

## KARST ROCKY DESERTIFICATION INFORMATION EXTRACTION BASED ON THE DECISION TREE

Chengjie Su<sup>1,2</sup>, Tao Yue<sup>1\*</sup>, Lin Jiang<sup>2</sup>, Xiangmin Li<sup>2</sup>, Wenguan Wang<sup>2</sup>

1.Guangxi Key Laboratory of Spatial Information and Geomatics, Guilin University of Technology, Guilin, Guangxi, 541004, China

2.Department of Civil Engineering, Guangxi Polytechnic of Construction, Nanning, Guangxi, 530007, China

**Keyword:** karst rocky, decision tree, vegetation cover, slope

**Abstract:** Rocky desertification is a common geo-ecological disasters in China are mainly distributed in southwest karst region, and a wide range of further deterioration. Based on the theory of decision tree Guangxi rocky information extraction, selection of experimental data of Guangxi Zhuang Autonomous Region in 2005 TM image. First of remote sensing images after geometric correction image registration and other pretreatment. Secondly based on binary model of pixel, the Guangxi Zhuang Autonomous Region *NDVI* values and vegetation cover and slope analysis combining the results of Guangxi Zhuang Autonomous Region, the use of decision tree classification of remote sensing images, and finally get different levels of Guangxi Zhuang Autonomous Region rocky area and spatial distribution. The experimental results showed that: 2005 Guangxi rocky area of about 22,000 km<sup>2</sup>, accounting for 9% of the total land area in Guangxi, accounting for 24.30% of the karst area the overall classification accuracy of 89.03%, Kappa coefficient was 0.8417. From the classification results and the accuracy evaluation shows that the use of the information extracted rocky achieve better results.

---

\*Corresponding author. Tel:+86-15878346093

E-mail :yuetao@glut.edu.cn

Sponsored by The project was funded by Guangxi Key Laboratory of Spatial Information and Geomatics (17-259-16-13) and 2019 Guangxi Young and Middle-aged Teachers' Research Fundamental Ability Enhancement Project (2019KY1363)

## 1. INTRODUCTION

Karst rocky is recognized in recent years to a new geo-ecological disasters, the impact on the economy, environment and society is growing, widespread national attention (Li-Yangbing 2004). Advantages of using remote sensing technology that can accurately, quickly and economically to get the status of karst rocky desertification, distribution, etc. for monitoring desertification, evolution and subsequent analysis of rocky provide effective governance scientific basis (Kuang-Shunda 2009). The commonly used information extraction methods are mainly rocky obtain information based on vegetation index. Tong-Liqiang (2003) using TM5/TM4 computing the ratio of information to enhance the rocky, rocky index generated image. Li-Zhaoyang et al (2007) using the ratio of enhanced vegetation index (EVI) to extract rocky information, EVI is designed to avoid the saturation ratio based vegetation index. While coupling the atmosphere and soil resistant vegetation index adjusted vegetation index, reducing the impact of atmospheric and soil background EVI synthesis. Kuang-Shunda. (2009) the normalized difference vegetation index transformation and enhanced vegetation index improved enhanced vegetation index method, this model makes full use of vegetation, soil, and water bodies, such as bare rock band in blue, red and near-infrared bands band spectral response characteristics, increasing the contrast between soil, vegetation, bare rock and water between vegetation conducive to extract information. Xia-Xueqi et al (2006) Puding County, Guizhou Province IKONOS remote sensing data, to geometric index to characterize rocky strength. On the basis of multi-dimensional vector space spectrum analysis, put forward the concept of "vegetation line" and the definition of "rocky geometry index" (GRI) is a multi-dimensional vector space varying degrees of rocky pixel to "vegetation line" normalized distance. Chen-qiuwei et al (2003) also proposed a method based on the spectral characteristics of karst rocky desertification of remote sensing information extraction. In addition, S.-J.WANG (2004) and Li -Yangbing (2009) are based on information extracted

rocky land use.

