Environ Geol (2006) 49: 1173–1179
DOI 10.1007/s00254-005-0162-4

ORIGINAL ARTICLE

Huang Qiu-hao
Cai Yun-long

Assessment of karst rocky desertification using the radial basis function network model and GIS technique: a case study of Guizhou Province, China

Received: 14 July 2005
Accepted: 21 November 2005
Published online: 4 January 2006

H. Qiu-hao · C. Yun-long (✉)
Department of Resources, Environment and Geography, College of Environmental Sciences, Peking University,
100871 Beijing, Peoples Republic of China
E-mail: qhhuang@pku.edu.cn
E-mail: caiyl@urban.pku.edu.cn
Tel.: +86-10-62758845
Fax: +86-10-62751187

Abstract Karst rocky desertification in karst areas of Southwest China is an obstacle for the sustainable development of the locality. While many researchers have studied this issue, this study quantified the various indicators affecting desertification, using GIS technology to extract spatial data and to construct a RBFN (radial basis function network) model for Guizhou province. By calibrating sample data of the different levels of karst rocky desertification, the model parameters were established, and then the assessment of karst rocky desertification. Results show that in counties of Southwestern Guizhou Province karst rocky desertification is severe, counties in northern Guizhou have moderate desertification, and southeastern counties are affected lightly. Comparison of the results with other research shows conformity with actual conditions, proving the reasonability and applicability of the RBFN model.

Keywords Karst rocky desertification · Radial basis function network (RBFN) · Assessment · Guizhou province · China

Introduction

Karst rocky desertification is a progressive process of land degradation where soil is seriously or thoroughly eroded. Bedrock exposure is widespread, productive capability of land declines and finally, a landscape similar to desert appears under human impact on a vulnerable eco-geo-environment (Yuan 1993; Yang 1995; Xiong 2002). This major type of desertification results in extensive exposure of bedrock. In the southwest mountain area of Guizhou, the karst area covers about 620,000 km^2, which is the largest known karst area in the world (Cai 1994).

Much research has been done on the process of karst rocky desertification including distribution (Su 2002; Xiong 2002), the ecological–environmental effects (Wang 2003), cause analysis (Wang and Li 2003; Zhang 2001; Zhang et al. 1999) and integrated rehabilitation (Cai 1994; Cai and Meng 1999; Huang 2000). Yet how to assess karst rocky desertification is still difficult because there are many factors to weigh such as percentage of vegetation, percentage of area with slope >25° and annual precipitation.

This study uses the GIS technique to extract spatial data, train data to identify the coefficients, and assess the 82 counties' karst rocky desertification levels in Guizhou Province using the resulting RBFN model. The result is in conformity with actual conditions, which indicates that GIS coupled with RBFN model is a flexible and powerful approach for land assessment. The aim of this study is to devise an efficient and simple method for quickly classifying karst rocky desertification by means of GIS and RBFN model.

Study area

The study area, Guzihou province, at approximatly 24°–29°N and 103°–109°E, is in the southwest of China (Fig. 1) and covers about 176,000 km^2, of which karst

