Surface River Water Quality Interpretation Using Environmetric Techniques: Case Study at Perlis River Basin, Malaysia

Mohd Saiful Samsudin[#1], Hafizan Juahir[#2], Sharifuddin M. Zain[*3], Nur Hazirah Adnan[#4]

[*]Department of Chemistry, Faculty of Science,50603, Kuala Lumpur, Malaysia

[#]Faculty of Environmental Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[1]saifulsamsudin294@gmail.com

[2]hafizan@env.upm.edu.my

Abstract- **Environmetric techniques, cluster analysis (CA), discriminant analysis (DA) and principle component analysis (PCA) were applied to the data on water quality of Perlis River Basin, generated from 2003 until 2007. There are eleven monitoring stations with different sites for 30 parameters. Three spatial clusters are designated as downstream and upstream of Perlis River regions. Forward and backward stepwise DA managed to discriminate ten and thirteen water quality variables, respectively, from the original 30 variables. The final results for this study showed that hydrological observations are supported by principle component analysis (PCA). Additionally, this joint analysis makes it possible to observe the significance of the pollutent sources which contribute to pollution. Nine PCs with eigenvalues greater than 1 explaining 77% of the total variance in the water-quality data set. The investigation from PCA showed nine main pollution sources on Perlis River which are a mineral component of the river water, surface runoff sources, anthropogenic pollution sources, municipal sewage and wastewater treatment plants, leachate from industrial activities, seasonal impact of discharge and temperature, agricultural waste, oil waste from restaurant and road runoff. Finally, the application of environmetric methods can result in significant cost reduction and will allow more effective and efficient river quality management activities.**

***Keywords-* Environmetric; Perlis River; Water quality; Cluster analysis; Discriminant analysis; Principle component; factor analysis**

I. INTRODUCTION

Surface water hydrochemistry depends on multiple natural factors such as intensity and composition of rain water, chemical reactions between water and soil or sediment, biochemical reaction, and surface water-groundwater interactions, as well as on human activities. In particular, such factors can significantly alter these systems, either by polluting them or by modifying the hydrological cycle [1, 2].

Surface waters are most vulnerable to pollution due to their easy accessibility for disposal of wastewaters. Both the natural processes, such as precipitation inputs, erosion, weathering of crustal materials, as well as the anthropogenic influences viz. urban, industrial and agricultural activities, increasing exploitation of water resources, together determine the quality of surface water in a region[3]. Rivers play major role in assimilation or carrying off the municipal and industrial wastewater and run-off is a seasonal phenomenon, largely affected by climate in the basin.

The discharge of industrial and municipal wastewaters, whether treated or not, can be considered a constant polluting source that modifies surface water hydrochemistry. Additionally, surface waters can be affected by non-point source pollution, caused by surface and subsurface runoff from croplands and urban areas, which is seasonal and highly affected by climate variability. These quality problems can be worsened since seasonal variations of precipitation, surface runoff, interflow, groundwater flow and pumped inflow and outflow have a strong effect on flow rates, and hence on the concentration of pollutants in the stream water[2, 4].

Perlis River is classified as a Class III river. It is currently experiencing heavy erosion at its river banks and has become very shallow. Rubbish that has been thrown into the river is therefore very visible and it is not accessible to boats. There is a landfill located in Kuala Perlis and this directly affects the water quality of the river. Squatters located near the river reserve area are also causing pollution problems. Other pollutant sources include shrimp livestock ponds, Kangar Wet Market, Sungai Perlis Esplanade, food stalls and the Kuala Perlis Fisherman Jetty. Nonetheless, the river was acknowledged with MS ISO 14001 (environmental management) for the Management of Perlis River. In this study, secondary data were taken into consideration.

Application of environmetric techniques approach such as multivariate analysis like cluster analysis (CA), discriminant analysis (DA) and factor analysis (FA)/ principal component analysis (PCA) for interpretation of the complex databases offers a reliable and better understanding of the hydrochemistry and hydrochemical processes in the study region. These techniques also permit identification of the possible factors/sources that influence the water systems and are responsible for the variations in water quality, thus offers valuable tool for developing appropriate strategies for effective management of the water resource.

