JCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-4 E-ISSN: 2347-2693

Complex analysis of classified of Soil parameters and its relationship identification using PCA

M.V.Mawale[1*], V.N.Chavan[2],

[1*]Dept of Computer Science. Adarsha Science, J.B.Arts & Birla Commerce Mahavidyalaya, Amravati, India
[2]Dept. Of Computer Science &IT, Seth Kesarimal Porwal College, Kamptee, Nagpur, India

**Corresponding Author: naanka@rediffmail.com*

Available online at: www.ijcseonline.org

Received: 16/Mar/2018, Revised: 24/Mar/2018, Accepted: 12/Apr/2018, Published: 30/Apr/2018

Abstract— This study was carried out to predict meaningful information from large data set of soil parameters and representation in graphical manner to make its clear understanding This analysis help in determining role of dependent variable and independent variable in the system and their relationships, their dependability for designing any prediction system.
A field study is carried out to collect information for assessing soil parameter. Soil parameters analysis is done on 902 soil samples collected from KrushiVighan Kendra, Ghatkhed, Amravati. The values of C, N, P, K, Mg, C, Fe, Cu, Zn, B, Mo, Lime, Saline, CEC, Mn, OM and pH of soil sample collected for the year 2011-2012 and 2012-2013 andPrinciple Component Analysis (PCA) is used to predict these soil parameters as a dependent and independent parameter that have direct/indirect effects on productivity.

Keywords— complex analysis.soilparameter,Principle Component Analysis,Cu_copper, Fe_iron; ; K_potassium; Mn_manganese; OC_organic content: P_ phosphorus; Zn_zinc.

I. INTRODUCTION

Real world problem needs lots of data and information to handle situation so its size and scale are larger and testing is done on such a large scale is quite difficult.Complex analysis is used to predicate meaningful information from large data and representation of that data in graphical manner. Plant growth is the result of a complex process. Many biological, chemical and physical factors are responsible for determining soil quality [1]. As soil having different texture for different area , the soil properties of soil are in variation and its operation at a range of scales, show intermittent effects, and fluctuating more insome regions than in others.and prediction of such soil properties become complex[2].With the help of complex analysis larger data and information is converted into meaningful information. More number of quality variables are good sign for better accuracy in data analysis as each variable shows its significance individually and relationship between each other having same properties.Complex analysis using Principle Component Analysis (PCA) is used to predict meaningful information from large data set of soil parameters and representation in graphical manner to make its clear understanding[3]. This analysis help in determining role of dependent variable and independent variable in the system and their relationships, their dependability for designing any prediction system[4].

COMPLEX ANALYSIS

During soil collection, the process of soil testing in laboratory is well studied. Testing of each macronutrient, micronutrient and other factor is carried out. As result of laboratory analysis, values of 902 soil sample are taken. To assess the variables relevant for the decision making the framework of complex analysis is used to predicate meaningful information from large data and representation of that data in graphical manner. Soil parameters are identified such that direct variable considered for fuzzy model and indirect variables are used for making platform for direct or relevant variables. Choosing of relevant variables is most crucial process in determining precision decision. Ec, N, P, K, Mg, C, Fe, Cu, Zn, pH are dependent variables which are applied directly in fuzzy model and Saline, Organic matter, CEC and Lime are independent variables[5].
The PCA technique of XLSTAT is used to analyses relationship between various variables by Correlation matrix (Pearson (n)), Eigenvalues, factor loading contribution and square cosines of variable and observations. It also plot different chart like scree plot, Bi-plot etc. that helps in analyzing the data and draw some valid inference.

