Pol. J. Environ. Stud. Vol. 32, No. 1 (2023), 79-97
DOI: *10.15244/pjoes/152975* ONLINE PUBLICATION DATE: 2022-12-05

Original Research

Temporal Seasonal Variations and Source Apportionment of Water Pollution in Melaka River Basin Using Multivariate Statistical Techniques

Ang Kean Hua[1]*, Paran Gani[2]

[1]Geography Programme, Faculty of Social Sciences and Humanities, Universiti Malaysia Sabah (UMS), Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia
[2]School of Applied Sciences, Faculty of Integrated Life Sciences, Quest International University, No. 227, Jalan Raja Permaisuri Bainun, 30250 Ipoh, Perak, Malaysia

Received: 23 May 2022
Accepted: 24 August 2022

Abstract

This study aims to determine the spatial and temporal variation on water quality and to identify the main pollutant sources in Melaka River basin. Based on nine monitored sampling stations, twenty variables of water quality data were used in this study for year 2016. Multivariate statistical techniques include Pearson correlation analysis, principal component analysis (PCA), and cluster analysis (CA) is applied to analyze the water quality data. The results indicate water quality in several stations were over the limit of Malaysian water standards. PCA used to compute the pattern within the analysis of water samples, and recognized on five factors in dry and wet season with the total of variation is 73.22% and 76.58% in the river water data respectively. These suggest concentrations of pollutant sources are associated with point source (municipal and domestic waste, sewage treatment plants, and industrial effluents), non-point source (agriculture, livestock), and natural process (weathering of soil and rock). CA explained rainfalls and other sources including surface runoff and wastewater discharges which become main contamination indicator to relatively affect surface water quality through spatial and seasonal variation. Overall, municipal wastes and industrial effluents required serious treatment due to potential impact on water and potential hazard to consumer's health. Therefore, this study suggests anthropogenic activities in Melaka River basin required wise management.

Keywords: water quality, temporal and spatial variation, multivariate techniques, Melaka River basin

*e-mail: angkeanhua@ums.edu.my

Introduction

Rivers play a major role to supply water resources for consumption of human, industrial activities, as well as culture irrigation. To effectively manage the water resources, accurate information regarding to the river water quality with its variability is required. It is essential to focus on this matter especially the developing country like Malaysia of Melaka River basin, whereby exploitation towards water resources increase due to rapid population growth and these resources are unable to be renewed. Natural processes plus industrial discharges, domestic wastewater, as well as agricultural drainage of anthropogenic activities had indirectly enhanced the chances of deterioration of river water quality. The majority of studies agreed that the source of pollutants in the river is originated from industrial, domestic wastewater, and agricultural runoff [1-4]. Natural processes and anthropogenic activities through temperature and precipitation will be affected by seasonal variations in the river water quality. Different variations of wet and dry seasons will lead to the different attribute of water pollution in the river [5]. Hence, water resources can only be managed by having the frequent monitoring and assessment activities on the river water quality [6]. In the purposes of monitoring the river water quality, sampling stations become an important of information sources on the local area as well as temporal status of water quality in the river. To be more specific, the selected of sampling stations network provide an information regarding on the geographical of temporal condition and evolution which happen to the ecosystem surrounding [7]. Therefore, the most popular and appropriate methods to process and analysis the large data are by using multivariate statistical techniques [1-2, 8].

Nowadays, to understand better the status of water quality and ecological perspective is by using multivariate statistical techniques, which consider favorite approach especially required to deal with a number of monitoring site of spatial and temporal data that exist in greater volume. According to the previous studies, cluster analysis (CA), discriminate analysis (DA), factor analysis (FA), as well as principal component analysis (PCA), which grouped into different techniques of statistical analysis is likely to be apply into this research because of its capability to assess the spatial and temporal variation of water quality in the river, and advantages in determining the possible sources of water pollutant [1-3, 8]. For example, Mohamed et al. [9] and Nasir et al. [10] applied FA and PCA to identify the variation of spatial/temporal of river water quality in Klang River basin, Malaysia. Meanwhile, Aris et al. [11] using CA, DA and PCA techniques into hydrochemical data to investigate the variability of spatial of the surface water quality in Langkawi Geopark (Malaysia). Correlation analysis and PCA/FA techniques had been highlight by Prasanna et al. [12] in the study of heavy metal pollution index and river water quality of Miri city (Malaysia), which is important to determine the potential sources of river water contamination. Moreover, these techniques had been widely used in other studies to classify and clustered the monitored sampling stations as well as to interpret the latent source of river water pollution [13-16]. Based on above literature studies, majority research using correlation analysis to verify relationships between the studied parameter of river water quality; while CA technique is to cluster the monitored stations that have similar characteristics based on water quality parameters; as well as FA/PCA technique is to define the source of water pollutants in the river. Hence, the advantages and benefits of multivariate statistical techniques through correlation analysis, CA, and FA/PCA, are being applied into the water quality of Melaka River.

For the last few decades, various studies have been conducted in Melaka River basin especially in identify the contamination sources, which indicated the river water quality are deteriorated due to anthropogenic activities [1, 8, 17-18]. Specifically, the main factor to initiate depletion of river water quality's vulnerability is due to industrial activities of metal processing, color pigments of textile, and coal combustion; as well as agricultural activities that originated from the usage of fertilisers, manure, and pesticide. Since the water quality of Melaka River continues to deplete, therefore, frequent monitoring environmental program for surface water had been conducted by Department of Environment (DOE) Malaysia within the basin with at least more than two decades, which provided massive and complex data of water quality parameters. Beginning with the long-term monitoring of biological, chemical and physical parameters, generated large amount of data will be problematic in handling, whereby reduction methods using statistical analysis toward the data is required.

Based on the above consideration, the objective of this study is to interpret and determine the river water quality dataset based on the seasonal variations (e.g. dry and wet season) for selected monthly monitored program of the year 2016 in Melaka River basin. In archiving the objective, multivariate statistical techniques of correlation analysis, CA, and PCA/FA is used into seasonal variations data to (1) identify the monitored sampling of similarities plus differences of the stations; (2) verify the surface water quality that contributed by the temporal variations of water quality parameters; as well as (3) determine the potential sources of contamination which affecting the water quality. Expected result from this paper will not only benefits the state government in understanding the different location of main pollution sources, but also advantages in verifyng the evolution of spatial-temporal of water quality in Melaka River basin.

