A data-driven approach for landslide susceptibility mapping: a case study of Shennongjia Forestry District, China

Wei Chen, Hongxing Han, Bin Huang, Qile Huang and Xudong Fu

Department of Geotechnical and Bridge Engineering, School of Civil Engineering, Wuhan University, Wuhan, China

ABSTRACT
The main purpose of this study was to establish a data-driven approach and assess its potential for shallow landslide susceptibility mapping of the Shennongjia Forestry District, China. For the data processing, Fisher segmenting was used for the classification of topographical factors (gradient, relief amplitude, plan curvature, and normalized difference vegetation index) and the improved Otsu method was used to determine the buffer length thresholds of the locational factors (proximity to roads, rivers and faults). The computations show that these two objective methods help to avoid the uncertainty caused by the subjective judgments and the obtained thresholds are more consistent with actual conditions. Considering the different data types, the locational and topographical factors were evaluated by fuzzy comprehensive evaluation while the geological factors were evaluated by frequency index, and then the three separate evaluations were combined using radar chart analysis. The validation result shows that the susceptibility map based on a data-driven approach has a high accuracy of prediction and this approach will be salutary for risk mitigation and land-use management in mountainous regions.

1. Introduction
Due to the complex geographical conditions and expanding human construction, landslides often occur during the rainy season in the mountainous regions of central and southern China. Among these landslides, shallow landslides have a wide distribution and occur frequently, which poses serious threats to residents and causes huge economic losses. (Li et al. 2012; Pastor et al. 2014; Xu et al. 2015).

Landslide susceptibility mapping plays an essential role in risk mitigation and land-use management in mountainous areas. A landslide susceptibility map emphasizes static landslide-prone conditions, making it easier to predict the spatial distribution and the probability of landslides (Lee et al. 2008; Eeckhaut et al. 2009). Recently, the Geographic Information System (GIS) has made it possible to apply theoretical methods to the regional landslide susceptibility mapping because of its spatial analysis and data manipulation (Pourghasemi et al. 2013; Crozier 2017; Hong et al. 2018), such as fuzzy (Jamaludin et al. 2006; Bui et al. 2012), artificial neural network (Ayalew & Yamagishi 2005; Rossi et al. 2010), logistic regression (Pradhan 2010; Nejad et al. 2016), and linear weighted combination (Ayalew et al. 2004; Jamaludin et al. 2006; Soltani et al. 2013; Shahabi & Hashim 2015). More recently, comparisons between different methods were conducted. For instance, Yilmaz (2009) assessed landslide susceptibility with four methods (conditional probability, logistic regression,
artificial neural network, and support vector machine) and concluded that the difference in the prediction accuracy of the four different methods was within 0.2%. Eker et al. (2014) compared five methods for landslide susceptibility evaluation (linear discriminant analysis, quadratic discriminant analysis, artificial neural network, logistic regression, and spatial regression) and drew similar conclusions that the five different methods have a similar prediction accuracy with a largest difference of less than 0.5%. As the comparison analysis revealed, with the same data processing, there is little difference in the prediction accuracy between the above methods (Melchiorre et al. 2008; Zhang et al. 2016).

Classification and the determination of buffer length threshold are vital components of data processing. However, because there are no universal standards or guidelines, most scholars used subjective decisions commonly, which even led to different results within a same project. For instance, for the buffer length thresholds of roads, Xu et al. (2015) took 250 m as the maximum threshold, while Hong et al. (2016) used 2000 m. A smaller threshold cannot ensure that all the regions influenced by a certain factor are considered, whereas with a greater threshold, there would be an increase in the cost of the risk mitigation and land-use management (Li et al. 2010; Wang et al. 2014). Moreover, different data sources lead to different data types: most topographical factors derived from the digital elevation model (DEM) are continuous; lithological characters, such as types of aquifer are generally discrete, are commonly digitized from the thematic map directly. In most cases, a single method cannot deal sufficiently with the situation of different data types.

For these reasons, in the current paper, considering the different data types of selected influence factors, a data-driven approach is established. Fisher segmenting and the improved Otsu method were employed for data processing and the selected factors were evaluated by fuzzy comprehensive evaluation and frequency index (FI) separately, and then combined using radar chart analysis. Subsequently, based on this data-driven approach, a landslide susceptibility map of the Shennongjia Forestry District, China, where shallow landslides frequently occur, was produced.

2. Study area

The study area, Shennongjia Forestry District, is located within latitude 31°15′ and 31°75′ N and longitude 109°56′ and 110°58′ E, covering an area of about 3250 km² (Figure 1). The district is bounded by the Ta-pa Mountains with altitudes ranging from 420 to 3100 m and decreasing from...
the southwest to the northeast. The average annual temperature is 12 °C, and the rainy season is mainly between May and October. The types of aquifer are mainly carbonate and clastic, where channels are well-formed for the migration of groundwater due to the development of karst pipelines. There are more than 300 mountain rivers in the district. The annual amount of groundwater is about 840 million m$^3$. The length of the road network is 1500 km. The main north–south travel route is the China National Highway 209 (G209), and the Hubei Provincial Route 307 (S307) runs through the northeastern part. Buildings and tourist attractions are widely distributed along the rivers and roads.