Use of PCA in the research

i. Analysis of independent variable of a data table representing observations described by several dependent variables and their inter-correlation.
ii. Its goal is to extract the important information from the data table and to express this information as a set of new orthogonal variables called principal components.
iii. PCA also represents the pattern of similarity of the observations and the variables by displaying them as points in maps.
iv. Principal components are obtained as linear combinations of the original variables. The first principal component is required to have the largest possible variance. The second component is computed under the constraint of being orthogonal to the first component and to have the largest possible inertia.
v. The other components are computed likewise. The values of these new variables are factor scores and these factors scores are interpreted geometrically as the projections of the observations onto the principal components.
vi. Figure No. 1 Interlinking between nutrients and analyzing their dependability

- dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per
- (Source: Compiled by Researcher)
- Soil parameters are classified into number of set. Each set consist of number of parameter according their dependability, correlation properties. Maximum number of set is prepared with the help of combination of variables.

Table No. 1 Classification of variables

Dependent Variables	Independent Variables
pH, EC, N, P, K, Mg, Zn, Cu, Fe, Mn, B, Mo, C	Lime, OM, Saline, CEC
No. of variables=13	No. of variables=4

(Source: Compiled by Researcher)

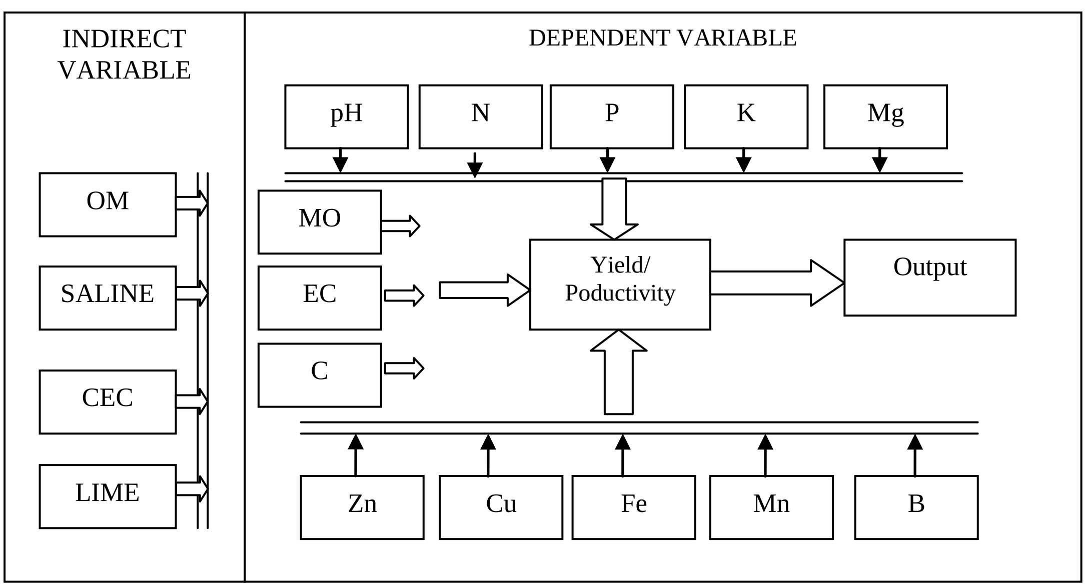


Table No. 2 Preferred combinations used for designing the system

Indirect Variables			Direct Variables			Mix variables		
Cii	Variables	No. of Rule	Cdi	Variables	No. of Rule	Cmi	Variables	No. of Rule
Ci1	{OM}	3	Cd1	{p}	06	Cm1	{Zn,OM}	2
Ci2	{OM,CEC}	1	Cd2	{Mn}	01	Cm2	{pH,CEC}	1
Ci3	{Lime}	1	Cd3	{K}	06	Cm3	{Mg,CEC}	1
			Cd4	{EC}	06	Cm4	{Lime,B}	1