Common information extraction method is mainly rocky vegetation index or improving the use of vegetation index. Relatively simple method in the classification, the classification results may not be ideal. As a key of decision tree classification method based on knowledge, the algorithm complexity low and high efficiency, widely used in the field of remote sensing, which is characteristic variables to extract information and select the node threshold (Mahesh Pal 2003). In this paper, the above methods, the use of decision tree classification method for remote sensing images in Guangxi 2005 TM data rocky classified according to *NDVI*, vegetation cover and slope analysis.

## 2. THEORETICAL MODEL OF DECISION TREE

Decision tree is an inductive learning training samples to generate decision rules, and decision rules for the use of new data through mathematical method of classification. A decision tree is a tree structure consists of a root node, consisting of a series of internal nodes and leaf nodes of three parts, a parent of two or more child nodes and nodes constitute each node is connected via a branch between nodes. Decision tree is an intuitive knowledge representation method, which uses information theory mutual information (information gain) to find property field with the maximum amount of information in the database, create a node tree, and then created based on different values of the field branch of the tree; concentrated repetition to build the tree at each branch and sub-branch of the lower node processes (HansenM, 1996, DudaRO, 2003). In addition to the tree in the form of a tree (Figure 1), but also can be expressed as a set of IF-THEN rules generated form. Each tree in the path from the root to the leaf corresponds to a rule, the rule is to determine the condition of all the property values of the nodes on this path, the conclusion is the rule class properties on this path leaf nodes.

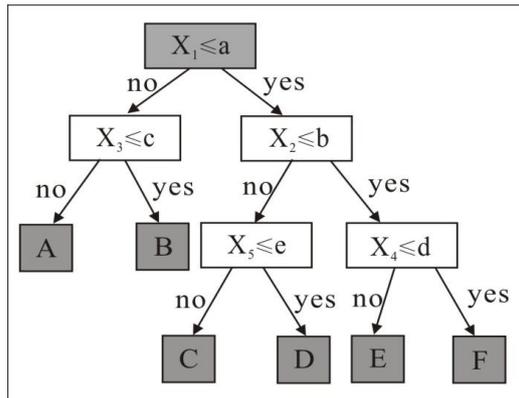


Figure 1. The tree structure of decision tree

$X_i$  is a characteristic;  $a, b, c, d$  and  $e$  is the decision threshold value;  $A, B, C, D, E$  and  $F$  of the classification results.

Decision tree classification methods including decision tree learning and decision tree classifier two processes. Decision tree learning is through inductive learning training samples to generate decision tree classification rules expressed in the form of machine learning process. Tree usually top-down recursively generated, using some method to select the best attribute as a node of the tree, were compared to determine property values down from a branch of the node in the node, in each repeat subset establish lower branch nodes and branches. Under certain conditions, to stop the growth of the tree, the leaf node classification decision tree. Decision tree classification is by learning the rules of training samples to generate decision trees, then an unknown sample can be classified according to the threshold set for each rule.

Decision tree for remote sensing image classification can be divided into three stages:

- (1) Data preparation phase, the remote sensing image data preprocessing, feature determines the type of statistical training in the region around the object type information (spectral and non-spectral information).
- (2) The construction of rule phase, the larger the amount of information the property containing the rules as to form an internal node tree until the same category included within each leaf node so far.
- (3) Decision tree trimming stage, check the classification results can meet the needs of, if not, you need to adjust the tree (pruning and add nodes) until

the establishment of a proper tree so far.

### 3. ROCKY EXTRACTION PROCESS

#### 3.1 Study area

Guangxi is a typical area of karst landforms, the region's distribution area of 89,500 carbonatite  $\text{km}^2$ , accounting for 37.8% of the total area; bare rocky mountain area of 78,800  $\text{km}^2$ , accounting for 33.3% of the total area; Guangxi 95 county (city), the karst area of more than 30% the proportion of 36, accounting for 37.9% (Li-Shuiming 2006). These areas are well developed karst landforms, vegetation is sparse, serious soil erosion, bare rock, causing very serious rocky land. Rocky desertification reduced survival and development in karst regions of the people, the destruction of the ecological environment, the loss of land resources, intensifying conflicts between man and nature, directly affects the economic and social development in karst areas.