With the acceleration of the town construction, the destruction of the natural environment has led to frequent landslides. It is observed from the landslide inventory that there were 285 shallow landslides over the last decade in the district (Table 1). The existing landslides are mainly small and medium scale (88.4% of the total), the minimum scale is 150 m$^3$, whereas the largest landslide reaches about 2500 × 10$^4$ m$^3$ (Figure 2). The frequent landslides have posed serious threats to residents and causes huge economic losses. For instance, Landslide Yaolan occurred twice due to heavy rainfall on 5 and 22 August 2012. Referring to the field survey, this landslide was about 120 m long and 63 m wide with an average thickness of 6.5 m. Landslide Yaolan led to traffic interruption of G209 for days and economic losses of about US$0.75 million, threatening the safety of 10 households as well as the passing vehicles. Therefore, it is necessary for decision-makers to take into account susceptibility to shallow landslides during risk mitigation and land-use management.

3. Data and methodology

3.1. Data used

3.1.1. Landslide inventory map

A landslide inventory is an important component of the spatial database, which records the details of the existing landslides and helps in the analysis of the connections between landslide occurrences
and influence factors (Ercanoglu et al. 2004; Reis et al. 2009; Mohammady et al. 2012; Martha et al. 2013). By means of extensive field surveys and spatial analysis, the landslide inventory, which contains existing 285 shallow landslides, was digitized by the School of Urban Design of Wuhan University. In the current paper, 199 landslides (70% of the total) were randomly selected for modelling, and the remaining 86 landslides (30% of the total) were used for validation (Figure 1).

3.1.2. Topographical factors
The topographical factors of the gradient, relief amplitude (RA), and plan curvature (PC) were derived from the DEM that was provided by the School of Urban Design, Wuhan University. RA is a measure of the surface cutting degree, which is defined as the maximum elevation difference within a certain range. PC describes the curvature of a contour line in the horizontal direction and is an essential measurement of the slope erosion (Pourghasemi et al. 2012; Bui et al. 2012). Normalized difference vegetation index (NDVI) is a measure of surface reflectance and gives a quantitative estimate of the vegetation growth and biomass (Hall et al. 1995; Deng et al. 2007; Zhu et al. 2013; Dou et al. 2015). Using the satellite image of Landsat 8 from February 2017, NDVI was calculated: 
$$\text{NDVI} = \frac{(\text{NIR} - \text{RED})}{(\text{NIR} + \text{RED})},$$
where NIR and RED represent the spectral reflectance of the near-infrared and red bands of the electromagnetic spectrum, respectively.

3.1.3. Locational factors
The locational factors refer to the proximity to locational elements (roads, rivers, and faults). The influence of the locational factors to landslides can be characterized by the distance between mapping cells and the locational elements. Roads affect landslides in the form of excavation, additional load, etc. The region near rivers is likely to have a higher water content, which may erode the slopes or saturate the lower part of slopes, resulting in a decline in the support capacity. The soil close to a fault is considered to be relatively weak because of the altered surface material structures and the increased permeability (Kanungo et al. 2006; Yalcin & Bulut 2006). The road layer was digitized from the traffic map of the Shennongjia Forestry District with a 1:50,000 scale. The river layer was digitized from the water resource map of the Shennongjia Forestry District with a 1:50,000 scale. The fault layer was digitized from the geological map of the Hubei province with a 1:100,000 scale.

3.1.4. Geological factors
The types of aquifer determine the migration of groundwater, which is one of the important indicators of lithological character (Stegmann et al. 2011; Suzuki & Higashi 2012). The layer of aquifers was directly digitized from the published hydrological map of the Hubei province with a 1:100,000 scale. The aquifers have been divided into five types (metamorphic-carbonate, clastic, clastic-carbonate, carbonate, and carbonate-clastic) according to the aquifer medium and hydrodynamic features.

3.2. Procedures and methods

3.2.1. General approach
The data-driven approach can roughly be divided into a three-step procedure (Figure 3). The first step is the data processing, which includes the classification of topographic factors by Fisher segmenting and the determination of buffer length thresholds for locational factors by the improved Otsu method. The second is the separate evaluations using fuzzy comprehensive evaluation and frequency index. The last step is the combination of the three separate evaluations using radar chart analysis to produce the landslide susceptibility map.

