International Research Journal of Pure & Applied Chemistry

22(4): 64-72, 2021; Article no.IRJPAC.69675
ISSN: 2231-3443, NLM ID: 101647669

Principal Component Analysis on Soil Fertility Parameters in Vegetable Growing Locations of Kottayam 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 done in collaboration among all the three 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/2021/v22i430404
Editor(s):
(1) Dr. Richard Sawadogo, Research Institute for Health Sciences, Burkina Faso.
Reviewers:
(1) Hasiah Salleh, Universiti Malaysia Terengganu (UMT), Malaysia.
(2) Mukesh Kumar Malviya, Guangxi Academy of Agricultural Sciences, China.
Complete Peer review History: http://www.sdiarticle4.com/review-history/69675

Original Research Article

Received 10 April 2021
Accepted 16 June 2021
Published 24 June 2021

ABSTRACT

Present study utilizes Principal Component Analysis (PCA) of 13 soil testing variables obtained from 28 vegetable growing locations of Kottayam district and there were a total of 718 samples for analysis. Thirteen Principal Components (PCs) were generated and five PCs could explain the major share of variance (80%). Score plot was drawn based on PCA and the results indicated that none of the variables was predominant in Bharananganam, Kadanadu, Kozhuvanal, Kidangoor and Pallikkathode and also these panchayats had positive scores on both F1 and F2 when factor analysis was conducted. Boron (B), Copper (Cu) and Zinc (Zn) were predominant in Akalakkunnam, Kadalpalamattom, Meeaachil, Melukavu, Poonjar and Ramapuram panchayats. Elikulam, Erumeli, Karoor, Mundakkayam, Mutholi, Poonjar south, Thalapalm and Vakathanom were those panchayats where the contribution of Magnesium (Mg), Potassium (K) and pH was more. All other elements viz, Oxidisable Organic Carbon (OC), Sulphur (S), Phosphorus (P), Calcium (Ca), Manganese (Mn) and Iron (Fe) had significant importance in Ayarkkunnam, Aymanam, Chempu, Kaduthuruthy, Kurichi, Manjoor, Maravanthuruth, Puthuppally and Thalayazham panchayats.

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

Keywords: Principal component analysis; factor analysis; principal component; oxidisable organic carbon; score plot.

1. INTRODUCTION

Kottayam, a district of God's own country is surrounded by hills on the east and Vembanadu Lake on the west and it spans over an area of 2208 sq km. (Fig. 1). It experiences the tropical climate. As Vaikom and Upper Kuttanadu lie near to sea, paddy is the dominant crop in these areas. The production and productivity of crops is directly related to the soil health which in turn results in better livelihood for the people of Kerala [1]. Plant growth is mainly influenced by the fertility status of soil prevailing in the region. Plants need primary nutrients (Nitrogen (N), Phosphorus (P) and Potassium (K)), secondary nutrients (Calcium (Ca), Magnesium (Mg) and Sulphur (S)) and micro nutrients (Boron (B), Iron (Fe), Manganese (Mn), Zinc (Zn) and Coppper (Cu)) in adequate quantities for growth and development and they are absorbed from the soil [2]. Other soil parameters viz. pH and Electrical conductivity (EC) also affect plant growth and nutrient availability [3].

Principal component analysis 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 [4]. PCA can be adopted to distinguish agricultural plots as a function of soil management and to determine the most important soil parameters to characterise them [5]. PCA was reported to be one of the best methods to identify the variations in both physical and chemical soil parameters [6]. PCA can be used to establish geochemical groupings of soils [7].

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 [8]. Factor analysis and PCA were applicable only if the sample size was adequate.

The purpose of this paper is to show how much PCA and factor analysis were apt 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.

Fig. 1. Map of Kottayam district, Kerala

2. MATERIALS AND METHODS

Soil samples were collected from different panchayats of Kottayam district as part of the project titled “Revolving fund mode project- Soil testing lab” during 2014-2015 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.

Soil samples collected from different panchayats of Kottayam district 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. Rubber is cultivated as a commercial crop in Kottayam. In addition to this vegetables are also grown in the district.

The data on thirteen soil fertility parameters of 28 panchayats of Kottayam viz., electro chemical parameters (pH and EC), Primary nutrients (OC, P and K); secondary nutrients (Ca, Mg and S) and micro nutrients (B, Fe, Mn, Zn and Cu) were available. Each panchayat is having different sample sizes and altogether sample size comes around 788.

