Integration of Information Theory, Fractal Theory and Statistical Analyses for Landslide Susceptibility Mapping at Regional Scales

Xiaolong Deng (✉ 844256111@qq.com)  
zhong jiao gong lu gui hua she ji yuan you xian gong si: CCCC Highway Consultants Co Ltd

Guangji Sun  
zhong jiao gong lu gui hua she ji yuan you xian gong si: CCCC Highway Consultants Co Ltd

Naiwu He  
zhong jiao gong lu gui hua she ji yuan you xian gong si: CCCC Highway Consultants Co Ltd

Yonghua Yu  
zhong jiao gong lu gui hua she ji yuan you xian gong si: CCCC Highway Consultants Co Ltd

Research Article

Keywords: landslide susceptibility, Shannon's entropy index, variable fractal dimension method, frequency ratio, comparison and validation

Posted Date: January 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1092127/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Integration of information theory, fractal theory and statistical analyses for landslide susceptibility mapping at regional scales

Xiaolong Deng · Guangji Sun · Naiwu He · Yonghua Yu

China Highway Engineering Consulting Corporation, Beijing 10089, China

lyttney@foxmail.com
Abstract A new model, integrating information theory, fractal theory and statistical model for accurate landslide susceptibility mapping (LSM) at regional scales, has been proposed. In this model, landslide conditional factors are firstly classified with an optimal number of classes, which is determined by maximizing their information coefficients estimated from Shannon’s entropy model. The spatial association between influencing factors and induced landslides has been measured by introducing the variable fractal dimension method (VFDM). The VFDM approach fully considers the characteristics of landslide fractal distribution. Then the fractal dimensions ($D$) are calculated to provide multiple factors with various numerical weights. The proposed model eventually combines the landslide frequency ratio ($fr$) of each factor with corresponding weight to achieve spatial prediction of landslides, illustrated by an example area in China. In the study area, 500 landslides have been identified by aerial photograph interpretation, extensive field investigations, historical and bibliographical landslide data. In the model, these landslides are randomly split into a training dataset (70 %) and a validating dataset (30 %). Seven factors are recognized and analyzed by frequency ratio (FR) method, including lithology, distance to fault, altitude, slope, aspect, distance to stream and distance to the road. The receiver operating characteristic curve (AUROC) has been adopted to compare and validate the model results. Results show that the proposed landslide model achieved a more accurate prediction with AUROC equal to 0.8467, over-performing than the conventional frequency ratio method (AUROC=0.8088). According to the final prognostic landslide susceptibility map, 16.37 % of the study area shows very high and high susceptibility, accounting for 63.55 % of the entire landslides. Evaluation of relative factor importance based on a one-by-one factor removal test indicates that the lithology factor contributes unique information for landslides. In conclusion, the example demonstrates that the proposed framework is promising for further improvement of LSM.

Keywords landslide susceptibility · Shannon’s entropy index · variable fractal dimension method · frequency ratio · comparison and validation
Introduction

Landslide describes the downslope transport of soil/rock, resulting from materials different geodynamic processes (Guzzetti et al. 2005). The landslide also represents a potentially significant natural hazard, contributing to the tremendous economic and social loss. Much research efforts have been devoted to denominating potential landslide hazard zones by evaluating responsible factors (Shahabi and Hashim 2015). Various methods being successfully applied in susceptibility assessment can be classified into qualitative (e.g., Regmi et al. 2010; Yalcin et al. 2011) and quantitative (e.g., Pradhan et al. 2011; Goetz et al. 2015) groups. Among quantitative approaches, statistical methods are favored in most publications (Chung and Fabbri 1999; Lee and Pradhan 2006). They were proven to be better options for comparatively large and complex areas (Cardinali et al. 2002). In addition, the recent development in machine learning and data mining have introduced new approaches for landslide susceptibility modeling such as support vector machine (Tien Bui et al. 2012), random forests (Catani et al. 2013) and decision tree (Pradhan 2013), some of these methods were reported to outperform conventional ones (Tien Bui et al. 2013). These approaches work well when handling large amounts of data. However, the model may not work well if considering all possible influencing factors, since the problem of over-fitting may occur. Screening these factors could decrease this phenomenon, but this could result in high computational cost and would be a time-consuming process (Wang et al. 2017).

Except for analyzing the physical properties of slope materials, the basic idea of landslide susceptibility assessment is to identify possible influencing factors that control landslide spatial distributions (Liucci et al. 2015). The highly heterogeneous character of landslide spatial distribution has been realized (Li et al. 2012). Literature reviews indicate that many landslides exhibit fractal distribution (Sezer 2010). Both the frequency-area statistics and frequency-size statistics satisfy the power-law
function (Guzzetti et al. 2002; Malamud et al. 2004; Guthrie et al. 2007; Trigila et al. 2010; Ghosh et al. 2012). Also, geoscientists explored the spatial association between landslides and corresponding influencing variables (Brunetti et al. 2009). Some related and significant conclusions have been drawn include: 1) fine geological features leading to a high fractal dimension of landslide distribution (Iwahashi et al. 2003), 2) demonstrated fractal relation between the cumulative frequency of landslides and their triggering rainfall level (Li et al. 2011), and 3) fractal relation between landslide density and distances to conditional factors (Zou and Carranza 2017). Thus, the landslide is a nonlinear system and its spatial distribution is heterogeneous (Rouai and Jaaidi 2003). However, a retrospective of the previous studies showed that the fractal analyses used to characterize landslide spatial distribution are mostly based on the box-counting fractal dimension (BCFD) algorithms (e.g., Zou et al. 2009; Li et al. 2012), which does not consider the variable-dimension fractal characteristics of natural hazards. Remarkable work from Liu et al. (2019) introduced a variable fractal dimension method (VFDM) to measure the spatial association between influencing factors and landslides. It concluded that the calculated fractal dimension values could be used as weights of their corresponding factors in LSM. The fractal dimension values/weights vary with different class numbers of the factors.

The objective of this study is to propose an alternative approach for landslide susceptibility mapping (LSM). The proposed new method integrates information theory, fractal theory and statistical analyses. Firstly, the optimal number of classes for the identified landslide conditional factors are determined by Shannon’s entropy index. Next, the variable fractal dimension method (VFDM) is introduced into the model, which considers the heterogeneous fractal structure of landslide spatial distribution. A fractal measure of spatial association between landslide occurrences and their conditional variables has been adduced to assign the conditional factors with numerical weights. Finally, the fractal-based weights are
developed into the frequency ratio model. The proposed model for accurate LSM will be illustrated and validated in an example area in China.

