

Environmental factors of road slope stability in mountain area using principal component analysis and hierarchy cluster

Huiqin He · Shaocai Li · Hailong Sun ·
Ting Yang

Received: 2 May 2009 / Accepted: 6 February 2010 / Published online: 9 March 2010
© Springer-Verlag 2010

Abstract Nine environmental factors of 147 roadside soil samples were administered in Sichuan Basin of China and principal component analysis was conducted using the Pearson correlation matrix. The results show that the first four principal components whose eigenvalue is over 1.00 can be extracted. The first principal component which is consisted of rock type, soil type, weathering degree, and soil depth is the most important factor of all. The geographical position which is consisted of altitude, longitude, and latitude is included in the second and the third principal components. The fourth principal component shows that the terrain factor influences the rock slope stability. The hierarchy cluster shows that rock type and soil type play the maximum positive correlation, while the slope and the aspect present the maximum negative correlation.

Keywords Environment variables · Rock roadside · Mountain region · Hierarchy cluster · PCA

Introduction

As the vegetation restoration is becoming stronger in rock slope stability and roadside landscape restoration, it is taken on as the effective measure in controlling the erosion and

stabling on rock slope (Tinker et al. 1998). The rock slope stability is influenced by many factors, including the geological structure, development situation, soil property, slope status, and the geographical position, etc. For the high correlations of variables, it is difficult to decide the most important one and how it influences the slope stability degree. Yet the roadside slope is a crucial question of ecological restoration. Principal component analysis (PCA) is the most important part of multivariate statistical analysis. By compressing the variable numbers and diminishing the co-linearity of original data, it can transform the highly correlated variables into just fewer variables (Thorpe 1988). The basic approach of PCA is to compute the eigenvectors of the covariance matrix and approximate the original data by a linear combination of the leading eigenvectors (Annoni 2007). This has been proved to be effective in overcoming the instability and ill-conditioned matrix in structural analysis, and lead to the enough information of the original variables and accurate results reflected by the model (Sousa et al. 2007). PCA has been widely used in social sciences such as astronomy (Ronen et al. 1998), geography, physics, chemistry, and life sciences (Fievez et al. 2003), and great progress has been made in data compression, image processing (Calder et al. 2001), visual-isolation, exploratory data analysis, pattern recognition (Doty et al. 1994; Guo et al. 2004) and time series prediction (Berrar et al. 2003). In environmental field, PCA is used to analyze the relationship between fertilizer and biomass (Morillo et al. 2008; Seabloom et al. 2003); some are concentrated on the relationship between vegetation and environment (O'Lenic and Liverzey 1988; Inger et al. 2008; White and Hood 2004), and species diversity and its site (ter Braak 1983). Recently, a few researches have investigated the stability of roadside slope in cognitive approaches (ter Braak 1989, Wiser et al. 1996).

H. He · S. Li (✉) · H. Sun
College of Life Sciences, Sichuan University,
29 Wangjiang Street, Chengdu 610064, Sichuan, China
e-mail: Lizist@vip.sina.com

H. He
e-mail: huiqinhe@yahoo.com.cn

H. He · T. Yang
Department of Life Sciences and Food Engineering, Yibin
University, 8 Jiusheng Road, Yibin 644000, Sichuan, China

In this paper, two different methods, PCA and hierarchy cluster, are applied to analyze the main factors influencing the rock slope stability in a large scale in Sichuan basin of southwest China. In addition, the models of the rock stability will reveal the relationships of the environmental factors from two aspects, qualitative and quantitative.

Site description

Sichuan basin is one of the four famous basins in China. Dominated by the faults trending in NE–SW and north east direction, it is divided into two particular parts, marginal mount and bottom basin, which cover about 260,000 km² with particular geomorphology including 7% plain, 52% hill, and 41% low mount. All sites are distributed in this area between latitudes 29°19'40"–32°15'34" and longitudes 102°57'52"–105°33'30" (Fig. 1). Field data were collected along the roadside slope with the altitude from 280 up to 1020 m above the sea level. Due to the westerly circulation and southwest monsoon, there has a subtropical climate. The annual average temperature is 17.5°C, with the lowest average monthly temperature ranging from 5°C to 8°C and the highest average monthly temperature ranging from 25°C to 29°C. The average annual rainfall in the area ranges from 1,000 to 1,300 mm and about 75% of the rainfall is concentrated from June to October. The natural vegetation belongs to the subtropical evergreen broad-leaved forest. The dominant rock types are red sandstone and shale, and the major soil type is the purple loam in this area.