The environmetric techniques, such as CA, DA and PCA have successfully been applied earlier for interpretation of surface and groundwater quality data bases and identification of the anthropogenic influences. Environmetrics is esteemed to be the best approach to circumlocute misinterpretation of large amounts of complex environmental monitoring data. Environmetric methods have been widely used in conveying

DOI 10.5963/IJEP0105001

meaningful information from assemblages of environmental data. These methods have often been used in inquisitive data analysis tools for the classification of samples or sampling stations and the identification of pollution sources. Environmetrics have also been applied to characterize and evaluate the surface and freshwater quality, as well as verifying spatial and temporal variations caused by natural and anthropogenic factors. Nowadays, multivariate analysis methods have become an important tool in environmental sciences to reveal and evaluate complex relationships in a wide variety of environmental applications. PCA technique extracts the eigenvalues and eigenvectors from the covariance matrix of original variables. The PCs are the uncorrelated (orthogonal) variables, obtained multiplying of the original correlated variables with the eigenvector, which is a list of coefficients (loadings or weightings). Thus, the PCs are weighted linear combinations of the original variables. PC provides information on the most meaningful parameters, which describe whole data set affording data reduction with minimum loss of original information [4, 5]. Factors analysis further reduces the contribution of less significant variables obtained from PCA and the new group of variables known as varifactors is extracted through rotating the axis defined by PCA.

There are several problems that were focused on this study. The abundance of secondary data from Department of Environment (DOE) and other agencies are need to utilize for exact identification and distribution of pollution for more effective and optimized sampling strategy. This study also focused on knowledge on the risks of rivers being unable to withstand current. Futhermore, future pollution loading needs to be ascertained and need more comprehend study. The objectives of this study are to classify the river class, find the significant parameters that contribute to clustering groups and to identify the source pollution of Perlis River.

II. METHODOLOGY

A. Study Site

The Perlis River basin is situated at north of peninsular Malaysia. The river basin originates from a natural reservoir in the forested area. Eleven water quality stations were selected from the upstream to downstream of the Perlis River Basin. The secondary data were collected from monitoring stations under the river water quality monitoring program by the Department of Environment (DOE) from 2003 to 2007. Thirty physico-chemical parameters were involved. The samples were taken at these stations which are in different area and location as shown as in Fig. 1. There are 11 stations which are the sub-rivers include Perlis (SPS01),Arau (SPS02), Ngulang (SPS03), Serai (SPS04), Jernih (SPS05 & SPS06), Tasoh (SPS07), Jarum (SPS08), Kok Mak (SPS09), Peralit (SPS10) and Kechor River (SPS11). Futhermore, not all the stations were consistently sampled throughout the sampling years[6].

B. The Data

The multivariate analysis which were used in this study is consists of 30 water quality parameters and 281 observations data sets [data matrix : 30 x 281]. The samples were analysed for all the parameters which include dissolved oxygen (DO), biological oxygen demand (BOD) , chemical oxygen demand (COD), suspended solid (SS), pH, ammonia calnitrogen (NH3-NL), temperature, electrical conductivity, salinity, turbidity, dissolved solid (DS), total solid (TS), nitrogen (NO3), chlorine (Cl), phosphate (PO4), arsenic (As), mercury (Hg), cadmium (Cd), cromium (Cr), plumbum (Pb), zinc (Zn), calcium (Ca), iron (Fe), potassium (K), magnesium (Mg), sodium (Na), oil and grease (OG), detergen (MBAS), E.coli and coliform. These data are arranged in sequence by date throughout the years of 2003 until 2007 from upstream to downstream of eleven monitoring stations along Perlis River.

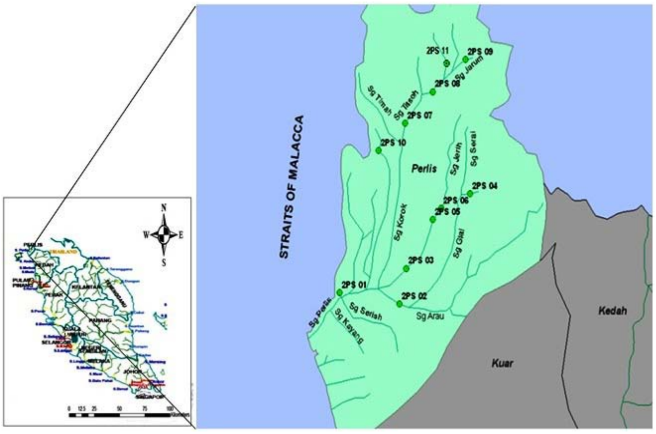

Fig. 1 Map of Perlis River with 11 water quality monitoring stations along the main river. (Source: Alam Sekitar Malaysia Sdn Bhd)

Preliminary work was done on the data matrix which included the assembling and transformation of the data. Data which were below the detection limit were complemented with values equal to half the detection limit. Normal distribution tests were carried out with W (Shapiro-Wilk) test ; the agreement of the distribution of the physico-chemical parameters of water with the normal distribution was checked. Variables with a distribution other than normal were subjected to a transformation. In the case of variables, where their post-transformation distribution differed significantly from the normal one, those parameters were not taken into consideration when making the environmentric analysis[6].

