Regional terrain complexity evaluation based on GIS and K-means clustering model: a case study of Ningdu County, China

Wenjing Kang, Faming Huang, Yaoxian Du, Dalin Liu and Zhongshan Cao*
School of Civil Engineering and Architecture, Nanchang University, Nanchang, China

*Corresponding author e-mail: cuczs@sina.com

Abstract. In order to accurately quantify the terrain complexity, a simple and accurate terrain complexity assessment (TCA) model is proposed. Taking Ningdu county in Jiangxi Province of China as an example, firstly, six terrain factors (named slope, topographic relief degree, surface cutting depth, surface roughness, elevation variation coefficient and topographic factors) of Ningdu county are extracted based on the Digital Elevation Model (DEM) with 30 m resolution and ARCGIS 10.2 software. Secondly, terrain complexity indexes of Ningdu County are obtained using k-means clustering. Results show that a current and effective spatial distribution characteristic of topographic complexity in Ningdu county is produced, and the very low and low terrain complexity indexes account for 33.28%, 28.35% respectively. The terrain complexity can be evaluated effectively by k-means clustering model. The terrain complexity can be provided for environmental protection and land use planning.

1. Introduction
Terrain complexity can be used as an important reference index to describe the degree of surface fragmentation and the intensity of surface soil erosion. Establishing topographic complexity index evaluation model will provide decision information for environmental protection, construction land evaluation and other applications[1-7]. There are fragile geological environment and complicated terrain in the mountainous areas of southern China, such as many counties in Jiangxi Province. Therefore, it is important to quantify the terrain complexity objectively for the field of digital terrain analysis in Jiangxi province.

Terrain complexity is difficult to quantify, as a result, it is very important to select appropriate quantization methods for terrain complexity modeling. At present, many scholars have studied the methods of quantifying terrain complexity. These methods include terrain complexity quantization based on wave theory [8, 9], fractal theory[10], grid DEM data [8]. These methods have their own emphasis, most of them can better reflect the terrain complexity. However, there are many problems in terrain complexity modeling based on the above methods, such as the complexity of the evaluation process, the single selection of terrain factor, and the inability to evaluate the terrain complexity from both the whole and the local aspects. Therefore, it is very necessary to propose a simple and comprehensive terrain complexity evaluation method, which can take into account various terrain factors.
Clustering analysis is an important technical means of data mining. It is a process of dividing a data set into several subsets and an unsupervised learning method [11]. The k-means clustering is one of the classical clustering algorithms, which was proposed by MacQueen in 1967 and has the advantages of simplicity and speed in processing large scale data [12]. It has been widely used in various fields of science, including engineering machinery, statistical decision, ecological sensitivity and other evaluation analysis work [13].

Digital elevation model (DEM) is the core data of spatial data processing and terrain analysis, and has been widely used in spatial data mining, such as terrain feature extraction, watershed water system analysis [14-17]. Extracting terrain factors from DEM data is the basis of quantifying terrain complexity. This study has six topographical factors, including slope and so on, are extracted from DEM.

To sum up, this paper takes Ningdu County of Jiangxi Province as the study area. Based on the essential features of terrain, six single factors are selected to describe the complexity of terrain. Cluster these single factors by k-means clustering. Finally, the terrain complexity distribution characteristics of the region are obtained and evaluated.

2. Evaluation and Analysis of terrain complexity

2.1. Evaluation process

The processes of topographic complexity evaluation in this paper mainly include:

1) Based on the essential characteristics of topography, the topographic single factor used to evaluate the complexity of terrain is selected.
2) Using ArcGIS to extract various topographic factors, the local shape factors are graded and mapped.
3) The k-means clustering iteration of each single factor is carried out, and the terrain complexity is divided according to the clustering results obtained.
4) The distribution characteristics of terrain complexity in this study area are evaluated.

