International Archive of Applied Sciences and Technology
Int. Arch. App. Sci. Technol; Vol 10 [2] June 2019 : 84-89
© 2019 Society of Education, India
[ISO9001: 2008 Certified Organization]
www.soeagra.com/iaast.html

IAAST
ONLINE ISSN 2277- 1565
PRINT ISSN 0976 - 4828

CODEN: IAASCA ORIGINAL ARTICLE

DOI: .10.15515/iaast.0976-4828.10.2.8489

Statistical evaluation of physico-chemical properties of Soils of Coimbatore district using Dimensionality Reduction Technique

S.Vishnu Shankar[1], M.Radha[*], R.Kumaraperumal[2]and S.R. Naffees Gowsar[3]
[1,3]Agricultural Statistics, Department of Physical Science & IT, Tamil Nadu Agricultural University,Coimbatore.
[2]Assistant Professor in Remote Sensing & GIS, Tamil Nadu Agricultural University, Coimbatore.
[*]Corresponding Author: Email: radhamyilsamy@gmail.com

ABSTRACT
Soil series is undoubtedly an important criterion to be classified for better management of crops and land use. Different parameters are used to determine a soil series. Hence there is a need to find the parameter which causes more variation. It can be achieved by Principal Component Analysis, which is a Multivariate technique used as Statistical and Machine Learning tool. It is a data reduction technique which helps to sort the most influencing variable of the group. Thereby it reduces the complexity of data and makes the analysis easier. Here the study concentrates on finding out the major variation causing variable in the soil series with help of Principal Component Analysis (PCA). The resulted components of PCA produce a total variation of about 75.66%. Thus, the Principal component analysis found to be useful in determining the high variation inducing components of soil series.
Keywords: Principal Component Analysis, Descriptive statistics, Correlation, Soil series.

Received 10.02.2019 Revised 08.03.2019 Accepted 19.04.2019

CITATION OF THIS ARTICLE
S.Vishnu Shankar, M.Radha, R.Kumaraperumal and S.R. Naffees Gowsar. Statistical evaluation of physico-chemical properties of Soils of Coimbatore district using Dimensionality Reduction Technique.Int. Arch. App. Sci. Technol; Vol 10 [2] June 2019: 84-89

INTRODUCTION

The soil is a soul of infinite living. A good physical, chemical and biological properties of the soil determines the growth of plants. Over a period of time, the pattern and type of soils at different places alters due to biotic and abiotic factors. There is also a change in the nature and distribution of soil from place to place and time to time. Hence there is a need to find out the soil series at different places and classify them. Soil classification also helps in determining the best possible use and management of soils. Under Comparable weather and climatic condition, soil groups are developed with similar soil profile characteristics which are known as soil series and a different soil series are found all over the world. Soil series is used for determining the cropping pattern [7] of the area. It also gives the information about the properties of soil such are organic content in soil, micronutrients content, the structure of soil etc. Further soil series are essential in producing Digital Soil Maps, which in turns avoids the difficulties of convention soil mapping. Classification soil series for a large area with a large number of physical, biological and chemical properties [8] is difficult. Hence the soil properties which determine the soil series are needed to be found out. For finding out the important principal soil properties, Principal Component analysis can be used.

Principal Component Analysis is dimensional reduction techniques which are highly used for facial recognition, Image Compression, and Computer Vision. It is also used for finding patterns in the field of finance, data mining, bioinformatics, psychology, etc. where there is high dimensionality of data. Principal Component Analysis is a statistical procedure that *uses* an orthogonal transformation [10] to convert a set of observations of possibly correlated variables [2] into a set of values of linearly uncorrelated variables called principal components. The principal components are orthogonal in nature. The variation found in the first principal components is higher than others i.e. variation decrease with the decrease in principal components. Based on the Eigenvalue and Eigenvectors the principal component [3] variable is determined.

