Environment Conservation Journal 14(3)9-16 , 2013
ISSN 0972-3099 (Print) 2278-5124 (Online)
Abstracted and Indexed

Source apportionment and quality assessment of surface water using principal component analysis and multiple linear regression statistics

Mohammed Amjed Hossain, Nasly Binti Mohamed Ali✉ and Mir Sujaul Islam

Received: 28.08.2013 **Revised: 25.09.2013** **Accepted: 10.10.2013**

Abstract

Principal component analysis (PCA) and multiple linear regressions (MLR) analysis were applied on the data set of surface water quality for source identification of pollution and their contribution on the variation of water quality. Results revealed that, most of the water quality parameters were found to be toxic compare to the national standard of Malaysia. PCA identified the sources as, ionic groups of salts, soil erosion and agricultural runoff, organic and nutrient pollutions from domestic wastewater, industrial sewage and wastewater treatment plants. MLR investigated the contribution of every variable with R= 0.968 and R^2=0.934 and it was highly significant (p<0.01).

Keywords: *Multiple linear regressions*, *principal component analysis*, *source apportionment*, *varimax rotation; water pollution*, *water quality.*

Introduction

Surface water pollution is a major concern in the present world. Population growth, urbanization and incremental commercial and industrial activities exert tremendous pressure on the existing surface waters and through disposal of pollutants. Rivers which are the main reservoir of surface water, serve as the recipients of excessive amount of wastes generated and discharged vide anthropogenic activities all over the world (Khanna *et al*., 2011; Milovanovic, 2007). Anthropogenic activities have negatively influenced water quality and are the major source of pollution for water quality deterioration (Satheeshkumar & Khan, 2012). Water pollution is ever-increasing with industrial and commercial development, further causing deterioration of water quality and threatening human health, aquatic ecosystem, economic and social welfare (Milovanovic, 2007). Surface water management and control of pollution is a big challenge for the environmentalist. It is obvious that for better management and control of surface water pollution authentic and reliable information of its sources is required (Singh *et al*., 2005).

Author's Address
Faculty of Civil Engineering and Earth Resources
University Malaysia Pahang, Kuantan, Malaysia
Email: nasly.ump@gmail.com

Apportioning the sources of pollution of surface water can provide information to policy makers and environmentalist to set priorities for sustainable water resources management (Huang *et al*., 2010). Identification of sources of pollutants and the contribution of the parameters is one of the major challenges in assessing surface water quality (Mustapha & Abdu, 2012). Use of principal component analysis (PCA) in identifying pollution sources is now a widely used unbiased statistical method (Satheeshkumar & Khan, 2012) and eventually multiple linear regressions is useful in estimating the contribution of parameters; these two techniques can expose the potential pollution sources and those are practically usable in various types of data (Praveena *et al*., 2012). PCA helps in interpretation of complex data in the simplified way (Pejman *et al*., 2009) for better understanding by investigating the structural information of confusing data (Ragno *et al*., 2007). On the other hand, multiple linear regressions is the tool to examine the relationship between single dependent variable and a set of independent variables to best represent relationship in the each factor(Mustapha & Abdu, 2012).Shrestha & Kazama (2007); Huang *et al*.,(2010); and Juahir *et al*.,(2011) in their research they used PCA to identify water quality sources apportionment; they studied spatial variation of surface water and source of pollution and classified the water quality. In their study, Onojake *et al*. (2011) used PCA to identify the latent factor and found that Rivers in Delta State of Nigeria were heavily polluted due to anthropogenic source of pollution.Similarly, recent study conducted by Koklu *et al*. (2010) revealed that, multiple regressions analysis identified important parameters to determine the major sources of pollution. In this study PCA were employed to evaluate the surface water pollution sources and the multiple regression analysis was done to estimate the contribution of the significant parameters towards water quality variation.

Material and Methods

Study area and selection of monitoring stations: River Tunggak is situated in between 3°56′06″ to 3°59′44″ and 103°22′42″ to103°24′47″E adjacent to the Gebeng industrial estate (GIE) located in Pahang state with geographical coordinate 3°58′N and 103°26′E (Fig.1). Gebeng has rapid industrial development with increase the mass of effluents discharge (Nasly *et al*., 2013). The river Tunggak carries almost all effluents of GIE. It originates at the upper end of Gebeng and after passing through the industrial area it falls into South China Sea with another river Balok. Normal tides usually occur twice daily and the tidal water goes up to 3km upstream.

