

*Original Research*

# Spatial Variation Assessment of River Water Quality Using Environmetric Techniques

Ang Kean Hua<sup>1</sup>, Faradiella Mohd Kusin<sup>1</sup>, Sarva Mangala Praveena<sup>2</sup>

<sup>1</sup>Department of Environmental Sciences, Faculty of Environmental Studies,  
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor

<sup>2</sup>Department of Environmental and Occupational Health, Faculty of Medicine and Health Sciences,  
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor

*Received: 30 April 2016*

*Accepted: 5 July 2016*

## Abstract

The Malacca River provides water resources, recreational activities, and habitat for aquatic animals, as well as serving as a tourist attraction. Nevertheless, the Malacca has experienced water quality changes as a result of urbanization and population growth. Environmetric techniques incorporating hierarchical cluster analysis (HCA), discriminant analysis (DA), and principal component analysis (PCA) have been applied to understand the spatial variation of water quality in nine sampling locations within the river basin. HCA has grouped the nine sampling locations into two clusters based on physico-chemical and biological water quality data and trace elements in water. DA analysis indicated that temperature, salinity, coliform, EC, DO, BOD, COD, As, Hg, Cd, Cr, and Zn are the most significant parameters that reflect the overall river water quality for discrimination in clusters 1 and 2. PCA resulted in six components in cluster 1 and eight components in cluster 2. Agricultural activities and residential areas are the main sources of pollution within cluster 1, while a sewage treatment plant and industrial activities are the main sources of pollution in cluster 2. This study has provided useful information for identifying and investigating the pollutant sources through the water quality variations in the river. However, continuous evaluation of river water quality will help in greater understanding of river water quality for a more holistic management of the river basin.

**Keywords:** spatial variation, hierarchical cluster analysis, discriminant analysis, principal component analysis, Malacca River water quality

## Introduction

Every day, two million tons of industrial and agricultural waste are discharged globally into water, in which the estimated amount of wastewater produced annually is about 1,500 [1]. The National Geographic Portal [2] has reported that developing countries produce

70% of industrial wastes that are dumped untreated into waters, and that an average of 99 million pounds (45 million kg) of fertilizers and chemicals are used each year. The deterioration of water quality of rivers is due to growing population, rapid urban development, anthropogenic inputs (e.g., municipal and industrial wastewater discharges, agricultural runoff), natural processes (e.g., chemical weathering and soil erosion) [3-5], human and ecological health, drinking water availability, and further economic development [6-8].

---

\*e-mail: angkeanhua@yahoo.com

Since 2008 Malacca state has served as an official historical tourism center [9]. Nevertheless, increased development and urbanization in certain areas within the Malacca River basin has led to undesirable effects toward natural resources such as river water quality. The Malaysian Department of Environment [10] has classified the Malacca as class III, which means slightly polluted and which can adversely affect aquatic species, even leading to death [11-16]. The degradation of water quality has altered species composition and decreased the overall health of aquatic communities within the river basin [17-19]. Therefore, a practical and reliable assessment of water quality is required for sustainable water resource use with respect to ecosystem health and social development concerning prevention and control of water pollution [20-21].

Techniques that include regression analysis, discriminant analysis (DA), hierarchical cluster analysis (HCA), and principle component analysis (PCA) are the best for data classification and modeling so as to avoid misinterpretation of environmental monitoring data [22]. These techniques have advantages of visualization of large amounts of raw analytical measurements and extraction of additional information about possible sources of pollution [23]. Environmetric techniques have also been applied to characterize and evaluate river water quality as well as identifying spatial variation caused by natural and anthropogenic factors [24-25]. Hence, the increase volume of literature on environmetric techniques and applications have proven that HCA, DA, and PCA are practical in various types of hydrochemistry data [26-28]. HCA provides an average of the clusters made by individual participants that represent the result of the group as a whole [22, 29]. HCA organizes observation into discrete classes or groups such that within a group similarity is maximized and among-group similarity is minimized according to some objective criteria [30]. It assesses the relationship within a single set of variables where no attempt is made to define the relationship between a set of independent variables and one or more dependent variables, etc. [31]. Meanwhile, DA is able to discriminate variables between two or more naturally occurring groups [4]. DA is used to identify water quality variables responsible for spatial and temporal variations in river water quality [24, 32]. PCA describes the correlated variables by reducing the numbers of variables and explaining the same amount of variance with fewer variables (principle components). In others words, the goals of PCA are to extract the most important information from the data table, compress the size of the data set by keeping only the important information, simplify the description of the data set, and analyze the structure of the observations and the variables [33].

Therefore, this study was performed to evaluate the spatial variations in river water quality data using environmetric techniques by incorporating HCA, DA, and PCA. HCA was used to classify 20 variables into clusters with respect to the nine sampling stations. Furthermore, DA was used to identify significant variables that have discriminant patterns between clusters provided from HCA

analysis, and PCA was applied to normalize the 20 variables separately based on the clustering in HCA technique and to obtain the pollutant sources based on spatial variation. Hence, the output of environmetric techniques via HCA, DA, and PCA within the river basin will provide valuable insights on spatial variation of pollutants and areas needing attention in future environmental management plans [34]. In fact, a catchment-scale water quality study would be essential for pollutant characterization and to understand the extent to which the water has been contaminated [35-36].

## Materials and Methods

### Study Area

Malacca state is located southwest of peninsular Malaysia with geographical coordinates of 2°23'16.08"N to 2°24'52.27"N and longitude of 102°10'36.45"E to 102°29'17.68"E. The increasing local population has led to increasing public facilities like transportation, healthcare, accommodation, sewage, and water supply services [37]. However, rapid development in the Strait of Malacca has caused several changes, especially from a land-use perspective. Historically, the Strait of Malacca has become the busiest shipping route between China and India, causing most local citizens to live nearer the Malacca River to gain benefits like water and food sources, transportation, and purchase of imported materials or items from abroad. Land use has continuously developed until today, which is in line with the vision and mission of a sustainable tourism sector for the state. Indirectly, this has contributed to economic growth, political changes, and strengthening social and cultural relationships, but has also created environmental consequences – especially regarding water quality of the Malacca River.