Materials and Methods

Study Area

Located at the south-west of peninsular Malaysia, the Melaka River basin lies between latitude 2°23'16.08"N to 2°24'52.27"N and longitude 102°10'36.45"E to 102°29'17.68"E. The Melaka State is experienced with an equatorial climate of hot and humid condition for the entire year, which having the mean value of maximum and minimum temperature are 32.6ºC (February) and 22.5ºC (August) respectively, with the average annual range between 31.5ºC to 22.9ºC per year [19]. Meanwhile, the mean humidity for daily is range within 75% to 86%, while the average per year is from 60% to 90% [19]. Specifically, the annual rainfall's mean is 162.53 cm, and possible to reach a maximum of 390 cm for the whole year [19]. Overall, the wet season is commonly during May to September due to heavy rainfall that link with the southwest monsoon, while dry season is during November to March [19]. In other words, majority of Melaka population will experience the weather of sunshine in the day as well as rainfall in the evening.

The Melaka River basin covers approximately 670 km^2 of the area with 80 km length of Melaka River flow across the Alor Gajah district and Melaka Tengah district. A 20 km^2 of catchment is located between both districts is a Durian Tunggal Reservoir, which become main water supply for the whole population of Melaka State, including for the industrial and domestic activities. Melaka River plus tributary in the basin are characterized by industrial (metal processing and coal combustion), agricultural (livestock), and residential, as well as commercial activities. Increasing population growth indirectly enhance on the rapid urbanization and uncontrolled development, which resulted an expansion of anthropogenic activities (e.g. residential, industrial, commercial) into northward of 20 km, as well as towards the west and east for 10 km respectively [1, 17]. Continuous unorganized development within the basin is expected to occur, and these circumstances increase wastewater system in Melaka River basin. Therefore, Melaka River with its tributaries becomes final end point to be deposited by the industrial effluents, municipal sewage, and agricultural runoff.

The Monitored Sampling Stations and Water Pollutants

Department of Environment (DOE) Malaysia monitored the river water quality in different sampling stations by measuring multiple parameters at periodic time along the Melaka River as well as its tributaries. In order to determine the river water quality's evolution in Melaka River basin, it is essential to conduct this study to define and elucidate the data received from the DOE department of monitoring stations. Selected nine sampling stations are targeted to collect water samples along Melaka River based on the year of 2016 (Table 1 and Fig. 1). The water quality data is obtained based on seasonal basis viz., during wet season (May to October) as well as during dry season (November to April). Selected stations is based on the contamination vulnerability, for instance, monitored station at river source is suspected to have lower impaired from urban contamination (S1 to S3), while sampling station located at downstream near the city is assumed to have higher impaired by these contaminations (S7 to S9) (Table 1). Considering S4, S5, and S6 are located at sub-urban area; as well as S7, S8, and S9 are situated at urban area, these sampling stations is suspected to have greater contamination which originated from point source and non-point source pollution along the Melaka River. Although S1, S2, and S3 are relatively unaffected by domestic pollution due to the monitored stations are at upstream, however, it is assumed to be exposed to the nonpoint source pollution in the river basin.

The twenty parameters of water quality are used in this study, namely physical parameter (i.e. pH, temperature, electrical conductivity (EC), salinity, turbidity, total suspended solids (TSS), total dissolved solid (TDS)), chemical parameter (i.e. dissolved oxygen (DO), biochemical oxygen demand (BOD),

Table 1. The latitude and longitude of sampling stations.

Station	Latitude	Longitude	Regions
S1	02°21'57.41"N	102°13'7.10"E	Rural Area
S2	02°21'30.16"N	102°13'18.20"E	Rural Area
S3	02°20'49.52"N	102°14'36.44"E	Rural Area
S4	02°19'41.70"N	102°15'27.30"E	Sub-urban Area
S5	02°17'48.86"N	102°15'39.51"E	Sub-urban Area
S6	02°15'46.55"N	102°14'10.72"E	Sub-urban Area
S7	02°14'5.02"N	102°15'24.67"E	Urban Area
S8	02°13'14.33"N	102°14'35.01"E	Urban Area
S9	02°12'23.42"N	102°15'0.80"E	Urban Area

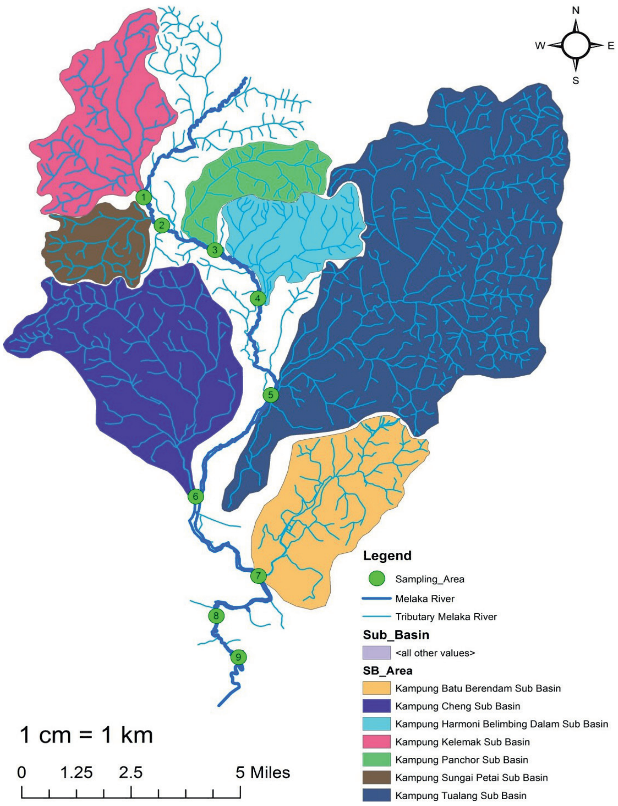

Fig. 1. Sampling stations along Melaka River basin.