Data and materials

Study area

Located in northeast China, the study area, covering an area of about 1.17×10^5 km^2, is between longitudes 85°31′37″E ~ 91°01′15″E and latitudes 45°59′35″N ~ 49°10′45″. Altitudes in the study area range from 365 m to 4330 m above sea level (Fig. 1).

The topographic inclination follows the NE–SW direction. The terrain surface is relatively deep-inclined with slope angles ranging from 0° to 83.97°. Areas with slopes less than 15° cover 82.64 % of the total area, while areas with slopes larger than 45° account for only 2.06 %. Besides, areas with slopes between 15° and 45° account for 15.30 %.

The climate belongs to type B (Arid), as defined in the Köppen climatic classification (Köppen 1884).
According to the National Meteorological Information Center (http://data.cma.cn). The average temperature is 10.6 °C. Mean annual rainfall for the period 1998–2015 ranges between 83.9 mm and 430.3 mm. During the rainy seasons (May–August), the average rainfall varies between 224.5 mm and 418.4 mm per month.

The lithology map used in this study is constructed based on the Geological and Mineral Resources maps at 1:200,000 scale. Rock types observed include volcanic, sedimentary and metamorphic varying in age from Sinian to Quaternary. Data collected from the China Geology Survey (CGS) (http://www.cgs.gov.cn) indicate that More than 50 geologic groups and units have been recognized. All the lithology types have been classified into eight classes based on clay composition, weathering degree, estimated physical (density) and strength parameters (Fig. 3a, Table 3). Around 53.9 % of the study area is covered by metamorphic rocks (schist, gneiss, and metamorphic limestone).

Data acquisition and processing

Landslide inventory map

The unique mountain-basin landscape, complex geologic condition and frequent extreme weather are responsible for the widespread landslide distribution in this area. Landslide inventory has been performed to analyze the homogeneous population. According to statistics, over 85 % of slope movements in the study area are rockfalls. Thus, only rockfalls have been selected. A total of 600 landslides were registered. Exact occurrence dates of them are mostly unknown. According to documentary reports and publications, landslides mainly occurred during and after the heavy rainfalls. Statistics of these landslides show that the largest landslide size is approximately 39,420 m$^2$, the smallest around 792 m$^2$. The average landslide size is estimated to be 10,733 m$^2$. Fig. 2 illustrates the cumulative frequency distribution of landslides areas.
Fig. 2 Main characteristics of landslide distribution

For preparing a landslide inventory map, landslides are usually included as singular points or as polygons (Dou et al. 2015). The choice of points or polygons is comprehensively determined by the map’s scale, data availability and research objectives (Wang et al. 2014). Comparisons of these sampling strategies can be found in some published articles (e.g., Oliveira et al. 2015). However, this issue is beyond the scope of this study. In this study, the 600 landslides are depicted as points in GIS shapefile format. The landslide inventory map has been then produced as a point map (Fig. 1).

Landslide conditional factors

Seven landslide conditional factors are recognized in the study area, including lithology, distance to fault, altitude, slope, aspect, distance to stream and distance to the road. The inclusion of these factors draws heavily on extensive literature reviews, experience gained from studying landslide phenomena, and sufficient data availability. Some ancillary data have been used to extract these conditional factors, as listed in Table 1 below. These factors and the landslide inventory map have been extracted with a grid size of 30 m×30 m to match the digital elevation model (DEM).

| No. | Data | Source |
|-----|------|--------|

Table 1 Data source for the study area
Geological and Mineral Resources map at 1:200,000 scale

1. Lithology
   [http://www.cgs.gov.cn](http://www.cgs.gov.cn)

2. Distance to fault
   [http://www.cgs.gov.cn](http://www.cgs.gov.cn)

   DEM and derived factors

3. ASTER GDEM data [http://www.gscloud.cn](http://www.gscloud.cn)
   (altitude, slope, aspect)

4. Distance to stream
   River and stream network map

5. Distance to road
   Road network map

Several classification methods are available to classify the continuous factors into different intervals, including Equal Interval, Quantile, Natural Breaks ([Jenk 1963](#)), Geometrical Interval and Standard Deviation available. Among them, the Natural Breaks (NB) and Geometrical Intervals (GI) are most widely used in LSM. Choosing which method to adopt is based on the distribution type of the data. In this study, both the NB and GI classification methods have been adopted to classify the continuous conditional factors ([Table 2](#)).

**Methodology**

**Preparation of training and validating datasets**

Landslide observations are usually divided into a training dataset for model building and a validating dataset for model evaluation. Different partition techniques, including time-partition, space-partition and random-partition, might be utilized to obtain the independent dataset ([Chung and Fabbri 2003](#)). Since the specific dates of landslides are mostly unknown in this study, the landslides have been randomly split into two datasets with a 3:1 ratio. The first 450 landslide points are used for model construction, and the left 150 landslide points are used for model validation. Also, the same number of non-landslide points
are randomly sampled from the landslide-free areas.

**Determination of class numbers**

In the treatment of classifying continuous landslide conditional factors, Shannon’s entropy index has been adopted to determine optimal class numbers by maximizing the information coefficient of each factor. Shannon’s entropy index, which was initially proposed by Claude Shannon (1948) for the entropy model, has been widely used to quantify the uncertainty and disorder in Information Theory. The earth science literature also contains a number of studies using the entropy method as a tool for integrated environmental assessments of natural processes (e.g., debris flows, sandstorms) (Yang and Qiao 2009; Yang et al. 2010).

