July 2009	Renhua	113.756085, 25.085997	High	Dangerous
30 July 2009	Lechang	113.057948, 25.296335	Highest	Dangerous
10 May 2010	Wengyuan	113.840344, 24.464006	High	Dangerous
6 March 2012	Qingxin	112.734159, 23.884612	High	Dangerous
7 March 2012	Shixing	114.19523, 24.998393	Highest	Dangerous
18 August 2013	Ruyuan	113.292056, 25.018527	High	Dangerous
16 May 2013	Yingde	112.576285, 24.037433	Moderate	Non-dangerous
20 May 2014	Huaiji	112.033168, 23.553957	Low	Non-dangerous
23 January 2015	Wengyuan	113.702506, 24.547739	High	Dangerous
18 April 2016	Shixing	114.079513, 24.837004	High	Dangerous
12 August 2016	Guangning	114.079513, 24.837004	Moderate	Non-dangerous
12 August 2016	Yangshan	112.571464, 24.204037	Highest	Dangerous
9 June 2018	Ruyuan	113.354181, 25.0129	Highest	Dangerous

Table 6. Verification sample occurred in the Beijiang River Bain after May, 2005.

5. Discussion

To further verify the rationality of RFW, another objective weight (OW), i.e., entropy weight (EW), and a subjective weight determined by AHP were applied as comparisons. With normalized values of the eight indices, the 120 spots that classified as first category and marked with "1" were applied to calculate the EW, while the 120 non-landslide spots classified as second category and marked with "0" did not add to the calculation because the entropy method could not differentiate the landslide and the non-landslide spot data if mixed together. Table 4 shows the EW considers AWC and SL as the most and least critical indexes with values of 0.3315 and 0.0346, respectively. For the weight of AHP, a two-level analysis model with one criterion level (eight indices) and one goal level (weight) was constructed and ten experienced experts were invited to score. The average weight of the ten experts was determined as a final weight (Table 4) featuring SL (0.2344) and SRC (0.0541) as the most and least important indexes, respectively. Overall, the weights based on the tree methods vary considerably.

The landslide susceptibility maps based on EW and weight of AHP were generated and five categories were classified by the quantile method. Except the very-high-susceptibility areas, spatial distributions of the other four categories are visually different among the three maps (Figures 7 and 8). Among the three maps, approximately 53.92% areas have the same susceptibility level between weight of AHP and RFW while only occupying 29.97% between EW and RFW. The 61 historical landslide spots were also employed to validate susceptibility of the two weights. Approximately 44 (AHP, 72.13%) and 36 (EW, 59.02%) historical landslide spots, respectively, locate in the dangerous zones, which are fewer than those found with RFW. Different index weights then were typically observed to produce large differences among the three landslide susceptibility maps.

The EW belongs to OW and its disadvantages, i.e. it only reflects the data law of landslide spots and it fails to reflect the information of non-landslide spots, are obvious. In this case, the EW has difficulty measuring the internal law between landslide and non-landslide and thus results in relatively worse performance. In this case, the AHP has good performance, featuring a total of 44 verification spots (72.13%) located in the dangerous zones, implying the experts of this case were well experienced and grasped key points. However, this weight is experience-dependent and strongly determined by the decision maker's intentions, which means the more experience and information the experts have, the more reasonable is the weight that will be obtained. Conversely, a ridiculous weight may be obtained if the expert's experience and level of understanding could not meet the requirements.



Figure 8. Landslide susceptibility maps based on entropy weight and weight determined by AHP.

The Random Forest method is an efficient and straightforward classifier applied in this study to calculate index weights to provide a new reference for the weighted method. Unlike the common OW and SW, the RF method is able to deal with multi-category samples and is able to identify the internal law between landslide and non-landslide spots and provide contribution rate of each index to susceptibility. This function provides the possibility for determining a comprehensive weight that reflects the real contribution of each index. Additionally, the RF method could be expediently implemented by the open source software R. Only two key parameters are required for proofreading during the assessment, the number of classification trees and the number of variables tried at each split, unlike other machine learning algorithms [61]. The application effect is satisfactory, most importantly, due to the high validation precision. This method is a novel approach for landslide susceptibility assessments; however, certain issues remain. The weight calculation of this study requires many historical landslide sites, for example, and increasing the number of spots would significantly improve accuracy of the results as data limitations restricted the spots to only 181. Some of the indices may not provide accurate enough data, for example, proportional scale of index DF is only 1:250,000 and the accuracy may be improved if using a more accurate DF. Although the quantile method was used to classify into five susceptibility levels, whether there is a better way to reduce the error rate (24.59% in this study) is still worthy of research. Additionally, we only evaluated the basin with drainage area of 46,649 km², whether it would be more effective for the study area with a larger or smaller spatial scale requires further discussion.

