Journal of the Indian Society of Soil Science, Vol. 63, No. 4, pp 379-383 (2015)

Multivariate Approaches for Soil Fertility Characterization of Lower Brahmaputra Valley, Assam, India

S.K. Reza*, Utpal Baruah[1] and S.K. Singh[2]

ICAR-National Bureau of Soil Survey and Land Use Planning, Sector-II, DK-Block, Salt Lake, Kolkata, West Bengal

The aim of this study was to characterize the soil fertiltiy like pH, organic carbon (OC), available nitrogen (N), available potassium (P) and available phosphorus (K), and DTPA extractable iron (Fe), manganese (Mn), zinc (Zn) and copper (Cu) of lower Brahmaputra valley of Assam using multivariate statistics (principal component, correlation matrix and cluster analysis). A total of 2753 soil samples from a depth of 0-25 cm at an approximate interval of 1 km were collected from Barpeta, Bongaigaon and Nalbari districts of Assam. Soil properties showed large variability with greatest variation was observed in DTPA-Zn (120%), whereas the smallest variation was in pH (17.5%). The principal component analysis (PCA) applied on the investigated soil properties identified three components with eigen values greater than 1, which explained 65% variability and same grouping was also obtained from cluster analysis. Cluster 1 includes Zn, Cu, pH, Fe, P and Mn, which has highest loading in PC1 and also showed strong significant relationship. Cluster 2, which contains OC and N, and had highest loading in PC2 and also showed the significant positive relationship with each other. Cluster 3 contain only K, which is equally distributed both in PC1 and PC2 and also significantly and positively correlated with pH, Fe, Zn and Cu of cluster 1 and OC and N of cluster 2.

Key words: Lower Brahmaputra valley, soil properties, principal component analysis, correlation matrix, cluster analysis

The use of multivariate methods in the field of soil science is emerged recently for finding a relationship between the soils classes and the nutritional state of some crops (Wadt 2005); deducing the categorical or continuous properties using the cluster technique (Simbahan and Dobermann 2006); determining the minimum number of soil properties that represent a class or category (Bockheim 2008). Soil pH, organic matter and electric conductivity (EC) were the important soil parameters out of 10 variables for sustained agricultural production in different soil classes as found by Vitharana *et al.* (2008). The other use of multivariate statistics in this field is to use it as a mechanism for the determination of soil quality index Mandal *et al.* (2008). Du *et al.* (2008) used the principal component analysis (PCA) and neural networks in order to classify the soil of a particular zone, whereas Deka *et al.* (2012) evaluated the fertility potential class of the rice growing soils of the northern Brahmaputra valley of Assam using PCA. Results clearly showed that these techniques can be used for the rapid soil identification and they also can be developed for a future use for the evaluation of the important soil properties affecting crop productivity by lowering the variables to be quantified as well as reducing the time and cost of the soil analysis where a lot of variables are determined.

The PCA also known as factor analysis, is a statistical device to reduce the number of variables to a smaller number of indices. The transformation of the raw soil properties data using principal factor analysis can result in new values that are often more interpretable than the original data (Norris 1972). Principal factors are, in fact the eigen vectors or characteristics / proper vectors of the covariance (or correlation) matrix and the variances of the principal factors are the corresponding eigen values. Principal

*Corresponding author (Email: reza_ssac@yahoo.co.in)
Present address
[1]ICAR-National Bureau of Soil Survey and Land Use Planning, Jamuguri Road, Jorhat, Assam
[2]ICAR-National Bureau of Soil Survey and Land Use Planning, Nagpur, Maharashtra

factor analysis compares the information content of a number of soil properties into few groups. Such reduction in dimensionality is an important economic consideration, especially if the potential information recoverable from the transformed data is just as good as the original data (Anderson and Furley 1975).

Cluster analysis also proved useful tool in grouping diverse soil properties into distinct classes. Katyal *et al.* (1985) used non-hierarchical Euclidian cluster analysis to subdivide a group of 57 Indian benchmark soils into different micronutrient availability classes. Li and Mahler (1992) and Li *et al.* (1992) also used cluster analysis to establish relationships between soil mapping units and geographic areas (drainage districts); to predict tendencies for micronutrient deficiency and sufficiency; and to evaluate the effects of pH and organic carbon on DTPA-extractable copper, manganese and zinc.

