Eco. Env. & Cons. 30 (May Suppl. Issue) : 2024; pp. (S305-S311)
Copyright@ EM International
ISSN 0971–765X

DOI No.: http://doi.org/10.53550/EEC.2024.v30i03s.053

Interpretation of Water Quality Using Principal Component Analysis – A Case Study from A Tropical Ramsar Wetland Site Adjacent to the Seafood Processing Facilities

V. Vidya*[1] and G. Prasad[2]

[1]Post Graduate Department of Zoology and Research Centre, Sanatana Dharma College, Alappuzha, University of Kerala, India
[2]Department of Zoology, University of Kerala, Kariavattom Campus, Thiruvananthapuram 695 581, Kerala, India

(Received 8 September, 2023; Accepted 3 November, 2023)

ABSTRACT

Cluster analysis and PCA were applied for the evaluation of water quality data of vembanad wetland adjacent to the seafood processing facilities. The water samples were collected from ten different stations for a period of two years (2010-2012) on a monthly basis. The stations, S_1-S_9 were closely associated with the seafood processing discharge outlets and the S_{10}, was kept as a reference site, which is free from the seafood processing discharge. The physico-chemical parameters (atmospheric and water temperature, TDS, pH, EC, salinity, DO, BOD, alkalinity, COD, hardness and nutrients such as nitrate, phosphate, ammonia and silica) were studied. Hierarchical cluster analysis was performed and dendrogram was generated, which grouped all 10 sampling sites into three statistically significant clusters. Thus, the water quality around the site may be categorized as relatively less polluted, moderately polluted and highly polluted. PCA could extract 6 principal components and they account for 65.79% of the total variance of the original data. The parameters that contributed to the first component include alkalinity, free CO_2, BOD, COD and nutrients. The organic pollution and dumping of waste from the seafood industry altered the water quality significantly. This study illustrates the benefit of multivariate statistical techniques for analyzing and interpretation of complex data sets, and to plan for future studies.

Key words : *Tropical Ramsar wetland, Principle component analysis, Pollution from seafood processing plant*

Introduction

The water contamination is a burning issue and its consequences are gaining attention worldwide. Anthropogenic intervention is rapidly degrading the surface water quality. Urban and industrial effluents are considered as being major sources of chemical and nutrients to aquatic ecosystem. The concentration of toxic chemicals and biologically available nutrients in excess can lead to diverse problems such as toxic algal blooms, loss of oxygen, fish kill, loss of biodiversity and loss of aquatic plants beds and coral reefs (Voutsa *et al*., 2001).

The quality of water is identify by its physical,

(*Assistant Professor, [2]Professor and Head)
ORCI D- Vidya Vijayakumari http://orcid.org/0000-0002-0352-8308

chemical and biological properties; associated with analysis the large number of measured variables (Bayacioglu, 2006). Long term surveys and monitoring programs of water quality are an adequate approach to a better knowledge of river hydrochemistry and pollution, but they produce large sets of data which are often difficult to interpret (Dixon and Chiswell, 1996). Multivariate statistical techniques are powerful tools for analysing large numbers of samples collected in surveys, classifying assemblages and assessing human impacts on ecosystem conditions (Gauch 1982; Reynoldson *et al.*, 2001; Cao *et al.*, 2002; Iyer *et al.*, 2003). Several studies have used factor analysis to identify primary sources of contamination (Reisenhofer *et al.*, 1995; Borovec 1996; Vega *et al.*, 1998; Singh *et al.*, 2005) and to find associations between parameters so that the number of measured parameters can be reduced (Parinet *et al.*, 2004; Zeng and Rasmussen, 2005).

The application of different multivariate statistical techniques, such as cluster analysis (CA) and principal component analysis (PCA) helps in the interpretation of complex data matrices to better understand the water quality and ecological status of the studied systems, allows the identification of possible factors that influence water environment systems and offers a valuable tool for reliable management of water resources (Simeonov *et al.*, 2003; Praus, 2005; Shrestha and Kazama, 2007). Zhou *et al.* (2007) and Boyacioglu and Boyacioglu (2007) use the PCA and CA to classify the sampling sites and to identify the latent pollution source. Mendiguchía *et al.* (2004) used the CA to divide a watershed to four zones with different water quality except using PCA to identify the main pollutants.

