International Research Journal of Pure & Applied Chemistry

21(10): 58-65, 2020; Article no.IRJPAC.58410
ISSN: 2231-3443, NLM ID: 101647669

Principal Component Analysis on Soil Fertility Parameters of Vegetable Growing Panchayats/ Locations in Ernakulam District of Kerala

A. Muhsina[1*], Brigit Joseph[1] and Vijayaraghava Kumar[1]

[1]Department of Agricultural Statistics, College of Agriculture, Vellayani, Thiruvananthapuram, Kerala, India.

Authors' contributions

This work was carried out in collaboration among all authors. Author AM designed the study, performed the analysis and wrote the first draft of the manuscript. Authors BJ and VK supervised the study. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/IRJPAC/2020/v21i1030208
Editor(s):
(1) Dr. Wolfgang Linert, Vienna University of Technology, Austria.
Reviewers:
(1) Ismael Ramza Haman, University of Ngaoundere, Cameroon.
(2) Aliyu Danladi, Gombe State University, Nigeria.
Complete Peer review History: http://www.sdiarticle4.com/review-history/58410

Original Research Article

Received 17 April 2020
Accepted 24 June 2020
Published 04 July 2020

ABSTRACT

The present paper used Principal Component Analysis (PCA) on 13 soil fertility parameters including soil pH and electrical conductivity of 17 vegetable growing panchyat/locations in Ernakulam district of Kerala based on 583 soil samples. Soil pH of panchayats varied from 4.2- 5.8 with a coefficient of variation 3.16-12.23 per cent and it was inferred that most of the panchayats in the district had very strongly acidic (pH: 4.2-5) and strongly acidic soils (pH: 5-5.5). High level of organic carbon content was noticed in most of the panchayats except in four panchayats. The results of PCA revealed that five PC's together explained a total variability of 80 per cent and the remaining PCs accounted for 20 per cent of the variability in the data which has been discarded from further analysis. First principal component accounted for 25 per cent variance followed by PC 2(21%), PC 3(14%), PC 4(10%) and PC 5(10%). Factor analysis generated five factors and they explained 85 per cent of variability. Score plot drawn as part of PCA showed that Chengamanadu, Manjapra and Thirumaradi panchayats had high content of soil available S and B. EC was also found to be higher in these panchayats. Amount of OC, Fe and Mn were more in Kalady, Keerampara and Mudakkuzha of Ernakulam district whereas Thuravur, Piravom and Pothanikkad

**Corresponding author: E-mail: rizzwana99@gmail.com;*

had highly acidic and Mg rich soils. Amount of Zn was more in Vengoor panchayat. Available K, Ca, P and Cu were found to be higher in Kakkad, Nedumbassery, Vengola and Kadungalloor. Based on the fertility status of each panchayats, they could be classified into different groups.

Keywords: Principal component analysis (PCA); factor analysis; principal component (PC); oxidisable organic carbon (OC); electrical conductivity; score plot.

1. INTRODUCTION

Ernakulam, a central district of Kerala spans over an area of 2407 square kilometers. It is surrounded by four districts viz., Thrissur in the north, Idukki in the east, Alappuzha and Kottayam in the south (Fig. 1). As Ernakulam is situated below mean sea level, paddy is the only crop possible for cultivation. Mainly rice – fish farming system is practiced in the Vyttila region of Ernakulam district [1]. Soil health plays a crucial role in the production and productivity of crops [2]. Soil fertility has an amazing influence on the plant growth and development. Primary nutrients (N, P, K); secondary nutrients (Ca, Mg, S) and micro nutrients (B, Fe, Mn, Zn, Cu) are required in sufficient quantity for proper growth. Moreover, soil conductivity and salinity have influence on plant growth [3].