			Cd4	{Zn}	01	Cm5	{K,Lime}	1
			Cd5	{N}	02	Cm6	{OM,B}	1
			Cd5	{Cu}	01	Cm7	{pH,OM}	4
			Cd6	{B}	05	Cm8	{C,Lime}	1
			Cd7	{C}	01	Cm9	{P,OM}	1
			Cd7	{pH}	04	Cm10	{Cu,Lime}	1
			Cd8	{MO}	01	Cm11	{pH,OM,CEC}	1
			Cd9	{K,Mn}	01	Cm12	{K,Cu,Lime}	1
			Cd10	{K,Zn}	01	Cm13	{pH,N,P,K,OM,CEC}	1
			Cd11	{MO,B}	01	Cm14	{Cu,Zn, Mg, Saline, MO}	1
			Cd12	{pH,K}	33	Cm15	{pH,K,Mg,CEC}	1
			Cd13	{pH,N}	33	Cm16	{pH,saline,OM,Lime}	17
			Cd14	{pH,P}	33	Cm17	{Cu, Zn, Mg,Saline, MO}	1
			Cd15	{EC,N}	06	Cm18	{pH, N, K,OM, CEC}	1
			Cd16	{Cu,MO}	01	Cm19		
			Cd17	{pH,B}	01			
			Cd18	{N,B}	01			
			Cd19	{pH,Cu}	03			
			Cd20	{N,P,Zn}	01			
			Cd21	{P,Cu,MO}	01			
			Cd22	{pH,Mn,MO}	01			
			Cd23	{pH,K,Fe,Mn,Zn}	02			
			Cd24	{Cu,Fe,Mn}	01			
			Cd25	{Cu,Zn,MO}	01			
			Cd27	{pH,Cu,P,Cu,Fe,Mn,Mo,Zn}	01			
			Cd28	{pH,Ec,N,P,K,C}	02			

(Source: Compiled by Researcher)

Table No 3: Correlation coefficient of soil parameter

	Lime	Saline	OM	CEC	pH	EC	N	P	K	C	Cu	Fe	MN	Zn	Mg	B	MO
Lime	1.00																
Saline	0.29	1.00															
OM	0.33	0.56	1.00														

CEC	-0.55	-0.05	0.00	1.00													
pH	0.24	-0.02	0.04	0.32	1.00												
EC	-0.36	-0.21	-0.32	-0.40	-0.84	1.00											
N	0.38	-0.06	-0.12	0.16	0.31	-0.50	1.00										
P	0.14	0.00	0.10	0.30	0.83	-0.75	0.30	1.00									
K	0.15	-0.09	-0.50	-0.05	-0.04	0.03	0.72	0.00	1.00								
C	0.35	-0.13	-0.19	0.22	0.50	-0.64	0.96	0.48	0.63	1.00							
Cu	0.45	0.32	0.04	0.08	0.49	-0.58	0.62	0.16	0.38	0.61	1.00						
Fe	0.44	0.36	0.47	0.35	0.58	-0.74	0.45	0.27	-0.02	0.45	0.75	1.00					
MN	0.29	0.21	0.24	-0.38	0.17	-0.22	0.09	0.20	-0.20	0.16	0.08	0.04	1.00				
Zn	0.24	0.10	0.09	0.21	0.13	-0.40	0.88	0.17	0.66	0.78	0.51	0.43	0.00	1.00			
Mg	-0.22	-0.42	-0.40	0.02	-0.14	0.34	0.01	0.20	0.31	-0.01	-0.52	-0.52	-0.44	0.05	1.00		
B	-0.40	-0.52	-0.70	-0.03	-0.13	0.51	0.10	-0.10	0.36	-0.09	-0.37	-0.53	-0.22	-0.07	0.68	1.00	
MO	0.05	0.30	0.66	0.37	0.07	-0.47	0.27	0.11	-0.24	0.21	0.28	0.51	0.08	0.37	-0.50	-0.79	1.00

(Source: Compiled by Researcher)