3.2.2. Fisher segmenting
The Fisher segmenting method is an effective clustering analysis of ordered samples based on the minimum sum of variance of all classes. For a group of ordered samples X = \{x_i, x_{i+1}, \ldots, x_j\} (i<j), P = \{i, i+1, \ldots, j\}, the average \(x_p\), the diameter, D(i, j), and the objective function, B(n, k), can be
determined via the following equations:

\[ x_p = \frac{1}{j-i-1} \sum_{t=1}^{j} x_t, \]  
\[ D(i,j) = \sum_{t=1}^{j} (x_t - \bar{x}_p)^T (x_t - \bar{x}_p), \]  
\[ B(n,k) = \sum_{t=1}^{k} D(i_t, i_{t+1} - 1), \]

where \( n \) represents the number of samples and \( k \) represents the number of classes. For a specific sample and a fixed class number, the smaller the \( B(n,k) \) is, the more reasonable the classification.

### 3.2.3. Improved Otsu method

In the GIS analysis, buffers are built to consider the influence of locational factors on landslides. Before building the buffers, there is a need to determine the buffer length threshold. The Otsu
algorithm is an effective algorithm for image binarization in computer vision and image processing (Otsu 1979). The algorithm searches for the optimum threshold separating the image into the foreground and background so that intra-class variance is minimal and inter-class variance is maximal:

\[ \sigma^2_w(t) = w_0(t) \cdot \sigma_0^2(t) + w_1(t) \cdot \sigma_1^2(t), \] (4)

where \( \sigma^2_w(t) \) is the sum of variances of the foreground and background, \( w_0(t) \) and \( w_1(t) \) are the proportions of the two classes separated by a threshold \( t \), and \( \sigma_0^2(t) \) and \( \sigma_1^2(t) \) are the variances of these two classes, respectively. When the sum of variances \( \sigma^2_w(t) \) is at its maximum, the threshold \( t \) is the optimal threshold.

The traditional Otsu method only classifies the sample into two categories, so it cannot meet the requirements of multiple classes. To build the buffers, \( L_i \) (where \( i \) is the number of classifications) is assumed as the distance threshold that divides the existing landslides \( C \) into influenced class \( c_1^{(i)} \) and relatively less influenced class \( c_0^{(i)} \). The sum of variances is defined by the following equation:

\[ \sigma^2(L^{(i+1)}) = (w_0 \cdot \sigma_0^2)_{c \in c_0^{(i)}} + (w_1 \cdot \sigma_1^2)_{c \in c_1^{(i)}}, \] (5)

where \( L^{(0)} \) is the maximum distance between the locational factors and the landslides (\( L_{\text{max}} \)), and \( c_1^{(0)} \) is equal to \( C \). When \( i \) is 0, Equations (4) and (5) are the same, which indicates that the first buffer is in the range greater than the threshold \( L^{(1)} \), and similarly, the second is \( L^{(2)} \) to \( L^{(1)} \), and the \( i \)-th buffer is \( L^{(i)} \) to \( L^{(i-1)} \) (Figure 4).

3.2.4. Fuzzy comprehensive evaluation

Fuzzy comprehensive evaluation refers to the overall evaluation of a phenomenon affected by multivariate factors and gives a nonnegative evaluation value (Feng et al. 2014). By performing the fuzzy composite operation of the weights and memberships, a fuzzy comprehensive evaluation is established:

\[ E = W \times F = (w_1, w_2, \ldots, w_m) \times \begin{pmatrix} f_{11} & \cdots & f_{1n} \\ \cdots & \ddots & \cdots \\ f_{m1} & \cdots & f_{mn} \end{pmatrix}, \]

where \( W \) represents the weight vectors of factors, \( F \) represents the fuzzy membership matrix, \( m \) is the factor number, \( n \) is the class number, and \( f_{mn} \) is the fuzzy membership for factor \( m \) of class \( n \).

Figure 4. Buffer length thresholds of the locational factors (produced via Microsoft Visio 2010).
The final evaluation is determined by the maximum membership principle (Liang et al. 2003; Lv et al. 2013).

Fuzzy membership $r_A(x)$ is used to characterize the extent of element $x$ belonging to fuzzy set $A$, whose values lie between 0 and 1. The higher the value is, the greater the degree of belonging to the fuzzy set and vice versa (Devkota et al. 2012). The membership functions that are commonly used include linear, exponential, hyperbolic, and inverse hyperbolic. In the current paper, the fuzzy membership was assumed to follow a Gaussian distribution (Figure 5) and can be calculated using the following equations (Su et al. 1993):

$$r_{ij} = e^{-\left(\frac{x_i - m}{c}\right)^2},$$  \hspace{1cm} (7)

$$m = \frac{1}{2} (x_i + x_{i+1}),$$ \hspace{1cm} (8)

$$c = 0.6 \cdot (x_{i+1} - x_i),$$ \hspace{1cm} (9)

where $x_{i+1}$ and $x_i$ represent the upper and lower bounds of the class interval.