chemical oxygen demand (COD), ammoniacal nitrogen (NH_3N)), trace metal elements (i.e. mercury (Hg), cadmium (Cd), chromium (Cr), arsenic (As), zinc (Zn), lead (Pb), iron (Fe)), and biological parameter (i.e. total coliform and *Escherichia coli* form (*E. coli*)). The monthly monitored stations in 2016 are obtained from DOE Malaysia, and the samples are analysed using APHA [30] method. Specifically, onsite measurement involved with pH are measured using a SevenGo Duo pro probe (Mettler Toledo AG); turbidity with a portable turbidity meter (Handled Turbidimeter Hach 2100); as well as temperature, EC, DS, salinity, and DO using a multi-parameter probe (Orion Star Series Portable Meter). Meanwhile, NH_3N 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 µm Whatman filter paper and acidified with nitric acid (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). In this study, all statistical data were made using IBM SPSS Ver. 23. Overall, the statistical data of 108 dataset (12 data set per stations x 9 stations x 1 year) and total number of 2160 dataset (12 data set per stations x 9 stations x 1 year x 20 variables). The overall data point for missing value is very small, which is approximately less than 5%. Nearest neighbour method is identify two and closest point distance, whereby the gaps between two points are used to overcome all the missing value [20-21]. Therefore, the nearest neighbour is applied in this study to facilitate the data analysis. The equation is shown as Eq. (1);

$$y = y_2 \; if \; x \leq x_1 + \left(\frac{x_2 - x_1}{2}\right) \quad \text{or} \quad y = y_1 \; if \; x > x_1 + \left(\frac{x_2 - x_1}{2}\right) \qquad (1)$$

where, y refers to interpolant, x refers to interpolant's time point, y_1 and x_1 refer to coordinate of the gap for beginning points, and y_2 and x_2 refer to coordinate of the gap for endpoints.

Apart from evaluating the relationship between water quality variables, Pearson correlation analysis was employed in the dataset will also benefits in minimizes the effect of between-sampling-station correlation relationships campaigns. When the p value is less than 0.05 at the significant, a correlation with value closets to 1 or -1 indicate relationship with strong positive or negative between two variables, respectively; whereas the number is near to 0 shows no linear relationship for the two variables involve [22]. It is important to assign the similar behavior and uniform distribution between the variables before the coefficient's correlation is examined [23]. Then, Kaiser-Meyer-Olkin (KMO) and Bartlett's Sphericity tests are used to check the normality data whether the variable is appropriate for factor analysis. KMO are able to compare the correlation values among variables as well as the partial correlation. If the value is near to 1 in KMO, the variable in PCA is consider relevant; and if the value is near to 0, the variable is considered non suitable. However, the KMO satisfactory value in factor analysis should be at least 0.5 and above. Therefore, the KMO value in this study is 0.715. Meanwhile, Bartlett's Sphericity test is applied to determine the possible intercorrelation matrix in uncorrelated variables in the null hypothesis. When the value is 0.05 of significant level will reject the null hypothesis, however, this study indicate the result of 0.00 whereby the value is too small to eliminate the null hypothesis. Hence, the PCA is used in present study for the determination in the interrelationship among the water quality variables.

After correlation analysis, the standardized data is applied into PCA to extract information regarding on the water samples in Melaka River basin [1]. The aims to reduce the differences of variance in variables, Z-scale are standardized including the mean and the variance within the value of zero until one are applied for adjustment on the inconsistency in measurement units and the variable sizes [7]. The variable data is standardized before calculate using the correlation matrix based parameter that measured to obtained the PCA scores. Then, to maximize the square loadings for variance by using varimax rotation method, before the factors is determine based on the Kaiser's criterion [24]. In cluster analysis, the method is used to identify similarity and dissimilarity of sampling stations. This study employed cluster analysis for temporal and spatial variability by using Ward's method of squared Euclidean distances in measuring the similarity [11, 21]. The method of Ward's technique indicates the groups are categorized accurately especially an increment of the squared error for two clusters with their proximity is increase [25]. When the twenty parameters are tested, the variables is centralized and reduced for the purpose of clustering method. The provided outcome of the cluster and their relation is demonstrated using dendogram which is from the reduction of original data. Detail procedures for the application of correlation analysis, PCA and CA analysis can be found elsewhere [1-2, 9, 11, 21].

Quality Assurance and Quality Control

To remove contaminants and traces of cleaning reagent, the laboratory apparatus and polyethylene bottles were soaking overnight using 5% (v/v) of nitric acid before carry out laboratory analysis [30]. For BOD analysis, the BOD bottles were wrapped with aluminum foil. The river water samples can be analyzed within one month if preserved using 1% (v/v) nitric acid (HNO_3) for trace metal. 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%. Before used, all probe meters and instruments are compulsory to calibrate prior to analysis. During analysis, the river water samples are treated the same way by employed the standards and blanks to minimize the matrix interference. 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 Treatment

The data received are compulsory to conduct for the normality test of each variable by kurtosis and skewness statistical test. Meanwhile, the raw data are also subjected to be treated by logarithmatic transformation (logscaling), column scaling and column auto-scaling. After log-transformation, the raw data will be valued within 0 to 0.060 in resulted for kurtosis and skewness, while this allow z-scale standardized for every variables of log-transformed to be minimize in the effects of different units as well as variance of variables and equally to transform the data dimensionless [6, 18]. Data transformation enables to normalize all set of data in order to fulfil the prediction of cluster plus factor analysis [24]. As stated before, all calculations that performed in this study using IBM SPSS Ver. 23.

Results and Discussion

Description of River Water Quality

Table 2 (i) and Table 2 (ii) show the value of mean and standard deviation of twenty variables of Melaka River water quality at nine sampling stations for dry and wet season in 2016. One-way ANOVA analysis proved that the studied variables are varied significantly among the stations ($p<0.05$). It is supported by the coefficient of variance (CV) for all variables were above 50%, except for the temperature, cadmium and lead in January to June; as well as cadmium, chromium, and lead in July to December. The CV was calculated based on the sum value of standard deviation from each studied parameters divided by its mean value. A high CV percentage indicated a high variation between sampling stations.

Based on Table 2 (i), the results indicate pH, temperature, as well as trace metal (i.e. As, Hg, Cd, Cr, Pb, Zn, Fe) is in class 1, considered as clean status in the Melaka River basin. Continue by salinity, electrical conductivity, and total dissolved solid, the highest mean value is detected in S7 with 7.458%, 17777.33 μS/cm,and 6516.50 mg/L respectively, to resulted as class 5; while class 4 occur in S1 with the mean value of 3.85 %, 5804.83 μS/cm, and 2892.50 mg/L respectively, in the river basin. This situation is assumed to have associated with agriculture and livestock activities that carried out along the Melaka River. Simultaneously, others monitored stations remain clean condition. Subsequently, total suspended solid is expected to have slightly polluted in Melaka River for S3 (TSS = 144.83 mg/L), S4 (TSS = 147.50 mg/L), S5 (TSS = 245.83 mg/L), S6 (TSS = 283 mg/L), S9 (TSS = 216.17 mg/L), as well as others sampling stations which resulted in class 2 to remain as clean river. Only turbidity are assumed to be high value of mean at S3 (Tur = 161.33 NTU), S4 (Tur = 182.07 NTU), S5 (Tur = 210.35 NTU), S6 (Tur = 252.83 NTU), and S9 (Tur = 198.78 NTU). S1, S2, S7 and S8 are still clean with class 2. The greater mean value in turbidity and total suspended solid is suspected to be involved with several activities attributed to land clearing, deforestation, river dredging and chennelisation within the Melaka River basin [1, 8].