### Field Sampling

A total of nine sampling stations were chosen along the Malacca River, where every station is located at the confluence of each sub-basin and the Malacca River within the river basin (Fig. 1). The locations of sampling stations were recorded using a GPS device. The collection of water quality samples was carried out monthly from January to December 2015. The purpose of primary data collection was to obtain recent water quality data and for field data verification. Additionally, secondary data from 2001 to 2012 was obtained from Malaysia's Department of the Environment. The river water quality data consists of the physico-chemical parameters: pH, temperature, electrical conductivity (EC), salinity, turbidity, total suspended solids (TSS), dissolved solids (DS), dissolved oxygen (DO), biological oxygen demand (BOD), chemical oxygen demand (COD), ammoniacal-nitrogen (NH<sub>3</sub>-N), trace elements (i.e., mercury, cadmium, chromium, arsenic, zinc, lead, and iron), and biological parameters (i.e., *Escherichia coli*form and total coliform).

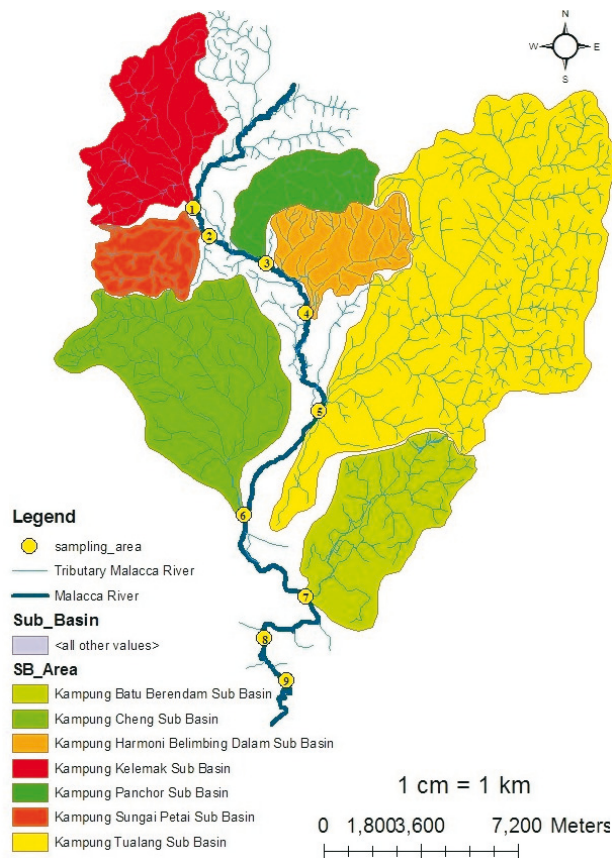


Fig. 1. Map of sampling locations along the Malacca River.

### Water Quality Analysis

*In-situ* measurements include measurement of pH, turbidity, temperature, EC, TDS, salinity, and DO. pH was measured using a SevenGo Duo pro probe (Mettler Toledo AG), turbidity using a portable turbidity meter (Handled Turbidimeter Hach 2100), and temperature, EC, DS, salinity, and DO using a multi-parameter probe (Orion Star Series Portable Meter). On the other hand,  $\text{NH}_3\text{N}$  was analyzed using a spectrophotometer at a specific wavelength using Hach Method 8038, while COD was measured using the APHA 5220B open reflux technique, BOD using APHA 5210B (or Hach Method 8043), and TSS using the APHA 2540D method. Both E-coli and total coliform were analyzed using the membrane filtration method based on APHA 9221B. For trace metal analysis, water samples of 500 mL were filtered through a 0.45  $\mu\text{m}$  Whatman filter paper and acidified with nitric acid ( $\text{HNO}_3$ ) to pH lower than 2, and analyzed using inductive-coupled plasma-mass spectrometry (ICP-MS, ELAN DRC-e, Perkin Elmer, which required 40 MHz in frequency and 1,600 watts for conducting the analysis).

### Quality Assurance and Quality Control

Before conducting the laboratory analysis, the laboratory apparatus and polyethylene bottles were washed using 5% (v/v) of nitric acid and soaking overnight to

remove contaminants and traces of cleaning reagent [38]. On the other hand, for BOD analysis the BOD bottles were wrapped with aluminum foil. The river water samples were preserved using 1% (v/v) nitric acid ( $\text{HNO}_3$ ) for trace metals and analyzed within one month. Each sample was analyzed in triplicate before calculating the mean value, and standard deviation (SD) was used as an indication of the precision of each parameter measured with less than 20%. All the probe meters and instruments used were first calibrated prior to analysis. In all cases, the standards and blanks were treated in the same way as the representative river water samples to minimize matrix interference during analysis. Accuracy of ICP-MS performance is based on the diluting preparation using ICP Multi-Element Mixed Standard III (Perkin Elmer) into concentration with the same acid mixture used for sample dissolution. The recovery of samples for all target elements fell within the standard requirements (90-110%).

### Data Analysis

River water quality data were analyzed using the Statistical Package for Social Sciences version 19 (SPSS 19) for descriptive analysis and environmetric techniques using HCA, DA, and PCA.

### Hierarchical Cluster Analysis

HCA has an advantage of sorting different objects into the same group based on similarities and associations between the objects, which involves several procedures:

1. Ward's method, which uses variance analysis to evaluate the distance between clusters with minimized sum of squares (SS) for any two clusters that are formed at each step [24, 39].
2. Measuring similarity by squared Euclidean distance, which is to provide the similarity between two samples and a distance that can be represented by differences between analytical values from the samples [24, 32, 39].
3. Results from dendrogram that have the ability to group high similarity with small distances between clusters while dissimilarity between groups is represented by the maximum of all possible distances between clusters [40]. In this study, HCA was employed to investigate the grouping of the sampling sites (spatial).

### Discriminant Analysis

Discriminant analysis determines variables that discriminate between two or more groups or clusters. It constructs a discriminant function (DF) for each group [41], which can be defined using:

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

...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  $w_{ij}$  is the weight coefficient assigned by DF analysis (DFA) to a given parameter ( $P_{ij}$ ). In this study, DA was used to define whether the mean of variables differ within the groups and the variables will be used to predict the group pattern. Based on the grouping of HCA results, the raw data are applied into DA using standard, forward stepwise, and backward stepwise modes to develop the DFs in evaluating spatial variations of river water quality. Generally, the stations (spatial) are assigned as dependent variables (referred to as grouping), and all parameters are independent variables.