C. Cluster Analysis

Cluster analysis is one of the method that was applied in independently pattern recognition. It is an independently pattern recognition technique that uncovers constitutional structure or underlying behavior of a data set without making an assumption about data, to classify the objects of the system into categories or clusters based on their nearness or similarity[4]. The agglomerative hierarchical cluster (AHC) analysis according to Ward (1963)[7] was applied to detect multivariate similarities between sampling sites in different sampling plots for different sampling days.

Hierarchical clustering is the most similar pair of objects and forming higher clusters step by step. The Euclidean distance usually gives the similarity between two samples and a 'distance' can be presented by the 'difference' between analytical values from both the samples[8]. AHC analysis was performed on the normalized data set by means of the

DOI 10.5963/IJEP0105001

Ward's method, using Euclidean distances as a measure of similarity. This method uses the analysis of variance approach to evaluate the distance between clusters, attempting to minimize the sum of squares of any two clusters that can be performed at each step. Cluster analysis was applied to the water quality index data set with a view to group the similar sampling sites (spatial variability) spread over the river. The result is illustrated by a dendogram,presenting the clusters and their proximity. The Euclidean distance (linkage distance) is reported as Dlink/Dmax, which represents the quotient between the linkage distance divided by the maximal distance. The quotient is usually multiplied by 100 as a way to standardized the linkage distance represented by the y-axis[9].

D. Discriminant Analysis (DA)

Discriminant analysis determine the variable that discriminate between two or more naturally occuring clusters. It constructs a discriminant function (DF) for each group. DFs were calculated using equation as in Eq. (1) :

$$f(Gi) = ki + \sum_{j=1}^{n} wj\ Pij \qquad (1)$$

where i is the number of groups (G), ki the constant inherent to each group, n the number of parameters used to classify a set of data into a given group and wj is the weight coefficient assigned by DF analysis (DFA) to a given parameter (Pj). In this study, DA was applied to determine whether groups differ with regards to the mean of a variable, and to use that variable to predict group membership[10].

DA was devoted to the variable to predict cluster interrelation. DA was applied to the raw data by using the standard, forward stepwise and backward stepwise modes. These were used to construct DFs to evaluate pollution source of variations in the river water quality. The identified pollution sources were the grouping (dependent) variables, while all the measured parameters constitute the independent variables. In the forward stepwise mode, variables were included step-by-step beginning from the most significant variable until no significant changes were obtained. In the backward stepwise mode, variables were removed step-by-step beginning with the less significant variable until no significant changes were obtained[11].

E. Principal Componen Analysis (PCA)

In PCA, eigenanalysis of the experimental data was performed to extract principal components (PCs) of the measured data, using two selection criteria : the screen plottest and corrected average eigenvalue[4]. A varimax factor (VF) can include unobservable, hypothetical, latent variables, while a PC is a linear combination of observable water-quality variables [4, 5, 12]. The characteristic roots (eigenvalues) of the PCs are a measure of their associated variances, and the sum of eigenvalues coincides with the total number of variables. Correlation of PCs and original variables is given by loadings, and individual transformed observations are called scores. Design of experiment (DOE) method was employed to identify the interaction between the seasonal effected to the water quality parameter. Statistical analysis was carried out by using XLSTAT2010 Excel add-in software for Windows packages.

III. RESULT AND DISCUSSION

A. Clustering of Sampling Station Based on Water Quality Data

AHC is used to classify water quality stations based on their similarity level. AHC was performed on the water quality data set to examine spatial variation among sampling stations. The classification procedure procreated three group in a very convincing way, as the stations in these groups have similar characteristic and natural ground.

Three clusters that formed according to their similarity are Cluster 1 (SPS01 and SPS02), Cluster 2 (SPS03, SPS04, SPS06, SPS07, SPS08, SPS09, SPS10 and SPS11) and Cluster 3 (SPS05). Based on Fig.2, Cluster 1 assigned as moderate pollution source (MPS), Cluster 2 assigned as ow polution source (LPS) and Cluster 3 assigned as high polution source (HPS). This method shows that only one station in each cluster is needed to represent a reasonable accurate spatial assessment of the water quality for the whole network. This result implies for rapid assesment of water quality and efficient river quality management[9].

Based on Singh et al., (2004)[13], the number of sampling sites in the monitoring network reduced without losing any significance of the outcome. This method also useful in offering reliable clustering of water quality of the whole region and can be used to design future spatial sampling strategies in an optimal manner[14].