When there is large number of variables, it may be difficult to interpret the data as there will be redundancy of information among the correlated dependent variables. Principal component analysis is one of the multivariate techniques in which interdependency of variables are investigated without specifying them as dependent or independent. Simultaneous analysis of several factors is possible with this technique [4]. PCA is a tool for identifying relationships within a single set of variables by transforming the original variables to a number of independent linear combinations of original variables. Principal components should retain as much of the information contained in the original variables as possible [5]. PCA can be adopted to distinguish agricultural plots as a function of soil management and to determine the most important soil parameters to characterize them [6]. PCA was reported to be one of the best methods to identify the variations in both physical and chemical soil parameters [7]. PCA can also be used to establish geochemical groupings of soils [8]. It also helps to eliminate the variables which are least important and contribute little in terms of variation [9].

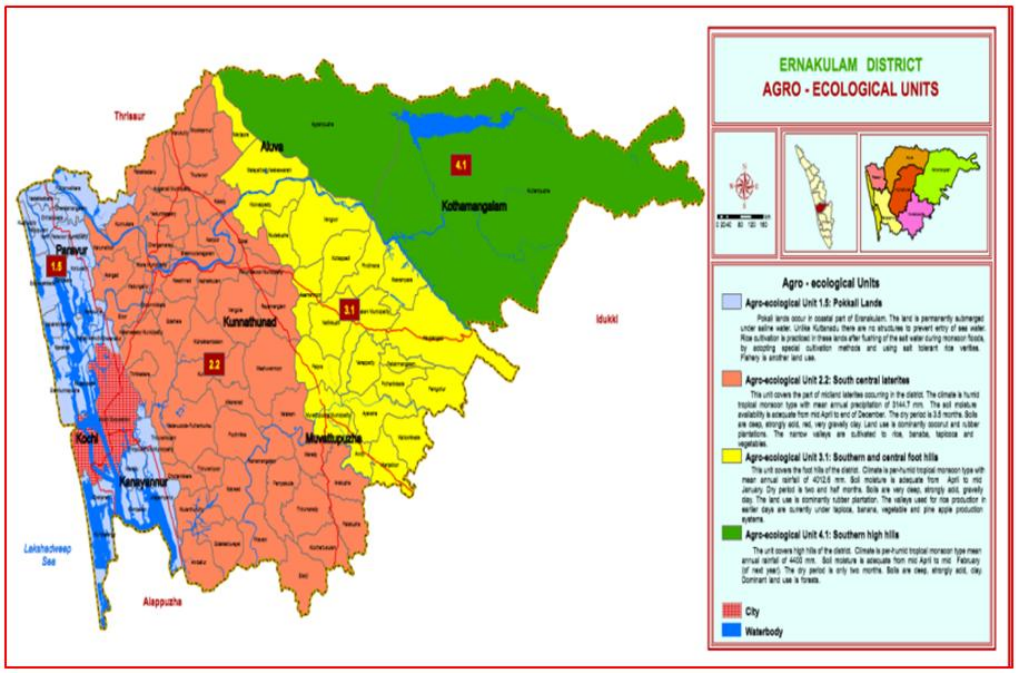

Fig. 1. Map of Ernakulam district, Kerala

Like PCA, factor analysis is also a data reduction technique wherein the original variables are transformed into linear combination of some unobservable factors. PCA and factor analysis could be employed as effective methods to reduce the dimension of soil variables for an easy interpretation [10]. Factor analysis and PCA were applicable only if the sample size was adequate.

The objective of this paper is to show how PCA and factor analysis were employed in reducing the dimensionality of the data set as well as to assess the soil fertility status of different locations and to identify the important parameters in terms of variation noticed.

2. MATERIALS AND METHODS

Soil samples collected from different panchayats of Ernakulam district as part of the project titled "Revolving fund mode project- Soil testing lab" during 2014-15 analysed by Department of Soil science & Agricultural Chemistry, College of Agriculture, Vellayani, Kerala and the data maintained was utilized for the present study. Samples were collected by the farmers themselves from their own vegetable growing plots at a depth of 15 cm using spade. About 10-15 samples from random locations were collected from each sampling plot; mixed them and reduced to 0.5 Kg by the method of quartering. It was observed that panchayats show variation in soil properties with cropping patterns and cultivation practices. Moreover, some of the soil parameters are more prevalent in certain region and important for certain crops.