In this paper, the Principal component analysis is carried out for soil series of Coimbatore district. Coimbatore district is situated in the Western Agro-climatic zone with minimum temperature prevailing of 18°C and maximum temperature of 35°C. Loamy soil, clayey soil, and Calcareous black cotton soil are the soil types found in the district.About 62 soil series were found in the Coimbatore district. Coconut is the major plantation crop cultivated in Coimbatore about an area of about 8.5831 ha. The Agricultural crops cultivated are Millets, Pulses, Oilseeds, Cotton and Sugarcane. Major horticulture crops include fruits crops like mango and banana, vegetables like tomato, brinjal, bhendi and onion, spices like turmeric and flowers like tuberose, and jasmine.Initially, descriptive statistics is carried out to extract the statistical information of the data. It is followed by the correlation matrix to find the correlation between the variables. Later, Principal Component Analysis is performed for the soil data and its corresponding principal components are found.

MATERIAL AND METHODS

Data

The study area is Coimbatore district. The soil survey database of Coimbatore district was collected from the department of Remote Sensing and GIS, Tamil Nadu Agricultural University, Coimbatore consists of soil series at association level and along with the soil physico-chemical properties. The soil properties are Soil Depth, Sand, Silt, Clay, ph, Electrical conductivity, Organic carbon, Exchangeable-Calcium, Magnesium Sodium, Potassium, Cation Exchange Capacity (CEC), Base Saturation, Maximum Infiltration Rate and Exchangeable Sodium Percentage (ESP). Each soil properties have a unique feature which differ one soil series from another.

Statistical analysis

Descriptive statistics provides the summary of the data which allows interpreting the data in an easier way [6]. It is a commonly used statistical method that breaks the complexity in the data and present in a simpler way. There are different measures of central tendency and measures of variability and dispersion. Measures of central tendency include the mean, median and mode, while measures of variability include the standard deviation (or variance), the minimum and maximum values of the variables, kurtosis, and skewness.

Table 1.Summary Statistics of physico-chemical properties of Soils

	Mean	SD	Kurtosis	Skewness	Range	Minimum	Maximum
Soil depth (cm)	39.31	17.78	8.72	2.10	113.00	14.00	127.00
Sand (%)	56.65	16.08	0.50	-0.46	72.50	11.00	83.50
Silt (%)	13.98	6.33	-0.14	0.22	27.10	0.90	28.00
Clay (%)	28.43	12.89	0.88	0.68	58.90	8.10	67.00
pH	6.79	1.30	-1.26	-0.27	4.40	4.20	8.60
EC (dS m^{-1})	0.29	0.36	5.32	2.24	1.73	0.01	1.74
OC (%)	0.66	0.45	7.35	2.16	2.70	0.12	2.82
Ca (cmol p(+) kg^{-1})	9.27	7.04	5.45	1.96	38.00	2.00	40.00
Mg (cmol p(+) kg^{-1})	3.95	3.67	3.17	1.74	17.44	0.10	17.54
Na (cmol p(+) kg^{-1})	0.85	1.46	17.81	4.03	8.42	0.00	8.42
K (cmol p(+) kg^{-1})	0.39	0.74	31.33	5.37	5.16	0.02	5.18
CEC (cmol p(+) kg^{-1})	18.22	10.92	2.16	1.41	53.60	4.50	58.10
BS (%)	75.36	18.07	-0.38	-0.02	92.83	34.15	126.98
ESP	4.48	4.76	7.27	2.38	26.31	0.00	26.31
IR (cm/h)	3.44	3.34	-0.70	0.92	10.24	0.24	10.48

Correlation

Followed by descriptive statistics, Pearson correlation is calculated. Correlation is a statistical technique that shows the degree of linear relationship between the two variables[5]. It also provides the magnitude of relationship between the variables and classified as Poor (0.0-0.3), Moderate (0.4-0.6) and Strong (0.7-1.0). By this way, the highly correlated parameters of soil series can be found.