Landslides are the result of the interaction processes of materials and energies with the environment. Thus, they can be measured and characterized by the entropy method (Pourghasemi et al. 2012; Shi and Jin 2009). Shannon’s entropy index determining the optimal class numbers involves the computation of landslide density within each class per factor. The information coefficients \( I_j \) can be estimated each time for a different number of classes by selecting the maximum information coefficient (Tsangaratos et al. 2017). The calculation can be implemented using a series of formulas given below (cf., Bednarik et al. 2010; Constantin et al. 2011):

\[
p_{ij} = \frac{N_{ij}}{A_{ij}} \tag{1}
\]

\[
(p_{ij}) = \frac{p_{ij}}{\sum_{j=1}^{n_j} p_{ij}} \tag{2}
\]

\[
H_j = -\sum_{j=1}^{n_j} (p_{ij}) \log_2(p_{ij}) \tag{3}
\]

\[
H_{j\text{max}} = \log_2 n_j \tag{4}
\]

\[
I_j = \frac{H_{j\text{max}} - H_j}{H_{j\text{max}}} \tag{5}
\]

where \( N_{ij} \) and \( A_{ij} \) denote the landslide percentage and area percentage for the \( j \)-th class of \( i \)-th factor,
\( n_j \) represents the total the number of classes, \( p_{ij} \) represents the probability density, \( H_j \) and \( H_{j\text{max}} \)
represent entropy values. Since the entropy values are above or equal to 0, the information coefficient is
restricted to the domain of \([0, 1]\), with values closer to 0 indicating less information and values to 1
indicating more information extracted.

**Spatial prediction modeling of landslides**

**Frequency ratio**

The frequency ratio (FR) method has been commonly used to predict landslide spatial distribution for its
simple input process, calculations, output process and intelligible principle (Lee and Pradhan 2007; Yilmaz 2009). Specifically, FR is the ratio of the area where landslides occurred to the total studied area
(Lee and Talib 2005). It can be concluded that the item of “\( p_{ij} \)” in Eq. (1) represents the concept of
“frequency ratio” (“\( f_{rij} \)” ) when recalling the entropy model mentioned above. In bivariate statistical
analysis, larger frequency ratio values indicate a stronger correlation between landslide occurrence and
the factor’s class and vice versa (Pradhan 2010).

**Fractal dimension-weight of factors**

The process of weighting followed after the interpretation of results obtained from the frequency ratio.
This phase involves estimating landslide spatial distribution within the defined class number of each
conditional factor based on fractal dimension. The fractal dimension will be used in the Fractal Theory
as a measure, which is initially proposed by Benoit Mandelbrot to quantify the complication degree of
seacoast shapes with statistical self-similarity (Mandelbrot 1967). The fractal distribution gives the
formulas as:

\[
N(r) \propto r^{-D}
\]

(6)

where \( r \) represents a typical linear scaling such as distance, time, etc., the parameter \( D \) represents the
fractal dimension. Expression $N(r)$ is the resulting increase connected with $r$, such as stock index, price, monitoring data, etc. (Lu et al. 2012).

A constant-dimension fractal object would be obtained if the fractal dimension $D$ kept constant with various linear scaling $r$. However, such a fractal relationship does not strictly exist in nature objects (Davis 2002; Shen 2002). Thus, some sophisticated phenomena such as the geological point process (e.g., mineral deposits, geohazards) cannot be well interpreted by traditional fractal methods (Agterberg 2013; Zou 2016).

Given that landslide distribution with its conditional factors is not strictly self-similar, the fractal analysis used here is the variable fractal dimension method (VFDM) (Liu et al. 2019). Concerning the $i$-th factor, let the linear scaling $r_j$ denote the $j$-th class, $N_j$ be landslide points within the $j$-th class, $n_i$ be the number of classes determined by Shannon’s entropy index. The fractal dimension $D_i$ derived by the VFDM can be defined by four steps as follows (cf., Lu et al. 2011; Sun et al. 2016):

- **Draw the original data series** $(r_j, N_j)$ ($j = 1, 2, \ldots, n_i$) in double logarithmic coordinate. Where
  \[
  \{N_j\} = \{N_1, N_2, \ldots, N_{n_i}\}
  \]
  is the basic data sequence.

- **Construct cumulative-sum sequences with the basic data sequence** $\{N_j\}$. We define the 1-order cumulative-sum sequence $\{S1_j\}$, the 2-order cumulative-sum sequence $\{S2_j\}$, and so on, as follows:
  \[
  \begin{align*}
  \{S1_j\} &= \{N_1, N_1 + N_2, \ldots, N_1 + N_2 + \cdots + N_{n_i}\} \\
  \{S2_j\} &= \{S1_1, S1_1 + S1_2, \ldots, S1_1 + S1_2 + \cdots + S1_{n_i}\} \\
  \vdots & \\
  \{S(k+1)\} &= \{S1_1, S1_2 + Sk_2, \ldots, S1_k + Sk_k + \cdots + S1_{n_i}\} \\
  \end{align*}
  \]
  where $k = 1, 2, \ldots, n$.

- **Build the $n$-order variable-dimension fractal sequence**. Taking the 1-order cumulative-sum sequence for example, we can obtain the segmental fractal dimension $D_{1,j,j+1}$ from the slope of double logarithmic coordinate by data points $(r_j, S1_j)$ and $(r_{j+1}, S1_{j+1})$. For $n_i$ data pairs, $n_i - 1$
segmental fractal dimension can be obtained. Therefore, the 1-order segmental fractal dimension sequence can be arranged as:

\[
\{D_{1,j+1}\} = \{D_{1,1,2}, D_{2,3}, \ldots, D_{n_i-1,n_i}\}
\]  \(8\)

while \(D_{n_i,j+1}\) \((j = 1,2, \ldots, n_i)\) represents the segmental fractal dimension for \(n\)-order cumulative-sum sequence.

- Select the optimal transform and determine the corresponding fractal dimension. By comparing every order cumulative-sum sequence, the one with the best linear fit effect is chosen. Thus, the fractal dimension \(D_i\) can be calculated as the slope of the linearly fitted line.

The fractal dimension measures the spatial associations between landslide conditional factors and landslide occurrence (Liu 2019), while the relative weights of the factors control the spatial pattern of landslide phenomenon (Liucci et al. 2015). Therefore, the fractal dimension values can be used as proxies for relative weights of the factors.