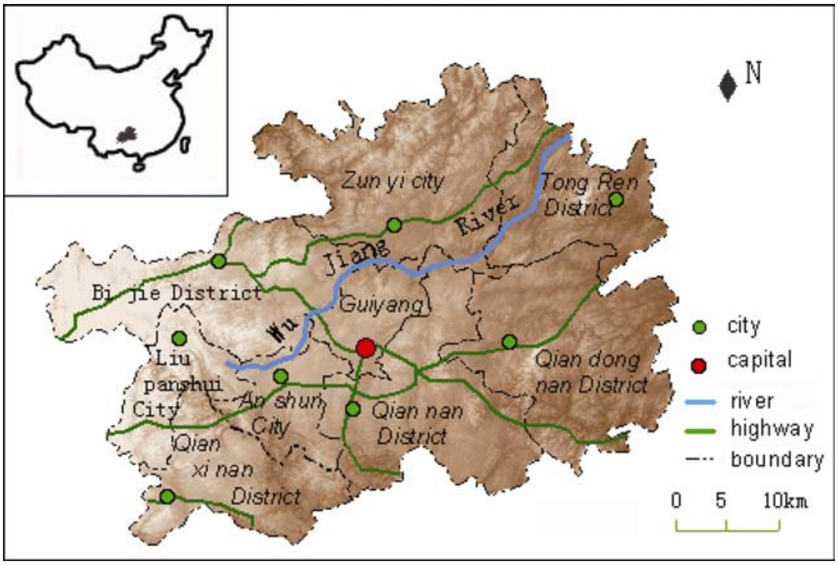

Fig. 1 Location of study area

areas account for 61.9%. The region has a warm–moist climate with an annual temperature ranging from 12 to 16°C. Terrain in this hilly province has an average elevation of 1,100 m above sea level. The amount of annual precipitation totals about 1,000 mm, most of which falls during the summer months between June and September. Guizhou Province has moderate populated density averaging 170 individuals per km^2. The main industries are agriculture and energy production (Cai 1990).

This area was selected because it is the province that the most serious karst rocky desertification has taken place. Assessment of karst rocky desertification may help the local government to develop relevant regional development policies.

Methodology

Sturcture of radial basis function network

A RBFN, which is multi-layer and feed-forward, is often used for classification. The term 'feedforward' means that the neurons are organized in the form of layers in a layered neural network (Park 1991; Haykin 1991). The basic architecture of a three-layered neural network is shown in Fig. 3. A RBFN has three layers including input layer, hidden layer and output layer. The input layer is composed of input data. The hidden layer transforms the data from the input space to the hidden space using a non-linear function. The output layer, which is linear, yields the response of the network. The argument of the activation function of each hidden unit in an RBFN computes the Euclidean distance between the input vector and the center of that unit. The network has better accuracy only when the early stop technique (Park 1991) is appropriately used.

In the structure of RBFN, the input data X is an I-dimensional vector, which is transmitted to each hidden unit. The activation function of hidden units is symmetric in the input space. The output of each hidden unit depends only on the radial distance between the input vector X and the center for the hidden unit. The output of each hidden unit, h_j, $j = 1, 2, 3, \ldots, J$; is given by

$$h_j(x) = \phi(\|x - c_j\|) \quad (1)$$

where $\|\ \|$ is the Euclidean Norm, c_j is the center of the neuron in the hidden layer, and $\phi()$ is the activation function, which is a non-linear function and has many types, for example, Gaussian, multiquadric, thinspline and exponential functions. If the form of the basis function is selected in advance, then the trained RBFN

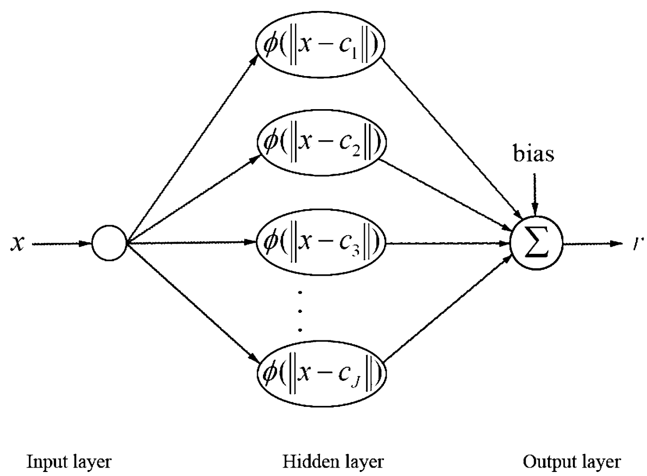

Fig. 2 The structure of RBFN

will be closely related to the clustering quality of the training data towards the centers. The popular Gaussian function was used in this study as the transform function in the hidden layer. The Gaussian activation function can be written as

$$\phi_j(x) = \exp\left[-\frac{\|x - c_j\|^2}{2\rho^2}\right] \quad (2)$$

where x is the training data, and ρ is the width of the Gaussian function. A center and a width are associated with each hidden unit in the network. The weights connecting the hidden and output units are estimated using least mean square method. Finally, the response of each hidden unit is scaled by its connecting weights to the output units and then summed to produce the overall network output. Therefore, the kth output of the network $\hat{y}_k$ is

$$\hat{y}_k = w_0 + \sum_{j=1}^{M} w_{jk}\phi_j(x) \quad (3)$$

where $\phi_j(x)$ is the response of the jth hidden unit, w_{jk} is the connecting weight between the jth hidden unit and the kth output unit, and w_0 is the bias term.