Information entropy is a frequently-used objective weighting method based on the characteristics of the original data, which avoids the deviations that arise from subjective decisions and can better characterize the importance of the influence factors (Devkota et al. 2012; Pourghasemi et al. 2012).

### 3.2.5. Radar chart analysis

A radar chart is a graphical tool for displaying multivariate data in the form of a two-dimensional chart and it is useful for displaying multivariate observations with an arbitrary number of variables (Chambers et al. 1983; Friendly 2009). Radar chart analysis involves simple calculations and is easy to understand, and it has been widely used in the fields of product comparison and control of quality improvement (Chang et al. 2012; Kalonia et al. 2013).

As shown in Figure 6, a radar chart consists of equal-angular axes, each representing the susceptibility class of the separate evaluation; lines are drawn connecting data points on adjacent axes, forming a characteristic polygon for the comprehensive evaluation of a distinct mapping cell. According to the area of the polygon, the score function of the comprehensive evaluation is established as follows:

$$ES = f(e_1, e_2, e_3) = \sqrt{3} \cdot \frac{1}{4} (c_1 \cdot c_2 + c_1 \cdot c_3 + c_2 \cdot c_3),$$ \hspace{1cm} (10)

where $c_i$ represents the susceptibility class of the separate evaluations.

### 4. Results and validation

#### 4.1. Data processing

The factor data were processed via the Fisher and improved Otsu method, and the FIs in each class interval were investigated. The FI refers to the ratio of landslide scale frequency ($u_{1,5}$) to the grid cell.
The FI can objectively reveal the density of landslide occurrences in a certain region, and the regions with relatively higher FI are considered to be more susceptible to landslides. As shown in Table 2, the FIs are the largest in the gradient interval of 13.5° to 21°, in the RA interval of 15 to 25 m, in the PC interval of −0.1 to 0.1, and in the NDVI interval of more than 0.36. Using improved Otsu method, buffer length thresholds were determined (Table 3). The maximum road buffer is slightly greater than the maximum river buffer and much smaller than the maximum fault buffer. Moreover, in the regions within 250 m of the locational elements, the FIs of roads are generally higher.

Table 2. Frequency indexes of the topographical factors.

| Intervals     | $u_{LS}$ (%) (a) | $u_{cell}$ (%) (b) | Frequency index (a/b) |
|--------------|-----------------|-------------------|----------------------|
| Gradient (°)  |                 |                   |                      |
| ≤13.5        | 14.07           | 12.81             | 1.10                 |
| 13.5–21      | 29.65           | 19.19             | 1.54                 |
| 21–29        | 23.62           | 26.62             | 0.89                 |
| 29–41.5      | 28.14           | 32.38             | 0.87                 |
| >41.5        | 4.52            | 8.99              | 0.50                 |
| RA (m)       |                 |                   |                      |
| ≤15          | 13.57           | 11.91             | 1.14                 |
| 15–25        | 32.66           | 21.51             | 1.52                 |
| 25–40        | 28.64           | 35.54             | 0.81                 |
| 40–65        | 22.11           | 26.65             | 0.83                 |
| >65          | 3.02            | 4.77              | 0.63                 |
| PC           |                 |                   |                      |
| ≤−0.4        | 5.53            | 20.58             | 0.27                 |
| −0.4–0.1     | 27.14           | 20.19             | 1.34                 |
| −0.1–0.1     | 26.63           | 17.17             | 1.55                 |
| 0.1–0.35     | 22.61           | 17.58             | 1.29                 |
| >0.35        | 18.09           | 24.49             | 0.74                 |
| NDVI         |                 |                   |                      |
| ≤0.14        | 6.03            | 5.86              | 1.03                 |
| 0.14–0.21    | 30.15           | 37.15             | 0.81                 |
| 0.21–0.28    | 33.67           | 35.96             | 0.94                 |
| 0.28–0.36    | 23.62           | 17.81             | 1.33                 |
| >0.36        | 6.53            | 3.22              | 2.03                 |

Figure 6. Radar schematic chart (Kalonia et al. 2013).
4.2. Mapping

The results of weights determined by information entropy included the weights of the topographical factors and the weights of locational factors (Table 4). Among the selected topographical factors, both gradient and RA have a high percentage of the weights (almost three times the weight of PC and twice the weight of NDVI). As for the locational factors, the roads have the greatest weight, the rivers come second, and the faults the last.

Using the classified layers, the fuzzy memberships of the mapping cells were calculated via Equations (7)—(9). As shown in Figure 8(a), the memberships of mapping cells follow a Gaussian distribution. As the gradient is about 17.25°, the mapping cells have the greatest extent of belonging to class V, and with the increasing gradient, the extent of belonging to class V gradually decreases while the extent of belonging to class III increases. Similarly, as shown in Figure 8(f), with increasing of the distance to roads, the memberships of mapping cells trend to belong to a lower class.