The data on thirteen soil fertility parameters of 17 panchayats of Ernakulam viz., electro chemical parameters (pH and Electrical conductivity (EC)), Oxidisable Organic Carbon (OC), Phosphorus (P), Potassium (K)), secondary nutrients (Calcium (Ca), Magnesium (Mg), Sulphur (S)) and micro nutrients (Boron (B), Iron (Fe), Manganese (Mn), Zinc (Zn), Copper (Cu)) were available. Each panchayat is having different sample sizes and altogether sample size comes around 583.

2.1 Principal Component Analysis

Data reduction and summarization of a large set of variables is possible with the technique of Principal component analysis. Even though this technique primarily concerns with data reduction, condensation of information contained within a large set of variables with minimum loss of information is obtained through PCA [11]. Artificial variables called principal components are generated by the linear transformation of original variables and are uncorrelated to each other. Principal components maintain most of the information contained in the original study variables. Each PC corresponds to certain amount of variability and first two PCs usually provide most of the variation within the data [12].

PCA was selected as the investigative procedure as it is commonly used to reduce the dimensionality of the data and will provide more information than single element distribution. In view of the fact that the variables under study were correlated, PCA is the effective method to be applied to reveal variability of the data set [13]. Scores were given to different locations based on their soil fertility status by the method of PCA and locations with nutrient deficiency and sufficiency were identified.

2.1.1 Definition of principal components

Let $X^T = [X_1 \ X_2 \ \ldots \ X_p]$ be some random vector consisting of p random variables with mean vector μ and variance matrix Σ. In PCA the original variables are transformed into linear combination which is given by,

$$PC1 = a_1 X^T = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p = \sum_{i=1}^{p} a_{1i} x_i, \quad (1)$$

subjected to the condition $a_1^T a_1 = 1$.

Here coefficient matrix is denoted as a_1 in case of PC1, a_2 in case of PC2 and so on.

Similarly a number of such PC_s can be generated as the number of original variables which will be uncorrelated and orthogonal to each other. PCA is recommended only when the original variables are correlated. If the original variables are uncorrelated, there is no point of doing PCA as PCA generate new variables which are uncorrelated to each other. Only first few PCs will be enough to explain necessary amount of variability in the dataset. Variance of PC is calculated as follows,

Var (*PC1*) = λ_1, Var *(PC2*) = λ_2 and so on.

Total variance explained by these PC_S is,

$\lambda_1 + \lambda_2 + \ldots + \lambda_p$, where λ_1, λ_2 …are characteristic roots.

Proportion of variance explained by i^{th} *PC* is,

$$\frac{\lambda i}{\lambda 1 + \lambda 2 + \ldots + \lambda p}$$

2.1.2 Score plot

Scores of each location is calculated as the sum of product of principal component loadings with their corresponding means. Scores of each observation is plotted on score plot.

2.2 Factor Analysis (FA)

Unlike PCA, in factor analysis each variable is written as the linear combination of unobservable random component known as factors. These factors represent the importance of each variable and are orthogonal to each other. It explains the interrelationship among the large set of variables.

Let $X_1, X_2, \ldots X_p$ be some *p* random variables and X^T denotes the vector with mean μ and variance Σ. Each variable is linearly related to factors and error variance.

$$X_1 - \mu_1 = l_{11} F_1 + l_{12} F_2 + \ldots + l_{1m} F_m + \mathcal{E}_1 \qquad (2)$$

Similarly *p* variables can be written. The coefficients l_{ij} is called as the factor loadings. Statistical problem can be explained by using the relevant factors obtained from factor analysis. Number of factors retained is based on the Kiser criteria which say that factors with eigen values greater than unity is important.

3. RESULTS AND DISCUSSION

Principal Component Analysis was carried out on 583 soil samples from the 17 vegetable growing panchayats of Ernakulam district based on 13 soil testing variables to generate smaller number of new variables termed as Principal Components (PCs). Since the data contained 13 soil testing variables, there were 13 PCs (Table 1). This was in conformity with the results of Zambon et al. [14] and Baggie et al. [15]. There were five principal components which had eigen values greater than unity. These five PCs together explained a total variability of 80 per cent and the remaining PCs accounted for 20 per cent of the variability in the data which has been discarded for further analysis. First principal component accounted for 25 per cent variance followed by PC 2(21%), PC 3(14%), PC 4(10%) and PC 5(10%). Maximum variance was recorded by the first principal component followed by the PC 2, PC 3 etc. The first two PCs together accounted for a total variance of 46 per cent.