Based on the rainfall pattern, terrain and soil characteristics, Assam has been delineated into six agro-climatic zones *viz*., north bank plain zone, upper Brahmaputra valley zone, central Brahmaputra valley zone, lower Brahmaputra valley zone, Barak valley zone and hill zone. Quantitative evaluation of soil properties to obtain from large number of variables using PCA has immense utility. A few reports have been obtained to use the multivariate statistics for evaluation of soil fertility in Brahmaputra Valley, Assam. An attempt was made (i) to group the soil properties affecting crop productivity using multivariate analysis and (ii) to study the interrelationship among soil properties in lower Brahmaputra valley zone.

Materials and Methods

Site Characteristics

The Brahmaputra valley of Assam is a part of vast Indo-Gangetic plains and covers an area of 56,578 km^2. Due to the differences in sediment deposited by the tributaries, wide variability in soil properties has been noticed across the Brahmaputra valley. The studied area is the lower Brahmaputra valley zone of Assam covering an area of 20,222 km^2. The topography of the lower Brahmaputra valley zone is generally characterized by an almost flat plain except for few low forested hills that break the monotony of the terrain. The climate of is subtropical with warm humid summer and cool dry winter. The average maximum temperature ranges from 23.6 to 31.7 °C and minimum temperature varies from 10.0 to 24.2 °C. Annual rainfall is 1000-3100 mm and about 80% of rainfall is from south west monsoon.

Soil Sampling, Processing and Analysis

In autumn of 2010, a total of 2753 soil samples were collected from the plow layer (0-15 cm) on grid basis over the entire lower Brahmaputra valley covering Borpeta, Chirang, Baksa, Bongaigaon, Goalpara, Kokrajhar, Nalbari districts of Assam. Soil samples were air-dried and ground to pass through a 2-mm sieve. Soil pH in 1:2 soil:water suspension was determined using pH meter. Available potassium (K) was extracted with 1 *N* NH_4OAc and then measured by flame photometer. Available phosphorus (P) was extracted by Bray-1 (Bray and Kurtz 1945) and then determined by a spectrophotometer. Available nitrogen (N) was determined the method given by Subbiah and Asija (1956). Organic carbon (OC) and DTPA extractable micronutrients were determined by Walkley and Black (1934) and Lindsay and Norvell (1978), respectively.

Statistical Analysis

The soil properties data were analyzed statistically using SPSS software (version 16.0 for Windows). The statistical parameters, including mean, median, standard deviation (SD), coefficient of variance (CV), and extreme maximum and minimum values, which are generally accepted as indicators of the central tendency and of the data spread, were analyzed. Soil properties were grouped using PCA that is a method or technique which provides a means of identifying and measuring the relationship or basic patterns in a data set. This technique greatly facilitates in reducing a large number of interdependent variables to a smaller set of more meaningful and nearly uncorrelated new variables or components known as principal components. Cluster analysis (CA) was applied to identify different groups, clustering the samples with similar properties. The CA was formulated according to the Ward-algorithmic method, and the squared Euclidean distance was employed for measuring the distance between clusters of similar soil properties. Pearson's correlation matrix was used to identify the relationship among soil properties and support the results obtained by multivariate analysis.

Results and Discussion

Descriptive Statistic of Soil Properties

The summary of the statistics for soil properties are shown in table 1. The median of each soil

Table 1. Descriptive statistics of soil properties of lower Brahmaputra valley, Assam (n=2753)

Soil properties	Minimum	Maximum	Mean	Median	SD*	CV (%)**
pH	4.0	8.4	5.7	5.4	1.0	17.5
Organic carbon (g kg^{-1})	0.2	45.0	8.8	8.1	4.6	52.3
Available N (mg kg^{-1})	12.1	448.6	163.0	149.9	65.2	40.0
Available P (mg kg^{-1})	0.7	59.1	9.6	7.7	6.1	63.5
Available K (mg kg^{-1})	3.0	299.0	50.9	45.0	27.5	54.0
DTPA-Fe (mg kg^{-1})	1.0	258.0	59.6	53.2	40.1	67.3
DTPA-Mn (mg kg^{-1})	0.3	161.6	21.3	15.8	17.8	83.6
DTPA-Zn (mg kg^{-1})	0.1	9.5	0.5	0.4	0.6	120.0
DTPA-Cu (mg kg^{-1})	0.1	14.6	3.0	2.8	1.6	53.3

