Assessment of Surface Water through Multivariate Analysis

Abbas F.M. Alkarkhi (Corresponding Author)

Environmental Division, School of Industrial Technology

Universiti Sains Malaysia

11800 Penang, Malaysia

Tel:60-4-653-2107   E-mail: abbas@usm.my

Anees Ahmad

School of Industrial Technology, Environmental Division

Universiti Sains Malaysia

11800 Penang, Malaysia

E-mail: anees@usm.my

Norli Ismail

School of Industrial Technology, Environmental Division

Universiti Sains Malaysia

11800 Penang, Malaysia

E-mail: Norlii@usm.my

Azhar mat Easa

School of Industrial Technology, Food Division

Universiti Sains Malaysia

11800 Penang, Malaysia

E-mail: azhar@usm.my

Khalid Omar

School of medical physics

Universiti Sains Malaysia

11800, Penang, Malaysia

E-mail: khalhadithi@yahoo.com

Abstract

Multivariate statistical techniques such as factor analysis (FA) and Discriminant analysis (DA), were applied for the evaluation of spatial variations and the interpretation of a large complex water quality data set of two rivers (Juru and Jejawi) in Malaysia, monitoring 10 parameters at 10 different sites each. Factor analysis resulted in two factors explaining more than 82% of the total variance in water quality data set. The factors indicate that the possible variances in water quality may be due to either sources of anthropogenic origin or due to different biochemical processes that are taking place in the system. The first factor called pseudo anthropogenic factor explained 59.29% of the total variance. The second factor called anthropogenic explained 23.03%. DA gave the best result to identify the relative contribution for all parameters in discriminating (distinguishing) the two rivers affording 100 % correct assignations. This study illustrates the benefit of multivariate statistical techniques for analyzing and interpretation of complex data sets, and to plan for future studies.

Keywords: Discriminant analysis, Factor analysis, Water quality, Multivariate

1. Introduction

The rivers are the important sources of surface water and boon of nature to the human beings. They are the inseparable component of ecology on the earth. A river, with its tributaries, is a system that sustains fish and the other aquatic life. It does one way transport of a significant load of dissolved matter and particulate material from different sources (Shrestha and Kazama, 2007) in the direction of its flow. Rivers play a major role in assimilation or transportation of the municipal and industrial wastewater discharge of a constant as well as occasional or seasonal polluting source. The surface runoff is a seasonal phenomenon which is largely affected by climate within the river basin (Singh et al., 2004). The seasonal variation in precipitation, surface runoff, interflow, groundwater flow and pumped in and outflows have a strong effect on river discharge and subsequently on the concentration of pollutants in river water (Vega et al., 1998). Rivers constitute the main inland water resources for domestic, industrial and irrigation purposes; it is inevitable to prevent and control the rivers pollution and to have reliable information on quality of water for effective management. The possible variances in water quality may be due to anthropogenic activities, natural variances during months (season) due to various biochemical or chemical processes. Monitoring programs result in a huge and complex data matrix consist of a large number of physico-chemical parameters (Chapman, 1992).

The application of multivariate methods such as Cluster analysis (CA), principal analysis (PCA), factor analysis (FA), and discriminant analysis (DA) has increased tremendously in recent years for analyzing environmental data and drawing meaningful information (Vega et al., 1998; Lee et al.,2001;. Wunderlin, et al., 2001 ; Reghunath, et al., 2002: Saadia, et al.,2005). In this paper we report our findings of the study of water quality of two rivers of Malaysia and their statistical analysis. The analysis is done to explore the extent of resemblance among the sampling sites, to identify the variables responsible for spatial variations in river water quality, to locate the hidden factors explaining the structure of the database, and to quantify the influence of possible natural and anthropogenic sources on the water parameters of the two selected rivers.