The above steps are repeated to calculate the fractal dimension \(D\) for all the conditional factors, then the weight of the \(i\)-th factor \((W_i)\) can be defined by:

\[
W_i = D_i / \sum_i D_i
\]  \(9\)

The calculated fractal dimension \(D_i\) form the \(W_i\) value in landslide susceptibility assessment. Then the fractal-based \(W_i\) is embedded into the conventional frequency ratio model. We express landslide susceptibility by employing a Calibrated Landslide Susceptibility Index (CLSI), with its equation shown as:

\[
\text{CLSI} = \sum W_i \times fr_i
\]  \(10\)

**Validation and comparison strategy**

**Receiver operating characteristic curve**
The receiver operating characteristic (ROC) curve can evaluate landslide susceptibility models (Zweig and Campbell 1993). Specifically, in each mapping unit, the stochastic methods produce an estimate of the probability (between 0 and 1) of an unstable condition. Once an optimal cutoff value has been derived, the positive prediction and the negative would be given. Correspondingly, the positive prediction refers to landslide prediction, and the negative prediction refers to stable slope prediction. Comparing the predicted with the observed status, estimating the accuracy of the model can be achieved by counting the true and false positive/negative prediction.

The ROC curve depicts the relation between true positive (TP) and false positive (FP), which is a plot of 1-Specificity (X-axis; false positive rate (FPR)) versus Sensitivity (Y-axis; true positive rate (TPR)). These statistical quality parameters are measured as below:

\[ 1 - \text{Specificity} = \text{FPR} = \frac{FP}{FP+TN} \quad (11) \]

\[ \text{Sensitivity} = \text{TPR} = \frac{TP}{TP+FN} \quad (12) \]

where TP and TN represent the number of correctly and incorrectly classified landslides, respectively. Similarly, the parameter FP and FN represent the number of correctly and incorrectly classified non-landslides. Specificity denotes the proportion of non-landslides that are correctly classified as non-landslides. Sensitivity denotes the proportion of landslides that are correctly classified as landslides (Shirzadi et al. 2017; Tien Bui et al. 2016).

The area under the ROC curve (AUROC) is used for quantitative comparison. The AUROC can be interpreted as the probability that a randomly chosen positive example (landslide) is ranked higher than a randomly chosen negative example (non-landslide) (Fawcett 2006). The AUROC can be calculated as:

\[ \text{AUROC} = \frac{\sum TP + \sum TN}{P+N} \quad (13) \]

where P/N is the total number of landslides/non-landslides. A non-informative model has an AUROC
value close to 0.5, so that the AUROC values ranges from 0.5 to 1.0 commonly. An ideal model performs an AUROC value of 1.0 (perfect prediction) (Swets 1988).

Relative importance of conditional factors

The relative importance of landslide conditional factors to landslide models is strongly connected with the modeling approaches and data sampling methods. Factors with a high contribution in a specific model might be useless for another one (Tien Bui et al. 2016). Besides, if we do this with the same dataset that has already being used in model construction (i.e., training dataset), the contribution cannot be considered as indicators for factor importance. Instead, it implies the percent of a specific factor being engaged in the modeling process (Kornejady et al. 2017). Therefore, each conditional factors’ relative importance and contribution are evaluated with one-by-one factor removal. The overall accuracy (AUROC) with the validating dataset can then be calculated. In such a context, the difference between the AUROC between these factors indicates their relative importance.

Results

Number of classes for landslide conditional factors

Following the procedure described above, the optimal number of classes for the five continuous landslide conditional factors is determined by maximizing their corresponding information coefficients, except for the two factors of lithology and aspect which were considered categorical factors. The analysis was performed from four to eight classes, using both the Geometrical Intervals (GI) and Natural Breaks (NB) classification methods. The information coefficients were calculated and listed in Table 2.

It can be observed from the table that for the distance to the fault factor, the information coefficient has the highest value of 0.1756 when classified into four classes by the GI method. Divided into five classes by the GI method, the three conditional factors of altitude, distance to stream and distance to road
maximize their information coefficients to 0.2804, 0.1985 and 0.2296. Finally, the slope factor presents
the highest information coefficient value (0.2467) with five classes classified by the GI method. The
most informative appears to be the altitude, followed by slope and distance to road factors when
comparing the information coefficients among these conditional factors. The distance to the fault factor
gives the least information.

Optimal class numbers of five continuous factors are given in Fig. 3. The two categorical factors of
lithology and aspect were also illustrated in this figure.

| Conditional factors | Information Coefficient |
|---------------------|-------------------------|
|                     | 4 Classes | 5 Classes | 6 Classes | 7 Classes | 8 Classes |
| Distance to fault   | GI 0.1756  | NB Null   | GI 0.2804 | NB 0.1695 | GI 0.2043 |
| Altitude            | 0.2443    | 0.1580    | 0.2804    | 0.1695    | 0.2043    |
| Slope               | 0.1986    | 0.1061    | 0.2467    | 0.0788    | 0.2232    |
| Distance to stream  | 0.1614    | Null 0.1985 | Null 0.1940 | Null 0.1940 | Null 0.1940 |
| Distance to road    | 0.2241    | 0.2010    | 0.2296    | 0.1334    | 0.1437    |

Note: The highest information coefficient value for each factor was indicated in bold text.

It can be concluded that the GI serves as a more suitable classification method for the current case, for
the information coefficient values derived from this method are generally higher than those derived from
the NB method. Note that “Null” represents non-calculate, indicating no landslide points in specific factor class/interval. For the factors of distance to lineaments (fault, stream, road), most of the information coefficient values were assigned as “Null” when classified by the NB method. This is mainly because the values of these factors are distributed in broad ranges, resulting in less or even none landsides in certain classes when using the NB method (Dou et al. 2015). In contrast, the GI method has the advantage of classifying not normally distributed data, especially when the data are heavily skewed (Tsagaratos et al. 2017).
Fig. 3 The classified landslide conditional factor maps

Landslide frequency ratio

After the determination of the optimal class number, the calculation of $f_{r_{ij}}$ followed. Table 3 sets out the complete calculation of $f_{r_{ij}}$ for each class of factors.