The training strategy

To construct a network system, the **I** vectors may be used to define **I** radial basis functions. This yields a number of hidden neurons equals to that of the data points. When the number of data points is high, the computational time is expensive. Another drawback is that the network may be over-trained. In the neural network, it is often assumed that the number of basis functions is significantly less than that of data points. Therefore, the determination of the number of neurons is an important problem for construction of a RBFN, as is the placement of neurons and the calculation of the adjusted parameters. There are different learning strategies in the design of a RBFN depending on how the centers of the radial basis functions of the network are specified. Two learning strategies are often used, namely, hybrid-learning algorithm (Moody and Darken 1989; Musavi et al. 1992) and fully supervised learning algorithm (Chen et al. 1990, 1991).

In this study, the fully supervised learning algorithm is used for parametric estimation of the network. First, the network begins with no hidden units. As input and output data are received during training, they are respectively used for generating new hidden units. The location of the first center may be chosen from the training data set, and the standard deviation ρ (i.e., width) of the jth neuron is

$$\rho = \sqrt{\frac{d_{\max}^2}{j+1}} \quad (4)$$

where $d_{\max}$ is the maximum distance between the training data. When the above steps are finished, a training data point is chosen as the new hidden unit. Then the single hidden layer RBFN linearly combines the output of the hidden units. After output of the network has been obtained, the relative root mean square error (RRMSE) is calculated. The RRMSE is defined as

$$\mathrm{RRMSE} = \sqrt{\frac{1}{P}\sum_{k=1}^{P}\left[\frac{\hat{y}_k - y_k}{y_k}\right]^2} \quad (5)$$

where k is the dummy time variable, P is the number of data elements in the period for which computations are made, and y_k and $\hat{y}_k$ are the observation and the forecast at time k, respectively. In a like manner, other data points are chosen sequentially. The training data point having the minimum value of RRMSE is added to the hidden layer until all training data points are tested.

However, the model with the best-performing parameter values may end up overfitting the validation data. The network becomes well trained, but its generalization ability is not adequate. That is the network does not learn enough about the past to generalize in the future. For this reason, the early stopping technique is used. A standard tool in statistics known as crossvalidation is used herein as a stopping criterion (Haykin 1991), in which training, validation and testing data sets are needed. The training data are used to find an optimal set of connecting weights, and the testing data are used to choose the best network configuration. Once an optimal network has been found, validation is required to test the true generalization ability of the model.

The early stopping technique is applied as follows: when the RRMSE's of the last two consecutive neurons are smaller than that of the former neuron, then the training of the network is stopped. Using the RRMSE's of two consecutive neurons can avoid fluctuation of the validation data. The decision as to whether an input–output (x_k, y_k) should give rise to a new hidden unit depends on the novelty of the data and is decided using the following two criteria

$$\mathrm{num}_{\mathrm{unit}} \le \mathrm{num}_{\max} \quad (6)$$

$$\mathrm{RRMSE} \le e_{\min} \quad (7)$$

where $\mathrm{num}_{\mathrm{unit}}$ is the number of units in hidden layer. Both $\mathrm{num}_{\mathrm{unit}}$ and $e_{\min}$ are thresholds to be appropriately selected. If the above two conditions are satisfied, then the data are deemed to have novelty and a new hidden unit is added. The first condition means that the number of units in the hidden layer must be limited and the