The application of RF for weight determination demonstrates significant potential in this study, despite a few drawbacks. RFW is then recommended for use in the field of landslide susceptibility assessment and other fields dealing with hazard assessment.

6. Conclusions

Landslide susceptibility assessment is an appropriate approach for predicting and analyzing the spatial distribution of susceptibility. Determination of a suitable index weight is a key step for assessment results accuracy, thus a new weight-determining method based on random forest (RF) was proposed in this study. Eight indices were utilized to construct the susceptibility index system utilizing the Beijing River Basin as a case study. In total, 240 training samples, including 120 landslide spots and 120 non-landslide spots, were utilized to calculate the weight based on RF. Landslide susceptibility

was calculated by the weighted linear combination (WLC) method employing the RF weight as an index weight. EW and weight determined by AHP were also applied for comparisons to demonstrate the reasonability and feasibility of the RFW. Results indicate that: (1) Average training and testing error rates of the 240 samples are 18.12% and 15.83%, respectively, suggesting that the RF model can be considered rational and credible. (2) The RF model ranks EL, SL, M1DP and DF as the Top 4 most critical of the eight indices, occupying 73.24% of the total weight, while the indices, RC, NDVI, SRC and AWC are less consequential, with an index importance degree of only 26.76% of the total. (3) The landslide susceptibility map based on RFW was exceptionally different from the maps based on EW and weight of AHP; a total of 46 spots among the 61 validation spots are located in dangerous areas based on RF weight with the accuracy rate reaching 75.41%; however, only 59.02% and 72.13% of the spots are in the dangerous areas based on the other two weights, respectively. Sixteen other landslide samples occurred after May 2005, further verifying the rationality of the susceptibility map based on RFW. The proposed weighted method could be expediently implemented with few parameters while producing a satisfactory practical application. Application of the weight based on RF to landslide susceptibility assessment provides a scientific reference for weight definition and reveals significant potential.

Author Contributions: P.W. and X.B. conceived and designed the research; P.W. performed the data analysis and wrote the manuscript; Y.H. proposed some constructive suggestions about the research; X.W. helped collect and analyze the data; and H.Y. and B.X.H. assisted in revising the R1 version. All authors read and approved the manuscript.

Funding: This research was funded by the National Key R&D Program of China (2017YFC1502706), the National Natural Science Foundation of China (Grant Nos. 51709127 and 51509040), the Natural Science Foundation of Guangdong Province, China (Grant No. 2017A030310172), the Beijing Municipal Natural Science Foundation (8181001), the Open Project Program of Chongqing Key Laboratory of Karst Environment (Grant No. Cqk 201702), and Young Scientists Academic Innovation Project of Hainan Association for Science and Technology (HAST201629).