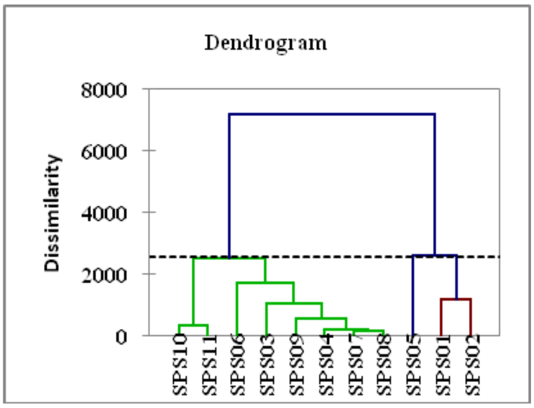

Fig. 2 Dendogram showing different clusters of sampling sites located at Perlis River Basin based on water quality parameters.

B. Discrimant Analysis of River Water Quality

From this study, the discriminant functions were performed based on the raw data, and similar classification matrixes were obtained using standardised data sets. To find the discriminating variables, the data were subjected to standard, forward stepwise and backward stepwise DA. The accuracy of spatial classsification using these methods were 84.29% (17 discriminant variables) by using standard method, 83.93% (10 discriminent variables) by using standard forward method and 83.93% (13 discriminant variables) by using standard backward method by each one (Table 1). The Wilk's lambda test for standard mode gave a Lambda value of 0.31 and $p < 0.01$. The null hypothesis states that the means of vectors of the 3 cluster (HPS, MPS, LPS) are similar with each other. The alternative hypothesis, on the

DOI 10.5963/IJEP0105001

other hand, states that at least one of the means of vectors is different to another. Since the enumerated p-value is lower than significant level of alpha is 0.05, one should reject the null hypothesis and accept the alternative hypothesis. The risk of rejecting the null hypothesis while it is true is lower than 0.01% . Thus the 3 clusters are indeed diffrent from one another because of different characteristic.

Discriminant functions (DFs) and classification matrices (CMs) obtained from standard, forward stepwise and backward stepwise modes of DA are shown in Table 1 and Table 2. Based on Table 2, the standard DA mode construct DFs including 30 parameters. The coefficients for the *E.coli* and coliform bacteria group were zero. In forward stepwise mode, variables are included one-by-one beginning with the more significant until no significant changes are obtained, whereas in backward stepwise mode, variables are removed one-by-one beginning with the less significant until no significant changes are obtained[13].

There are 10 significant variables that can be found by using forward stepwise mode which are COD, pH, NH3-NL , temperature, turbidity, Cl, Ca, K, Mg and Na. This results showed us that these parameters have high variation in terms of their spatial distribution. Backward stepwise mode on the other hand included DO, conductivity, salinity and TS as their other parameters to have a high spatial variation and leaving chloride behind. Box and whisker plots of these water quality parameters over five year period since 2003 until 2007 as shown in Fig.3. Ten selected parameters which gave high variations by forward stepwise DA are used for further analysis.

TABLE 1 CLASSIFICATION MATRIX BY DA FOR SPATIAL VARIATIONS IN PERLIS RIVER BASIN

		Regions Assigned by DA		
Sampling Regions	**% correct**	**MPS**	**LPS**	**HPS**
Standard DA mode				
MPS	50	15	15	0
LPS	94.2	7	195	5
HPS	60.47	0	17	26
Total	84.29	22	227	31
Forward stepwise mode				
MPS	36.67	11	19	0
LPS	95.17	5	197	5
HPS	62.79	0	16	27
Total	83.93	16	232	32
Backward stepwise mode				
MPS	43.33	13	17	0
LPS	93.72	8	194	5
HPS	65.12	0	15	28
Total	83.93	21	226	33

C. Principal Component Anaysis

PCA was applied to the normalized data to compare the compositional patterns between the analyzed water samples and also to identify the factors that influenced each one. Rotation of the axis defined by PCA produced a new set of factors, each one involving primarily a subset of the original variables are divided into groups somewhat independent of each other[13]. An eigenvalue gives a measure of the significance of the factor which the factors with the highest eigenvalues are the most significant. Eigenvalues of one or greater are considered significant[15]. PCA of the entire data set (Table 5) involved nine PCs with eigenvalues greater than one explaining about 77% of the total variance in the water-quality data set. Projections of the original variables on the subspace of the PCs are called loadings and coincide with the alternation coefficients between PCs and variable. Shrestha et al., (2007)[16] classified the factor loading as ‘strong’, ‘moderate’ and ‘weak’, corresponding to absolute loading values of greater than 0.75, 0.75-0.50 and 0.50-0.30, respectively.