In India, Iyer *et al.* (2003) constructed a statistical, model, which based on the PCA for coastal water quality data from the Cochin coast in southwest India, which explain the relationships between the various physicochemical variables that have been monitored and environmental conditions effect on the coastal water quality. In this study principal components analysis (PCA) are used in order to interpret and grouping the water quality parameter.

In India especially in Kerala, the environmental problems associated with the discharge of wastewater from seafood processing facilities are gaining attention only recently. The processing of seafood products can require significant volume of water, which has a direct consequence in the generation of contaminated effluents. The seafood industries in Kerala are particularly limited to the narrow coastline areas of the Vembanad Lake aremainly due to the ease of transport. All the processing plants discharge their effluents together with solid waste into this lake. The volume of effluent discharged into this wetland system is unknown and no routine monitoring is done to check the uncontrolled discharge of the seafood processing waste. In the present study, CA was used to identify several zones with different water quality and PCA was done to understand the most important factors influencing the water pollution.

Materials and Methods

Study Site

The Vembanad wetland system and its associated drainage basins lie in the humid tropical region between 09˄00′ -10˄40′N and 76˄00′-77˄30′E. The Vembanad wetland system covers an area of over 2033.02 km² and extends between 09°00'-10°40' N latitude and 76°00'-7°30' E longitude, is the largest and the most important tropical wetland in Kerala. The lake is bordered by Alappuzha, Kottayam and Ernakulam. The Vembanad Lake is approximately 14 kilometres wide at its widest point. The lake is a part of Vembanad-Kol wetland system which extends from Alappuzha in the south to Azheekkode in the north, making it by far, India's longest lake at just over 96.5 km in length. The lake is fed by 10 rivers flowing into it including the six major rivers of central Kerala namely the Achenkovil, Manimala, Meenachil, Muvattupuzha, Pamba and Periyar. The total area drained by the lake is 15,770 km², which accounts for 40% of the area of Kerala. Its annual surface runoff of 21,900 mm accounts for almost 30% of the total surface water resource of the state.

The present study conducted in Cherthala-Aroor-Edakochi coastal belt, where most seafood processing plants are functioning. The study was conducted for two years (October 2010-September 2012) from ten preselected stations (S_1 - S_{10}) on monthly basis. They are Pattanakadu (S_1), Parayakadu (S_2), Shankaranthodu (S_3), Konkeri bridge (S_4), Chandiroor (S_5), Edakochi I (S_6), Edakochi II (S_7), Aroormukkam (S_8), Aroorkutti bridge (S_9), Panavally (S_{10}). Nine of the selected stations near the effluent outlet of the seafood processing plants and one kept as the reference site (S_{10}) which is free from seafood effluent discharge (Table 1 and Fig.1).

Table 1. Characteristics of study sites

Sl. No	Name	Geographic position	Source of pollution	Land	Colour utilisation	Remarks
1	Pattanakadu	9°44'22N, 76°19'07E	Discharge from Sea food processing industry	Panchayath area/ Residential	Turbid water	Interconnected canal
2	Parayakadu	9°47'13N,76°18'20E	Discharge from Sea food processing industry	Aquaculture farm	Turbid water	Interconnected canal
3	Shankaranthodu	9°47'22N,76°18'12E	Discharge from Sea food processing industry	Panchayatharea/ Residential	Heavily turbid water	Interconnected canal
4	Konkeri Bridge	9°49'22N,76°18'22E	Discharge from Sea food processing industry	Panchayath area/ Residential	Turbid water	Interconnected canal
5	Chandiroor	9°50'28N,76°18'29E	Discharge from Sea food processing industry	Panchayath area/ Residential	Heavily turbid water	Interconnected canal
6	Edakochi I	9°54'47N,76°17'06E	Discharge from Sea food processing industry	Panchayath area/ Residential	Heavily turbid water	Main water body
7	Edakochi II	9°54'07N,76°17'42E	Discharge from Sea food processing industry	Boat building yard	Heavily turbid water	Main water body
8	Aroormukkam	9°53'23N,76°17'49E	Discharge from Sea food processing industry	Fallow land	Turbid water	Main water body
9	Aroorkutti	9°52'12N,76°19'01E	Discharge from Sea food processing industry	Panchayath area/ Residential	Heavily turbid water	Main water body
10	Panavally/ (Reference site)	9°49'13 N, 76°21'33E	Free from Sea food processing waste discharge	Agriculture and Aquaculture	Clear water	Main water body

The first five stations are in interconnected channels and the remaining including the reference station are in the main water body. The collections were made on a monthly basis in all stations.

