Article

Regional Terrain Complexity Assessment Based on Principal Component Analysis and Geographic Information System: A Case of Jiangxi Province, China

Faming Huang 1, Jianbo Yang 1, Biao Zhang 1, Yijing Li 1*, Jinsong Huang 2 and Na Chen 3

1 School of Civil Engineering and Architecture, Nanchang University, Nanchang 330031, China; faminghuang@ncu.edu.cn (F.H.); jiawu_chen@email.ncu.edu.cn (J.Y.); 416114417186@email.ncu.edu.cn (B.Z.)
2 ARC Centre of Excellence for Geotechnical Science and Engineering, University of Newcastle, Newcastle, NSW 2308, Australia; jinsong.huang@newcastle.edu.au
3 School of Civil Engineering and Architecture, Hubei University of Technology, Wuhan 430068, China; cn_research@hbut.edu.cn
* Correspondence: ejinn@ncu.edu.cn

Received: 13 August 2020; Accepted: 7 September 2020; Published: 8 September 2020

Abstract: Regional terrain complexity assessment (TCA) is an important theoretical foundation for geological feature identification, hydrological information extraction and land resources utilization. However, the previous TCA models have many disadvantages; for example, comprehensive consideration and redundancy information analysis of terrain factors is lacking, and the terrain complexity index is difficult to quantify. To overcome these drawbacks, a TCA model based on principal component analysis (PCA) and a geographic information system (GIS) is proposed. Taking Jiangxi province of China as an example, firstly, ten terrain factors are extracted using a digital elevation model (DEM) in GIS software. Secondly, PCA is used to analyze the information redundancy of these terrain factors and deal with data compression. Then, the comprehensive evaluation of the compressed terrain factors is conducted to obtain quantitative terrain complexity indexes and a terrain complexity map (TCM). Finally, the TCM produced by the PCA method is compared with those produced by the slope-only, the variation coefficient and K-means clustering models based on the topographic map drawn by the Bureau of Land and Resources of Jiangxi province. Meanwhile, the TCM is also verified by the actual three-dimensional aerial images. Results show that the correlation coefficients between the TCMs produced by the PCA, slope-only, variable coefficient and K-means clustering models and the local topographic map are 0.894, 0.763, 0.816 and 0.788, respectively. It is concluded that the TCM of the PCA method matches well with the actual field terrain features, and the PCA method can reflect the regional terrain complexity characteristics more comprehensively and accurately when compared to the other three methods.

Keywords: terrain complexity assessment; digital elevation model; principal component analysis; variation coefficient method; K-means clustering; geographic information system

1. Introduction

Regional terrain complexity assessment (TCA) is a meaningful foundation for identifying regional geological features, estimating groundwater level and river information and effectively utilizing the land resources in order to promote social and economic development [1–4]. Hence, it is significant to strengthen the TCA research and produce terrain complexity maps (TCM) in the areas with mountainous and hilly landforms [5–7].
The modeling of TCA includes multiple processes, such as acquisition of data sources, assessment unit division, selection of terrain factors and assessment model building. Recently, the technologies of remote sensing (RS) and geographic information systems (GIS) provide accurate and abundant data for regional TCA; therefore, the terrain factors used in this study are also acquired based on RS and GIS [8–11]. Furthermore, it is very important to select an appropriate model for TCA. In response to this issue, some scholars have proposed a lot of TCA-related models in recent years. Some of these models describe the regional terrain complexity only using a single terrain factor; for example, Tianwen, et al. [12] and Huaxing, et al. [13] selected slope angle as the terrain factor of TCA; Long et al. [14] used the fractal dimension value to describe the terrain complexity; Ashenfelter and Eberhard [15] adopted the angle between the space planes to describe the terrain complexity. These studies are unable to comprehensively reflect the terrain features of the study areas because they select limited terrain factors for TCA. On the other hand, some scholars used several types of terrain factors to determine the regional terrain features. For example, Chambers, et al. [16] pointed out that the regional differentiation laws of terrain complexity can be reflected by the synthesis calculations of terrain factors; Huaxing, Liu and Tang [13] explored the composite terrain factors based on a digital elevation model (DEM) to reflect the change characteristics of topographic relief and folds respectively from the holistic and local perspectives; Rawat and Joshi [17] studied the regional terrain characteristics in the land suitability classification; Thompson, et al. [18] analyzed the effects of DEM with different spatial resolutions on TCA and land landscape simulation.

In any case, there are still several problems in the previous models (variation coefficient method, k-means clustering method, etc.) which need to be solved, such as the need to introduce more terrain factors to TCA models, the lack of consideration of redundant information between terrain factors and the quantitative computation of the terrain complexity indexes. That is to say, the previous method cannot deal with the redundant information between the ten terrain factors. As a result, the TCA accuracy may be reduced. Furthermore, it is difficult for the previous methods to calculate the terrain complexity indexes quantitatively. Therefore, it is necessary and significant to propose a novel model which is appropriate for quantitatively and accurately analyzing a lot of terrain factors and can solve multifactor collinearity problems.

In this study, an excellent and widely acknowledged multivariate statistical method, namely principal component analysis (PCA), is proposed to work with TCA to overcome the drawbacks of the previous models. The PCA method can effectively calculate the component scores from the extracted principal components. Generally, the calculated component scores can be considered as the quantitative terrain complexity indexes, which are helpful to introduce regional topographic features into engineering applications. Then, the comprehensive assessment results can be accurately calculated by the PCA method [19]. The PCA model has been successfully used in various research fields [20–23]. For example, Li et al. [24] proposed the PCA and high-dimensional model representation techniques to estimate the probabilistic power flow; Geng, et al. [25] used the PCA based long short-term memory network to predict short-term wind speed time series; Schwartz, et al. [26] proposed a robust PCA method to conduct change detection in ultrawideband synthetic aperture radar images; Gao, et al. [27] researched network intrusion detection using extreme learning machine and adaptive PCA methods. However, the PCA method has received little attention in the field of regional TCA. Therefore, the PCA method is used to deal with the terrain factor compression and quantitative calculation of TCA.