In chemical parameter, majority variables are shift from clean to slightly polluted which detected along the Melaka River. For example, ammoniacal nitrogen have high mean in S1 with 5.53 mg/L, S2 with 3.58 mg/L, S3 with 2.09 mg/L, S4 with 0.34 mg/L, S7 with 4.45 mg/L, and S8 with 7.74 mg/L. The remaining monitored stations are still clean as class 2. Biochemical oxygen demand valued in S2 with 7.33 mg/L, S7 with 6.33 mg/L, and S8 with 8.17 mg/L, to provide the result as class 4;

Table 2 (i). Mean and standard deviation values of water quality data along Melaka River between January to June of 2016.

Category	Unit		S1	S2	S3	S4	S5	S6	S7	S8	S9	CV (%)
pH	-	Mean	7.13	6.89	6.95	7.20	7.09	7.28	7.25	7.45	7.06	63.26
		SD	0.64	0.23	0.13	0.27	0.16	0.06	0.12	0.28	0.17	
Temp	ºC	Mean	28.83	29.02	29.61	29.10	29.06	28.77	29.32	30.09	28.71	42.13
		SD	1.18	1.10	1.65	1.50	1.07	0.46	0.73	0.50	1.31	
Sal	%	Mean	3.85	0.16	0.09	0.05	0.05	0.03	7.458	0.17	0.04	69.32
		SD	2.05	0.04	0.03	0.01	0.01	0.01	6.35	0.04	0.00	
EC	µS/cm	Mean	5804.83	328.83	185.33	102.00	95.83	68.17	17777.33	388.17	89.17	74.94
		SD	3272.58	87.62	67.76	13.34	14.76	18.34	10680.92	32.63	11.00	
TSS	mg/L	Mean	32.83	38.17	144.83	147.50	245.83	283.00	30.83	36.17	216.17	54.90
		SD	8.33	12.51	59.38	89.41	338.42	165.97	10.87	6.31	180.03	
TDS	mg/L	Mean	2892.50	213.33	230.17	193.83	288.17	305.33	6516.50	1917.33	254.67	76.02
		SD	1812.44	47.58	33.13	87.13	331.47	153.48	6613.57	4092.00	172.27	
Tur	NTU	Mean	23.98	44.62	161.32	182.07	210.35	252.83	14.62	17.80	198.78	73.32
		SD	5.70	29.06	72.07	62.49	276.89	148.57	2.26	3.02	171.90	
BOD	mg/L	Mean	5.67	7.33	5.17	4.67	4.33	4.67	6.33	8.17	4.50	59.80
		SD	1.63	1.63	2.14	2.16	1.63	1.63	1.51	0.75	1.05	
COD	mg/L	Mean	42.17	28.00	28.33	20.50	29.83	22.17	32.33	58.50	21.17	65.55
		SD	6.91	9.01	17.95	8.87	12.75	4.26	6.86	17.85	8.16	
DO	mg/L	Mean	2.15	3.16	5.35	6.00	6.46	6.71	3.69	6.41	6.49	66.01
		SD	0.72	1.16	0.63	0.48	0.10	0.28	0.53	3.14	0.62	
NH_3N	mg/L	Mean	5.53	3.58	2.09	0.34	0.25	0.23	4.45	7.74	0.29	71.36
		SD	1.60	1.54	1.32	0.26	0.14	0.07	1.89	0.73	0.17	
As	mg/L	Mean	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	52.69
		SD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	
Hg	mg/L	Mean	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	55.56
		SD	LOD	LOD	LOD	LOD	LOD	LOD	LOD	LOD	LOD	
Cd	mg/L	Mean	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	46.28
		SD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	
Cr	mg/L	Mean	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	0.01	<LOD	51.99
		SD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	
Pb	mg/L	Mean	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	48.19
		SD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	
Zn	mg/L	Mean	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	53.56
		SD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	0.01	<LOD	

Table 2 (i). Continued.

Fe	mg/L	Mean SD	0.06 0.06	0.28 0.29	0.35 0.15	0.24 0.05	0.233 0.04	0.21 0.08	0.40 0.33	0.27 0.12	0.31 0.07	69.73
Total Coliform	count/100 mL	Mean SD	731500.00 665865.98	284833.33 180992.17	158500.00 118525.52	70666.67 63108.37	58966.67 28662.29	66850.00 25350.64	195000.00 113534.14	287000.00 49735.30	72250.00 48709.09	84.75
Ecoli	count/100 mL	Mean SD	96500.00 45469.77	32033.33 38212.39	27500.00 17784.82	9166.67 15197.59	6933.33 2026.49	11616.67 4485.72	23333.33 17362.80	32333.33 17772.64	14666.67 11927.56	94.25
WQI			III	III	III	II	II	II	III	III	II	-

(Tur = Turbidity; TDS = Total Dissolved Solids; Con = Electrical Conductivity; Sal = Salinity; Temp = Temperature; DO = Dissolved Oxygen; BOD = Biochemical Oxygen Demand; COD = Chemical Oxygen Demand; TSS = Total Suspended Solids; pH = Acidic or Basic water; NH_3N = Ammoniacal Nitrogen; *E coli* = *Escherichia* Coliform; Coli = Coliform; As = Arsenic; Hg = Mercury; Cd = Cadmium; Cr = Chromium; Pb = Lead; Zn = Zinc; Fe = Iron; SD = Standard Deviation; S1 to S9 = Station 1 to Station 9; WQI = Water Quality Index; <LOD = Limit Of Detection; CV means one-way Coefficients of Variance (one-way ANOVA))

Table 2 (ii). Mean and standard deviation values of water quality data along Melaka River between July to December of 2016.

