Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

ISSN: 0972-6268	Vol. 12	No. 4	pp. 685-690	2013

Original Research Paper

Water Quality Analysis in Jining City Using Clustering Methods

Yunxin Zhang and Changjun Zhu*
College of Water Conservancy and Hydropower, Hebei University of Engineering, Handan, 056038, China
*College of Urban Construction, Hebei University of Engineering, Handan, 056038, China

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 14-5-2013
Accepted: 8-6-2013

Key Words:
Clustering analysis
Water pollution indicators
SPSS software
Water quality assessment

ABSTRACT

According to the existing water quality monitoring data from the monitoring sites in Jining city, the clustering analysis method of mathematical statistical analysis is employed to analyse and evaluate the water quality of Jining city. First of all, by using the statistical software SPSS, we aggregate and classify the monitoring sites; at the same time, the water quality objectives are aggregated and classified and the main water pollution indicators are selected. Finally we made a preliminary assessment of water quality situation in Jining city.

INTRODUCTION

Surface water environment is an important part of the environment, whose evaluation can predict the water quality and provide the scientific basis for the prevention of water pollution. Surface water environment is a complex system with multiple factors. Some factors are known, some are unknown, which is a grey system. So the grey theory can be used to evaluate the water environment. In recent years, in addition to the comprehensive index and fuzzy comprehensive evaluation method, grey theory has been successfully used.

Chen Dongjing et al. (2002), Yang Wei et al. (2007) and Zhang caixiang et al. (2005) have used the factor analysis method to evaluate water quality to get the water pollution factor which has a good diagnosis in water quality evaluation. There are many methods in water quality evaluation and management, such as single pollution index, comprehensive pollution index, Nemerow pollution index, neural network, fuzzy clustering and fuzzy comprehensive evaluation, etc. These methods have higher precision, which can be used for water quality evaluation, but in the calculation process, high empirical is still needed, such as selection of implicit function, and creation of membership functions. The application of principal component analysis is broad, but not used much in water quality evaluation.

In this paper, authors used the grey clustering to find the main influencing factors to evaluate the water quality in Baoying country. The study of this paper has an important role in water evaluation and water control plan.

In the paper, we used cluster analysis method to determine the main factors affecting water quality assessment. Jining is a water abundant city, but the water quality is worse. So the water quality evaluation is imminent. By clustering analysis, we can determine the main monitoring site to reduce unnecessary sites.

STUDY AREA

Jining city is located in the southwest of Shandong province and the grand canal of Beijing to Hangzhou is through Jining, known as the reputation of the canal city. With the development of society, the pollution in Jining city is becoming more and more serious, and at the same time, because of the start of South to North Water Transfer Project, the strict requirements to water quality in Jining city have been put forward. The rivers are distributed in each aspect, which are different in every section. The monitoring site covered the entire city, which eventually led to multifarious difficulties in water quality evaluation.

Monitoring data: The selected data are the monitoring data of water quality in Jining city in 2002, including a total of 36 monitoring stations. The monitoring indicators include Cl, SO_4, total hardness, BOD5, CODMn, ammonia nitrogen, and phenol. The data are presented in Table 1.

CLUSTER ANALYSIS AND SPSS IMPLEMENTATION

Cluster analysis is an individual classification method according to the itself characteristics. The individual in the same category is very similar and individual differences are very large. There are many kinds of methods in clustering analysis. Multivariate statistical analysis includes system clustering method, dynamic clustering method and decomposition. While in fuzzy mathematics, clustering analysis