2.2. k-means clustering

The basic idea of k-means clustering is to randomly select k data objects as the initial clustering center from the dataset containing a large number of data objects. First, we calculate the distance between each data object and the k cluster center, and divide all data into classes represented by the cluster center closest to it. Then we update the K Clustering center based on the mean of the newly generated data objects in each category. By clustering the changes of the clustering center, the topographic factors with similar topography complexity are grouped into one class, which provides a way of thinking for the comprehensive evaluation of topographic complexity.

(1) Obtain research samples and evaluation indexes: set with a total of n samples, each sample has an m indicator consisting of a matrix as shown in Eq. (1).

\[
\begin{pmatrix}
X_1', X_2', \ldots, X_n'
\end{pmatrix} = \begin{pmatrix}
x_1'(1) & x_2'(1) & \cdots & x_n'(1) \\
x_1'(2) & x_2'(2) & \cdots & x_n'(2) \\
\vdots & \vdots & \ddots & \vdots \\
x_1'(m) & x_2'(m) & \cdots & x_n'(m)
\end{pmatrix}
\]

(2) Cluster center: K-mean clustering algorithm to divide Data Sets X into K cluster classes as shown in Eq. (2), each cluster center is \( c_j (j = 1, 2, \ldots, k) \).

\[
C = \{C_1, C_2, \ldots, C_k\}
\]

(3) Description of the steps to run the algorithm:
1) Initialize K Clustering Center points Randomly: \( c_1, c_2, \cdots, c_k \)

2) Specify the number of iterations and repeat the following actions:

For each object \( x'_i \) in the dataset:

\[
C_j = \arg \min_j \| x'_i - c_j \|^2
\]  

(3)

For each cluster center \( c_j \):

\[
c_j = \frac{\sum_{i=1}^{m} 1[C_i = j] x'_i}{\sum_{i=1}^{m} 1[C_i = j]}
\]

(4)

Eq. (3) means dividing each data object \( x'_i \) into a category corresponding to the cluster center \( c_j \) closest to it; Formula (4) Moves the cluster center \( c_j \) to a somewhat mean in the current clustering.

3. Organization of the Text

3.1. General situation of Ningdu County, Jiangxi Province

Ningdu is a county in Ganzhou city of Jiangxi Province, located in the southeast of Jiangxi Province. And it is located 26°05′18″ to 27°08′13″ N, 115°40′20″ to 116°17′15″ E. Its length is of 117.2 kilometers from north to south, its width is 61 kilometers from east to west Its total area is of 4053.16 Km\(^2\), belonging to the subtropical monsoon humid climate zone. The plum river, which originated in the north of Ningdu, flows through 11 townships from north to south, with a total length of 145 km and a basin area of nearly 3000 Km\(^2\). And plum river is the most extensive, longest length and the largest runoff in the Gan River basin. Abundant water resources have given birth to a beautiful ecological environment [6, 17].

![Study area location map](image-url)
3.2. Selection of topographic factors in Ningdu County

3.2.1 Data source. In this paper, the terrain factors such as slope, terrain fluctuation, surface cutting depth, surface roughness, elevation coefficient of variation and profile curvature are extracted using DEM with 30 m resolution. The 30 m resolution raster can not only effectively characterize the topographic and geomorphological features of the study area, but also not lead to excessive model computation due to a large number of grids [18]. Therefore, the topographic complexity evaluation based on the 30 m resolution raster is carried out in this paper.

3.3. Single-factor evaluation.

(1) Slope is the ratio of the vertical distance from the highest point to the horizontal plane of the slope [15, 19]. The DEM data is imported into ArcGIS and the slope of the study area is calculated by the slope command. As shown in Table 1 and Figure 2 (a).

(2) Terrain undulation is the difference between the highest and lowest points in a particular region [20]. By using the neighborhood analysis method, according to the topography characteristics of southern China, the mesh size of the statistical unit is 5×5, and the fluctuation degree of the terrain is obtained. As shown in Table 1 and Figure 2 (b).