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

References

- Cai, W.J.; Santoso, A.; Wang, G.J.; Weller, E.; Wu, L.X.; Ashok, K.; Masumoto, Y.; Yamagata, T. Increased frequency of extreme Indian Ocean Dipole events due to greenhouse warming. *Nature* 2014, *510*, 254. [CrossRef] [PubMed]
- Wang, Z.L.; Li, J.; Lai, C.G.; Zeng, Z.Y.; Zhong, R.D.; Chen, X.H.; Zhou, X.W.; Wang, M.Y. Does drought in China show a significant decreasing trend from 1961 to 2009? *Sci. Total Environ.* 2016, 579, 314–324. [CrossRef] [PubMed]
- Masselink, G.; Scott, T.; Poate, T.; Russell, P.; Davidson, M.; Conley, D. The extreme 2013/2014 winter storms: Hydrodynamic forcing and coastal response along the southwest coast of England. *Earth Surf. Process. Landf.* 2016, 41, 378–391. [CrossRef]
- 4. Chen, W.; Pourghasemi, H.R.; Kornejady, A.; Zhang, N. Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma* **2017**, *305*, 314–327. [CrossRef]
- 5. Blenkinsop, S.; Lewis, E.; Chan, S.C.; Fowler, H.J. Quality-control of an hourly rainfall dataset and climatology of extremes for the UK. *Int. J. Climatol.* **2017**, *37*, 722–740. [CrossRef] [PubMed]
- 6. Wang, Z.L.; Li, J.; Lai, C.G.; Wang, R.Y.; Chen, X.H.; Lian, Y.Q. Drying tendency dominating the global grain production area. *Glob. Food Secur.* **2018**, *16*, 138–149. [CrossRef]
- Toride, K.; Cawthorne, D.L.; Ishida, K.; Kavvas, M.L.; Anderson, M.L. Long-term trend analysis on total and extreme precipitation over Shasta Dam watershed. *Sci. Total Environ.* 2018, 626, 244–254. [CrossRef] [PubMed]
- 8. Tian, Y.; Xu, Y.P.; Wang, G.Q. Agricultural drought prediction using climate indices based on Support Vector Regression in Xiangjiang River basin. *Sci. Total Environ.* **2018**, 622, 710–720. [CrossRef] [PubMed]
- 9. Wang, Z.L.; Zhong, R.D.; Lai, C.G.; Zeng, Z.Y.; Lian, Y.Q.; Bai, X.Y. Climate change enhances the severity and variability of drought in the Pearl River Basin in South China in the 21st century. *Agric. For. Meteorol.* **2018**, 249, 149–162. [CrossRef]

