ISSN print: 1224-5984
ISSN online: 2248-3942

Rom Biotechnol Lett. 2023; 28(3): 3969-3975
doi: 10.25083/rbl/28.3/3969.3975

Received for publication, October, 14, 2023
Accepted, October, 21, 2023

Original article

A statistical method for source approtionment of soil pollution

DEVENDRA MANI TRIPATHI[1]* AND SMRITI TRIPATHI[2]

[1] Institute of Environment and Development Studies, Bundelkhand University, Jhansi – 284128, India
[2] Dept. of Microbiology, Bundelkhand University, Jhansi – 284128, India

Abstract The present study was conducted to evaluate the source apportionment of soil pollution and its impact on soil quality using multivariate statistical analysis tools. Multivariable statistical techniques such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and correlation offers superior interpretation of complicated data sets to better understand the soil quality. Soil samples were analyzed for physicochemical parameters (pH, temperature, electrical conductivity, organic carbon, moisture content, phosphate, potassium and sulphar) and for some important heavy metals (zinc, iron, manganese and copper). Sampling stations have been classified into three groups using Cluster Analysis (CA) in a convincing way. Results revealed that among all the sampling stations, three significantly different groups: (1) stations having highest pollution sources (HW and MW), (2) Moderate polluted area (SW and IW) & (3) control site (VW) of the study stretch. The PCA generated 2 significant factors having eigenvalues >1 which explain 82.78% of the variance in the dataset. Each soil quality parameter with strong correlation coefficient value was considered to be significant ($p<0.01$). Therefore, the present study revealed that mining activities, dumping of municipal sewage, disposal of biomedical waste and industrial effluent including solid waste (e.g fly ash) have greater influence on the soil quality. Results also showed that the concentrations of heavy metals (Cu, Mn and Zn) in highly polluted soil are beyond the permissible limit. The pollutant levels were significantly different in the three groups of soil pollution character samples, which were confirmed by ANOVA analysis.

Keywords Soil, Cluster Analysis, Factorial Analysis, PCA, mining activity, source apportionment

*Corresponding author: e-mail: smritienvs@gmail.com

Introduction

Soil is one of the important and valuable resources of the nature. It is composed of 50% of organic and inorganic matters, and 50% of air and water which fills existing vacant spaces of the soil and keeps live organisms of the soil [9]. Soil quality is controlled by complex anthropogenic activities and natural factors [1].

Population growth, unplanned urbanization and increasing industrialization lead to enormous pollution sources to degrade the soil quality. Mining has been one of the most common activities since ancient times and continues to remain so in the modern world. Mining is an important part of our economy. The continued advancements in industrialization and the ever-increasing demand for energy resources and minerals, have led to a spurt in mining activities, bringing in its wake imbalances in ecological equilibrium and many environmental hazards

Mining activities such as crushing, grinding, washing, smelting and all other processes used to extract, concentrate generate waste products such as mine overburden and mine tailings (waste soil) which directly affect the cultivated land, forest or grazing land, and the overall loss of production. Inorganic fertilisers are also frequently used in conventional agriculture to attain high crop yields. However, the intensive application of these chemical inputs can decrease the quality of agricultural soils and increase the risks of environmental pollution [21]

Another significant element is the complexity of pathways determined by emission sources, interactions with soil surfaces, and changes over time in the chemical and biological conditions in the environment. Soil ecosystem includes inorganic and organic constituents, and the microbial groups. In suburban areas, the use of industrial or municipal wastewater for irrigation purpose is common practice in many parts of the world [22], including India [18,19]. An additional source of waste that finds its way from hospitals and clinics has long term effect on environment [29]. In developing countries open dumpsites are common, due to the low budget for waste disposal and non-availability of trained manpower, it also poses serious threat to soil quality. The contamination of soil can cause adverse effects on human health, animals and soil productivity [20]. It is depriving our ecosystem of the natural balance and bear result beyond any repair. Assessment of soil pollution becomes difficult when contaminants belong to different sources and their products are variably distributed [28].

The application of multivariable statistical techniques such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA), and correlation offers superior interpretation of complicated data sets to better understand soil quality [26]. Therefore, the present study was conducted for the evaluation of pollution sources and its impact on soil quality using multivariate statistical analysis tools.