Category	Unit		**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**CV (%)**
pH	-	Mean SD	7.21 0.49	6.74 0.12	6.71 0.26	7.01 0.19	6.88 0.10	6.97 0.13	7.15 0.68	6.85 0.31	6.87 0.14	57.10
Temp	°C	Mean SD	28.56 0.79	29.09 1.31	28.82 1.09	29.03 1.46	28.69 1.04	29.30 0.98	28.75 0.65	28.77 1.23	28.42 0.86	75.96
Sal	%	Mean SD	17.15 12.22	0.17 0.06	0.11 0.04	0.05 0.00	0.05 0.01	0.06 0.03	19.70 10.99	0.16 0.05	0.05 0.01	60.19
EC	µS/cm	Mean SD	24734.83 20427.98	372.50 143.03	227.67 81.33	105.33 8.50	100.33 9.33	131.50 64.70	30569.33 16318.27	379.00 118.22	99.83 7.47	52.56
TSS	mg/L	Mean SD	36.83 9.58	118.17 116.21	121.50 82.82	168.17 100.58	149.17 59.58	134.00 35.05	28.83 5.27	53.17 12.16	135.17 41.67	58.14
TDS	mg/L	Mean SD	14134.50 12227.70	329.17 57.37	235.17 44.65	216.33 98.07	197.50 54.18	206.50 41.77	18876.33 10268.32	274.50 35.84	183.83 42.56	55.52
Tur	NTU	Mean SD	21.82 9.14	142.40 142.54	133.87 93.06	206.50 130.00	153.15 75.02	106.97 29.74	28.38 8.18	39.52 9.60	145.08 44.53	80.45

Table 2 (ii). Continued.

BOD	mg/L	Mean SD	4.67 1.03	8.50 2.43	8.17 3.06	3.83 1.72	3.50 1.22	4.17 1.33	5.83 1.72	8.00 1.41	4.17 1.72	53.06
COD	mg/L	Mean SD	21.83 7.68	34.33 14.43	31.33 15.20	18.33 7.12	20.83 6.27	17.17 5.91	21.83 5.56	29.17 8.84	24.67 8.26	76.89
DO	mg/L	Mean SD	4.56 1.02	3.47 0.73	4.35 1.52	6.73 0.38	6.19 0.40	6.56 0.24	5.21 0.87	3.42 0.81	5.97 0.66	70.75
NH_3N	mg/L	Mean SD	1.33 0.72	3.49 1.87	1.94 0.77	0.46 0.09	0.39 0.13	0.36 0.07	1.10 0.57	3.96 1.30	0.51 0.20	63.71
As	mg/L	Mean SD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	57.75
Hg	mg/L	Mean SD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	58.00
Cd	mg/L	Mean SD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	43.98
Cr	mg/L	Mean SD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	<LOD <LOD	48.21
Pb	mg/L	Mean SD	0.01 <LOD	0.01 <LOD	0.01 <LOD	0.01 <LOD	0.01 <LOD	0.01 <LOD	0.01 <LOD	0.01 <LOD	0.01 <LOD	39.23
Zn	mg/L	Mean SD	0.01 0.03	0.01 0.01	0.02 0.01	0.01 <LOD	0.01 <LOD	0.01 <LOD	0.02 0.01	0.02 0.01	0.01 <LOD	71.07
Fe	mg/L	Mean SD	0.03 0.04	0.16 0.10	0.19 0.11	0.19 0.07	0.21 0.10	0.19 0.04	0.06 0.09	0.20 0.13	0.24 0.08	69.42
Total Coliform	count/100 mL	Mean SD	100750.00 126964.77	196000.00 115751.46	159700.00 117335.33	70666.67 43458.79	235700.00 308261.80	114833.33 44821.50	77950.00 81881.86	194666.67 27522.11	175833.33 63666.06	74.64
Ecoli	count/100 mL	Mean SD	14150.00 19312.77	34233.33 27374.56	21216.67 28360.92	8733.33 10568.95	17300.00 14971.31	12566.67 1286.34	19466.67 33199.20	34000.00 17029.39	29983.33 16698.79	81.54
WQI			III	III	III	II	II	II	III	III	II	-

(Tur = Turbidity; TDS = Total Dissolved Solids; Con = Electrical Conductivity; Sal = Salinity; Temp = Temperature; DO = Dissolved Oxygen; BOD = Biochemical Oxygen Demand; COD = Chemical Oxygen Demand; TSS = Total Suspended Solids; pH = Acidic or Basic water; NH_3N = Ammoniacal Nitrogen; *E coli* = *Escherichia* Coliform; Coli = Coliform; As = Arsenic; Hg = Mercury; Cd = Cadmium; Cr = Chromium; Pb = Lead; Zn = Zinc; Fe = Iron; SD = Standard Deviation; S1 to S9 = Station 1 to Station 9; WQI = Water Quality Index; <LOD = Limit Of Detection; CV means one-way Coefficients of Variance (one-way ANOVA))

while S1, S3, S4, S5, S6, and S9 are indicated as class 3. In chemical oxygen demand and dissolved oxygen, the mean value showed as slightly polluted which suspected to be detected in S1 (COD = 42.17 mg/L; DO = 2.15 mg/L), S2 (COD = 28 mg/L; DO = 3.16 mg/L), and S7 (COD = 32.33 mg/L; DO = 3.69 mg/L) respectively, while others monitored stations remain as clean in the river basin. Overall, anthropogenic activities such as residential, municipal and domestic, industrial, agriculture, and livestock is carried out within the Melaka River basin [1, 11, 13, 26]. In other words, these activities might be considered as major contribution of contamination to occur in the Melaka River. Lastly, biological parameters of total coliform and *Escherichia* coliform are resulted as class 5 in all sampling station. These contaminations are expected to be linked with sewage treatment plant, agriculture and livestock activities, which cause the point and non-point source pollution of surface runoff to end up in the river [1, 8].

Almost the same as Table 2 (i), the results for pH, temperature, and trace metal (i.e. As, Hg, Cd, Cr, Pb, Zn, Fe) are in class 1 to remain as clean river in the basin (Table 2 (ii)). According to Table 2 (ii), the physical parameter which includes salinity, electrical conductivity, and total dissolved solid was found to be high in S1 with the mean value of 17.15 %, 24734.83 μS/cm, and 14134.50 mg/L; as well as S7 with a mean value of 19.70 %, 30569.33 μS/cm, and 18876.33 mg/L, respectively. The results provided are having class 5 conditions, which are suspected to have agriculture and livestock activities carried out within the S1 and S7 sub-basin. Meanwhile, other stations are still considered clean with the result are in class 1. Then, total suspended solid are class 2 in S1 (TSS = 36.83 mg/L) and S7 (TSS = 28.83 mg/L), while others monitored stations are in class 3. The turbidity is class 5 from S2 to S6 and S9, as well as class 2 in S1, S7 and S8. As explained before, both variables in the river might be associated with land clearing, deforestation, river dredging and chennelisation activities in the basin [1, 8].