Fig. 1 Spatial distribution diagram of the 147 roadside slope plots in Sichuan Basin, China

Objects and methods

147 samples from 80 sites of the rock roadsides around basin were established and studied in October, 2007. The site was set at least 15 m long along the road and 10 m high along the slope in order to meet to the minimum scale. The samples were set in the middle of the investigation area. Taking the community into account, grass sample was set 1 m × 1 m, 2 m × 2 m for shrub and 5 m × 5 m for tree, and 147 samples were settled in total. These variables including slope, aspect, and rock type, weathering degree, soil type, soil depth, altitude, latitude, and longitude were recorded and calculated in total. The methods and classification standards of the nine variables were listed as follows: Rock type was qualitative to three types according to the formation reason. Rock weathering degree was quantified by the ratio of weathering porosity, namely the ratio of inhaled water quality of weathered rock to dry weathering rock. The rock weathering degree was then divided into four types, complete weathering, strong weathering, weak weathering and slight weathering according to the ratio. Soil type was measured with the hydrometer, named by the soil texture classification of the ratio of physical sand to physical clay. The soil depth was measured with the earth auger from surface soil vertical downward to cane with the unit centimeter. When the soil thickness was thin, the soil profile was dug to be measured directly. Geography position and terrain were measured by portable global position system (GPS) and compass, respectively.

Data analysis

Principal component analysis

Nine variables of the 147 roadside slope plots were conducted using statistical package for social science (SPSS) version 16.0 for all calculation. The main steps were listed as follows: (1) Original data standardization. Owing to the data types, data formats and data units of nine variables were different and it was unreasonable to analyze them directly. The original data were standardized with the z-score transformation to the same scale, producing new variables with a mean of zero and a standard deviation of one. These new variables were independent linear combinations and retained the maximum possible variance of the original set, and could be added to the working data file for further analysis. (2) Coefficient matrix calculation and the significance level test. The Kaiser–Meyer–Olkin (KMO) test of sampling adequacy and Bartlett's test of sphericity were used to assess the appropriateness of the correlation matrices. (3) Eigenvalue, contribution rate, cumulative contribution rate and eigenvector. (4) Principle component

extraction, the construction of the extracted principal component models, and the slope stability evaluation model.

Hierarchy cluster

Hierarchical clustering referred to the formation of a recursive clustering of the data points: a partition into two clusters, each of which is itself hierarchically clustered. It can be used to evaluate the relationships between the environment stressors on the plant community effecting from the chemical and physical system (Lipkovich et al. 2008). The standardized variables were induced to hierarchy cluster based on the Pearson correlation. The cluster results were obtained according to their relative correlation.

Results

Correlation analysis

Based on the standardized variables, the correlation matrix of the 147 samples was measured. Before conducting principal component analysis, Barlett’s test of sphericity and KMO test of sampling adequacy were initially performed to confirming the appropriateness of conducting PCA (Sousa et al. 2007). The Bartlett’s test for sphericity showed that the correlation matrix was at an appropriate level to perform principal component analysis reaching a significance level of $p < 0.001$. The KMO measure provided a value between 0 and 1. Small value for the KMO indicated that a factor analysis of the variables may not be appropriate. Value higher than 0.5 was considered satisfactory for principal component analysis (Norusis 1990). In this paper, the KMO test was 0.726, which manifested that the samples were adequate to principal component analysis. Both of the two tests supported that the principal

component analysis were appropriate. Table 1 showed that many of the standardized variables were relatively well correlated with one another. Both negative correlation and positive correlation were observed, where the correlation coefficient between soil depth and soil type was the maximum 0.912.