- 10. Keefer, M.C.; Larsen, M.C. Assessing Landslide Hazards. Science 2007, 316, 1136–1138. [CrossRef] [PubMed]
- 11. Witze, A. Mappers rush to pinpoint landslide risk in Nepal. Nature 2015, 521, 133–134. [CrossRef] [PubMed]
- 12. EM-DAT. Disaster Profiles. The OFDA/CRED International Disaster Database. Available online: http://www.emdat.be/database (accessed on 19 December 2016).
- Schuster, R.L. Socioeconomic significance of landslides. In *Landslides: Investigation and Mitigation,* Special Report 247; Turner, A.K., Schuster, R.L., Eds.; Transportation Research Board, National Research Council, National Academy Press: Washington, DC, USA, 1996; pp. 12–35.
- 14. Liu, X.P.; Li, X.; Liu, L.; He, J.Q.; Ai, B. An Innovative Method to Classify Remote Sensing Images Using Ant Colony Optimization. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 4198–4208. [CrossRef]
- 15. Goetz, J.N.; Guthrie, R.H.; Brenning, A. Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology* **2011**, *129*, 376–386. [CrossRef]
- Wang, Z.L.; Zeng, Z.Y.; Lai, C.G.; Lin, W.X.; Wu, X.S.; Chen, X.H. A regional frequency analysis of precipitation extremes in Mainland China with fuzzy c-means and L-moments approaches. *Int. J. Climatol.* 2017, 37, 429–444. [CrossRef]
- 17. Jimenez-Peralvarez, J.D.; Irigaray, C.; El Hamdouni, R.; Chacon, J. Landslide-susceptibility mapping in a semi-arid mountain environment: An example from the southern slopes of Sierra Nevada (Granada, Spain). *Bull. Eng. Geol. Environ.* **2011**, *70*, 265–277. [CrossRef]
- 18. Brabb, E.E. Innovative approaches to landslide hazard mapping. In Proceedings of the 4th International Symposium on Landslides, Toronto, ON, Canada, 16–21 September 1984; Volume 1, pp. 307–324.
- Reichenbach, P.; Busca, C.; Mondini, A.C.; Rossi, M. The Influence of Land Use Change on Landslide Susceptibility Zonation: The Briga Catchment Test Site (Messina, Italy). *Environ. Manag.* 2014, 54, 1372–1384. [CrossRef] [PubMed]
- 20. Ließ, M.; Glaser, B.; Huwe, B. Functional soil-landscape modelling to estimate slope stability in a steep Andean mountain forest region. *Geomorphology* **2011**, *299*, 287–299. [CrossRef]
- 21. Carrara, A.; Giovanni, C.; Frattini, P. Geomorphological and historical data in assessing landslide hazard. *Earth Surf. Process. Landf.* **2003**, *28*, 1125–1142. [CrossRef]
- 22. Balteanu, D.; Chendes, V.; Sima, M.; Enciu, P. A country-wide spatial assessment of landslide susceptibility in Romania. *Geomorphology* **2010**, *124*, 102–112. [CrossRef]
- 23. Thouret, J.C.; Enjolras, G.; Martelli, K.; Santoni, O.; Luque, J.A.; Nagata, M.; Arguedas, A.; Macedo, L. Combining criteria for delineating lahar- and flash-flood-prone hazard and risk zones for the city of Arequipa, Peru. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 339–360. [CrossRef]
- 24. Catani, F.; Lagomarsino, D.; Segoni, S.; Tofani, V. Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2815–2831. [CrossRef]
- 25. Roodposhti, M.S.; Rahimi, S.; Beglou, M.J. PROMETHEE II and fuzzy AHP: An enhanced GIS-based landslide susceptibility mapping. *Nat. Hazards* **2014**, *73*, 77–95. [CrossRef]
- Imaizumi, F.; Sidle, R.C.; Togari-Ohta, A.; Shimamura, M. Temporal and spatial variation of infilling processes in a landslide scar in a steep mountainous region, Japan. *Earth Surf. Process. Landf.* 2015, 40, 642–653. [CrossRef]
- 27. Ciurleo, M.; Cascini, L.; Calvello, M. A comparison of statistical and deterministic methods for shallow landslide susceptibility zoning in clayey soils. *Eng. Geol.* **2017**, 223, 71–81. [CrossRef]
- Schiliro, L.; Montrasio, L.; Mugnozza, G.S. Prediction of shallow landslide occurrence: Validation of a physically-based approach through a real case study. *Sci. Total Environ.* 2016, 569, 134–144. [CrossRef] [PubMed]
- 29. Vieira, B.C.; Fernandes, N.F.; Augusto, O.; Martins, T.D.; Montgomery, D.R. Assessing shallow landslide hazards using the TRIGRS and SHALSTAB models, Serra do Mar, Brazil. *Environ. Earth Sci.* **2018**, 77. [CrossRef]
- 30. Ohlmacher, G.C.; Davis, J.C. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Eng. Geol.* **2003**, *3*–4, 331–343. [CrossRef]
- Parise, M.; Jibson, R.W. A seismic landslide susceptibility rating of geologic units based on analysis of characteristics of landslides triggered by the 17 January, 1994 Northridge, California earthquake. *Eng. Geol.* 2000, 3–4, 251–270. [CrossRef]
- 32. Lee, S.; Choi, U.C. Development of GIS-based geological hazard information system and its application for landslide analysis in Korea. *Geosci. J.* **2003**, *7*, 243–252. [CrossRef]