From Table 3, VF1 which explains 30.3% of total variance has strong positive loadings on conductivity, salinity, dissolved solid, Cl, Ca, K, Mg and Na, which can be interpreted as a mineral component of the river water. Vega et. al (1998)[4] said this clustering variables points to a common origin for these minerals, likely from dissolution of limestone and gypsum soils. VF2, which explain 11.9% of total variance has strong positive loading of SS, turbidity and Fe. This VF represents surface runoff sources because of its contains. This factor loaded with solids indicates towards their origin in run-off from the fields with high load of solids and waste disposal activities. VF3 explained 7.6% of strong variance has strong positive loading of BOD and COD which represent anthropogenic pollution sources. This first factor could be explained by considering the chemical components of various anthropogenic activities that constitute point source pollution exspecially from industrial, domestic, and commercial and agricultural runoff areas[6]. This VF can be explained that high levels of dissolved organic matter and biological organic matter are coming from agricultural activities (paddy) and industrial activies based on observation along Perlis River. VF4 (6.9% of total variance) has strong positive loading of *E.coli* and coliform. *E.coli* and coliform are strongly related to municipal sewage and wastewater treatment plants along the river[17]. VF5 (4.9% of total variance) has one strong positive loading which is chromium. Chromium a specific pollutant providing evidence of industrial pollution like dyeing or paint operations [18]. From the site survey along Perlis River, the main activities on this river are fishery and agriculture. On the other hand, the presence of Cr can be related to the anti-fouling paint from fishermen’s boats. Anti-fouling paint is made for boat bottom to prevent the build-up of algae and other marine life. of dissolved oxygen. This VF represent the seasonal impact of discharge and temperature. This factor can be atttributed to seasonal changes. This fact is also supported by studies of Shrestha and Kazama (2007)[16].

VF6 explained 4.3% of a strong positive loading of temperature and a strong negative loading VF7 explain 3.9% of a strong positive loading of NO3 which represent agricultural waste based on nitrate that can be found in the Perlis River. Based on Kazama and Yoneyama (2002)[19], this factor represents the contribution of non-point source pollution from paddy field and agricultural areas. In these areas, farmers use the nitrogenous fertilizer, which undergo nitrification processes, and the rivers receive nitrate nitrogent via groundwater leaching.

DOI 10.5963/IJEP0105001

VF8 contributed 3.7% of total variance has one strong negative loading which is oil and grease. This VF represent non-point source pollution and can be assumed as there had an oil waste from restaurant in Perlis River. Fast food restaurants typically produce a low volume but high grease/chemical oxygen demand (COD) containing wastewater, generated by their daily kitchen activities, for which there is currently no acceptable treatment technology. VF9 (3.9% of total variance) has two moderate negative loading of Pb and Cd. Laxen and Harison (1977)[20] said although studies performed before the widespread use of leaded petrol have shown considerable Pb deposition on and near to highways, there can be little doubt that the motor car is now the major source of Pb in the highway environment. According to the land use activities, there is a ferry terminal located at the mouth of the Perlis River. Thus, the ships would have a stop at the terminal which the ships maintenance and equipment repair are believed as the main sources of Pb in water body. Meanwhile, the fossil fuel combustion which occurred during the shipping may lead to the presence of Cd. And thus, this can be attributed to the shipping waste pollution.

TABLE 2 CLASSIFICATION FUNCTION EQ.(1) FOR DISCRIMINANT ANALYSIS OF TEMPORAL VARIATION IN PERLIS RIVER BASIN

Parameters	**Standard mode**			**Forward Stepwise mode**			**Backward Stepwise mode**		
	LPS	**HPS**	**MPS**	**LPS**	**HPS**	**MPS**	**LPS**	**HPS**	**MPS**
DO	2.791	2.497	2.366				2.11	1.83	1.636
BOD	0.209	0.215	0.206						
COD	0.035	0.026	0.026	0.067	0.06	0.06	0.073	0.066	0.064
SS	0.059	0.057	0.042						
pH	59.826	62.144	60.198	40.752	42.775	41.31	39.281	41.793	40.5
NH3-NL	2.977	2.662	5.31	0.319	0.114	3.013	0.75	0.546	3.138
TEMP	10.68	10.699	11.272	8.273	8.302	8.92	9.083	9.084	9.6
COND	-0.032	-0.036	-0.034				-0.01	-0.014	-0.012
SAL	39.642	40.356	34.94				1.983	3.025	-2.027
TUR	-0.039	-0.037	-0.026	0.021	0.023	0.034	0.016	0.014	0.023
DS	0.056	0.062	0.051						
TS	-0.073	-0.073	-0.057				0.002	0.006	0.011
NO3	7.502	8.41	7.291						
Cl	0.026	0.022	0.023	-0.004	-0.008	-0.004			
PO4	-57.82	-55.175	-59.49						
As	-239.7	-232.66	-249.7						
Hg	157692	157807	158942						
Cd	80080	78429	82465						
Cr	-75.08	-73.568	-68.63						
Pb	9378.8	9541.3	9477.8						
Zn	597.49	602.41	600.26						
Ca	-0.324	-0.344	-0.376	-0.12	-0.137	-0.175	-0.076	-0.092	-0.134
Fe	-6.005	-6.66	-5.928						
K	-0.052	0.121	-0.009	-0.284	-0.122	-0.26	0.01	0.145	0.014
Mg	-0.417	-0.595	-0.664	-0.577	-0.718	-0.759	-0.563	-0.716	-0.783
Na	0.115	0.139	0.149	0.093	0.112	0.119	0.114	0.135	0.145
OG	3.163	3.584	2.545						
MBAS	2169.6	2141.2	2066						
E.coli	0	0	0						
Coliform	0	0	0						

