Environ Monit Assess (2011) 173:625–641
DOI 10.1007/s10661-010-1411-x

Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques

Hafizan Juahir · Sharifuddin M. Zain · Mohd Kamil Yusoff · T. I. Tengku Hanidza · A. S. Mohd Armi · Mohd Ekhwan Toriman · Mazlin Mokhtar

Received: 13 December 2008 / Accepted: 24 February 2010 / Published online: 27 March 2010

Abstract This study investigates the spatial water quality pattern of seven stations located along the main Langat River. Environmetric methods, namely, the hierarchical agglomerative cluster analysis (HACA), the discriminant analysis (DA), the principal component analysis (PCA), and the factor analysis (FA), were used to study the spatial variations of the most significant water quality variables and to determine the origin of pollution sources. Twenty-three water quality parameters were initially selected and analyzed. Three spatial clusters were formed based on HACA. These clusters are designated as downstream of Langat river, middle stream of Langat river, and upstream of Langat River regions. Forward and backward stepwise DA managed to discriminate six and seven water quality variables, respectively, from the original 23 variables. PCA and FA (varimax functionality) were used to investigate the origin of each water quality variable due to land use activities based on the three clustered regions. Seven principal components (PCs) were obtained with 81% total variation for the high-pollution source (HPS) region, while six PCs with 71% and 79% total variances were obtained for the moderate-pollution source (MPS) and low-pollution source (LPS) regions, respectively. The pollution sources for the HPS and MPS are of anthropogenic sources (industrial, municipal waste, and agricultural runoff). For the LPS region, the domestic and agricultural runoffs are the main sources of pollution. From this study, we can conclude that the application of environmetric methods can reveal meaningful information on the spatial variability of a large and complex river water quality data.

Keywords Environmetric · Water quality · Cluster analysis · Discriminant analysis · Principal component analysis · Factor analysis

H. Juahir (✉) · M. K. Yusoff · T. I. T. Hanidza · A. S. M. Armi
Department of Environmental Science, Faculty of Environmental Studies, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
e-mail: hafizan@env.upm.edu.my

A. S. M. Armi
e-mail: armyfor@yahoo.com

S. M. Zain
Department of Chemistry, Faculty of Science, University of Malaya, 56000 Kuala Lumpur, Malaysia

M. E. Toriman
School of Social Development and Environmental Studies, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

M. Mokhtar
Institute of Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

Introduction

The State of Selangor, Malaysia, has a long history of river pollution problems associated with land use changes. The Langat River is one of the principal river draining a densely populated and developed area of Selangor. Over the past 40 years, it has served about half of the population of Selangor and is a source of hydropower and control of flood discharges. More than two third of the Selangor population inhabit the floodplain, which provides highly fertile land for agriculture and land for housing, recreation, and industrial developments. This scenario has brought humans into conflict of harmony between human development and river environment and increases the degree of pollution into river channels. According to Aiken et al. (1982), 42 tributaries in Peninsular Malaysia have been categorized as very polluted, including the Langat River. Until 1999, there were 13 polluted tributaries all over Malaysia with 36 polluted rivers due to human activities such as industry, construction, and agriculture at the tributaries (Department of Environment 1999). In 1990, there were 48 clean rivers compared to only 32 rivers in 1999 that could still be classified as clean (Rosnani 2001).

Almost 60% of the major rivers are regulated for domestic, agricultural, and industrial purposes (Department of Irrigation and Drainage 2001). According to Rosnani (2001), the major pollution sources affecting rivers in Malaysia are sewage disposal, discharges from small- and medium-sized industries that are still not equipped with proper effluent treatment facilities and land clearing and earthworks activities. In 1999, 42% of the river basins were recorded to be polluted with suspended solids (SS) resulting from poorly planned and uncontrolled land clearing activities, 30% with biological oxygen demand (BOD) from industrial discharges, and 28% with ammoniacal nitrogen (AN) from animal husbandry activities and domestic sewage disposal.

Surface water pollution is identified as the major problem affecting the Langat River Basin in Malaysia. The increase of developing areas within the river basin increases pollution loading into the Langat River. As an effort to avoid the Langat River from becoming more polluted, the Department of Environment (DOE) of Malaysia, Ministry of Natural Resources and Environment of Malaysia, has installed telemetric stations along the river basin to continuously monitor its water quality. Based on the water quality data, the water quality index (WQI) was developed to evaluate the water quality status and river classification. WQI provides a useful way to predict changes and trends in the water quality by considering multiple parameters. WQI is formed by six selected water quality variables, namely, dissolved oxygen (DO), BOD, chemical oxygen demand (COD), SS, AN, and pH (DOE 1997).

Rapid urbanization along the Langat River plays an important role in the increase of point source (PS) and non-point source (NPS) pollution loading. The water quality in the basin has been deteriorating over the years, as evidenced from the water quality database compiled for 15 years. The recorded WQI ranged from 58.1 to 75, which corresponds to polluted (WQI, 0–59) and moderately polluted (WQI, 60–80). Based on the average values taken from the 2002 survey (Table 1 and Fig. 1), the major pollutants in the Langat River Basin expressed as percentage of stations exhibiting quality corresponding to class 3 and above are as follows (figures in parenthesis indicate percent sampling stations): AN (94%), TSS and BOD (71%), COD (65%), and DO (53%; UPUM 2002).

From this survey, the NPS pollution is seen as the main contributor to the pollution load compared to PS pollution (UPUM 2002). Due to the exponential increase in urbanization within the Langat River Basin, a review of WQI, is required. In this light, spatial temporal evaluation of river water quality variations along the Langat River is initiated.

Spatial analysis is conducted to evaluate the most significant water quality parameters, taking into account land use activities that affect the health of the river. The transformation of a particular type of land use, such as agriculture and forest area into industrial or municipal area, will change the types of pollutant loadings into the river system. Since there are numerous PS and NPS pollution sources along the Langat River, it is quite a challenge to identify the origin of each pollutant. Regular monitoring by DOE provides the

Table 1 Comparison between total point source and non-point source pollution load contribution in the Langat River Basin based on the event sampling approach

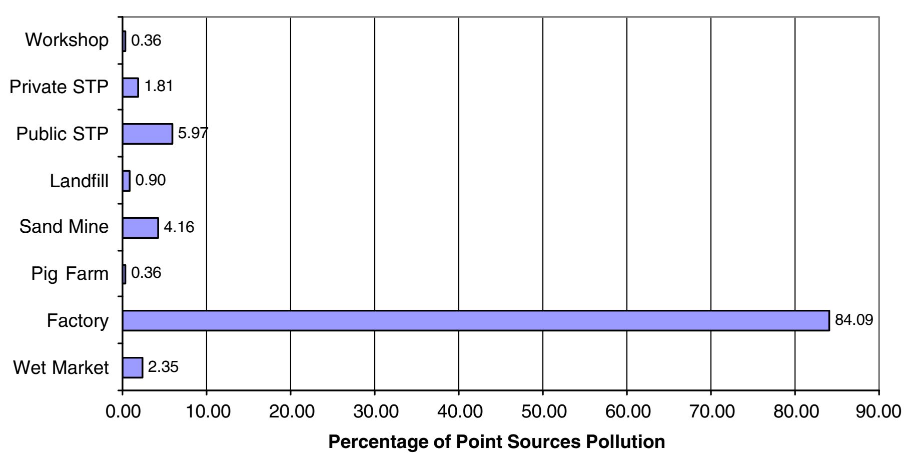
Pollution source	Pollution loading (t/day)				
	COD	BOD	TSS	TM	AN
Industry	22.650	5.536	14.414	0.2440	1.513
Wet market	1.070	0.384	0.409	0.0050	0.056
Pig farm	1.020	0.201	0.460	0.0056	0.725
Public sewage treatment plants	20.680	5.047	5.692	0.0890	3.071
Private sewage treatment plants	4.340	1.060	1.200	0.0200	0.640
Individual sewage treatment plants	17.260	7.140	8.354	0.0000	2.784
Landfill	5.650	1.048	8.149	0.0310	0.664
Sand mining	20.880	0.820	206.670	2.7780	0.057
Total PS	93.550	21.240	245.350	3.1700	9.510
Total NPS (t/day)	614.950	132.690	2,791.030	15.2500	12.670
% PS	13.2	13.8	8	17.2	42.8
% NPS	86.8	86.2	92	82.8	57.2

UPUM (2002)

available spatial variation data. However, interpreting the huge amount of data is a challenge requiring one to use correct methods of data interpretation (Chapman 1992; Dixon and Chiswell 1996).