(3) There is the difference between the average elevation of the neighborhood range at a point on the ground and the minimum elevation in that neighborhood range [21]. Similar to terrain fluctuation, the application of ArcGIS neighborhood analysis, the same selection of 5×5 statistical window, get the surface cutting depth. The results are shown in Table 1 and Figure 2 (c) below.

(4) Surface roughness is the ratio of the spatial curvature area of the surface element to the area mapped on the horizontal ground, which reflects the fluctuation degree of the surface [22]. The surface roughness is calculated according to Eq. (5), where $S$ is the slope factor. As shown in Table 1 and Figure 2 (d).

$$ R = \frac{1}{\cos(\frac{3.14}{180}S)} $$  

(5) The elevation variation coefficient is the ratio of elevation standard difference to average in the region [23], which is used to indicate the size of ground elevation change and ground fluctuation frequency. Similar to terrain fluctuation, the application of ArcGIS neighborhood analysis, the same selection of 5×5 statistical window, to obtain elevation coefficient of variation. According to the formula (6), the VCE is the elevation coefficient of variation, $S$ is the elevation standard deviation, and the mean represents the average elevation. The calculation results are shown in Table 1 and Figure 2 (e).

$$ VCE = \frac{S}{\text{mean}} $$  

(6) The profile curvature is the derivative of the elevation change rate of the slope along the maximum direction of slope. We import DEM data into ArcGIS and extract profile curvature using the twice slope command. The calculation results are shown in Table 1 and Figure 2 (f).
Figure 2. Single factor evaluation classification chart.
Table 1. Single factor statistical table of terrain complexity evaluation.

| single-factor evaluation | attribute value | grid       | proportion |
|--------------------------|-----------------|------------|------------|
| Slope                    |                 |            |            |
| 0-5.20                   | 1470867         | 0.32       |
| 5.20-10.60               | 1269060         | 0.28       |
| 10.60-16.63              | 963706          | 0.21       |
| 16.63-24.12              | 602472          | 0.13       |
| 24.12-53.02              | 222177          | 0.05       |
| Terrain Undulation       |                 |            |            |
| 0-15.19                  | 1526438         | 0.34       |
| 15.19-29.70              | 1318571         | 0.29       |
| 29.70-47.00              | 977933          | 0.22       |
| 47.00-69.75              | 539601          | 0.12       |
| 69.75-176.11             | 165739          | 0.04       |
| Surface Cutting Depth    |                 |            |            |
| 0-7.39                   | 1595507         | 0.35       |
| 7.39-14.79               | 1294420         | 0.29       |
| 14.79-23.66              | 948236          | 0.21       |
| 23.66-35.49              | 528953          | 0.12       |
| 35.49-94.27              | 161166          | 0.04       |
| Surface Roughness        |                 |            |            |
| 1-1.02                   | 2786534         | 0.62       |
| 1.02-1.05                | 1080489         | 0.24       |
| 1.05-1.10                | 466601          | 0.10       |
| 1.10-1.18                | 160813          | 0.04       |
| 1.18-1.66                | 33845           | 0.01       |
| Elevation Variation Coefficient |                 |            |            |
| 0-0.01                   | 1121654         | 0.25       |
| 0.01-0.02                | 1473386         | 0.33       |
| 0.02-0.03                | 1119257         | 0.25       |
| 0.03-0.04                | 621646          | 0.14       |
| 0.04-0.14                | 192339          | 0.04       |
| Profile Curvature        |                 |            |            |
| 0-2.04                   | 1657875         | 0.37       |
| 2.04-4.33                | 1430947         | 0.32       |
| 4.33-7.12                | 896840          | 0.20       |
| 7.12-11.07               | 425362          | 0.09       |
| 11.07-32.44              | 117258          | 0.03       |

3.4. Evaluation and results of terrain complexity in Ningdu County.
According to the natural break point grading method [2], each topographic factor is divided into 5 levels of very low, low, medium, high and very high complexity, which are assigned to 1, 3, 5, 7 and 9 respectively according to its grade. The grading results, as shown in Table 2, count the proportion of topographic factor area under each complexity, and the results are shown in Table 3. After evaluating 6 single factors, K mean clustering method is used to evaluate the topography complexity of Ningdu County.
Table 2. Classification of terrain complexity factor.