2. Methods

2.1 Study area

The two rivers of Malaysia selected for the study are Juru and Jejawi located in the North West coast of peninsular Malaysia, in the state of Penang and within a coastal mudflat in the Juru and Bukit Tambun district (Fig. 1). The sites are located adjacent to industrial areas which were reclaimed from mangrove. The types of industry presently in operation include: electronics; textiles; basic and fabricated metal products; food processing and canning; processing of agricultural products; feed mills; chemical plants; rubber based industry; timber based wood products; paper products and printing works; and transport equipment. Other main activities that are operating in vicinity of the cultured area are a ships' harbour with petroleum unloading and a red earth quarry which extends right up to the coastline.

2.2 Analytical Methods

The water quality of monitoring sites comprising 10 water quality parameters were monitored monthly over one year (2006) and analysed as given below:

The temperature and conductivity were measured using HACH portable pH meter, Dissolves oxygen (DO) was measured with YSI 1000 DO meter. biochemical oxygen demand (BOD), chemical oxygen demand (COD), total nitrate and total phosphate concentrations were analyzed using Spectrophotometer (HACH/2010). Turbidity was measured using Nephlometer. Total suspended solids (TSS) were analyzed gravimetrically at the laboratory. APHA Standard Methods for the Examination of Water and Wastewater were applied for the analysis of concentration of above mentioned parameters.

2.3 Multivariate statistical methods

2.3.1 Discriminant function

Discriminant analysis is a multivariate technique used for two purposes, the first purpose is description of group separation in which linear functions of the several variables (discriminant functions (DFs)) are used to describe or elucidate the differences between two or more groups and identifying the relative contribution of all variable to separation of the groups. Second aspect is prediction or allocation of observations to group in which linear or quadratic functions of the variable (classification functions (CFs)) are used to assign an observation to one of the groups (Richard&Dean, 2002; Alvin, 2002). SPSS version 12 software was used for carrying out the statistical analysis of the data.

2.3.2 Factor analysis

Factor analysis (FA) is designed to transform the original variables into new uncorrelated variables called factors, which are linear combinations of the original variables. The FA is a data reduction technique and suggests how many variates are important to explain the observed variances in the data. Principal components method (PCA) is used for extraction of different factors. The axis defined by PCA is rotated to reduce the contribution of less significant variables (Vega et

al., 1998; Helena et al., 2000). This treatment provides a small number of factors that usually account for approximately the same amount of information as the original set of observations. The FA can be expressed as:

$$Z_{ij} = a_1 f_{1j} + a_2 f_{2j} + ... + a_m f_{mj} + e_{ij}$$

(1)

where z is the measured variable, a is the factor loading, f is the factor score, e the residual term accounting for errors or other source of variation.

3. Result and discussion

3.1 Factor analysis

Factor analysis was carried out on the data set (10 variables) to compare the compositional patterns between analyzed water samples and to identify the various factors that influence each of them. Two factors were extracted explaining more than 82 % of the total variance in the water quality data set. Eigenvalues >1 were taken as criterion for the extraction of the principal components required for explaining the source of variances in the data set .The eigenvalues for different factors, percentage variance accounted and cumulative percentage variance are given in Table 1. The Scree plot is shown in Fig.3 to clarify the method of extraction of different factors.

The factor analysis was actually performed on the correlation matrix between different parameters followed by Varimax rotation and the same has been used to examine their inter relationship.

The parameter loadings for the two identified factors from the factor analysis of the data are given in Table 2. The factor 1 accounts for 59.29 % of the total variance. It is positively correlated (loading > 0.75) with turbidity, temperature and nitrate concentration while negatively correlated with BOD, and phosphate concentration. This factor appears to be originated from the combined effect of anthropogenic activities accompanied with partial ecological recovery system of the river. So this factor may be called as *pseudo anthropogenic* factor.

Factor 2, on the other hand, explains 23.03% of the total variance and is positively loaded with conductivity and COD. Since the causes of these two parameters are based on excessive industrial activities and are not compensated/ removed instantaneously by the natural recovery system so might be termed as *anthropogenic* only.