Environmetrics can be considered as a branch of environmental analytical chemistry that uses multivariate statistical modeling and data treatment also known as chemometrics; (Simeonov et al. 2000). Environmetric is deemed to be the best approach to avoid misinterpretation of a large complex environmental monitoring data (Simeonov et al. 2002). Environmetric methods have been widely used in drawing meaningful information from masses of environmental data. These methods have often been used in exploratory data analysis tools for classification (Brodnjak-Voncina et al. 2002; Kowalkowski et al. 2006) of samples (observations) or sampling stations and the identification of pollution sources (Massart et al. 1997; Vega et al. 1998; Shrestha and Kazama 2007). Environmetrics have also been applied to characterize and evaluate the surface and freshwater quality as well as verifying spatial variations caused by natural and anthropogenic factors (Helena et al. 2000; Singh et al. 2005; Juahir et al. 2008). Recently, environmetric methods have become an important tool in environmental sciences (Brown et al. 1994, 1996) to reveal and evaluate complex relationships in a wide variety of environmental applications (Alberto et al. 2001). The most common environmetric methods used for clustering are the hierarchical agglomerative cluster analysis (HACA) and the principal components analysis (PCA) with factor analysis (FA; Kannel et al. 2007). These methods are

Fig. 1 Percentage contribution of each type of point source of pollution in the Langat River Basin



Source: UPUM (2002)

commonly supported by discriminant analysis (DA) as a confirmation for HACA and PCA and are usually referred to as pattern recognition methods (Adams 1998). The application of different pattern recognition techniques to reduce the complexity of large data sets has proven to give a better interpretation and understanding of water quality data (Brown et al. 1980; Qadir et al. 2007).

The objective of this study is to evaluate the spatial variations in the river water quality data matrix taken from the Langat River (Peninsular Malaysia) using environmetric methods. The data are taken from the river quality monitoring program of 1995–2002. Environmetric methods were used to identify the influence of land use activities on the spatial variations of Langat River water quality. Based on the information obtained from this study, a critique on the WQI methodology will be presented.

Methodology

Description of study area

The Langat River has a total catchment area of approximately 1,815 km^2. It lies within latitudes 2°40′ M 152″ N to 3°16′ M 15″ and longitudes 101°19′ M 20″ E to 102°1′ M 10″ E. The catchment is illustrated in Fig. 2. The Langat River Basin formed by 15 sub-basins, namely, Pangsoon, Hulu Lui, Hulu Langat, Cheras, Kajang, Putrajaya, Hulu Semenyih, Semenyih, Batang Benar, Batang Labu, Beranang, Bangi Lama, Rinching, Teluk Datok, and Teluk Panglima Garang. The main river course length is about 141 km, mostly situated around 40 km east of Kuala Lumpur. The Langat River has several tributaries with the principal ones being the Semenyih River, the Lui River, and the Beranang River. There

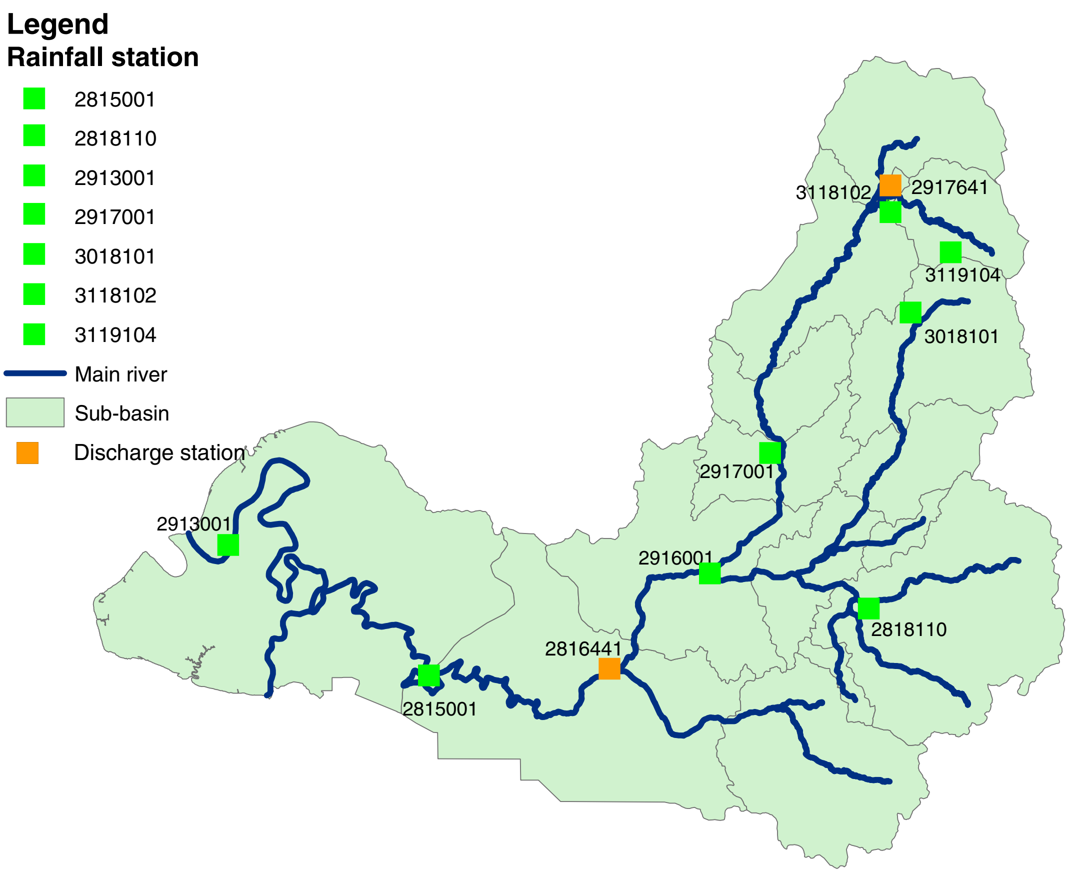


Fig. 2 Seven water quality stations (*Sb*) were selected in this study along the main river

are two reservoirs, the Langat Reservoir and the Semenyih Reservoir catchments, respectively. The Langat Reservoir, built in 1981, has a catchment area of 54 km^2, while the Semenyih Reservoir, built in 1982 with the purpose to supply domestic and industrial water, has a catchment area of 41 km^2. For the Langat Reservoir, it is also used to generate power supply at moderate capacity for the population within the Langat Valley.

This climate in the study area is characterized by high average and uniform annual temperatures, high rainfall, and high humidity. This climate has a dominant impact on the hydrology and geomorphology of the study area. Generally, the study area experiences two types of season: the wet season in April to November and a relatively drier period from January to March. The weather is very much influenced by the SW monsoon that blows across the Straits of Malacca. Flooding is common in the study area.

The study areas include selected impacted and non-impacted stretches of the Sungai Langat Basin. Starting from the Semenyih and Hulu Langat Dams, down to the lowest area at the estuary near Permatang Pasir, with a total distance of 78 km. Due to immense area (1,987.7 km^2), it is crucial that critically impacted areas be identified.

The data and monitoring sites

The water quality data in this study were obtained from seven stations along the main Langat River (Fig. 2). The water quality monitoring stations are manned by the Department of Environment (DOE), Ministry of Natural Resource and Environment of Malaysia. The selected stations are illustrated in Table 2. All the stations were identified based on the availability of recorded data from 1995 to 2002. The data collected from September 1995 to May 2002 were selected for this study. Referring to the sampling sites manned by DOE in the present study, a total of seven sites represent the eight Langat sub-basin, namely, Teluk Panglima Garang (site 7), Teluk Datok (site 6), Putrajaya (site 5), Kajang (site 4), Cheras (site 3), Hulu Langat (site 2), Pangsoon, and Ulu Lui (site 1). Sites 3–7 are located in the region of high pollution load, as there are several wastewater drains situated in the middle and downstream of the Langat River Basin. Site 2 is partly situated in the middle stream region of moderate river pollution. Site 1 and a part of site 2 are located upstream of the Langat River, in the area of relatively low river pollution. It is worth mentioning that some stations have missing data and not all stations were consistently sampled.

Although there are 30 water quality parameters available, only 23 consistently sampled parameters were selected. A total of 254 samples were used for the analysis. The 23 water quality parameters are DO, BOD, electrical conductivity (EC), COD, AN, pH, SS, temperature (T), salinity (Sal), turbidity (Tur), dissolved solid (DS), total solid (TS), nitrate (NO_3^-), chlorine (Cl), phosphate (PO_4^-), zinc (Zn), calcium (Ca), iron (Fe), potassium (K), magnesium (Mg), sodium (Na), *Escherichia coli*, and coliform. The descriptive statistics of the measured 8-year data set are summarized in Table 3.

