Check for updates

Asian Journal of Water, Environment and Pollution, Vol. 17, No. 4 (2020), pp. 23-29.
DOI 10.3233/AJW200047

Water Quality Assessment of a Water Body using Principal Component Analysis – Sanjay Lake, New Delhi, India

S.K. Singh, Harshit Sachdeva*, Harshit Joon and Mridul Narang
Department of Environmental Engineering, Delhi Technological University, Bawana Road, Delhi-42, India
✉ sachdevaharshit50@gmail.com

Received May 14, 2020; revised and accepted May 30, 2020

Abstract: The present study aims at assessing the water quality at Sanjay Lake, Delhi, India using multivariate statistical techniques. A total of 16 physiochemical variables such as pH, conductivity, TDS, COD, DO, chloride, total hardness, magnesium hardness, phosphate, suspended solids, sulphate, nitrate, fluoride, sodium, potassium and calcium were analysed in water samples collected on January, 2020 from 10 sampling sites. The data were evaluated using IBM Statistics software SPSS 25 for principal component analysis, which limits the multiple data dimensions without the loss of vital information for better understanding. Five principal components were defined to be responsible for the data collection, indicating 93% of the total variance of data collection; of which 36.38% by PC1, 24.84% by PC2, 15.13% by PC3, 11.26% by PC4, and 6.36% by PC5 indicated that the domestic sewage, the municipal sewage waste and the untreated industrial effluent discharges may be affecting the water quality. The present study shows that PCA techniques are useful resources for the identification of important surface water quality parameters.

Key words: Sanjay lake, multivariate statistical analysis, principal component analysis.

Introduction

Water is the most essential natural resource required for the life on earth to thrive. Whilst water is a renewable source, it is also a restricted supply in the meantime. But its demand is rising day by day due to growing populations, rapid urbanisation, industrial development, irrigation and other activities. With the planet's second largest population at 1.3 billion and further expecting a surge of 1.8 billion by 2060, India is on the verge of a severe water crisis (Singh and Kumar, 2015). According to the Ministry of Water Resource, India possesses 4% of the world's freshwater resources of 2.5% of the global landmass. This is, however, a matter of grave concern for India which is home to about 16% of world's population, hence, it is difficult to cater to the basic needs of its citizens. One of the country's main sources of fresh water is the surface flow. Surface sources of water are those in which water flows over the surface of the earth, and are thus directly available in the form of water supplies. Lakes and ponds are one of the major sources of surface water (Kishor et al., 2005). The lakes and ponds are differentiated based on the size of depression and the sources of collection of water. Lakes may be natural or man-made. In India, there are around 17 considerable artificial lakes and over 60 natural lakes. Lake ecosystems are crucial resources for aquatic wildlife and human needs, therefore, any change to their environmental characteristics and water renewal rates has extensive ecological and societal

*Corresponding Author

impacts (Jhamnani and Singh, 2009). Declining water quality in the lakes and reservoirs is a major problem, especially in developing countries like India (Vyas et al., 2006). The main causes behind this decline are rapid industrialisation, urbanisation, population explosion, destructive agricultural activities, irresponsible consumption and wastage of water (Deepika and Singh, 2015). These detrimental activities lead to various problems such as loss of oxygen, toxic algal growth, loss of biodiversity and aquatic life too (Vousta et al., 2001).

The authors studied Sanjay Lake, a man-made lake, situated in Trilokpuri in East Delhi, India. It was developed into a recreational area by the Delhi Development Authority (DDA). The lake is located in the middle of a 170 acres forest area, popularly known as Sanjay Van or Sanjay Park. The lake itself is extended over an area of approximately 42 acres. The lake was developed in the late 1970s and was inaugurated in 1982. It is a hub for various migratory birds and many indigenous trees. But for the past few years, the lake is paying the price for rapid urbanisation.

It was seen that the residential colonies of Mayur Vihar and Patparganj consume water from the rainwater catchment region, which is the primary source of water supply in the lake. An embankment was built along the Yamuna River, which lies barely 1 kilometre away from the lake, to obstruct the river water from flooding into the region.

Various factors influence the environment of the lake, which can be studied after knowing its physical and chemical parameters. The determination of these parameters can conclude the condition of the lake. Thereafter, appropriate measures can be taken to clean the lake. Also, its water can be used for different chores such as gardening, sanitation and recreational activities, etc.