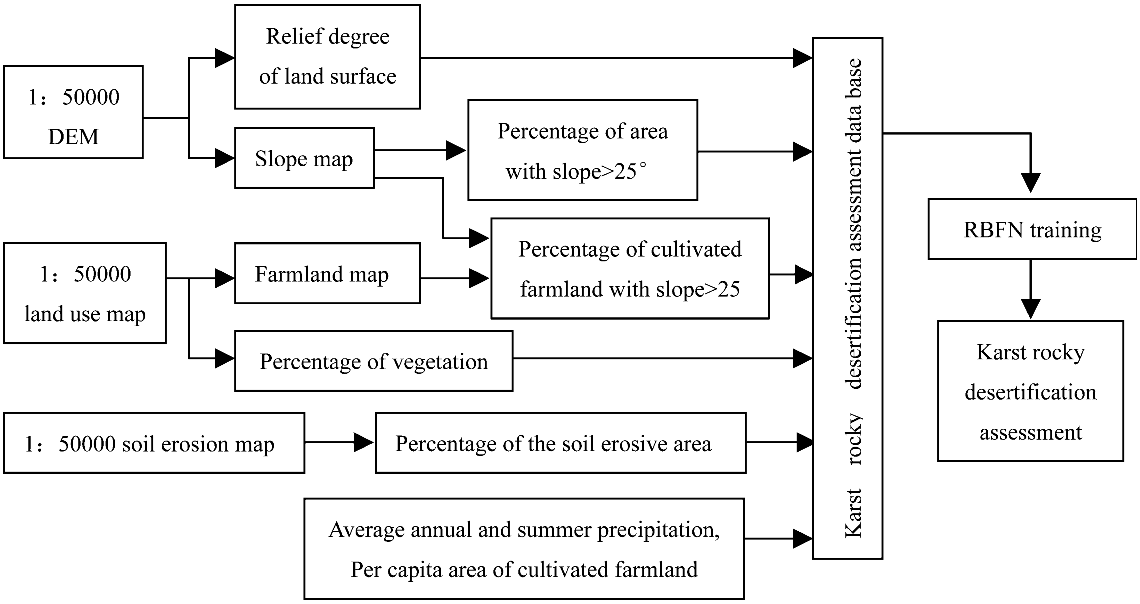

Fig. 3 Flow chart of karst desertification assessment using RBFN model

second condition means that the error between the network output and the target output must be significant. The value of $e_{\min}$ represents the desired approximation accuracy of the network output.

Main factors forming karst rocky desertification

According to relevant research (Li and Wang 2004; Xiong 2002), factors listed below were selected to assess karst rocky desertification.Percentage of vegetation (%): the most distinct feature of karst rocky desertification is deforestation.Index of average relief degree of land surface: topography has close correlation with karst rocky desertification. In the flat area the intensity of karst rocky desertification is weak, and the distribution of karst rocky desertification is sporadic. In rugged and fragmented areas, karst rocky desertification is more closely distributed. The formula for index of relief degree of land surface is

$$\mathrm{RDLS} = \frac{\max(h) - \min(h)}{\max(H) - \min(H)} \times \frac{1 - P(A)}{A} \tag{8}$$

h is the average altitude of each county, H is the average altitude of the Guizhou Province, A is the area of each country, and the $P(A)$ is the flat area of each country.Percentage of slope $>25°$(%): usually the steeper the slope, the more erosive the soil. The average slope of Guizhou Province is about 21°. When the slope $>25°$, soils are easily eroded.Percentage of erosion area (%): soil erosion is the direct cause of karst rocky desertification. In the fragile karst ecosystem, the soil is very thin and conditions for soil formation are poor. When soils are intensively eroded and deforested, karst rocky desertification is widespread.Average annual precipitation (mm): precipitation is one basic momentum for

Table 1 Classification standard of assessment of karst desertification in Guizhou

	The standard of the karst rocky desertification			
	Slightly	Moderately	Severely	Very severely
Percentage of vegetative unit (%)	> 50	35–50	20–35	< 20
Index of average relief degree of land surface	< 0.1	0.1–0.15	0.15–0.2	> 0.2
Percentage of area with slope > 25°	< 40	40–50	50–60	> 60
Percentage of the soil erosive area	< 20	20–30	30–40	> 40
Average annual precipitation (mm)	< 1,060	1,060–1,200	1,200–1,300	> 1,300
Average summer Precipitation (mm)	< 480	480–600	600–700	> 700
Percentage of cultivated farmlands with slope > 25°	< 15	15–20	20–25	> 25
Per capita area of cultivated farmland (hm^2/ per capital)	> 0.9	0.7–0.9	0.6–0.7	< 0.6