#### 3.2 Data Introduction

In this paper, the use of remote sensing data classification images of Landsat-5 TM acquired imaging sensors and data processing through ASTER GDEM come digital elevation (DEM) product (data from the International Computer Network Information Center, Chinese Academy of Sciences data mirroring website (<http://www.gscloud.cn>)). The basic parameters of the experimental data shown in Table 1.

Table 1 The basic parameters of the experimental data

	Spatial resolution	Projection	Datum plane
TM images	30m	UTM	WGS-84
DEM	30m	UTM	WGS-84

### 3.3 Experiment

**3.3.1 Data preprocessing:** Using Envi4.8 the image preprocessing, including geometric correction and image registration and image mosaic, and finally get the whole TM images Guangxi Administrative vector file according to the completed mosaic image cropping. Figure 2 is a band with 5, 4, 3 false color composite image of the entire Guangxi TM.

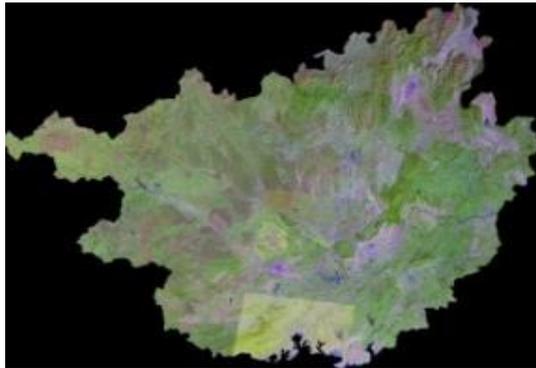


Figure 2. Guangxi TM false color composite image

#### 3.3.2 Rocky information classification decision tree:

##### 1) Vegetation coverage calculation

Assuming as an information element can be divided into two parts, soil and vegetation, the information  $S$  observed by remote sensors, the composition of the soil can be expressed by  $S_s$  and the contribution of the information contributed by the vegetation component  $S_v$ , two parts.

$$S = S_s - S_v \quad (1)$$

For a mix of soil and vegetation in a two-part pixel, pixel area at a ratio of vegetation cover for the pixel of vegetation cover  $f_c$ , while the ratio of the area of soil cover  $(1-f_c)$ . Let the whole of vegetation cover resulting pure pixel sensing information  $S_{veg}$ . Dollar mixed as the vegetation component of the information contributed by  $S_v$  can be expressed as:

$$S_v = S_{veg} * f_c \quad (2)$$

Similarly, the information contributed by the soil composition  $S_s$  can be expressed as:

$$S_s = S_{soil} * (1 - f_c) \quad (3)$$

The obtained of formula (2) and (3) into equation (1):

$$S = f_c * S_{veg} + (1 - f_c) * S_{soil} \quad (4)$$

Formula (4) is transformed to obtain a formula for the calculation of vegetation cover:

$$f_c = (S - S_{soil}) / (S_{veg} - S_{soil}) \quad (5)$$

The normalized difference vegetation index (NDVI) can be approximated into equation (4) as follows:

$$NDVI = f_c * NDVI_{veg} + (1 - f_c) * NDVI_{soil} \quad (6)$$

By the formula (6) can be obtained vegetation cover are:

$$f_c = \frac{NDVI - NDVI_{soil}}{NDVI_{veg} - NDVI_{soil}} \quad (7)$$

$NDVI_{soil}$  should not change with time, for most types of bare ground, the theoretical value should be close to zero  $NDVI_{soil}$ . However, changing in the surface is affected by humidity atmospheric conditions,  $NDVI_{soil}$  change over time. In addition, due to the different surface moisture, roughness, soil type and soil conditions such as color,  $NDVI_{soil}$  will vary with the space.  $NDVI_{veg}$  represents the maximum pixel full of vegetation cover. Due to the different types of vegetation and seasonal variations, determine  $NDVI_{veg}$  values also exist with  $NDVI_{soil}$  similar situation,  $NDVI_{veg}$  value will change over time and space. Usually need to experience to judge the general  $NDVI_{veg} = NDVI_{max}$ ,  $NDVI_{soil} = NDVI_{min}$ .