3.1 PC 1

PC 1 explained 25 per cent of total variability in the data. Loadings of each soil fertility variable on 1st PC have been described using the graph given in Fig. 2. High loadings was observed for B, Fe, OC and S (> 0.2) while high negative loadings were noticed (< -0.2) for pH, Ca and K. Since Fe has chelating property, it will reduce the availability of Ca in soil and hence are negatively related. Similar findings were also reported by Dawes and Goonetilleke [16].

3.2 PC 2

For the second principal component PC2; P, EC and S was found to be the soil testing variables with high loadings (> 0.2) and high negative loadings were recorded for the variables Mn and Zn(< -0.2). Loadings of each variable on the PC2 are given in Fig. 3. Mihai [17] also used PCA to identify the important variables out of many under study.

3.3 Principal Component Scores

Scores of each panchayat was estimated by multiplying the loadings of each variable on PCs with corresponding mean values of thirteen soil fertility parameters and the standardized scores of the first two PCs are presented in Table 2. Range of scores in PC1 and PC 2 for the panchayats was from -3.83 to 3.49 and -3.14 to 3.03 respectively. Highest score based on PC 1 was recorded for Ayyampuzha irrespective of its sign and Vengoor with respect to PC2.

3.4 Factor Analysis on Soil Fertility Parameters

Factor analysis by the method of principal component factor was used to reduce the

dimensionality. There were five factors which had characteristic roots greater than unity and they were used to represent each soil testing variable as a linear combination of five unobservable factors and these five factors together accounted 85 per cent of total variance. Factor 1 had high loadings on the variables pH (-0.65), K (-0.74), Ca (-0.65) and B (0.78), Fe (0.53) and S (0.53). Similarly, a high positive loading was recorded for P (0.82), EC (0.55) and S (0.51) and negative loading for Zn (-0.75) and Mn (-0.56). Factor 1 explained a variance of 27 per cent followed by factor 2 (22%), factor 3 (15%), factor 4(11%) and factor 5 (10%). Factor 1 and 2 together accounted for a total variance of 50 per cent approximately. The findings of Hammer et al. [18] are in accordance with the present result.

The scores of first two PC's were further plotted along with factor loading of the thirteen soil fertility parameters (Fig. 4). Score plot drawn as part of PCA showed that Chengamanadu, Manjapra and Thirumaradi panchayats had high content of soil available S and B (Fig. 4). Length of each vector representing the soil fertility variable indicated the relative contribution of that particular variable on panchayat. Identical results were also expressed by Lammel et al. [19]. It was seen that K, P, B and pH had a greater contribution in the unobservable factors.

Table 1. Principal components generated through PCA in Ernakulam district

Variable	PC 1	PC 2	PC 3	PC 4	PC 5
pH	-0.35	-0.15	0.29	-0.23	-0.40
EC	0.16	0.36	0.47	0.15	0.06
OC	0.31	-0.13	0.38	0.17	0.22
P	-0.05	0.48	0.21	-0.07	0.12
K	-0.42	0.15	0.31	-0.04	0.21
Ca	-0.37	0.10	0.28	0.12	-0.10
Mg	-0.15	-0.19	0.20	-0.27	0.63
S	0.28	0.32	0.21	-0.13	-0.46
B	0.44	0.08	-0.07	-0.34	0.17
Fe	0.33	-0.20	0.33	0.24	-0.08
Mn	0.17	-0.38	0.30	-0.10	0.05
Zn	-0.08	-0.46	0.15	0.31	-0.20
Cu	-0.04	0.18	-0.16	0.72	0.20
Eigen value	3.22	2.72	1.82	1.36	1.27
Proportion of variance (%)	25.00	21.00	14.00	10.00	9.00
Cumulative variance (%)	25.00	46.00	60.00	70.00	80.00