Table 2 DOE sampling station at the study area

DOE station no.	Study code	Distance from estuary (km)	Grid reference	Location
2814602	Sb07	4.19	2°52.027′ 101°26.241′	Air Tawar Village
2815603	Sb06	33.49	2°48.952′ 101°30.780′	Telok Datuk, near Banting town
2817641	Sb05	63.43	2°51.311′ 101°40.882′	Bridge at Dengkil Village
2918606	Sb04	81.14	2°57.835′ 101°47.030′	Near West Country Estate
2917642	Sb03	86.94	2°59.533′ 101°47.219′	Kajang Bridge
3017612	Sb02	93.38	2°02.459′ 101°46.387′	Junction to Serdang, Cheras at Batu 11
3118647	Sb01	113.99	3°09.953′ 101°50.926′	Bridge at Batu 18

Table 3 Mean values of water quality measurement along the Langat River during 1995–2002

Variable	Station 1 (Sb01)		Station 2 (Sb02)		Station 3 (Sb03)		Station 4 (Sb04)		Station 5 (Sb05)		Station 6 (Sb06)		Station 7 (Sb07)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DO	2.54	1.63	3.85	1.52	3.85	1.52	3.71	1.95	3.03	0.29	6.01	1.19	7.54	0.52
BOD	8.13	13.72	6.28	3.94	6.28	3.94	15.37	15.03	13.31	1.74	8.51	10.51	2.36	1.98
COD	219.27	678.9	44.97	29.24	44.97	29.24	60.64	33.75	60.57	5.76	43.47	39.42	14.3	12.82
SS	357.39	594.33	564.31	603.1	564.31	603.1	528.92	885.68	417.1	115.98	468.5	609.28	27.76	67.84
pH	6.44	0.9	6.45	0.32	6.45	0.32	6.39	0.59	6.62	0.06	6.6	0.23	7.09	0.37
AN	0.83	1.07	1.79	2.44	1.79	2.44	2.02	1.5	2.13	0.19	1.68	1.33	0.12	0.14
Temp.	30.71	1.57	29.06	1.55	29.06	1.55	29.01	1.99	28.78	0.28	28.41	2.02	28.26	1.86
EC	20,128.33	12,275.44	115.91	54.45	115.91	54.45	127.87	77.75	158.43	15.02	82.4	44.75	32.78	7.05
Sal	14.14	7.85	0.06	0.09	0.06	0.09	0.03	0.04	0.05	0.01	0.01	0	0.01	0.02
Tur.	516.3	765.09	721.6	773.76	721.6	773.76	506.61	502.58	346.67	65.59	756.93	1,402.88	13.97	20.9
DS	15,311.19	8,255.71	68.01	31.65	68.01	31.65	77.16	36.75	87.93	6.77	58.72	29.75	30.7	21.03
TS	15,668.59	8,123.29	632.32	596.63	632.32	596.63	606.07	877.67	505.03	113.46	527.22	604.9	58.7	67.64
NO_3^-	0.99	1.55	1.78	2.32	1.78	2.32	0.81	0.92	1.01	0.24	1.66	2.23	0.81	0.74
Cl	8,197.19	4,294.78	7.33	5.12	7.33	5.12	7.11	4.78	10.84	1.45	4.89	3.26	1.22	0.38
PO_4^-	0.07	0.14	0.11	0.18	0.11	0.18	0.11	0.17	0.16	0.07	0.12	0.18	0.09	0.16
Zn	0.05	0.07	0.06	0.1	0.06	0.1	0.04	0.05	0.04	0.01	0.04	0.06	0.04	0.05
Ca	156.91	84.96	8.54	2.92	8.54	2.92	9.07	3.63	11.68	0.74	6.62	3.48	2.29	1.01
Fe	0.48	0.87	1.97	2.16	1.97	2.16	1.92	3.1	1.98	0.81	1.92	2.54	0.34	0.24
K	169.64	100.53	4.54	1.38	4.54	1.38	4.44	1.36	4.69	0.29	3.46	1.03	2.24	0.47
Mg	477.53	293.36	1.16	0.51	1.16	0.51	0.93	0.38	1.22	0.09	0.72	0.38	0.59	0.28
Na	4,240.21	2,424.17	7.79	5.23	7.79	5.23	8.53	4.02	10.22	1.18	6.04	2.9	2.82	1.07
E. coli	1,518.55	4,473.37	31,943.61	48,068.55	31,943.61	48,068.55	87,789.72	77,439.75	83,715.24	12,069.49	54,866	53,593.15	12,477.67	34,289.89
Coliform	24,639.33	104,144.69	92,621	104,669.35	92,621	104,669.35	192,275	245,384.11	164,450	39,805.99	959,673.33	4,353,732.5	109,659.7	289,274.59

Cluster analysis

In this study, HACA was employed to investigate the grouping of the sampling sites (spatial). HACA is a common method to classify (Massart and Kaufman 1983) variables or cases (observations/samples) into clusters with high homogeneity level within the class and high heterogeneity level between classes with respect to a predetermined selection criterion (McKenna 2003). Ward's method using euclidean distances as a measure of similarity (Willet 1987; Adams 1998; Otto 1998) within HACA has proved to be a very efficient method. The result is illustrated by a dendogram, presenting the clusters and their proximity (Forina et al. 2002). The euclidean distance (linkage distance) is reported as $D_{\text{link}}/D_{\text{max}}$, which represents the quotient between the linkage distance divided by the maximal distance. The quotient is usually multiplied by 100 as a way to standardize the linkage distance represented by the y-axis (Singh et al. 2004, 2005; Shrestha and Kazama 2007).

Discriminant analysis

Discriminant analysis determines the variables that discriminate between two or more naturally occurring groups/clusters. It constructs a discriminant function (DF) for each group (Johnson and Wichern 1992). DFs are calculated using Eq. 1:

$$f(G_i) = k_i + \sum_{j=1}^{n} w_{ij} P_{ij} \tag{1}$$

where i is the number of groups (G), k_i is the constant inherent to each group, n is the number of parameters used to classify a set of data into a given group, and w_j is the weight coefficient assigned by DF analysis (DFA) to a given parameter (p_j).

In this study, DA was applied to determine whether the groups differ with regard to the mean of a variable and to use that variable to predict group membership. Three groups for spatial analysis (three sampling regions represents upstream, middle stream, and downstream), which were determined from CA, were selected. The DA was applied to the raw data using the standard, forward stepwise, and backward stepwise modes. These were used to construct DFs to evaluate spatial variations in the river water quality. The stations (spatial) were the grouping (dependent) variables, while all the measured parameters constitute the independent variables. In the forward stepwise mode, variables are included step by step beginning with the most significant variable until no significant changes were obtained. In the backward stepwise mode, variables are removed step by step beginning with the less significant variable until no significant changes were obtained.

Principal component analysis/factor analysis

The most powerful pattern recognition technique that is usually coupled with HACA is the PCA. It provides information on the most significant parameters due to spatial and temporal variations that describes the whole data set by excluding the less significant parameters with minimum loss of original information (Singh et al. 2004, 2005; Kannel et al. 2007). The principle component (PC) can be expressed as

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \ldots + a_{im}x_{mj} \tag{2}$$

where z is the component score, a is the component loading, x is the measured value of the variable, i is the component number, j is the sample number, and m is the total number of variables.

The FA is usually applied as a method to interpret a large complex data matrix and offers a powerful means of detecting similarities among variables or samples (Reghunath et al. 2002). The PCs generated by PCA are sometimes not readily interpreted; therefore, it is advisable to rotate the PCs by varimax rotation. Varimax rotations applied on the PCs with eigenvalues more than 1 are considered significant (Kim and Mueller 1987) in order to obtain new groups of variables called varimax factors (VFs). The number of VFs obtained by varimax rotations is equal to the number of variables in accordance with common features and can include unobservable, hypothetical, and latent variables (Vega et al. 1998). The VF coefficients having a correlation greater than 0.75

are considered as "strong"; 0.75–0.50, as "moderate"; and 0.50–0.30, as "weak" significant factor loadings (Liu et al. 2003).

Source identification of different pollutants was made on the basis of different activities in the catchment area in light of previous literatures. The basic concept of FA is expressed as

$$z_{ij} = a_{f1} f_{1i} + a_{f2} f_{21} + ... + a_{fm} f_{mi} + e_{fi} \quad (3)$$

where z is the measured value of a variable, a is the factor loading, f is the factor score, e is the residual term accounting for errors or other sources of variation, i is the sample number, j is the variable number, and m is the total number of factors.