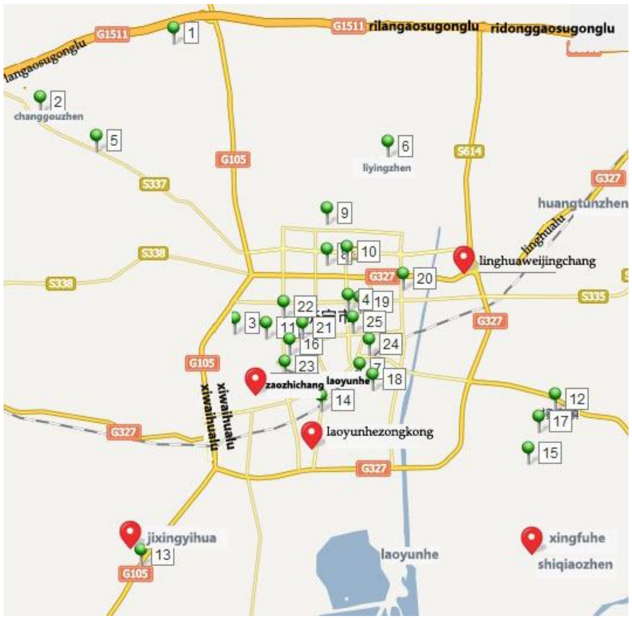

Fig. 1: The study area.

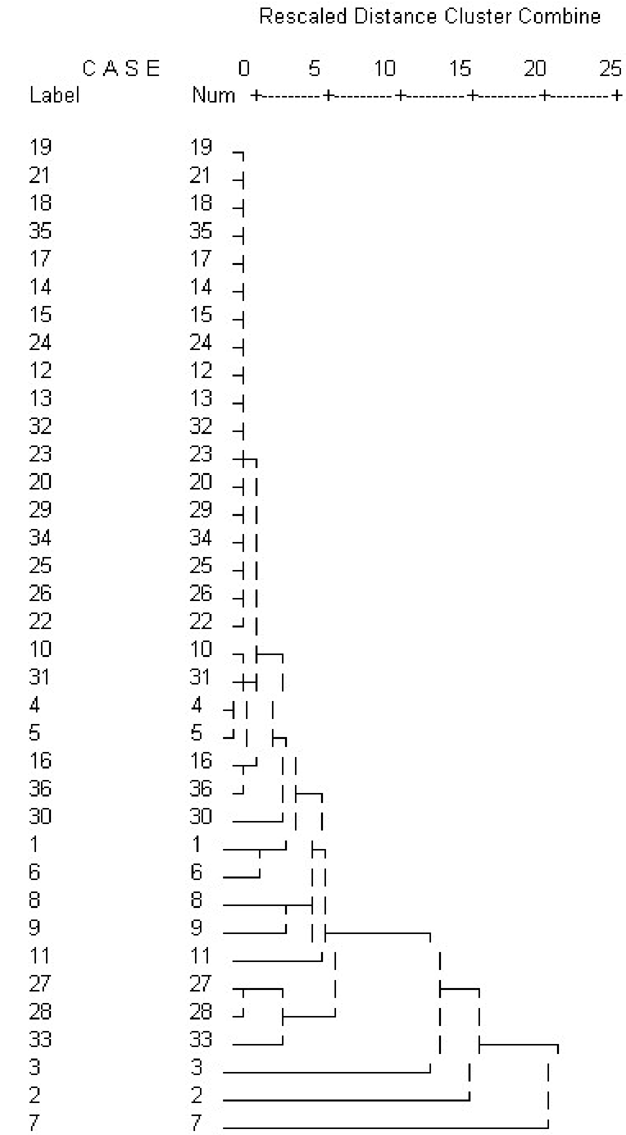

Fig. 2: Tree chart of water quality clustering.

includes maximum tree, transitive closure method and netting method. Clustering analysis method mainly has two kinds, one kind is "fast clustering method" (K-means clustering analysis), another is " hierarchical clustering analysis method" (hierarchical cluster analysis). The form of clustering analysis can be divided into two species, on the classification of samples, the type can be called as Q clustering, on the classification of observation variables of study object, the type can be called as R clustering. In this paper, Q cluster in fast clustering method is adopted. Supposing in a statistical problem, there is *m* variables and *n* samples, then,

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \qquad ...(1)$$

Where x_{ij} is the observation data of jth variable in ith sample. The observation of every sample $x_i = (x_{i1}, x_{i2}, \cdots x_{ik} \cdots x_{im})$ can be regarded as one point in *m* dimensional space. The distance d_{ij} between point to point in *m* dimensional space can be used to express the intimacy between samples. The calculation method of d_{ij} include Euclidean Distance Squared, Euclidean Distance Chebychev, Block distance, Minkowski distance, Customized distance. In this paper, the Euclidean distance is adopted

$$d_{ij} = \sqrt{\sum_{k=1}^{m}(x_{ik} - x_{jk})^2} \qquad ...(2)$$

A. Fast clustering analysis process and realization in SPSS

1. Quantification of sample index and standardized treatment.
2. According to the actual requirements, the cluster can be determined to classify many types.
3. Determination of initial central points of k classes. M dimension data in k group can be specified by the users, and also according to the sample data, k representative samples data can be as the initial center point.
4. Calculating the Euclidean distance of k center point. According to shortest distance principle of center points, all the samples are assigned to each center point class to form a new class k to complete an interactive process.