3.2 Source Identification

An attempt was made to study the relationship between factor scores and the samples from different sites. The scores for the first factor are shown in Fig. 3. It is observed that turbidity and nitrate concentrations recorded were low where as BOD and phosphate concentrations were higher in Jejawi River for all sites. The high BOD and phosphate concentrations indicate relatively high waste dumping activity in the Jejawi river. Since the high value of BOD accounts for higher micro organism concentration which in tern may consume nitrate and can cause the precipitation of suspended and colloidal particles in water causing thereby reduction in turbidity. It also indicates that Jejawi river has low pollution than Jeru river due to probably lesser industrial activities and easy natural recovery process in the former river. The physical assessment of industrial area shows that more industries are located at Jeru River and thus provides more strength to the conclusion drawn above.

The scores for the second factor are presented in Fig. 4 and appear to show opposite behavior in the two rivers. The more pollution load from industrial activities in Jeru River probably weakens its natural recovery system which in turn is normal in Jejawi River. Thus the two rivers are almost opposite to each other in terms of industrial pollution and natural recovery system. For the same reason all parameters correlated with the second factor (Fig. 5) are exchanged their level among the sites in the two rivers.

3.3 Discriminant analysis (DA)

Variation in water quality parameters was evaluated through DA. The DA applied on raw data consisted of ten parameters. Only one DF was found to discriminate the two rivers as shown in Table.3. Wilk's Lambda test showed that DF is statistically significant as shown in Table 4. Furthermore 100% of the total variance between the two locations was explained by only one DF. The relative contribution for each parameter is given in Eq.2.

$$Z = 1.60\ Tem. + 0.97\ pH + 1.86\ Do - 1.48\ BOD + 0.83\ COD - 1.00\ TSS + 1.41\ Cod. + 2.18\ Tur.$$
$$+ 0.03\ phos. - 0.34\ Nit.$$

(2)

It can be seen that, Turbidity, Temperature, Do, BOD, Conductivity and Tss exhibited strong contribution in discriminating the two locations and account for most of the expected variations in water quality, while other parameters showed less contribution in explaining the variation between Juru and Jejawi River. The relative contribution for water quality parameters can be arranged in the order;

Turbidity > Do > Temperature > BOD > Conductivity > TSS > pH > COD>Nitrate> phosphate.

The classification matrix (Table 5) showed that more than 100% of the cases were correctly classified to their respective groups. The results of classification also showed that significant differences existed between these two rivers, which are expressed by in term of one discriminant function.

4. Conclusion

The multivariate statistical techniques, namely, cluster analysis and factor analysis are important analytical techniques for the processing of water quality parameters and power full tools for classification as well as identification of possible sources of pollution. The techniques are also helpful in providing the possible mechanism with justification, by simple reasoning, to the causes of variation in water quality parameters.

Acknowledgement

References

Alvin, C. R., (2002).   Methods of Multivariate Analysis. John Wiley & Sons.INC. USA.

Champan, D., (1992). Water quality assessment. In; Chapman, D. (Ed.), on behalf of UNESCO, WHO and UNEP, Chapman&Hall, London.

Lee, J.Y., Cheon, J.Y., Lee, K.K.,Lee,S.Y., Lee, M.H., (2001). Statistical evaluation of geochemical parameter distribution in a ground water system contaminated with petroleum hydrocarbons. Journal of Enviromental Quality 30, 1548-1563.

Otto, M., (1998). Multivariate methods. In: Kellner, R., Mermet, J. M., Otto, M., Widmer, H.M.(Eds), Analytical Chemistry. Wiley-VCH, Weinheim, Germany .

Reghunath, R., Murthy, T.R.S., Raghavan, B.R.,2002. The utility of Multivariate statistical techniques in hydrogeo-chemical studies: an example from Karanataka, India. Water research, 36, 2437-2442.

