International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:6, No:5, 2012

Principal Component Analysis for the Characterization in the Application of Some Soil Properties

Kamolchanok Panishkan, Kanokporn Swangjang, Natdhera Sanmanee, and Daoroong Sungthong

Abstract**—**The objective of this research is to study principal component analysis for classification of 67 soil samples collected from different agricultural areas in the western part of Thailand. Six soil properties were measured on the soil samples and are used as original variables. Principal component analysis is applied to reduce the number of original variables. A model based on the first two principal components accounts for 72.24% of total variance. Score plots of first two principal components were used to map with agricultural areas divided into horticulture, field crops and wetland. The results showed some relationships between soil properties and agricultural areas. PCA was shown to be a useful tool for agricultural areas classification based on soil properties.

Keywords**—**soil organic matter, soil properties, classification, principal components

I. Introduction

PRINCIPAL component analysis (PCA) is a multivariate analysis technique and is also known as eigenvector analysis. It is usually applied in environmental and agricultural studies [1- 4]. In this work, PCA was used to cluster patterns of 67 soil samples collected from different agricultural areas in Thailand based on their properties. Agricultural areas are divided into three crop types including horticulture, field crops and wetland, which have the different agricultural practices such as rates of fertilizers and pesticides. Some key parameters related to soil features are observed. Those are organic matter and nitrogen contents determined as soil fertility, cation exchange capacity determined as soil chemical characteristics, bulk density, percentage of silt and clay determined as soil physical characteristics. The aims of this study were (1) to determine contents of soil samples included Organic Mater (OM), Total Nitrogen (TN), Cation Exchange Capacity (CEC), Bulk Density (BD), Percentage of Silt (Silt) and Percentage of Clay (Clay) (2) to compare differences of means of contents between agricultural areas divided into horticulture, field crops and wetland by using analysis of variance.(3) to study classification of soil samples according to their properties by using principal component analysis.

K. Panishkan is with the Department of Statistics, Faculty of Science, Silpakorn University, Nakorn Pathom, Thailand (e-mail:kamolcha@su.ac.th).
K. Swangjang, N. Sanmanee and D. Sungthong are with the Department of Environmental Science, Faculty of Science, Silpakorn University, Nakorn Pathom,Thailand (e-mail: knokporrn@su.ac.th, sanmanee@su.ac.th and daoroong@su.ac.th).

II. Materials And Methods

A. Soil samples used in this study

The soil samples were collected from three agricultural locations in the western region of Thailand; Nakorn Pathom, Samut Sakorn, and Samut Songkram provinces and 67 sites in all. The soil samples were examined and soil properties were measured that contain 6 parameters included Organic Mater (OM), Total Nitrogen (TN), Cation Exchange Capacity (CEC), Bulk Density (BD), Percentage of Silt (Silt) and Percentage of Clay (Clay) .

B. Reduction of the Number of Original Variables

Principal component analysis (PCA) [5] is a technique to reduce the number of variables and eliminate the relations among input variables by developing a set of new variables that are linear functions of the original variables. This set will retain properties of the original ones, provided that the number of new variables will not exceed the original number. That is, if the original number of variables is p and the number of new variables is m, then $m \leq p$. The number of variables m is chosen components to sufficiently explain the variation of the data.

C. Data Analysis

The data were analyzed by SPSS for windows version 18 using one-way ANOVA analysis followed by Duncan test or Kruskal-Wallis test, depending on whether the normality assumptions were met. Differences of means were calculated to compare significant effects at the 5% significant level. PCA is used for characterization of soil samples according to soil properties.

III. Results And Discussions

Basic statistics of soil properties included Organic Mater (OM), Total Nitrogen (TN), Cation Exchange Capacity (CEC), Bulk Density (BD), Percentage of Silt (Silt) and Percentage of Clay (Clay) and are shown in Table 1. The data were analyzed using a one-way ANOVA or Kruskal-Wallis test, depending on whether the assumptions of normality is met. The significant difference between agricultural areas was showed in Table 2. The results of multiple pairwise tests are identified by letters followed by the means. Means with the same letters do not differ significantly from each other ($P>0.05$).