Statistical analysis

Multivariate statistics such as principal component analysis and cluster analysis were employed to understand the principle components responsible for variance and group the stations based on their similarities.

Results and Discussion

The determined initial Principal Component (PC) and its eigen value and percent of variance contributed in each PC were represented at table 80. In this present study, the PCA could extract 6 principal components (eigen value>1) and the percentages of the total variances of the 6 extracted components are accumulated, these six principal components account for 65.79% of the total variance of the original data (Table 2 and 3). This suggested that the 6 extracted components has been the reason for the majority of the variance.

Table 2. Principal component analysis

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.803
Bartlett's Test of Sphericity	Approx. Chi-Square	1742.14
	df	190
	Sig.	.000

The PC1 accounts for 21.97% of the total variance. The parameters that contributed to the first component includes alkalinity (0.649), free CO2 (0.744), BOD (0.732), COD (0.679) and nutrients like phosphate (0.757), nitrate (0.432), ammonia (0.759) and silica (0.653) (Table 4). Among these the parameters alkalinity, free CO2, BOD, COD, phosphate, silica and ammonia showed moderate factor loadings while nitrate showed weak factor loadings. The PC2 contributed 18.10% of the total variance. The parameters included are TDS (0.816), EC (0.891), salinity (0.834) and hardness (0.628). Among these TDS, EC and salinity showed strong factor loadings while hardness showed moderate factor loadings. The PC3 con-

Fig. 1. Location Map

tributed 8.30% of the total variance and the parameters included are pH (-0.534), transparency (-0.670) and DO (-0.732). Here the parameters showed moderate negative factor loadings. In PC4, water temperature (0.847) is the only parameter contributed 6.64% of the total variance and it showed strong factor loadings. In PC5 biological components contributed 5.73% of the total variance. The phytoplankton (0.723) and zooplankton (0.806) showed strong factor loadings. The light intensity came under the PC6 and it contributed 5.02% of the total variance and it showed (0.776) strong factor loadings.

PCA was done to understand the influence of the effluent discharge from seafood processing industries onto the nearby wetlands. In the present study, the PCA could extract 6 principal components (eigen value>1) and the percentages of the total variances of the 6 extracted components are accumulated and these six principal components account for 65.79% of the total variance of the original data. This suggested that the 6 extracted components have been the reason for the majority of the variance. Component loadings (correlation coefficients), which measure the degree of closeness between the variables and the PC, the largest loading either positive or negative, suggests the meaning of the dimen-

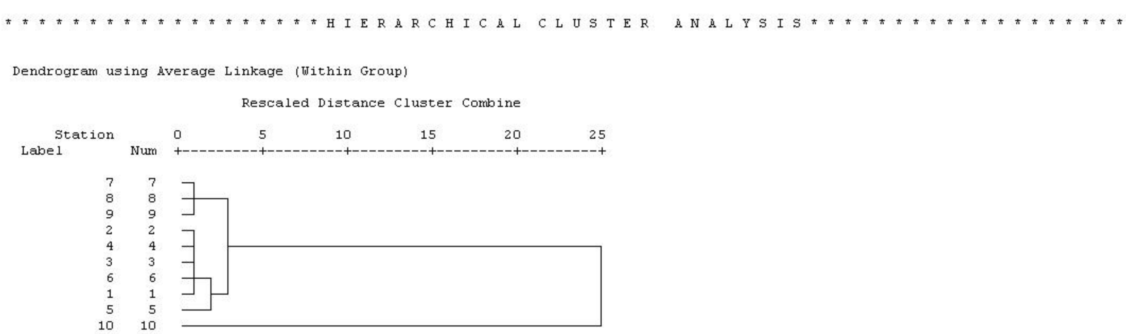

Fig. 2. The clustering of stations using Hierarchical cluster analysis

sions; positive loading indicate that the contribution of the variables increases with the increasing load in dimension; and negative loading indicates a decrease (Lawrence and Upchruch, 1982). According to Mazlum *et al*. (1999), the most significant variables in the components are represented by high loadings, which contributed primarily to the total variance.