Based on the obtained fuzzy memberships and weights, the topographical and locational factors were evaluated separately (Figure 9(a,b)). Meanwhile, the geological factor, namely types of aquifer, was also evaluated into five (I–V) categories via the FI (Figure 9(c)). As shown in Table 5, the FIs of each separate evaluation are similar: the FIs increase with the increasing susceptibility class. In particular, the FIs of the locational evaluation change evenly in class I to class III, and increase sharply as the class increases from III to IV, then become even in class IV and class V. The FIs of the geological evaluation stay fairly low in classes I to IV, but increase sharply as the class increases from IV to V.

| Intervals         | \( u_{LS} \) (%) (a) | \( u_{cell} \) (%) (b) | Frequency index (a/b) |
|-------------------|----------------------|------------------------|-----------------------|
| Proximity to roads (m) |                      |                        |                       |
| \(<55\)          | 29.65                | 10.60                  | 2.80                  |
| 55–115           | 25.13                | 8.23                   | 3.05                  |
| 115–245          | 22.11                | 12.32                  | 1.80                  |
| 245–710          | 15.58                | 28.24                  | 0.55                  |
| \(>710\)         | 7.54                 | 40.62                  | 0.19                  |
| Proximity to rivers (m) |                      |                        |                       |
| \(<50\)          | 15.08                | 9.02                   | 1.67                  |
| 50–105           | 16.08                | 10.83                  | 1.49                  |
| 105–240          | 31.66                | 21.56                  | 1.47                  |
| 240–590          | 25.13                | 33.48                  | 0.75                  |
| \(>590\)         | 12.06                | 25.11                  | 0.48                  |
| Proximity to faults (m) |                      |                        |                       |
| \(\leq305\)      | 16.58                | 10.96                  | 1.51                  |
| 305–635          | 9.05                 | 9.88                   | 0.92                  |
| 635–1470         | 25.13                | 19.15                  | 1.31                  |
| 1470–3145        | 16.08                | 26.13                  | 0.62                  |
| \(>3145\)        | 33.17                | 33.87                  | 0.98                  |

Table 3. Frequency indexes of the locational factors.

Table 4. Weights of the topographical and locational factors.
On the basis of radar chart analysis, the above three evaluations were combined. The comprehensive evaluation was classified into the categories I–V with the segmenting points of 3, 12, 27, and 48, and the susceptibility map was generated (Figure 10). As the map shows, 46% of the district's total area was in susceptibility classes IV and V, which is where more than 70% of existing landslides are distributed.

Figure 7. Map layers of the selected influence factors: (a) gradient, (b) RA, (c) PC, (d) NDVI, (e) roads, (f) rivers, (g) faults, and (h) types of aquifers (produced via Arcgis 10.2).
4.3. Validation

The radar charts of the mapping cells in typical regions (Figure 10(a–d)) were analysed comparatively. For the region a, the three lower separate evaluation classes were combined into a low comprehensive evaluation class, and vice versa for the region b. Moreover, a single low or high class could not make an extreme evaluation class, such as the regions c and d (Table 6).

The validation results indicate that the FIs of both the data for the modelling and the data for the validation increase with the increasing susceptibility class similar to the FIs of the separate evaluations (Table 7). Of the existing landslides, 71.4% in the data for modelling and 72.1% in the data for validation are located in regions with susceptibility classes IV and V (46.8% of district’s total area). Meanwhile, in the same susceptibility class, the FIs of both are similar.

5. Conclusion and discussion

5.1. Conclusion

A data-driven approach was established for regional landslide susceptibility mapping. This approach could be roughly divided into a three-step procedure: (i) data processing, which includes the classification of topographic factors by Fisher segmenting and the determination of buffer length thresholds for locational factors by the improved Otsu method; (ii) separate evaluations using fuzzy
comprehensive evaluation and frequency index; (iii) combination of the separate evaluations by radar chart analysis.

The inaccurate thresholds of the locational factors determined by the subjective decisions could lead to an incomplete consideration or a higher cost and, therefore, the improved Otsu method was established. The buffer length thresholds of roads are 55, 115, 245, and 710 m, the buffer length thresholds of rivers are 50, 105, 240, and 590 m, and the buffer length thresholds of faults are 205, 365, 1470, and 3415 m.