The Jiangxi province of China is taken as the research area because the terrain of Jiangxi province is very complex, with types of high mountains and hills and so on. Moreover, complex terrain environments limit the local geological survey and land use planning. Hence, this paper firstly extracts several terrain factors using the DEM data in the GIS software and proposes the PCA method to calculate the terrain complexity indexes of Jiangxi province. Then, the influences of the extracted terrain factors on terrain complexity are explored. Moreover, the TCA results of PCA are compared with those of the slope-only method (considering only the slope map as the TCA map), variation coefficient method and k-means clustering model to discuss the modeling performance of PCA.
2. Materials and Methods

2.1. Introduction of Jiangxi Province and Data Sources

Jiangxi province is located in Eastern China, between 24°29′ N–30°04′41″ N and 113°34′36″ E–118°28′58″ E, with a length of 620 km from north to south and 490 km from east to west, covering an area of 166,900 km² (Figure 1). Jiangxi province is mainly hilly and mountainous, with extensive basins and valleys. This province is relatively flat in the northern part, surrounded by mountains on three sides in the eastern, western and southern parts, and is hilly in the middle part, making it a huge basin inclined toward Poyang Lake and opening to the north. The Ganjiang river, Fuhe river, Xinjiang river, Xiuhe river and Raohe river are the five major rivers that flow through Jiangxi province, and Poyang Lake is the largest freshwater lake in China. This province has a subtropical humid climate, with annual average rainfall ranging from 1341 to 1940 mm.

Figure 1. Aerial images of Jiangxi province.

2.2. Description of Terrain Factors

The data sources used for TCA modeling include (1) the topography data acquired through field survey; (2) the aerial images taken by the unmanned aerial vehicle on 1st February 2018 (freely downloaded from Google Earth 7.1.8.3036 (32-bit) with spatial resolution of 1.09 m); (3) DEM data with 100 m resolution taken on 11th February 2000 for extracting terrain factors (freely downloaded from http://www.gscloud.cn/) [28,29]; (4) the local topographic map with 1:250,000 scale. The Jiangxi province is divided into 16,713,615 grid cells with 100 m grid resolution—this is because this resolution can effectively characterize the topographical features of the study area, and it will not lead to too much computation as a result of mass grid cells [30].

In this study, ten terrain factors (including elevation (m), slope (°), plan curvature, profile curvature, relief amplitude (m), surface roughness, surface cutting depth (m), gully density (m/m²), elevation variation coefficient and slope length (m)) from the above data sources are extracted by the spatial analysis function of ArcGIS 10.2 software. All the ten terrain factors are respectively divided into five classes using the natural breakpoint method [31–33]. This is because the natural breakpoint method can effectively identify the frequency distribution characteristics of terrain factors to obtain the best
class division scheme, and it has been successfully used in the research of the class divisions by many other scholars [34–37]. The purpose of class division is to better represent the distribution rules of these terrain factors (Figure 2). In fact, the standardized results of terrain factors are the final input variables of all the TCA models.

(1) Elevation
The elevation [36,38] of the study area ranges from 2 to 2147 m. The mountains with high elevation mainly distribute around the Jiangxi province, while the central area of Jiangxi province has relatively low elevation. Generally speaking, the higher the elevation in the southern hilly area, the greater the terrain complexity (Figure 2a).

(2) Slope
According to the slope classification standard and the actual slope characteristics, the slopes of the study area are divided into five classes as flat slope, gentle slope, moderate slope, steep slope and acute slope [39–42]. The area with gentle slope covers around 33.88% of the total study area. The greater the slope, the more complex the terrain. To produce TCM based on the slope-only model, the flat slope, gentle slope, moderate slope, steep slope and acute slope in Jiangxi province are respectively considered as very low, low, moderate, high and very high complex terrain levels.

(3) Plan curvature and profile curvature
Plan curvature and profile curvature respectively reflect the change characteristics of concave and convex terrains from the horizontal direction and vertical direction [43]. The plan curvature (Figure 2b) can be determined by calculating the slope aspect of a slope map. Meanwhile, the profile curvature (Figure 2c) can be determined by calculating the slope twice on the basis of the DEM map [44]. In general, the greater the profile curvature or plan curvature, the higher the terrain complexity. However, compared with profile curvature, plan curvature has less correlation with terrain complexity.

(4) Relief amplitude
The relief amplitude of the study area can be obtained through the statistical test and the maximum height difference method. The greater the relief amplitude, the higher the terrain complexity [45]. According to Liu [46], the study area can be divided into five grades: plain region, tableland, hill, small rolling mountain and mountain. The results in Figure 2d show that the main terrain relief types of Jiangxi province are hill and small rolling mountain.

(5) Surface cutting depth
Surface cutting depth is defined as the difference between average elevation and minimum elevation in a certain area of one point on the ground surface [47]. In this study, the calculation radius of surface cutting depth is set to be 12 grid cells and the calculation results are shown in Figure 2e. Generally, a greater surface cutting depth means higher terrain complexity [48].