Using Envi4.8 calculated NDVI values TM image, we can see  $NDVI_{max} = 0.555997$ ,  $NDVI_{min} = -0.009116$  based on the results. And then put these two values into the formula (7), to obtain the distribution of vegetation covering the Guangxi region (Figure 3).

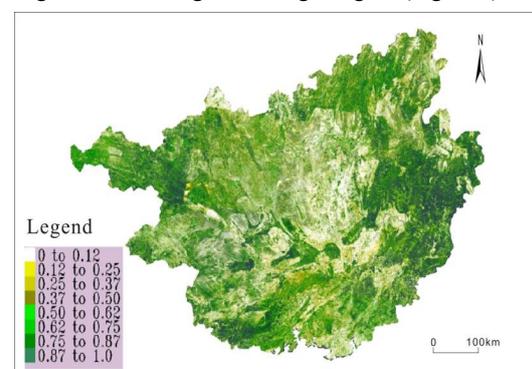


Figure 3. Guangxi vegetation cover map

##### 2) Slope analysis

Download complete coverage of the entire DEM data

Guangxi, and inlaid with ArcGIS get the whole DEM data Guangxi, and then use ArcGIS carried slope analysis and the results shown in Figure 4.

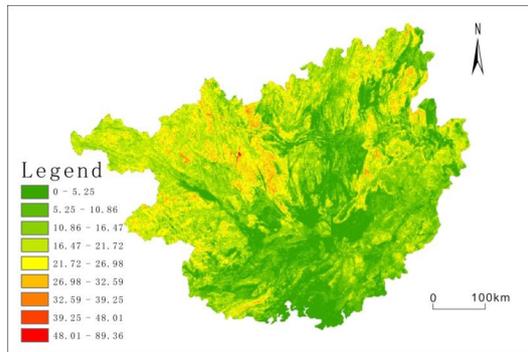


Figure 4. Guangxi slope analysis diagram

### 3) Rocky Information Classification

In this study, first using GPS, undertaking field surveys, to obtain correspondence between vegetation coverage and *NDVI* vegetation index in the karst region, combined with the relational model vegetation cover and vegetation index analysis results, calculated for each coverage classification threshold. According to results of the analysis combined with field gradient field survey karst region, the calculated levels of different rocky slope threshold (Table 2).

Table 2 in the study area rocky graded index

Rocky grade	<i>NDVI</i>	The vegetation coverage (%)	Slope (°)
R6	>0.32	>65	<19
R5	0.15~0.32	50~65	>19
R4	-0.05~0.15	30~50	>24
R3	-0.27~-0.05	20~30	>28
R2	-0.48~-0.27	9~20	>35
R1	<-0.48	<9	>45

Note: R1-R6 represent extremely rocky desertification, strength rocky desertification, moderate rocky desertification, slight rocky desertification, potential rocky desertification and non-rocky desertification.

In Envi4.8 in the above-mentioned indicators as a condition of rocky decision tree classification, decision tree classification rule base (Figure 5), run the initial classification decision tree to get results.

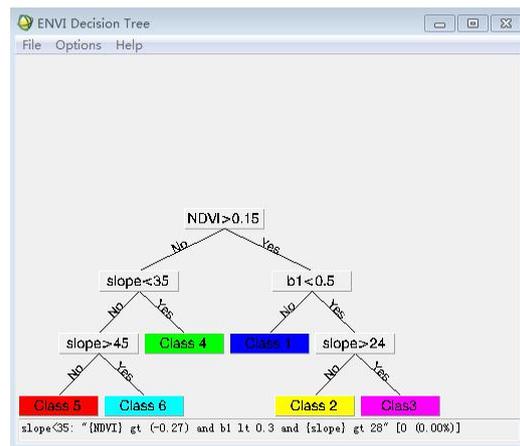


Figure 5. Decision tree classification rule base

Which, *b1* and slope, respectively, the vegetation coverage and slope, 1-6 rocky represent different levels, in turn means no rocky, potentially rocky, rocky mild, moderate desertification, strength rocky and extremely rocky. As it relates to the whole area in Guangxi, the situation in the classification process misclassification, so the need for initial classification of the image can be corrected. Using 5, 4, 3 false color composite image at the right or wrong of correction, and then combined with Google Earth corrected. The main areas misclassified as bare soil mountain region, image features in such areas and image features in rocky desertification areas shown in Figure 6. Results of the initial classification by Google Earth would be wrong type of correction, the final classification results combined into a map (Figure 7).