### Principal Components Analysis

PCA has the ability to provide information on most significant parameters due to spatial and temporal variations that define the whole data set by excluding less significant parameters with minimum loss of original information [4, 24, 32]. PCA can be explained as:

$$Z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \dots + a_{im}x_{mj} \quad (2)$$

...where  $z$  is the component score,  $a$  is 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. General procedures used in PCA are: 1) the hypothesis in an original data group are then reduced to dominant components or factors (source of variation) that influence the observed data variance and 2) the whole data set is extracted through eigenvalues and eigenvectors from the square matrix produced by multiplying the data matrix [42]. Eigenvalues greater than 1 are considered significant enough [43] to perform a new group of variables, namely varimax factors (VFs). VF coefficients that have a correlation greater than 0.75 are considered "strong," 0.75 to 0.50 as "moderate," and 0.50 to 0.30 as "weak" [44] (only factor loadings above 0.6 were taken into account). In this study, PCA was applied to the normalized data set (20 variables) separately based on different spatial regions obtained from the HCA technique.

### Results and Discussion

The mean and standard deviation values of Malacca River water quality data for physico-chemical and biological parameters and the trace elements from 2001 to 2015 are presented in Table 2. Table 2 indicates that most physico-chemical parameters are in class 1, such as temperature (S1 to S9), salinity (S4 to S6 and S8 to S9), electrical conductivity (S2 to S6 and S8 to S9), and dissolved solids (S3 to S6 and S8 to S9) (Table 1). However, there are some stations that continue to be polluted with respect to turbidity (S1 to S6) from class 2 to class 5, and total suspended solid (S2 to S6) from class

1 to class 3; while some stations are classified as being less polluted with respect to dissolved solids (S1 to S3) and salinity (S1 to S4) from class 5 to class 1 (Table 1). The possible causes of changes in increasing turbidity and total suspended solids are the increasing land clearing for agricultural activities, while decreasing in dissolved solids and salinity are potentially due to reduction from big to small-scale animal husbandry activities. Meanwhile, parameters like chemical oxygen demand (S1 to S7), biological oxygen demand (S1 to S7), and dissolved oxygen (S1 to S3 and S7 to S8) were mostly in class 3. There was an increasing level of pollution with respect to ammoniacal nitrogen (S6 to S8) from class 2 to class 5, but improved levels in station 2 to station 4 from class 5 to class 3. The reason for this is because of the changes in land use, i.e., residential activities that change from upstream and the middle parts to the downstream part of the river. All the trace metals are in class 1, but biological parameters are classified as class 5 (Tables 1 and 2).

HCA results showed that two clusters were identified from the nine sampling stations (Fig. 2). Cluster 1 consists of S1, S2, S7, and S8, while cluster 2 consists of S3, S4, S5, S6, and S9. The results showed that S1, S2, S7, and S8 are considered moderate-pollution sources (MPS), while S3, S4, S5, S6, and S9 are considered high-pollution sources (HPS). The areas that constitute MPS are Kampung Kelemak sub-basin (S1), Kampung Sungai Petai sub-basin (S2), and Kampung Batu Berendam sub-basin (S7), while the HPS are from Kampung Panchor sub-basin (S3), Kampung Harmoni Belimbing Dalam sub-basin (S4), Kampung Tualang sub-basin (S5), and Kampung Cheng sub-basin (S6). Based on land uses in the Malacca River basin, the potential sources of pollution within cluster 1 (MPS) result from the widely used land for agricultural activities and residential areas, while in cluster 2 (HPS) the sources may result from effluent discharge from sewage.

Discriminant analysis (DA) was used to further evaluate the spatial variation of two main clusters resulting from HCA output. The results show that spatial classification for both clusters in standard mode are 92% with 20 variables, forward stepwise are 81% with six variables, and backward stepwise are 85% with 12 variables. Therefore, the mode shows that temperature, salinity, coliform, EC, DO, BOD, COD, As, Hg, Cd, Cr, and Zn are found to be the most significant parameters having high variation in terms of their spatial distribution. The results indicate that temperature, salinity, coliform, EC, DO, BOD, COD, Cr, and Zn in cluster 1 have higher values than in cluster 2 (except for As, Hg, and Cd, which have almost similar values in both clusters). Fig. 3 shows box and whisker plots of these water quality parameters for 13 years (2001 to 2012 and 2015).

Principal component analysis was applied to compare composition patterns between the water quality parameters and to determine the factors that influence the identified regions (clusters 1 and 2). In cluster 1, six PCs were obtained with eigenvalues larger than 1 with 54% of total variance, while cluster 2 indicated eight PCs with eigenvalues

Table 1. National Water Quality Standards for Malaysia (source: DOE Malaysia report, 2012).

Category	Unit	Class					
		I	IIA	IIB	III	IV	V
pH	-	6.5-8.5	6 – 9	6 – 9	5 – 9	5 – 9	-
Temp	°C	-	Normal+2°C	-	Normal+2°C	-	-
Sal	%	0.5	1	-	-	2	-
EC	µS/cm	1000	1000	-	-	6000	-
TSS	mg/L	25	50	50	150	300	300
DS	mg/L	500	1000	-	-	4000	-
Tur	NTU	5	50	50	-	-	-
BOD	mg/L	1	3	3	6	12	>12
COD	mg/L	10	25	25	50	100	>100
DO	mg/L	7	5 – 7	5 – 7	3 – 5	< 3	<1
NH <sub>3</sub> N	mg/L	0.1	0.3	0.3	0.9	2.7	>2.7
As	mg/L	-	0.05	0.05	0.4 (0.05)	0.1	-
Hg	mg/L	-	0.001	0.001	0.004(0.0001)	0.002	-
Cd	mg/L	-	0.01	0.01	0.01 (0.001)	0.01	-
Cr	mg/L	-	0.05	0.05	1.4 (0.05)	0.1	-
Pb	mg/L	-	0.05	0.05	0.02 (0.01)	5	-
Zn	mg/L	-	1	1	3.4	0.8	-
Fe	mg/L	-	1	1	1	1(leaf)5(others)	-
Total Coliform	count/100mL	100	5000	5000	5000(20000)	5000(20000)	>50000
Ecoli	count/100mL	10	5000	5000	50000	50000	>50000

Tur means Turbidity; DS means Dissolved Solid; Con means Electrical Conductivity; Sal means Salinity; Temp means Temperature; DO means Dissolved Oxygen; BOD means Biological Oxygen Demand; COD means Chemical Oxygen Demand; TSS means Total Suspended Solids; pH means Acidic or Basic water; NH<sub>3</sub>N means Ammoniacal Nitrogen; E coli means *Escherichia Coliform*; Coli means Coliform; As means Arsenic; Hg means Mercury; Cd means Cadmium; Cr means Chromium; Pb means Lead; Zn means Zinc; Fe means Iron.

more than 1 having 62% of total variance. Corresponding principal components, variable loadings, and variance are explained based on Table 3.