2.1 Principal Component Analysis

Principal component analysis is a widely used procedure for data reduction and summarization of large set of variables. Compression of information contained within a large set of variables without too much loss of information is gained through PCA [9]. Principal components are the artificial variables 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. Certain amount of variability is exhibited by each PC and first two PCs generally provide most of the variation within the data [10].

PCA was selected as the multivariate technique as it is normally used to shrink the dimensionality of the data and will provide more information than single element distribution. Since the variables under study were correlated, PCA is found to be the effective method to be applied to disclose variability of the data set [11]. Based on the soil fertility status, scores were given to different locations by the method of PCA and locations with nutrient deficiency and sufficiency were recognized.

2.1.1 Definition of principal components

Let $X_1\ X_2\ ...\ X_p$ be some p random variables and X^T denotes the vector with mean μ and variance Σ.

$$X^T = [X_1\ X_2\ ...\ X_p]$$

Linear combination of the original variables is the PC which is given by,

$$PC1 = a_1X^T = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p = \sum_{i=1}^{p} a_{1i}X_i, \qquad (1)$$

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

Similarly number of such PCs 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 PCs is,

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

Proportion of variance explained by each PC is,

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

2.1.2 Score plot

Scores of each panchayat is calculated as the sum of product of principal component loadings with their corresponding means. Scores of each observation is plotted on score plot which is a two dimensional plot with PC 1 on *x* axis and PC 2 on *y* axis as these two PCs account for maximum variability.

Score obtained for a panchayat = $\sum_{i=1}^{p} PC_i \bar{X}_i$

Where PC_i is the principal component loading of each variable and $\bar{X}_i$ is the mean of each soil fertility parameter taken for analysis. Usually scores are calculated based on first two PCs as they together account for maximum variability in the dataset and are plotted on two dimensional area.

When the factor loadings are superimposed on the score plot, we get information on the relationship between variables and locations.

2.2 Factor Analysis (FA)

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.

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

Similarly p variables can be written. The coefficients l_{ij} is called as the factor loadings where i is the number of variables and j is the number of factors. 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

PCA was employed in order to identify the most relevant variables among the 13 variables under study. Five PCs were sufficient to explain 80 per cent variance in the data of Kottayam district. Similar study was conducted by Zambon et al. [12] and Baggie et al. [13]. Remaining 20 per cent was explained by eight PCs together. PC 1 accounted for 29 per cent of total variance followed by PC 2 (17%), PC 3 (15%), PC 4 (11%) and PC 5 (8%). PC 1 and PC 2 jointly explained 46 per cent of available variance (Table 1). Loading values of each variable on the PC represented the relative contribution of that particular variable on the formation of principal component.

Ca, Mg and EC were observed to be the variables which had high positive loadings on PC 1 (>0.2) while all the variables except Ca, Mg, Zn and pH showed positive loading on PC 2. Both PC 1 and PC 2 explained 46 per cent of variance. Principal component scores were calculated for each panchayat in the district and were plotted by means of score plot.

Table 1. Principal components generated through PCA in Kottayam district

Variable	PC 1	PC 2	PC 3	PC 4	PC 5
pH	0.232	-0.130	-0.249	0.459	-0.091
EC	0.322	0.273	0.051	0.292	0.314
OC	0.098	0.231	-0.065	-0.592	0.133
P	0.208	0.414	0.042	0.241	0.270
K	0.033	0.040	-0.552	0.147	0.388
Ca	0.386	-0.038	0.065	-0.102	0.363
Mg	0.378	-0.126	-0.009	0.041	-0.334
S	-0.410	0.353	-0.189	0.094	-0.048
B	-0.411	0.354	-0.179	0.107	-0.025
Fe	0.239	0.362	0.196	0.160	-0.488
Mn	0.172	0.505	0.153	-0.281	-0.046
Zn	-0.177	-0.153	0.522	0.029	0.405
Cu	-0.205	0.078	0.467	0.369	0.051
Eigen value	3.710	2.150	1.910	1.380	1.030
Proportion (%)	29	17	15	11	8
Cumulative (%)	29	46	61	72	80

3.1 PC 1

PC 1 accounted for 29 per cent of total variability in the whole data set. It was so clear that Ca, Mg which were the secondary nutrients possessed high positive loading (> 0.2) on PC1 while S and B were found to be the most negatively loaded variable i.e. nearly -0.4 (Fig 2). Fe and pH showed somewhat same loading value on PC 1 (0.239 and 0.232 respectively). Dawes and Goonetilleke [14] had also reported similar findings.