DOI 10.5963/IJEP0105001

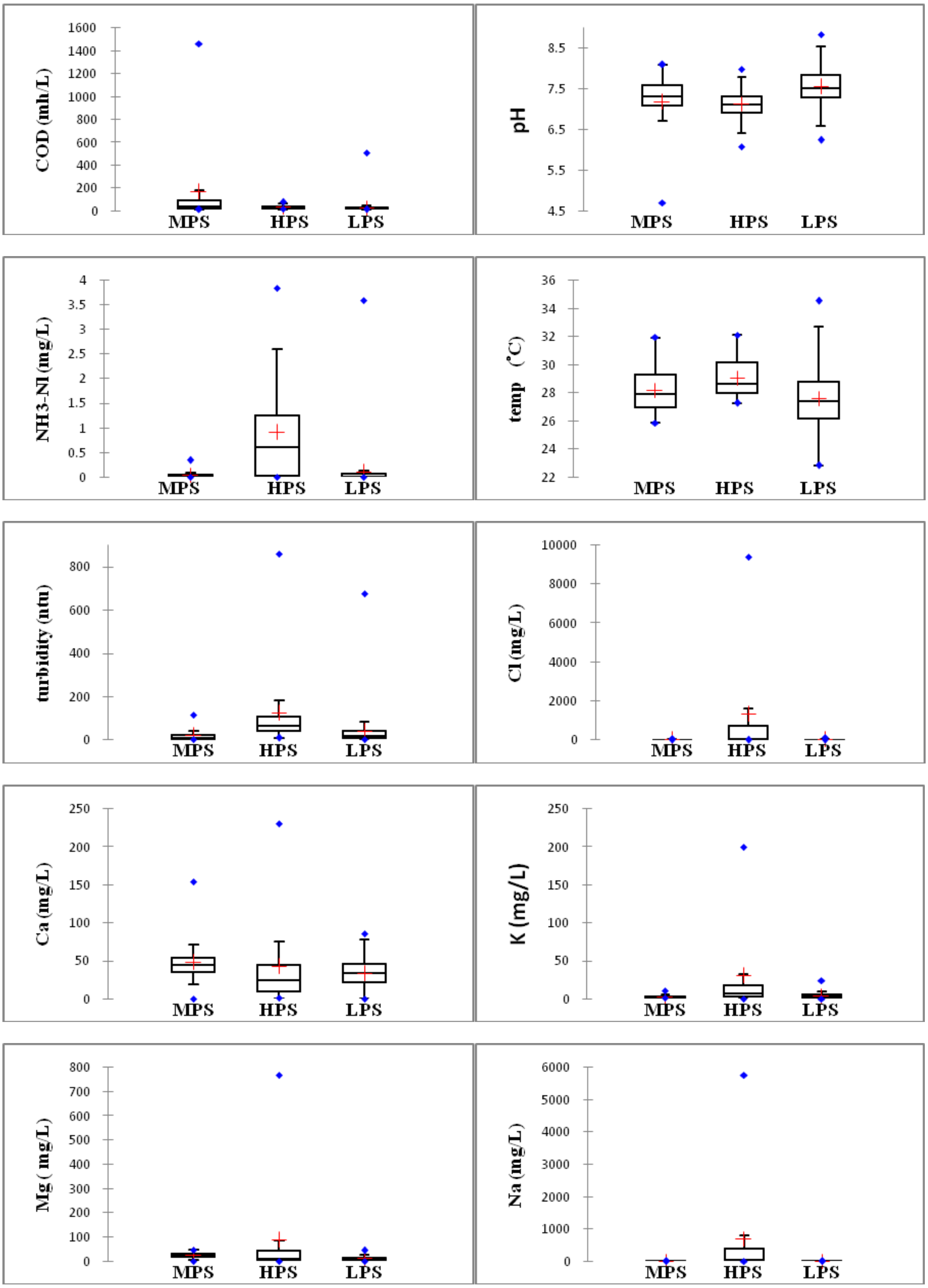

Fig. 3 Box and whisker plots of some parameters separated by spatial DA associated with the water quality data of Perlis River Basin.