| Topographic complexity factor grading | very low complexity | low complexity | medium complexity | high complexity | Very high complexity |
|---------------------------------------|--------------------|----------------|-------------------|----------------|-------------------|
| Valuation                             |                    |                |                   |                |                   |
| Slope                                 | 0-5.20             | 5.20-10.60     | 10.60-16.63       | 16.63-24.12    | 24.12-53.02       |
| Terrain Undulation                    | 0-15.19            | 15.19-29.70    | 29.70-47.00       | 47.00-69.75    | 69.75-176.11      |
| Surface Cutting Depth                 | 0-7.39             | 7.39-14.79     | 14.79-23.66       | 23.66-35.49    | 35.49-94.27       |
| Surface Roughness                     | 1-1.02             | 1.02-1.05      | 1.05-1.10         | 1.10-1.18      | 1.18-1.66         |
| Elevation Variation Coefficient       | 0-0.01             | 0.01-0.02      | 0.02-0.03         | 0.03-0.04      | 0.04-0.14         |
| Profile Curvature                     | 0-2.04             | 2.04-4.33      | 4.33-7.12         | 7.12-11.07     | 11.07-32.44       |

Table 3. Analysis of terrain complexity factors.

| Topographic complexity factor grading | very low complexity | low complexity | medium complexity | high complexity | Very high complexity |
|---------------------------------------|--------------------|----------------|-------------------|----------------|-------------------|
| Slope                                 | 32.48%             | 28.03%         | 21.28%            | 13.30%         | 4.91%             |
| Terrain Undulation                    | 33.71%             | 29.12%         | 21.60%            | 11.92%         | 3.66%             |
| Surface Cutting Depth                 | 35.23%             | 28.59%         | 20.94%            | 11.68%         | 3.56%             |
| Surface Roughness                     | 61.54%             | 23.86%         | 10.30%            | 3.55%          | 0.75%             |
| Elevation Variation Coefficient       | 24.77%             | 32.54%         | 24.72%            | 13.73%         | 4.25%             |
| Profile Curvature                     | 115.86%            | 31.60%         | 19.81%            | 9.39%          | 2.59%             |

3.4.1. Evaluate terrain complexity by k-means clustering. Because the sample dataset is large, the number of iterations is set at 100. The initial clustering center and final clustering center are respectively shown in Table 4 and Table 5. The final result is graded according to its cluster center size, the 1st class is very low complexity, the 2nd class is extremely high complexity, the 3rd class is low complexity, the 4th class is medium complexity, and the 5th class is high complexity, as shown in Table 6.

Table 4. Initial clustering center of local shape factors.

| Initial clustering center | 1   | 2   | 3   | 4   | 5   |
|---------------------------|-----|-----|-----|-----|-----|
| Slope                     | 0.00| 44.50|8.12 |24.02|44.74|
| Terrain Undulation        | 0.00|175.27|53.83|79.07|126.63|
| Surface Cutting Depth     | 0.00| 90.78| 8.96|58.38| 59.11|
| Surface Roughness         | 1.00| 1.40 | 1.01|1.09 |1.41 |
| Elevation Variation Coefficient|0.00|0.04 |0.05 |0.03 |0.06 |
| Profile Curvature         | 0.00| 6.48 |20.01|25.19| 0.73 |
Table 5. The final clustering center of local shape factors.

| final clustering center | 1     | 2     | 3     | 4     | 5     |
|-------------------------|-------|-------|-------|-------|-------|
| Slope                   | 3.03  | 27.42 | 8.07  | 13.76 | 19.94 |
| Terrain Undulation      | 3.95  | 40.80 | 10.40 | 17.92 | 27.14 |
| Surface Cutting Depth   | 1.00  | 1.13  | 1.01  | 1.03  | 1.07  |
| Surface Roughness       | 0.01  | 0.04  | 0.02  | 0.03  | 0.03  |
| Elevation Variation Coefficient | 1.72 | 5.72  | 3.83  | 4.84  | 5.31  |