(6) Surface roughness
The surface roughness, which is an important index reflecting surface relief and soil erosion degree, is defined as the ratio of surface area to its projected area in a certain grid cell (Figure 2f). The surface roughness can be determined using the grid calculator of ArcGIS software according to Equation (1), where \( R_{ou} \) is the surface roughness value, and \( S \) is the slope value of the study area. The greater the surface roughness, the higher the terrain complexity [49].

\[
R_{ou} = \frac{1}{\cos(S \times 3.14159/180)}
\]  

(1)

(7) Gully density
The gully density (Figure 2g) is defined as the total length of the river network in a unit area [50,51]. The river network in a study area can be extracted by the hydrological analysis tool of ArcGIS 10.2 software. Then, we can calculate the gully density through the grid calculator, as shown in Equation (2), where \( D_S \) is the gully density, and \( L \) is length of the river networks in a unit area \( A \).

\[
D_S = \sum \frac{L}{A}
\]  

(2)
(8) Elevation variation coefficient

Elevation variation coefficient (Figure 2h) can be expressed as the ratio of the standard deviation of elevations to their average value in a certain area. This terrain factor is in direct proportion to the terrain complexity [52,53].

(9) Slope length

Slope length (Figure 2i) suggests the projected length of the maximum ground distance on a horizontal plane, from a point on the ground to its starting point in the direction of water flow [54,55]. Slope length affects terrain complexity indirectly and its formula is expressed in Equation (3), where $m$ indicates the flow length on the ground surface in the direction of water flow, and $S$ is the slope value of the ground surface.

$$L = m \times \cos(S)$$  \hspace{1cm} (3)

---

**Figure 2.** Terrain factors: elevation (a), plan curvature (b), profile curvature (c), relief amplitude (d), surface roughness (e), surface cutting depth (f), gully density (g), elevation variation coefficient (h), slope length (i).
2.3. Modeling Procedures of TCA

The modeling procedures of TCA using the PCA method are shown in Figure 3:
(1) All ten types of terrain factors are extracted and graded from the DEM data using spatial analysis tools and data management tools in ArcGIS 10.2 software.
(2) The PCA method is used to conduct quantitative calculation of topographic complexity indexes based on the correlation coefficient matrix of the above terrain factors. The principal components with cumulative contribution rates greater than 90% are adopted to calculate the component scores, which are considered the terrain complexity indexes.
(3) The slope-only, variation coefficient and k-means clustering methods are respectively used to assess the topographic complexity for comparisons. Their TCA results are also classified into five levels, namely very low, low, moderate, high and very high terrain complexity.
(4) The TCMs of Jiangxi province are produced in ArcGIS 10.2 by the above four methods.
(5) The accuracy of these TCMs is evaluated using the correlation coefficients between the TCMs and the topographic map and also evaluated through matching with the actual three-dimensional aerial images.

Figure 3. Flow chart of principal component analysis method for terrain complexity assessment.

2.4. Description of the Methods

2.4.1. Principal Component Analysis

The PCA method aims to use the idea of dimensionality reduction to transform multiple factors into a few principal components. Each principal component can reflect a piece of information of the original variables [56]. Furthermore, the contained information of each principal component is not repeated with that of other principal components. The PCA converts a given set of related variables into another set of unrelated variables through a linear transformation. Then, these new uncorrelated variables are arranged in descending order of variance. Initially, when using the PCA method, the dimensionality of all terrain factors should be unified. Supposing the original data set \( x_{ij}, i = 1, 2, \ldots, n; j = 1, 2, \ldots, m \) is composed by \( n \) grid cells with \( m \) types of terrain factors, the data set \( x_{ij} \) can be standardized by column as

\[
\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, m
\] (4)
\[ \sigma_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2} \]  

where \( x_{ij}^* \) is the standardized result of \( x_{ij} \), \( \bar{x}_j \) is the mean value of \( x_j \), and \( \sigma_j \) is the standard deviation of \( x_j \). The correlation coefficient matrix of \( x_{ij}^* \) is first calculated. Then, the eigenvalues \( \lambda_i \) and their corresponding eigenvectors are calculated based on the correlation coefficient matrix. The contribution rate of each eigenvalue can be obtained using Equation (6), where \( k \) is the number of eigenvalues.

\[ V_i = \frac{\lambda_i}{\sum_{i=1}^{k} \lambda_i} \]  

According to the eigenvalue and its eigenvector, the physical significance of the principal components is explained. According to the load coefficient value of the selected principal component in each terrain factor, the weight coefficient value of each terrain factor in the selected principal component is calculated by the following formula:

\[ a_{ij} = \frac{l_{ij}}{\sqrt{\lambda_i}} \]  

where \( a_{ij} \) and \( l_{ij} \) are respectively the weight coefficient value and load coefficient value of the \( j \)-th terrain factor in the \( i \)-th principal component; \( \lambda_i \) is the eigenvalue corresponding to the \( i \)-th principal component. Then, the comprehensive factor of each principal component can be calculated as

\[ F_i = \sum_{j=1}^{m} a_{ij} \times x_j \]  

where \( F_i \) is the comprehensive factor of all the terrain factors, and \( m \) is the number of terrain factors. Finally, the comprehensive scores of the selected principal components are obtained by Equation (9), where \( V \) is the total contribution rate of the selected principal components; \( T \) is the number of retained principal components; \( V_i \) is the contribution rate of the \( i \)-th principal component.

\[ F = \sum_{i=1}^{T} \frac{V_i}{V} \times F_i \]  

2.4.2. Variation Coefficient Method

The variation coefficient, which is an objective weighting method, directly uses the information contained in each factor to calculate the weight of the factor [57,58]. The basic idea of this method is that, in the evaluation factor system, the greater the difference of the values contained in a factor, the more the influence of a factor on the evaluated object. The differences between the values contained in a factor can be measured using the variation coefficient of this factor as

\[ VC_j = \frac{\sigma_j}{\bar{x}_j}, j = 1, 2, 3, \ldots, m \]  

where \( VC_j \) is the variation coefficient of the \( j \)-th factor, also known as the standard deviation coefficient; \( \sigma_j \) and \( \bar{x}_j \) are respectively the standard deviation and mean values of the \( j \)-th factor. Then, the weight of each factor is calculated as

\[ W_j = \frac{VC_j}{\sum_{j=1}^{m} VC_j} \]
2.4.3. K-Means Clustering Model

The K-means clustering model [59,60] firstly sets the number of clusters $K$ and the initial center point of each cluster and then updates the center point and optimizes the clustering results in an iterative way. The specific steps are as follows:

1. We set a training sample $S = \{x_1, x_2, \cdots, x_m\}$, where $x_m \in \mathbb{R}^n$ and $m$ represents the number of samples. $K$ clusters of $(SD_1, SD_2, \cdots, SD_K)$ are set as the output. Meanwhile, $K$ clustering center points are selected and denoted as $u_1, u_2, \cdots, u_K$.

2. Empty $K$ clusters in turn as $SD_j = \{\}$, $j = 1, 2, \cdots, k$. For each sample $x_p$ in the training set $S$, its distance to the $K$ center points is respectively calculated, and then it is classified into the corresponding cluster with the smallest distance to the center point of the cluster. In Equations (12) and (13), $C_p$ denotes the cluster which $x_p$ belongs to

$$C_p = \arg\min_j ||x_p - u_j||$$

$$SD_{C_p} = \{SD_{C_p}, x_p\}$$

3. For each cluster $SD_i$, its center point $\mu_i$ is recalculated according to Equations (14) and (15), where the $\text{sum}\_\text{feature}(SD_i)$ is the sum of the characteristic values of all samples in cluster $SD_i$, and the $\text{sum}\_\text{number}(SD_i)$ is the number of samples in cluster $SD_i$.

$$\mu_i = \frac{\text{sum}\_\text{feature}(SD_i)}{\text{sum}\_\text{number}(SD_i)} = \frac{\sum_{p=1}^{m} x_p SD_{i,p}}{\sum_{p=1}^{m} SD_{i,p}}$$

$$SD_{i,p} = \begin{cases} 1, & C_p = i \\ 0, & C_p \neq i \end{cases}$$

4. Repeat steps 2 and 3 until the cost function is less than the given threshold. Equation (16) is set as the cost function, which denotes the sum of Euclidean distances between all samples in the sample set $S$ and the center points of their corresponding clusters.

$$J(u, c) = \sum_{p=1}^{m} ||x_p - u_{C_p}||^2$$

3. Results

3.1. TCA Using Principal Component Analysis

3.1.1. Terrain Factor Analysis

The selected terrain factors are standardized, and then the Kaiser-Meyer-Olkin (KMO) test and Bartlett spherical test of the PCA method are performed based on the standardized data. Results show that the KMO test value of the standardized data is 0.790 (greater than 0.6) and its Bartlett spherical test value is less than 0.05, indicating that the PCA method is appropriate for dealing with the selected terrain factors. The PCA method is conducted on all the grid cells in Jiangxi province through the covariance matrix of the selected terrain factors. Then, the eigenvalues and explanatory contribution rates of all principal components are obtained, as shown in Table 1.
Table 1. Eigenvalues and explanatory contribution rates of all principal components.

<table>
<thead>
<tr>
<th>Principal Component</th>
<th>Eigenvalue</th>
<th>Explanatory Contribution Rate (%)</th>
<th>Cumulative Contribution Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.518</td>
<td>55.179</td>
<td>55.179</td>
</tr>
<tr>
<td>2</td>
<td>1.184</td>
<td>11.843</td>
<td>67.022</td>
</tr>
<tr>
<td>3</td>
<td>1.007</td>
<td>10.072</td>
<td>77.094</td>
</tr>
<tr>
<td>4</td>
<td>0.900</td>
<td>9.000</td>
<td>86.094</td>
</tr>
<tr>
<td>5</td>
<td>0.569</td>
<td>5.687</td>
<td>91.781</td>
</tr>
<tr>
<td>6</td>
<td>0.373</td>
<td>3.735</td>
<td>95.515</td>
</tr>
<tr>
<td>7</td>
<td>0.260</td>
<td>2.605</td>
<td>98.120</td>
</tr>
<tr>
<td>8</td>
<td>0.145</td>
<td>1.449</td>
<td>99.569</td>
</tr>
<tr>
<td>9</td>
<td>0.024</td>
<td>0.235</td>
<td>99.804</td>
</tr>
<tr>
<td>10</td>
<td>0.020</td>
<td>0.196</td>
<td>100</td>
</tr>
</tbody>
</table>

According to the results in Table 1, the principal components that can be used to describe the terrain factors are selected. Relevant studies suggest that the main information contained in primitive variables can be effectively interpreted by the corresponding principal components when their cumulative contribution rate reaches 90% [61]. It can be seen from Table 1 that the explanatory cumulative contribution rate of the first five principal components $F_1 \sim F_5$ can reach 91.78%. Therefore, the $F_1 \sim F_5$ (Table 2) are used to describe the terrain characteristics of Jiangxi province.

Table 2. Component matrix of the ten terrain factors.