*Standard deviation; **Coefficient of variation

properties was lower than the mean, which indicates that the effects of abnormal data on sampling value were not great. There was difference in the CV of the soil properties. The greatest variation was observed in DTPA-Zn (120%), whereas the smallest variation was in pH (17.5%). Similar observation was also documented with a smaller variation of soil pH compared to other soil properties (Reza *et al.* 2012). This may be attributed to the fact that pH values are log scale of proton concentration in soil solution, there would be much greater variability if soil acidity is expressed in terms of proton concentration directly. The OC, available P, available K, and DTPA-extractable Fe, Mn, Zn and Cu exhibit a high variation (>50%), while available N having medium variation (25-50%) and pH low variation (<25%) according to guidelines provided by Warrick (1998).

Multivariate Analysis of Soil Properties

The fertility potential of the studied soils was carried out using principal factor analysis. Varimax rotation (Gotelli and Ellison 2004) was used to maximize the sum of the variance of the factor coefficients. This technique clusters variables into groups, such that variables belonging to one group are highly correlated with one another. Since the variability was distributed among the 9 studied soil characters, only three number of eigen values were extracted (Table 2). The sum of these seven eigen values explained 65% of the total variance. If the number of characters being considered were less to explain 100% variability, the number of eigen values would have been less, as the total variability would have got distributed within few characters only (Anderson and Furley 1975). The factor loadings for the first three components which are the correlation of individual characters with the respective components are also given in table 2.

Table 2. Rotated component matrix of the three principal component accounting for most of the total variance

Soil properties	PC1	PC2	PC3
pH	-0.311	-0.104	**0.660**
Organic C	0.031	**0.803**	0.002
Available N	0.018	**0.715**	-0.250
Available P	-0.010	-0.156	**0.684**
Available K	0.196	0.452	**0.502**
DTPA-Fe	**0.786**	0.046	-0.190
DTPA-Mn	**0.591**	-0.150	-0.387
DTPA-Zn	**0.614**	0.085	0.090
DTPA-Cu	**0.551**	0.174	0.371
Eigen values	2.205	1.405	1.335
Percentage of variance	29.50	20.62	14.84
Cumulative per cent	29.50	50.12	64.96

In details, principal component 1 (PC1), which had highest positive loadings of Fe, Mn, Zn, Cu and moderate negative loadings of pH and accounts for 29% of variance (Table 2), thus, the first factor was strongly influenced by inherited mechanical components. So, the first factor was easily interpretable and could be designated as ‘inherent fertility factor’. The PC1 could be better explained by correlation study (Table 3) as there are highly significant negative relationships of soil pH with Fe (r = -0.379**), Mn (r = -0.249**) and Zn (r = -0.134**). The PCA and correlation study showed that the soil pH and DTPA-extractable Fe, Mn and Zn had strong negative interrelationship with each other. A similar correlation was also reported by Deka *et al.* (2012) for the northern Brahmaputra valley soils of Assam. The PC2 had high positive loadings of OC and available N, and moderate positive loading of available K, accounts for 21% of variance. This factor could be referred as ‘nitrozinc factor’ (Deka *et al.* 2012). Correlation study also showed that OC highly significantly positively correlated with available N (r = 0.498**) and K (r = 0.119**). A positive correlation

Table 3. Correlation coefficients among soil properties and their level of significance