The combination of the separate evaluations by radar chart analysis solves the problem of different data types caused by the different data sources. The FIs of the data for the modelling and the

Table 5. Frequency indexes of each evaluation.

| Separate evaluations | $u_{LS}$ (%) (a) | $u_{cell}$ (%) (b) | Frequency index (a/b) |
|----------------------|------------------|-------------------|----------------------|
| **Topographical evaluation** | | | |
| I | 5.03 | 8.08 | 0.62 |
| II | 34.67 | 46.02 | 0.75 |
| III | 12.56 | 12.03 | 1.04 |
| IV | 16.08 | 12.36 | 1.30 |
| V | 31.66 | 21.51 | 1.47 |
| **Locational evaluation** | | | |
| I | 9.55 | 41.25 | 0.23 |
| II | 16.08 | 25.86 | 0.62 |
| III | 27.64 | 16.47 | 1.68 |
| IV | 28.14 | 10.07 | 2.79 |
| V | 18.59 | 6.35 | 2.93 |
| **Geological evaluation** | | | |
| I (Meta-Carb) | 0.00 | 1.04 | 0.00 |
| II (Carbonate) | 5.53 | 12.72 | 0.43 |
| III (Clas-Carb) | 8.04 | 10.22 | 0.79 |
| IV (Carb-Clas) | 66.83 | 70.43 | 0.95 |
| V (Clastic) | 19.60 | 5.59 | 3.51 |
data for validation have common distribution features and similar values, which indicates that the obtained susceptibility map has a high accuracy of prediction.

Since this is a prototype study performed in landslide-prone region in Shennongjia Forestry District, China, the findings may potentially be utilized for regions with similar topographical and geological conditions.

5.2. Discussion

For selection of suitable influence factors, the quality of the landslide inventory map, the difficulties in data acquisition, the credibility of the obtained data, and the connections between landslides and influence factors should be fully considered.
A high-quality landslide inventory map should cover the occurrence time and shape feature of landslides, the physical and mechanical properties of the main soil and rock, and the changes in the main triggering factors, etc. However, the obtained landslide inventory only records the locations and scales. To be compatible with the frequency index, in this paper, the circular buffers were built according to the recorded scales, where the average values were extracted from the factor data layers as the attributes of the existing landslides.

Precipitation is one of the main trigger factors for landslides (Tarolli et al. 2011). The obtained precipitation data is the average annual precipitation (AAP) of the last 10 years from eight meteorological observation stations. Using the ordinary Kriging method, a spatial interpolation method, the AAP layer was produced. However, the AAP layer shows that the AAP appears to be uniform throughout the district, which makes the weight of the AAP calculated via information entropy too small to characterize the importance of AAP and, therefore, the AAP was not considered in the current paper. The same can be said for the land-use.

It seems not that the greater the gradient is, the higher the probability of landslide is, which may due to the locations of constructions and the fact that steep slopes stay relatively stable under natural conditions. As the flat terrain is scarce in the district, constructions are mainly concentrated on slopes, leading to a change in the slope structures. Under natural conditions, the stability of the slope is mainly determined by the geological conditions and mechanical properties. When the surrounding environment changes drastically or slopes are subject to constructions, the slopes are relatively unstable.

The diversity of the data sources leads to different data types. For data processing, there is a need to classify the DEM-derived data and build buffers for locational elements. The Fisher segmenting and the improved Otsu method solve these issues. Meanwhile, other objective methods, such as K-means, Jenks, etc., are alternative solutions. The performances of these methods should be investigated in further studies.

A landslide susceptibility map is usually estimated using area under the receiver operating characteristic curve (AUC) (Lee et al. 2008; Chen et al. 2014; Hong et al. 2017). However, since the susceptibility indexes calculated by radar chart analysis are still discrete, the AUC is not suitable for map validation. As shown in Table 7, the FIs of the data for the modelling are very similar with the FIs of the data for validation in the same susceptibility class. This is a fairly good result, which indicates that the data-driven approach has a good ability to predict landslides.

In addition to the grid, the mapping cells come in various forms, such as unique-condition units, sub-basin, slope units, etc. Among these, a slope unit is the basic terrain unit for landslide occurrence and can reveal shape features and keep the differences between the units (Guzzetti et al. 1999; Abella & Westen 2008). However, the flow threshold, which is the key to generating the slope units, is sensitive to hydrological characteristics and difficult to determine. Recently, a generation method of slope units established by our team presents a high consistency between the generated slope units and the actual slopes. If this is reasonable, the generation method of slope units could be applied in landslide susceptibility mapping to better predict the shape features of the landslides.

Based on the discussion above, the future work of our team is mainly focused on the filtering of influence factors and the application of slope units. Meanwhile, some other evaluation method, including the mechanics method, will be considered and introduced, and the detailed geological investigations will be conducted for obtaining the geotechnical parameters at the site-specific scale, where the local geological heterogeneities may predominate.

Disclosure

No competing financial interests exist.

Funding
The study is funded by the National Science and Technology Support Project of China [grant number 2014BAL05B07].
Acknowledgments

The authors would like to express their gratitude to the Government of Shennongjia Forestry District and the School of Urban Design of Wuhan University, and they also give thanks to the comments from the editors and reviewers.