<table>
<thead>
<tr>
<th>Terrain Factors</th>
<th>$F_{11}$</th>
<th>$F_{12}$</th>
<th>$F_{13}$</th>
<th>$F_{14}$</th>
<th>$F_{15}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>elevation ($E$)</td>
<td>0.780</td>
<td>-0.477</td>
<td>0.019</td>
<td>0.180</td>
<td>0.036</td>
</tr>
<tr>
<td>slope ($S$)</td>
<td>0.961</td>
<td>0.036</td>
<td>-0.085</td>
<td>0.066</td>
<td>-0.091</td>
</tr>
<tr>
<td>plan curvature ($Plc$)</td>
<td>-0.238</td>
<td>-0.107</td>
<td>0.927</td>
<td>0.198</td>
<td>-0.140</td>
</tr>
<tr>
<td>profile curvature ($Prc$)</td>
<td>0.724</td>
<td>0.044</td>
<td>0.251</td>
<td>-0.019</td>
<td>0.534</td>
</tr>
<tr>
<td>relief amplitude ($Rel$)</td>
<td>0.956</td>
<td>0.059</td>
<td>0.068</td>
<td>0.062</td>
<td>-0.130</td>
</tr>
<tr>
<td>surface roughness ($Rou$)</td>
<td>0.944</td>
<td>0.021</td>
<td>0.063</td>
<td>0.090</td>
<td>-0.147</td>
</tr>
<tr>
<td>surface cutting depth ($Scd$)</td>
<td>0.859</td>
<td>0.014</td>
<td>-0.131</td>
<td>0.159</td>
<td>-0.317</td>
</tr>
<tr>
<td>gully density ($Gd$)</td>
<td>-0.272</td>
<td>0.477</td>
<td>-0.127</td>
<td>0.812</td>
<td>0.134</td>
</tr>
<tr>
<td>elevation variation coefficient ($Evc$)</td>
<td>0.316</td>
<td>0.840</td>
<td>0.187</td>
<td>-0.335</td>
<td>-0.103</td>
</tr>
<tr>
<td>slope length ($L$)</td>
<td>0.829</td>
<td>0.067</td>
<td>0.001</td>
<td>-0.118</td>
<td>0.294</td>
</tr>
</tbody>
</table>

It can be seen from Tables 1 and 2 that the explanatory contribution rate of $F_{11}$ is 55.179%, and the maximum load in $F_1$ is $S$, with a load coefficient of 0.96, followed by the $Rel$ and $Rou$, with load coefficients of 0.956 and 0.944, respectively. Therefore, $F_1$ is mainly used to describe the slope and relief amplitude of the study area. The explanatory contribution rate of $F_2$ is 11.843%, $Evc$ is the maximum load in $F_2$, with a load coefficient of 0.840, and $Gd$ is the second maximum load in $F_2$, with a load coefficient of 0.477, indicating that $F_2$ is mainly used to describe the elevation variation coefficient and gully density of the study area. In addition, the explanatory contribution of $F_3$ is 10.072% and the maximum load in $F_3$ is $Plc$, with a load coefficient of 0.927, which suggests that $F_3$ is mainly used to describe the plan curvature. Based on Equation (8), the comprehensive factors of the selected five principal components $F_{11} \sim F_{15}$ are calculated as follows:

$$F_{11} = \frac{0.332 \times E + 0.409 \times S - 0.101 \times Plc + 0.308 \times Prc + 0.407 \times Rel + 0.366 \times Rou + 0.402 \times Scd - 0.116 \times Gd + 0.134 \times Ecv + 0.353 \times L}{0.020 \times Plc + 0.041 \times Prc + 0.055 \times Rel - 0.012 \times Rou + 0.020 \times Scd + 0.438 \times Gd + 0.772 \times Ecv + 0.062 \times L}$$ (17)

$$F_{12} = \frac{-0.438 \times E + 0.033 \times S - 0.099 \times Plc + 0.041 \times Prc + 0.055 \times Rel - 0.012 \times Rou + 0.020 \times Scd + 0.438 \times Gd + 0.772 \times Ecv + 0.062 \times L}{0.063 \times Scd - 0.126 \times Gd + 0.186 \times Ecv + 0.001 \times L}$$ (18)

$$F_{13} = \frac{0.018 \times E - 0.084 \times S + 0.924 \times Plc + 0.250 \times Prc + 0.068 \times Rel - 0.131 \times Rou + 0.063 \times Scd - 0.126 \times Gd + 0.186 \times Ecv + 0.001 \times L}{0.033 \times S - 0.099 \times Plc + 0.041 \times Prc + 0.055 \times Rel - 0.012 \times Rou + 0.020 \times Scd + 0.438 \times Gd + 0.772 \times Ecv + 0.062 \times L}$$ (19)
which can be mainly described as middle and low mountains supplemented by hilly areas. Meanwhile, the terrain complexity of Jiangxi province is at a low and middle level on the whole. The overall distribution rule of both TCMs is consistent. The TCA produced by the slope-only method is shown in Figure 4b for comparison, suggesting that the TCA results of Jiangxi province using the PCA method are shown in Figure 4a. The areas with very high and high terrain complexity levels are mainly distributed in the surrounding areas of Jiangxi province with continuous mountain ranges, accounting for 1.5% and 6.0% of the total area, respectively. These very high and high terrain complexity indexes are mainly affected by the great values of slope, elevation, relief amplitude, surface cutting depth, surface roughness and profile curvature. On the contrary, the very low and low complexity regions account for the largest area (36.6% and 40.7%, respectively) and are mainly distributed in central parts of Jiangxi province with small values of slope, elevation, relief amplitude, profile curvature and gully density. Furthermore, the moderate terrain complex areas account for around 36.64%, 40.72%, 15.14%, 6.04% and 1.46% of the total area, respectively.