Table 2.Correlation between of physico-chemical properties of Soils

	Soil depth	**Sand**	**Silt**	**Clay**	**pH**	**EC**	**OC**	**Ca**	**Mg**	**Na**	**K**	**CEC**	**BS**	**ESP**	**IR**
Soil depth	1														
Sand	-0.08	1													
Silt	0.00	-0.53	1												
Clay	0.04	-0.81	0.11	1											
pH	0.20	0.07	-0.23	0.16	1										
EC	0.01	0.10	0.06	-0.11	**0.42**	1									
OC	-0.27	-0.12	0.13	0.04	-0.48	-0.23	1								
Ca	0.02	-0.39	0.05	**0.45**	**0.57**	0.21	-0.12	1							
Mg	0.15	-0.32	0.08	**0.39**	**0.62**	0.31	-0.25	**0.64**	1						
Na	0.27	-0.20	0.06	0.16	0.27	0.29	-0.11	0.11	0.44	1					
K	0.19	-0.17	-0.04	0.20	0.31	0.12	-0.08	0.19	**0.45**	**0.65**	1				
CEC	0.09	-0.37	0.05	**0.48**	**0.60**	0.25	-0.11	**0.89**	**0.81**	**0.42**	**0.48**	1			
BS	0.18	-0.25	0.01	0.13	**0.56**	0.34	-0.34	**0.53**	**0.54**	0.24	0.17	**0.38**	1		
ESP	0.26	-0.16	0.07	-0.02	-0.05	0.18	-0.04	-0.14	0.15	**0.85**	**0.40**	0.06	0.19	1	
IR	-0.25	0.54	-0.03	-0.74	-0.23	0.20	-0.03	-0.35	-0.35	-0.17	-0.20	-0.41	-0.10	0.01	1

Soil Depth, Sand, Silt, Clay, pH, Electrical conductivity (EC), Organic carbon (OC), Ca-Calcium, Mg-Magnesium, Na-Sodium, K-Potassium,Cation Exchange Capacity (CEC), Base Saturation (BSP), Maximum Infiltration Rate (IR) and Exchangeable Sodium Percentage (ESP).

Principal Components Analysis

Principal Component Analysis is a frequently used method to reduce the real dimension underlying a data set from p random variables to s<<p linear function of those p random variables and their observed values [1]. The basic idea behind a PCA is to transform the p-dimensional observations Y in Rp into a set of s-dimensional functions called principal component PC with s<<p in Rp. Specifically, the principal components, PC, is given by

$PC_{\upsilon}=e_{\upsilon 1}Y_1+\ldots+e_{\upsilon p}Y_p, \quad \upsilon=1\ldots p,$

Where υ *and* $e_{\upsilon} = (e_{\upsilon 1\ldots,}e_{\upsilon p})$are the υth eigenvalue and υth eigenvector, respectively. The principal components are the Eigenvectors [4] of covariance matrix and they are orthogonal in nature. In principal component analysis, the first step is to calculate the covariance matrix of the data. Next calculate the eigen values and eigenvectors from covariance matrix

$det(\lambda I - A) = 0 \& (\lambda I - A)\upsilon = 0I$

Where λ is an eigenvalue, υ *is an* eigen vector, A is square matrix and*det* is the determinant of the matrix. The selection of principal components [9] can be done with help of scree plot. In scree plot, the X-axis represents the Principal components and Y-axis denotes the Variance in components.

Table 3.Results of eigenvalues of principal component analysis

Principal Component	Eigenvalue	Percentage of variance	Cumulative Percentage of Variance
1	4.964	31.026	31.026
2	2.448	15.298	46.324
3	2.153	13.455	59.779
4	1.371	8.566	68.345
5	1.171	7.317	75.662
6	0.789	4.931	80.592
7	0.672	4.201	84.793
8	0.632	3.948	88.741
9	0.553	3.453	92.194
10	0.381	2.382	94.576
11	0.337	2.108	96.684
12	0.268	1.676	98.359
13	0.178	1.114	99.474
14	0.049	0.307	99.780
15	0.029	0.182	99.963
16	0.006	0.037	100.000