Principal component analysis

The correlation matrix among the nine standardized variables was subjected to principal component analysis, a procedure that, although similar to the factor analysis, did not suffer from the factor in determinacy of factor analysis. The most common stopping rule in PCA was based on the average value of the eigenvalues >1.0 (i.e., the Kaiser-Guttman criterion; Guttman 1954; Cliff 1988; Jackson 1993). According to the rule, the first four principal components whose eigenvalues >1.00 were extracted which accounted for 75.552% of the standardized variance of the data. The eigenvalue of the first principal component was 2.782, and it explained 30.914% of the standardized variance, in which rock type, weathering degree, soil type, and soil depth were the major contributing variables. The eigenvalue of the second principal component was 1.851, and the latitude and longitude were the major factors with the contribution rate of 20.563%. The eigenvalue of the third principal component was 1.111, and altitude loaded the contribution rate 12.349%. Aspect and slope were the important variables of the fourth principal component, and the eigenvalue was 1.055 with the contribution rate 11.726%. The first four principal component whose eigenvalue >1.00 embraced all the standardized variables, which manifested the principal component analysis was effective in rock slope stability analysis. According to the component eigenvalue of the standardized variables, it was significant to name the first four principal components: the first principal component named as parent material factor,

Table 1 Correlation coefficient of the standardized variables

	Slope	Aspect	Rock type	Weathering degree	Soil type	Soil depth	Altitude	Latitude	Longitude
<i>Correlation Matrix</i>									
Correlation									
Slope	1.000								
Aspect	0.055	1.000							
Rock type	0.413	-0.058	1.000						
Weathering degree	-0.192	-0.011	-0.283	1.000					
Soil type	-0.214	0.097	-0.472	0.426	1.000				
Soil depth	-0.250	0.117	-0.538	0.476	0.912	1.000			
Altitude	-0.016	0.016	-0.080	0.146	-0.009	0.028	1.000		
Latitude	-0.036	-0.044	0.136	-0.035	-0.043	-0.080	0.478	1.000	
Longitude	-0.057	-0.104	0.109	-0.028	0.102	0.036	-0.031	0.692	1.000

and the second and third principal components named as geographical factor, and the fourth principal component named as terrain factor.

Based on the principal component definition, four linear combinations of the principal component eigenvectors of the nine standardized variables were obtained (Table 2). The eigenvectors of the principal component were calculated according to the coefficients of the standardized variables. Accordingly, the *z* score functions of the four extracted principal component were listed as follows:

$$F_1 = -0.279 \times Z_{x_1} + 0.071 \times Z_{x_2} - 0.439 \times Z_{x_3} + 0.377 \times Z_{x_4} + 0.525 \times Z_{x_5} + 0.549 \times Z_{x_6} + 0.041 \times Z_{x_7} - 0.068 \times Z_{x_8} - 0.014 \times Z_{x_9}$$

$$F_2 = -0.08 \times Z_{x_1} - 0.102 \times Z_{x_2} + 0.065 \times Z_{x_3} + 0.049 \times Z_{x_4} + 0.049 \times Z_{x_5} + 0.019 \times Z_{x_6} + 0.373 \times Z_{x_7} + 0.700 \times Z_{x_8} + 0.587 \times Z_{x_9}$$

$$F_3 = 0.217 \times Z_{x_1} + 0.464 \times Z_{x_2} - 0.036 \times Z_{x_3} + 0.147 \times Z_{x_4} - 0.100 \times Z_{x_5} - 0.039 \times Z_{x_6} + 0.712 \times Z_{x_7} + 0.078 \times Z_{x_8} - 0.435 \times Z_{x_9}$$

$$F_4 = 0.518 \times Z_{x_1} + 0.602 \times Z_{x_2} + 0.253 \times Z_{x_3} - 0.100 \times Z_{x_4} + 0.282 \times Z_{x_5} + 0.222 \times Z_{x_6} - 0.306 \times Z_{x_7} + 0.044 \times Z_{x_8} + 0.268 \times Z_{x_9}$$

where *Z* indicated the standardized variables, and *x* denoted the variables, and *i* denoted the order of the variables.