The present study uses the multivariate statistical analysis tool to explain the intricate water quality data. It helps to analyse the complicated datasets and derive useful information about the possible impact of the surrounding environment on the water quality. It is very useful in analysing data comprising of a large number of variables. The multivariate method is used in the present study to gain information about the most significant characteristics of the physicochemical variables by reducing the original data sets into the most relevant data sets (De Bartolomeo et al., 2004; Kazi et al., 2009; Weinning and Erickson, 1994).

The present study intends to analyse the physio-chemical water quality criteria in water samples collected from the Sanjay Lake basin in Delhi, India. The data sets obtained from field measurement are subjected to the cluster principal component analysis (PCA) and factor analysis (FA) techniques to define the natural and anthropogenic origin and to estimate the contributions of different sources on concentrations of determined parameters (Kim et al., 2009; Singh et al., 2005).

Materials and Methods

Study Area

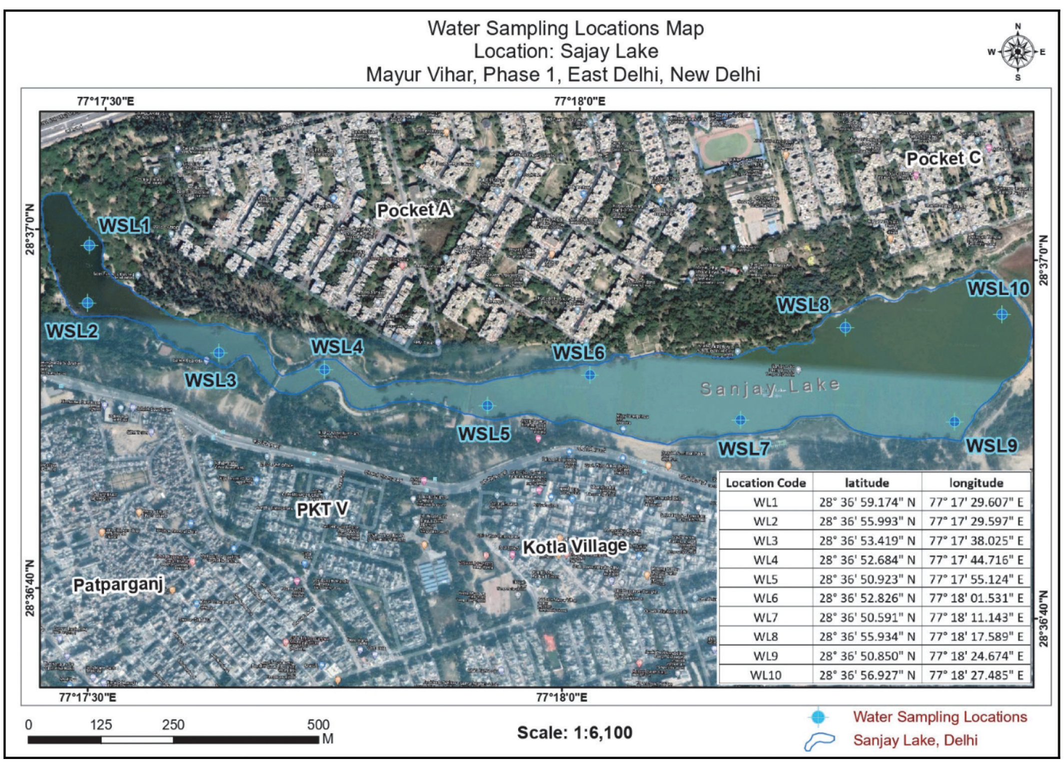
Sanjay lake (28º36’51” N 77º18’14” E) is an artificial lake developed by Delhi Development Authority (DDA) in Trilokpuri in East Delhi, India (Jain, 2009), adjoining Mayur Vihar residential area (Mohan, 2007). The location specific details are given in Figure 1. Surface water samples were collected during January 2020 from 10 different locations covering an area of about 17 hectares. 2L PET bottles were pre-rinsed and used for the collection of samples. The parameters such as pH, electrical conductivity (EC), total suspended solids (TDS), and salinity were determined on the site using Star A320 model multi-parameter metre. The samples were stored in an ice box during transfer to prevent degradation. Whereas the determination of other parameters was completed within the next 48 hours (APHA, 2005). Total hardness (TH), chloride (Cl), bicarbonate (HCO_3^-) and carbonate (CO_3^{2-}) were analysed using volumetric analysis. Spectrophotometry technique was applied for the determination of phosphate (PO_4^{3-}), sulphate (SO_4^{2-}), and nitrate (NO_3^-) using LabIndia make UV3092 model double-beam UV-VIS spectrophotometer. Sodium (Na^+), potassium (K^+), calcium (Ca^{2+}) and lithium (Li^+) were determined using Systronics make 128 μC model flame photometer.

