Exploring the Determinants of Soil and Water Conservation Measures with Data Mining Techniques

Dai Fuqiang[1,2,3], Liu Gangcai[1]
[1] Institute of Mountain Hazard and Environment, Chinese Academy of Science & Water Resources Ministry (CAS & WRM), Chengdu 610041, China
[2] Land and Resources College, China West Normal University (CWNU), Nanchong 637002, China
[3] Graduate School of the Chinese Academy of Sciences (GSCAS), Beijing 100039, China
daifq80@gmail.com, liugc@imde.ac.cn

ABSTRACT: This study explores different socioeconomic and environmental factors influencing the adoption of SWC (soil and water conservation) measures. As a consortium of national and international institutions, WOCAT (World Overview of Conservation Approaches and Technologies) has developed a database system to record specific details about the environmental and socioeconomic setting in soil and water management. Factor analysis, as a data mining technique, was used in the present work and the results show that adoption of SWC measures were influenced primarily by four factors including economic efficiency, human setting and land use, natural environment and soil quality. The first factor named as economic efficiency explained 18.888% of the total variance which comprised of variables like short term returns to establishment, long term returns to establishment, short term returns to maintenance and long term returns to maintenance. The second factor called human setting and land use explained 17.546% (comprised of variables like land ownership, off-farm income, market orientation, wealth, production subsidy and size of crop land per household) and the third factor called natural environment explained 14.945% (comprised of variables like average annual rainfall and slope) of the total variance, respectively. Furthermore, forth factor called soil quality explained 13.483% of the total variance comprised of soil fertility and topsoil.

KEYWORDS: determinant; adoption; soil and water conservation; factor analysis; data mining

I. INTRODUCTION

A common problem in data analysis occurs when a large number of features or variables hinder our investigating some patterns present in data [1]. As the presence of redundant or irrelevant features may results in a mask of underlying patterns of data, data mining is becoming an important tool to assist in the analysis of collections of observations. To be brief, data mining is to explore databases in search of additional useful information that has not yet been exploited [2].

Factor analysis seeks to reduce the dimensions of a data by principal component analysis finding orthogonal linear combinations of the variables that explain as much variance in the data as possible. Factor analysis played an important role as a data mining exploratory tool to extract the characteristics of data and to decrease high dimensionality of data [1, 3, 4].

WOCAT (World Overview of Conservation Approaches and Technologies) database is a tool for documenting and evaluating SWC (soil and water conservation) activities. Collection of information involves personal contact and sharing of knowledge between land users and SWC specialists. Nevertheless, few studies have described models for exploring the determinants of SWC measures with the large knowledge in the database, since considerable researches have been devoted to evaluate the on-site benefits the SWC measures have shown.

The main purpose of this paper is to investigate the influential factors of SWC adoption and provide implications for decision makers. This paper begins with a brief introduction of factor analysis approach and data collection. Next, the implications of results are discussed after presenting the results of factor analysis. Conclusions are made in the final section.

II. METHODS AND DATA

A. Factor Analysis

Factor analysis is a statistical method based on the correlation analysis of multiple variables to reduce the dimensionality of a set of data. The basic model of factor analysis is as follow.

$$\begin{aligned} x_1 &= \lambda_{11} f_1 + \lambda_{12} f_2 + \ldots + \lambda_{1m} f_m + \varepsilon_1 \\ x_2 &= \lambda_{21} f_1 + \lambda_{22} f_2 + \ldots + \lambda_{2m} f_m + \varepsilon_2 \\ &\vdots \\ x_p &= \lambda_{p1} f_1 + \lambda_{p2} f_2 + \ldots + \lambda_{pm} f_m + \varepsilon_p \end{aligned} \tag{1}$$

Where $x_1, x_2, \ldots, x_p$ are the observed variables and $f_1, f_2, \ldots, f_m$ are the factors; λ_{jk} ($j = 1, 2, \ldots, p$; $k = 1, 2, \ldots, m$) are constants called the factor loading; e_j ($j = 1, 2, \ldots, p$) are error terms.

Principal component analysis (PCA) has been widely used as the part of factor analysis. There are four main steps in factor analysis.

978-0-7695-3930-0/09 $26.00 © 2009 IEEE
DOI 10.1109/IFCSTA.2009.215

(1) Correlation matrix is calculated for each of the selected variables, which a $p \times p$ array of the correlation coefficients of the variables with each other. The following is the calculation equation.

$$r_{ij} = \frac{\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{kj} - \bar{x}_i)^2 \sum_{k=1}^{n}(x_{kj} - \bar{x}_j)^2}} \quad (2)$$

Where r_{ij} ($i, j = 1, 2, \ldots, p$) are the correlation coefficient between x_i and x_j of the selected variables.