Table 1: the monitoring data of water quality in Jining.

Section	Cl	SO_4	Total hardness	BOD_5	COD_{Mn}	NH_4-NH_3	Nitrite nitrogen	Nitrate nitrogen	Phenol
1	156	240	460	120.2	40.8	6.96	0.003	0.32	0.37
2	228	211	392	430.2	200.8	170	0.003	0.05	0.103
3	283	211	444	19.2	41.1	0.05	0.461	9.27	0.526
4	108	183	326	4.2	15.5	0.05	0.105	1.22	0.002
5	127	204	408	1.7	16.3	0.05	0.187	0.6	0.002
6	182	167	430	287.7	147.5	0.05	0.003	0.42	0.093
7	314	215	344	850.2	608.3	0.05	0.003	0.35	0.322
8	282	553	396	100.2	76.3	0.05	0.003	0.6	0.042
9	401	188	328	23.7	29.9	0.05	0.003	0.54	0.002
10	119	92.2	396	8.9	9.6	9.15	0.174	0.18	0.002
11	284	232	422	6.9	28.8	19	0.393	0.25	0.006
12	25.5	5.76	262	0.3	1	0.05	0.003	0.4	0.002
13	26.6	15.4	260	0.8	1.9	0.05	0.003	0.4	0.003
14	30.1	19.2	276	0.4	1.3	0.05	0.003	0.29	0.002
15	26.6	26.9	276	0.1	0.8	0.05	0.003	0.34	0.002
16	85.4	188	573	0.2	1	0.05	0.003	3.18	0.002
17	18.8	9.61	302	1.4	1.2	0.05	0.003	3.25	0.002
18	31.2	23.1	300	0.4	1	0.05	0.003	1.51	0.002
19	17.7	30.7	288	0.9	1.9	0.05	0.003	2.37	0.002
20	28.7	19.2	106	0.3	3.3	0.05	0.003	0.32	0.002
21	17.7	28.8	286	0.5	1.7	0.05	0.003	2.4	0.002
22	82.2	42.3	400	1.7	0.9	0.05	0.003	0.87	0.005
23	13.5	26.9	240	0.4	1.1	0.05	0.003	0.09	0.002
24	32.3	23.1	284	0.4	1	0.05	0.003	0.05	0.002
25	64.5	117	362	0.4	0.7	0.05	0.003	3.29	0.002
26	51	110	398	0.2	1.5	0.05	0.003	1.28	0.002
27	88.3	99.9	685	0.6	1.3	0.05	0.017	18.7	0.002
28	141	19.2	524	0.2	0.7	0.05	0.003	13	0.002
29	76.2	38.4	368	0.2	1	0.05	0.003	6.18	0.002
30	231	23.1	20	0.2	1	0.05	0.003	7.73	0.002
31	36.5	48	370	0.3	1	0.05	0.16	0.46	0.002
32	11	23.1	280	0.2	0.6	0.05	0.003	0.3	0.002
33	187	121	867	1.8	1	0.05	0.003	30.9	0.005
34	85.4	46.1	424	0.2	1.2	0.05	0.003	6.76	0.002
35	35.4	23.1	296	0.2	0.8	0.05	0.003	1.64	0.002
36	89	215	536	0.3	1.2	0.05	0.003	0.5	0.002

5. Determination of center point of k class. Mean value of each variable can be calculated by SPSS which can be as a new center point.
6. Repeat step 4 and step 5, the calculating process can be terminated until meeting the requirements demand.