Methodology

Principal component analysis (PCA) is the most popular technique for interpreting large datasets by reducing the dimensionality in an interpretable way without losing most of the information in the data. It is an adaptive data analysis technique as it involves generating variables that are functions of those in the dataset, rather than being pre-defined.

Pearson’s coefficient (*r*) measures the magnitude and direction of the statistical relationship between two variables which involve the use of the method of covariance in the dataset. The values of Pearson’s coefficient range from +1 to -1, with 0 indicating no association between the variables. The positive

Figure 1: Location map.



Location Code	latitude	longitude
WL1	28° 36' 59.174" N	77° 17' 29.607" E
WL2	28° 36' 55.993" N	77° 17' 29.597" E
WL3	28° 36' 53.419" N	77° 17' 38.025" E
WL4	28° 36' 52.684" N	77° 17' 44.716" E
WL5	28° 36' 50.923" N	77° 17' 55.124" E
WL6	28° 36' 52.826" N	77° 18' 01.531" E
WL7	28° 36' 50.591" N	77° 18' 11.143" E
WL8	28° 36' 55.934" N	77° 18' 17.589" E
WL9	28° 36' 50.850" N	77° 18' 24.674" E
WL10	28° 36' 56.927" N	77° 18' 27.485" E

Figure 2: Sampling sites in the study area.

correlation ($+1 \geq r > 0$) indicates that as the value of one variable increases so does the value of the second variable. Similarly, the negative correlation ($0 > r \geq -1$) indicates that if the value of one variable increases, then the value of the second variable decreases.

The 10 sampling sites were uniformly selected to map the 16 parameters to determine the water quality of the lake. These sites and their coordinates have been indicated in Figure 2.

The Pearson coefficient and principal component analysis were conducted using the IBM SPSS Statistics 25 programme.

Results and Discussion

Pearson correlation matrix for physico-chemical water parameters in the research area is shown in Table 1**.** A correlation matrix prepared between 16 different parameters of water quality showed only few parameters had major correlation relationships. The highest positive correlation at P<0.01 was observed between TDS and EC with r=0.991 and between Na^+ and Ca^{2+} with r=0.992. Sodium and EC with r=0.679 and 0.671, respectively, are positively correlated with potassium. There is also a strong negative correlation between fluoride and sulphate (-0.840). In other physicochemical parameter, the correlation was found from being moderate to insignificant.

Table 2 listed the initial principal component (PC) and its eigen value corresponding to the percentage of the variance in each PC. Figure 3 represents a scree plot of the eigen value for each component in which five principal components were defined with eigenvalue >1 constituting almost 93.96% of the total variance in the water dataset. Eigenvalue representing the five PC is the most essential component which represents more than 90% variance in the water quality. Total 36.38% by PC1, 24.84% by PC2, 15.13% by PC3, 11.26% by PC4 and 6.36% by PC5.

Communalities explain the amount of variation in each variable that has been taken into account. Initial communalities evaluate the change in each variable recitable in by all the factors. This should be equal to 1.0 under principal component extraction for correlation analyses. Extraction communalities are the assessments of variance in each variable considered in by the components. The communalities are high for all cases which implies that the variables are illustrated well by the extracted components which is shown in Table 3**.**

Component loading and communalities for each variable in five selected components before varimax rotation and after varimax are explained in Tables 4 and 5. Communalities provide the efficiency value for the reduced set of variables and the degree of participation of each variable in the selected five components. Various water quality parameters falling under the