3.1.2. Terrain Complexity Assessment

According to Equation (9) and the five principal components $F_{i1} \sim F_{i5}$, the terrain complexity index $F$ of each grid cell in Jiangxi province can be calculated as

$$F_i = 0.601 \times F_{i1} + 0.129 \times F_{i2} + 0.110 \times F_{i3} + 0.098 \times F_{i4} + 0.062 \times F_{i5}$$

The TCA results of Jiangxi province using the PCA method are shown in Figure 4a. The areas with very high and high terrain complexity levels are mainly distributed in the surrounding areas of Jiangxi province with continuous mountain ranges, accounting for 1.5% and 6.0% of the total area, respectively. These very high and high terrain complexity indexes are mainly affected by the great values of slope, elevation, relief amplitude, surface cutting depth, surface roughness and profile curvature. On the contrary, the very low and low complexity regions account for the largest area (36.6% and 40.7%, respectively) and are mainly distributed in central parts of Jiangxi province with small values of slope, elevation, relief amplitude, profile curvature and gully density. Furthermore, the moderate terrain complex areas account for 15.1% of the total area and are evenly distributed in the study area. The terrain complexity of Jiangxi province is at a low and middle level on the whole. The overall terrain complexity of the study area is in line with its topographic and geomorphic characteristics, which can be mainly described as middle and low mountains supplemented by hilly areas. Meanwhile, the TCA produced by the slope-only method is shown in Figure 4b for comparison, suggesting that the overall distribution rule of both TCMs is consistent.

![Figure 4. TCMs produced by the PCA (a) and slope-only (b) methods.](image)

### 3.2. TCA Using Variation Coefficient Method

According to the selected terrain factors, the variable coefficient of each terrain factor is determined. In addition, the weight values of all terrain factors are calculated based on their variable coefficients, as shown in Table 3. Then, these weight values are put into Equation (11) to calculate the terrain

$$F_{i4} = 0.190 \times E + 0.070 \times S + 0.209 \times Plc - 0.020 \times Prc + 0.066 \times Rel + 0.168 \times Rou + 0.094 \times Scd + 0.856 \times Gd - 0.354 \times Ecv - 0.124 \times L$$

$$F_{i5} = 0.048 \times E - 0.120 \times S - 0.186 \times Plc + 0.708 \times Prc - 0.173 \times Rel - 0.421 \times Rou - 0.195 \times Scd + 0.178 \times Gd - 0.1377 \times Ecv + 0.390 \times L$$
complexity indexes of Jiangxi province. The produced TCA results are shown in Figure 5a. It can be seen from Table 5 that the very low, low, moderate, high and very high terrain complexity areas account for around 36.64%, 40.72%, 15.14%, 6.04% and 1.46% of the total area, respectively.

<table>
<thead>
<tr>
<th>Terrain Factor</th>
<th>E</th>
<th>S</th>
<th>Plc</th>
<th>Prc</th>
<th>Rel</th>
<th>Rou</th>
<th>Scd</th>
<th>Gd</th>
<th>Ecv</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable coefficient</td>
<td>0.923</td>
<td>0.967</td>
<td>0.680</td>
<td>0.981</td>
<td>0.920</td>
<td>0.964</td>
<td>0.030</td>
<td>0.577</td>
<td>0.701</td>
<td>0.232</td>
</tr>
<tr>
<td>Weight value</td>
<td>0.132</td>
<td>0.139</td>
<td>0.097</td>
<td>0.141</td>
<td>0.132</td>
<td>0.138</td>
<td>0.004</td>
<td>0.083</td>
<td>0.101</td>
<td>0.033</td>
</tr>
</tbody>
</table>

### 3.3. TCA Using K-Means Clustering Model

The selected standardized terrain factors are put into the k-means clustering model. First, the number of clusters is set to 5, the clustering center point is randomly determined, and the number of iterations is set to 25. Then, each grid cell with ten standardized terrain factors is classified into the corresponding cluster with the smallest distance to the final center point of the cluster through 25 iterative calculation. The five final center points of the clusters are shown in Table 4. Finally, all the grid cells are classified into five types of different terrain complexity levels, as shown in Figure 5b and Table 5.

<table>
<thead>
<tr>
<th>Terrain Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>elevation (E)</td>
<td>2.122</td>
<td>−0.489</td>
<td>−0.654</td>
<td>0.444</td>
<td>1.171</td>
</tr>
<tr>
<td>slope (S)</td>
<td>2.855</td>
<td>0.138</td>
<td>−0.801</td>
<td>0.190</td>
<td>1.462</td>
</tr>
<tr>
<td>plan curvature (Plc)</td>
<td>−0.706</td>
<td>−0.054</td>
<td>0.174</td>
<td>0.016</td>
<td>−0.391</td>
</tr>
<tr>
<td>profile curvature (Prc)</td>
<td>1.215</td>
<td>0.281</td>
<td>−0.681</td>
<td>0.305</td>
<td>1.095</td>
</tr>
<tr>
<td>relief amplitude (Rel)</td>
<td>2.752</td>
<td>0.240</td>
<td>−0.794</td>
<td>0.168</td>
<td>1.424</td>
</tr>
<tr>
<td>surface roughness (Rou)</td>
<td>2.765</td>
<td>0.146</td>
<td>−0.769</td>
<td>0.162</td>
<td>1.429</td>
</tr>
<tr>
<td>surface cutting depth (Scd)</td>
<td>3.940</td>
<td>−0.109</td>
<td>−0.548</td>
<td>−0.116</td>
<td>1.212</td>
</tr>
<tr>
<td>gully density (Gd)</td>
<td>0.407</td>
<td>1.745</td>
<td>−0.410</td>
<td>−0.251</td>
<td>0.284</td>
</tr>
<tr>
<td>elevation variation coefficient (Evc)</td>
<td>−0.372</td>
<td>0.322</td>
<td>0.259</td>
<td>−0.366</td>
<td>−0.301</td>
</tr>
<tr>
<td>slope length (L)</td>
<td>1.165</td>
<td>0.578</td>
<td>−0.979</td>
<td>0.749</td>
<td>0.959</td>
</tr>
</tbody>
</table>

Figure 5. TCMs produced by variation coefficient method (a) and K-means clustering model (b).
A greater correlation coefficient value means that the produced TCM and the topographic map matches better, which further suggests higher TCA accuracy.