In cluster 1 (Fig. 4a), principal component 1 loadings with 17.3% of total variance have strong positive loadings on DS, EC, and salinity, but moderate negative loadings of NH<sub>3</sub>N. The existence of some physical parameter contaminations can be connected with the erosion of riverbanks due to dredging in the river and the agricultural runoff from non-point source pollution [45]. Meanwhile, salinity and NH<sub>3</sub>N pollution can be connected with pesticide usage for agriculture in oil palm rubber plantations and animal husbandry (chicken, cow, and goat), which are carried out by some local residents along the Malacca River. This condition has resulted in non-point source pollution that leads to surface runoff and water flows into the nearby sub-basin before entering the river. On the other hand, principal component 2 loadings explain strong positive loadings on turbidity but moderate

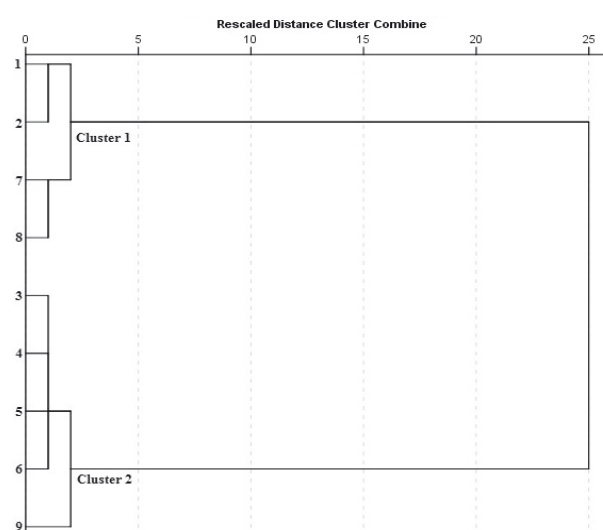


Fig. 2. Hierarchical cluster analysis (HCA) using the Ward linkage method to generate a dendrogram.

Table 2. Mean (and standard deviation) values of water quality data along the Malacca River, 2001-15 (n = 20).

Category	Unit		S1	S2	S3	S4	S5	S6	S7	S8	S9
pH	-	Mean	7.14	6.98	6.78	6.92	6.91	6.88	7.18	7.00	6.85
		SD	0.44	0.49	0.49	0.45	0.38	0.40	0.44	0.44	0.36
Temp	°C	Mean	28.62	29.03	28.99	28.82	28.31	28.32	28.86	29.07	28.33
		SD	1.33	1.72	1.74	1.72	1.49	1.52	1.37	1.67	1.33
Sal	%	Mean	16.14	1.47	0.95	0.09	0.04	0.04	16.13	0.34	0.04
		SD	12.15	2.86	1.89	0.17	0.02	0.02	17.82	1.38	0.018
EC	µS/cm	Mean	22968.82	964.68	867.89	124.10	107.13	182.12	23004.56	733.15	173.63
		SD	17163.09	1538.90	1504.59	89.59	47.17	1004.82	16578.12	2603.44	316.07
TSS	mg/L	Mean	66.12	41.27	91.72	133.45	162.22	201.95	85.19	43.58	149.10
		SD	65.09	54.61	108.28	198.42	204.51	217.57	77.23	41.34	191.06
DS	mg/L	Mean	13709.49	828.52	387.66	74.49	48.17	50.56	12135.32	494.62	59.30
		SD	10390.65	1385.04	612.81	94.79	16.06	33.51	10677.17	1811.15	61.52
Tur	NTU	Mean	72.24±	63.78	259.29	167.16	203.20	239.52	59.84	111.83	187.00
		SD	81.32	80.88	301.95	215.62	287.67	277.92	54.89	121.98	253.93
BOD	mg/L	Mean	5.55	8.94	6.27	3.58	3.63	4.10	5.72	9.46	3.75
		SD	3.69	5.50	4.64	1.94	2.13	2.19	3.79	4.44	2.26
COD	mg/L	Mean	44.63	45.19	33.97	25.76	26.26	26.24	40.28	51.44	25.64
		SD	21.38	20.53	11.89	11.58	11.27	10.19	21.17	19.21	10.22
DO	mg/L	Mean	3.68	3.61	4.76	6.00	6.39	10.41	3.97	4.64	6.33
		SD	1.65	2.10	1.83	1.16	0.82	50.10	1.51	2.59	0.70
NH <sub>3</sub> N	mg/L	Mean	2.42	4.47	2.60	0.48	0.31	0.28	2.16	4.56	0.41
		SD	1.90	2.16	2.24	0.64	0.35	0.27	1.67	2.30	0.36
As	mg/L	Mean	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
		SD	0.03	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00
Hg	mg/L	Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		SD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cd	mg/L	Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
		SD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cr	mg/L	Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		SD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pb	mg/L	Mean	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		SD	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Zn	mg/L	Mean	0.05	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.03
		SD	0.04	0.02	0.02	0.03	0.09	0.03	0.03	0.03	0.03
Fe	mg/L	Mean	0.19	0.41	0.53	0.77	0.83	0.77	0.14	0.52	0.84
		SD	0.29	0.45	0.48	0.59	0.58	0.61	0.24	0.53	0.61
Total Coliform	count/100mL	Mean	227798.72	201677.05	145835.77	136941.03	121105.13	137667.31	218566.67	199536.79	92893.59
		SD	335116.17	355626.54	227435.27	357839.14	205600.13	190798.06	699525.07	280631.55	84720.20
Ecoli	count/100mL	Mean	59652.63	38655.77	19718.63	10639.62	11399.97	22899.17	30401.30	42620.19	15029.76
		SD	71282.59	49068.00	28255.10	30401.26	19905.04	39283.23	41822.87	54646.54	19584.80

Tur means Turbidity; DS means Dissolved Solid; Con means Electrical Conductivity; Sal means Salinity; Temp means Temperature; DO means Dissolved Oxygen; BOD means Biological Oxygen Demand; COD means Chemical Oxygen Demand; TSS means Total Suspended Solids; pH means Acidic or Basic water; NH<sub>3</sub>N means Ammoniacal Nitrogen; E coli means *Escherichia* Coliform; Coli means Coliform; As means Arsenic; Hg means Mercury; Cd means Cadmium; Cr means Chromium; Pb means Lead; Zn means Zinc; Fe means Iron; Max means Maximum; Min means Minimum; SD means Standard Deviation; S1 to S9 means Station 1 to Station 9.