In this study, PCA/FA was applied to the normalized data sets (23 variables) separately for the three different spatial regions, HPS, MPS, and LPS, as delineated by the CA technique. The input data matrices (variables × cases) for PCA/FA were 23 × 69 for HPS, 23 × 120 for MPS, and 23 × 65 for LPS regions.

Results and discussion

Classification of sampling station based on historical water quality data

This section examines the historical values of water quality parameters in order to classify the water quality station based on their similarity level using HACA. HACA was performed on the water quality data set to evaluate spatial variation among the sampling sites. This analysis resulted in the grouping of sampling stations into three clusters/groups (Fig. 3).

Cluster 1 (stations Sb01 and Sb02) represents the low-pollution source (LPS) from the subbasin Pangsoon and Ulu Lui, cluster 2 (stations Sb03, Sb04, and Sb05) represents the moderate-pollution sources (MPS) from Cheras and Hulu Langat, and cluster 3 (stations Sb06 and Sb07) represents the high-pollution sources (HPS) from sub-basin Kajang, Putrajaya, Teluk Datok, and Teluk Panglima Garang.

This result implies that for rapid assessment of water quality, only one station in each cluster is needed to represent a reasonably accurate spatial assessment of the water quality for the whole network. The CA technique reduces the need for numerous sampling stations. Monitoring from 3 stations that represent three different regions is sufficient. It is evident that the HACA technique is useful in offering reliable classification of surface water for the whole region and can be used to design future spatial sampling strategies in an optimal manner. Figure 4 shows the three regions given by HACA and the possible pollution sources within the study regions. The clustering procedure generated three groups/clusters in a very convincing way, as the sites in these groups have similar characteristics and natural backgrounds.

Spatial variations of river water quality

To study the spatial variation among the different stream regions, DA was applied on the raw data post grouping of the Langat Basin into three main clusters/groups defined by CA. Groups (HPS,

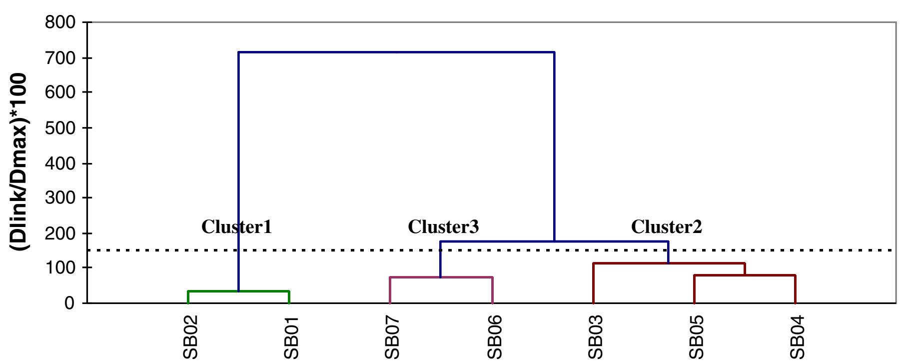


Fig. 3 Dendogram showing different clusters of sampling sites located on the Langat River Basin based on water quality parameters

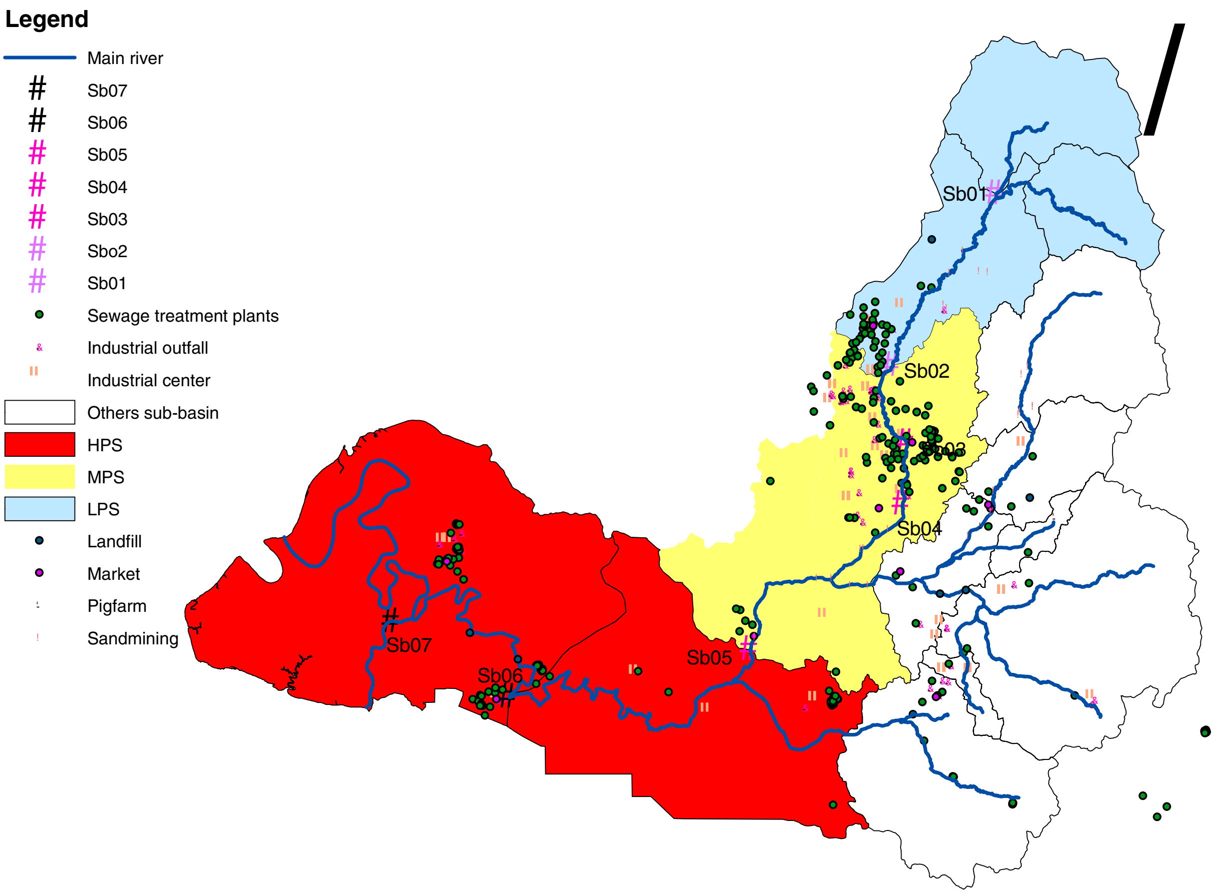


Fig. 4 Classification of regions due to surface river water quality by HACA for the Langat River Basin

MPS, and LPS) were treated as dependent variables, while water quality parameters were treated as independent variables. DA was carried out via standard, forward stepwise, and backward stepwise methods. The accuracy of spatial classification using standard, forward stepwise, and backward stepwise mode DFA were 90.5% (23 discriminant variables), 88.1% (six discriminant variables), and 88.9% (seven discriminant variables), respectively (Table 4). Using forward stepwise, DA, DO, BOD, pH, AN, Cl, and *E. coli* were found to be the significant variables. This indicates that these parameters have high variation in terms of their spatial distribution. Backward stepwise mode on the other hand included coliform as the seventh parameter to have a high spatial variation. Box and whisker plots of three of these water quality parameters over the 8-year period (1995–2002) are shown in Fig. 5. Seven selected water

Table 4 Classification matrix for DA of spatial variations in Langat River

Sampling regions	% Correct	Regions assigned by DA		
		HPS	LPS	MPS
Standard DA mode (23 variables)				
HPS	95.65	66	1	2
LPS	90.63	0	58	6
MPS	87.50	2	13	105
Total	90.51	68	72	113
Forward stepwise mode (6 variables)				
HPS	91.30	63	1	5
LPS	92.19	0	59	5
MPS	84.17	3	16	101
Total	88.14	66	76	121
Backward stepwise mode (7 variables)				
HPS	91.30	63	1	5
LPS	93.75	0	60	4
MPS	85.00	3	15	102
Total	88.93	66	76	111