In order to analyze the TCA accuracy of the PCA, slope-only, variable coefficient and k-means clustering models, the correlation coefficients between the TCMs produced by the four models and the local topographic map obtained from the government field surveys are calculated, respectively. A greater correlation coefficient value means that the produced TCM and the topographic map matches better, which further suggests higher TCA accuracy.

In this study, the topographic map with a scale of 1:250,000, surveyed by the Bureau of Land and Resources of Jiangxi province in 2015, is used to evaluate the TCA accuracy of the four models. In the topographic map of Jiangxi province, the topographic types are classified as five categories of plain, hill, low mountain, middle mountain and mountain, which are assigned values of 1, 2, 3, 4, 5, respectively (Figure 6). In order to compare the TCMs with the topographic map, the produced TCMs are classified into five terrain complexity levels of very low, low, moderate, high and very high, which are also assigned values of 1, 2, 3, 4, 5, respectively. Then, the correlation coefficients between the topographic map and the TCMs produced by the PCA, slope-only, variable coefficient and K-means clustering models are calculated as 0.894, 0.763, 0.816 and 0.788, respectively. The results show that the TCM produced by the PCA method has the highest correlation coefficient with the topographic map compared to those produced by the slope-only, variable coefficient and k-means clustering models. That is to say, the PCA model has much higher TCA accuracy than the other three types of models.

### Table 5. TCA results of the four models.

<table>
<thead>
<tr>
<th>TCA Level</th>
<th>Area/km²</th>
<th>Percentage/</th>
<th>Area/km²</th>
<th>Percentage/</th>
<th>Area/km²</th>
<th>Percentage/</th>
<th>Area/km²</th>
<th>Percentage/</th>
</tr>
</thead>
<tbody>
<tr>
<td>very low</td>
<td>61,156</td>
<td>36.64</td>
<td>83,442</td>
<td>50.00</td>
<td>66,395</td>
<td>39.78</td>
<td>74,493</td>
<td>44.63</td>
</tr>
<tr>
<td>low</td>
<td>67,954</td>
<td>40.72</td>
<td>56,538</td>
<td>33.88</td>
<td>48,384</td>
<td>28.99</td>
<td>19,004</td>
<td>11.39</td>
</tr>
<tr>
<td>moderate</td>
<td>25,274</td>
<td>15.14</td>
<td>22,409</td>
<td>13.43</td>
<td>32,383</td>
<td>19.40</td>
<td>45,098</td>
<td>27.02</td>
</tr>
<tr>
<td>high</td>
<td>10,078</td>
<td>6.04</td>
<td>4271</td>
<td>2.56</td>
<td>15,495</td>
<td>9.28</td>
<td>23,247</td>
<td>13.93</td>
</tr>
<tr>
<td>Very high</td>
<td>2438</td>
<td>1.46</td>
<td>240</td>
<td>0.14</td>
<td>4243</td>
<td>2.54</td>
<td>5058</td>
<td>3.03</td>
</tr>
</tbody>
</table>

#### 4. Discussion

##### 4.1. Accuracy Assessment of the Four Models

In order to analyze the TCA accuracy of the PCA, slope-only, variable coefficient and k-means clustering models, the correlation coefficients between the TCMs produced by the four models and the local topographic map obtained from the government field surveys are calculated, respectively. A greater correlation coefficient value means that the produced TCM and the topographic map matches better, which further suggests higher TCA accuracy.

In this study, the topographic map with a scale of 1:250,000, surveyed by the Bureau of Land and Resources of Jiangxi province in 2015, is used to evaluate the TCA accuracy of the four models. In the topographic map of Jiangxi province, the topographic types are classified as five categories of plain, hill, low mountain, middle mountain and mountain, which are assigned values of 1, 2, 3, 4, 5, respectively (Figure 6). In order to compare the TCMs with the topographic map, the produced TCMs are classified into five terrain complexity levels of very low, low, moderate, high and very high, which are also assigned values of 1, 2, 3, 4, 5, respectively. Then, the correlation coefficients between the topographic map and the TCMs produced by the PCA, slope-only, variable coefficient and K-means clustering models are calculated as 0.894, 0.763, 0.816 and 0.788, respectively. The results show that the TCM produced by the PCA method has the highest correlation coefficient with the topographic map compared to those produced by the slope-only, variable coefficient and k-means clustering models. That is to say, the PCA model has much higher TCA accuracy than the other three types of models.
4.2. TCA Compared with Three-Dimensional Aerial Images

In order to more intuitively reflect the consistency between the TCA performance and the actual terrain features, the TCM produced by the PCA method is compared with three-dimensional aerial images of the study area. Eight locations in the TCM are randomly selected, and the three-dimensional aerial images of these locations (Figure 7a–h) are acquired correspondingly. Taking Figure 7a–e as examples,

1. The topographic complexity levels of Figure 7a,d are mainly high and very high, which is highly consistent with the three-dimensional aerial images of Figure 7a,d. Generally, the topography types of Figure 7a,d are dominated by mountains and valleys.

2. The topographic complexity levels in both Figure 7b,c are low and the corresponding elevation variations are small. Meanwhile, the topographic features shown in the three-dimensional aerial images of Figure 7b,c are plain and hill. Hence, the matching degree between the TCA result and the three-dimensional aerial image is very high.

3. A mixture of several kinds of topographic complexity levels is shown in Figure 7e, which is also very consistent with the topographic features of rolling hills and low mountains shown in the aerial image of Figure 7e. The above analysis shows that the TCA accuracy of the PCA method is high and the terrain complexity maps match well with the actual three-dimensional aerial images.