Note: the classification standard is abstracted from Xiong (2002) and Li and Wang (2003)

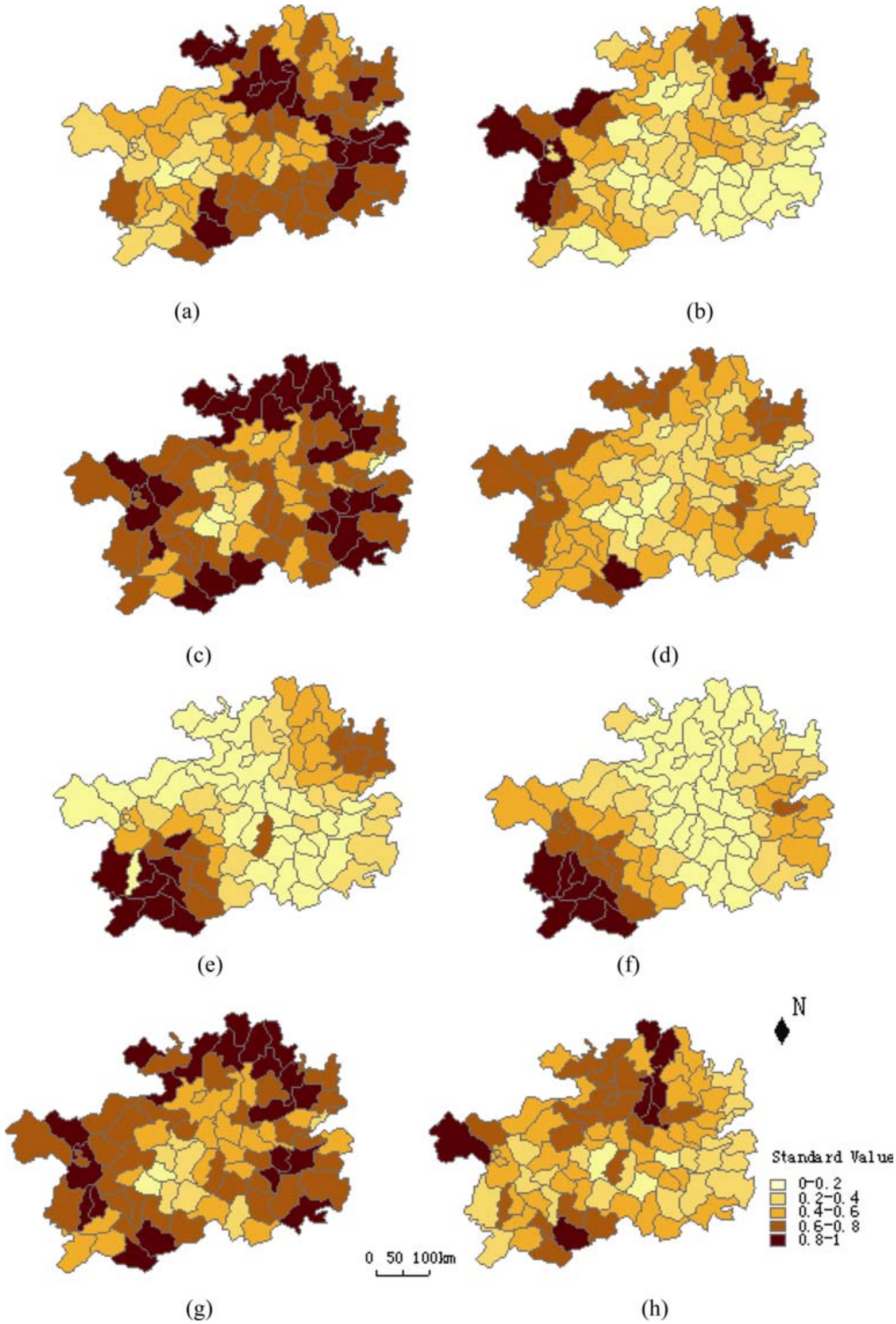

Fig. 4 Maps of standard value of the affecting factors of karst rocky desertification in Guizhou (**a** Percentage of vegetation, **b** index of average relief degree of land surface, **c** percentage of the soil erosion area, **d** percentage of area with slope >25°, **e** average annual precipitation, **f** average summer precipitation, **g** percentage of cultivated farmlands with slope >25°, **h** per capita area of cultivated farmland)