Ammoniacal nitrogen are consider having the condition of slightly polluted in the river basin, whereby class 3 in S4 (0.46 mg/L), S5 (0.39 mg/L), S6 (0.36 mg/L), and S9 (0.51 mg/L); class 4 in S1 (1.33 mg/L), S3 (1.94 mg/L), and S7 (1.10 mg/L); as well as class 5 in S2 (3.49 mg/L) and S8 (3.96 mg/L). Meanwhile, chemical oxygen demand and dissolved oxygen resulted in class 3 are S2 (COD = 34.33 mg/L; DO = 3.47 mg/L), S3 (COD = 31.33 mg/L; DO = 4.35 mg/L) and S7 (COD = 29.17 mg/L; DO = 3.42 mg/L), while other sampling stations are in class 2. Lastly, biochemical oxygen demand, total coliform, and *Escherichia* coliform are having class 5 in majority monitored stations. As explained before, anthropogenic activities are suspected to act as pollutant sources to cause contamination in the Melaka River [1, 8].

The Evaluation within the River Water Quality Variables

According to correlation analysis in Table 3(i), the result shows salinity is positively correlated with electrical conductivity (r = 0.924, $p<0.01$) and total dissolved solid (r = 0.878, $p<0.01$), as well as negative correlated with turbidity (r = -0.287, $p<0.05$). Meanwhile, electrical conductivity is positively correlated with total dissolved solid (r = 0.779, $p<0.01$) and negatively correlated with turbidity (r = -0.296, $p<0.05$). Then, turbidity is positively correlated with total suspended solid (r = 0.969, $p<0.01$), as well as negatively correlated with total dissolved solid (r = -0.285, $p<0.05$). Nevertheless, beginning in July of monitored water sample, majority physical parameters are significantly correlated between each other (Table 3(ii)). For instance, salinity is positively correlated with electrical conductivity (r = 0.983, $p<0.01$) and total dissolved solid (r = 0.973, $p<0.01$); as well as negatively correlated with total suspended solid (r = -0.429, $p<0.01$) and turbidity (r = -0.412, $p<0.01$). Continue by electrical conductivity, whereby this variable is positive correlation with total dissolved solid (r = 0.994, $p<0.01$), as well as negative correlation with total suspended solid (r = -0.422, $p<0.01$) and turbidity (r = -0.406, $p<0.01$). Total suspended solid is significantly correlated in positive with turbidity (r = 0.947, $p<0.01$) but significantly correlated in negative with total dissolved solid (r = -0.428, $p<0.01$). Lastly, total dissolved solid is a negative correlation with turbidity (r = -0.413, $p<0.01$).

Based on the overall analysis of physical parameters, the result shows that several physical activities are involved in the Melaka River basin. For instance, land clearing and deforestation activities are suspected to occur in Kampung Sungai Petai sub-basin and Kampung Panchor sub-basin. These activities are transforming the forest field into agriculture and residential area. Simultaneously, several disturbances on the river physical characteristic through the activities of dredging, channelisation and irrigation would increase the riverbank erosion and indirectly enhance the physical contamination in the river [11, 13, 26]. During wet season, this circumstance would boost the amount of contamination in water which is due to the rapid surface runoff from elsewhere to end up in the river. It is assumed to occur along the river especially beginning from Kampung Tualang sub-basin to Kampung Kelemak sub-basin. Ultimately, pH variable is positively correlated with total dissolved solid (r = 0.363, $p<0.01$); while temperature is negatively correlated with total suspended solid (r = -0.316, $p<0.05$) and turbidity (r = -0.307, $p<0.05$) (Table 3 (i)). Nonetheless, correlation between variable increases, especially in pH parameter during the raining season for July to December 2016, whereby the pH variable correlate with salinity (r = 0.668, $p<0.01$) and electrical conductivity (r = 0.654, $p<0.01$) in positive condition;

Table 3 (i). PCC among water quality variables in Melaka River basin for January to June of 2016.

	pH	Temp	Sal	EC	TSS	TDS	Tur	BOD	COD	DO	NH3N	As	Hg	Cr	Zn	Fe	Tcoli	*Ecoli*
pH	1																	
Temp	.114	1																
Sal	.157	.058	1															
EC	.169	.057	.924**	1														
TSS	.008	-.316*	-.239	-.244	1													
TDS	.363**	.135	.878**	.779**	-.185	1												
Tur	-.024	-.307*	-.287*	-.296*	.969**	-.285*	1											
BOD	-.221	.158	.230	.238	.253	.254	-.241	1										
COD	.504**	-.007	.078	.049	-.153	.213	-.221	.558**	1									
DO	.406**	.071	-.380**	-.354**	.344**	-.086	.350**	-.163	.018	1								
NH_3N	.295*	.157	.234	.221	-.495**	.318*	-.562**	.514**	.752**	-.392**	1							
As	.475**	.091	.157	.230	-.256	.315*	-.288*	.453**	.611**	-.087	.631**	1						
Hg	-.056	-.148	.252	.221	-.128	.211	-.136	.002	.135	-.238	.283*	.122	1					
Cr	.329*	.141	-.131	-.131	-.190	-.029	-.219	.318*	.523**	.112	.456**	.332*	-.074	1				
Zn	-.077	.132	-.035	.017	-.216	-.062	-.243	.351*	.125	-.303*	.264	.236	-.067	.050	1			
Fe	.046	.160	.178	.238	.059	.247	.049	.334*	-.036	.242	-.168	.025	-.263	-.065	.044	1		
TColi	-.047	.002	.290*	.267	-.287*	.239	-.319*	.267	.256	-.444**	.445**	.178	.309*	.020	.272*	-.124	1	
EColi	-.005	.092	.209	.167	-.247	.131	-.288*	.194	.218	-.567**	.364**	.165	-.029	.113	.284*	-.164	.527**	1

**Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed)

Table 3 (ii). PCC among water quality variables in Melaka River basin for July to December of 2016.