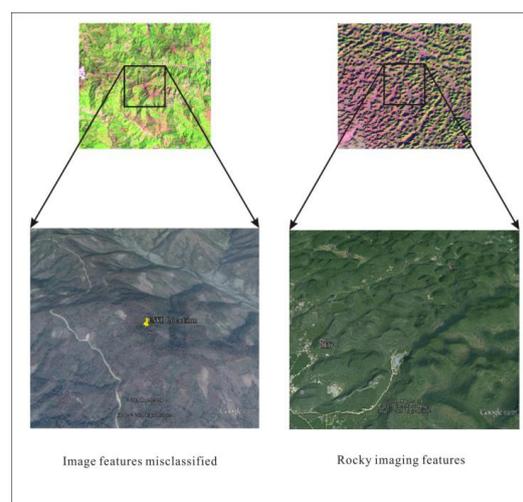


Figure 6. Google Earth corresponding image features

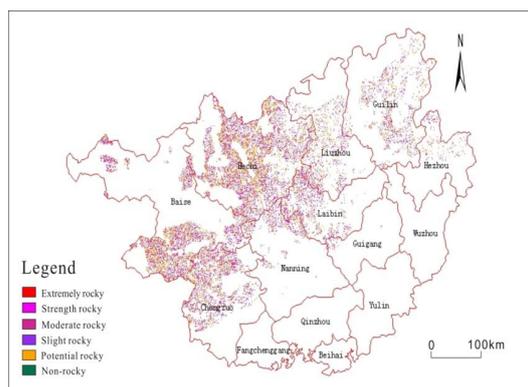


Figure 7. Guangxi rocky hierarchical diagram

**3.3.3 The results:** The statistical results of each grading rocky classification (Table 3) showing that, the whole Guangxi rocky area of about 22,000 km<sup>2</sup>, accounting for about 9% of the land area of Guangxi.

Table 3 rocky situation in Guangxi in 2005

Rocky grade	Area (km <sup>2</sup> )	Proportion of land area in Guangxi (%)
R6	215174.6397	90.82
R5	8287.1253	3.50
R4	4370.2605	1.84
R3	4628.3553	1.95
R2	3071.7990	1.30
R1	1390.0518	0.59
The total	236922.2316	100.00

Note: R1-R6 represent extremely rocky desertification, strength rocky desertification, moderate rocky desertification, slight rocky desertification, potential rocky desertification and non-rocky desertification.

Karst area of about 89,500 km<sup>2</sup> (Yu-Su 2003), accounted for the entire five graded rocky karst area is approximately 24.30% of its statistical results are shown in Table 4.

Table 4 Share of karst rocky desertification area case

Rocky grade	Area (km <sup>2</sup> )	Proportion of karst area in Guangxi (%)
R5	8287.1253	9.26
R4	4370.2605	4.88
R3	4628.3553	5.17
R2	3071.7990	3.43
R1	1390.0518	1.55
The total	21747.5919	24.30

Note: R1-R5 represent extremely rocky desertification, strength rocky desertification, moderate rocky desertification, slight rocky desertification, potential rocky desertification.

According Envi4.8 class confusion matrix reports

(Table 5) can be obtained, the overall classification accuracy of 89.03%, Kappa coefficient was 0.8417. From the overall classification accuracy and Kappa coefficient shows based on *NDVI*, The vegetation coverage and slope analysis, decision tree classifier to achieve better results.