The PC1 accounts for 21.97% of the total variance and was loaded primarily with phosphate (0.757), free CO2 (0.744), BOD (0.732), COD (0.679), silica (0.653), alkalinity (0.649), and nitrate (0.432). PC1 indicates pollutions of anthropogenic origin. The incorporation of COD and BOD as the first component implies that both organic and inorganic pollution which was due to anthropogenic intervention. In addition to it, the high loadings of nutrients represents pollution from point sources such as discharges from wastewater treatment plants, domestic waste water, agricultural activities and industrial effluents (Shrestha and Kazama, 2007; Juahir *et al.*, 2011). The PC2 contributed 18.10% of the total variance and was loaded by TDS (0.816), EC (0.891), salinity (0.834) and hardness (0.628). These factors also indicate pollutants of anthropogenic origin. The high contribution of salinity, TDS and hardness in PC2 points out the influence of seafood waste discharge directly into these water bodies. Fish processing industries use salt for the preservation of seafood products and hence the chloride content is increased in discharged water (Sankpal and Naikwade, 2012; Vaghela *et al.*, 2015). Moreover, hardness in water is due to the CO_2 released in bacterial action and by metabolic ions dissolved in

Table 3. Total variance explained in PCA

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.395	21.973	21.973	4.395	21.973	21.973
2	3.621	18.107	40.080	3.621	18.107	40.080
3	1.662	8.308	48.388	1.662	8.308	48.388
4	1.328	6.640	55.028	1.328	6.640	55.028
5	1.146	5.732	60.759	1.146	5.732	60.759
6	1.005	5.026	65.785	1.005	5.026	65.785

Table 4. Rotated Component Matrix of PCA

	Component					
	1	2	3	4	5	6
Air temperature	-.020	.196	-.031	.825	-.006	-.099
Water temperature	.056	.090	.110	.847	-.071	.017
Ph	.044	-.270	-.534	.111	.076	.533
TDS (ppt)	-.129	0.816	.215	.113	-.076	.038
EC (ms)	-.011	0.891	.090	.210	.022	-.015
Salinity (ppt)	-.059	0.834	.031	-.068	-.020	-.073
Turbidity (ppt)	-.305	-.057	-.670	-.131	.001	-.114
Light intensity (klux)	-.001	.119	.278	-.124	-.032	.776
Free CO2 (mg/l)	.744	-.259	.182	-.071	-.059	-.112
Alkalinity(mg/l)	.649	.042	.383	.138	.207	.005
Hardness (mg/l)	.036	0.628	.204	.385	.062	.266
DO (mg/l)	-.124	-.257	-0.732	.010	-.102	-.133
BOD (mg/l)	0.732	.164	-.123	-.035	.005	.108
COD (mg/l)	0.679	.026	.231	.085	.111	.215
Phosphate (mg/l)	0.757	-.087	.232	-.025	-.165	-.091
Nitrate (mg/l)	0.432	.106	.320	.214	.089	.167
Ammonia (mg/l)	0.759	-.068	.025	-.155	.054	-.179
Silica (mg/l)	0.653	-.245	-.135	.217	.221	.096
Phytoplankton (ml)	.142	.063	.146	-.233	.723	-.187
Zooplankton (ml)	.007	-.081	-.027	.101	.806	.144

wastewater discharged from the industries (Chauhan and Sagar, 2013). PC1 and PC2 have the largest proportion of the total variance. The PC3 contributed 8.30% of the total variance and the parameters included are pH (-0.534), transparency (-0.670) and DO (-0.732). Here the parameters showed negative factor loadings. The negative loading indicates that the contribution of the variables decreases with the increasing load in dimension. This implies that the discharge from the seafood industries reduced the DO and transparency of the water bodies and made them more acidic. In PC4, water temperature (0.847) is the only parameter contributed 6.64% of the total variance. In PC5 biological components contributed 5.73% of the total variance. The phytoplankton (0.723) and zooplankton (0.806) showed strong factor loadings. The dumping of waste from the seafood industry altered the water quality, which favoured the growth of pollution indicators among phytoplankton and zooplankton. The light intensity came under the PC6 and it contributed 5.02% of the total variance and it showed (0.776) strong factor loadings. Thus, PCA identifies the important parameters responsible for the majority of the variance and hence avoiding the monitoring of the insignificant water quality parameters. This technique helps to focus on the core of the problem thus saving sufficient energy and money.