(2) An appropriate number of factors to be extracted depend on calculating eigenvalue and eigenvector, together with calculating the variance percentage of the principal component and the cumulative variance percentage.

(3) The factor loading can be calculated as the equation below.

$$\lambda_{ij} = p(f_i, x_j) = \sqrt{\gamma_i} e_{ij},\ i, j = 1, 2, \ldots, p \quad (3)$$

(4) To clarify the relationship between the variables and factors, the factors are regularly rotated since the factors can not be interpreted clearly. There are various methods for carrying out this process of rotation. The most widely used is known as the varimax method.

B. Data

The data used in this study were obtained from the WOCAT database. WOCAT was established as a global network of SWC specialists from all over the world to realize a share of their valuable knowledge in soil and water management. It has developed tools to document, monitor and evaluate local SWC activities that are filled into a database for storage and use of the wealth of SWC experience, and to disseminate it around the globe in order to facilitate exchange of experience. The database is accessible in the form of books, CD-ROMs and maps, or via the internet to improve decision making. The WOCAT process and tools ensure systemic recording and collecting valuable information related to SWC activities as well as the environmental and socio-economic setting. A large data warehouse, providing basic information for study on evaluation of SWC measures in depth, has been developed with case studies on technologies and approaches by WOCAT, as well as geographic data. A representative sample of 221 SWC measures all over the world were selected in this study.

III. RESULTS AND DISCUSSION

A. Initial Solution

Factor analysis was used to explore the relationship between various factors and SWC measures. The task of first importance is to select variables because it is important for the reliability of the factor analysis. Thus, the correlation matrix was generated firstly between SWC measure adoption and each of indices involving natural environment, human environment, land use, and economic analysis by SPSS software. Fifteen variables were selected based on the high correlation influencing SWC measure adoption in TABLE I.

The correlation matrix results showed that these fifteen variables had a strong correlation with the others and they had considerable explanation strength. The KMO value of 0.622 meant that the degree of common variance among the fifteen variables was 'mediocre' and if a factor analysis is conducted, the factors extracted will account for a sufficient amount of variance.

TABLE I. VARIABLES MEASURED IN DETERMINING SOIL AND WATER CONSERVATION ADOPTION

Variables	Observation numbers
X_1: land ownership	203
X_2: market orientation	137
X_3: average annual rainfall	221
X_4: slope	221
X_5: soil fertility	215
X_6: topsoil	218
X_7: wealth	194
X_8: off-farm income	209
X_9: production subsidy	156
X_{10}: size of crop land per household	133
X_{11}: short term returns to establishment	201
X_{12}: long term returns to establishment	202
X_{13}: short term returns to maintenance	211
X_{14}: long term returns to maintenance	209
X_{15}: population density	169

B. Extracting and Rotating Factors

Principal component analysis, as the most commonly used method, was carried out to extract the various components from the intercorrelation matrix. As shown in Fig. 1, the scree plot including eigenvalues and percentage variance gives an idea on how the different principal components were extracted. TABLE II gives the initial solution values of unrotated factors, i.e. eigenvalues, percentage variance and cumulative percentage variance. The eigenvalues more than 1 were taken as a criterion for the extraction of the principal components required for explaining the sources of variances in the data. According to the criterion, it is found that four factors would be extracted in this study. The cumulative percentage of variance explained by these four factors was 64.862%. To make the interpretation of the factors more clear, the varimax rotation was performed for factor values. TABLE II and TABLE III reflect the difference of the eigenvalues and variance percentage corresponding to the principal components both before varimax rotation and after varimax rotation. However, this did not affect the cumulative percentage of the variances and the ranking of the factors.

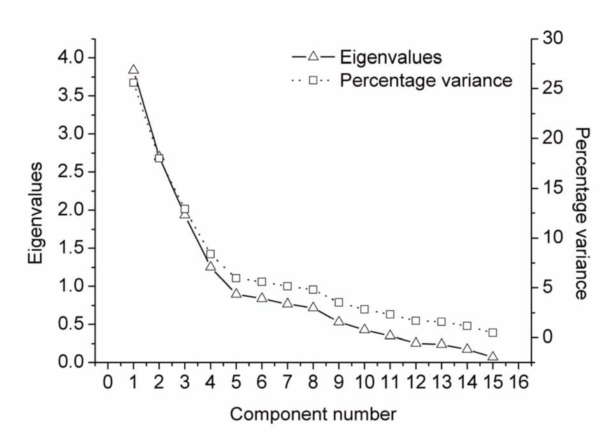

Figure 1. Scree plot of eigenvalues and percentage variance.