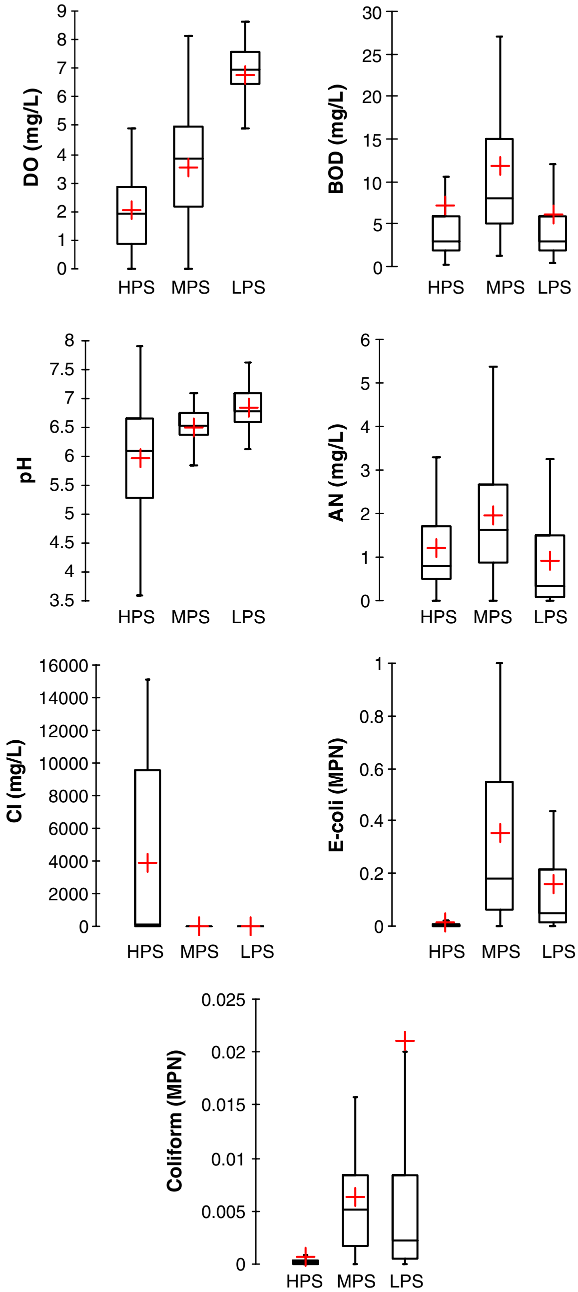


Fig. 5 Box and whisker plots of some parameters separated by spatial DA associated with water quality data of Langat River

quality parameters that gave high variations (the most significant) by backward stepwise DA were then used for further discussion.

PCA was employed on the data set to compare the compositional patterns between the examined water parameters and to identify the factors that influence each of the identified regions (HPS, MPS, and LPS). Seven PCs were obtained for the HPS region; and six PCs, for the other two regions (MPS and LPS), with eigenvalues larger than 1 summing almost 80.8%, 71.0%, and 78.7% of the total variance in the data set, respectively. Equal numbers of VFs were obtained for the three regions via FA performed on the PCs. Corresponding VFs, variable loadings, and variance explained are presented in Table 5.

1. HPS
   For HPS, among the seven VFs, VF1 accounts for 36.9% of the total variance, showing strong positive loadings on EC, DS, TS, Cl, Ca, K, Mg, and Na. This factor contains chemical parameters that are responsible for the water hardness (Ca and Mg) and those attributed to products from anthropogenic activities (DS, TS, and Cl). DS and TS can be identified to originate from both wastewater treatment plants (PS) and NPS pollution sources (USGS 1999; Ha and Bae 2001). The presence of Ca, Mg, K, and Na (Zampella 1995) may be possibly linked with parent rock materials (earth crust) of the HPS region. VF2, explaining 8.2% of the total variance, has strong positive loadings on organic pollution parameters, BOD and AN, thus representing the influence of organic pollutants from point sources such as discharge from wastewater treatment plants, domestic wastewater, and industrial effluents. The presence of BOD and AN in the HPS region, also possibly contributed by pollution loading from pig farms that contribute 200.8 kg/day of BOD and 725.4 kg/day of AN, attributed to the absence of a treatment system (UPUM 2002). VF3, explaining 10.0% of the total variance, has strong positive loadings on SS and Tur. This factor, within the HPS region, can be attributed to runoffs from fields with high load of soil and waste disposal activities.
   VF4 explains 6.2% of the total variance and has strong positive loading on Zn. This can be explained by considering the large number of houses and buildings in the area that use metallic roofs coated with zinc. These, when

Table 5 Loadings of environmental variables on the varimax-rotated PCs for water quality data collected from HPS, MPS, and LPS of the Langat River Basin

Variables	LPS						MPS						HPS						
	VF1	VF2	VF3	VF4	VF5	VF6	VF1	VF2	VF3	VF4	VF5	VF6	VF1	VF2	VF3	VF4	VF5	VF6	VF7
DO	−0.766	−0.227	−0.288	0.1	−0.092	−0.217	−0.649	0.058	0.061	−0.057	0.139	−0.422	0.297	−0.415	−0.362	−0.13	−0.12	−0.353	0.348
BOD	0.442	0.103	0.821	−0.003	0.135	0.066	0.375	−0.046	−0.094	−0.026	−0.191	0.789	0.098	0.887	0.095	−0.094	−0.085	−0.134	0.04
COD	0.376	0.299	0.771	−0.055	−0.024	0.005	0.176	0.352	−0.201	0.11	−0.121	0.686	0.224	−0.056	−0.008	0.009	−0.013	0.056	0.82
SS	0.117	0.948	0.019	0.065	−0.053	−0.014	−0.157	0.951	−0.005	0.04	0.014	0.022	−0.121	0.038	0.95	0.006	−0.004	−0.025	0.027
pH	−0.519	−0.368	0.165	−0.226	0.007	−0.037	0.101	−0.079	−0.036	0.084	0.822	−0.284	0.683	0.148	0.027	−0.108	0.113	0.104	0.271
AN	0.95	0.047	0.109	0.107	0.145	−0.076	0.703	−0.161	−0.032	−0.186	0.082	0.08	−0.212	0.739	−0.071	0.174	0.367	0.156	−0.087
T	0.199	−0.233	0.045	−0.734	0.022	−0.012	0.564	−0.242	−0.123	0.078	−0.45	−0.176	0.43	0.097	0.101	0.441	−0.291	−0.016	0.475
EC	0.869	0.109	0.209	−0.162	0.055	0.2	0.822	−0.173	−0.154	0.025	0.223	0.119	0.941	0.025	−0.111	−0.067	0.003	−0.058	−0.002
Sal	−0.121	−0.02	0.019	−0.147	0.72	0.054	0.385	−0.107	0.655	−0.262	0.118	−0.114	0.239	0.158	−0.011	−0.183	0.346	0.308	0.377
Tur	−0.007	0.878	0.13	−0.043	0.046	−0.015	−0.318	0.707	−0.053	−0.175	−0.082	0.129	−0.167	0.027	0.953	−0.01	0.101	−0.007	−0.009
DS	0.699	0.053	0.3	0.322	−0.009	0.07	0.772	−0.189	0.086	−0.14	−0.025	0.137	0.975	0.029	−0.098	−0.048	−0.028	−0.051	0.079
TS	0.153	0.942	0.034	0.082	−0.053	−0.01	−0.119	0.956	0	0.033	0.013	0.029	0.979	0.032	−0.024	−0.048	−0.029	−0.054	0.082
NO_3^-	0.292	−0.038	−0.224	0.565	0.503	−0.156	−0.103	−0.007	0.865	−0.109	−0.083	−0.199	−0.154	0.38	0.12	0.523	0.623	−0.002	−0.025
Cl	0.91	−0.004	0.059	−0.025	0.048	0.198	0.833	−0.074	−0.041	0.106	0.078	0.114	0.965	0.007	−0.087	−0.064	−0.048	−0.051	0.131
PO_4^-	0.252	−0.097	0.424	0.155	0.722	−0.103	0.123	−0.149	0.219	−0.508	0.22	0.279	0.033	0.348	−0.079	−0.146	0.695	−0.211	0.065
Zn	0.063	−0.141	0.829	0.027	0.081	−0.111	0.043	0.165	0.612	0.252	0.112	0.407	−0.091	−0.057	−0.068	0.838	0.099	0.015	−0.004
Ca	0.857	0.173	0.258	0.039	−0.072	0.262	0.79	−0.087	0.021	0.245	−0.028	0.023	0.966	−0.091	−0.069	0.051	−0.089	0.067	0.059
Fe	0.351	0.223	0.542	0.444	0.068	−0.151	0.071	0.574	0.422	−0.094	−0.181	−0.059	−0.135	−0.135	0.143	0.097	0.857	0.014	−0.066
K	0.884	0.15	0.168	0.144	−0.072	−0.017	0.824	−0.079	0.163	−0.037	−0.051	0.212	0.976	−0.056	−0.088	0.001	−0.018	0.065	0.05
Mg	0.486	−0.055	0.283	0.542	−0.219	−0.024	0.878	−0.002	0.144	0	−0.036	−0.063	0.945	−0.056	−0.068	0.048	−0.071	0.111	0.078
Na	0.898	−0.038	0.152	−0.025	0.052	−0.202	0.822	−0.177	−0.017	−0.155	0.04	0.259	0.97	−0.05	−0.072	0.008	−0.067	0.087	0.06
E. coli	0.527	0.03	−0.185	−0.494	−0.231	−0.258	0.091	−0.022	−0.098	0.806	0.219	0.19	−0.141	−0.042	0.479	−0.312	−0.022	0.333	−0.165
Coliform	0.2	−0.043	−0.062	−0.011	−0.043	0.94	−0.151	−0.099	−0.062	0.58	−0.303	−0.22	0.153	−0.062	−0.035	0.012	−0.096	0.811	0.079
Eigenvalue	8.721	2.834	2.365	1.75	1.25	1.182	7.215	2.983	2.299	1.443	1.263	1.125	8.987	2.718	2.134	1.366	1.254	1.075	1.039
Variability (%)	32.218	13.287	13.026	8.177	6.443	5.556	28.184	13.285	8.456	7	5.567	8.495	36.894	8.172	9.977	6.167	8.827	4.922	5.795
Cumulative (%)	32.218	45.505	58.531	66.707	73.15	78.707	28.184	41.47	49.926	56.926	62.492	70.988	36.894	45.066	55.043	61.21	70.037	74.958	80.754