3.4.2. Analysis of comprehensive evaluation results of terrain complexity. As the Table 6 shows, the areas with very low complex terrain and low complex terrain are large in Ningdu county, accounting for 33.28% and 28.35%, respectively; Areas with medium complex terrain account for 21.4% of the total area in the study area. High complex areas with high and very high complex terrain involve a small area of 12.82% and 4.15%, respectively. The above data shows that the terrain of Ningdu county is generally flat. Comprehensive analysis of Figure 3: Shodian Township, Dongshao Township and Xiao Bu Town, Cai Jiang Township, Tsing Tong Town and Cham Township and other places due to peak staggered and high slope, belong to the areas with very high complex terrain. The distribution range of high complex area is wide. Luofang, changsheng Towns and other southern regions are complex areas. The terrain complexity in southwestern Ningdu is very low, due to the existence of mountain valleys in this part of the area.

Table 6. Evaluation Result of Terrain Complexity.

| complex degree | Extremely low complexity | low complexity | medium complexity | high complexity | Very high complexity |
|----------------|--------------------------|----------------|-------------------|-----------------|---------------------|
| Cluster grid   | 1                         | 3              | 4                 | 5               | 2                   |
| proportion(%)  | 1506832                  | 1283662        | 968996            | 580668          | 188124              |
|                | 33.28%                    | 28.35%         | 21.40%            | 12.82%          | 4.15%               |

Figure 3. Classification of terrain complexity in Ningdu county.
4. Conclusion
Taking Ningdu county as research area and DEM as data source, the k-means clustering method is used to analyze and evaluate the terrain complexity synthetically. Some conclusions can be obtained as follows:

(1) The k-mean clustering method is an efficient modeling method, which can synthesize many kinds of terrain factors to evaluate regional terrain complexity.

(2) The terrain complexity maps in Ningdu county show that, the area with very low terrain complexity occupies the largest area, accounting for 33.28% of Ningdu county; the area with low terrain complexity accounts for 28.35% of the total area in Ningdu county; and the area with very high terrain complexity accounts for minimum proportion as 4.15%. The spatial differentiation rules of terrain complexity is in accordance with the actual topographic characteristics of Ningdu County.

Acknowledgments
This work was financially supported by National Natural Science Foundation of China (NO.41807285).