positive loadings on TSS, with a total variance of 8.4%. This can be related to soil erosion – especially interruption from human activities toward hydrologic modifications (such as dredging, water diversions, and channelization) causing disruption in the river [11]. A small percentage of discharge from urban development areas – including land clearings [46] and erosion of road edges due to surface runoff [47] – can also happen within residential areas close to urban areas. Principal component 3 shows moderate positive loadings on COD of 8.1% of total variance,

which is usually related to the discharge of municipal wastes [25]. These results from residential areas located within the Kampung Batu Berendam sub-basin, Kampung Kelemak sub-basin, and Kampung Sungai Petai sub-basin. Principal component 4 loadings indicate that 8% of total variance have moderately positive DO loadings and moderately negative E. coli loadings. The existence of E. coli can originate from animal husbandry and municipal wastes. Principal component 5 loadings described 6.1% of total variance being moderately negative on Cd. The

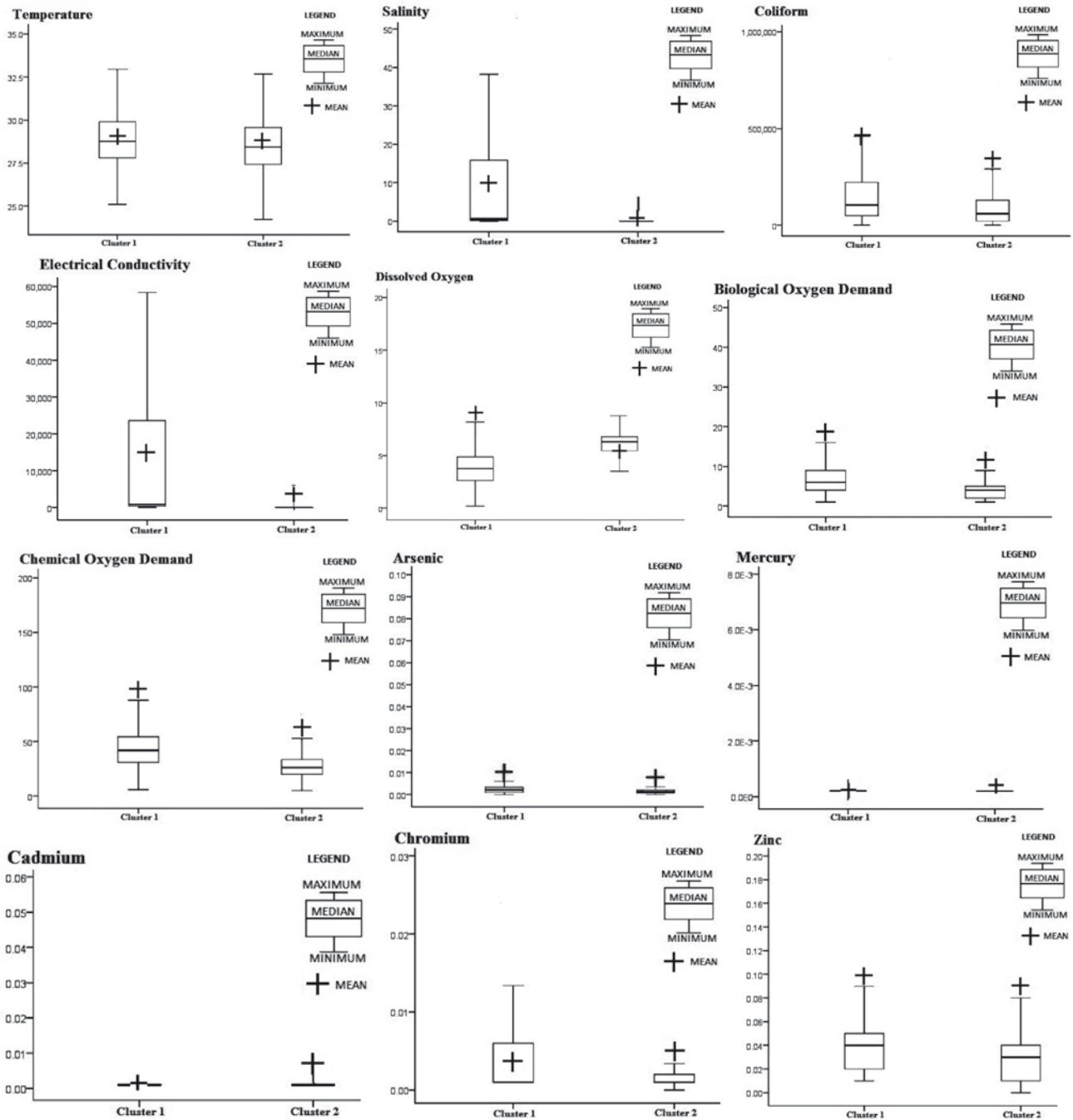


Fig. 3. Plots of some parameters separated from DA associated with water quality data of the Malacca River.

presence of Cd is potentially from agricultural activity through fertilizer applications and leachate from a nearby dumpsite [48]. Based on site observations, we found that there is a dumpsite located near the river in close vicinity to the residential area. Lastly, principal component 6 showed moderate positive loadings of Pb and Zn total variance of 6%. The connection with Zn pollution may be due to the large number of houses and building areas that use metallic roofs coated with Zn, which can be mobilized into the atmosphere and waterways when in contact with acid rain or smog [25], while the existence of Pb can be attributed to agricultural activities [42].

In the case of cluster 2 (Fig. 4b), principal component 1 indicated 15.3% of total variance for strong positive loadings on DO, EC, and salinity, and moderately positive loadings of  $\text{NH}_3\text{N}$ . Principal component 2 loadings showed strong positive turbidity and TSS loadings of 9.2%. As discussed earlier, the principal component 1 and 2 loadings for DO, EC, turbidity, and TSS originate from riverbank erosion and interruption of human activities toward hydrological modifications, causing the river to be polluted through the Kampung Panchor sub-basin. Meanwhile, salinity and  $\text{NH}_3\text{N}$  can be connected with effluent from sewage treatment plants that are located

Table 3. Varimax rotation PCs for water quality data based on two clusters within the Malacca River basin.