Material and methods

Study area

The study area lies in southwest portion of Uttar Pradesh state in India between 25^0 30’ and 25^0 57’ N latitude and 78^0 40’ and79^0 25’ E longitudes. The present study was divided in 5 division (on the basis of solid waste discharge i.e. Hospital (HW), Household sector (SW), Vegetation area (VW), mining area (MW) and industrial area (IW) of the city Jhansi. The 10 sites were divided as HW comprises Medical Hospital (MH) and Germany Hospital (GH), IW includes Parichha Dam (PcD), Pahuj Dam (PD) and Thermal Power Plant (TPP), SW contains Bundelkhand University (BU) and Sipri city (S), MW impose Gora Machia (GM) and Bhagwantpura (BP) mining areas whiles VW embrace Pichor vegetation area as a control site (Fig.1).

Sampling and analysis

Soil sampling was done manually on fortnightly basis and transferred to the laboratory, preserved and stored for further analytical determinations and treatment. Soil samples from the depth range of 2 to 20 cm were collected and loaded in sterile envelopes. Biological activity such as microbial respiration, chemical activity such as precipitation or pH change, and physical activity such as aeration or high temperature must be kept to a minimum. Methods of preservation include cooling, pH control, and chemical addition. Air dried and ground up to pass through a 2-mm sieve for soil particle size distribution, a 1-mm sieve for soil pH, and a 0.25-mm sieve for pH, electrical conductivity (EC), OC, K, S, Zn, Fe and Mn, Temp, PO_4^{-2} and Cu using standard methods. Soil was digested by sulfuric acid–perchloric acid and then used by molybdenum– antimony colorimetry to measure Phosphate (PO_4^{-3}).

Organic matter was detected by oxidation with a potassium dichromate-titration of $FeSO_4$. Procedures for the determination of soil basic properties were standard methods recommended by American Society for Testing Material (ASTM, 1985). Heavy metals, Zn, Fe, Mn and Cu were analyzed by using AAS (Perkin Elmer, USA, 2009). For data accuracy, reagent blanks and standards were analyzed at beginning and end of the measurement. For analytical precision, the samples were analyzed in triplicates. The reproducibility was within ±5 % in all measurements.

Statistical analysis

Multivariate analyses were performed through hierarchical agglomerative CA and PCA. After standardization of the data (z-score transformation), CA was performed on all the six stations by single linkage method using Euclidean distance as a measure of similarity. PCA was performed to obtain significant principal components (PCs) from the data of wastewater from all the stations and groups obtained from CA with a view to assess spatial differences in ground water quality. All the statistical and mathematical calculations were conducted using SPSS 16 software.

Results and Discussion

General description of soil quality

The descriptive statistics of physical and chemical parameters analyzed in the soil samples collected from different depths in the study site are given in Table 1. Results revealed that pH value ranged from slightly acidic to mildly alkaline in all the sampling stations (6.60 – 7.34). The EC and pH of this soil samples are within the permissible limits (WHO, 1984). The permissible limit of organic carbon is 0.8% and the values ranged between 0.378 – 1.132%. Results showed that the concentrations of heavy metals (Cu, Mn and Zn) are beyond the permissible limit. The values of Cu, Mn and Zn ranged between 0.298 – 0.398, 1.27 – 8.08 and 1.684 – 7.998 respectively. The sulphur values found between 0.0073 – 0.635, which is much lesser than the permissible limit prescribed by BIS. The nutrients (K and PO_4^{-3}) present in soil samples are also within the permissible limit which depicts in the range of 147.2 – 417.4 and 4.262 – 17.41 respectively.

Pattern recognition of soil quality

Figure 1 revealed the characteristic features and types of pollution sources in soil. Sampling stations have been classified into three groups using Cluster Analysis (CA) in a convincing way. Clustering analysis is an unsupervised multivariate technique used to classify objects into categories or clusters based on their nearness or similarity[23,18]. Among all the ten stations, three significantly different groups: (1) stations having highest pollution sources (HW and MW), (2) Moderate polluted area (SW and IW) &(3) control site (VW) of the study stretch were formed by CA. In group 1, the sites of MH, GH, GM and BP having higher polluted soil, the pollution is due to the dumping of hospital waste, biomedical waste and active mining activities. Group 2 comprising of PcD, TPP, PD, S and BU comprises moderate soil pollution due to discharge of municipal waste, influence of Dams and fly ash dumping from thermal power plant. However, in group 3 the soil quality is good because of having vegetation area. Previous worker [11]; also reported the similar results in different