Richard A. Johnson and Dean W. Wichern, (2002). Applied Multivariate statistical analysis. Prentice-Hall, London.

 Shrestha, S., Kazama, F., (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of Fuji river basin. Japan. Enviromental Modelling & software, 22, 464-475.

Saadia R. Tariq, Munir H. Shah, N. Shaheen, A. Khalique, S. Manzoor, M.anzoor, M. Jaffar, (2005). Multivariate analysis of trace metal levels in tannery effluents in relation to soil and water: A case study from Peshawar, Pakistan. Journal of environmental management.

Singh. K.P., Malik,A., Mohan, D., Sinha, S., (2004). Multivariate statistical technique for the evaluation of spatial temporal variation in water quality of Gomti River (India): a case study. Water research 38,3980-3992.

Vega, M., Pardo, R., Barrado, E., Deban, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water research, 32, 3581-3592.

Wunderlin, D. A.,Diaz, M.P., Ame, M.V., pesce, S.F., Hued A.C., Bistoni, M.A., (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (cordoba-Argentina). Water Res. 35,2881-2894.

Table 1. Extracted values of various factor analysis for water quality parameters

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.93	59.29	59.29	5.93	59.29	59.29	4.79	47.90	47.90
2	2.30	23.03	82.32	2.30	23.03	82.32	3.44	34.42	82.32

Extraction Method: Principal Component Analysis.

Table 2. Results of the factor analysis for water quality parameters

Parameter	factor	
	F1	F2
Turbidity	0.97	0.09
Temperature	0.93	-0.09
BOD	-0.91	-0.40
Phosphate	-0.85	-0.45
Nitrate	0.75	0.26
TSS	-0.54	0.03
PH	-0.02	0.96
DO	-0.03	0.95
Conductivity	0.44	0.83
COD	0.63	0.70

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 3 iterations.

Table 3. Eigen-value of DF for the two locations

Function	Eigen-value	% of Variance	Cummulative %
1	416.779	100	100

Table 4. Wilks' Lambda for testing discriminant function validity

Test of Function	Wilks' Lamda	P-value
1	0.002	< 0.0001

Table 5. Classification results for discriminant analysis of the two rivers

Locations	% correct	Predicted group membership	
		Juru River	Jejawi River
Juru River	100	10	0
Jejawi River	100	0	10

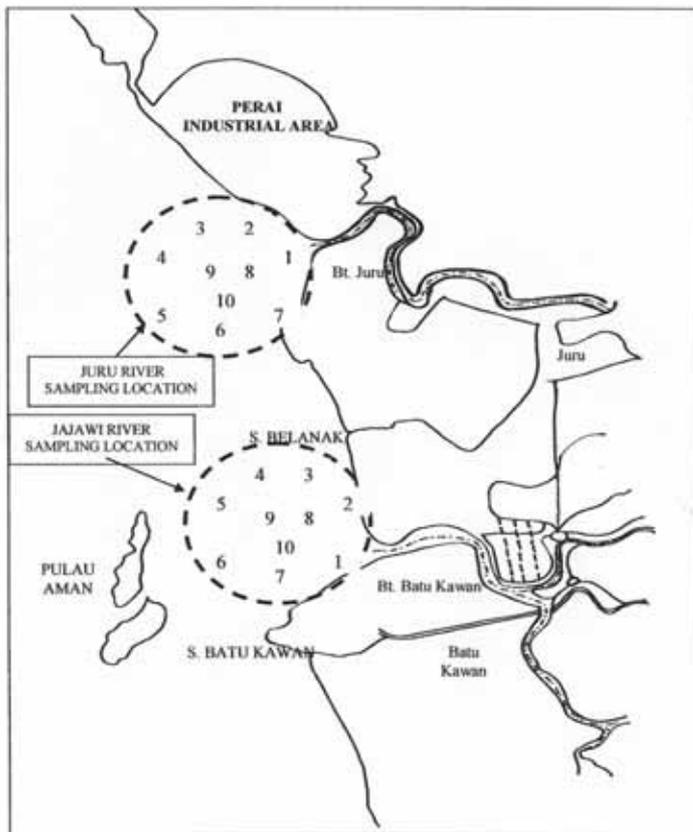a   100% of original grouped cases correctly classified.