Table 3 The process of frequency ratio and information coefficient calculation

| Conditional factor class ($r_j$) | Class description                  | Landslide points ($N_j$) | $A_{ij}$  | $f_{r_{ij}}/p_{ij}$ | $I_j$  |
|--------------------------------|------------------------------------|--------------------------|-----------|---------------------|--------|
| 0                             | Waters and snow areas              | 0                        | 0.0000    | 0.0138              | 0.0000 |
| 1                             | ophiolite                          | 2                        | 0.0044    | 0.0032              | 1.3958 |
|                               | Litology                           |                          |           |                     |        |
|                               | Metamorphic rocks (schist, gneiss, metamorphosed sandstone) |                          |           |                     |        |
|                               | Intrusive volcanic rocks           |                          |           |                     |        |
| 3                             | (granite, diorite, gabbro)         | 142                      | 0.3156    | 0.1796              | 1.7575 |

Table includes class descriptions, number of landslide points, and calculations for $A_{ij}$, $f_{r_{ij}}$, and $I_j$.
| Conditional factor | Number of class \( (r_j) \) | Class description | Landslide points \( (N_j) \) | \( N_{ij} \) | \( A_{ij} \) | \( f_{r_{ij}}/p_{ij} \) | \( I_j \) |
|-------------------|---------------------------|-----------------|-----------------|-----------|----------|-----------------|-----|
| Extrusive volcanic rocks | 4 | 47 | | 0.1044 | 0.0744 | 1.4034 |
| (tuff, breccia, basalt) | | | | | | | |
| Sedimentary rocks | 5 | 5 | | 0.0111 | 0.0105 | 1.0582 |
| (sandstone, thinly bedded fine sandstone) | | | | | | | |
| Soft sedimentary rocks | 6 | 105 | | 0.2333 | 0.1823 | 1.2798 |
| (mudstone, shale, coal line) | | | | | | | |
| Quaternary deposits (sand, soil, mud, gravel) | 7 | 19 | | 0.0422 | 0.1416 | 0.2982 |
| Distance to fault | 1 | <1,895 m | 93 | | 0.2067 | 0.1709 | 1.2093 |
| fault | 2 | 1,895 m–8,571 m | 176 | | 0.3911 | 0.3531 | 1.1078 |
| | 3 | 8,571 m–32,088 m | 179 | | 0.3978 | 0.3482 | 1.1423 |
| | 4 | >32,088 m | 2 | | 0.0044 | 0.1278 | 0.0348 |
| Altitude | 1 | <687 m | 35 | | 0.0778 | 0.2734 | 0.2845 |
| | 2 | 687 m–807 m | 8 | | 0.0178 | 0.1242 | 0.1431 |
| | 3 | 807 m–1,130 m | 75 | | 0.1667 | 0.2290 | 0.7277 |
| | 4 | 1,130 m–1,997 m | 284 | | 0.6311 | 0.2239 | 2.8190 |
| | 5 | >1,997 m | 48 | | 0.1067 | 0.1495 | 0.7136 |
| Slope | 1 | <0.15° | 47 | | 0.1044 | 0.1450 | 0.7204 |
| | | | | | | | 0.2467 |
| Conditional factor | Number of class \((r_j)\) | Class description | Landslide points \((N_j)\) | \(N_{ij}\) | \(A_{ij}\) | \(fr_{ij}/p_{ij}\) | \(I_j\) |
|-------------------|-----------------|-------------------|-----------------|---|---|---|---|
| 2                 | 0.15°–0.66°     | 1                 | 0.0022          | 0.0218 | 0.1022 |
| 3                 | 0.66°–2.32°     | 31                | 0.0689          | 0.3205 | 0.2150 |
| 4                 | 2.32°–7.75°     | 66                | 0.1467          | 0.2489 | 0.5893 |
| 5                 | 7.75°–25.57°    | 179               | 0.3978          | 0.1660 | 2.3968 |
| 6                 | >25.57°         | 126               | 0.2800          | 0.0979 | 2.8589 |
| 0                 | Flat            | 0                 | 0.0000          | 0.2119 | 0.0000 |
| 1                 | North           | 19                | 0.0711          | 0.0729 | 0.9754 |
| 2                 | Northeast        | 36                | 0.0800          | 0.0715 | 1.1191 |
| 3                 | East            | 86                | 0.1911          | 0.1221 | 1.5647 |
| Aspect            | Southeast       | 49                | 0.1089          | 0.0671 | 1.6238 | 0.0646 |
| 5                 | South           | 102               | 0.2267          | 0.1474 | 1.5379 |
| 6                 | Southwest       | 45                | 0.1000          | 0.0861 | 1.1608 |
| 7                 | West            | 71                | 0.1578          | 0.1467 | 1.0757 |
| 8                 | Northwest       | 29                | 0.0644          | 0.0743 | 0.8669 |
| 1                 | <4,700 m        | 169               | 0.3756          | 0.1705 | 2.2029 |
| 2                 | 4,700 m–14,797 m| 123               | 0.2733          | 0.2628 | 1.0400 |
| 3                 | 14,797 m–36,483 m| 121              | 0.2689          | 0.2694 | 0.9981 | 0.1985 |
| 4                 | 36,483 m–83,066 m| 34               | 0.0756          | 0.1663 | 0.4543 |
| 5                 | >83,066 m       | 3                 | 0.0067          | 0.1310 | 0.0509 |
| 1                 | <1,118 m        | 190               | 0.4222          | 0.1663 | 2.5391 | 0.2296 |
| Conditional factor | Number of class \( r_j \) | Class description | Number of landslide points \( N_{ij} \) | \( A_{ij} \) | \( f_{r_{ij}}/p_{ij} \) | \( I_j \) |
|-------------------|-----------------|--------------------|-----------------|------|-----------------|------|
| Distance to road  | 2               | 1,118 m–3,414 m    | 92              | 0.2044 | 0.2736 | 0.7473 |
|                   | 3               | 3,414 m–8,128 m    | 124             | 0.2756 | 0.3139 | 0.8779 |
|                   | 4               | 8,128 m–17,808 m   | 42              | 0.0933 | 0.1823 | 0.5119 |
|                   | 5               | >1,7808 m          | 2               | 0.0044 | 0.0639 | 0.0695 |

Note: For categorical factors of lithology and aspect, the “\( p_{ij} \)” values of class number “0” were not involved in the calculation of information coefficients.

The calculated \( f_{r} \) values for each class/interval of the conditional factors were plotted in Fig. 4.