Table No 4: Correlation coefficient of soil parameter

	Lime	Saline	OM	CEC	pH	EC	N	P	K	C	Cu	Fe	MN	Zn	Mg	B	MO
Lime	23.55																
Saline	0.04	0.00															
OM	0.01	0.00	0.00														
CEC	-0.65	0.00	0.00	0.06													
pH	0.34	0.00	0.00	0.02	0.08												
EC	-0.12	0.00	0.00	-0.01	-0.02	0.00											
N	86.59	-0.07	-0.04	1.86	4.11	-1.64	2208.73										
P	17.06	0.00	0.02	1.80	5.84	-1.29	349.13	603.91									
K	9.16	-0.07	-0.08	-0.31	-0.27	0.05	892.80	-1.07	705.00								
C	0.10	0.00	0.00	0.00	0.01	0.00	2.67	0.69	0.99	0.00							
Cu	3.24	0.01	0.00	0.03	0.21	-0.06	43.31	5.69	15.04	0.05	2.18						
Fe	4.70	0.02	0.01	0.19	0.36	-0.11	46.35	14.35	-1.15	0.06	2.43	4.85					
MN	13.46	0.06	0.01	-0.89	0.46	-0.15	41.25	46.96	-51.01	0.09	1.17	0.89	94.46				
Zn	1.67	0.00	0.00	0.07	0.05	-0.04	58.69	6.08	25.00	0.07	1.06	1.35	-0.03	2.01			
Mg	-0.05	0.00	0.00	0.00	0.00	0.00	0.03	0.22	0.37	0.00	-0.03	-0.05	-0.19	0.00	0.00		
B	-0.16	0.00	0.00	0.00	0.00	0.00	-0.37	-0.19	0.78	0.00	-0.04	-0.09	-0.18	-0.01	0.00	0.01	
MO	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.02	-0.04	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00

Table No. 3 shows the correlation coefficient of the two soil parameters ranges. It determines the relationship between two Soil properties with the help of correlation coefficient equation. Table determines the degree to which two variable's movements are associated. It gives the strength and direction of a linear relationship between two variables. The table shows that all soil Parameter shows moderate uphill and positive linear relation and almost all parameter shows positive relationship except EC, Fe, and Cu as these soils parameter availability is very low.

Table No. 4 shows covariance, which is the average of the products of deviations for each data point of soil parameters pair and are used for determining the relationship between two data sets. From table it is observed that how much two soil variables change together. The variables tend to show similar behavior, the covariance is positive in the opposite case, when the greater values of one variable mainly correspond to the smaller values of the other, the covariance is negative. The sign of the covariance therefore shows the tendency in the linear relationship between the variables, thus it is concluded that all soil parameter has positive covariance.

Principal component analysis is used to find out relationship between various soil parameters and classified them into dependent and independent variables.

Distribution of soil characteristics in relation to the two first PCA axes are shown in Figure. The eigenvalues and the proportion of variance explained by the axes are listed in Table No. 5 The proportion of the variance is simply the eigenvalue for that axis divided by the total variance, i.e. the sum of the diagonal of the cross-products matrix. Graph No. 5.2 shows Soil characteristics, occupied different regions of the diagram. Table No. 5also shows Pearson and Kendal correlation coefficients among soil variables, and between them and main axes of PCA.

Table No. 5 PCA Analysis (Min, Max, Mean, SD)

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
Lime	15	0	15	3.73	14.01	9.62	5.02
Saline	15	0	15	0.65	0.75	0.72	0.03
OM	15	0	15	0.23	0.26	0.25	0.01
CEC	15	0	15	10.17	11.14	10.71	0.25
pH	15	0	15	7.60	8.75	7.73	0.29
EC	15	0	15	0.32	0.66	0.53	0.07
N	15	0	15	267.77	468.47	314.97	48.65
P	15	0	15	20.25	128.35	43.58	25.44
K	15	0	15	722.62	826.04	754.23	27.48
C	15	0	15	0.40	0.63	0.47	0.06
Cu	15	0	15	2.84	7.05	4.63	1.53
Fe	15	0	15	5.21	13.68	9.55	2.28

MN	15	0	15	19.12	59.91	30.05	10.06
Zn	15	0	15	0.91	6.78	1.80	1.47
Mg	15	0	15	21.08	21.31	21.20	0.05
B	15	0	15	0.03	0.36	0.06	0.08
MO	15	0	15	0.33	0.36	0.35	0.01

(Source: Compiled by Researcher)

Table No. 6 PCA Analysis (Correlation matrix (Pearson (n)))