Figure 7. Topographic complexity map and its corresponding three-dimensional aerial images.
4.3. Limitations and Research Prospects

Although the PCA method used in this study shows better TCA performance than the other three types of methods, there are still some limitations in the processes of PCA modeling. Firstly, it is necessary to further test the accuracy of the original DEM data and strive to obtain DEM data with higher accuracy on this basis. Secondly, the DEM spatial resolution of 100 m is relatively rough; hence, DEM data with higher spatial resolution is needed in order to implement TCA modeling in further research. Thirdly, only ten kinds of terrain factors are extracted from the DEM data in this study; it is significant to extract more terrain factors from DEM data, such as ridge lines, valley lines, slope forms, etc. In the next study, more terrain factors will be introduced as the input variables of the PCA method. Fourthly, the main features extracted by the PCA are the linear correlations among all these terrain factors, and the subsequent research will extract more nonlinear features among the terrain factors. Therefore, kernel principal component analysis \[62\] or other nonlinear data compression methods can be considered in order to carry out this study. Finally, the accuracy of terrain complexity assessment should be analyzed and verified by combining it with field investigation in the next research, and the engineering applications of TCA results in various fields should be explored to enhance the research significance of this study.

In addition, for the use of the PCA method, all the data sets are put into the PCA method to calculate terrain complexity indexes. However, there is a lack of statistical validation of the results, as a result, an estimate of the goodness of the model itself is not present in this study. Hence, compared to the unsupervised models used in this study, some supervised mathematical models will be used to implement TCA in the next study. For the supervised mathematical models, some of the data sets are used to create a model and the remaining are used to give an estimate of the goodness of the model itself. Furthermore, it is necessary to make sure that the accuracy of the description of the territory using only the new variables is homogeneous over the whole territory. Hence, in the next study, we need to repeat the construction of the models many times using each time a subset of the available data (typically two or three of the datasets) extracted randomly.

5. Conclusions

There are some disadvantages in the previous TCA studies, such as few terrain factors used for TCA modeling, the lack of information compression between terrain factors and the qualitative TCA modeling processes. In order to overcome these disadvantages, a PCA method is introduced to carry out TCA in this study. The Jiangxi province of China with complex terrain features is used as the study area. Ten terrain factors including elevation, slope, plan curvature, profile curvature, relief amplitude, surface roughness, surface cutting depth, gully density, elevation variation coefficient and slope length are selected as input variables of PCA. The TCM produced by PCA is compared with the results of slope-only, variation coefficient and K-means clustering models and is also verified by aerial images obtained by field survey.

It can be concluded that the PCA method can accurately and reliably describe the terrain complexity distribution features; the produced TCM is very consistent with the actual terrain features. In addition, the proposed PCA method has considerably higher accuracy than the other three models. The main contributions of this study include the comprehensive consideration of ten related terrain factors to address TCA, the removal of redundant information in these terrain factors and the quantitative calculation of terrain complexity indexes. Hence, the TCM produced in this study can be applied to improve the local geological investigation and land planning utilization.

Author Contributions: Conceptualization: Faming Huang, Yijing Li and Jianbo Yang; methodology: Faming Huang and Jianbo Yang; software: Jianbo Yang; validation: Jinsong Huang, and Biao Zhang; formal analysis: Na Chen; investigation: Jianbo Yang; resources: Faming Huang; data curation: Yijing Li; writing—original draft preparation: Faming Huang, Jianbo Yang, and Yijing Li; writing—review and editing: Faming Huang and Jinsong Huang; visualization: Biao Zhang; supervision: Yijing Li; project administration: Faming Huang;
funding acquisition: Faming Huang and Yijing Li. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the National Natural Science Foundation of China (Nos. 41807285, 41972280, 40501454 and 41501454), National Science Foundation of Jiangxi Province of China (NO. 20192BAB216034), the Postdoctoral Science Foundation of Jiangxi Province of China (NO. 2019KY08), the China Postdoctoral Science Foundation (Nos. 2019M652287 and 2020T130274) and Science and Technology Project of Water Resources Department of Jiangxi Province (NO. KT201544).

**Acknowledgments:** This study was supported by the National Natural Science Foundation of China (Nos. 41807285, 41972280, 40501454 and 41501454), National Science Foundation of Jiangxi Province of China (No. 20192BAB216034), the Postdoctoral Science Foundation of Jiangxi Province of China (NO. 2019KY08), the China Postdoctoral Science Foundation (Nos. 2019M652287 and 2020T130274) and Science and Technology Project of Water Resources Department of Jiangxi Province (NO. KT201544).

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

**References**

6. Farmakis-Serebryakova, M.; Hurni, L. Comparison of relief shading techniques applied to landforms. *ISPRS Int. J. Geo-Inf.* 2020, 9, 253. [CrossRef]
11. Prasannakumar, V.; Vijith, H.; Geetha, N. Terrain evaluation through the assessment of geomorphometric parameters using DEM and GIS: Case study of two major sub-watersheds in Attapady, South India. *Arab. J. GeoSci.* 2013, 6, 1141–1151. [CrossRef]
18. Thompson, J.A.; Bell, J.C.; Butler, C.A. Digital elevation model resolution: Effects on terrain attribute calculation and quantitative soil-landscape modeling. *Geoderma* 2015, 100, 67–89. [CrossRef]


27. Gao, J.; Chai, S.; Zhang, B.; Xia, Y. Research on network intrusion detection based on incremental extreme learning machine and adaptive principal component analysis. *Energies* 2019, 12, 1223. [CrossRef]


34. Huang, F.; Cao, Z.; Guo, J.; Jiang, S.-H.; Li, S.; Guo, Z. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *CATENA* 2020, 191, 104580. [CrossRef]


© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).