3.2 PC 2

Except for Ca, Mg, pH and Zn, all other soil testing variable had positive loading on PC 2 and it can be seen that Mn was having highest positive loading (0.5) followed by P, Fe, B and S. EC and OC also had significant loading values (> 0.2) (Fig. 3). PCA was also used as a tool to identify the important variables by Mihai [15].

3.3 Principal Component Scores

Scores were determined for each panchayat based on PC loadings. Standardized scores were calculated for each panchayat based on PC 1 and PC 2 (Table 2). Scores based on PC 1 ranged from -3.19 to 4.80 and that of PC 2 ranged from -2.83 to 2.53. Maximum score was obtained for Poonjar (4.80) and minimum score was recorded for Karoor (-0.36) based on PC 1. Based on PC 2, Thalayazham panchayat got high score (-2.83) and Puthuppally got minimum score (-0.15).

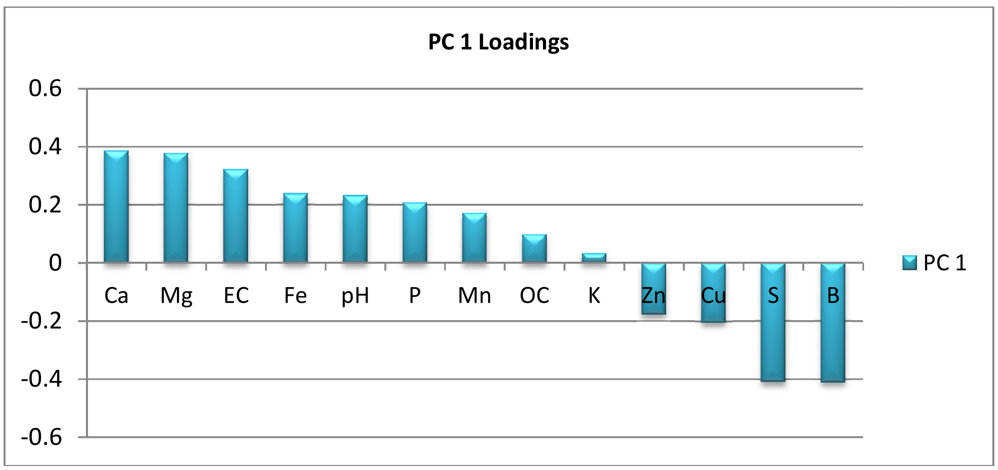

Fig. 2. Element loadings of PC 1 in Kottayam

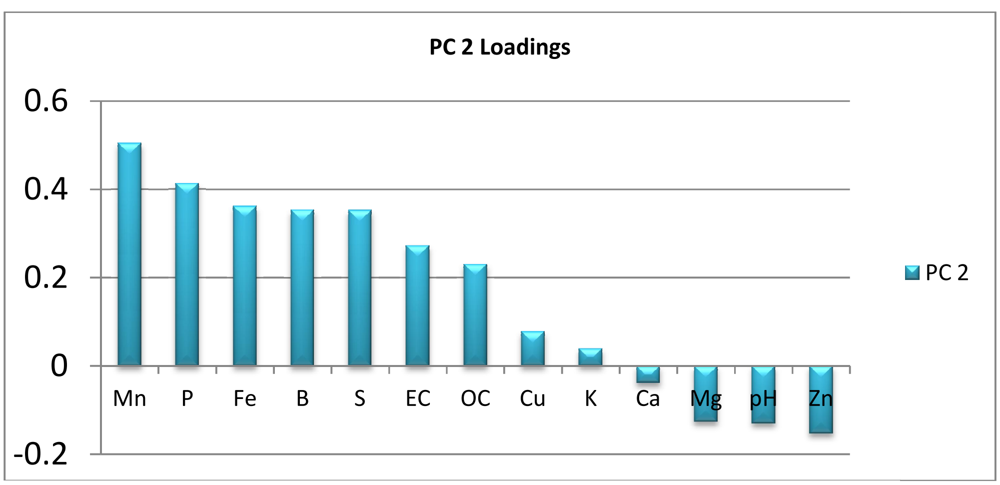

Fig. 3. Element loadings of PC 2 in Kottayam

Table 2. Standardized scores of different panchayats of Kottayam based on PC 1 and PC 2