TABLE II. TOTAL VARIANCE EXPLANATION PERCENTAGES OF THE COMPONENTS FOR UNROTATED VALUES

Component	Eigenvalues	Variance %	Cumulative variance %
1	3.838	25.586	25.586
2	2.700	18.001	43.587
3	1.935	12.901	56.488
4	1.256	8.373	64.862
5	0.894	5.959	70.821
6	0.840	5.598	76.419
7	0.771	5.140	81.559
8	0.720	4.799	86.358
9	0.531	3.542	89.900
10	0.428	2.854	92.755
11	0.351	2.337	95.091
12	0.253	1.689	96.781
13	0.238	1.589	98.369
14	0.173	1.156	99.526
15	0.071	0.474	100.000

TABLE III. TOTAL VARIANCE EXPLANATION PERCENTAGES OF THE COMPONENTS FOR ROTATED VALUES

Component	Eigenvalues	Variance %	Cumulative variance %
1	2.833	18.888	18.888
2	2.632	17.546	36.434
3	2.242	14.945	51.379
4	2.022	13.483	64.862

C. Naming the Factors

The parameter loading for the four factors from the principal component analysis of the data set are given in TABLE IV. It is obviously that most of the variables associated with each factor are well defined and have little contributions to the other factors, which helps in the interpretation of results. The names were given as follow.

TABLE IV. ROTATED COMPONENT MATRIX

Variables	Component			
	1	*2*	*3*	*4*
X_{12}	**0.845**	0.172	0.253	-0.183
X_{14}	**0.827**	0.240	0.187	-0.176
X_{13}	**0.816**	-0.120	0.073	0.271
X_{11}	**0.681**	-0.392	-0.068	0.056
X_1	0.063	**0.713**	-0.161	0.044
X_8	-0.028	**0.662**	-0.308	-0.089
X_2	-0.146	**0.632**	-0.086	0.495
X_7	0.078	**0.593**	0.107	-0.024
X_9	-0.136	**0.506**	-0.467	0.328
X_3	0.129	-0.037	**0.801**	0.060
X_{15}	0.313	-0.025	**0.715**	0.061
X_{10}	-0.125	**0.530**	-0.558	0.270
X_4	-0.345	-0.309	**0.552**	0.128
X_5	0.064	-0.113	0.065	**0.859**
X_6	-0.032	0.178	0.036	**0.827**

- Factor 1, which had the strongest variance explanation level of 18.888%, is related to the four variables including short term returns to establishment, long term returns to establishment, short term returns to maintenance and long term returns to maintenance that had strong factor loads. The four variables provided information on the short-term and long-term returns in comparison with the establishment costs as well as maintenance costs from land user's perspective. Thus, the factor was named as 'economic efficiency'.
- Factor 2 explains 17.546% of the total variance and is positively loaded with land ownership, off-farm income, market orientation, wealth, production subsidy and size of crop land per household. This factor was related with land use rights, relative level of wealth, significance of off-farm income, government support and land size and might be termed as 'human setting and land use'.
- Factor 3 explains 14.945% of the total variance and loaded with average annual rainfall, population density and slope. This factor was related with the climatic condition and topography and was thus named as 'natural environment'.
- Factor 4 explains 13.483% of the total variance and is loaded positively with soil fertility and topsoil. This factor was related with the soil condition which is sufficient to support plant life and was named 'soil quality'.

D. Implication of the Factors

Results of the factor analysis indicate that the adoption of soil and water conservation measures was influenced by economic efficiency factor, human setting and land use factor, natural environment factor and soil quality factor.

Economic efficiency is positive and significantly correlated with adoption of SWC measures. These results suggest that peoples are prone to invest in soil and water practices if the benefits are positive in comparison with establishment cost or maintenance cost in a short and long term. This finding is reinforced by the work of Lockie who stated that farmers were aware of the effects of farming practices on profitability and sustainability, but were equally concerned with risk minimization and income stability [5]. Other studies have also found that producers will continue to adopt SWC systems as an economically viable means of sustaining their soil resource when these systems provide a yield advantage [6, 7].

With regard to human setting and land use, the results suggest that variables including land ownership, off-farm income, market orientation, wealth, production subsidy and size of crop land per household significantly influence the adoption of SWC. Previous studies revealed that adoption of SWC was affected by security of land tenure, the size of farms, off-farm income and economic assistance [8, 9].