Table 5 Guangxi rocky classification accuracy

evaluation results in 2005					
	R1	R2	R3	R4	R5
R1	76	0	0	1	3
R2	3	54	1	1	0
R3	2	7	50	1	1
R4	3	3	5	51	0
R5	6	3	0	1	98
R6	7	3	8	2	8
Total	97	70	64	57	110
P (%)	78.35	77.14	78.13	89.47	89.09

Note: R1-R6 represent extremely rocky desertification, strength rocky desertification, moderate rocky desertification, slight rocky desertification, potential rocky desertification and non-rocky desertification. U and P represent respectively user accuracy and produced accuracy.

#### 4. CONCLUSION

(1) In this paper as a decision tree to determine the conditions in accordance with *NDVI*, The vegetation coverage and slope classification results combined with Google Earth will be corrected in 2005 to get the rocky area in Guangxi region about 22,000 km<sup>2</sup>. The overall classification accuracy of 89.03%, Kappa coefficient was 0.8417, indicating that the method used to extract information rocky achieve better results.

(2) The extent and evolution of rocky the vegetation coverage was negatively correlated positively correlated with the slope is presented. Guangxi rocky area of about 22,000 km<sup>2</sup> in 2005, accounting for 9% and 24.30% of the total land area of Guangxi and Guangxi, the total area of karst. Guangxi rocky serious extent, mainly due to the karst region of Guangxi and more affected by natural conditions and human factors, and promote rocky process.

## REFERENCES

- Li-yangbing, Wang-shijie and Rong-li. Discussion on the Concepts of Karst Rock Desert and Rock Desertification. *Chinese Desert*, 2004, 24(6), 689-695.
- Kuang-Shunda, Dai-Chuang, Wang-Shangyan et al. Discussion on Remote Sensing Information Enhancement Technology for Karst Rocky Desertification. *Guizhou Geology*, 2009(1), 44-48.
- Tong-Liqiang. Research on Automatic Extraction Technology of Rocky Desertification Information in Karst Mountain Area of Southwest China. *Remote Sensing of Land and Resources*, 2003(4), 35-39.
- Li-Zhaoyang, Kuang-Shunda, Li-Zhizhong et al. Remote Sensing Information Enhancement and Monitoring Technology for Karst Rocky Desertification. *Journal of Guilin University of Technology*, 2007, 27(1), 30-35.
- Xia-Xueqi, Tian-Qingjiu and Du-Fenglan. Research on Remote Sensing Information Extraction Method of Rocky Desertification Degree. *Journal of Remote Sensing*, 2006, 10(4), 469-475.
- Chen-Qiwei, Lan-Anjun, Xiong-Kangning et al. Extracting Information of Karst Rocky Desertification Based on Remote Sensing Spectral Features. *Journal of Guizhou Normal University*, 2003, 21(4), 82-87.
- Li-Shuiming, Shu-Ning, Wang-Guochong et al. Analysis of the Causes and Development Trends of Rocky Desertification in Guangxi. *Journal of Guangxi Academy of Sciences*, 2006, 22(3), 193-196.
- Yu-Su. Current Status and Characteristics of Rocky Desertification Distribution in China. *Southwest Forestry Survey Planning*, 2003, 22(2), 53-55.
- HansenM, DubayahR, DeFrieSR. ClassificationTreeS, AnAlternative to Traditional Land Cover Classifiers. *InternationalJournalofRemoteSensing*, 1996, 17, 1075-1081.
- Quinlan J R. C4.5: Programs for Machine Learning. SanMateo. CA: Morgan Kaufman Publishers, 1993.
- Quinlan J R. Simplifying Decision Tree. *International Journal of Human-Computer Studies*, 1999, 51, 497-510.
- S.-J. WANG, Q.-M. LIU AND D.-F. ZHANG. Karst Rocky Desertification in Southwestern China: Geomorphology, Landuse, Impact and Rehabilitation, *Land Degradation & Development*, 2003, 15, 115-121.
- Li-Yangbing, Shao-Jingan, Yang- Hua et al. The Relations Between Land Use and Karst Rocky Desertification In A Typical Karst Area, China. *Environ Geol*, 2009, 57, 621-627.
- Mahesh Pal, Paul M. Mather. An Assessment of the Effectiveness of Decision Tree Methods for Land Cover Classification. *Remote Sensing of Environment*, 2003, 86, 554-565.