Table 1. Descriptive Statistics

Parameters	Minimum	Maximum	Mean	Std. Deviation	Variance
pH	6.60	7.34	7.15	0.209	0.044
Temp (0C)	23.54	25.08	24.61	0.472	0.222
OC (%)	0.378	1.132	0.807	0.225	0.051
PO4-3	4.262	17.41	8.359	5.084	25.847
K	147.2	417.40	263.58	86.72	7.520E3
S	0.0073	0.6248	0.128	0.198	0.039
Zn (ppm)	1.684	7.998	4.389	2.068	4.277
Fe (ppm)	1.854	5.610	2.5086	1.132	1.281
Mn (ppm)	1.270	8.080	2.963	2.038	4.154
Cu (ppm)	0.298	0.398	0.346	.0347	0.001

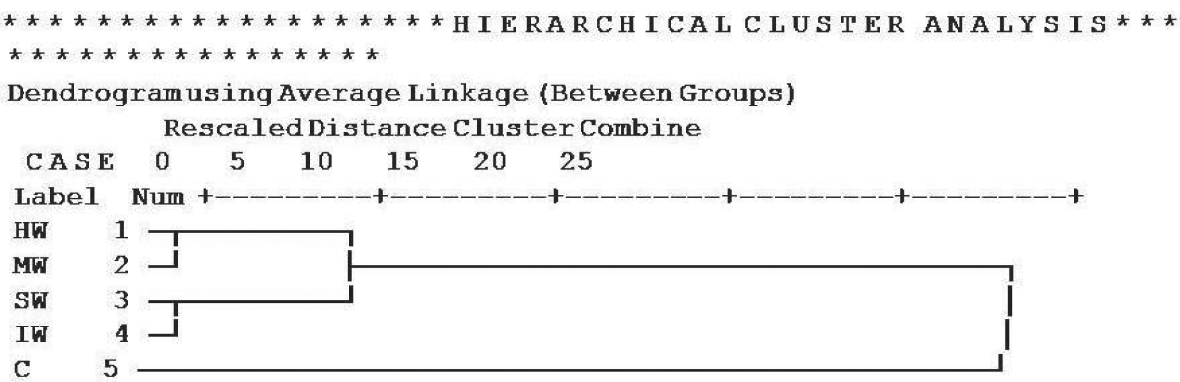

Fig. 1. Dendrogram based on hierarchical agglomerative clustering of sites

areas. The values obtained in this study which are above the maximum allowable contents of metals for agricultural purposes as proposed by Blankenship et al. 1994 in these soils are indicative of anthropogenic action on the soil total elemental composition.

Comparison among parameters and sites

Factor analysis provides information regarding the most meaningful parameters, which describe whole data set rendering for data reduction with minimum loss of original information [23, 24]. It is a powerful technique for pattern recognition and attempts to explain the variance of a large set of inter-correlated variables and transform it into a smaller set of independent variables [17, 27].

Principal component analysis/factor analysis was done on standard data set. An eigenvalue gives a measure of the significance of the factor: the factors with the highest eigenvalues are the most significant. Eigenvalues of 1.0 or greater are considered significant [17] Classification of factor loading is thus 'strong', 'moderate' and 'weak', corresponding to absolute loading values of > 0.75, 0.75-0.50 and 0.50-0.30, respectively [10].

The PCA generated 2 significant factors having eigenvalues >1. These factors, eigenvalues and proportion of variance explained are presented in Table 2. These two factors explain 82.78% of the variance in the dataset. The factor scores are mapped out in fig. 2. Each soil quality parameter with strong correlation coefficient value was considered to be significant (p<0.01). Scatter plot of scores (fig. 2) for the principal components, PC1 and PC2 was obtained for soil samples from all sites (highest pollution, moderate pollution and control site). Therefore, the present study revealed that mining activities, dumping of municipal sewage, disposal of biomedical waste and industrial effluent including solid waste (e.g. fly ash) have greater influence on the soil quality. The factor analysis results including the loadings, variance contribution rate of each VF and cumulative variance contribution rate are presented in Table 3. The loading plot (fig. 2) for the two groups showed the relationships

Table 3. Loading of 10 experimental variables on significant variance factor

Parameter	PC1	PC2
pH	0.281	***0.802***
Temp	-0.444	***0.794***
OC	***0.854***	-0.495
PO4	***0.898***	-0.053
K	0.630	0.496
S	0.543	***0.803***
Zn	-0.584	-0.483
Fe	***-0.785***	-0.562
Mn	***0.787***	-0.556
Cu	***0.787***	-0.556
% Variance	46.981	46.981
Cumulative % variance	35.793	82.775

(Bold and italic values indicate strong and moderate loadings, respectively.)