TABLE I
Basic Statistics of Soil Properties

Property	Unit	Mean	SD.
OM	%	1.2363	0.8836
TN	%	0.1524	0.0543
CEC	meq/100g	26.7396	9.2178

BD	g/cc	1.1563	0.1610
Silt	%	31.0079	6.2557
Clay	%	67.6007	10.9348

TABLE II
MEANS OF SOIL PROPERTIES OF THREE AGRICULTURAL AREAS

	OM	TN	CEC	BD	Silt	Clay
horticulture	1.137^{b}	0.156^{b}	30.96^{a}	1.159^{b}	32.23^{a}	73.69^{a}
field crops	0.641^{c}	0.102^{c}	15.97^{b}	1.282^{a}	27.52^{b}	57.54^{b}
wetland	2.594^{a}	0.217^{a}	26.64^{a}	0.941^{c}	31.58^{ab}	58.71^{b}

The results of one-way ANOVA presented in Table II show that means of all soil properties do significantly differ among agricultural areas ($P < 0.05$). These results are very informative and confirm that soil properties are related to agricultural areas which have the different agricultural practices such as rates of fertilizers and pesticides.

To apply PCA, the variables were standardized due to the difference in the units of measurement. The applicability of the PCA to the data sets used in this study was verified through the application of Bartlett's sphericity test. The null hypothesis states that the population correlation matrix is an identity matrix. If the obtained chi-square value is significant, then PCA should then be applied. The result from the hypothesis test showed that the chi-square value was equal to 182.12 ($P < 0.05$). Rejecting the hypothesis means that the strength of the relationships among the variables are strong and appropriate for PCA. To determine the number of components to retain, one of the most commonly-used criterion, called the eigenvalue-one criterion, was applied. With this criterion, the first two principal components with an eigenvalue greater than one were retained and were shown in Fig. 1. Table III indicates the loading values of the first two principal components. These loadings explain the contribution of each variable in a principal component. The bold number means the variable loads on that component (loadings > 0.4). The first principal component explains 41.57% of total variation and the second one 30.67%; a two component model thus accounts for 72.24% of the total variance. Fig.2 indicates that the first PC showed high loadings of OM and TN with positive effect and BD with negative effect. The second PC was associated with CEC and Clay with both positive effects. Silt has little effect on both PC1 and PC2. Fig.4 clearly shows that the samples are clustered into three different groups which correspond to the agricultural areas.

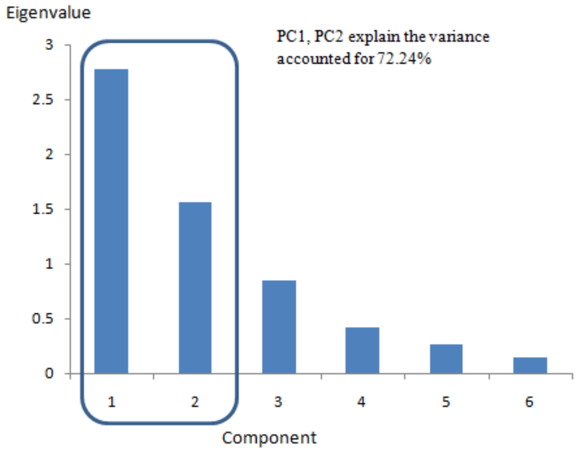

Fig. 1 A scree plot for all components

TABLE III
LOADING VALUES OF THE FIRST 2 PC's FROM SOIL SAMPLES

Component	PC1	PC2
OM	**0.894**	0.011
TN	**0.899**	0.241
CEC	0.223	**0.869**
BD	**-0.842**	-0.067
Silt	0.343	0.384
Clay	-0.100	**0.936**
Eigenvalue	2.494	1.840
Accumulated variance	41.57	72.24

Samples from field crop are characterized by negative values of PC1 and PC2, taking into account higher contribution of BD by negative value of PC1.

Samples from wetland are characterized by positive values of PC1, taking into account higher contribution of OM and TN, and negative values of PC2.