Soil Parameter	Lime	Saline	OM	CEC	pH	EC	N	P	K	C	Cu	Fe	MN	Zn	Mg	B	MO
Lime	1.00	0.29	0.33	-0.55	0.24	-0.36	0.38	0.14	0.15	0.35	0.45	0.44	0.29	0.24	-0.22	-0.40	0.05
Saline	0.29	1.00	0.6	-0.05	-0.02	-0.21	-0.06	0.00	-0.09	-0.13	0.32	0.36	0.21	0.10	-0.42	-0.52	0.30
OM	0.33	0.56	1.00	-0.004	0.04	-0.32	-0.12	0.10	-0.50	-0.19	0.04	0.47	0.24	0.09	-0.40	-0.70	0.66
CEC	-0.55	-0.05	0.00	1.00	0.32	-0.40	0.16	0.30	-0.05	0.22	0.08	0.35	-0.38	0.21	0.02	-0.03	0.37
pH	0.24	-0.02	0.04	0.32	1.00	-0.84	0.31	0.83	-0.04	0.50	0.49	0.58	0.17	0.13	-0.14	-0.13	0.07
EC	-0.36	-0.21	-0.32	-0.40	-0.84	1.00	-0.50	-0.75	0.03	-0.64	-0.58	-0.74	-0.22	-0.40	0.34	0.51	-0.47
N	0.38	-0.06	-0.12	0.16	0.31	-0.50	1.00	0.30	0.72	0.96	0.62	0.45	0.09	0.88	0.01	-0.10	0.27
P	0.14	0.00	0.10	0.30	0.83	-0.75	0.30	1.00	0.00	0.48	0.16	0.27	0.20	0.17	0.20	-0.10	0.11
K	0.15	-0.09	-0.50	-0.05	-0.04	0.03	0.72	0.00	1.00	0.63	0.38	-0.02	-0.20	0.66	0.31	0.36	-0.24
C	0.35	-0.13	-0.19	0.22	0.50	-0.64	0.96	0.48	0.63	1.00	0.61	0.45	0.16	0.78	-0.01	-0.09	0.21
Cu	0.45	0.32	0.04	0.08	0.49	-0.58	0.62	0.16	0.38	0.61	1.00	0.75	0.08	0.51	-0.52	-0.37	0.28
Fe	0.44	0.36	0.47	0.35	0.58	-0.74	0.45	0.27	-0.02	0.45	0.75	1.00	0.04	0.43	-0.52	-0.53	0.51
MN	0.29	0.21	0.24	-0.38	0.17	-0.22	0.09	0.20	-0.20	0.16	0.08	0.04	1.00	0.00	-0.44	-0.22	0.08
Zn	0.24	0.10	0.09	0.21	0.13	-0.40	0.88	0.17	0.66	0.78	0.51	0.43	0.00	1.00	0.05	-0.07	0.37
Mg	-0.22	-0.42	-0.40	0.02	-0.14	0.34	0.01	0.20	0.31	-0.01	-0.52	-0.52	-0.44	0.05	1.00	0.68	-0.50
B	-0.40	-0.52	-0.70	-0.03	-0.13	0.51	-0.10	-0.10	0.36	-0.09	-0.37	-0.53	-0.22	-0.07	0.68	1.00	-0.79
MO	0.05	0.30	0.66	0.37	0.07	-0.47	0.27	0.11	-0.24	0.21	0.28	0.51	0.08	0.37	-0.50	-0.79	1.00