Parameter	pH	OC	Avail N	Avail P	Avail K	DTPA-Fe	DTPA-Mn	DTPA-Zn
OC	-0.267**	1.000						
Avail N	-0.341**	0.354**	1.000					
Avail P	0.375**	-0.074**	-0.131**	1.000				
Avail K	0.046*	0.203**	0.099**	0.096**	1.000			
DTPA-Fe	-0.379**	0.091**	0.080**	-0.080**	0.050**	1.000		
DTPA-Mn	-0.249**	-0.030	0.077**	-0.146**	0.001	0.358**	1.000	
DTPA-Zn	-0.134**	0.097**	0.091**	0.026	0.096**	0.268**	0.205**	1.000
DTPA-Cu	-0.007	0.111**	0.038*	0.022	0.184**	0.308**	0.029	0.170**

**Correlation is significant at P=0.01; *Correlation is significant at P=0.05

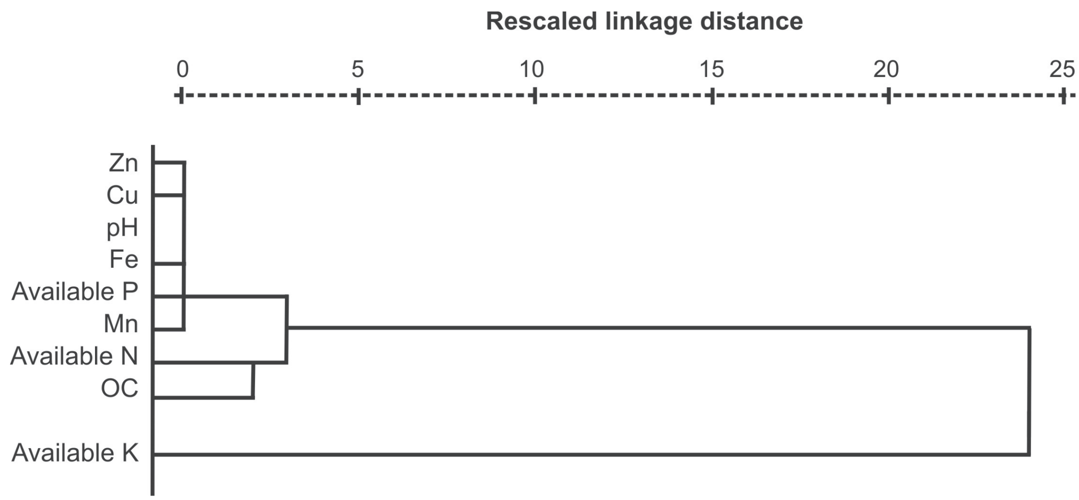

Fig. 1. Dendrogram obtained by hierarchical clustering analysis for soil properties

was also found between available N and K (r = 0.063*). The PC3 had high positive loadings of pH, available P and K and accounts for 15% of variance. Correlation study also indicated that soil pH highly significantly correlated with available P (r = 0.375**). Similar correlation was also observed by Reza *et al.* (2011) in Jorhat soils of Assam. So, this factor could be termed as 'acidic factor' (Deka *et al.* 2012).

Based on information assessed from hierarchical cluster analysis was performed (Gotelli and Ellison 2004). Three main clusters can be distinguished in the dendrogram obtained from the cluster analysis performed on the analyzed parameters with Ward's method and the squared Euclidean distance as a similarity measure (Fig. 1). Cluster 1 includes Zn, Cu, pH, Fe, available P and Mn which in the previous section was identified strong significant relationship. Cluster 2, which contains OC and available N, and had significant positive relationship. Cluster 3 contain only available K, which is equally distributed both in PC1 and PC2 and also significantly positively correlated with pH, Fe, Zn and Cu of cluster 1 and OC and available N of cluster 2.

Conclusions

Different useful tools and methods have been employed to group the soil properties of lower Brahmaputra valley. The summary statistics for soil properties showed that there was difference in the CV of the soil properties. Multivariate analysis (principal component and cluster analysis) and correlation matrix used in this study provide important tools for better understanding the soil fertility potential. The PCA applied on the investigated soil properties identified three components and same grouping was also obtained from cluster analysis. A significant correlation is observed among Zn, Cu, pH, Fe, available P and Mn, and between OC and available N.

Acknowledgments

This research was funded by the Directorate of Agriculture, Government of Assam, under the project "Assessment and mapping of some important soil parameters for the thirteen priority districts of Assam state towards optimum land use planning".