DOI 10.5963/IJEP0105001

TABLE 3 LOADINGS OF EXPERIMENTAL VARIABLES (30) ON THE FIRST NINE ROTATED PCS FOR COMPLETE DATA SET

Variables	Mineral Component	Surface Runoff	Anthro - pogenic pollution	Sewage/ Wastewater Pollution	Anti-Fouling Paint	Seasonal Changes	Agriculture Waste	Food Waste	Shipping Waste
DO	-0.136	-0.07	-0.225	-0.139	0.097	-0.775	0.042	-0.121	0.172
BOD	-0.008	0.005	0.957	-0.006	-0.025	0.049	-0.026	0.008	0.027
COD	-0.002	-0.004	0.94	0.002	-0.022	0.027	0.001	-0.021	-0.023
SS	0.036	0.932	-0.032	0.104	0.025	-0.001	-0.026	-0.033	-0.05
pH	-0.084	-0.176	-0.475	-0.048	0.526	-0.269	0.211	-0.164	-0.096
NH3-NL	0.487	0.081	-0.14	0.122	-0.314	0.429	0.078	0.086	0.014
TEMP	0.093	0.087	-0.018	0.022	0.204	0.764	-0.01	-0.253	0.169
COND	0.995	-0.01	-0.005	0.008	-0.009	0.032	-0.005	0.001	-0.001
SAL	0.995	-0.006	-0.003	0.008	-0.009	0.008	-0.004	-0.006	0.005
TUR	-0.025	0.926	-0.024	0.064	0	0.023	0.02	-0.014	-0.081
DS	0.995	-0.008	-0.004	0.006	-0.005	0.025	-0.004	-0.001	0
TS	0.995	0.026	-0.005	0.01	-0.004	0.027	-0.006	-0.001	0
NO3	-0.097	0.076	-0.056	0.001	-0.143	-0.021	0.763	0.071	0.059
Cl	0.99	-0.004	-0.015	0.004	-0.035	0.034	-0.004	0.009	-0.007
PO4	-0.009	0.426	0.078	-0.095	-0.013	0.185	0.157	0.336	0.532
As	0.447	0.006	-0.019	-0.089	0.438	0.102	0.195	0.046	0.293
Hg	0.18	-0.039	-0.011	-0.059	0.348	0.244	0.226	0.161	0.012
Cd	-0.031	-0.007	0.054	-0.095	0.002	0.097	0.39	0.15	-0.6
Cr	-0.024	-0.077	-0.054	0.206	0.732	0.112	-0.294	0.088	0.057
Pb	0.037	0.321	-0.037	-0.123	-0.093	0.041	-0.185	0.142	-0.624
Zn	0.083	0.448	0.22	-0.183	-0.305	0.045	-0.196	0.274	0.274
Ca	0.766	-0.107	0.21	0.056	0.308	0.078	-0.1	0.073	-0.013
Fe	-0.1	0.745	0.126	-0.011	-0.223	0.161	0.082	0.07	0.171
K	0.983	0.018	-0.016	-0.002	0.018	0.08	0.025	0.032	0.034
Mg	0.985	-0.014	0.009	0.001	0.03	0.031	-0.026	-0.008	-0.01
Na	0.993	0.004	-0.011	0.009	-0.037	0.011	-0.016	-0.001	0.003
OG	-0.032	0.005	0.015	-0.093	-0.042	0.08	-0.042	-0.834	0.059
MBAS	-0.028	0.177	0.277	-0.098	-0.092	0.227	-0.414	0.298	0.168
E.coli	0.075	0.096	0.018	0.932	0.031	0.054	0.013	0.015	0.037
Coliform	-0.027	0.03	-0.016	0.94	0.045	0.046	-0.017	0.04	0
Eigenvalue	9.098	3.585	2.299	2.079	1.481	1.288	1.173	1.114	1.027
% Total variance	30.327	11.951	7.664	6.93	4.937	4.294	3.909	3.712	3.422
Cumulative % variance	29.969	39.583	47.251	53.742	58.947	64.548	68.722	72.669	77.146

IV. CONCLUSION

In this study, different multivariate statistical methods were used to evaluate variations and type of water quality of the Perlis River basin. Water quality monitoring programs generate complex data that need multivariate statistical treatment. AHC analysis helped to cluster the eleven sampling stations into three clusters of similar characteristic based on water quality characteristics and pollution sources. Extracted clustering information can be used in reducing the number of sampling sites on the river wihout missing much information for future monitoring program. DA gave the best results for both the temporal and spatial analysis. For eleven different sampling sites of the basin, it yielded an important data reduction, as used only ten parameters (COD, pH, NH3-NL, temperature, turbidity, Cl, Ca, K, Mg and Na) to discriminate between the seasons with 84% correct assignations (83% reduction). Eventhough the FA/PCA did not result in a significant data responsible for variations in

DOI 10.5963/IJEP0105001

river water quality at eleven sampling sites. However, nine varifactors obtained from PCs indicate that the parameters responsible for water quality variations are mainly related to soluble salts (natural) and organic pollution load (anthropogenic).