in contact with acid rainwater and smog, could readily mobilize zinc into the atmosphere and waterways. Moderate positive loading on NO_3^- is possibly due to agricultural runoff because nitrogen and potassium fertilizers are commonly used. The positive correlation of NO_3^- and agricultural land in the HPS region (located downstream) is consistent with many others studies (Hill 1978; Neill 1989; Johnson and Gage 1997; Tufford et al. 1998). There are also weak positive and negative loadings on T, and *E. coli* temperature is most possibly related to seasonal effects, while *E. coli* is strongly related to municipal sewage and wastewater treatment plants (Frenzel and Couvillion 2002) along the river in the HPS region.

VF5 explains 4.9% of the total variance and has strong positive loading on Fe and moderate positive loadings on NO_3^- and PO_4^{3-}. The presence of Fe basically represents the metal group originating from industrial effluents. The presence of NO_3^- and PO_4^{3-} are due to agricultural runoff such as livestock waste and fertilizers (Buck et al. 2003), industrial effluents, municipal sewage, and existing sewage treatment plants because PO_4^{3-} is an important component of detergents (Vega et al. 1998). VF6 explains 4.9% of the total variance and has strong positive loading on coliform, while VF7 explained 5.8% of the total variance with strong loading on COD. These are mainly due to municipal sewage and sewage treatment plants.

2. MPS

In the case of MPS, VF1 explains 28.2% of the total variance and has strong loadings on AN, EC, DS, Cl, Ca, K, Mg, and Na; moderate positive and negative loadings on temperature and DO; weak positive loadings on BOD and Sal; and weak negative loading on Tur. This first factor could be explained by considering the chemical components of various anthropogenic activities that constitute point source pollution especially from industrial, domestic, and commercial and agricultural runoff areas located at the Hulu Langat, Cheras, and Kajang districts. VF2 explains 13.3% of the total variance and shows strong positive loadings on SS, TS, and Tur, which are related to discharge from urban development areas involving clearing of lands (USGS 2007), the erosion of road edges due to surface runoff (Goonetilleke et al. 2005), as well as agricultural runoff (Schlosser and Karr 1981). The conversion of forest or agriculture land to urban areas has indeed caused large negative impacts to the ecosystem (Wahl et al. 1997) of Langat Basin in the form of mud flood, land slide, and river floods. Urbanization is, until now, actively pursued, in line with various developmental plans proposed by the government within the MPS area (Shah et al. 2002). Moderate loading on Fe is possibly generated from industrial activities such as electroplating. Weak loading on COD is related to the discharge of municipal and industrial waste.

VF3 explains 8.5% of the total variance and shows strong positive loading on NO_3^-, moderate positive loading on Zn and Sal, and weak positive loading on Fe. Strong positive loading on NO_3^- is suspected to originate from agricultural fields (Vega et al. 1998), where irrigated horticultural crops are grown and the use of inorganic fertilizers (usually as ammonium nitrate) is rather frequent. This practice could also explain the high levels of ammonia, but this pollutant may also originate from decomposition of nitrogen containing organic compounds via degradation process of organic matters (USGS 2007), such as proteins and urea occurring in municipal wastewater discharges. VF4 and VF5 explain 7% and 5.6% of the variance, respectively, and have strong positive loadings on *E. coli* and pH, which are related to municipal wastes, oxidation ponds, and animal husbandry. The presence of *E. coli* in river water consumes large amount of oxygen, and as the amount of available DO decreases, they undergo anaerobic fermentation processes leading to the production of ammonia and organic acids. Hydrolysis of these acidic materials causes a decrease of water pH values. VF6 explains 8.5% of the variance and has strong positive loadings on BOD and COD that, as explained before, are related to anthropogenic pollution sources and are suspected to come from point sources pollution such as sewage treatment plants and industrial effluents.

3. LPS

Finally, for the LPS region, the first VF (VF1) explains about 32.2% of the total variance and has strong positive loadings on AN, COND, Cl^-, Ca, K, and Na, and moderate positive and negative loadings on *E. coli* and pH. The presence of AN is related to the influence of domestic waste and agricultural runoff (Fisher et al. 2000; Osborne and Wiley 1988). McFarland and Hauck (1999), in their study, found that higher nitrogen levels were detected in agricultural waters, where fertilizers, manure, and pesticide have been applied. Loadings on EC, Cl^-, Ca, K, and Na are probably due to the mineral component of the river water (Barnes et al. 1981; Dahlgren and Singer 1994; Holloway and Dahlgren 2001). This assumption is reasonable, as the water quality in this region is good and land use activities are mostly limited to agriculture and forest areas. Strong negative loadings on DO can be related to high levels of dissolved organic matter consuming large amounts of oxygen, suspected to come from agricultural activities and forest areas that are the dominant land use type within this region.

VF2, which explains 13.2% of the total variance, has strong positive loading on SS, Tur, and TS. Lately, urban developments have been carried out within the LPS region (Siwar et al. 2004), and at times, especially during wet seasons, solids, especially mud, and soil follow runoff into the river. Agricultural runoff also contributes toward this loading as well as construction. Farming and construction were more frequent near stream areas and sediment deposited as a result of these activities. Thus, SS, Tur, and TS increments may be due to overland inputs, increased streambank erosion, and increased entrainment of bedload sediments during stormflow (Bolstad and Swank 1997) especially in forested area (Yusoff and Haron 1999; Yusoff et al. 2006).

VF3 explains about 13% of the total variance and has strong positive loadings on BOD, COD, and Zn. BOD and COD are considered organic factors (Simeonov et al. 2003) and may be interpreted as representing influences from NPS such as agricultural activities and forest areas. Presence of Zn, as explained before, is due to village houses with zinc roofs.

VF4, VF5, and VF6, which explain about 8.2%, 6.4%, and 5.6% of the total variance, respectively, have strong negative loading on temperature and strong positive loadings on Sal, PO_4 and coliform. Salinity and phosphate have their origin in soils due to the use of phosphate fertilizers in this region as well as high salt content. Arheimer and Swank (2000) and Hill (1981), in their studies, conclude that agricultural land use strongly influences stream phosphorous. The presence of coliform is due to discharge into the river via surface runoff of domestic waste and fertilizer (animal waste) used in agricultural activities. According to Bolstad and Swank (1997), the transport of coliform is probably primarily through the soil or direct input by a warm-blooded animal (e.g., livestock). This region, formed by the Pangsoon and Ulu Lui subbasins, is dominated by agriculture, forest, and recreation activities along the river.

The application of WQI for river classification

The DOE started the river monitoring program since 1978. There are 927 manual stations located within 120 river basins throughout Malaysia. Water quality data were used to determine the water quality status and to classify the rivers based on WQI and the Interim National Water Quality Standards for Malaysia. Although WQI is an effective tool for the purpose of monitoring, there have been questions of its relevance in particular parameters, namely, DO, BOD, COD, AN, pH, and SS. Despite the use of AN in the WQI calculation, AN is not listed in the standard A and B, the Environmental Quality Acts. However, recent study reported that 43% of the rivers are polluted with AN (Hashim 2001). SS is another pollutant that is not monitored in the DOE-approved programs, but the level of SS that originated from NPS increases during a storm event. From the critical evaluation of the DOE data, pH has minimal influence of the water quality. Therefore its weightage in the WQI calculation should be reduced (Azni Idris et al. 2003; Juahir et al. 2004).