According to the thematic map of lithological units with landslides, the most affected areas are built of ophiolite and metamorphic rocks. Although they represent only 4 % of the total area of the studied territory, more than 12 % of the total landslides occur in these two categories. For the altitude factor, slope failures mainly occurred in class 4 (1,130 m–1,997 m), reaching 63 % of the total landslides. As for aspect factor, slope deformations exhibit an approximating normal distribution. The landslide inventory well represents the link between landslides and slope. It can be seen in Fig. 4 that landslides are more likely to occur with an increasing slope gradient. The most critical slopes range between 7.75°-25.57° (class 5), with more than 39 % landslides. Additionally, unlike distance to stream and distance to road factors, which landslide occurrence decreases with increasing distances, landslide occurrences are distributed evenly with varying distances to the fault.
Fractal-based weights

The measurement of spatial associations between landslides and the conditional factors using fractal dimension has been performed by the VFDM. Fig. 5 presents the linear fitting results of the relationships between the conditional factors and the transformed landslide cumulative-sum sequences. The figure shows that the paired datasets \((r_j, S_2_j)\) of six factors can be well linear fitted after the two-order cumulative-sum transform, except for the slope factor.
From the slopes of linearly fitted lines, the fractal dimension values can be derived. Table 4 provides the \( D_i \) values that are used to weight these factors. The highest \( W_i \) (0.2071) of lithology factor indicates that the dominant effect upon landslide occurrence and distribution is posed by lithology formation, followed by altitude (0.1620), slope (0.1405) and aspect (0.1324) factors. Landslide occurrence is least susceptible to the distance to stream with the lowest \( W_i \) (0.0960).

Table 4 Weights of the factors calculated from fractal dimension

| Conditional factors | Linear regression formulation | Correlation coefficient \((R^2)\) | \( D_i \) | \( W_i \) |
|---------------------|------------------------------|-------------------------------|---------|--------|
| Lithology           | \( \log S_2 = 3.1770 \log r + 0.6197 \) | 0.9549                        | 3.1770  | 0.2071 |
| Distance to fault   | \( \log S_2 = 1.9026 \log r + 1.9775 \) | 0.9984                        | 1.9026  | 0.1240 |
| Altitude            | \( \log S_2 = 2.4858 \log r + 1.4335 \) | 0.9687                        | 2.4858  | 0.1620 |
| Slope               | \( \log S_2 = 2.1554 \log r + 1.5749 \) | 0.9767                        | 2.1554  | 0.1405 |
| Aspect              | \( \log S_2 = 2.0315 \log r + 1.4531 \) | 0.9970                        | 2.0315  | 0.1324 |
| Distance to stream  | \( \log S_2 = 1.4732 \log r + 2.2278 \) | 0.9995                        | 1.4732  | 0.0960 |
| Distance to road    | \( \log S_2 = 2.1156 \log r + 1.5640 \) | 0.9969                        | 2.1156  | 0.1379 |
Different from commonly used BCFD algorithm to measure the spatial association between landslides and conditional factors (e.g., Li et al. 2012; Liucci et al. 2015; Zou and Carranza 2017; Liu et al. 2019), this paper first defines the optimal class number of landslide conditional factor by Shannon’s entropy model, measures the heterogeneous fractal distribution of landslides among the each factor’s class by the VFDM, and uses the derived fractal dimensions to provide the conditional factors with numerical weights.

**Model results and analysis**

The final landslide susceptibility map can be created by summing weighted multiplications of the classified factor maps. Based on Eq. (10), the equation used to create a landslide susceptibility map can be described as:

\[
CLSI = 0.2071 \times fr_{\text{Lithology}} + 0.1240 \times fr_{\text{Distance to fault}} + 0.1620 \times fr_{\text{Altitude}} + 0.1405 \times \]

\[
fr_{\text{Slope}} + 0.1324 \times fr_{\text{Aspect}} + 0.9060 \times fr_{\text{Distance to stream}} + 0.1379 \times \]

\[
fr_{\text{Distance to road}} \quad (14)
\]

The result of this summation shows continuous values ranging from 0.1361 to 2.5896. Typically, the values should be divided into three or five classes expressing the different levels of landslide susceptibilities. In this study, the Natural Breaks classification method has been adopted to divide the calculate \( CLSI \) values into five susceptibility levels, according to Eq. (11) as:

- Very Low (VL) [0.0361–0.7038]
- Low (L) [0.7038–1.0983]
- Moderate (M) [1.0983–1.4831]
- High (H) [1.4831–1.8776]
- Very High (VH) [1.8776–2.5896]

For visualization, the final landslide susceptibility map was illustrated in Fig. 6a.
According to the statistical result of the landslide susceptibility map (Table 5), the VL susceptibility level accounts for 34.80 % of the total studied region, but only 0.89 % of landslides falls in this level. The L, M and H susceptibility zones donate at a value of 26.71 %, 22.12 %, 12.52 %, respectively. Only 3.85 % of the study region is classified as presenting VH susceptibility, yet it includes 24.00 % of the total number of landslides. It can also be observed that when moving from the very low susceptibility class successively to very high susceptibility class, the corresponding landslide density showed an increasing trend.

| Susceptibility | Areas (cell numbers) | Percentage of domain (%) | Number of landslides | Percentage of landslides (%) |
|---------------|----------------------|--------------------------|----------------------|-----------------------------|
| VL            | 45,158,011           | 34.80                    | 4                    | 0.89                        |
| L             | 34,657,849           | 26.71                    | 54                   | 12.00                       |
| M             | 28,704,258           | 22.12                    | 106                  | 23.56                       |

Fig. 6 The landslide susceptibility maps produced by the proposed method (a) and the FR method (b)
Validation and comparison

The prediction performance of the landslide model has been evaluated using the ROC curve. The ROC analysis shows that the proposed method (red line in Fig. 7) provides a satisfying prediction accuracy, with an AUROC value of 0.8467.

Fig. 7 ROC curves based on the validating dataset

The second landslide susceptibility map (Fig. 6b), together with the related ROC curve (blue line in Fig. 7), has been added using the same weight for all the conditional factors. Its ROC shows relatively lower prediction performance (AUROC=0.8088). Applying the same weight values is the same as merely summing the $fr$ values. In this study, applying the fractal-based weights on frequent ratio gives a more accurate prediction for landslide susceptibility. The improved FR model shows higher prediction than the conventional FR model. This finding is consistent with similar comparative results (e.g., Yilmaz 2009; Wang et al. 2014; Kavzoglu et al. 2015; Liu et al. 2019).

Discussions
Factor relative importance

Several influencing factors of landslides have been identified and used to construct the LSM model. The evaluation of factor importance needs to be further discussed, presented in Fig. 8. Assessment of the AUROC values shows that when using all the seven conditional factors, the AUROC gives the highest value of 0.8467. Removing the lithology factor, the most significant reduction in AUROC occurs, indicating that lithology formation as a natural factor provides more information than others. The AUROC, by excluding the altitude factor, also exhibits considerable reduction, followed by the slope. These two factors also act as significant roles pointing out the geomorphology configuration on the propensity to landslides. Distance to stream produces the lowest reduction in AUROC when excluding it in the model, meaning that landslides in the study region do not respond much to this factor.