Fig. 1 Map of the study areas showing sampling stations

A total of 10 Monitoring stations were selected throughout the river basin based on the land use-pattern, point-sources of pollution, vegetation and river network. Global positioning system (GPS) was used to locate the stations.

Sampling and sample analysis

Water samples were collected monthly from pre-selected 10 stations for 12 months starting from February 2012-January 2013. Grab samples were collected at 10cm below the water level using 1000ml HDPE bottles and for BOD samples the dark BOD bottles (300 ml) were used. Collected samples were immediately preserved in ice-boxes under low temperature conditions and quickly transported to the laboratory maintaining standard methods (Andrew, 2005) and HACH (2005). In the laboratory samples was stored at 4° C until further analysis. Samples were analyzed to identify the status of physiochemical parameters. A total of 14 parameters were observed; of which temperature, pH, dissolved oxygen (DO), total dissolved solids (TDS), conductivity, salinity and turbidity were detected at the field condition (*in-situ*) during sampling using multi-parameters monitoring instrument YSI (YSI incorporated, Yellow Spring Ohio, USA). Five-day biochemical oxygen demand (BOD_5), chemical oxygen demand (COD), total suspended solids (TSS), nitrogen in the form of ammonia (NH_3) and nitrate (NO_3), phosphates (PO_4^{3-}) and sulphate (SO_4). All analysis was done according to Andrew (2005) and HACH (2005) following standard procedure. For BOD_5 determination, the DO concentration was read out before and after the incubation. BOD samples were incubated for 5 days at 20 ± 3^0C in BOD bottle and after the incubation period BOD_5 was calculated with the final reading of DO. COD determination was done in reactor digestion method using HACH spectrometer 5000. Nitrogen (NH_3) was measured as per nessler method; nitrate was estimated with cadmium reduction method, SO_4 measurement was done in sulfavar 4 method and PO_4^{3-} was determined in ascorbic acid method. In the measurement of all those parameters, calorimetric method was used (Andrew, 2005).

Principal Component Analysis (PCA)

Principal component analysis (PCA) was applied on the measured concentration of the physicochemical parameters of the Tunggak River. PCA is one of the best multivariate statistical techniques to extract linear relationships among a set of diverse variables (Mustapha & Abdu, 2012). It is the method that converts a set of possibly correlated variables into a set of linearly uncorrelated variables (principal components now called varifactor, VF). It can be used to reduce the number of variables and to explain the variance of large data set with a smaller set of variables (Wang *et al*., 2007) and also provides a better understanding of variables correlation and grouping (Ermakov *et al*., 2012). VFs are the linear combinations of actual data and the eigenvectors. In this study PCA were applied on the estimated data standardized through z-scale transformation, so that misclassification due to wide differences in data dimensionality can be avoided (Wang *et al*., 2007; Liu *et al*., 2003; Simeonov *et al*., 2003).The z-scale equation below is:

$$Zij= a_{i1}x_{1j}+a_{i2}x_{2j}+\text{------------}+a_{im}x_{mj} \quad (1)$$

Where, Z is the component score, a is the component loading, x is the measured value of a variable, i is the component number, j is the sample number, and m is the total number of variables. For all statistical analysis SPSS 16.0 software was used.

Multiple Linear Regressions

Multiple linear regressions analysis is a statistical method to predict the relationship between a dependent variable and a set of explanatory (independent) variables (several predictors) (Koklu *et al*., 2010). It can be used for both explanatory and predictive purposes within experimental design and also in non-experimental design (Mustapha & Abdu, 2012). In multiple linear regressions the relationship between the dependent variable and the explanatory variables can be expressed by the following equation:

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\text{-------------}+\beta_mX_m+\varepsilon \quad (2)$$

Where Y represents the dependent variable, X_1 …….X_m represent the several independent variables, β_0 is the constant term and β_1…….β_m represent the regression coefficient and εrepresent the random error.