erosion.Average summer precipitation (mm): in Guizhou province, the summer precipitation usually accounts for 50% of the whole year. The great volume of summer precipitation affects the soil erosion greatly.Percentage of cultivated farmlands with slope >25°(%): quite a few parts of karst rocky desertification comes from cultivated land with steeper slope that has been over-exploited for agriculture. Because the cultivated farmland resources are limited in Guizhou province, nearly all the county has the cultivated farmlands with some slope. The higher the slope, the more vulnerable the land karst rocky desertification.Per capita area of cultivated

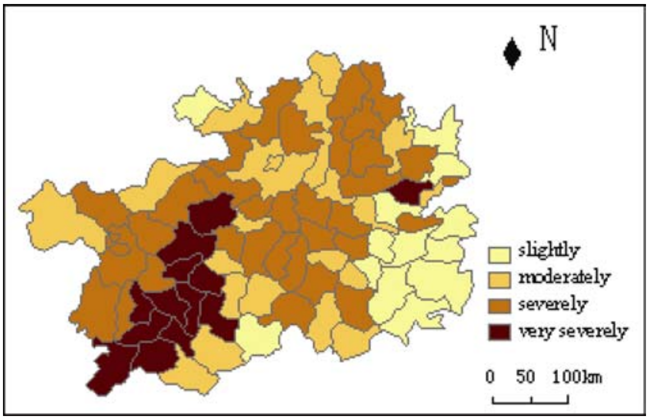

Fig. 5 Map of different levels of karst rocky desertification

farmland (hm^2 per capital): this indicator reflects population pressures to land resources in the karst area.Figure 3 summarizes the overall analysis procedure followed in this study. The GIS software used in is ARCGIS Desktop 8.3, and the RBFN model is run by Matlab 6.5.

Data source and preparation

Data source

The DEM (Tor 1999) data used is at a scale of 1:50,000 from Guizhou Province's DEM, and the land use data and the soil erosion data were mapped in 2000 at a scale of 1:50,000 (Arc/info format). All data are from the Chinese Academy of Science (CAS), Beijing, P. R. China. Other kinds of data, such as the annual precipitation, social–economical data are all from statistical annals of China (Guizhou Bureau of Statistics 2000).

Data preparation

The data preparation included two steps: first unify the projection, coordinate and resolution of spatial data, and then use GIS software to extract the spatial data (% of vegetation, % of area with slope >25°, soil erosion data etc). All the data are standardized for the same unit using the formula

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}. \quad (9)$$

Train sample data

Based on these results classification standards were applied to the sample data. As defined by Li and Wang (2003) and Xiong (2002), four land types were used: (I) slightly karst rocky desertification county, (II) moderately karst rocky desertification county, (III) severely karst rocky desertification county, (IV) very severely karst rocky desertification county. The classification standard is shown in Table 1.

Results

The GIS was used to extract the spatial data from the basic data. Figure 4 shows maps of the standard value of the relevant factors. Different gray degree means the different scores of standard value, the darker the county, the relative score is higher.

Then 20 cases of data in Table 1 (each level using 5 different cases) were randomly selected to train and set stops when the RRRSE $= 10^{-4}$. When the error rate is $< 10^{-6}$, the train stops, and the relevant coefficients of the RBFN model are established.

Next data for the 82 counties were input for assessment. Figure 5 documents the results.

From the map in the southwest part of Guizhou Province karst rocky desertification is severe, especially along the northeast–southwest line. In southeastern Guizhou, the desertification situation is better. In northwestern and northeast Guizhou, counties vary between moderate and severe desertification.

In fact, according to Li et al. (2003), the severely karst rocky desertification counties are mainly located in two areas. One is along the Wu Jiang Rivers, and the other is in the *Qian Xinan District*. Zhao also concluded that the middle part of the Province has only widely distributed desertification, while the southeastern part of the Province has moderate karst desertification, and the southeast part is prone to slight karst rocky desertification because of the clastic rock geology. The results of the model are comparatively accurate. The RBFN model is useful in assessing karst rocky desertification.