References

Abella EAC, Westen CJV. 2008. Qualitative landslide susceptibility assessment by multicriteria analysis: a case study from San Antonio del Sur, Guantánamo, Cuba. Geomorphology. 94(3–4):453–466.
Ayalew L, Yamagishi H, Ugawa N. 2004. Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. Landslides. 1(1):73–81.
Ayalew L, Yamagishi H. 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology. 65(1–2):15–31.
Bui DT, Pradhan B, Lofman O, Revhaug I, Dick OB. 2012. Landslide susceptibility assessment in the Hoa Binh province of Vietnam: a comparison of the Levenberg–Marquardt and Bayesian regularized neural networks. Geomorphology. 171–172:12–29.
Chambers JM, Cleveland WS, Kleiner B, Tukey PA. 1983. Graphical methods for data analysis. Boston (MA): Duxbury Press.
Chang YC, Chang CJ, Chen KT, Lei CL. 2012. Radar chart: scanning for satisfactory QoE in QoS dimensions. IEEE Netw. 26(4):25–31.
Chen W, Li W, Hou E, Bai H, Chai H, Wang D, Cui X, Wang Q. 2014. Application of frequency ratio, statistical index, and index of entropy models and their comparison in landslide susceptibility mapping for the Baozhong Region of Baoji, China. Arab J Geosci. 8(4):1829–1841.
Crozier MJ. 2017. A proposed cell model for multiple-occurrence regional landslide events: implications for landslide susceptibility mapping. Geomorphology. 295:480–488.
Deng Y, Chen X, Chuvieco E, Warner T, Wilson JP. 2007. Multi-scale linkages between topographic attributes and vegetation indices in a mountainous landscape. Remote Sens Environ. 111(1):122–134.
Devkota KC, Regmi AD, Pourghasemi HR, Yoshida K, Pradhan B, Ryu IC, Dhillon MR, Althuwaynee OF. 2012. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. Nat Hazards. 65(1):135–165.
Dou J, Bui DT, Yunus AP, Jia K, Song X, Revhaug I, Xia H, Zhu Z. 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.
Eeckhaut MVD, Reichenbach P, Guzzetti F, Rossi M, Poesen J. 2009. Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardenne, Belgium. Nat Haz Earth Syst Sci. 9(2):507–521.
Eker AM, Dikmen M, Cambazoglu S, Düzgün ŞHB, Akgün H. 2014. Evaluation and comparison of landslide susceptibility mapping methods: a case study for the Ulus district, Bartın, northern Turkey. Int J Geogr Inf Sci. 29(1):132–158.
Ercanoglu M, Gokceoglu C, Asch TWJV. 2004. Landslide susceptibility zoning north of Yenice (NW Turkey) by multivariate statistical techniques. Nat Hazards. 32(1):1–23.
Feng L, Zhu X, Sun X. 2014. Assessing coastal reclamation suitability based on a fuzzy-AHP comprehensive evaluation framework: a case study of Lianyangung, China. Mar Pollut Bull. 89(1–2):102–111.
Friendly M. 2009. Milestones in the history of thematic cartography, statistical graphics, and data visualization. http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf.
Guzzettia F, Carrarab A, Cardinalia M, Reichenbacha P. 1999. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology. 31(1–4):181–216.
Hall FG, Townshend JR, Engman ET. 1995. Status of remote-sensing algorithms for estimation of land-surface state parameters. Remote Sens Environ. 51(1):138–156.
Hong H, Chen W, Xu C, Youssef AM, Pradhan B, Bui DT. 2016. Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. Geocarto Int. 32(2):1–16.
Hong H, Ilia I, Tsangaratos P, Chen W, Xu C. 2017. A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the Wuyuan area, China. Geomorphology. 290:1–16.
Hong H, Liu J, Bui DT, Pradhan B, Acharya TD, Pham BT, Zhu AX, Chen W, Ahmad BB. 2018. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China), Catena. 163:399–413.
Jamaludin S, Huat BBK, Omar H. 2006. Evaluation and development of cut–slope assessment systems for Peninsular Malaysia in predicting slaldsides in granitic formation. Jurnal Teknologi. 44(1):31–46.
Kalonia C, Kumru OS, Kim JH, Middaugh CR, Volkin DB. 2013. Radar chart array analysis to visualize effects of formulation variables on IgG1 particle formation as measured by multiple analytical techniques. J Pharm Sci. 102(12):4256–4267.

Kanungo DP, Arora MK, Sarkar S, Gupta RP. 2006. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. Eng Geol. 85(3–4):347–366.

Lee CT, Huang CC, Lee JF, Pan KL, Lin ML, Dong JJ. 2008. Statistical approach to storm event-induced landslides susceptibility. Nat Hazard Earth Syst. 8(4):941–960.

Liang YM, Zhai HC, Chang SJ, Zhang SY. 2003. Color image segmentation based on the principle of maximum degree of membership. Acta Phys Sin. 52(11):2655–2659.

Li N, Xu J, Qin Y. 2012. Research on calculation model for stability evaluation of rainfall-induced shallow landslides. Rock Soil Mech. 33(5):209–214.