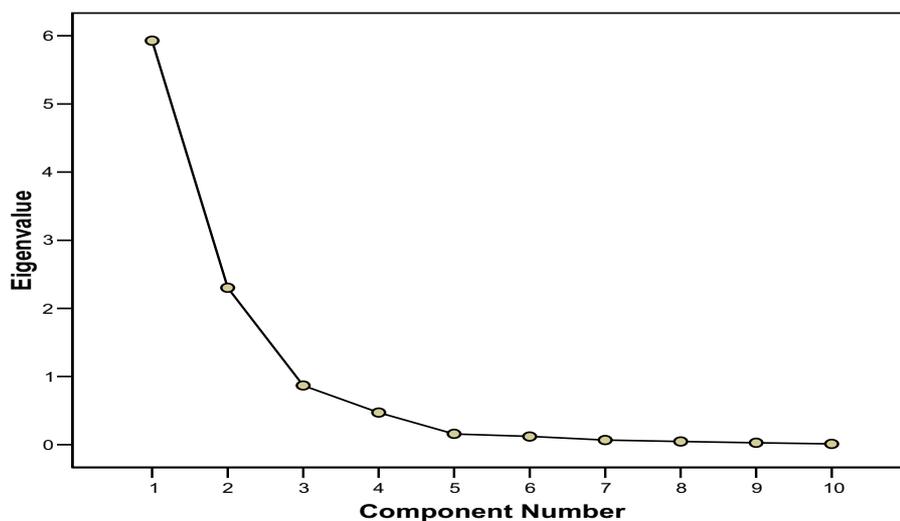Figure 1. Map of sampling locations for study areas
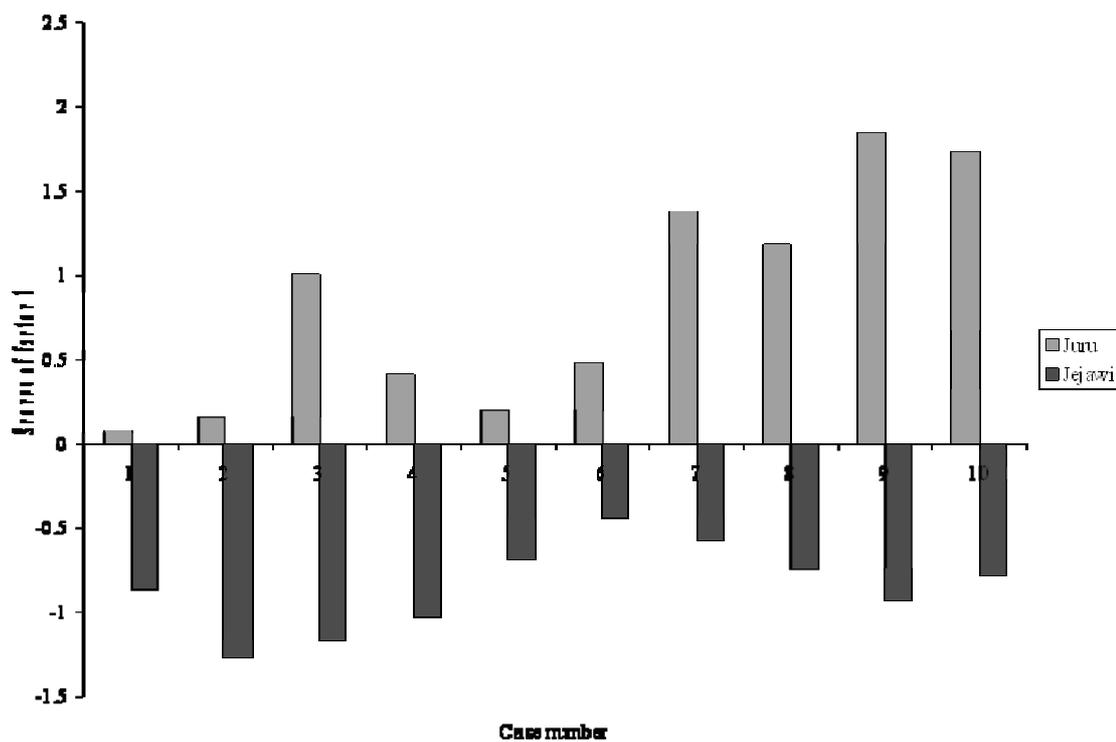
Scree Plot