Since the validating dataset is obtained from the random-partition technique, the AUROC from excluding a factor indicates its more transferability when using future data or employing the model for another region (Kornejady et al. 2017).

![AUROC values for the proposed model using validating dataset](image)

Comparison of typical BCFD and proposed VFDM method

By following the conclusion drawn by Liu (2019), the spatial distribution of landslides is usually a
heterogeneous fractal structure. We believe that the landslide spatial distribution in the study area also
possesses heterogeneous fractal properties, characterized by the VFDM, instead of the most commonly
used BCFD algorithm. Below we discuss this hypothesis.

First, landslides in the study area have been analyzed using the BCFD algorithm (cf., Li et al. 2012).
Table 5 sets out the computed number of grid cells occupied by landslides $N_j$ (Row 3), with different
grid size $r_j$. The parameters $N_j$ versus $r_j$ have been plotted in a double logarithmic coordinate (Fig.
9a). It can be seen that these points are best fitted by a polynomial curve rather than a straight line. This
result indicates that the spatial distribution of landslides presents variable dimension fractal
characteristics, meaning the fractal dimension values change with the linear scaling $r$. In other words,
the fractal structure of landslide distribution is a heterogeneous one.

Furthermore, the VFDM has been introduced to process the calculated data in Table 5. The landslide
grid number $N_j$ was transformed to obtain the 1-order cumulative-sum sequence $S1_j$ (Row 4), then the
data series ($r_j, S1_j$) were plotted in double logarithmic coordinate (Fig. 9b). For comparison, these data
series can be well fitted by a single straight line, from which we can obtain a constant fractal dimension
value.

These outcomes comparisons support our hypothesis and prove the feasibility of VFDM for the current
research.

Table 5 Results of the BCFD algorithm

| Grid size (km) $r_j$ | 0.1 | 0.5 | 1.0 | 2.0 | 3.0 | 4.0 | 5.0 | 6.0 | 7.0 | 8.0 | 9.0 | 10.0 |
|---------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Total Grid number   | 11,798,635 | 469,796 | 117,986 | 29,763 | 13,331 | 7,564 | 4,871 | 3,416 | 2,524 | 1,955 | 1,552 | 1,265 |
This paper presented a novel framework for landslide susceptibility mapping by integrating information theory, fractal theory and statistical analysis. Seven conditional factors were selected to perform statistical analyses, including lithology, distance to fault, altitude, slope, aspect, distance to stream and distance to the road. Significant steps of the proposed framework could be achieved by determining the optimal class number of landslide conditional factors by Shannon’s entropy model and identifying the weights of landslide conditional factors via the variable fractal dimension method (VDFM). The proposed framework was illustrated using an example area in China.

Experimental results from this research showed that the proposed framework had a relatively high prediction performance. Specifically, the proposed method in this work outperformed the conventional FR method, showing a higher prediction accuracy in terms of AUROC values (0.8467) than that estimated for the FR method (0.8088). According to the final prognostic landslide susceptibility map, 16.37% of the study area showed very high and high susceptibility, including the 63.55% of the total number of landslides. The most susceptible areas were located in the northeast mountainous areas, while...
low and very low susceptible zones characterized the central-south areas. Evaluation of relative factor importance based on a one-by-one factor removal test indicated that the lithology factor provides unique information compared to the rest.

Furthermore, for many regions where the landslides distribution presents fractal characteristics with heterogeneous structure, using the framework described here can make a promising landslide susceptibility assessment.

Acknowledgements The authors would like to thank the Editor-in-Chief and anonymous reviewers for their helpful and insightful comments to improve the manuscript. This research was funded by Science and Technology Project of China Highway Engineering Consulting Corporation (No. zzkj-2017).

Declarations

Conflict of interest The author declare no competing interest.

References

Agterberg FP (2013) Fractals and spatial statistics of point patterns. J Earth Sci 24(1):1-11

Bednarik M, Magulova B, Matys M (2010) Landslide susceptibility assessment of the Kral’ovany-Liptovsky’ Mikuláš railway case study. Phys Chem Earth 35:162-171

Brunetti MT, Guzzetti F, Rossi M (2009) Probability distribution of landslide volumes. Nonlin Processes Geophys 16:179-188

Cardinali M, Carrara A, Guzzetti F, et al (2002) Landslide hazard map of the Upper Tiber River basin, central Italy. Geomorphology 96(1):105-122

Catani F, Lagomarsino D, Segoni S, et al (2013) Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. Nat Hazards Earth Syst 13:2815-2831

Chung CJF, Fabbri AG (1999) Probabilistic prediction models for landslide hazard mapping. Photogramm Eng Remote Sens 65:1389-1399
Chung CJF, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. Nat Hazards 20:451-472

Constantin M, Bednarik M, Jurchescu MC, et al (2011) Landslide susceptibility assessment using the bivariate statistical analysis and the index of entropy in the Sibiciu Basin (Romania). Environ Earth Sci 63:397-406

Davis JC (2002) Statistics and Data Analysis in Geology, Third ed. Wiley, USA, 638

Dou J, Tien Bui D, Yunus AP, et al (2015) Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. PLOS ONE 10(7):e0133262

Dou J, Yamagishi H, Pourghasemi HR, et al (2015) An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. Nat Hazards 78:1749

Fawcett T (2006) An introduction to ROC analysis. Pattern Recog Lett 27:861-874

Ghosh S, van Westen CJ, Carranza EJM, et al (2012) Generating event-based landslide maps in a data-scarce Himalayan environment for estimating temporal and magnitude probabilities. Eng Geol 128:49-62

Goetz JN, Brenning A, Petschko H, et al (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput Geosci 81:1-11

Guthrie RH, Deadman PJ, Cabrera AR, et al (2007) Exploring the magnitude-frequency distribution: a cellular automata model for landslides. Landslides 5:151-159

Guzzetti F, Malamud BD, Turcotte DL, et al (2002) Power-law correlations of landslide areas in central Italy. Earth Planet Sci Lett 195:169-183