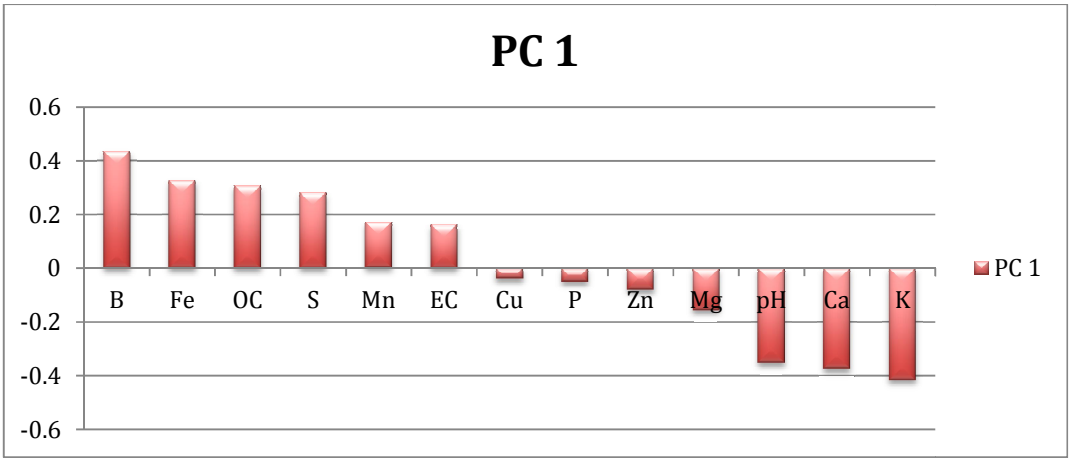

Fig. 2. Element loadings on PC 1 in Ernakulam

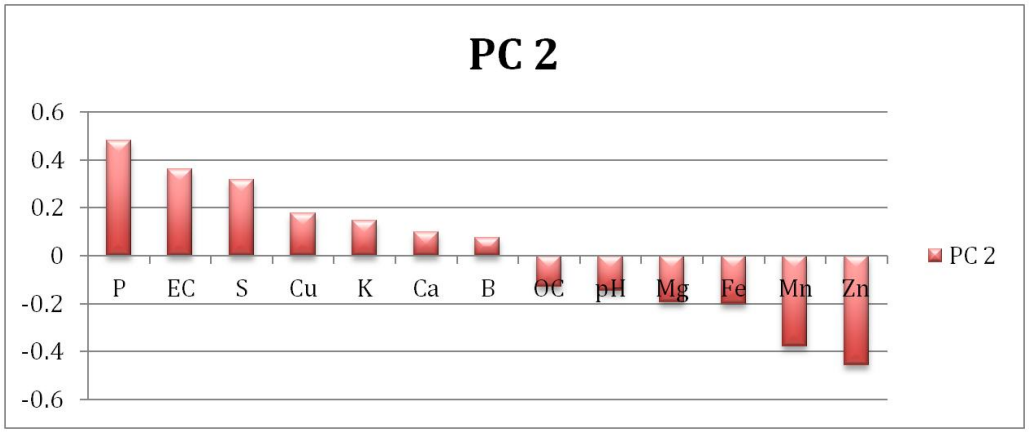

Fig. 3. Element loadings on PC 2 in Ernakulam

Table 2. Standardized scores of different panchayats of Ernakulam based on PC1 and PC 2

Panchayat		PC1	PC2
Ayyampuzha	(E1)	-3.83	-0.07
Chengamanadu	(E2)	0.92	3.03
Kadungalloor	(E3)	-0.75	1.48
Kakkad	(E4)	-0.96	0.47
Kalady	(E5)	0.51	-0.56
Keerampara	(E6)	1.67	-2.43
Manjapra	(E7)	1.92	1.87
Mudakkuzha	(E8)	1.98	-1.69
Nedumbassery	(E9)	-0.51	0.59
Pampakuda	(E10)	2.00	-0.04
Piravom	(E11)	-0.04	-0.80
Pothanikkad	(E12)	-0.81	-0.76
Puthenvelikkara	(E13)	-2.38	2.68
Thirumaradi	(E14)	3.49	0.52
Thuravur	(E15)	-1.08	-1.68
Vengola	(E16)	-0.41	0.56
Vengoor	(E17)	-1.70	-3.14