Figure 2. Scree plot of Eigen-values

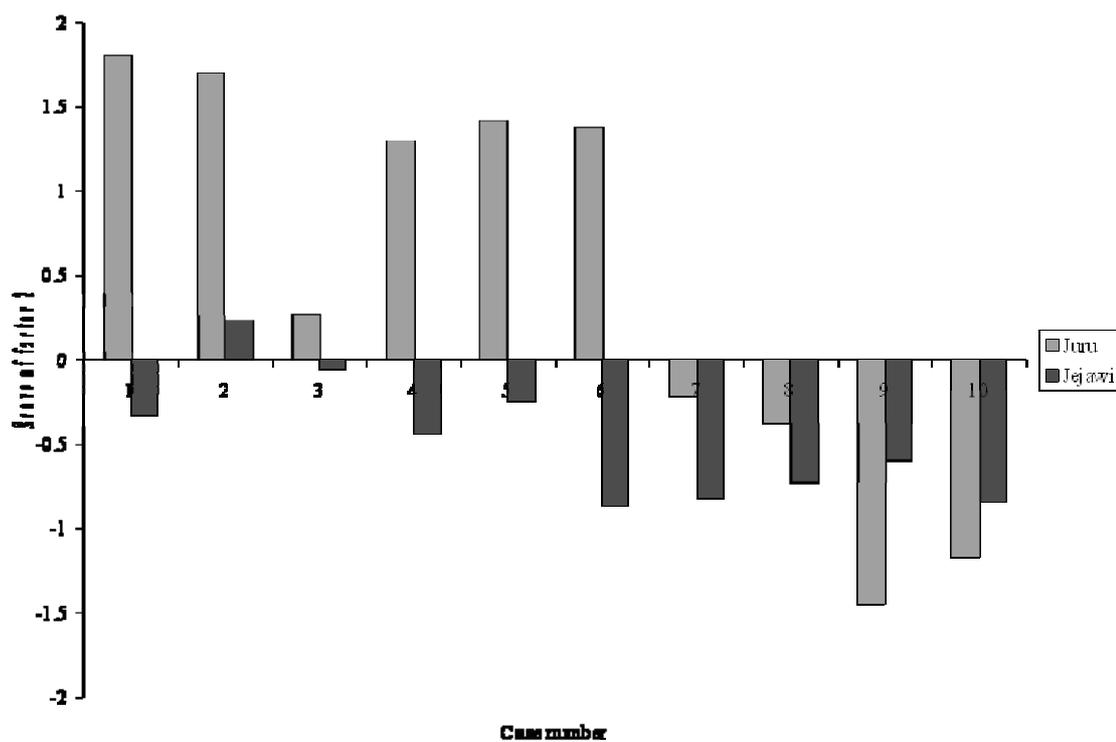Figure 3. Factor score for factor 1 for different sites of Juru and Jejawi Rivers

Figure 4. Factor score for factor 2 for different sites of Juru and Jejawi Rivers