Table 1: Pearson's correlation matrix

	Correlations															
	EC	*TDS*	*pH*	*S.S*	*COD*	*DO*	*Cl^-*	*TH*	*Mg^{2+}*	*PO_4^{3-}*	*SO_4^{2-}*	*NO_3^-*	*F^-*	*Na^+*	*K^+*	*Ca^{2+}*
EC	1															
TDS	.991	1														
pH	.202	.263	1													
S.S	-.288	-.292	.275	1												
COD	-.573	-.597	-.638	.032	1											
DO	.227	.320	.326	.038	-.436	1										
Cl^-	-.409	-.421	.038	.143	-.326	-.115	1									
TH	-.390	-.347	-.196	.028	.324	.597	-.221	1								
Mg^{2+}	-.240	-.189	-.278	-.369	.203	.598	.002	.750	1							
PO_4^{3-}	-.521	-.520	-.522	-.254	.389	-.328	.439	.214	.394	1						
SO_4^{2-}	.130	.102	.219	.559	.194	-.199	-.606	-.024	-.538	-.599	1					
NO_3^-	.470	.446	-.184	-.508	-.112	.041	-.281	-.303	-.121	-.498	-.030	1				
F^-	.213	.225	-.248	-.716	-.269	-.017	.322	-.124	.452	.587	-.840	.026	1			
Na^+	.416	.414	.163	-.767	-.350	.011	-.019	-.264	.053	-.143	-.464	.563	.475	1		
K^+	.671	.614	.130	-.415	-.252	-.294	-.408	-.308	-.293	-.336	.140	.337	.142	.667	1	
Ca^{2+}	.386	.390	.214	-.744	-.337	.050	-.064	-.186	.073	-.153	-.423	.506	.425	.992	.679	1

Table 2: Total variance explained

Total Variance Explained			
Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	5.820	36.376	36.376
2	3.975	24.843	61.219
3	2.421	15.133	76.351
4	1.801	11.256	87.608
5	1.017	6.356	93.964
6	.750	4.689	98.653
7	.216	1.347	100.000
8	5.297E-16	3.310E-15	100.000
9	2.569E-16	1.606E-15	100.000
10	1.517E-16	9.484E-16	100.000
11	6.634E-17	4.146E-16	100.000
12	-1.993E-17	-1.245E-16	100.000
13	-1.897E-16	-1.185E-15	100.000
14	-1.944E-16	-1.215E-15	100.000
15	-2.901 E-16	-1.813E-15	100.000
16	-1.425E-15	-8.906E-15	100.000

Extraction Method: Principal Component Analysis

Table 3: Initial and extraction communalities

Communalities		
	Initial	Extraction
Conductivity	1.000	.938
TDS	1.000	.903
pH	1.000	.975
S.S	1.000	.920
COD	1.000	.974
DO	1.000	.995
Chloride	1.000	.782
Total Hardness	1.000	.996
Mg Hardness	1.000	.991
Phosphate	1.000	.877
Sulphate	1.000	.962
Nitrate	1.000	.862
Fluoride	1.000	.940
Sodium	1.000	.972
Potassium	1.000	.966
Calcium	1.000	.983

Extraction Method: Principal Component Analysis

first component will not be influenced by the second component and vice versa.

In general, components loading more than 0.45 can be taken into account when analysing, in other words, the most substantial variables throughout the components are high loads, which have been taken into account when calculating the components (Mazlum et al., 1996).

The first PC explains 36.38% of the cumulative variance with an eigenvalue of 5.82. It has a positive correlation (indicates that an increase in one variable would increase other variables) with variables such as fluoride, sodium (Na^+), potassium (K^+), calcium (Ca^+)