Guzzetti F, Reichenbach P, Cardinali M, et al (2005) Landslide hazard assessment in the Staffora basin, northern Italian Apennines. Geomorphology 72:272-299

Iwahashi J, Watanabe S, Furuya T (2003) Mean slope-angle frequency distribution and size frequency distribution of landslide masses in Higashikubiki area, Japan. Geomorphology 50(4):349-364

Jenk GF (1963) Generalization in statistical mapping. Ann Assoc Am Geogr 53(1):15-26
Kavzoglu T, Sahin EK, Colkesen I (2015) An assessment of multivariate and bivariate approaches in landslide susceptibility mapping: a case study of Duzkoy district. Nat Hazards 76:471-496

Köppen W (1884) The thermal zones of the Earth according to the duration of hot, moderate and cold periods and of the impact of heat on the organic world. Meteorol Z 1:215-226

Kornejady A, Ownegh M, Bahremand A (2017) Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. Catena 152:144-162

Lee S, Pradhan B (2007) Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides 4:33-41

Lee S, Talib JA (2005) Probabilistic landslide susceptibility and factor effect analysis. Environ Geol 47:982-990

Li CJ, Ma TH, Zhu XS, et al (2011) The power-law relationship between landslide occurrence and rainfall level. Geomorphology 130(3/4):221-229

Li CJ, Ma TH, Sun LL, et al (2012) Application and verification of a fractal approach to landslide susceptibility mapping. Nat Hazards 61:169-185

Liu LN, Li SD, Li X, et al (2019) An integrated approach for landslide susceptibility mapping by considering spatial correlation and fractal distribution of clustered data. Landslides 16:715-728

Liucci L, Mellelli L, Suteanu C (2015) Scale-invariance in the spatial development of landslide in the Umbria Region (Italy). Pure Appl Geophys 172:1959-1973

Lu J, Wu J, Yao H, et al (2011) Predicting river dissolved oxygen in complex watershed by using sectioned variable dimension fractal method and fractal interpolation. Environ Earth Sci 66:2129-2135

Malamu BD, Turcotte DL, Guzzetti F, et al (2004) Landslide inventories and their statistical properties. Earth Surf Proc Land 29(6):687-711

Mandelbrot B (1967) How long is the coast of Britain? Statistical self-similarity and fractal dimension. Science 156:636-638
Oliveira SC, Zêzere JL, Garcia RAC (2014) Structure and characteristics of landslide input data and consequences on landslide susceptibility assessment and prediction capability. In: Lollino G, et al (eds) Engineering Geology for society and territory, Springer International Publishing, Switzerland.

Pourghasemi HR, Mohammady M, Pradhan B (2012) Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. Catena 97:71-84

Pradhan B (2010) Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multiple logistic regression approaches. J Indian Soc Remote Sens 38:301-320

Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput Geosci 51:350-365

Pradhan B, Mansor S, Pirastech S, et al (2011) Landslide hazard and risk analyses at a landslide prone catchment area using statistical based geospatial model. Int J Remote Sens 32(14):4075-4087

Regmi NR, Giardino JR, Vitek JD (2010) Assessing susceptibility to landslides: Using models to understand observed changes in slopes. Geomorphology 122:25-38

Rouai M, Jaaidi EB (2003) Scaling properties of landslides in the Rif Mountains of Morocco. Eng Geol 68:353-359

Sezer E (2010) A computer program for fractal dimension (FRACEK) with application on type of mass movement characterization. Comput Geosci 36:391-396

Shahabi H, Hashim M (2015) Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. Sci Rep 5:9899

Shannon CE (1948) A mathematical theory of communication. Bull Syst Technol J 27:379-423

Shen GQ (2002) Fractal dimension and fractal growth of urbanized areas. Int J Geogr Inf Sci 16(5):419-437

Shi YF, Jin FX (2009) Landslide Stability Analysis Based on Generalized Information Entropy. 2009 International Conference on Environmental Science and Information Application Technology, 83-85
Shirzadi A, Tien Bui D, Pham BT, et al (2017) Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. Environ Earth Sci 76:60

Sun HZ, Wen ZP, Wang F, et al (2016) Dam structural behavior identification and prediction by using variable dimension fractal model and iterated function system. Appl Soft Comput 48:612-620

Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240:1285-1293

Tien Bui D, Pradhan B, Lofman O, et al (2012) Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and naive bayes models. Math Probl Eng 2012:1-26

Tien Bui D, Pradhan B, Lofman O, et al (2013) Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam. Nat Hazards 66:707-730

Tien Bui D, Tuan TA, Klempe H, et al (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13:361-378

Trigila A, Iadanza C, Spizzichino D (2010) Quality assessment of the Italian landslide inventory using GIS processing. Landslides 7(4):455-470

Tsangaratos P, Ilia L, Hong HY, et al (2017) Applying Information Theory and GIS-based quantitative methods to produce landslide susceptibility maps in Nancheng County, China. Landslides 14:1091-1111

Wang Q, Wang Y, Niu RQ, et al (2017) Integration of information theory, K-Means cluster analysis and the logistic regression model for landslide susceptibility mapping in the Three Gorges Area, China. Remote Sens 9:938

Wang XL, Zhang LQ, Wang SJ, et al (2014) Regional landslide susceptibility zoning with considering the aggregation of landslide points and the weights of factors. Landslides 11:399-409

Yalcin A, Reis S, Aydinoglu AC, et al (2011) A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistic and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. Catena 85:274-
Yang Z, Qiao J (2009) Entropy-based hazard degree assessment for typical landslides in the three Gorges Area, China. In: Wang F, Li T (Eds) Landslide Disaster Mitigation in Three Gorges Reservoir, China. Environmental Science and Engineering. Springer, Heidelberg, 519-529

Yang Z, Qiao J, Zhang X (2010) Regional Landslide Zonation based on entropy method in three Gorges Area, China. 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010), 1336-1339

Yilmaz I (2009) Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat-Turkey). Comput Geosci 35:1125-1138

Zou RG (2016) A nonlinear controlling function of geological features on magmatic-hydrothermal mineralization. Sci Rep 6(1):27127

Zou RG, Agterberg FP, Cheng QM, et al (2009) Fractal characterization of the spatial distribution of geological point process. Int J Appl Earth Obs 11:394-402

Zou RG, Carranza EJM (2017) A fractal measure of spatial association between landslides and conditional factors. J Earth Sci 28(4):588-594

Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 39:561-577