Taking the eigenvalues of the first four principal components into account, the weights were calculated as follows: 0.4092, 0.2722, 0.1634, and 0.1552. The slope stability evaluation model was expressed as

$$F = 0.4092 \times F_1 + 0.2722 \times F_2 + 0.1634 \times F_3 + 0.1552 \times F_4.$$

Hierarchy cluster

Based on the Pearson correlation of the standardized variables, hierarchy cluster was launched in SPSS 16.0. According to the hierarchy cluster analysis of the nine standardized variables, four positive correlations, and three negative correlations of Pearson correlations were observed. The correlation between rock type and soil type was the maximum positive correlation 0.680, while the slope and the aspect presented the maximum negative correlation -0.557. Nine standardized variables were classified into three categories: geographic location, parent material, and terrain factors from the rescaled distance 17.5–20.0. Category I included four variables, aspect, weathering degree, soil type, and soil depth. Category II included two variables, slope and rock type. The category III included altitude, latitude and longitude. The dendrogram using complete linkage of the hierarchy cluster is shown in detail in Fig. 2.

Discussion and conclusions

Principal component analysis has proved to be an exceedingly popular technique for dimensionality reduction (Tipping and Bishop 1999). As applied to the stability analysis of rock side slope, PCA serves a similar function: It identifies a limited number of factors that can represent the complex factor information in road side slope in a suitable form for slope stability. Two different methods were used to analyze the correlations of the standardized data. Principal component analysis extracted four principal components. The first principal component which was consisted of rock type, soil type, weathering degree, and soil depth was the most important factor of all whose cumulative contribute rate was 30.914%, which revealed the parent material factor influencing the roadside stability.

Table 2 Eigenvectors of the four components after standardization variables

	Component			
	1	2	3	4
Slope	-0.279	-0.080	0.217	0.518
Aspect	0.071	-0.102	0.464	0.602
Rock type	-0.439	0.065	-0.036	0.253
Weathering degree	0.377	0.049	0.147	-0.100
Soil type	0.525	0.049	-0.099	0.282
Soil depth	0.549	0.019	-0.039	0.222
Altitude	0.041	0.373	0.712	-0.306
Latitude	-0.068	0.699	0.078	0.044
Longitude	-0.014	0.587	-0.435	0.268

Four components extracted
Extraction method: principal component analysis

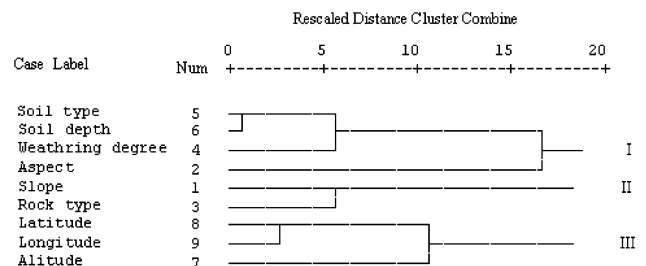


Fig. 2 Dendrogram of standardized variables using complete linkage of hierarchy cluster

In the second and third components, latitude, longitude, and altitude were important variables of geographical position. The fourth principal component revealed the terrain factors, slope, and aspect. Although four principal components were formed, they belong to three factors: material factor, geographical factor, and terrain factor. The conclusion was drawn that parent material factor was the most important component influencing the rock slope stability, and the geographical position was the second important factor.

Some differences exist between principal component analysis and hierarchy cluster of the standardized variables. In PCA, the first component showed that four factors contributed to the maximum contribution rate: rock type, weathering degree, soil type, and soil depth. But in hierarchy cluster, the cluster just contained three variables: weathering degree, soil type, and soil depth. In hierarchy cluster analysis, altitude, latitude, and longitude belong to the same cluster, while in PCA they were classified into two principal components. The reason for the difference could be explained by the data type of the two methods. Principal component analysis used the absolute value of the correlation coefficient, whereas the hierarchy cluster used the vector. Both of the two results provided that the environmental factors influencing the roadside stability were classified into three sorts, in which hierarchy cluster provided qualitative factors while PCA provided qualitative and quantitative factors of the rock slope stability.