- 33. Lee, S.; Pradhan, B. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* **2007**, *4*, 33–41. [CrossRef]
- Oliveira, S.C.; Zezere, J.L.; Lajas, S.; Melo, R. Combination of statistical and physically based methods to assess shallow slide susceptibility at the basin scale. *Nat. Hazards Earth Syst. Sci.* 2017, 17, 1091–1109. [CrossRef]
- 35. Patriche, C.V.; Pirnau, R.; Grozavu, A.; Rosca, B. A Comparative Analysis of Binary Logistic Regression and Analytical Hierarchy Process for Landslide Susceptibility Assessment in the Dobrovat River Basin, Romania. *Pedosphere* **2016**, *26*, 335–350. [CrossRef]
- Pistocchi, A.; Luzi, L.; Napolitano, P. The use of predictive modeling techniques for optimal exploitation of spatial databases: A case study in landslide hazard mapping with expert system-like methods. *Environ. Geol.* 2002, 41, 765–775. [CrossRef]
- 37. Lee, S.; Evangelista, D.G. Earthquake-induced landslide-susceptibility mapping using an artificial neural network. *Nat. Hazards Earth Syst. Sci.* **2006**, *6*, 687–695. [CrossRef]
- Marjanović, M.; Kovačević, M.; Bajat, B.; Vozenilek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* 2011, 123, 225–234. [CrossRef]
- 39. Su, C.; Wang, L.L.; Wang, X.Z.; Huang, Z.C.; Zhang, X.C. Mapping of rainfall-induced landslide susceptibility in Wencheng, China, using support vector machine. *Nat. Hazards* **2015**, *76*, 1759–1779. [CrossRef]
- 40. Voogd, H. *Multicriteria Evaluation for Urban and Regional Planning*, 1st ed.; Pion Ltd.: London, UK; Princeton University: Princeton, NJ, USA, 1983.
- 41. Jiang, H.; Eastman, J.R. Application of fuzzy measures in multi-criteria evaluation in GIS. *Int. J. Geogr. Inf. Sci.* **2000**, *14*, 173–184. [CrossRef]
- 42. Akgün, A.; Bulut, F. GIS-based landslide susceptibility for Arsin-Yomra (Trabzon, North Turkey) region. *Environ. Geol.* **2007**, *51*, 1377–1387. [CrossRef]
- 43. Sakkas, G.; Misailidis, I.; Sakellariou, N.; Kouskouna, V.; Kaviris, G. Modeling landslide susceptibility in Greece: A weighted linear combination approach using analytic hierarchical process, validated with spatial and statistical analysis. *Nat. Hazards* **2016**, *84*, 1873–1904. [CrossRef]
- 44. Lai, C.G.; Chen, X.H.; Chen, X.Y.; Wang, Z.L.; Wu, X.S.; Zhao, S.W. A fuzzy comprehensive evaluation model for flood risk based on the combination weight of game theory. *Nat. Hazards* **2015**, *77*, 1243–1259. [CrossRef]
- 45. Zou, Q.; Zhou, J.Z.; Zhou, C.; Song, L.X.; Guo, J. Comprehensive flood risk assessment based on set pair analysis variable fuzzy sets model and fuzzy AHP. *Stoch. Environ. Res. Risk Assess.* **2013**, 27, 525–546. [CrossRef]
- 46. Stefanidis, S.; Stathis, D.R. Assessment of flood hazard based on natural and anthropogenic factors using analytic hierarchy process (AHP). *Nat. Hazards* **2013**, *68*, 569–585. [CrossRef]
- 47. Jesmin, F.K.; Sharif, M.B. Weighted entropy for segmentation evaluation. Opt. Laser Technol. 2014, 57, 236–242.
- 48. Zhao, H.L.; Yao, L.H.; Mei, G.; Liu, T.Y.; Ning, Y.S. A fuzzy comprehensive evaluation method based on AHP and entropy for a landslide susceptibility map. *Entropy* **2017**, *19*, 396. [CrossRef]
- 49. Opricovic, S.; Tzeng, G.H. Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *Eur. J. Oper. Res.* **2004**, *156*, 445–455. [CrossRef]
- 50. Behzadian, M.; Otaghsara, S.K.; Yazdani, M.; Otaghsara, S.K.; Yazdani, M.; Ignatius, J. A state-of the-art survey of TOPSIS applications. *Expert Syst. Appl.* **2012**, *39*, 13051–13069. [CrossRef]
- 51. Jia, X.L.; Li, C.H.; Cai, Y.P.; Wang, X.; Sun, L. An improved method for integrated water security assessment in the Yellow River basin, China. *Stoch. Environ. Res. Risk Assess.* **2015**, *29*, 2213–2227. [CrossRef]
- 52. Diakoulaki, D.; Mavrotas, G.; Papayannakis, L. Determining objective weights in multiple criteria problems: The critic method. *Comput. Oper. Res.* **1995**, *22*, 763–770. [CrossRef]
- 53. Li, L.H.; Mo, R. Production task queue optimization based on multi-attribute evaluation for complex product assembly workshop. *PLoS ONE* **2015**, *10*, e0134343. [CrossRef] [PubMed]
- 54. Deng, Z.; Zhang, H.; Fu, Y.; Wan, L.; Lv, L. Research on intelligent expert system of green cutting process and its application. *J. Clean. Prod.* **2018**, *185*, 904–911. [CrossRef]
- 55. Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
- 56. Chen, X.; Ishwaran, H. Random forests for genomic data analysis. *Genomics* **2012**, *99*, 323–329. [CrossRef] [PubMed]
- 57. Smith, P.F.; Ganesh, S.; Liu, P. A comparison of random forest regression and multiple linear regression for prediction in neuroscience. *J. Neurosci. Methods* **2013**, 220, 85–91. [CrossRef] [PubMed]