References
[1] L. U. Huaxing, Xuejun Liu, Guoan Tang. Terrain Complexity Assessment Based on Multivariate Analysis[J]. Journal of Mountain Science, 2012,
[2] Faming Huang, Chi Yao, Weiping Liu, et al. Landslide susceptibility assessment in the Nantian area of China: a comparison of frequency ratio model and support vector machine[J]. Geomatics, Natural Hazards and Risk, 2018, 9(1): 919-938.
[3] Faming Huang, Kunlong Yin, Jinsong Huang, et al. Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine[J]. Engineering Geology, 2017, 223: 11-22.
[4] Faming Huang, Jinsong Huang, Shuihua Jiang, et al. Landslide displacement prediction based on multivariate chaotic model and extreme learning machine[J]. Engineering Geology, 2017, 218: 173-186.
[5] Weiping Liu, Xiaoyan Luo, Faming Huang, et al. Uncertainty of the Soil – Water Characteristic Curve and Its Effects on Slope Seepage and Stability Analysis under Conditions of Rainfall Using the Markov Chain Monte Carlo Method[J]. Water, 2017, 9(10): 758.
[6] Faming Huang, Xiaoyan Luo, Weiping Liu. Stability Analysis of Hydrodynamic Pressure Landslides with Different Permeability Coefficients Affected by Reservoir Water Level Fluctuations and Rainstorms[J]. Water, 2017, 9(7): 450.
[7] FM Huang, P Wu, YY Ziggh. GPS Monitoring Landslide Deformation Signal Processing using Time-series Model[J]. International Journal of Signal Processing, Image Processing and Pattern Recognition, 2016, 9(3): 321-332.
[8] I. S. Evans, N. J. Cox. Relations between land surface properties: Altitude, slope and curvature[M]. 1999.
[9] Yuanyao Li, Jinsong Huang, Shui-Hua Jiang, et al. A web-based GPS system for displacement monitoring and failure mechanism analysis of reservoir landslide[J]. Scientific reports, 2017, 7(1): 17171.
[10] Faming Huang, Jinsong Huang, Shui-Hua Jiang, et al. Prediction of groundwater levels using evidence of chaos and support vector machine[J]. Journal of Hydroinformatics, 2017, 19(4): 586-606.
[11] Ji Gui Sun, Jie Liu, Lian Yu Zhao. Clustering Algorithms Research[J]. Journal of Software, 2008, 19(19)
[12] Chuan Wang, Y. U. Lei. Investigation on Single Phase Water Reverse Flow in Inverted U-Tubes of Steam Generator under Condition of Natural Circulation[J]. Nuclear Power Engineering, 2011, 32(1): 58-62.
[13] Liping Jing, Michael K. Ng, Joshua Zhexue Huang. An Entropy Weighting k-Means Algorithm
for Subspace Clustering of High-Dimensional Sparse Data[J]. IEEE Transactions on Knowledge & Data Engineering, 2007, 19(8): 1026-1041.

[14] Huaxing Bi, Xiaoyin Li, Mengxia Guo, et al. Digital Terrain Analysis Based on DEM[J]. Journal of Beijing Forestry University, 2005, 1(1): 54-58.

[15] Faming Huang, Lixia Chen, Kunlong Yin, et al. Object-oriented change detection and damage assessment using high-resolution remote sensing images, Tangjiao Landslide, Three Gorges Reservoir, China[J]. Environmental Earth Sciences, 2018, 77(5): 183.

[16] Faming Huang, Kunlong Yin, Tao He, et al. Influencing factor analysis and displacement prediction in reservoir landslides—a case study of Three Gorges Reservoir (China)[J]. Tehnički vjesnik, 2016, 23(2): 617-626.

[17] Faming Huang, Kunlong Yin, Guirong Zhang, et al. Landslide displacement prediction using discrete wavelet transform and extreme learning machine based on chaos theory[J]. Environmental Earth Sciences, 2016, 20(75): 1-18.

[18] Binh Thai Pham, Dieu Tien Bui, Indra Prakash, et al. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS[J]. Catena, 2017, 149: 52-63.

[19] Tamotsu Matsui, Kaching San. FINITE ELEMENT SLOPE STABILITY ANALYSIS BY SHEAR STRENGTH REDUCTION TECHNIQUE[J]. Soils & Foundations, 2008, 32(1): 59-70.

[20] Hai Hui Han, Ting Gao, Y. I. Huan, et al. Extraction of Relief Amplitude Based on Change Point Method: A Case Study on the Tibetan Plateau[J]. Scientia Geographica Sinica, 2012, 32(1): 101-104.

[21] S. To, Y. H. Zhu, W. B. Lee. Effects of cutting depth on the surface microstructure of a Zn–Al alloy during ultra-precision machining[J]. Applied Surface Science, 2008, 254(6): 1559-1564.

[22] Chongya Jiang, Hongliang Fang, Shanshan Wei. Review of Land Surface Roughness Parameterization Study[J]. Advances in Earth Science, 2012, 27(3): 292-303.

[23] Shangmin Zhao, Weiming Cheng, Chenghu Zhou, et al. Accuracy assessment of the ASTER GDEM and SRTM3 DEM: an example in the Loess Plateau and North China Plain of China[J]. International Journal of Remote Sensing, 2011, 32(23): 8081-8093.