Li XW, Zhang LN, Liang C. 2010. A GIS-based buffer gradient analysis on spatiotemporal dynamics of urban expansion in Shanghai and its major satellite cities. Procedia Environmental Sciences.2(6): 1139–1156.

Lv Y, Liu JZ, Yang TT, Zeng DL. 2013. A novel least squares support vector machine ensemble model for NOx emission prediction of a coal-fired boiler. Energy. 55:319–329.

Martha TR, Westen CJV, Kerle N, Jetten V, Kumar KV. 2013. Landslide hazard and risk assessment using semi-automatically created landslide inventories. Geomorphology. 184:139–150.

Melchiorre C, Matteucci M, Azzoni A, Zanchi A. 2008. Artificial neural networks and cluster analysis in landslide susceptibility zonation. Geomorphology. 94(3–4):379–400.

Mohammady M, Pourghasemi HR, Pradhan B. 2012. Landslide susceptibility mapping at Golestan Province, Iran: a comparison between frequency ratio, Dempster–Shafer, and weights-of-evidence models. J Asian Earth Sci. 61:221–236.

Nejad SG, Falah F, Daneshfar M, Haghizadeh A, Rahmati O. 2016. Delineation of groundwater potential zones using remote sensing and GIS-based data-driven models. Geocarto Int. 32(2):1–21.

Otsu N. 1979. A threshold selection method from gray-level. IEEE Trans Syst Man Cybern. 9(1):62–66.

Pastor M, Blanc T, Haddad B, Petrone S, Morles MS, Drempetic V, Issler D, Crosta GB, Cascini L, Sorbino G, et al. 2014. Application of a SPH depth-integrated model to landslide run-out analysis. Landslides. 11(5):793–812.

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.

Pourghasemi HR, Pradhan B, Gokceoglu C, Moezzi KD. 2013. A comparative assessment of prediction capabilities of Dempster–Shafer and Weights-of-evidence models in landslide susceptibility mapping using GIS. Geomat Nat Haz Risk. 4(2):93–118.

Pourghasemi HR, Pradhan B, Gokceoglu C. 2012. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. Nat Hazards. 63(2):965–996.

Pradhan B. 2010. Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate logistic regression model on three test areas in Malaysia. Adv Space Res. 45(10):1244–1256.

Reis S, Nisanci R, Yomralioglu T. 2009. Designing and developing a province-based spatial database for the analysis of potential environmental issues in Trabzon, Turkey. Environ Eng Sci. 26(1):123–130.

Rossi M, Guzzetti F, Reichenbach P, Mondini AC, Peruccacci S. 2010. Optimal landslide susceptibility zonation based on multiple forecasts. Geomorphology. 114(3):129–142.

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

Soltani SR, Mahiny AS, Monavari SM, Alesheikh AA. 2013. Sustainability through uncertainty management in urban land suitability assessment. Environ Eng Sci. 30(4):170–178.

Stegmann S, Sultan N, Kopf A, Apprioual R, Pelleau P. 2011. Hydrogeology and its effect on slope stability along the coastal aquifer of Nice, France. Mar Geol. 280(1–2):743–754.

Su JY, Zhou XY, Fan S. 1993. A fuzzy set evaluation method of hazard degree of debris flow. J Nat Disasters. 2:83–90.

Suzuki K, Higashi S. 2012. Groundwater flow after heavy rain in landslide-slope area from 2-d inversion of resistivity monitoring data. Geophysics. 66(3):733–743.

Tarolli P, Borga M, Chang KT, Chiang SH. 2011. Modeling shallow landsliding susceptibility by incorporating heavy rainfall statistical properties. Geomorphology. 133(3–4):199–211.

Wang J, Yin K, Xiao L. 2014. Landslide susceptibility assessment based on GIS and weighted information value: a case study of Wanzhou district, Three gorges reservoir. Chin J Rock Mech Eng. 33(4):797–808.

Xu K, Guo Q, Li ZW, Xiao J, Qin YS, Chen D, Kong CF. 2015. Landslide susceptibility evaluation based on BPNN and GIS: a case of Guojiaba in the Three Gorges Reservoir Area. Int J Geogr Inf Sci. 29(7):1111–1124.

Yalcin A, Bulut F. 2006. Landslide susceptibility mapping using GIS and digital photogrammetric techniques: a case study from Ardesen (NE-Turkey). Nat Hazards. 41(1):201–226.
Yilmaz I. 2009. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environ Earth Sci. 61(4):821–836.
Zhang J, Yin K, Wang J, Liu L, Huang F. 2016. Evaluation of landslide susceptibility for Wanzhou district of Three Gorges Reservoir. Chin J Rock Mech Eng. 35(2):284–296.
Zhu G, Liu Y, Ju W, Chen J. 2013. Evaluation of topographic effects on four commonly used vegetation indices. J Remote Sens. 17(1):210–234.