Results and Discussion

Descriptive Statistics

A descriptive statistics of the surface water quality parameters are demonstrated in Table 1, which showed the range, minimum and maximum values, mean, standard deviation and variance of the physico-chemical parameters and also the Malaysian standard level of parameters in the study area.Statistics reveal that the concentration of conductivity, salinity, TDS, turbidity, NH_3-N, SO_4 and PO_4^{3-} were observed over the standard level of Malaysia (DOE, 2008) with higher mean values of 2538.80 µS/cm, 1.27%, 4242.40 mg/l, 17.45 mg/l, 1.66 mg/l, 135.5 mg/l and 1.24 mg/l respectively. This statistics ascertained that, those variables might have common source of origin (Mustapha & Abdu, 2012). However, the mean temperature (30.72^0C) was within the normal range ($27\text{-}31^0$C) (Saad *et al*., 2008) although at some stations it was observed maximum (38.45^0C). The maximum and minimum value of pH was recorded 9.12 and 4.15 respectively with a mean value of 6.70; which is almost in neutral level; even though some stations were highly acidic and some were highly alkaline. The mean concentration of DO (3.27 mg/l) was recorded very low compare to the standard level. It is likely due to discharge of industrial and domestic wastewater causing high organic pollution, and decrease in DO due to the decomposition of large OM(Satheeshkumar & Khan, 2012; Wang *et al.*, 2007). Again, at some station at industrial zone was found with high temperature because of less DO; as the inverse relationship between DO and temperature is a natural process (Wu *et al.*, 2010). Regarding the BOD_5 and COD values, the mean concentrations were 17.93 and 41.76 mg/l respectively; which were above the standard level. COD concentration was higher than BOD and both COD and BOD concentrations were higher than DO; this order of concentration reveal that anthropogenic pressure was associated with some natural pressure on the surface water quality (Mustapha & Abdu, 2012).

Table: 1. Descriptive statistics of the concentration of water quality parameters

WQ Parameters	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Malaysian standard (max)
BOD (mg/l)	38.25	0.10	38.35	17.93	12.02	144.51	1.00
COD (mg/l)	139.00	1.00	140.00	41.76	31.58	997.27	10.00
Conductivity (µS/cm)	2.71E4	16.00	27080.00	2.54E3	4756.67	2.26E7	1000.00
DO (mg/l)	6.55	0.55	7.10	3.27	1.67	2.78	7.00 (min)
NH_3-N (mg/l)	4.05	0.00	4.05	1.66	0.90	0.82	0.10
NO_3-N (mg/l)	4.50	0.00	4.50	0.34	0.71	0.51	7.00
pH	4.97	4.15	9.12	6.70	1.21	1.47	6.5-8.5
Phosphate (mg/l)	37.19	0.01	37.20	1.24	4.74	22.50	0.2
Salinity (%)	13.72	0.01	13.73	1.27	2.54	6.43	0.50
SO_4 (mg/l)	1220.00	0.00	1220.00	1.36E2	216.32	4.68E4	250
SS (mg/l)	74.00	1.00	75.00	15.46	12.915	166.80	25
TDS (mg/l)	7.46E4	7.70	74600.00	5.24E3	12096.05	1.46E8	500
Temp. (^{0}C)	13.09	25.36	38.45	30.72	2.49	6.20	27-31
Turbidity (NTU)	198.41	1.59	200.00	17.45	18.79	353.19	5

Principal component analysis

PCA was used to obtain composite variables VFs for identifying pollution factors that affect the water quality and latent pollution sources (Zhao *et al.*, 2011). The primary objective of this analysis was to create a new set of factor (VF) for reducing the contribution of less significant variables; that much smaller than the original data set in subsequent analysis (Mustapha & Abdu, 2012; Satheesh kumar & Khan, 2012). Before applying PCA, Kaisere-Meyere-Olkin (KMO) and Bartlett's test were performed to check the sampling adequacy. KMO test indicates the proportion of common variance and a value close to 1 denotes that PCA may be useful (Shrestha & Kazama, 2007). For KMO value there is a general thumb rule that it should be greater than 0.5 to precede a satisfactory PCA (Hinton *et al.*, 2004). In the present study KMO was found 0.663, indicating that the variables were correlated enough for appropriate PCA. Likewise, the Bartlett test of sphericity significant level was 0.00 ($p < 0.01$). In this study a total of 5 variance factors(VF) had been extracted by PCA based on Eigen value>1 and a varimax rotation were conducted to reduce the overlapping of genuine variables over every VF (Zhang *et al.*, 2011). Factor loading matrix is given in Table 2 and component plot of 14 variables are also shown in Fig. 2 The summary of the PCA result (Table 2) after rotation demonstrated parameters loadings, Eigen values, % variance of each component and cumulative variance. It can be seen that five (5) significance factors (VF) extracted by PCA with Eigen value>1 altogether explained 74.72% of total variance.