- Huang, J.H.; Xie, H.L.; Yan, J.; Lu, H.M.; Xu, Q.S.; Liang, Y.Z. Using random forest to classify T-cell epitopes based on amino acid properties and molecular features. *Anal. Chim. Acta* 2013, *804*, 70–75. [CrossRef] [PubMed]
- 59. Heung, B.; Bulmer, C.E.; Schmidt, M.G. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. *Geoderma* **2014**, 214–215, 141–154. [CrossRef]
- 60. Ai, F.F.; Bin, J.; Zhang, Z.M. Application of random forests to select premium quality vegetable oils by their fatty acid composition. *Food Chem.* **2014**, *143*, 472–478. [CrossRef] [PubMed]
- 61. Wang, Z.L.; Lai, C.G.; Chen, X.H.; Yang, B.; Zhao, S.W.; Bai, X.Y. Flood hazard risk assessment model based on random forest. *J. Hydrol.* **2015**, *527*, 1130–1141. [CrossRef]
- 62. Eastman, R. Multi-criteria evaluation and GIS. In *Geographical Information Systems*; Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W., Eds.; Wiley: New York, NY, USA, 1999.
- 63. Li, X.G.; Wei, X.; Huang, Q. Comprehensive entropy weight observability–controllability risk analysis and its application to water resource decision-making. *Water Res.* **2012**, *38*, 573–579. [CrossRef]
- 64. Shannon, C. A mathematical theory of communication. Bell Syst. Tech. J. 1948, 5, 3–53.
- 65. Saaty, T. *The Analytic Hierarchy Process*; McGraw-Hill: New York, NY, USA, 1980.
- 66. Saaty, T. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation;* RWS Publications: Pittsburgh, PA, USA, 1988; 287p.
- 67. Lai, C.G.; Chen, X.H.; Wang, Z.L.; Wu, X.S.; Zhao, S.W.; Wu, X.Q.; Bai, W.K. Spatio-temporal variation in rainfall erosivity during 1960–2012 in the Pearl River Basin, China. *Catena* **2016**, *137*, 382–391. [CrossRef]
- 68. Wang, Z.L.; Zhong, R.D.; Lai, C.G.; Chen, J.C. Evaluation of the GPM IMERG satellite-based precipitation products and the hydrological utility. *Atmos. Res.* **2017**, *196*, 151–163. [CrossRef]
- 69. Wu, C.H.; Huang, G.R.; Yu, H.J. Prediction of extreme floods based on CMIP5 climate models: A case study in the Beijiang River basin, South China. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 1385–1399. [CrossRef]
- 70. Wang, Z.L.; Zhong, R.; Lai, C.G. Evaluation and hydrologic validation of TMPA satellite precipitation product downstream of the Pearl River Basin, China. *Hydrol. Process.* **2017**, *31*, 4169–4182. [CrossRef]
- 71. Lai, C.G.; Shao, Q.X.; Chen, X.H.; Wang, Z.L.; Zhou, X.W.; Yang, B.; Zhang, L.L. Flood risk zoning using a rule mining based on ant colony algorithm. *J. Hydrol.* **2016**, *542*, 268–280. [CrossRef]
- Stumpf, A.; Kerle, N. Object-oriented mapping of landslides using Random Forests. *Remote Sens. Environ.* 2011, 115, 2564–2577. [CrossRef]
- Chen, W.T.; Li, X.J.; Wang, Y.X.; Chen, G.; Liu, S.W. Forested landslide detection using LiDAR data and the random forest algorithm: A case study of the Three Gorges, China. *Remote Sens. Environ.* 2014, 152, 291–301. [CrossRef]
- 74. Jiao, J.J.; Wang, X.S.; Nandy, S. Confined groundwater zone and slope instability in weathered igneous rocks in Hong Kong. *Eng. Geol.* **2005**, *80*, 71–92. [CrossRef]
- Miller, S.; Brewer, T.; Harris, N. Rainfall thresholding and susceptibility assessment of rainfall-induced landslides: Application to landslide management in St Thomas, Jamaica. *Bull. Eng. Geol. Environ.* 2009, 68, 539–550. [CrossRef]
- Lei, T.C.; Huang, Y.M.; Lee, B.J.; Hsieh, M.H.; Lin, K.T. Development of an empirical model for rainfall-induced hillside vulnerability assessment: A case study on Chen-Yu-Lan watershed, Nantou, Taiwan. *Nat. Hazards* 2014, 74, 341–373. [CrossRef]
- Wang, Z.L.; Li, J.; Lai, C.G.; Huang, Z.Q.; Zhong, R.D.; Zeng, Z.Y.; Chen, X.H. Increasing drought has been observed by SPEI_pm in Southwest China during 1962–2012. *Theor. Appl. Climatol.* 2018, 133, 23–38.
 [CrossRef]
- 78. Frodella, W.; Ciampalini, A.; Bardi, F.; Salvatici, T.; Di Traglia, F.; Basile, G.; Casagli, N. A method for assessing and managing landslide residual hazard in urban areas. *Landslides* **2018**, *15*, 183–197. [CrossRef]
- Dahal, R.K.; Hasegawa, S.; Nonomura, A.; Yamanaka, M.; Dhakal, S.; Paudyal, P. Predictive modelling of rainfall-induced landslide hazard in the Lesser Himalaya of Nepal based on weights-of-evidence. *Geomorphology* 2008, 102, 496–510. [CrossRef]
- Yalcin, A.; Reis, S.; Aydinoglu, A.C.; Yomralioglu, T. A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. *Catena* 2011, *85*, 274–287. [CrossRef]