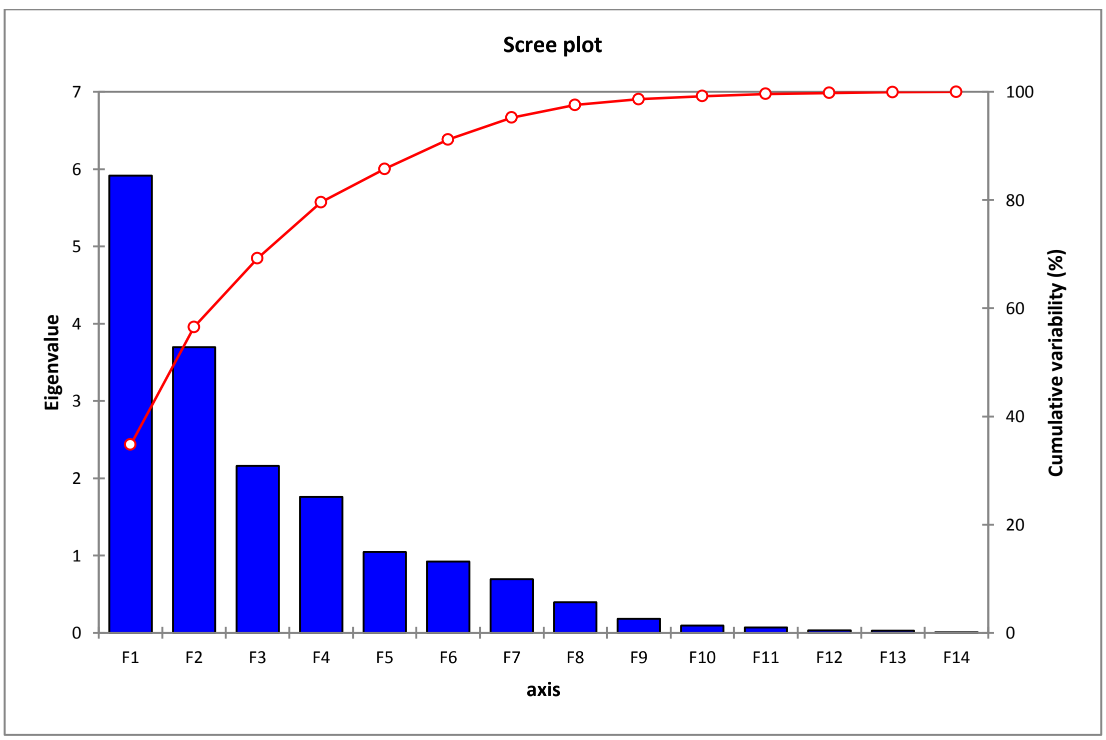
(Source: Compiled by Researcher)

Table No. 7 PCA Analysis (Eigenvalues)

Function Count	Eigenvalue	Variability (%)	Cumulative %
F1	5.91	34.79	34.79
F2	3.7	21.74	56.53
F3	2.16	12.71	69.24
F4	1.76	10.35	79.59
F5	1.05	6.16	85.75
F6	0.92	5.42	91.17
F7	0.69	4.08	95.25
F8	0.39	2.32	97.57
F9	0.18	1.06	98.63
F10	0.1	0.56	99.19
F11	0.07	0.41	99.6
F12	0.03	0.19	99.79
F13	0.03	0.16	99.95
F14	0.01	0.05	100