	pH	Temp	Sal	EC	TSS	TDS	Tur	BOD	COD	DO	NH3N	As	Hg	Zn	Fe	Tcoli	*Ecoli*
pH	1																
Temp	.204	1															
Sal	.668**	-.080	1														
EC	.654**	-.064	.983**	1													
TSS	-.269*	-.114	-.429**	-.422**	1												
TDS	-.428**	-.070	.973**	.994**	-.428**	1											
Tur	-.249	-.127	-.412**	-.406**	.947**	-.413**	1										
BOD	-.218	.179	-.128	-.111	-.298*	-.107	-.207	1									
COD	-.118	.209	-.114	-.106	-.288*	-.103	-.267	.712**	1								
DO	.389**	.026	.061	.063	.318**	.060	.269*	-.609**	-.363**	1							
NH_3N	-.222	.071	-.172	-.173	-.369*	-.178	-.320*	.680**	.546**	-.693**	1						
As	.158	.247	.379**	.374**	-.306*	.369**	-.340*	.223	.164	-.227	.192	1					
Hg	.056	-.048	-.062	-.061	.086	-.062	.132	.073	.048	.149	-.095	-.090	1				
Zn	-.372**	-.164	.037	.015	-.383*	.030	-.368**	.161	.070	-.433**	.334*	.192	-.066	1			
Fe	-.443**	.076	-.581**	-.563**	.041	-.565**	.009	-.087	.081	.055	.068	-.326*	.023	.056	1		
TColi	-.323*	.032	-.353**	-.344**	.110	-.344*	.035	.108	.219	-.176	.295*	-.073	-.130	.085	.100	1	
EColi	-.383**	.127	-.241	-.216	.069	-.210	.066	.317*	.124	-.418**	.339*	.124	-.104	.105	.046	.474**	1

**Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed)
(Tur = Turbidity; TDS = Total Dissolved Solids; Con = Electrical Conductivity; Sal = Salinity; Temp = Temperature; DO = Dissolved Oxygen; BOD = Biochemical Oxygen Demand; COD = Chemical Oxygen Demand; TSS = Total Suspended Solids; pH = Acidic or Basic water; NH_3N = Ammoniacal Nitrogen; *E coli* = *Escherichia* Coliform; TColi = Total Coliform; As = Arsenic; Hg = Mercury; Cd = Cadmium; Cr = Chromium; Pb = Lead; Zn = Zinc; Fe = Iron; S1 to S9 = Station 1 to Station 9)

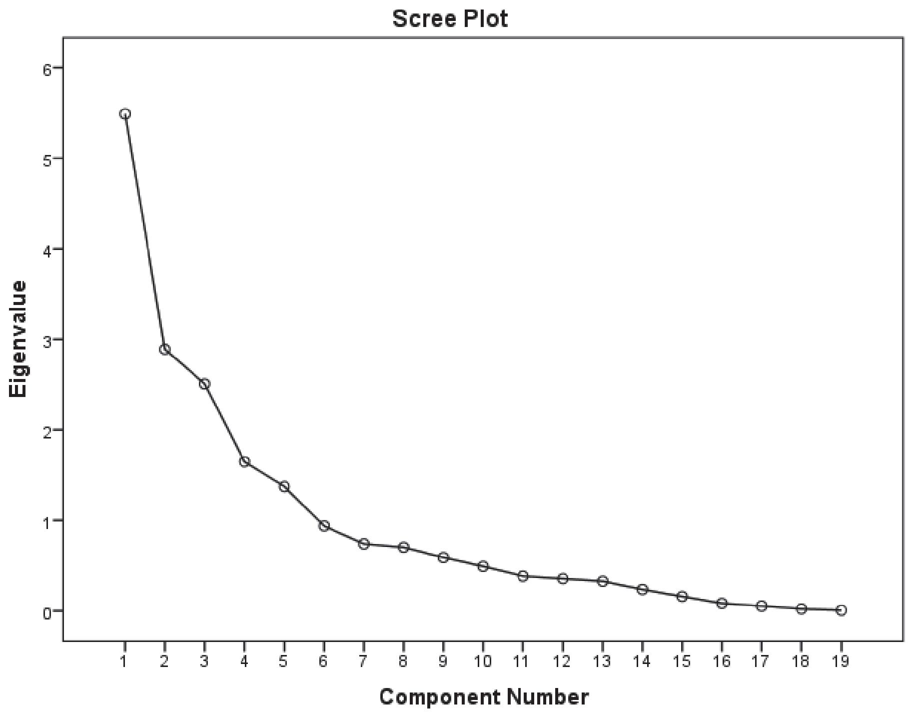

Fig. 2a (i). Scree plot of the eigenvalues for January to June of 2016.

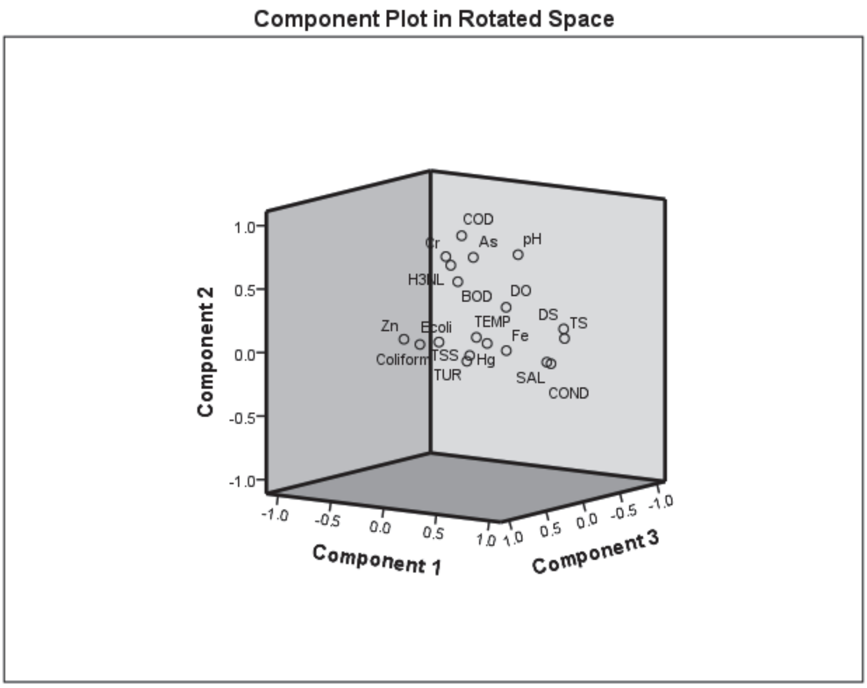

Fig. 2a (ii). Five principal components of January to June of 2016.

as well as the pH variable correlate with total dissolved solid (r = -0.428, p<0.01) and total suspended solid (r = -0.269, p<0.05) in negative condition (Table 3 (ii)). This condition proved that physical parameter in wet season will have greater contamination than dry season in the Melaka River basin.