The sampling station classification was performed by the use of cluster analysis and dendrogram was generated, which grouped all 10 sampling sites into three statistically significant clusters (Fig. 2 and 3). They are:

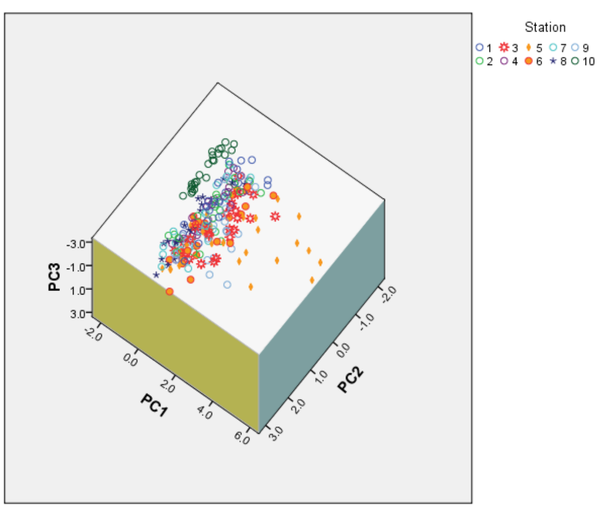

Fig. 3. Grouping of stations by cluster analysis

Cluster I: It includes station 7, station 8 and station 9.

Cluster II: The stations 1, 2, 3, 4, 5 and 6 constituted the second cluster.

Cluster III: It is seen from the dendrogram that cluster III is characterized by the biggest Euclidean distance to the other clusters (high significance of clustering). The site 10 comprised the third cluster.

In the study, hierarchical agglomerative CA was performed based on the normalized data set (mean of observations over the whole period) by means of the Ward's method using squared Euclidean distances as a measure of similarity (Zhang *et al*., 2009). The spatial variability of water quality in the whole river basin was determined from CA, which divides a large number of objects into a smaller number of homogenous groups based on their internal correlations (Hussain *et al*., 2008). CA grouped the sampling stations into three clusters of similar water quality characteristics. The cluster analysis revealed that the reference station is different from the other nine stations in terms of the water quality. The release of waste from the seafood industry deteriorated the water quality in nine stations (S_1-S_9) when compared to the reference station (S_{10}).

The waste discharge from the seafood processing industry is a major reason for the alarming rate of organic pollution and eutrophication. If the condition persists, water quality will degrade subsequently, questioning the existence of aquatic organisms including fishes. This will adversely affect the seafood industry.

Acknowledgements

The financial aid for the present study was obtained from the Directorate of Environment and Climate Change, Government of Kerala, which is greatly acknowledged. The author is also thankful to the UGC for providing a BSR/RFSMS fellowship for pursuing the work.

References

Bayacioglu, H. 2006. Surface Water Quality Assessment Using Factor Analysis. *Water SA*. 32(3): 389-394.

Borovec, Z. 1996. Evaluation of the Concentrations of Trace Elements in Stream Sediments by Factor and Cluster Analysis and the Sequential Extraction Procedure. *Sci. Total Environ.* 177: 237-250.

Boyacioglu, H. and Boyacioglu, H. 2007. Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin. *Turkish Environ. Geol.* 54: 275-282.

Cao, Y., Larsen, D.P., Hughes, R.M., Angermeier, P.L. and Patton, T.M. 2002. Sampling Effort Affects Multivariate Comparisons of Stream Assemblages. *J. North Am. Benthol. Soc.* 21 (4): 701–714.

Chauhan, B.S. and Sagar, S.K. 2013. Impact of Pollutants on Water Quality of River Sutlej in Nangal Area of Punjab India. *Biol. Forum Int. J.* 5 (1): 113-123.

Dixon, W. and Chiswell, B. 1996. Review of aquatic monitoring program design. *Water Research.* 30(9): 1935-1948.