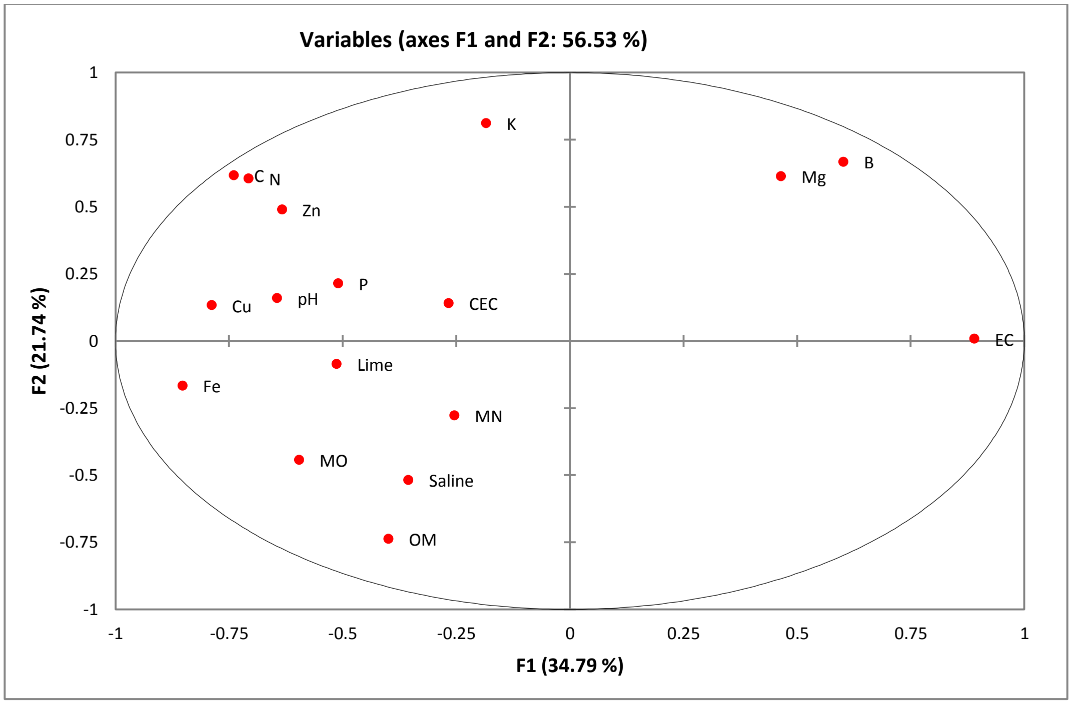
(Source: Compiled by Researcher)

Graph No. 1 Scree plots of eigenvalue, its variability among them, and cumulative variability in%



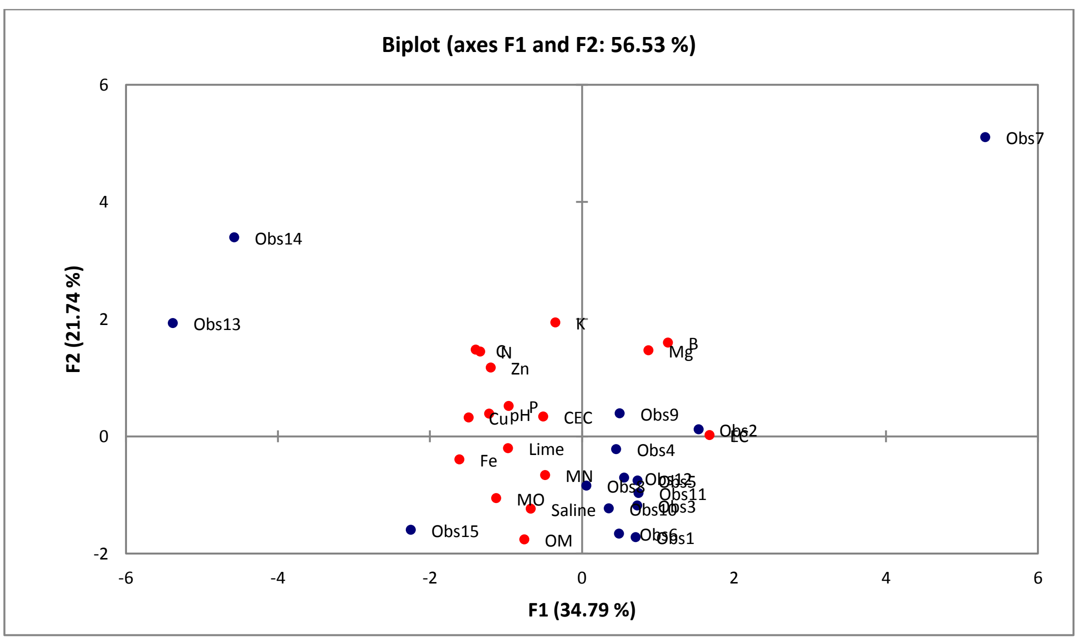
(Source: Compiled by Researcher)

Graph No. 2 Variables distribution space



(Source: Compiled by Researcher)

Graph No. 3 Biplot of PCA



(Source: Compiled by Researcher)

Principle Component Analysis of data which include mean, standard deviation, min-max values and data does not contain any missing value in total number of observation. Correlation coefficient among all soil parameter is calculated and it shows their positive uphill relationship among two parameters.