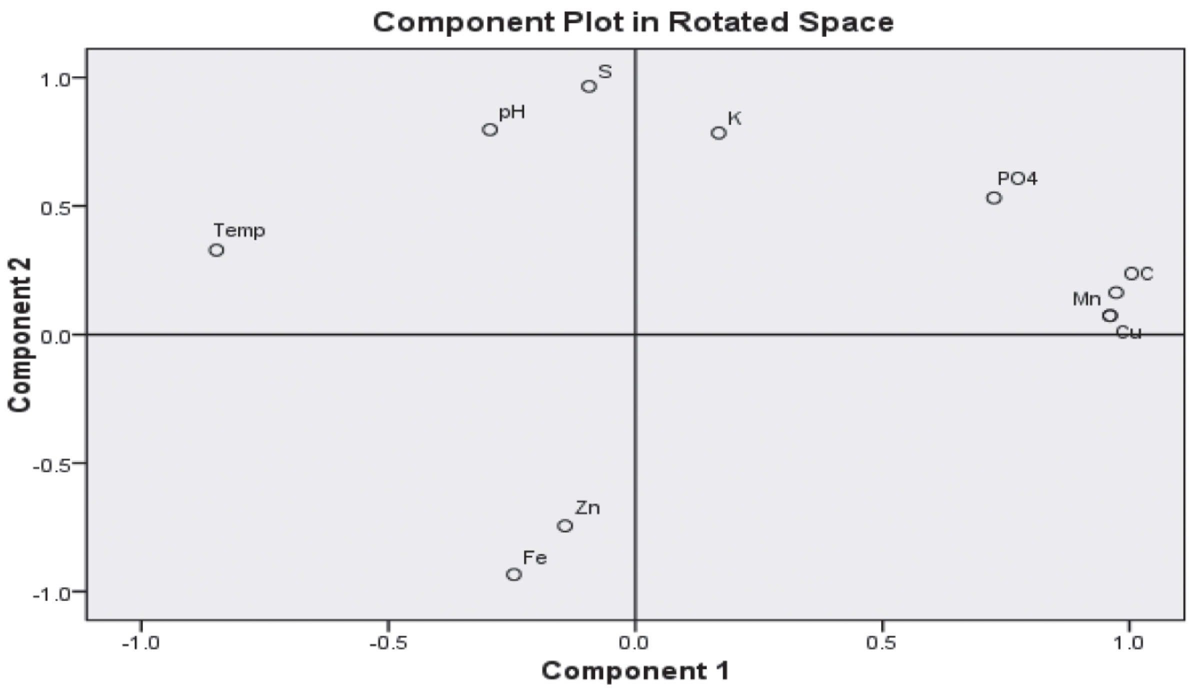

Fig. 2. Scatter plot of loading for PC1 and PC2

Table 4. Inter-elemental correlation analysis matrix of selected parameters in the soil samples

		pH	Temp	OC	PO_4^{-3}	K	S	Zn	Fe	Mn	Cu
pH	r^2	1									
	p										
Temp	r^2	0.133	1								
	p	0.649									
OC	r^2	0.162	0.199	1							
	p	0.581	0.496								
PO_4^{-3}	r^2	0.177	-0.236	**-0.684****	1						
	p	0.546	0.416	0.007							
K	r^2	-0.034	0.171	-0.327	0.**683****	1					
	p	0.908	0.559	0.254	0.007						
S	r^2	0.160	0.194	0.480	-0.067	0.096	1				
	p	0.584	0.506	0.082	0.821	0.744					
Zn	r^2	0.160	0.416	0.462	-0.186	0.322	0.**692****	1			
	p	0.586	0.139	0.096	0.524	0.261	0.006				
Fe	r^2	0.264	0.294	-0.081	-0.018	-0.290	-0.319	-0.485	1		
	p	0.362	0.308	0.784	0.950	0.315	0.266	0.079			
Mn	r^2	0.251	0.247	-0.334	0.116	-0.189	-0.460	-0.517	0.**918****	1	
	p	0.388	0.395	0.243	0.692	0.517	0.098	0.058	0.000		
Cu	r^2	0.261	**-0.542***	-0.491	0.**845****	0.346	-0.009	-0.191	-0.124	0.021	1
	p	0.368	0.045	0.075	0.000	0.226	0.975	0.512	0.673	0.942	

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

among the parameters; the smaller distance, and the stronger correlation between the parameters [15]. However, the pollutant levels were significantly different in the three groups of soil samples. Previous workers [30]; also reported the similar finding.