Panchayat		PC 1	PC 2
Akalakkunnam	(K1)	2.06	-1.17
Ayarkkunnam	(K2)	-1.42	-0.38
Aymanam	(K3)	-1.96	-0.74
Bharananganam	(K4)	1.77	0.47
Chempu	(K5)	-2.65	-0.70
Elikulam	(K6)	-1.50	0.93
Erumeli	(K7)	-0.98	0.76
Kadaplamattom	(K8)	0.64	-0.17
Kadanadu	(K9)	1.24	2.10
Kaduthuruthy	(K10)	-1.08	-0.80
Karoor	(K11)	-0.36	1.31
Kidangoor	(K12)	0.58	0.91
Kozhuvanal	(K13)	0.47	2.21
Kurichi	(K14)	-0.45	-0.72
Manjoor	(K15)	-1.70	-2.07
Maravanthuruth	(K16)	-0.48	-2.68
Meenachil	(K17)	4.10	-0.74
Melukavu	(K18)	1.14	-1.58
Mundakkayam	(K19)	-0.97	1.55
Mutholi	(K20)	-1.31	0.56
Pallikkathode	(K21)	1.41	1.29
Poonjar (S)	(K22)	-0.60	1.78
Poonjar	(K23)	4.80	-1.62
Puthuppally	(K24)	-0.58	-0.15
Ramapuram	(K25)	2.83	-0.81
Thalapalam	(K26)	-0.63	2.53
Thalayazham	(K27)	-3.19	-2.83
Vakathanam	(K28)	-1.18	0.75

Table 3. Results of factor analysis in Kottayam

Variable	Factor 1	Factor 2	Factor 3
pH	-0.292	0.285	0.638
EC	-0.553	-0.419	0.278
OC	-0.364	-0.053	-0.581
P	-0.563	-0.431	0.113
K	-0.400	0.606	0.035
Ca	-0.450	-0.196	-0.298
Mg	-0.490	0.170	0.330
S	-0.349	-0.027	0.347
B	0.801	-0.288	0.202
Fe	-0.506	-0.541	0.275
Mn	-0.623	-0.516	-0.376
Zn	0.668	-0.467	-0.067
Cu	0.417	-0.564	0.341
Eigen value	3.477	1.849	1.565
Proportion(%)	40.80	21.70	18.40
Cumulative (%)	40.80	62.40	80.80

3.4 Factor Analysis on Soil Fertility Parameters

Factor analysis revealed that only three factors were sufficient in order to explain 80 per cent of total variance available in the data (Table 3). Factor 1 (F 1) had high loading on B (0.801) followed by Zn (0.668) and Mn (-0.623). First two factors could explain 62 per cent of total variance which was more than half of the total variance present. Factor 2 alone accounted for 21 per cent of variability and K had a greater positive loading (0.606) followed by Cu (-0.564) and Fe (-0.541). Results are illustrated in Table 3. Present result is in accordance with the study of Hammer et al. [16].

Score plot was drawn based on PCA and the results indicated that none of the variables was predominant in Bharananganam, Kadanadu, Kozhuvanal, Kidangoor and Pallikkathode and also these panchayats had positive scores on both F1 and F2 (Fig. 4). Lammel et al. [17] had also used score plot as part of PCA. Based on the scores of each panchayat and loadings of each variable, they were positioned in the four quadrants of the score plot. Panchayats that lie near to a particular variable (red dot) in the score plot indicate that those panchayats had predominance of that particular variable in the soil.