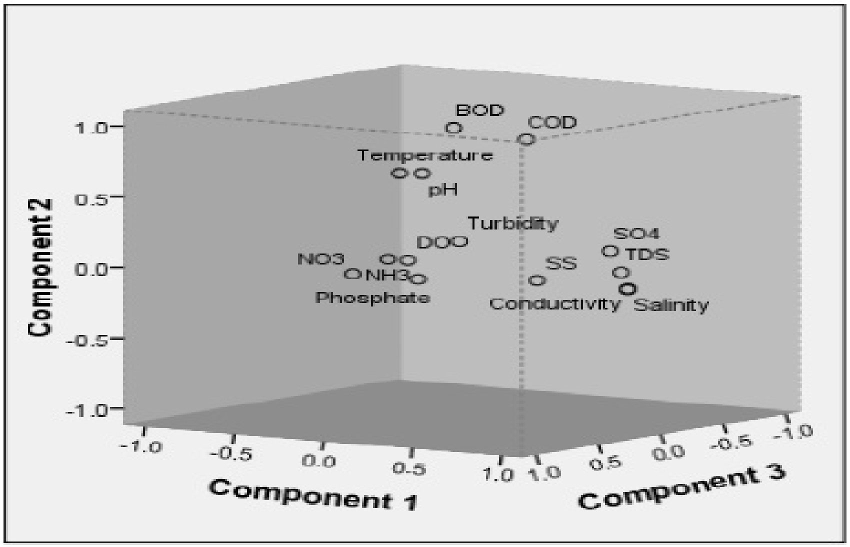

Fig. 2 Component plot of 14 variables in rotated space

The first factor (VF1) explained 26.832% of the total variance and showed the strong loading on conductivity, salinity, TDS and SO_4. This VF indicated that this factor group was highly and positively contributed from the natural sources related variables like ionic groups of salts in the basin from inflows, soil erosion and runoff (Varol *et al.*, 2012). This factor group also refers to as ionic pollution factor group which amounts a lots of ions and their compounds that leads to high loading of those variables (Zhang *et al.*, 2011). VF2 explained 18.49% of total variance and had strong positive loading of BOD_5, COD, temperature and pH. High loading of BOD_5 and COD represented organic pollutant from industrial sewage and domestic wastewater (Zheng *et al.*, 2008) and high loading of temperature and pH represented the physiochemical source of variables and also strongly influenced by climatologic and other environmental factors (Shrestha & Kazama, 2007; Wang *et al.*, 2007). VF3 had strong positive loading on NO_3-N and NH_3-N and explained 11.23% of total variance. This factor represented that source of these variables were attributed to the non-point source pollution from agricultural areas (Shrestha and Kazama 2007) and organic & nutrient pollutions from domestic wastewater, industrial sewage and wastewater treatment plants (Zheng *et al.*, 2008; Zhao *et al.*, 2011). VF4 explained 9.7% of total variance and had strong positive loading on turbidity and TSS. This factor explained soil erosion from upland areas(Shrestha and Kazama 2007) and diluted due to the river flow that increased the level of TSS and due to the significant positive correlation between TSS and turbidity, turbidity also increased. It indicated the common source of origin of this factor group. VF5 had strong positive loading of PO_4^{3-} ; strong negative loading of DO and explained 8.46% of total variance. High loading of inorganic phosphorus represented the influence from both point sources like industrial effluents, domestic wastewater (Shrestha & Kazama, 2007)and non-point sources like agricultural runoff (Varol *et al.*, 2012). Agricultural and stream runoff carry a lot of suspended solids bearing significant level of inorganic phosphorus. Inverse relationship between PO_4^{3-} and DO indicated huge difference in pollution sources, such as point sources and non-point sources. DO loading is extremely depends upon point sources like organic wastes bearing more organic matter and some natural process like temperature variability. On the other hand, inorganic phosphorus largely comes from non-point sources and also from point sources; but not the similar as DO.