Table 3. Results of factor analysis on soil fertility variables in Eranakulam district

Fertility parameters	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
pH	-0.65	-0.20	0.38	-0.24	-0.50
EC	0.29	0.55	0.57	-0.03	0.19
OC	0.54	-0.24	0.56	0.11	0.31
P	-0.06	0.82	0.26	0.14	0.01
K	-0.74	0.29	0.39	0.23	0.08
Ca	-0.65	0.20	0.33	-0.17	0.10
Mg	-0.30	-0.30	0.28	0.76	0.04
S	0.53	0.51	0.31	-0.39	-0.38
B	0.78	0.09	-0.08	0.36	-0.24
Fe	0.53	-0.33	0.42	-0.17	0.15
Mn	0.25	-0.56	0.36	0.09	-0.09
Zn	-0.18	-0.75	0.22	-0.37	0.18
Cu	-0.05	0.26	-0.20	-0.20	0.73
Eigen value	3.14	2.61	1.70	1.23	1.19
Proportion of variance (%)	27.00	22.00	15.00	11.00	10.00
Cumulative variance (%)	27.00	49.00	64.00	75.00	85.00

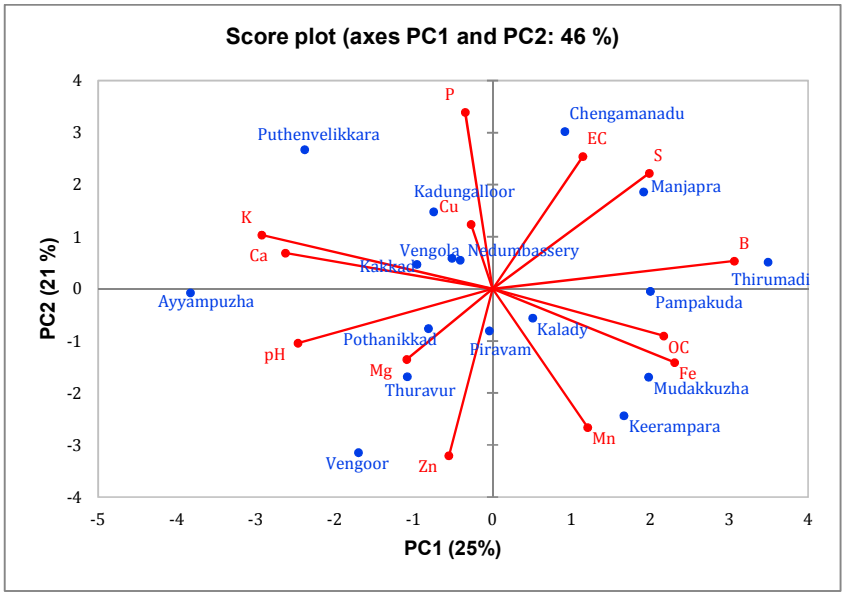

Fig. 4. Score plot of soil fertility parameters based on PCA in Ernakulam

4. CONCLUSION

The results obtained by applying PCA to 13 loactions based on 583 samples in the present study is in connection with the research work of Lomeling et al. [20] in exploring the spatial distribution of soil variables. Among the micro nutrients, Mn, Zn and Fe had relatively higher contribution as their vector length in score plot was comparatively more. EC was also found to be higher in these panchayats. Amount of OC, Fe and Mn were more in Kalady, Keerampara and Mudakkuzha of Ernakulam district whereas Thuravur, Piravom and Pothanikkad had highly acidic and Mg rich soils. Amount of Zn was more in Vengoor panchayat. Available K, Ca, P and Cu were found to be higher in Kakkad, Nedumbassery, Vengola and Kadungalloor. Both PCA and factor analysis were useful in order to draw conclusions on the spatial distribution of soil fertility parameters in the selected vegetable growing panchayat/ locations of Ernakulam district of Kerala.