B. Data Standardization

In order to eliminate the unwanted error caused by size dimension, the collected data can be applied to the model after standardization. The principle of standard treatment is as following. If there are *n* samples, and there are *m* data in every sample, then the variable can be marked as X_{ij} ($i = 1,2 \ldots n$, $j = 1,2 \ldots m$), and the variable after standardization can be expressed as:

$$X_{ij}^{'} = \frac{X_{ij} - \overline{X}_j}{S_j}$$

Where $\overline{X}_j$ is the average of j^th variable, and S_j is the standard deviation of j^th variable.

C . SPSS Implementation of Fast Clustering

By clustering analysis, the monitoring data can be classified as the following tree chart. Then according to the Fig. 1, the monitoring can be classified as 5 kinds. Then using the method , water quality indexes also can be classified, which can be seen in Fig. 2.

D. Optimization of Monitoring Index

The monitoring indicators are as the parameters and the monitoring points are as a case. The variables are clustered. So the clustering tree can be got as shown in Fig. 3.

WATER QUALITY ASSESSMENT IN JINING

Jining is a famous city with much water, having river port

Table 2: Measured concentrations of assessment factos of water quality.

Surveying section	COD_{Mn}	COD_{Cr}	BOD_5	DO	NH_4- NH_3	Grey relation model	ANN model	Fuzzy comprehensive model
1	7.00	25.50	5.10	1.70	7.13	IV	IV	IV
2	7.30	30.10	5.40	2.60	5.52	IV	V	IV
3	7.00	24.10	5.50	2.60	6.73	IV	IV	IV
4	6.60	38.60	4.80	4.40	6.14	V	V	V
5	7.20	35.80	4.60	2.30	6.73	IV	IV	IV
6	10.00	30.70	15.80	0.90	9.66	I	I	I
7	11.90	43.20	27.80	1.80	8.60	I	I	I
8	7.80	45.20	15.60	0.60	7.13	IV	IV	IV
9	7.40	50.70	6.50	1.70	8.87	IV	IV	V
10	7.40	26.70	4.50	3.20	5.84	IV	IV	IV
11	7.10	28.80	5.80	3.70	4.35	IV	IV	IV
12	8.00	26.40	6.40	4.00	5.37	IV	V	IV
13	7.10	14.80	4.60	3.90	5.40	V	V	V
14	6.90	25.50	4.40	3.50	6.01	IV	IV	IV
15	7.70	26.40	7.40	2.20	6.54	IV	IV	IV
16	7.30	17.90	6.00	1.60	7.14	IV	IV	IV
17	7.00	5.50	5.50	1.30	6.12	I	I	I
18	10.10	50.60	9.40	0.50	6.91	IV	IV	IV
19	10.50	36.50	17.60	1.05	8.16	IV	IV	IV
20	14.20	58.50	27.20	25.00	11.90	IV	IV	V

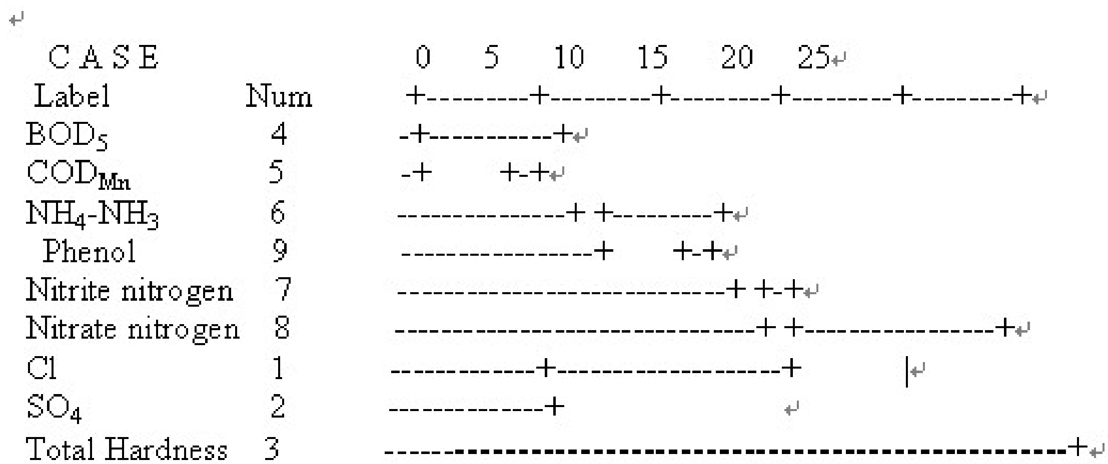

Fig. 3: Tree chart of water quality indicators clustering.