References

Anderson, K.E. and Furley, P.A. (1975) An assessment of the relationship between the surface properties of chalk soils and slope from using principal component analysis. *Journal of Soil Science* **26**, 130-143.

Bockheim, J.G. (2008) Functional diversity of soils along environmental gradients in the Ross Sea region, Antarctica. *Geoderma* **144**, 32-42.

Bray, H.R. and Kurtz, L.T. (1945) Determination of total organic and available forms of phosphorus in soil. *Soil Science* **59**, 39-45.

Deka, B., Baruah, T.C., Dutta, M. and Patgiri, D.K. (2012) Fertility potential classification of soils in different landscape units of the northern Brahmaputra valley zone of Assam. *Journal of the Indian Society of Soil Science* **60**, 92-100.

Du, C., Linker, R. and Shaviv, A. (2008) Identification of agricultural Mediterranean soils using mid-infrared photoacoustic spectroscopy. *Geoderma* **143**, 85-90.

Gotelli, N.J. and Ellison, A.M. (2004) *A Primer of Ecological Statistics*. Sinauer Associates, Sunderland, MA, USA,

Katyal, J.C., Doshi, S.P. and Malhorta, P.K. (1985) Use of cluster analysis for classification of benchmark soil samples from India in different micronutrient availability groups. *The Journal of Agricultural Science (Cambridge)* **104**, 421-424.

Li, G.R. and Mahler, R.J. (1992) Micronutrients in the Kootenai River Valley of northern Idaho. I. Effect of soil chemical properties on micronutrient availability. *Communications in Soil Science and Plant Analysis* **23**, 1161-1178.

Li, G.R., Mahler, R.J. and Everson, D.O. (1992) Micronutrients in the Kootenai River Valley of northern Idaho. II. Use of cluster and discriminant analysis to evaluate soil micronutrient status. *Communications in Soil Science and Plant Analysis* **23**, 1179-1194.

Lindsay, W.L. and Norvell, W.A. (1978) Development of a DTPA soil test for zinc, iron, manganese and copper. *Soil Science Society of America Journal* **42**, 421-428.

Mandal, K.U., Warrington, N.D., Bhardwaj, K.A., Bar-Tal, A., Kautsky, L., Minz, D. and Levy, J.G. (2008) Evaluating impact of irrigation water quality on a calcareous clay soil using principal component analysis. *Geoderma* **144**, 189-197.

Norris, J.N. (1972) The application of multivariate analysis to soil studies. II. Soil variation. *Journal of Soil Science* **23**, 62-75.

Reza, S.K., Baruah, U. and Sarkar, D. (2012) Spatial variability of soil properties in Brahmaputra plains of north-eastern India: A geostatistical approach. *Journal of the Indian Society of Soil Science* **60**, 108-115.

Reza, S.K., Baruah, U., Sarkar, D. and Dutta, D.P. (2011) Influence of slope positions on soil fertility index, soil evaluation factor and microbial indices in acid soil of humid subtropical India. *Indian Journal of Soil Conservation* **39**, 44-49.

Simbahan, C.G. and Dobermann, A. (2006) An algorithm for spatially constrained classification of categorical and continuous soil properties. *Geoderma* **136**, 504-523.

Subbiah, B.W. and Asija, G.L. (1956) A rapid procedure for estimation of available nitrogen in soils. *Current Science* **25**, 259-260.

Vitharana, A.W.U., Meirvenne, M.V., Simpson, D., Cock, L. and De Baerdemaeker, J. (2008) Key soil and topographic properties to delineate potential management classes for precision agriculture in the European loess area. *Geoderma* **143**, 206-215.

Wadt, S.P.G. (2005) Relationships between soil class and nutritional status of coffee plantations. *Revista Brasileira de Ciência do Solo* **29**, 227-234.

Walkley, A. and Black, I.A. (1934) An examination of the Digtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science* **37**, 29-38.

Warrick A.W. (1998) Spatial variability. In *Environmental Soil Physics* (D. Hillel, Ed.), pp. 655-675. Academic Press, New York.

Received 3 May 2014; Accepted 30 December 2015