ACKNOWLEDGEMENT

This research was supported by Kerala Agricultural University (KAU) and we thank our colleagues from KAU who provided insight and expertise that greatly assisted the research. We would also like to show our gratitude to all those who gave assistance and comments for completing this research.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Kerala Agricultural University. Package of Practices Recommendations: Crops (15th Ed.). Thrissur: Kerala Agricultural University; 2016.
2. Das DK. Introductory soil science (Indian reprint, 2016). New Delhi: Kalyani Publishers; 1996.
3. Bernstein L. Effects of salinity and sodicity on plant growth. Annu. Rev. Phytopathol. 1975;13:295-312.
4. Valladares GS, Camargo OA, Carvalho JRP, Silva AMC. Assessment of heavy metals in soils of avineyard region with the use of principal component analysis. Sci. Agric. (Piracicaba, Braz.). 2009;66(3):361-367.
5. SAS Institute Inc. The Analyst Application. 1st Ed. Cary, NC: SAS Institute Inc; 1999.
6. Sena MM, Frighetto RTS, Valarini PJ, Tokeshi H and Poppi RJ. Discrimination of management effects on soil parameters by using principal component analysis. Soil Till. Res. 2002;67:171-181.
7. Salehi A, Amiri GZ. Study of physical and chemical soil properties variations using principal component analysis method in

the forest, North of Iran. Caspian J. Env. Sci. 2005;3(2):131-137.
8. Dempster M, Dunlop P, Scheib A, Cooper M. Principal component analysis of the geochemistry of soil developed on till in Northern Ireland. J.Maps. 2013;9(3):373-389.
9. Wilks DS. Statistical Methods in the Atmospheric Sciences. USA: Academic Press; 2006.
10. Dahal H. Factor analysis for soil test data: A methodological approach in environment friendly soil fertility management; 2006. Accessed 12 December 2017 Available:http://www.ecostat.unical.it/tarsitano/Didattica/Anamul/Papers_ADMD_FC/Soils.pdf
11. Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate data analysis. Singapore and India: Pearson Education; 2003.
12. Reimann C, Filzmoser P, Garrett RG, Dutter R. Statistical data analysis explained (1st ed.). Chichester, UK: John Wiley & Sons Ltd; 2008.
13. Cheng Q, Jing L, Panahi A. Principal component analysis with optimum order sample correlation coefficient for image enhancement. Int. J. Remote Sens. 2006; 27(16):3387–3401.
14. Zambon I, Benedetti A, Ferrara C, Salvatic C. Soil matters? A multivariate analysis of socioeconomic constraints to urban expansion in Mediterranean Europe. Ecol. Econ. 2018;146:173–183.
15. Baggie I, Sumah F, Zwart SJ, Sawyerrd P, Bandabla T, Kamara CS. Characterization of the mangrove swamp rice soils along the Great Scarcies river in Sierra Leone using principal component analysis. Catena. 2018;163:54–62.
16. Dawes L, Goonetilleke A. Using multivariate analysis to predict the behavior of soils under effluent irrigation. Water Air Soil Pollut. 2006;172(1-4):109-127.
17. Mihai C. Leaf and twig morphological variability of romanianlinden species (tilia;tiliaceae): A case study. Proceedings of the Biennial International Symposium. Forest and sustainable development, Brașov, Romania, 24-25th October 2014. 2015. . Available:https://www.cabdirect.org/cabdirect/abstract/20153392994
18. Hammer RD, Philpot JW, Maatta JM. Applying principal component analysis to soil-landscape research-quantifying the subjective. Proceedings of Annual Conference on Applied Statistics in Agriculture; 1990. Available:http://newprairiepress.org/agstatconference/1990/proceedings/11
19. Lammel DR, Feigl BJ, Cerri CC, Nusslein K. Specific microbial gene abundances and soil parameters contribute to C, N, and greenhouse gas process rates after land use change in Southern Amazonian Soils. Front. Microbiol. 2015;6:1057.
20. Lomeling D, Otwari SM, Khater YM. Spatial characterization of cone index and some nutrients in a sandy loam soil (eutricleptosol) using the multivariate analysis. Am. J. Exp. Agric. 2015;7(2): 118-134.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/58410