interlocks and numerous lakes. Water surface in whole city is approximately 3607 square kilometres, which approximately composes the total area of 42.52%. The urban district water surface is approximately 24 square kilometres, composing the urban district area of 20.15%. Suzhou altogether has more than 4000 bigger lakelets, the big lake swings have 87 and altogether have size rivers 20,000 with total length of 1457km. Outside the moat the Suzhou old city area circle will become a relatively independent region, spreading across the river course and has formed “three horizontal three straight link” the urban river course network of rivers and lakes system with the city. The rivers height is equally 0.8-1.3m , the water depth is equally 3 meters, the gradient is zero nearly. The river water excretes weakly, therefore, the fluent speed of flow is very small, mean velocity of stream is 0.05m/s-0.1m/s.

Water quality is divided into five grades according to the surface water environmental quality standard (GB3838-2002) issued by the government of China. Table 1 shows the surface water environmental quality.

The sequence which is made up of by the value in Table 2 for non-dimensional treatment can be seen as the reference sequence X_i , threshold concentration in the water quality standards ((GB3838-2002) can be seen as compared sequence Y_j . According to the above process, the grey relational de-

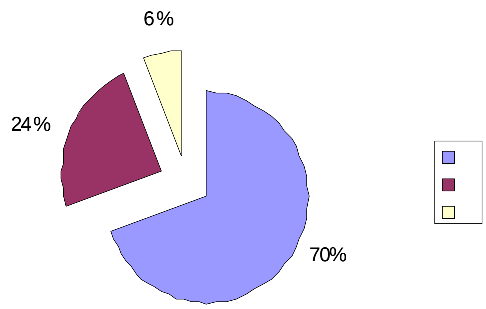

Fig. 4: Overview map of water quality of monitoring sites in Rencheng region during 1999-2002.

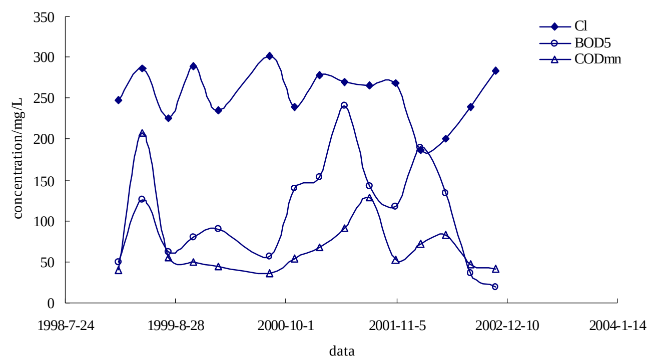

Fig. 6: Trend chart of water quality in Dayunhe river.

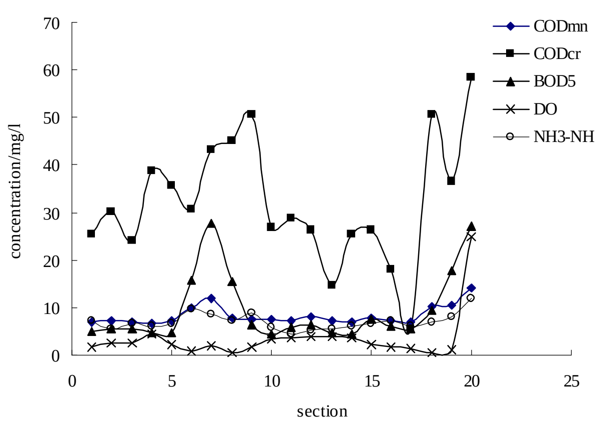

Fig. 8: Concentration in every section.

gree can be calculated and the evaluating result can be got. The results are given in Table 2. According to the communications of water quality, there are no one monitoring site to meet the I, II classification. The monitoring sites meeting III classification accounted for 70% of the monitoring sites. Monitoring sites of grade IV accounted for 24% of all monitoring sites and monitoring sites of grade V accounted for 6%. Statistical results are shown in Fig. 3.