Table: 2. Rotated component matrix

	Component				
Parameters	**VF1**	**VF2**	**VF3**	**VF4**	**VF5**
Conductivity	**0.952**	-0.096	0.032	0.130	0.061
Salinity	**0.949**	-0.105	0.029	0.081	0.077
TDS	**0.866**	0.000	-0.036	-0.051	0.080
SO_4	**0.814**	0.147	-0.024	0.150	-0.095
BOD	-0.121	**0.913**	-0.104	-0.069	-0.049
COD	0.224	**0.851**	-0.195	-0.010	0.070
pH	0.007	**0.665**	0.330	0.157	-0.108
Temperature	-0.143	**0.647**	0.296	0.043	-0.081
NO_3	-0.045	0.018	**0.813**	-0.047	-0.175
NH_3	0.110	0.130	**0.752**	0.028	0.414
Turbidity	-0.042	0.125	-0.036	**0.909**	0.034
SS	0.440	-0.091	0.029	**0.791**	-0.010
Phosphate	-0.122	-0.118	0.180	0.007	**0.751**
DO	-0.186	0.008	0.170	0-.015	**-0.700**
Eigen value	3.756	2.589	1.573	1.358	1.185
Variance (%)	26.832	18.489	11.233	9.701	8.462
***CV (%)**	26.832	45.321	56.554	66.255	74.717

Extraction method: principal component analysis; Rotation method: varimax with Kaiser Normalization.
***CV means cumulative variance**

Multiple linear regressions analysis

Multiple linear regression analysis was done with SPSS 16.0 statistical software to identify the contribution of variables to water quality of Tunggak river basin. To detect the best predictors and remove the less significant variables (predictors) of water quality variation stepwise multiple linear regressions model was used (Hinton *et al*., 2004). Classical assumptions of linear regressions were checked before the interpretation of MLR model results: normal p-p plot of regression standardized residuals were analyzed. It explained that all the observed values fall roughly along the straight line and indicated that the residuals are from normally distributed population. Furthermore, scatter plot of regression standardized predicted values against observed values also showed the linear relationship between the dependent variable and the predictors; and the residuals variances are equal or constant. Removing less significant variables using stepwise multiple linear regression model for best predictors namely BOD_5, COD, pH and NH_3-N were detected; meant that the maximum water quality variation of Tunggak River was explained by those four predictor variables. Model summary showed (Table 3) the R^2=0.934 revealed that 93.4% variation of water quality of the river was explained by the above mentioned four predictors. The coefficient of the predictors estimated in the model and presented in Table 3. As can be seen COD makes the strongest unique contribution in water quality variation with Beta coefficient value -0.567. The second highest Beta value was for BOD_5 (-0.500) followed by NH_3-N (-0.454). The least contributor was pH with a Beta value 0.223. The negative sign of Beta value indicated that water quality was negatively associated/correlated with those predictors (Nathans *et al*., 2012). The analysis of variance (ANOVA) was done and the ANOVA table showed that the *F*-statistics value is 18.344 with 4 df and the corresponding *p* value is 0.003 which is highly significant. This test indicated that, the estimated slop of regression model is not equal to zero; which confirmed the linear relationship between the predictors of the applied models.