Variables (Unit)	Cluster 1						Cluster 2							
	1	2	3	4	5	6	1	2	3	4	5	6	7	8
Turbidity (NTU)	-0.73	<b>.837</b>	.044	.038	.037	.025	.290	<b>.889</b>	.103	.019	-.042	-.007	.000	-.012
Dissolved Solid (mg/L)	<b>.884</b>	-.037	.043	.079	.012	.031	<b>.909</b>	.021	.083	.096	-.015	-.034	.002	-.010
Electrical Conductivity (uS)	<b>.914</b>	.001	.037	.102	.035	.065	<b>.850</b>	.104	-.027	-.007	.007	-.032	-.002	-.022
Salinity (ppt)	<b>.815</b>	.006	-.032	.047	-.035	.023	<b>.772</b>	.041	-.066	.113	-.039	-.056	-.042	-.052
Temperature (°C)	.088	-.067	-.086	.091	-.255	-.105	-.033	.114	-.086	.087	-.270	.008	-.028	.076
Dissolved Oxygen (mg/L)	.057	.003	.002	<b>.729</b>	.216	-.082	-.098	-.057	.272	-.101	.349	-.294	.092	-.586
Biological Oxygen Demand (mg/L)	-.496	.070	.486	.236	-.035	-.141	.357	.076	<b>.685</b>	.158	-.184	-.058	-.039	.059
Chemical Oxygen Demand (mg/L)	-.126	.099	<b>.722</b>	.157	-.124	.086	-.001	.226	<b>.702</b>	-.077	.147	.168	-.056	-.007
Total Suspended Solid (mg/L)	.035	<b>.654</b>	.374	.004	.136	.094	-.193	<b>.914</b>	.120	.051	-.081	.012	.020	-.016
Acidity/Alkalinity (pH)	.297	-.279	.247	.451	.379	.094	-.227	-.266	.310	-.107	-.356	.136	.466	.138
Ammoniacal Nitrogen (mg/L)	<b>-.644</b>	-.429	.278	-.034	.025	-.019	<b>.632</b>	-.209	.483	.104	-.204	.016	-.031	-.007
E-coli (cfu/100ml)	.024	-.042	.022	<b>-.664</b>	.112	.003	.052	.030	.096	<b>.736</b>	-.079	.004	.071	-.048
Coliform (cfu/100ml)	-.093	-.040	-.108	-.475	.370	-.170	.138	.025	-.062	<b>.741</b>	.029	-.077	-.103	-.018
Arsenic (mg/L)	.438	.105	.456	-.116	.057	-.177	.416	-.034	.170	.003	.204	.089	.082	.158
Mercury (mg/L)	.032	.185	.290	-.018	.013	-.254	-.029	-.046	.127	-.113	.113	-.127	.020	<b>.717</b>
Cadmium (mg/L)	.010	-.162	.202	-.062	<b>-.673</b>	.038	.069	.031	-.166	-.081	-.028	.192	<b>.690</b>	-.189
Chromium (mg/L)	.021	-.262	.455	-.272	.236	.233	-.042	-.010	.110	-.120	.004	<b>.780</b>	.045	-.028
Lead (mg/L)	.042	-.086	-.020	.031	.074	<b>.669</b>	.004	-.064	-.034	-.140	-.159	.374	-.589	-.225
Zinc (mg/L)	.067	.252	.046	-.035	-.047	<b>.713</b>	-.026	.056	.010	.324	.477	.506	.068	.068
Iron (mg/L)	-.331	.245	-.117	-.089	-.507	.028	-.065	.019	-.148	.011	<b>.760</b>	.034	-.053	.110
Initial Eigenvalue	3.476	1.686	1.618	1.613	1.221	1.207	3.063	1.840	1.528	1.345	1.322	1.214	1.089	1.021
% of Variance	17.378	8.429	8.089	8.063	6.106	6.036	15.313	9.198	7.638	6.726	6.608	6.072	5.445	5.103
Cumulative %	17.378	25.807	33.896	41.959	48.065	54.101	15.313	24.511	32.149	38.874	45.482	51.554	56.999	62.102

\*The bold values are factor loadings above 0.60 that were taken after Varimax rotation was performed.



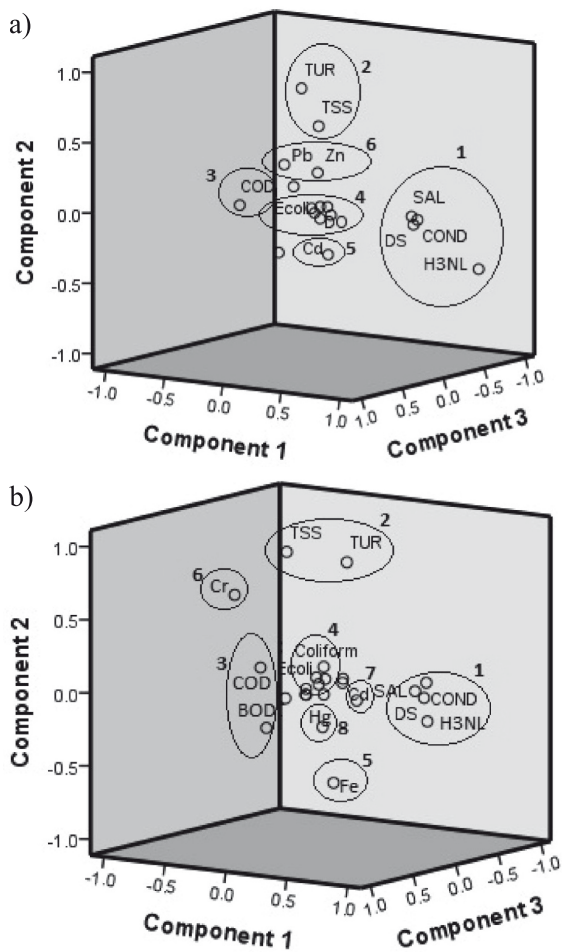


Fig. 4. a) Cluster 1 with six principal components, b) Cluster 2 with eight principal components.

within the Kampung Harmoni Belimbing Dalam sub-basin near the river. Meanwhile, principal component 3 has moderately positive loading on BOD and COD, with total variance of 7.6%. This can be related to anthropogenic sources and possibly comes from point source pollution like sewage treatment plants [25, 35]. Next, principal component 4 loadings highlighted the moderately positive of *E. coli* and coliform loadings with total variance of 6.7%, which are strongly connected with raw and municipal sewage from domestic use, poultry farms, surface runoff, and discharge from wastewater treatment plants [25, 49]. Principal component 5 explained the moderately positive Fe loading with 6.6% total variance, which is suspected from industrial effluents. Principal component 6 explained the strong positive Cr loading with 6.1% total variance, which has a connection with urban storm runoff [50]. Principal component 7 explained the moderately positive of Cd loading with 5.4% total variance and being subjected to leachate from a dumpsite near the residential area [48]. Principal component 8 explained moderately positive of Hg loading with 5.1% total variance, which is suspected of being linked with plastic waste [51]. Generally, principal components 5-8 are subjected to point-source pollution that discharges directly into the river.