The multivariate statistical techniques served as an excellent exploratory tool in analysis an interpretation of complex data set on water quality and in understanding their temporal and spatial variation. Additionally, the present of WQI has some disadvantages. It is generally unable to represent the water quality status of specifics locations. There is a need for intergrated approach where spatial analysis is one of the most important aspects. Hence, this study illustrates the environmetric techniques for analysis and interpretation of complex data, water quality assessment, identification of pollution sources, and investigating spatial variations of water quality as an effort toward a more effective river basin management. Since the activity of water quality monitoring, at least in the Perlis river basin, is an expensive endeavour, identifying redundant stations can result in significant cost reduction and allow more effective and efficient river quality management activities. Such information could povide opportunities for better river basin management to control river water pollution in Malaysia.

ACKNOWLEDGEMENT

The authors are gratefully greet to the Department of Environment (DOE) Malaysia for providing the data. Last but not least, the author extremely thanks to the lab members for the support and encorouge for making this paper.

REFERENCES

[1] Brunke, M., Gonser, T., The ecological significance of exchange processes between rivers and ground-water. Freshwater Biol **1997,** 37, 1-33.

[2] Sophocleous, M., Interactions between groundwater and surface water: the state of the science. Hydrogeology Journal **2002,** 10, 52-67.

[3] Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., Smith. V. H., Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecological Applications **1998,** 8, (3), 559–568.

[4] 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.

[5] Helena, B., Pardo, R., Vega, M., Barrado, E., Ferna´ndez, J. M., Ferna´ndez, L., Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. Water Research **2000,** 34, 807-816.

[6] Juahir, H., Zain, M. S.,Yusoff, M. K., Ismail, T. T. H., Samah, A. M. A., Toriman, M. E., Mokhtar, M., Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. Environ Monit Assess **2010b,** 173, 625–641.

[7] Ward, J., Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association **1963,** 58 236-244.

[8] Otto, M., Multivariate methods. In: Kellner, R., Mermet, J. M., Otto, M., Widmer, H. M. (Eds.). WileyeVCH: Weinheim, Germany, 1998.

[9] Juahir, H., Zain, M. S, Aris, A. Z., Yusoff, M. K., Mokhtar, M., Spatial assessment of Langat River water quality using chemometrics. . Journal of Environmental Monitoring **2010c** 12, 287-295.

[10] [Johnson, R. A., Wichern, D. W., Applied multivariate statistical analysis. Prentice-Hall: New Jersey, 1992; Vol. 3rd ed.

[11] Saim, N., Osman, R., Spian, D. R. S. A., Jaafar, M. Z., Juahir, H., Abdullah, M. P., Ghani, F. A., Chemometric approach to validating faecal sterols as source tracer for faecal contamination in water. Water Research **2009,** 43 5023–5030.

[12] 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.

[13] Singh, K. P., Malik, A., Mohan, D., Sinha, S., Multivariate statistical techniques for the evaluation of spatial and temporal variations in water qualityof Gomti River (India)—a case study. Water Research **2004,** 38 3980–3992.

[14] Juahir, H., Zain, M. S., Aris, A. Z., Yusof, M. K., Samah, M. A. A., Mokhtar, M., Hydrological Trend Analysis Due to Land Use Changes at Langat River Basin. Environment Asia **2010a** 3, 20-31.

[15] Kim, J.-O., Mueller, C. W., Introduction to factor analysis: what it is and how to do it. Sage University Press: Newbury Park, 1987.

[16] Shrestha, S., Kazama. F., Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. Environmental Modelling & Software **2007,** 22, 464-475.

[17] Frenzel, S., Couvillion, C., Fecal-Indicator Bacteria in Streams Along a Gradient of Residential Development. JAWRA **2002,** 2 38 (1. February.), 265-273.

[18] Kaushik, A., Kansal, A., Santosh, Meena, Kumari, S., Kaushik, C. P., Heavy metal contamination of river Yamuna, Haryana, India:Assessment by Metal Enrichment Factor of the Sediments. Journal of Hazardous Materials **2009,** 164, 265–270.

[19] Kazama, F., Yoneyama, M., Nitrogen generation in the Yamanashi prefecture and its effects on the groundwater pollution. International Environmental Science 1 **2002,** 5 (4), 293–298.

[20] Laxen, D. P. H., Harrison, R. M., The Highway as a Source of Water Pollution : An Appraisal With The Heavy Metal Lead. Water Research **1977,** 11, 1-11.

DOI 10.5963/IJEP0105001