Correlation analysis

Inter-elemental correlation analysis provides information about the source and pathway of metals and physicochemical parameters. Table 4 demonstrate the Pearson's correlation coefficient of all the parameters analyzed in the soil samples. Inter-elemental correlation analysis shows that PO_4^{-3} has strong positive correlation with K ($r^2 > 0.723$, $p<0.05$) and Cu ($r^2>0.841$, $p<0.001$), while strong negative correlation with OC ($r^2>0.660$, $p<0.05$). Likewise, S has positive correlation with Zn ($r^2>0.731$, $p<0.05$) and Fe has strong positive correlation with Mn ($r^2>0.922$, $p<0.001$). Similar results were reported previous workers [12, 16]. Heavy metals in the soil are binding with various oxides and classified as metals bound with carbonates, Mn-oxides, Fe-oxides, organic matter, sulfides and exchangeable and residuals [25]. In the study site, elements have good correlation with Fe and Mn may be bound with Fe and Mn oxides in the sediment matrices.

Major source identification

Present study revealed total 5 types of soil pollution sources in the study area including mining waste, hospital waste, industrial waste, solid waste disposal from house hold sectors and agricultural waste. Several studies have been conducted to identify pollution types. [6] found that soil quality degradation in Nigeria was mainly related to the dumping of hospital waste which is responsible for highest enrichment of metal content in soil. [3] discovered that open dumping of municipal solid waste in Islamabad city increase pH, organic content, conductivity and available heavy metal content in soil[2] found that Mining activities have both local and regional impacts on terrestrial ecosystems. Mines produce large quantities of waste-rock and tailings which must be disposed of on land and degrade the soil quality. The quantity of pollution may be determined by a number of variables and actual pollution levels and types [8, 27]. In addition, the different pollution types in the above results along with this work indicated that soil samples of each site had unique physical and chemical characteristics due to its different natural and anthropogenic features [8]. However, results suggested that pollution factors in soil that play important roles in influencing the quality in one environment may not be important in another.

Conclusions

In the present study, multivariate statistical analysis was used for the estimation of and apportionment of sources of soil pollutants and their impact on soil quality. PCA performed on soil samples revealed that in present study nine sampling stations were influenced by pollution excluding one control site. The possible reason is that mining activity, discharge of biomedical waste open dumping of house hold waste may affect the soil quality. However, the pollutant levels were significantly different in the three groups of soil samples, which were confirmed by ANOVA analysis and cluster analysis of the sampling stations. Cluster Analysis classified the soil received at sampling stations into three significant groups depending on similarities among them. The present results were also proved correct by using chemometric tools. In addition, the different pollution types in the above results along with present work indicated that soil samples of each site had unique physical and chemical characteristics and results suggested that pollution factors in soil that play important roles in influencing the quality in one environment may not be important in another.

Acknowledgement

The authors are thankful to Institute of Environment and Development Studies, Bundelkhand University, Jhansi – 284128, India, for providing laboratory facilities.