Table: 3. Estimated coefficients of the multiple linear model

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
(Constant)	60.043	5.216		11.511	0.000
BOD	-0.354	0.186	-0.500	-1.902	0.116
COD	-0.190	0.083	-0.567	-2.307	0.069
pH	1.227	1.136	0.223	1.080	0.329
NH_3-N	-4.337	1.412	-0.454	-3.072	0.028

Conclusion

In this study, principal component analysis (PCA) and multiple regression models were performed to assess Tunggak River basin water quality data sets. PCA yielded five VFs with 72.72% total variance corresponding to seven pollution sources namely: ionic, soil erosion run-off, industrial sewage, domestic wastewater, waste water treatment plant, dilution and agricultural run-off and natural & climatic reasons. Multiple linear regression supported PCA result and identified the contribution of each variable with significant values R =0.968, R^2= 0.936. From the analysis, it is clear that BOD_5, COD, NH_3-N, conductivity, salinity, TDS, SO_4, NO_3-N, turbidity, TSS, PO_4^{3-} and DO were found to be the most significant parameters responsible for water pollution in Tunggak River. Therefore, it is recommended that proper waste management (both industrial and domestic) should be taken including standard waste water treatment plant in the area and monitoring of anthropogenic activities should be ensured to confirm the least negative effects on the rivers.

Acknowledgment

Authors are grateful to the University Malaysia Pahang and the Faculty of Civil Engineering and Earth Resources to provide necessary funding and through the research project RDU 110354 and GRS 120363.

References

Andrew, D., 2005. *Standard Methods for the Examination of Water and Wastewater*. 12th ed. American Public Health Association.

Ermakov, V V, Bogomolov, A. and Bykov, D. E., 2012. Oil Sludge Depository Assessment Using Multivariate Data Analysis. ***Journal of environmental management***, 105: 144–51.

HACH. 2005. “Water Analysis Guide.”

Hinton, P, Brownlow, C. and McMurray, I., 2004. *SPSS Explained*. Routledge, Taylor and Francis group.

Huang, Fang, Xiaoquan Wang, Liping Lou, Zhiqing Zhou, and Jiaping Wu., 2010. Spatial Variation and Source Apportionment of Water Pollution in Qiantang River (China) Using Statistical Techniques. ***Water research***, 44(5): 1562–72.

Juahir, Hafizan, Sharifuddin M Zain, Mohd Kamil Yusoff, Tengku Hanidza, T. I., Mohd Armi, A. S., Mohd Ekhwan Toriman, and Mazlin Mokhtar., 2011.Spatial Water Quality Assessment of Langat River Basin (Malaysia) Using Environmetric Techniques. ***Environmental monitoring and assessment***, 173(1-4): 625–41.

Khanna, D.R, Bhutiani, R. and Kulkarni, DB., 2011. A Study on Pollution Status and Its Impact on Water Quality of River Ganga at Haridwar. ***Environment Conservation Journal***, 12: 9-15.

Koklu, Rabia, Bulent Sengorur, and Bayram Topal., 2010. Water Quality Assessment Using Multivariate Statistical

Methods—A Case Study: I
Water Resources Management, 24(5): 959–78.

Liu, Chen-Wuing, Kao-Hung Lin, and Yi-Ming Kuo., 2003.Application of Factor Analysis in the Assessment of Groundwater Quality in a Blackfoot Disease Area in Taiwan. ***Science of the total environment***, 313(1-3): 77–89.

Milovanovic, Mimoza, 2007. Water Quality Assessment and Determination of Pollution Sources Along the Axios/Vardar River, Southeastern Europe.***Desalination***, 213(1-3): 159–73.

Mustapha, Adamu, and Ado Abdu, 2012. Application of Principal Component Analysis & Multiple Regression Models in Surface Water Quality Assessment. ***Journal of Environment and Earth Science***, 2(2): 16–24.

Nasly, M A, Hossain, M A and Islam, M S, 2013. Water Quality Index of Sungai Tunggak: An Analytical Study. In ***3rd International Conference on Chemical, Biological and Environment Sciences (ICCEBS'2013)***, 40–45.

Nathans, LL, Oswald, FL and Nimon,K., 2012. Interpreting Multiple Linear Regression: A Guidebook of Variable *Importance*. ***Practical Assessment, Research & Evaluation***, 17(9): 1–19.

Onojake, M. C., Ukerun, S. O. and Iwuoha, G., 2011. A Statistical Approach for Evaluation of the Effects of Industrial and Municipal Wastes on Warri Rivers, Niger Delta, Nigeria.***Water Quality, Exposure and Health***, 3(2): 91–99.