Natural environment, such as slope and rainfall, plays an important role in adopting SWC measures. He stated that the slope of the field plays a critical role in the adoption of pasture crop rotation [10]. Similarly, the SWC measures were found to be profitable to farmers with low to medium opportunity costs of about on gentle to moderate slopes [11].

As the most popular indicator of soil quality, soil fertility is found to have a positive correlation with adoption of SWC measures. This is consistent with the finding that Investments geared towards extension with a focus on soil fertility maintenance have particular potential to raise incomes [12].

IV. CONCLUSION

The primary objective of this article is to explore the determinants of SWC measures adoption. To achieve this, factor analysis was applied to the WOCAT database. Four factors were identified and these results showed to measure economic efficiency, human setting and land use, natural environment and soil quality and proved to be important for the adoption of SWC measures. Based on the above discussion, it may be concluded that out of the 64.862% of variance explained by the four factors, it is the critical factor that explains most part of the observed variance in the data (18.888%). The parameters that are loaded in the factor include short term returns to establishment, long term returns to establishment, short term returns to maintenance and long term returns to maintenance. The importance of human setting and land use is taking a secondary role in the adoption of SWC measures. Similarly, natural environment and soil quality are regarded as important aspect in adopting SWC measures.

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers and thank for the financial support of National 973 program (2007CB407206) and National Sustentation Program (2006BAC01A11).

REFERENCES

[1] H. W. Cho, “A data mining-based subset selection for enhanced discrimination using iterative elimination of redundancy,” Expert Syst. Appl, vol. 36, Mar. 2009, pp. 1355–1361, doi:10.1016/ j.eswa.2007.11.020.

[2] P. Costantini, M. Linting, and G. C. Porzio, “Mining performance data through nonlinear PCA with optimal scaling,” Appl. Stoch. Model. Bus. Ind, May 2009, doi: 10.1002/asmb.771.

[3] D. T. Larose, Data Mining Methods and Models, New York, NY: Wiley, 2006.

[4] P. Giudici, Applied Data Mining for Business and Industrial Statistic, New York, NY: Wiley, 2000.

[5] S. Lockie, A. Mead, F. Vanclay, and B. Butler, “Factors encouraging the adoption of more sustainable crop rotations in south-east australia - profit, sustainability, risk and stability,” J. Sustain. Agr, vol. 6, Jul 1995, pp. 61–79, doi:10.1300/J064v06n01_06.

[6] R. S. Gray, J. S. Taylor, and W. J. Brown, “Economic factors contributing to the adoption of reduced tillage technologies in central Saskatchewan,” Can. J. Plant Sci, vol. 76, Oct 1996, pp. 661–668.

[7] F. H. D'Emden, R. S. Llewellyn, and M. P. Burton, “Factors influencing adoption of conservation tillage in Australian cropping regions,” Aust. J. Agr. Resour. Ec, vol. 52, Jun 2008, pp. 169–182, doi: 10.1111/j.1467-8489.2008.00409.x.

[8] M. J. Kipsat, “Socio-economics of soil conservation in Kericho district, Kenya,” Advances in Integrated Soil Fertility Management in Sub-Saharan Africa: Challenges and Opportunities, Springer, Oct 2007, pp. 1001–1012, doi: 10.1007/978-1-4020-5760-1_97.

[9] P. Winters, C. C. Crissman, and P. Espinosa, “Inducing the adoption of conservation technologies: lessons from the Ecuadorian Andes,” Environ. Dev. Ec, vol. 9, Oct 2004, pp. 695–719, doi: 10.1017/s1355770x04001482.

[10] J. He, H. W. Li, X. Y. Wang, A. D. McHugh, W. Y. Li, H. W. Gao, et al., “The adoption of annual subsoiling as conservation tillage in dryland maize and wheat cultivation in northern China,” Soil Till. Res, vol. 94, Jun 2007, pp. 493–502, doi: doi:10.1016/j.still.2006.10.005.

[11] A. J. Tenge, J. De graaff, and J. P. Hella, “Financial efficiency of major soil and water conservation measures in West Usambara highlands, Tanzania,” Appl. Geog, vol. 25, Oct 2005, pp. 348–366, doi: 10.1016/j.apgeog.2005.08.003.

[12] H. G. P. Jansen, A. Rodriguez, A. Damon, J. Pender, J. Chenier, and R. Schipper, “Determinants of income-earning strategies and adoption of conservation practices in hillside communities in rural Honduras,” Agr. Syst, vol. 88, Apr 2006, pp. 92–110, doi: 10.1016/j.agsy.2005.06.005.