Results from our study indicate that land use activities significantly influence water quality variations. Looking at the WQI, land use-based parameters have not been considered in the equation. Taking these into consideration, we feel that to better classify the river, parameters should be reviewed. For example, within MPS and HPS, which industrial activity dominates, there is COD contribution. COD exhibited positive correlation (0.82 and 0.7) for VF7 and VF6, which explained only 5.8% and 8.5% of the total variance for both regions. This study agreed, as reported by Juahir et al. (2004), where upon using multiple linear regression and artificial neural networks methods in predicting WQI at the Langat River, COD contributed little to the variance of the WQI. Hence, the WQI is not accurate in representing the surface water quality indicator for this region. WQI is less sensitive to the changes in pollutant types. Due to rapid changes in technology as well as a more diverse use of chemical, pollutant changes with respect to space and time are more drastic. WQI, if not revised, is unable to capture these drastic changes. We recommended that parameters used in the WQI equation be revised.

Conclusions

In this study, environmetric techniques were used to investigate the spatial variations of surface river water quality data of the Langat River. HACA successfully classified the seven monitoring stations into three different cluster regions namely DLR, MLR and ULR. With this classification, optimal sampling strategy can be designed, which could reduce the number of sampling stations and associated costs. For spatial variations, DA gives encouraging results in discriminating the seven monitoring station with six and seven discriminant variables assigning 89% cases correctly using forward and backward stepwise modes. Application of the PCA coupled with the FA on the available data for each of the identified regions resulted in seven parameters responsible for major variations in surface water quality along the Langat River contributed by the regions (DLR, MLR, and ULR). The main sources of variations come from industrial effluents, wastewater treatment plants, and domestic and commercial areas. For MLR and ULR regions, six parameters were found to be responsible for the major variations. The sources of these variations at MLR are mainly attributed to industrial areas, waste water treatment plants, domestic and commercial areas, and agricultural runoffs. For the ULR region, the major source of variations mainly comes from domestic, commercial, agricultural runoff, and forest areas. For a better Langat River Basin management, review of surface water quality changes within the developing regions of MLR and DLR was compared to that of the ULR region.

Additionally, the present WQI has some drawbacks. It is generally unable to represent the water quality status of specifics locations. There is a need for a holistic approach where spatial analysis is one of the most important aspects. Thus, this study illustrates the of environmetric techniques for the 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.

Acknowledgements The authors acknowledge the financial and technical support for this project provided by the Ministry of Science, Technology and Innovation and Universiti Putra Malaysia under the Science Fund project no. 01-01-04-SF0733. The authors wish to thank the Department of Environment and the Department of Irrigation and Drainage, the Ministry of Natural Resources and Environment of Malaysia, the Institute for Development and Environment (LESTARI), the Universiti Kebangsaan Malaysia, the Universiti Malaya Consultancy Unit (UPUM), and the Chemistry Department of Universiti Malaya, who have provided us with secondary data and valuable advice.

References

Adams, M. J. (1998). The principles of multivariate data analysis. In P. R. Ashurst & M. J. Dennis (Eds.), *Analytical methods of food authentication* (p. 350). London: Blackie Academic & Professional.

Aiken, R. S., Leigh, C. H., Leinbach, T. R., & Moss, M. R. (1982). *Development and environment in Peninsular Malaysia*. Singapore: McGraw-Hill International Book Company.

Alberto, W. D., Pilar, D. M. D., Valeria, A. M., Fabiana, P. S., Cecilia, H. A., et al. (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Squia River Basin (Cordoba-Argentina). *Water Research, 35*, 2881–2894. doi:10.1016/S0043-1354(00)00592-3.

Arheimer, P. V., & Swank, W. T. (2000). Nitrogen and phosphorus concentrations from agriculture catchments—Influence of spatial and temporal variables. *Journal of Hydrology (Amsterdam), 227*(1–4), 140–159. doi:10.1016/S0022-1694(99)00177-8.

Barnes, I., Kistler, R. W., Mariner, R. H., & Presser, T. H. (1981). *Geochemical evidence on the nature of the basement rocks of the Sierra Nevada, California*. U.S. Geological Survey Water Supply Paper, 2181.

Bolstad, P. V., & Swank, W. T. (1997). Cumulative impacts of land use on water quality in a southern Appalachian watershed. *Journal of the American Water Resources Association, 33*(2), 519–534. doi:10.1111/j.1752-1688.1997.tb03529.x.

Brodnjak-Voncina, D., Dobcnik, D., Novic, M., & Zupan, J. (2002). Chemometrics characterization of the quality of river water. *Analytica Chimica Acta, 462*, 87–100. doi:10.1016/S0003-2670(02)00298-2.

Brown, S. D., Blank, T. B., Sum, S. T., & Weyer, L. G. (1994). Chemometrics. *Analytical Chemistry, 66*, 315R–359R. doi:10.1021/ac00084a014.

Brown, S. D., Skogerboe, R. K., & Kowalski, B. R. (1980). Pattern recognition assessment of water quality data: Coal strip mine drainage. *Chemosphere, 9*, 265–276. doi:10.1016/0045-6535(80)90003-X.

Brown, S. D., Sum, S. T., & Despagne, F. (1996). Chemometrics. *Analytical Chemistry, 68*, 21R–61R. doi:10.1021/a1960005x.

Buck, O., Niyogi, D. K., & Townsend, C. R. (2003). Scale-dependence of land use effects on water quality of streams in agricultural catchments. *Environmental Pollution, 130*, 287–299. doi:10.1016/j.envpol.2003.10.018.

Chapman, D. (UNESCO, WHO, and UNEP) (1992). *Water quality assessment*. London: Chapman & Hall.

Dahlgren, R. A., & Singer, M. J. (1994). *Nutrient cycling in managed and non-managed oak woodland–grass ecosystems*. Land, Air and Water Resources Research Paper 100028, University of California, Davis, CA.

Department of Environment Malaysia (DOE) (1997). *Malaysia environmental quality reports, 1999*. Kuala Lumpur: Ministry of Science, Technology and Environment.

Department of Environment Malaysia (DOE) (1999). *Malaysia environmental quality reports, 1999*. Kuala Lumpur: Ministry of Science, Technology and Environment.

Department of Irrigation and Drainage (DID) (2001). *DID annual report*. Kuala Lumpur.

Dixon, W., & Chiswell, B. (1996). Review of aquatic monitoring program design. *Water Research, 30*, 1935–1948. doi:10.1016/0043-1354(96)00087-5.

Fisher, D. S., Steiner, J. L., Endale, D. M., Stuedemann, J. A., Schomberg, H. H., & Wilkinson, S. R. (2000). The relationship of land use practices to surface water quality in the Upper Oconee Watershed of Georgia. *Forest Ecology and Management, 128*, 39–48. doi:10.1016/S0378-1127(99)00270-4.

Forina, M., Armanino, C., & Raggio, V. (2002). Clustering with dendograms on interpretation variables. *Analytica Chimica Acta, 454*, 13–19. doi:10.1016/S0003-2670(01)01517-3.

Frenzel, S. A., & Couvillion, C. S. (2002). Fecal-indicator bacteria in streams along gradient of residential development. *Journal of the American Water Resources Association, 38*, 265–273. doi:10.1111/j.1752-1688.2002.tb01550.x.

Goonetilleke, A., Thomas, E., Ginn, S., & Gilbert, D. (2005). Understanding the role of land use in urban stormwater quality management. *Journal of Environmental Management, 74*, 31–42.

Ha, S. R., & Bae, M.-S. (2001). Effects of land use and municipal wastewater treatment changes on stream water quality. *Water, Air, and Soil Pollution, 70*, 135–151.

Hashim, D. (2001). Water pollution control in Malaysia—A regulator's perspective. Paper Presented in the Seminar on World Day for Water, 23–24 March 2001, Batu Pahat, Johor.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M., & Fernandez, L. (2000). Temporal evaluation of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. *Water Research, 34*, 807–816. doi:10.1016/S0043-1354(99)00225-0.

Hill, A. R. (1978). Factors affecting the export of nitrate-nitrogen from drainage basins in southern Ontario. *Water Research, 12*, 1045–1057. doi:10.1016/0043-1354(78)90050-7.