Samples from horticulture are characterized by positive values of PC2, taking into account higher contribution of CEC and Clay by positive value of PC2.

Considering these results, it appears that the first principal component might be considered as the indicator between field crop and wetland and the second principal component might be considered as the indicator between horticulture and the others. A plot of loadings in Fig.2 and score plots in Fig.3 showed some relationships among agricultural areas divided into horticulture, field crops and wetland.

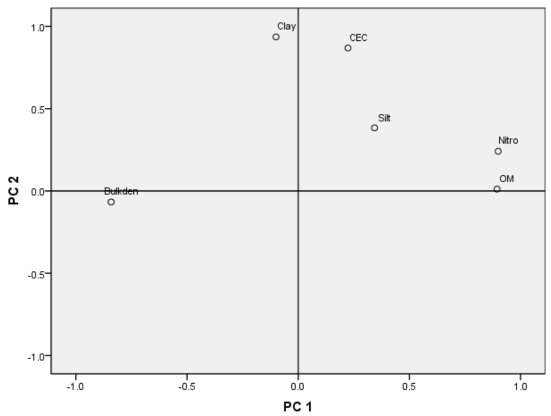

Fig. 2 Plot of loading corresponding to the first two principal components

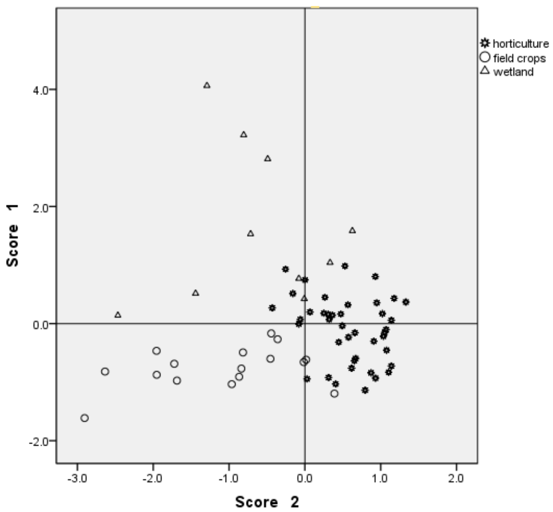

Fig. 3 Score plot of PC1 versus PC2 indicating the differentiation of soil samples according to agricultural areas

IV. Conclusion

The 67 soil samples were collected from three agricultural areas, divided into horticulture, field crops and wetland, in the western region of Thailand. Six different parameters were measured including OM, TN, CEC, BD, Silt and Clay. Principal component analysis is applied to define grouping of soil samples. The first two principal components were obtained. The first principal component explains 41.57% of total variation and the second one explains 30.67%. A model based on the first two principal components accounts for 72.24% of total variance.

In conclusion, the PCA helped to reveal some relationships between some soil properties and agricultural areas. It has proved to be useful approaches to characterization of soils based on their properties.

Acknowledgment

This research was funded by Thailand Annual Government Statement of Expenditure (2011) under Thailand Office of the National Research Council, and coordinated by Research and Development Institute, Silpakorn University, Thailand.

References

[1] Boruvka L., Vacek O. and Jehlicka J., 2005. Principal component analysis as a tool to indicative the origin of potentially toxic elements in soils. Geoderma 2005;128, pp.289-300.

[2] Dragovic S. and Onjia A. Classification of soil samples according to their geographic origin using gamma-ray spectrometry and principle component analysis. Journal of Environmental Radioactivity 2006; 89, pp.150-158.

[3] Sousa S.I.V., Fernando G. Matins, Maria C. Alvim-Ferra, and Maria C. Pereira. "Multiple linear regression and artificial neural networks based on principal component to predict ozone concentrations." Environmental Modelling & Software 22, 1 (January 2007), pp.97-103.

[4] Mico C., Recatala L., Peris M. and Sanchez J. Assessing heavy metal sources in agricultural soil of an European Mediteranean area by multivariate analysis. Chemoshere 2006; 65, pp.863-872.

[5] Jolliffe I.T. Principal component analysis. Springer-verlag, Newyork.1986.