Conclusions

Assessing previous research on the formation and distribution of karst rocky desertification combined with biophysical factors of Guizhou Province, this study constructed a RBFN model. The results of the model show that in counties of southwestern Guizhou Province, the karst rocky desertification is severe, especially along the southwest–northeast line. Northern Guizhou is moderate while southeastern is slight. The assessment result provides a scientific foundation for land resource planning, controlling the karst rocky desertification and eco-environmental protection.

In this combination of the GIS—RBFN model, the GIS function was utilized to analyze and extract relevant spatial data, then the RBFN model was used to assess the extent of the karst rocky desertification. The result is quite satisfactory. The combination of GIS—RBFN may effectively solve problems of the management and visualisation of spatial data.

The research here demonstrates the feasibility of using GIS and RBFN to assess karst rocky desertification. Only a few indicators were used to establish the model. In fact, with more relevant data sources, a more mature model could be established.

Acknowledgements The research was supported by the Key Project of National Natural Science Foundation of China (40335046). The authors are grateful to Dr. Li Shuncheng and the anonymous reviewer for their comments and suggestions. However, any errors or shortcomings in the paper are our responsibility.

References

Cai Y (1990) The territorial structure and resources development in Guizhou province. Ocean Press, Beijing

Cai Y (1994) The study of alleviating poor and sustainable development in southwest karst areas. Peking University Press, Beijing

Cai Y, Meng J (1999) Ecological reconstruction of degraded land: a social approach.Sci Geogr Sin 19:198–203

Chen SS, Billings A, Grant PM (1990) Recursive hybrid algorithm for non-linear system identification using radial basis function networks.Int J Control 55:1051–1070

Chen SC, Cowan FN, Grant PM (1991) Orthogonal least squares learning algorithm for radial basis function networks IEEE Trans Neural Netw 2(2):302–309

Guizhou Bureau of Statistics (2000) Guizhou statistical yearbook. China Statistical Press, Beijing

Haykin S, (1991) Neural networks: a comprehensive foundation Prentice Hall, Englewood Cliffs, pp 213–214

Huang B (2000) Advance of the karst ecological test in laibing country. Carslogica Sin 19: 314

Li R, Wang S, Zhou D (2003) The correlation between rock desertification and lithology in karst area of Guizhou. Acta Geogr Sin 58:314–320

Li R, Wang S (2004) A study on rocky desertification evaluation index system.Trop Geogr 24:145–149

Moody J, Darken C (1989) Fast learning in networks of locally tuned processing units.Neural Comput 4:740–747

Musavi MT, Ahmed W, Chan KH, Faris KB, Hummels DM (1992) On the training of radial basis function classifiers.Neural Netw 5:595–603

Park J, Sandberg IW (1991) Universal approximation using radial basis function network.Neural Comput 3:246–257

Su W (2002) Controlling model for rocky desertification of karst mountainous region and its preventing strategy in southwest, China.J Soil Water Conserv 16:29–32

Tor B (1999) Geographic information systems: an introduction Wiley, New York, pp 124–125

Wang S (2003) The most serious eco-geologically environmental problem in southwestern China-karst rocky desertification. Bull Mineral Petrol Geochem 22:120–126

Wang S, Li Y (2003) Karst rocky desertification: formation background, evolution and comprehensive taming.Quaternary Sci 2:657–666

Xiong K (2002) The study of karst rocky desertification using the GIS & RS tech, a case study of Guizhou province. Geology Press, Beijing, pp 123–128

Yuan D, (1993) The karst study of China. Geology Press. Beijing

Yang H, (1995) Karst desertification and assessment of its disasters.Mar Geol Quaternary Geol 15:137–147

Zhang D, Wang S, Zhou D (2001) Intrinsic driving mechanism of land rocky desertification in karst regions of Guizhou province Bull Soil Water Conserv 21:1–5

Zhang H, Zhao X, Cai Y (1999) The driving mechanism of human forces to the land-use change in the karst mountain area Geogr Res 18:136–142