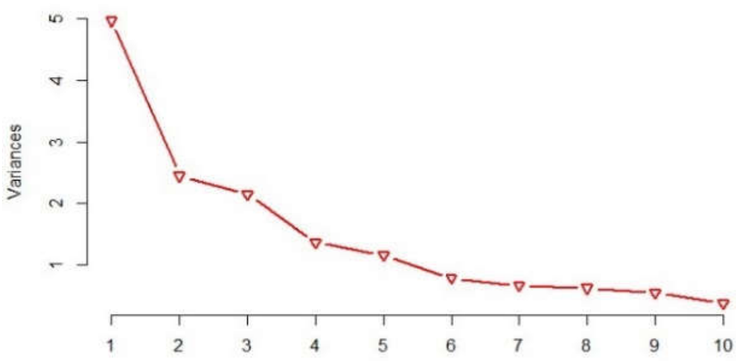

Figure 1.Scree plot of PCA

Table 4.Results of the eigenvectors for components

Properties	PC1	PC2	PC3	PC4	PC5
Soil depth	0.131	-0.11	-0.182	-0.249	-0.61
Sand	-0.234	-0.45	0.111	-0.251	0.194
Silt	0.041	**0.246**	-0.122	**0.612**	-0.193
Clay	0.266	**0.429**	0.01	-0.114	-0.056
pH	**0.305**	-0.268	**0.28**	-0.157	-0.014
EC	0.145	-0.338	0.083	0.373	0.095
OC	-0.122	0.31	-0.171	0.087	-0.494
Ca	**0.336**	0.091	**0.323**	0.064	0.156
Mg	**0.383**	-0.059	0.104	0.053	0.097
Na	0.272	-0.192	-0.455	0.06	0.099
K	0.258	-0.117	-0.274	-0.099	**0.232**
CEC	**0.391**	0.041	0.138	-0.003	0.275
Base Saturation	0.269	-0.177	0.182	0.23	-0.243
ESP	0.138	-0.191	-0.545	0.15	-0.051
Infiltration rate	-0.243	-0.356	0.026	0.337	0.164

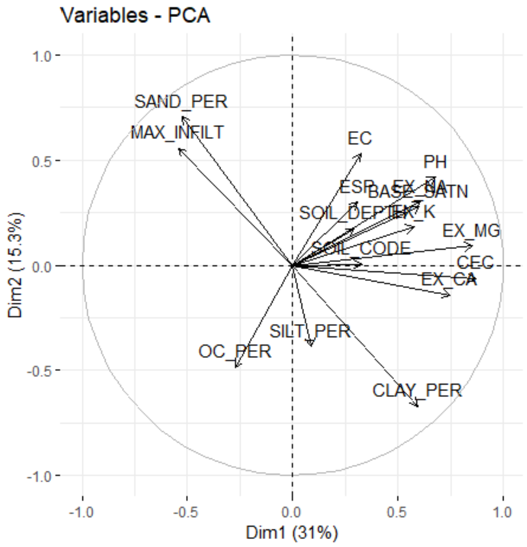

Figure 2. Correlation plot for variables

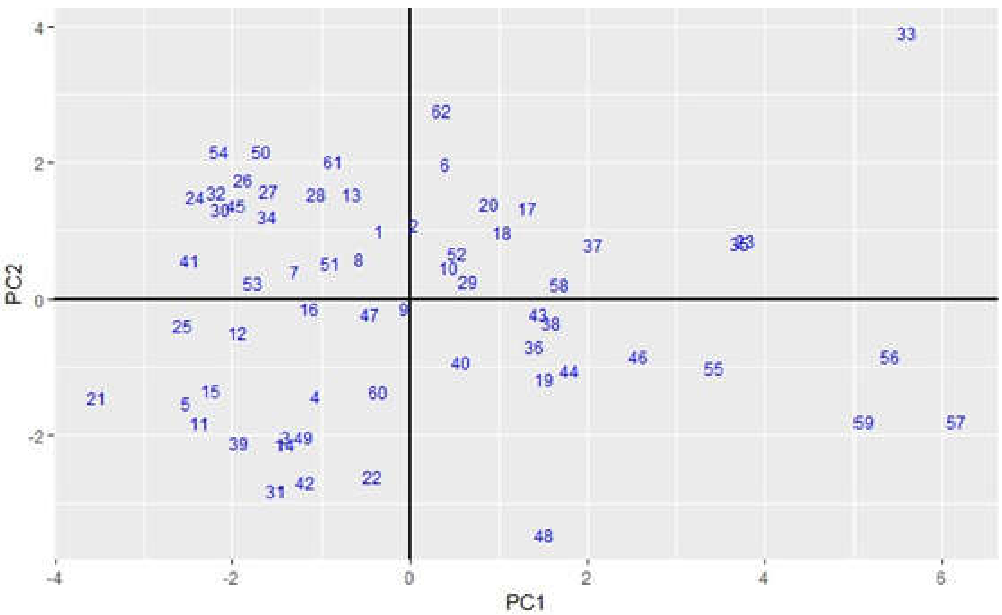

Figure 3. Dispersion of observations in Principal Components

RESULTS AND DISCUSSION

The Mean, Standard deviation (SD), Kurtosis, Skewness and Range of soil properties of whole Coimbatore district are given in Table 1. The Mean and Range value gives the status of physico-chemical properties of Soils in Coimbatore district. This helps in efficient application of fertilizer to the soil and also helps in water management. Based on the pH values, it is easy to separate the acidic, neutral, alkaline soils which can be used for better cropping pattern.