The eigenvalue table from the current analysis shows the eigenvalues for components 1 to 5 are 5.914, 3.695, 2.161, 1.760, and 1.047, respectively. Only these components demonstrated eigenvalues greater than 1.00, so the eigenvalue-one criterion would lead to retain and interpret only these components.

The first Principle Component (PC) accounts for the most variance (and hence have the highest eigenvalue), and the next component account for as much of the left over variance, and so on. Hence, each successive component will account for less and less variance.

- This variability in the percent of variance accounted for by each principal component show decreasing pattern for each pc.
- Cumulative % - This cumulative percentage of variance accounted for by the current and all preceding principal components. From Table No 5.11 observed that the first six components together account for 91.169% of the total variance. Scree plot is plotted with these three items. Bilot shows that first two component F1, F2 occupies 56.53% of total variance.

  Result
- PCA shows positive uphill for coefficient of correlation for soil parameters using independent and dependent variables representing orthogonal space in the scree graph. Scree plot is plotted with Eigen value, variability and cumulative variability for variables. Cumulative percentage of variance accounted for by principal components observes that the first six components together account for 91.169% of the total variance. Biplot shows that first two component F1, F2 occupies 56.53% of total variance.
- The Eigen value shows for components 1 to 5 are 5.914, 3.695, 2.161, 1.760, and 1.047, respectively. Only these components demonstrated Eigen values greater than 1.00, so the Eigen values lead to retain and interpret only these components. The variability is accounted for by each principal component show decreasing pattern for each principal component

Conclusion

As soil is complex system, its behavior is examined by making complex analysis using PCA because soil consists of number of components, number of macronutrient, micronutrient, related to each other and specifies the dependent variable and independent variable for making prediction system for productivity. This analysis is one of the way for judging yield production like regression analysis, simulation time series and neural network.

References

[1] Kevin Mc-Sweeney, Sabine Grunwald,,Soil Morphology, Classification, and Mapping, University of Wisconsin-Madison , 1999

[2] R. M. Larka, S. R. Kaffkab, D. L. Corwinca, Multi-resolution analysis of data on electrical conductivity of soil using wavelets, Journal of Hydrology Vol. 272, pp 276–290, 2003

[3] Abdi. H. & Williams L. J., Principal component analysis, Wiley Interdisciplinary Reviews Computational Statistics, Vol. 2, pp. 433– 459, 2010.

[4] Ali Salehi1 and G. ZahediAmiri, Study of Physical and Chemical Soil Properties Variations Using Principal Component Analysis Method in the Forest, North of Iran, Caspian J. Env. Sci. Vol. 3 No. 2, pp. 131-137, 2005.

[5] A Xing Zhu,Feng Qi,Amanda Moore,James E Burt,Prediction of soil properties using membership values Geoderma, Vol.158, pp.199–206, 2010.

[6] P. Bhargavi , Dr. S. Jyothi, Soil Classification Using Data Mining Techniques:A Comparative Study, International Journal of Engineering Trends and Technology- July to Aug,Vol.2, Issue 1,pp.55-59,2011.

[7] M. Kumar & A. L. Babel, Available Micronutrient Status and Their Relationship with Soil Properties of Jhunjhunu Tehsil, District Jhunjhunu, Rajasthan, India, Journal of Agricultural Science, Vol 3, No 2, pp 102-118, 2011.

[8] J.O. Ogunwole, E.N. O. Iwuafor, N.M. Eche, J. Diels, Effect of organic and inorganic soil amendments on soil physical and chemical properties in a west africa savanna agroecosystem,Tropical and Subtropical Agroecosystems, Vol. 12, No. 2, pp. 247-255, 2010.

[9] A.Akbarzadeh,R.Taghizadeh,Mehrjardi H. Rahimi Lake & H.Ramezanpour,Application of artificial intelligence in modeling of soil properties,Enviromental Research Journal Vol.3,No.2,pp.19-24,2009.