References

1. Abrahams PW. Soils: their implications to human health. Science of total environment. 2002:291 (3), 1 – 32.
2. Allan JR. Impact of Mining Activities on the Terrestrial and Aquatic Environment with Emphasis on Mitigation and Remedial Measures. Environmental Science. 1995;119 – 140.
3. Ali SM, Pervaiz A, Afzal B, Hamid N, Yasmin A. Open dumping of municipal solid waste and its hazardous impacts on soil and vegetation diversity at waste dumping sites of Islamabad city, Journal of King Saud University – Science. 2014;26, 59–65.
4. ASTM, 1985. Standard test method for particle-size analysis of Soils – D 422-463.
5. Blankenship APY, Chan AD, Jones PB. Toxic combustion by-products from incineration of chlorinated hydrocarbons and plastic. Chemosphere. 1994;28:183–196.
6. Ephraim IP, Ita A, Eusebius IO. Investigation of soils affected by burnt hospital wastes in Nigeria using PIXE. Springer Plus. 2013;2:208
7. Feigin A, Ravina I, Shalhevet J. Irrigation with treated sewage effluent. Berlin: Springer. 1991.
8. Huang F, Wang XQ, Lou LP, Zhou ZQ, Wu JP. Spatial variation and source apportionment of water pollutionin Qiantang River (China) using statistical techniques. Water Research. 2010;44: 1562–1572.
9. Khakbaz PPN, Mahdeloei S, Heidari A. Soil Pollution Control Management Techniques and Methods. Annals of Biological Research. 2012;3(7):3101-3109.
10. Liu CW, Lin KH, Kuo YM. Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. Science of the Total Environment. 2003;313:77–89.
11. Lu X, Wang L, Lei K, Huang J, Zhai Y. Contamination assessment of copper, lead, zinc, manganese and nickel in street dust of Baoji, NW China. Journal of Hazardous Materials. 2009;161(2–3):1058–1062.
12. Mico, C., Recatala, L., Peris, M., & Sanchez, J. 2006. Assessing heavy metal sources in agricultural soils of an European Mediterranean area by multivariate analysis. Chemosphere, 65(5), 863–872.
13. Oliver MA. Soil and human health: A review. Eur. J. Soil Sci. 1997;48:573–592.
14. Pal R, Rai JPN. Phytochelatins: Peptides involved in heavy metal detoxification. Applied Biochemistry Biotechnology. 2010;160 945-963.
15. Qu W, Kwlderman P. Heavy metal contents in the Delft canal sediments and suspended solids of the river Rhine: multivariate analysis for source tracing. Chemosphere. 2001;45: 919–925.
16. Rajmohan N, Prathapar SA, Jayaprakash M, Nagarajan R. Vertical distribution of heavy metals in soil profile in a seasonally waterlogging agriculture field in Eastern Ganges Basin. Environ Monit Assess. 2014;186:5411–5427.
17. Shrestha S, Kazama F. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. Environmental Modeling and Software. 2007;22:464– 475.
18. Singh KP, Malik A, Sinha S. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: a case study. Analytica Chimica Acta. 2005;538: 355–374.
19. Singh KP, Mohon D, Sinha S, Dalwani R. Impact assessment of treated/untreated wastewater toxicants discharge by sewage treatment plants on health, agricultural, and environmental quality in wastewater disposal area. Chemos. 2004;55, 227–255.
20. Smith CJ, Hopmans P, Cook FJ. Accumulation of Cr, Pb, Cu, Ni and Cd in soil following irrigation with treated urban effluent in Australia. Environ. Pollut. 1996;94(3):317–323.
21. Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S. Agricultural sustainability and intensive production practices. Nature. 2002;418:671–677.
22. Urie DH. The status of wastewater irrigation of forest. In: Cole, D.W., Henry, C.L., Nutter, W.L. (Eds.), The Forest Alternative for Treatment and Utilization of Municipal and Industrial Wastes. University of Washington Press, Seattle, WA, USA. 1986;26–40.
23. Vega M, Pardo R, Barrado E, Deban L. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research. 1998;32:3581–3592.
24. Wunderlin D A, Diaz M P, Ame M V, Pesce S F, Hued A C, Bistoni M A. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba, Argentina). Water Research. 2001;35:2881–2894.
25. Yu KC, Tsai LJ, Chen SH, Ho ST. Correlation analyses on binding behavior of heavy metals with sediment matrices. Water Research. 2001;35(10):2417–2428.

26. Zhang Y, Guo F, Meng W, Wang XQ. Water quality assessment and source identification of Daliao river basin using multivariate statistical methods. Environment Monitoring and Assessment. 2009;152:105–121.
27. Zhao J, Fu G, Lei K, Li Y. Multivariate analysis of surface water quality in the Three Gorges area of China and implications for water management. Journal of Environmental Sciences. 2011;23(9):1460–1471.
28. Partha V, Murthya NN, Saxena PR. Assessment of heavy metal contamination in soil around hazardous waste disposal sites in Hyderabad city (India): natural and anthropogenic implications. E3 J. Environ. Res. Manage. 2011;2(2):027-034.
29. Pattnaik S, Reddy MV. Assessment of municipal solid waste management in Puducherry (Pondicherry), India. Resourc Conservat Recycl. 2009;54:512-520
30. Pejman AH, Nabi Bidhendi GR, Karbassi AR, Mehrdadi N, Esmaeili Bidhendi M. Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques. Int. J. Environ. Sci. Tech. 2009;6(3):467-476.