From Table 2, it is found that CEC is highly correlated with Calcium (0.9) and Magnesium (0.8), Sodium is highly correlated with Potassium (0.7) and ESP (0.8). Whereas moderately correlated variables are Clay with Calcium, Magnesium, CEC; pH with Calcium, Magnesium, CEC, BS; Calcium with Magnesium, BS; Magnesium with BS; Sodium with CEC; Potassium with ESP and CEC with BS.

It is found that PC1, PC2, PC3, PC4 and PC5 are having eigen value greater than 1 and those components are considered for further analysis (Table 3). These five components produce a total variation of about 75.66% *i.e.* PC1=31.02%, PC2=2.45%, PC3=2.15%, PC4=1.37%, PC5=1.17%. Scree pot also confirms the selection of principal components by knee point (Figure 2). The dispersion of observations between PC1 and PC2 shows the relationship within the soil series (Figure 3). It is inferred from the Table 4 that PC1 has positive values for exchangeable Magnesium, Calcium, CEC, pH and slight negative values for Sand and Infiltration rate. PC2 have positive values for Silt, Clay and negative values for Electrical conductivity. Similarly, PC3 is contributed mainly by pH and Calcium, PC4 by Silt and PC5 by Potassium.

CONCLUSION

It is concluded that PC1, PC2, PC3, PC4, and PC5 have contributed a total variation of 75.66 per cent. These high variations are contributed by variables such as Magnesium, calcium, pH, Silt, Clay, and Potassium and they considered as potential indicators in determining the soil series. Thus, Principal Component Analysis helps in bringing out the principal components of soil series with the help of eigen values and eigen vectors.

REFERENCES

1. Bhuyan, K. C. (2005). *Multivariate Analysis & Its Applications*. New Central Book Agency.
2. Chinnadurai, C., Gopalaswamy, G., & Balachandar, D. (2014). Impact of long-term organic and inorganic nutrient management on the biological properties and eubacterial community diversity of the Indian semi-arid Alfisol. *Archives of Agronomy and Soil Science*, *60*(4), 531-548.
3. Fox, G. A., & Metla, R. (2005). Soil property analysis using principal components analysis, soil line, and regression models. *Soil Science Society of America Journal*, *69*(6), 1782-1788.
4. Ghaemi, M., Astaraei, A. R., Emami, H., Nassiri Mahalati, M., & Sanaeinejad, S. H. (2014). Determining soil indicators for soil sustainability assessment using principal component analysis of Astan Quds-east of Mashhad-Iran. *Journal of soil science and plant nutrition*, *14*(4), 1005-1020.
5. Gogtay, N. J., & Thatte, U. M. (2017). Principles of Correlation Analysis. The Journal of the Association of Physicians of India, Vol.65. 78-81.
6. Kaushik, M., & Mathur, B. (2014). Data Analysis of Students Marks with Descriptive Statistics. *International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), ISSN*, 2321-8169.
7. Panishkan, K., Swangjang, K., Sanmanee, N., & Sungthong, D. (2012). Principal component analysis for the characterization in the application of some soil properties. *World Acad Sci Eng Technol*, *6*, 727-729.
8. Parra-González, S. D., & Rodriguez-Valenzuela, J. Determination of the Soil Quality Index by Principal Component Analysis in Cocoa Agroforestry System in the Orinoco Region, Colombia.
9. Paul, L. C., Suman, A. A., & Sultan, N. (2013). Methodological analysis of principal component analysis (PCA) method. *International Journal of Computational Engineering & Management*, *16*(2), 32-38.
10. Skrbic, B., & Durisic-Mladenovic, N. (2007). Principal component analysis for soil contamination with organochlorine compounds. *Chemosphere*, *68*(11), 2144-2152.