Hill, A. R. (1981). Stream phosphorus exports from watersheds with contrasting land uses in southern Ontario. *Water Resources Bulletin, 17*(3), 627–634.

Holloway, J. M., & Dahlgren, R. A. (2001). Seasonal and event-scale variations in solute chemistry for four Sierra Nevada catchments. *Journal of Hydrology (Amsterdam), 250*, 106–121. doi:10.1016/S0022-1694(01)00424-3.

Idris, A., Mamun, A. A., Mohd, A. M. S., & Wan, N. A. S. (2003). Review of water quality standards and practices in Malaysia. *Pollution Research, 22*(1), 145–155.

Johnson, L. B., & Gage, S. H. (1997). Landscape approaches to the analysis of aquatic ecosystems. *Freshwater Biology, 37*, 113–132. doi:10.1046/j.1365-2427.1997.00156.x.

Johnson, R. A., & Wichern, D. W. (1992). *Applied multivariate statistical analysis* (3rd ed.). Prentice-Hall Int.: New Jersey.

Juahir, H., Ekhwan, T. M., Zain, S. M., Mokhtar, M., Zaihan, J., & Ijan Khushaida, M. J. (2008). The use of chemometrics analysis as a cost-effective tool in sustainable utilisation of water resources in the Langat River Catchment. *American-Eurasian Journal of Agricultural & Environmental Sciences, 4*(1), 258–265.

Juahir, H., Sharifuddin, M., Zain, M., Toriman, E., & Mokhtar, M. (2004). Use of artificial neural network in the prediction of water quality index of Langat River Basin. Malaysia. *Jurnal Kejuruteraan Awam, 16*(22), 42–55.

Kannel, P. R., Lee, S., Kanel, S. R., & Khan, S. P. (2007). Chemometric application in classification and assessment of monitoring locations of an urban river system. *Analytica Chimica Acta, 582*, 390–399. doi:10.1016/j.aca.2006.09.006.

Kim, J.-O., & Mueller, C. W. (1987). Introduction to factor analysis: What it is and how to do it. Quantitative applications in the social sciences series. Newbury Park: Sage University Press.

Kowalkowski, T., Zbytniewski, R., Szpejna, J., & Buszewski, B. (2006). Application of chemometrics in river classification. *Water Research, 40*, 744. doi:10.1016/j.watres.2005.11.042.

Liu, C. W., Lin, K. H., & Kuo, Y. M. (2003). Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. *The Science of the Total Environment, 313*, 77–89. doi:10.1016/S0048-9697(02)00683-6.

Massart, D. L., & Kaufman, L. (1983). *The interpretation of analytical data by the use of cluster analysis*. New York: Wiley.

Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., De Jong, S., Lewi, P. J., & Smeyers-Verbeke, J. (1997). *Handbook of chemometrics and qualimetrics: Data handling in science and technology* (Parts A and B, Vols. 20A and 20B). Elsevier: Amsterdam.

McFarland, A. M., & Hauck, S. L. (1999). Relating agricultural land uses to in-stream stormwater quality. *Journal of Environmental Quality, 28*(2), 836–844.

McKenna, J. E., Jr. (2003). An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. *Environmental Modelling & Software, 18*(2), 205–220. doi:10.1016/S1364-8152(02)00094-4.

Neill, M. (1989). Nitrate concentrations in river waters in the south-east of Ireland and their relationship with agricultural practice. *Water Research, 23*, 1339–1355. doi:10.1016/0043-1354(89)90073-0.

Osborne, L. L., & Wiley, M. J. (1988). Empirical relationships between land use/cover and stream water quality in an agricultural watershed. *Journal of Environmental Management, 26*, 9–27.

Otto, M. (1998). Multivariate methods. In R. Kellner, J. M. Mermet, M. Otto, & H. M. Widmer (Eds.), *Analytical chemistry*. Wenheim: Wiley-VCH.

Qadir, A., Malik, R. N., & Husain, S. Z. (2007). Spatio-temporal variations in water quality of Nullah Aik-tributary of the river Chenab, Pakistan. *Environmental Monitoring Assessment, 140*, 43–59.

Reghunath, R., Murthy, S. T. R., & Raghavan, B. R. (2002). The utility of multivariate statistical techniques in hydrogeochemical studies: An example from Karnataka, India. *Water Research, 36*, 2437–2442. doi:10.1016/S0043-1354(01)00490-0.

Rosnani, I. (2001). River water quality status in Malaysia. In *Proceedings national conference on sustainable river basin management in Malaysia*, 13–14 November 2000, Kuala Lumpur, Malaysia.

Schlosser, I. J., & Karr, J. R. (1981). Water quality in agricultural watersheds: Impact of riparian vegetation during base flow. *Water Resources Bulletin, 17*, 233–240.

Shah, A. H. H., Hadi, A. S., & Jahi J. M. (2002). Lembangan Langat Sebagai Pentas Kehidupan. In M. Mokhtar, Shaharudin Idrus, Ahmad Fariz Mohamed, Abdul Hadi Harman Shah, & Sarah Aziz (Eds.), *Langat Basin research symposium 2001. Proceedings of the 2001 Langat Basin research symposium* (pp. 9–20). Institut Alam Sekitar dan Pembangunan (LESTARI).

Shrestha, S., & 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*, 464–475. doi:10.1016/j.envsoft.2006.02.001.

Simeonov, V., Einax, J. W., Stanimirova, I., & Kraft, J. (2002). Envirometric modeling and interpretation of river water monitoring data. *Analytical and Bioanalytical Chemistry, 374*, 898–905. doi:10.1007/s00216-002-1559-5.

Simeonov, V., Stefanov, S., & Tsakovski, S. (2000). Environmetrical treatment of water quality survey data from Yantra River, Bulgaria. *Mikrochimica Acta, 134*, 15–21. doi:10.1007/s006040070047.

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., et al. (2003). Assessment of the surface water quality in Northern Greece. *Water Research, 37*, 4119–4124. doi:10.1016/S0043-1354(03)00398-1.

Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study. *Water Research, 38*, 3980–3992. doi:10.1016/j.watres.2004.06.011.

Singh, K. P., Malik, A., & Sinha, S. (2005). Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques: A case study. *Analytica Chimica Acta, 35*, 3581–3592.

Siwar, C., Shah, A. H. H., Hadi, A. S., Mohamed, A. F., & Idrus, S. (2004). Socioeconomic status and transformation of households in Langat Basin. In M. Mokhtar, Shaharudin Idrus & Sarah Aziz (Eds.), *Ecosystem health of the Langat Basin. Proceedings of the 2003 research symposium on ecosystem of The Langat Basin* (pp. 23–43). Institut Alam Sekitar dan Pembangunan (LESTARI).

Tufford, D. L., McKellar, H. N., & Hussey, J. R. (1998). In-stream non-point source nutrient predictions with land-use proximity and seasonality. *Journal of Environmental Quality, 27*, 100–111.

Universiti Malaya Consultancy Unit (UPUM) (2002). *Final report program Pencegahan dan Peningkatan Kualiti Air Sungai Langat*. Kuala Lumpur.

U.S. Geological Survey (USGS) (1999). *The quality of our nation's waters-nutrients and pesticides*. U.S. Geological Survey Circular 1225.

U.S. Geological Survey (USGS) (2007). Water quality in the Upper Anacostia River, Maryland: Continuous and discrete monitoring with simulations to estimate concentrations and yields, 2003–05. Scientific Investigations Report 2007-5142, USGS, Virginia.

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. doi:10.1016/S0043-1354(98)00138-9.

Wahl, M. H., McKellar, H. N., & Williams, T. M. (1997). Patterns of nutrient loading in forested and urbanized coastal streams. *Journal of Experimental Marine Biology and Ecology, 213*, 111–131. doi:10.1016/S0022-0981(97)00012-9.

Willet, P. (1987). *Similarity and clustering in chemical information systems*. New York: Research Studies Press, Wiley.

Yusoff, M. K., & Haron, A. R. (1999). Water quality status of Air Hitam forest reserve. *Pertanika Journal of Tropical Agricultural Science, 22*(1), 127–129.

Yusoff, M. K., Ramli, M. F., Juahir, H., Mustapha, S., Ismail, M. R., Mat Perak, Z., et al. (2006). Relationship between suspended solids and turbidity of river in forested catchment. *Malayan Forester, 69*(1), 155–162.

Zampella, R. A. (1995). Characterization of surface water quality along a watershed disturbance gradient. *Water Resources Bulletin, 30*, 605–611.