- Conoscent, C.; Ciaccio, M.; Caraballo-Arias, N.A.; Gomez-Gutierrez, A.; Rotigliano, E.; Agnesi, V. Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines: A case of the Bence River basin (Western Sicily, Italy). *Geomorphology* 2015, 242, 49–64. [CrossRef]
- 82. Lai, C.G.; Chen, X.H.; Wang, Z.L.; Xu, C.-Y.; Yang, B. Rainfall-induced landslide susceptibility assessment using random forest weight at basin scale. *Hydrol. Res.* **2017**. [CrossRef]
- Chen, W.; Xie, X.S.; Wang, J.L.; Pradhan, B.; Hong, H.Y.; Bui, D.T.; Duan, Z.; Ma, J.Q. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* 2017, 151, 147–160. [CrossRef]
- 84. Meinhardt, M.; Fink, M.; Tuenschel, H. Landslide susceptibility analysis in central Vietnam based on an incomplete landslide inventory: Comparison of a new method to calculate weighting factors by means of bivariate statistics. *Geomorphology* **2015**, *234*, 80–97. [CrossRef]
- 85. Food and Agriculture Organization of the United Nations (FAO); International Institute for Applied Systems Analysis (IIASA); International Soil Reference and Information Centre (ISRIC); Institute of Soil Science-Chinese Academy of Sciences (ISS-CAS); Joint Research Centre of the European Commission (JRC). *Harmonized World Soil Database (Version 1.1)*; FAO: Rome, Italy; IIASA: Laxenburg, Austria, 2009.
- Li, J.; Wang, Z.L.; Lai, C.G.; Wu, X.Q.; Zeng, Z.Y.; Chen, X.H.; Lian, Y.Q. Response of net primary production to land use and land cover change in mainland China since the late 1980s. *Sci. Total Environ.* 2018, 639, 237–247. [CrossRef] [PubMed]
- Wang, Z.L.; Xie, P.W.; Lai, C.G.; Chen, X.H.; Zeng, Z.Y.; Li, J. Spatiotemporal variability of reference evapotranspiration and contributing climatic factors in China during 1961–2013. *J. Hydrol.* 2017, 544, 97–108. [CrossRef]
- 88. Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).