## Conclusion

Environmetric techniques (HCA, DA, and PCA) were applied to explore and identify the spatial variation and potential sources of pollution in the Malacca River. HCA categorized the nine sampling stations into two clusters, in which cluster 1 comprises S1, S2, S7, and S8 (indicating MPS), and cluster 2 is S3, S4, S5, S6, and S9 (indicating HPS). The MPS occurs within Alor Gajah sub-basin (S1 and S2) and the middle part of Malacca Central sub-basin (S7 and S8), while the HPS are from the lower part of the Alor Gajah sub-basin and Malacca Central sub-basin. DA analysis showed that temperature, salinity, coliform, EC, DO, BOD, COD, As, Hg, Cd, Cr, and Zn are the most significant parameters reflecting the overall quality of the river water as determined from the backward stepwise mode. PCA showed that six components with 54% of total variance were extracted in cluster 1, while eight components with 62% of total variance were extracted in cluster 2. The major sources of pollution come from agricultural and residential areas along the Malacca, as well as from sewage treatment plants and industrial activities.

This study has provided useful information in identifying the pollution sources. Identification of problematic areas through spatial variation output will help in proper management and understanding of the river water quality within the basin in the coming future. Additionally, the study has also provided a water quality database for future references in developing water and land use policies.

## Acknowledgements

The authors would like to thank the Malaysian Department of the Environment (DOE), the Department of Irrigation and Drainage (JPS), and the Department of Town and Country Planning (JPBD) for providing water quality data, river information, and GIS map-based information – including land use activities in Malacca State.

## References

1. UN WWAP. United Nations World Water Assessment Programme. The World Water Development Report 3: Water in a Changing World. UNESCO: Paris, France, **2009**.
2. National Geographic Portal. Water Pollution. Retrieved from <http://environment.nationalgeographic.com/environment/freshwater/pollution/>, **2016**.
3. HOLLOWAY J.M., DAHLGREN R.A., HANSEN B., et al. Contribution of bedrock nitrogen to high nitrate concentrations in stream water. *Nature*, **395**, 785, **1998**.
4. SINGH K.P., BASANT N., GUPTA S. Support vector machines in water quality management. *Analytical Chimica Acta*, **703**, 152, **2011**.
5. SHIN J.Y., ARTIGAS F., HOBBLE C., et al. Assessment of anthropogenic influences on surface water quality in urban estuary, northern New Jersey: multivariate approach.

- Environmental Monitoring and Assessment, **185**, 2777, **2013**.
6. HOUSER J.N., RICHARDSON W.B. Nitrogen and phosphorous in the UpperMississippi River: transport, processing, and effects on the river ecosystem. *Hydrobiologia*, **640**, 71, **2010**.
  7. LI S., ZHANG Q. Spatial characterization of dissolved trace elements and heavy metals in the upper Han River (China) using multivariate statistical techniques. *Journal of Hazardous Materials*, **176**, 579, **2010**.
  8. MORSE N.B., WOLLHEIM W.M. Climate variability masks the impacts of land use change on nutrient export in a suburbanizing watershed. *Biogeochemistry*, **121**, 45, **2014**.
  9. UNESCO Official Portal. Melaka and George Town, Historic Cities of the Straits of Malacca. Retrieved from <http://whc.unesco.org/en/list/1223>, **2016**.
  10. Department of Environment Malaysia (DOE). Malaysia Environmental Quality Report 2012. Department of Environment, Ministry of Natural Resources and Environment, Kuala Lumpur, Malaysia, **2012**.
  11. DANESHMAND S., HUAT B.B., MOAYEDI H., ALI T.A.M. Study on water quality parameters of Linggi and Melaka rivers catchments in Malaysia. *Engineering Journal*, **15** (4), 41, **2011**.
  12. HUA A.K. An Indication of Policy Study towards Water Resources in Malacca State: A Case Study of Malacca River, Malaysia. *International Research Journal of Social Sciences*, **4** (6), 15, **2015**.
  13. HUA A.K. Law Enforcement and Water Resources of the Malacca River A Case Study of Public Perception. *International Journal of Scientific Research in Science & Technology*, **3** (1), 111, **2015**.
  14. NASBAH N.N. Sungai Melaka Tercemar. Utusan Online. Retrieved from [http://ww1.utusan.com.my/utusan/info.asp?y=2010&dt=0123&sec=Selatan&pg=ws\\_01.htm](http://ww1.utusan.com.my/utusan/info.asp?y=2010&dt=0123&sec=Selatan&pg=ws_01.htm), **January 23, 2010**.
  15. Metro Online. Sampel air sungai dianalisis. Retrieved from <http://www.hmetro.com.my/node/45800?m=1>, **April 23, 2015**.
  16. Sinar Harian Online. Kurang oksigen punca ikan mati di Sungai Melaka. Retrieved from <http://www.sinarharian.com.my/edisi/melaka-ns/kurang-oksigen-punca-ikan-mati-di-sungai-melaka-1.467732>, **January 1, 2016**.
  17. CAMPBELL D., BERGMAN M., BRODY R., KELLER A., LIVINGSTON-WAY P., MORRIS F., WATKINS B. SWIM plan for the Lower St. Johns River Basin. St. Johns River Water Management District. Palatka, FL, **1993**.
  18. DURELL G.S., SEAVEY J.A., HIGMAN J. Sediment quality in the Lower St. Johns River and Cedar-Ortega River Basin: chemical contaminant characteristics. March 2001. Battelle, **2004**.
  19. OUYANG Y., HIGMAN J., THOMPSON J., O'TOOLE T., CAMPBELL D. Characterization and spatial distribution of heavy metals in sediment from Cedar and Ortega rivers subbasin. *Journal of Contaminant Hydrology*, **54** (1), 19, **2002**.
  20. SIMEONOV V., STRATIS J.A., SAMARA C., et al. Assessment of the surface water quality in northern Greece. *Water Research*, **37**, 4119, **2003**.
  21. SHRESTHA S., KAZAMA F. Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji River basin, Japan. *Environmental Modeling and Software*, **22**, 464, **2007**.
  22. BOYACIOGLU H., BOYACIOGLU H. Surface water quality assessment by environmetric methods. *Environmental Monitoring and Assessment* **131**, 371, **2007**.
  23. SIMEONOV V., EINAX J.W., STANIMIROVA I., KRAFT J. Environmetric modeling and interpretation of river water monitoring data. *Analytical and bioanalytical chemistry*, **374** (5), 898, **2002**.
  24. SINGH K.P., MALIK A., SINGH V.K., MOHAN D., SINHA S. Chemometric analysis of groundwater quality data of alluvial aquifer of Gangetic plain, North India. *Analytica Chimica Acta*, **550** (1), 82, **2005**.
  25. JUAHIR H., ZAIN S.M., YUSOFF M.K., HANIDZA T.T., ARMI A.M., TORIMAN M.E., MOKHTAR M. Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environmental monitoring and assessment*, **173** (1-4), 625, **2011**.
  26. ARIS A.Z., PRAVEENA S.M., ABDULLAH M.H., RADOJEVIC M. Statistical approaches and hydrochemical modelling of groundwater system in small tropical island. *Journal of Hydroinformatics* **14** (1), 206, **2011**.
  27. PRAVEENA S.M., KWAN O.W., ARIS A.Z. Effect of data pretreatment procedures on principal component analysis: A case study for mangrove surface sediment datasets. *Environmental Monitoring and Assessment* **184** (11), 6855. doi:10.1007/s1066-1011-2463-2, **2011**.
  28. LIM W.Y., ARIS A.Z., PRAVEENA S.M. Application of the chemometric approach to evaluate the spatial variation of water chemistry and the identification of the sources of pollution inLangatRiver, Malaysia. *Arabian Journal of Geosciences* 1–11. doi:10. 1007/s12517-012-0756-6 (in press), **2012**.
  29. BARCELO D., PETROVIC M. Ebro River Basin: The Handbook of Environmental Chemistry. Heidelberg, Germany: Springer, 431, **2011**.
  30. BOYACIOGLU H., BOYACIOGLU H. Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. *Environmental Geology*, **54** (2), 275, **2008**.
  31. MCGARIGAL K., CUSHMAN S., STAFFORD S. *Multivariate Statistics for Wildlife and Ecology Research*–Springer-Verlag. New York, **2000**.
  32. SINGH K.P., MALIK A., MOHAN D., SINHA S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – a case study. *Water research*, **38** (18), 3980, **2004**.
  33. ABDI H., WILLIAMS L.J. *Principal component analysis*. Wiley Interdisciplinary Reviews: Computational Statistics, **2** (4), 433, **2010**.
  34. HUANG J., HO M., DU P. Assessment of temporal and spatial variation of coastal water quality and source identification along Macau Peninsula. *Stochastic Environmental Research and Risk Assessment* **25**, 353, **2011**.
  35. SHARIF S.M., KUSIN F.M., ASHA'ARI Z.H. Characterization of water quality conditions in the Klang River Basin, Malaysia using self organizing map and K-means algorithm. *Procedia Environmental Sciences*, **30**, 73, **2015**.
  36. KUSIN F.M., ZAHAR M.S.M., MUHAMMAD S.N., MOHAMAD N.D., ZIN Z.M. SHARIF S.M. Hybrid off-river augmentation system as an alternative raw water resources: the hydrogeochemistry of abandoned mining ponds. *Environmental Earth Sciences*, **75** (3), 1, **2016**.
  37. PEARCE D., PEARCE C., PALMER C. (Eds.). *Valuing the environment in developing countries: case studies*. Edward Elgar Publishing, **1**, **2002**.
  38. American Public Health Association (APHA). *Standard Methods for the Examination of Water and Wastewater*. 21<sup>st</sup> ed. Washington: American Water Works Association, Water Environment Federation, **2005**.