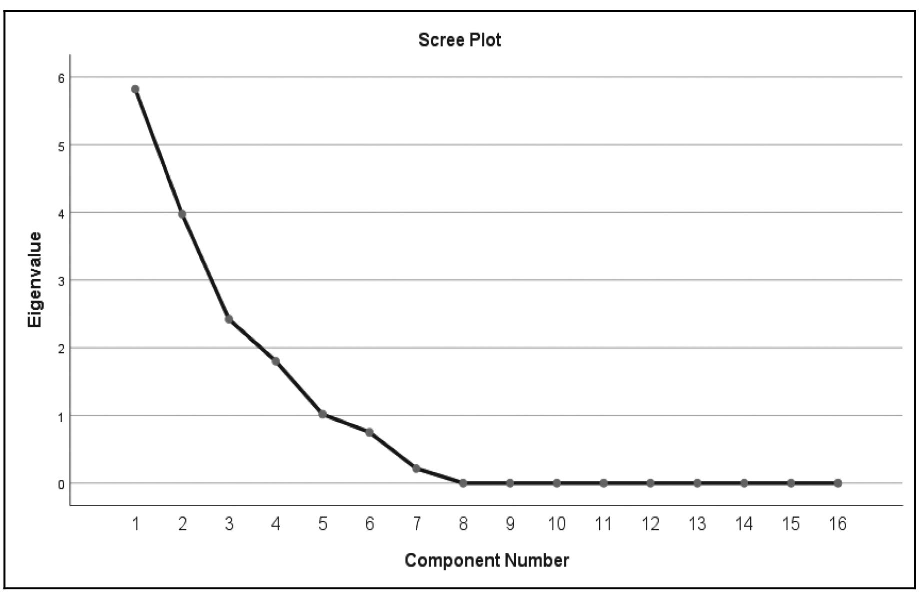


Figure 3: Scree plot of the eigenvalue with each component.

Table 4: Component matrix

Component Matrix[a]					
	Component				
	1	*2*	*3*	*4*	*5*
Conductivity	.823	-.370	.133	.084	-.314
TDS	.796	-.373	.235	.034	-.273
pH	.268	-.622	.414	.192	.554
S.S	-.669	-.610	.108	.297	.010
COD	-.620	.350	-.412	-.538	.082
DO	.085	.042	.986	-.013	-.119
Chloride	-.017	.127	-.127	.855	-.136
Total Hardness	-.459	.603	.629	-.117	.113
Mg Hardness	-.180	.765	.596	-.122	.060
Phosphate	-.313	.765	-.333	.286	-.018
Sulphate	-.485	-.805	-.106	-.243	.095
Nitrate	.603	-.058	-.061	-.565	-.414
Fluoride	.578	.695	-.098	.317	-.112
Sodium	.906	.246	-.053	-.157	.252
Potassium	.852	.022	-.344	.083	.338
Calcium	.882	.262	-.003	-.162	.332

Extraction Method: Principal Component Analysis.
[a]: 5 components extracted

Table 5: Rotated component matrix

Rotated Component Matrix[a]					
	Component				
	1	*2*	*3*	*4*	*5*
Conductivity	.321	.888	-.205	.063	-.004
TDS	.302	.884	-.113	.118	-.063
pH	.004	.318	-.029	.933	-.046
S.S	-.868	-.067	-.129	.358	.132
COD	-.163	-.785	-.037	-.410	-.402
DO	-.100	.465	.845	.226	-.062
Chloride	-.065	.048	-.100	-.026	.875
Total Hardness	-.101	-.382	.915	-.054	-.010
Mg Hardness	.184	-.270	.925	-.166	.029
Phosphate	.152	-.615	.133	-.463	.493
Sulphate	-.705	-.051	-.401	.343	-.429
Nitrate	.368	.529	-.129	-.389	-.528
Fluoride	.740	.096	.153	-.349	.488
Sodium	.936	.267	-.059	.088	-.115
Potassium	.831	.204	-.428	.213	.074
Calcium	.945	.220	-.006	.156	-.126

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.3
[a]: Rotation converged in 6 iterations

whereas sulphate and suspended solids show negative contribution (indicates that an increase in one variable will result in a decrease in another variable) to this variance. This may be due to point source such as industrial discharge because it is used in the production of dyes, glass, paper, chemicals, fertilisers and also used in metal and plating industries. Oxidation of biological material occurs under anaerobic conditions in which bacteria predominate resulting in the formation of hydrogen sulphide. Hydrogen sulphide forms when anoxic environment prevails in deep water above the sediments, and water is sufficiently acidic enough to precipitate the ions present.

The second PC explains the cumulative variance of 61.22% with an eigenvalue of 3.975, which is positively correlated with conductivity, TDS and nitrate whereas phosphate and COD are negatively correlated. Therefore, more concentration of dissolved organic matter will result in consumption of more amount of oxygen, which results in the decreased availability of dissolved oxygen, which undergoes anaerobic fermentation process leading to ortho-phosphate and organic acid. Phosphate is an important part of plant life but may exacerbate eutrophication if its concentration rises in water. The concentration of phosphorus rises with depth due to the degradation of biological content and the regeneration of the sediments under an anoxic environment. Phosphate are part of effluents obtained from industries like laundering and commercial cleaning fluids and municipalities, organic wastes in sewage. Soil erosion during strong wind is another source of phosphate pollution.