Gauch, H.G. 1982. *Multivariate Analysis in Community Ecology*. Cambridge University Press, Cambridge. 307 p.https://doi.org/10.1017/CBO9780511623332

Hussain, M., Ahmed, S.M. and Abderrahman, W. 2008. Cluster analysis and quality assessment of logged water at an irrigation project, eastern Saudi Arabia. *J. Environ. Manage*. 86(1): 297-307.

Iyer, C.S., Sindhu, M., Kulkarni, S.G., Tambe, S.S. and Kulkarni, B.D. 2003. Statistical analysis of the physico-chemical data on the coastal water of Cochin. *J. Environ. Monit*. 5: 324-327.

Juahir, H., Zain, M.S., Yusoff, M.K., Hanidza, T.I.T., Armi, A.S.M., Toriman, M.E. and Mokhtar, M. 2011 Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. Environ. *Monit. Assess.* 173(1-4): 625-641.

Lawrence, F.W. and Upchruch, S.B. 1982. Identification of recharge areas using geochemical factor analysis. *Groundw*. 20: 680-687.

Mazlum, N., Ozer, A. and Mazlum, S. 1999. Interpretation of water quality data by principal component analysis. *J. Environ. Eng. Sci*. 23: 19-26.

Mendiguchýìa, C., Moreno,C., Galindo-Riaño, M.D., Garcýìa-Vargas, M. 2004. Using chemometric tools to assess anthropogenic effects in river water: A case study: Guadalquivir River (Spain). *Analytica Chimica Acta.* 515 (1) : 143-149.

Parinet, B., Lhote, A. and Legube, B. 2004. Principal Component Analysis: *An Appropriate Tool for Water Quality Evaluation and Management – Application to a Tropical Lake System*. *Ecol. Modell.* 178 (3–4): 295–311.

Reynoldson, T.B., Rosenberg, D.M. and Resh, V.H. 2001. Comparison of Models Predicting Invertebrate Assemblages for Biomonitoring in the Fraser River Catchment, British Columbia. *Can. J. Fish. Aquat. Sci.* 58: 1395–1410.

Sankpal, T.S. and Naikwade, P.V. 2012. Physicochemical analysis of effluent discharge of fish processing industries in Ratnagiri India. *Biosci. Discov*. 3(1): 107-111.

Shrestha, S. and Kazama, F. 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. Environ. *Modell. Softw.* 22(4): 464-475.

Shrestha, S. and Kazama, F. 2007. Assessment of surface water quality using multi-variate statistical techniques: A case study of the Fuji River Basin, Japan. *Environ Model Softw*. 22 (4): 464-475.

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. and Kouimtzis, Th. 2003. *Assessment of the surface water quality in northern Greece*. *Water Res*. 37(17): 4119-4124.

Vaghela, D.T., Krishnakumar, P.K. and Dar, S.A. 2015. Physico-chemical characteristic of fish processing industry effluent discharge into Veraval harbor waters Gujarat India. *Ecol. Environ. Conserv*. 21(1): 575-580.

Vega, M., Padro, R., Barrado, E. and Deban, L. 1998. Assessment of Seasonal and Polluting Effects on the Quality of the River Water by Exploratory Data Analysis. *Water Res.* 32 (12): 3581–3592.

Voutsa, D., Manoli, E., Samara, C., Sofoniou, M. and Stratis, I. 2001. A Study of Surface Water Quality in Macedonia, Greece: Speciation of Nitrogen and Phosphorus. *Water Air Soil Pollut.* 129: 13–32.

Zeng, X. and Rasmussen, T.C. 2005. Multivariate Statistical Characterization of Water Quality in Lake Lanier, Georgia, USA. *J. Environ. Quality*. 34: 1980–1991.

Zhang, M., Cui, L., Sheng, L. and Wang, Y. 2009. Distribution and enrichment of heavy metals among sediments, water body and plants in Hengshuihu wetland of North China. *Ecol. Eng*. 35(4): 563-569.

Zhou, F., Liu,Y. and Guo, H. 2007. Application of multivariate statistical methods to water quality assessment of the water courses in north western new territories Hong Kong. Environ. *Monit. Assess.* 132 : 1-13.