Acknowledgments This study is supported by the National Natural Science Foundation of China (NSFC) (50974092), and the Bureau of Science and Technology of Yibin (200903030). Thanks are also to YANG Tao, Professor GAN Youmin, and YANG Guanghui, for the field data collection and plant recognition and identification.

References

- Annoni P (2007) Different ranking methods: potentialities and pitfalls for the case of European opinion poll. *Environ Ecol Stat* 14:453–471
- Calder AJ, Burton AM, Andrew PM et al (2001) A principal component analysis of facial expressions. *Vis Res* 41:1179–1208
- Cliff N (1988) The eigenvalues-greater-than-one rule and the reliability of components. *Psychol Bull* 103:276–279
- Doty RL, Smith R, Mckeown DA et al (1994) Tests of human olfactory function: principal components analysis suggests that most measure a common source of variance. *Percept Psychophys* 56(6):701–707
- Fievez V, Vlaeminck B, Dhanoa MS et al (2003) Use of principal component analysis to investigate the origin of heptadecenoic and conjugated linoleic acids in milk. *J Dairy Sci* 86:4047–4053
- Guo H, Wang T, Louie PKK (2004) Source apportionment of ambient non-methane hydrocarbons in Hong Kong: application of a principal component analysis/absolute principal component scores (PCA/APCS) receptor model. *Environ Pollut* 129:489–498
- Guttman L (1954) Some necessary conditions for common factor analysis. *Psychometrika* 19:149–161
- Inger A, Rydgren K, Økland RH (2008) Scale-dependence of vegetation-environment relationships in semi-natural grasslands. *J Veg Sci* 19(1):139–148
- Jackson DA (1993) Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* 74(8):2204–2214
- Lipkovich I, Smith EP, Ye K (2008) Detecting pattern in biological stressor response relationships using model based cluster analysis. *Environ Ecol Stat* 15:71–78
- Morillo E, Romero AS, Madrid L et al (2008) Characterization and sources of PAHs and potentially Toxic metals in urban environments of Seville (Southern Spain). *Water Air Soil Pollut* 187:41–51
- Norusis MJ (1990) *SPSS-X advanced statistics guide*. SPSS Incorporated
- O'Lenic EA, Liverzey RE (1988) Practical considerations in the use of rotated principal component analysis (RPCA) in diagnostic studies of upper-air height fields. *Mon Weather Rev* 116:1682–1689
- Ronen S, Aragón-Salamanca A, Lahav O (1998) Principal component analysis of synthetic galaxy spectra. *Mon Not R Astron Soc*, 1–11
- Seabloom EW, Borer ET, Boucher VL et al (2003) Competition, seed limitation, disturbance, and reestablishment of California native annual forbs. *Ecol Appl* 13(3):575–592
- Sousa SIV, Martins FG, Alivim-Ferraz MCM et al (2007) Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environ Model Software* 22:97–103
- ter Braak CJF (1983) Principal components biplots and alpha and beta diversity. *Ecology* 64(3):454–462
- ter Braak CJF (1989) The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetation* 69:69–77
- Thorpe RS (1988) Multiple group principal component analysis and population differentiation. *J Zool Lond* 216:37–40
- Tinker DB, Rosor CAC, Beauvais GP et al (1998) Watershed analysis of forest fragmentation by clear cuts and roads in a Wyoming Forest. *Landscape Ecol* 13:149–165
- Tipping ME, Bishop CM (1999) Probabilistic principal component analysis. *J R Stat Soc B* 61(3):611–622
- White DA, Hood CS (2004) Vegetation patterns and environmental gradients in tropical dry forests of the northern Yucatan Peninsula. *J Veg Sci* 15(2):151–161
- Wiser SK, Peet RK, White PS (1996) High-elevation rock outcrop vegetation of the southern Appalachian Mountains. *J Veg Sci* 7:703–722