39. ALBERTO W.D., PILAR D.M., VALERIA A.M., FABIANA P.S., CECILIA H.A., MARIA B. Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality. A Case Study: Suquia River Basin (Córdoba-Argentina). *Water research*, **35** (12), 2881, **2001**.
40. BIERMAN P., LEWIS M., OSTENDORF B., TANNER J. A review of methods for analyzing spatial and temporal patterns in coastal water quality. *Ecological Indicators*, **11** (1), 103, **2011**.
41. JOHNSON R.A., WICHERN D.W. Applied multivariate statistical analysis (Vol. 4). Englewood Cliffs, NJ: Prentice hall, **1992**.
42. ARIS A.Z., PRAVEENA S.M., ISA N.M., LIM W.Y., JUAHIR H., YUSOFF M.K., MUSTAPHA A. Application of environmetric methods to surface water quality assessment of Langkawi Geopark (Malaysia). *Environmental Forensics*, **14** (3), 230, **2013**.
43. KIM J.O., MUELLER C. W. Introduction to factor analysis: what it is and how to do it. Quantitative applications in the social sciences series. Newbury Park: Sage University Press, **1987**.
44. LIU C.W., LIN K.H., KUO Y.M. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the Total Environment*, **313** (1), 77, **2003**.
45. SCHLOSSER I.J., KARR J.R. Water quality in agricultural watersheds: Impact of riparian vegetation during base flow. *Water Resources Bulletin*, **17**, 233, **1981**.
46. U.S. Geological Survey (USGS). The quality of our nation's water-nutrients and pesticides. U.S. Geological Survey Circular 1225, **1999**.
47. GOONETILLEKE A., THOMAS E., GINN S., GILBERT D. Understanding the role of land use in urban stormwater quality management. *Journal of Environmental Management*, **74**, 31-42, **2005**.
48. World Health Organization (WHO). Guidelines for drinking water quality, (2<sup>nd</sup> ed.), **1**, Recommendations, Geneva, 188, **1993**.
49. FRENZEL S.A., COUVILLION C.S. Fecal-Indicator Bacteria in Streams along A Gradient of Residential Development. *Journal of the American Water Resources Association*, **38** (1), 265, **2002**.
50. SAAD F.N.M., RAHMAN N.N.N.R., KADIR M.O.A., OMAR F.M. Identification of pollution sources within the Sungai Pinang river basin. In Proceeding of the Malaysian Research Group International Conference, Manchester, United Kingdom 478, **2006**.
51. PAPAIOANNOU A., MAVRIDOU A., HADJICHRISTODOULOU C., PAPASTERGIOU P., PAPPAS O., DOVRIKI E., RIGAS I. Application of multivariate statistical methods for groundwater physicochemical and biological quality assessment in the context of public health. *Environmental monitoring and assessment*, **170** (1-4), 87, **2010**.