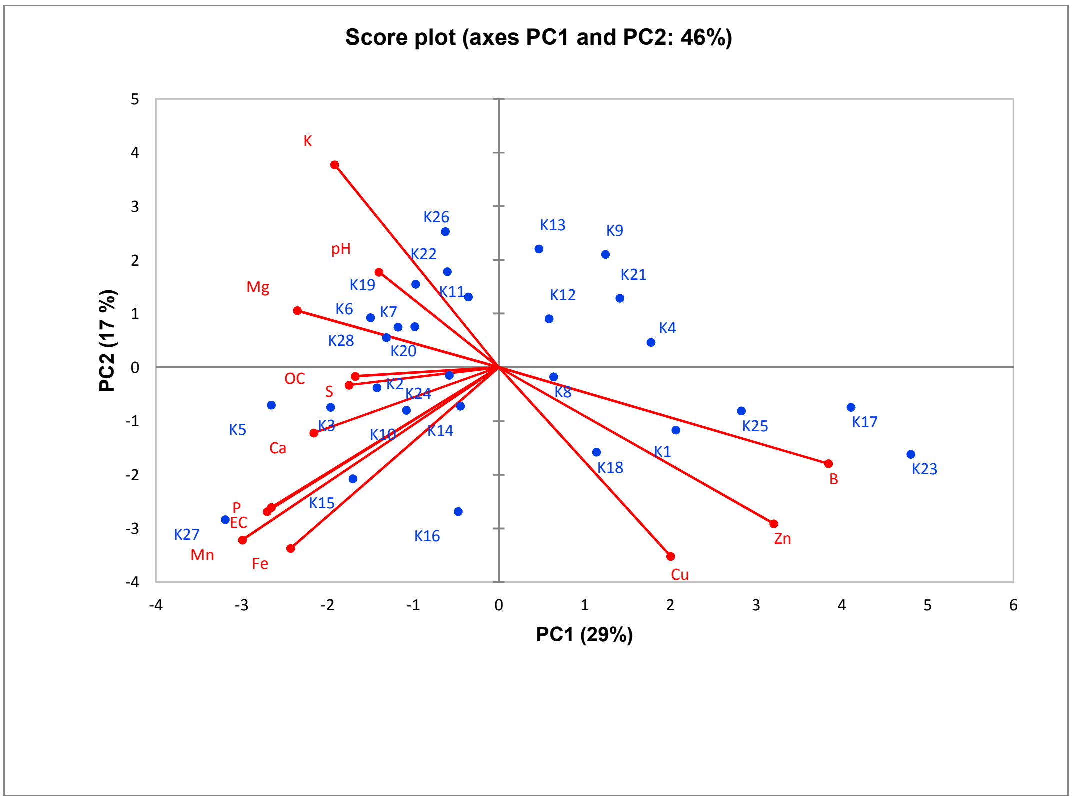

Fig. 4. Score plot based on PCA in Kottayam district

K1 – Akalakunnam	K11 – Karoor	K21 - Pallickathodu
K2 – Ayarkunnam	K12 – Kidangoor	K22 – Poonjar South
K3 – Aymanam	K13 – Kozhuvanal	K23 - Poonjar
K4 – Bharananganam	K14 - Kurichi	
K5 – Chempu	K15 - Manjoor	
K6 – Elikulam	K16 - Maravanthuruthu	
K7 – Erumeli	K17 - Meenachil	
K8 – Kadaplamattom	K18 - Melukavu	
K9 – Kadanadu	K19 - Mundakkayam	
K10 – Kaduthuruthy	K20 - Mutholi	

4. CONCLUSION

Findings obtained from the present study are in connection with the research work of Lomeling et al. [18] in exploring the spatial distribution of soil variables. Boron, Cu and Zn were predominant in Akalakkunnam, Kadalpalamattom, Meenachil, Melukavu, Poonjar and Ramapuram panchayats. Elikulam, Erumeli, Karoor, Mundakkayam, Mutholi, Poonjar south, Thalapalm and Vakathanom were those panchayats where the contribution of Mg, K and pH was more. All other elements *viz.,* OC, S, P, Ca, Mn and Fe had significant importance in Ayarkkunnam, Aymanam, Chempu, Kaduthuruthy, Kurichi, Manjoor, Maravanthuruth, Puthuppally and Thalayazham panchayats.

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 locations of Kottayam 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. Das DK. Introductory Soil Science (Indian reprint, 2016). New Delhi: Kalyani Publishers; 1996.
2. Kerala Agricultural University. Package of Practices Recommendations: Crops (15th Ed.). Thrissur: Kerala Agricultural University; 2016.
3. Bernstein L. Effects of salinity and sodicity on plant growth. Annu. Rev. Phytopathol. 1975;13:295-312.
4. SAS Institute Inc.The Analyst Application. 1st Ed. Cary, NC: SAS Institute Inc; 1999.
5. 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.
6. 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.
7. 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.
8. 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
9. Hair JF, Anderson RE, Tatham RL and Black WC. Multivariate Data Analysis. Singapore and India: Pearson Education; 2003.
10. Reimann C, Filzmoser P, Garrett RG and Dutter R. Statistical data analysis explained (1st ed.). Chichester, UK: John Wiley & Sons Ltd; 2008.
11. Cheng Q, Jing L and Panahi A. Principal component analysis with optimum order sample correlation coefficient for image enhancement. Int. J. Remote Sens. 2006; 27(16):3387–3401.
12. 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.
13. 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.
14. Dawes L. and Goonetilleke A. Using multivariate analysis to predict the behavior of soils under effluent irrigation. Water Air Soil Pollut. 2006;172(1-4):109-127.
15. 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.

16. 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.
17. 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.
18. Lomeling D, Otwari SM, and 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/69675