The third PC explains a cumulative variance of 76.35% with an eigenvalue of 2.421, which contains variables like DO, total hardness and magnesium hardness. All relationships between the variables show a positive correlation and represents the physicochemical source of variability.

Conclusions

Sanjay Lake has been a breeding ground for dengue causing mosquitoes (cseindia.org, 2015). Moreover, the lake has been reducing in size over the past years, however, a plan for revival for the lake has been laid out, which is being implemented since 2015. The present study assessed the ability of PCA to determine the most important indicator in the water quality analysis, environmental factors and the contamination sources responsible.

The PCA approach shows that the physicochemical source of variability, the municipal sewage waste from nearby residential areas and the untreated industrial effluent discharges from neighbouring small-scale industries are mainly affecting the water quality. If the effluent flowing into the lake is handled with the aid of a sewage treatment plant (STP) and also the flow of industrial effluent is prohibited, then the water quality can be significantly improved, hence the lake will be protected. It might enable the use of lake water for various industrial, agricultural and recreational activities.

References

APHA (2005). Standard methods for the examination of water and wastewater, 21st edition. American Public Health Association Washington, DC.

De Bartolomeo, A., Sanchini, G., Poletti, L., Sebastiani, B. and G. Morozzi (2004). Relationship among parameters of lake polluted sediments evaluated by multivariate statistical analysis. *Chemosphere*, **55:** 1323–1329.

Singh, D.S. (2015). Water Quality Index Assessment Of Bhalswa Lake, New Delhi. *International Journal of Advanced Research*, **3:** 1052–1059

Jain, A.K. (2009). Low Carbon City: Policy, Planning and Practice. ISBN: 9788183564687.

Jhamnani, B. and S.K. Singh (2009). Groundwater contamination due to Bhalaswa Landfill site in New Delhi. *International Journal of Environmental Engineering*, **3(3):** 181–185.

Kazi, T.G., Afridi, H.I., Arain, M.B., Jamali, M.K., Sarfraz, R.A., Baig, J.A. and A.Q. Shah (2009). Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. *Ecotoxicology and Environmental Safety*, **72:** 301–309.

Kim, J.H., Choi, C.M., Kim, S.B. and S.K. Kwun (2009). Water quality monitoring and multivariate statistical analysis for rural streams in South Korea. *Paddy and Water Environment*, **7:** 197–208.

Kishor, K. and B.D.D. Joshi (2005). Physico-chemical characteristics of pond water at Khanpur village in Bareilly district (U.P.). *Himalyan Journal of Environment and Zoology*, **19(8):** 89–92.

Mazlum, N., Ozer, A. and S. Mazlum (1996). Interpretation of water quality by principal component analysis. *Journal of Engineering and Environmental Science*, **23:** 19–26.

Mohan, I. (2007). Environment Education. ISBN: 9788131302828

Reporter (2015). cseindia.org/sanjay-lake-dying-silently-716

Singh, S.K. and L. Kumar (2015). Removal of chloride from groundwater by bio adsorption. *International Journal of Advance Research*, **3(5):** 140–154.

Singh, K.P., Malik, A. and S. Sinha (2005). Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: A case study. *Analytica Chimica Acta*, **538:** 355–374.

Voutsa, D., Manoli, E., Samara, C., Sofoniou, M. and I. Stratis (2001). A study of surface water quality in Macedonia, Greece: Speciation of nitrogen and phosphorus. *Water, Air, and Soil Pollution*, **129:** 13–32.

Vyas, A., Mishra, D., Bajapai, A., Dixit, S. and N. Verma (2006). Environment impact of idol immersion activity lakes of Bhopal, India. *Asian Journal of Experimental Sciences*, **20(2):** 289–296.

Wenning, R.J. and G.A. Erickson (2014). Interpretation and analysis of complex environmental data using chemometric methods. *Trends in Anaytical. Chemistry*, **13:** 446–457.