Pejman, A. H., Nabi Bidhendi, g. R., Karbassi, A. R., Mehrdadi, N. and Esmaeili Bidhendi, M., 2009. Evaluation of Spatial and Seasonal Variations in Surface Water Quality Using Multivariate Statistical Techniques.***Int. J. Environ. Sci. Tech***, 6(3): 467–76.

Praveena, Sarva Mangala, Ong Wei Kwan, and Ahmad Zaharin Aris, 2012. Effect of Data Pre-treatment Procedures on Principal Component Analysis: a Case Study for Mangrove Surface Sediment Datasets. ***Environmental monitoring and assessment***, 184(11): 6855–68..

Ragno, Gaetano, Michele De Luca, and Giuseppina Ioele, 2007. An Application of Cluster Analysis and Multivariate Classification Methods to Spring Water Monitoring Data.***Microchemical Journal,*** 87(2): 119–27.

Saad, F.N. Mohd, Rahman, N A and Nik Norulaini, 2008. Identification Of Pollution Within The Sungai Pinang River Basin.***eprints.usm.my***: 478–85.

Satheeshkumar, P, and Anisa B Khan, 2012. Identification of Mangrove Water Quality by Multivariate Statistical Analysis Methods in Pondicherry Coast, India.***Environmental monitoring and assessment***, 184(6): 3761–74.

Shrestha, S., and Kazama, F., 2007. Assessment of Surface Water Quality Using Multivariate Statistical Techniques: A Case Study of the Fuji River Basin, Japan.***Environmental Modelling & Software*,**22(4): 464–75.

Simeonov, V, Stratis, J A, Samara, C,Zachariadis, G, Voutsa, D, Anthemidis, A, Sofoniou, M and Kouimtzis, T H., 2003. Assessment of the Surface Water Quality in Northern Greece. ***Water research***, 37(17): 4119–24.

Singh, KP, Amrita Malik, and Sarita Sinha, 2005. Estimation of Source of Heavy Metal Contamination in Sediments of Gomti River (India) Using Principal Component Analysis. ***Water, Air, and Soil Pollution***, 166: 321–41.

Varol, Memet, Bülent Gökot, Aysel Bekleyen, and Bülent Şen, 2012. Spatial and Temporal Variations in Surface Water Quality of the Dam Reservoirs in the Tigris River Basin, Turkey.***Catena*,** 92: 11–21.

Wang, Xiao-long, Yong-long Lu, Jing-yi Han, Gui-zhen He, and Tie-yu Wang, 2007. Identification of Anthropogenic Influences on Water Quality of Rivers in Taihu Watershed.***Journal of environmental sciences (China),*** 19(4): 475–81.

Wu, Mei-Lin, You-Shao Wang, Cui-Ci Sun, Haili Wang, Jun-De Dong, Jian-Ping Yin, and Shu-Hua Han, 2010. Identification of Coastal Water Quality by Statistical Analysis Methods in Daya Bay, South China Sea.***Marine pollution bulletin***, 60(6): 852–60.

Zhang, Yunlin, Yan Yin, Longqing Feng, Guangwei Zhu, Zhiqiang Shi, Xiaohan Liu, and Yuanzhi Zhang, 2011. Characterizing Chromophoric Dissolved Organic Matter in Lake Tianmuhu and Its Catchment Basin Using Excitation-emission Matrix Fluorescence and Parallel Factor Analysis. ***Water research*,**45(16): 5110–22.

Zhao, Jian, Guo Fu, Kim Lei, and Yanwu Li, 2011. Multivariate Analysis of Surface Water Quality in the Three Gorges Area of China and Implications for Water Management.***Journal of Environmental Sciences***, 23(9): 1460–71.

Zheng, Bing-Hui, Cheng-Jin Cao, Yan-Wen Qin, and Min-Sheng Huang, 2008. [Analysis of Nitrogen Distribution Characters and Their Sources of the Major Input Rivers of Three Gorges Reservoir]. ***Huan jing ke xue= Huanjing kexue / [bian ji, Zhongguo ke xue yuan huan jing ke xue wei yuan hui "Huan jing ke xue" bian ji wei yuan hui.]*** 29(1): 1–6.