Fig. 4 and Fig. 5 are the trends of water quality in

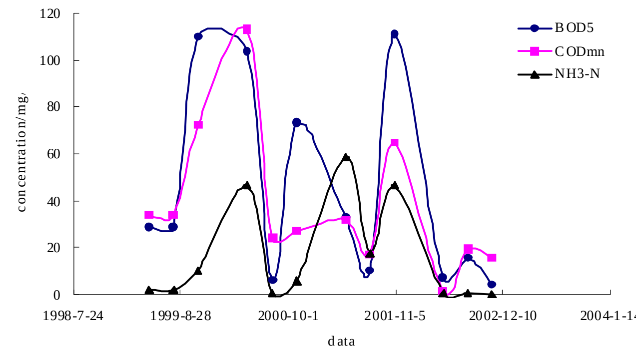

Fig. 5: Trend chart of water quality in Rencheng region.

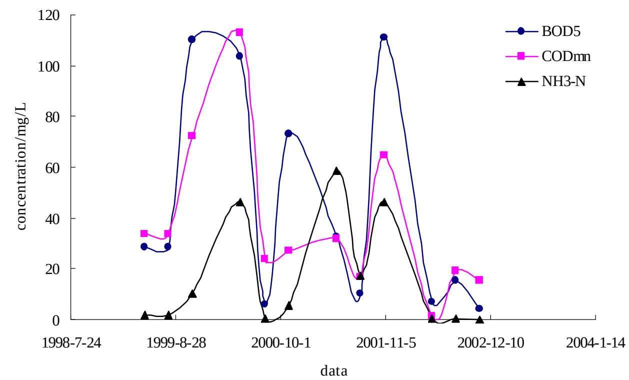

Fig.7: Trend chart of water quality in Houying of Dayunhe river.

Dayunhe river in Jining city. Seen from the Fig. 4 and Fig. 5, the water pollution is more serious with dry season having comparatively lower pollution.

In the years from 1999 to 2002, the concentration of the indicators are relatively more stable, in which the concentration of SO_4 increased with time (Fig. 5). In May 1999, COD_{Mn} and BOD_5 in May, 2001 have been the highest, probably due to decreased rainfall, because of which concentration in rivers increases.

Seen from Fig. 6, the concentration of most of the indicators is relatively stable, but the fluctuation is high. In August 2001, BOD_5 has the minimum value, and SO_4 and COD_{Mn} in December have the highest values. The value of total hardness increased with time and ammonia concentration increased in 2000, may be caused due to fertilizer use and the rain-wash.

Fig. 7 indicates large fluctuation of indicators' concentration from 1999 to 2002. Concentration of every indicator is lowest in August 2000, because August is the peak period, which diluted the river water, so that the concentration reduced. The concentration of SO_4 has a significant decline, indicating that the government intensified the pollution control.

Seen from Fig. 8, all indicators have relevance. Concentration in some sections is high, mainly because at upper river, there as waste discharge point.

CONCLUSIONS

Water quality indicators classification after cluster analysis, the selected main pollution index can distinguish various types of waters, and only the main pollution indices can be selected for the water quality assessment. By cluster analysis for monitoring site, the mean value of major pollution indices can represent the situation of the water quality. Using the mean value after classification, the water can be evaluated preliminarily. While the significant pollution of water need to be further chosen to analyse. By cluster analysis, surface water quality is poor. The vast majority of water quality exceeds the state environment standards. For surface water, the pollution is very serious.

ACKNOWLEDGEMENT

This study was funded by Open Foundation of State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering (2011491511), the program for Handan Science and Technology Research and Development (1123109066-4), and The Natural Science Foundation of Hebei Province (E2012402013).

REFERENCES

Chen Dongjing, Ma Anqing, Xu Zhongmin and Cheng Guodong 2002. The application of factor analysis in water quality evaluation. Hydrology, 22(3): 29-31.

Yang Wei, Lu Wenxi, Li Ping and Yang Zhongping 2007. Application of factor analysis method to the water quality evaluation of Yitong River. Research of Soil and Water Conservation, 14(1): 113-114.

Zhang Caixiang, Wang Yanxin and Zhang Zhaonian 2005. Application of factor analysis method to the water quality evaluation of lower Huanghaihe river. Water Resources Protection, 21(4): 11-15.