Meanwhile, except for individual parameter of chemical oxygen demand in Table 3 (i), the result shows biochemical oxygen demand is positively correlated with chemical oxygen demand (r = 0.558, p<0.01) and ammoniacal nitrogen (r = 0.514, p<0.01). Then, ammoniacal nitrogen is positively correlated with chemical oxygen demand (r = 0.752, p<0.01) and negatively correlated with dissolved oxygen (r = -0.392, p<0.01). Compared to the statistical data analysis in wet season, the results indicate all chemical parameters is

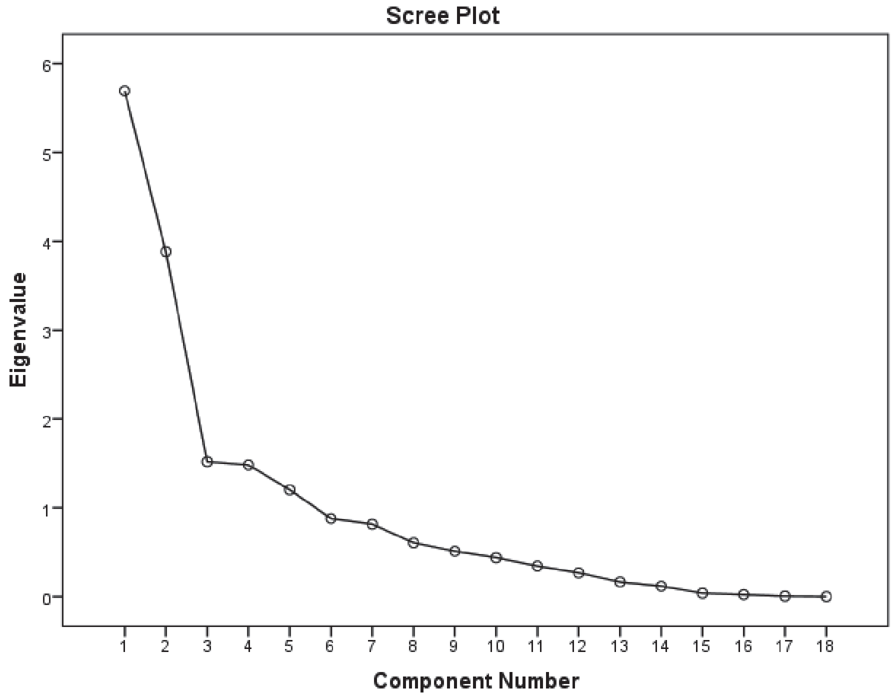

Fig. 2b (i). Scree plot of the eigenvalues for July to December of 2016.

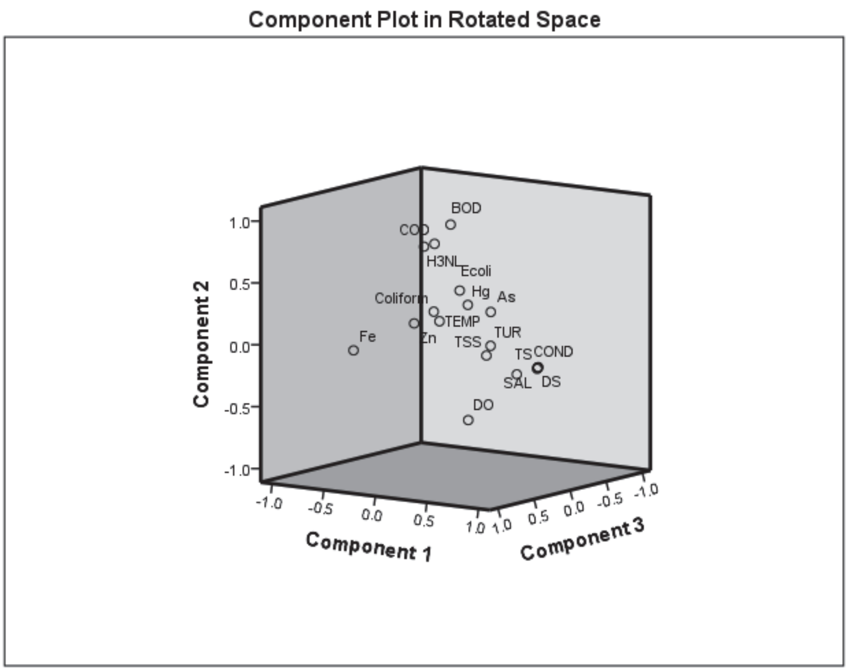

Fig. 2b (ii). Five principal components of July to December of 2016.

significantly correlated among each other (Table 3 (ii)). In biochemical oxygen demand, this variable is correlated positively with chemical oxygen demand (r = 0.712, p<0.01) and ammoniacal nitrogen (r = 0.680, p<0.01); as well as correlated negatively with dissolved oxygen (r = -0.609, p<0.01). Then, chemical oxygen demand is correlated with ammoniacal nitrogen (r = 0.546, p<0.01) in positive condition and dissolved oxygen (r = -0.363, p<0.01) in negative condition. Meanwhile, dissolved oxygen variable is negatively correlated with ammoniacal nitrogen variable (r = -0.693, p<0.01).

The results indicate major contamination is from dissolved oxygen and biochemical oxygen demand, followed by the ammoniacal nitrogen contamination in the basin area. This occurred during dry season.

Nevertheless, frequent raining at study area had led to dissolved oxygen, biochemical oxygen demand, and chemical oxygen demand to spread within the riverine system before the ammoniacal nitrogen began to grow within the basin. Generally, when organic matter is in the water, this will cause biological process to occur and increase the concentration of chemical oxygen demand in the river [13, 26]. At the same time, these processes will also lead to a large consumption amount of dissolved oxygen or biochemical oxygen demand to undergo the anaerobic fermentation process, before producing the ammoniacal nitrogen and organic acid [13, 26]. Surface runoff with high capacity during wet season are able to transport foreign substances (e.g. including municipal and domestic waste, agriculture waste, industrial effluents, etc.) into the river to result more microorganism activity, which indirectly increases the chemical oxygen demand, biochemical oxygen demand, dissolved oxygen, and ammoniacal nitrogen in basin area. According to Juahir et al. [13], these activities might be originated from municipal and domestic wastewater, agriculture runoff, and livestock activities which are likely to carried out adjacent to the Melaka River and its tributaries. Wide usage of pesticide in agriculture activities within the basin area will enrich the organic matter and indirectly increase chemical contamination in Melaka River. Meanwhile, the pH are also positively correlated with chemical oxygen demand (r = 0.504, $p<0.01$), dissolved oxygen (r = 0.406, $p<0.01$), and ammoniacal nitrogen (r = 0.295, $p<0.05$). However, pH variable are only correlated with dissolved oxygen (r = 0.389, $p<0.01$) during wet season.

In trace metal, only arsenic element is positively correlated with chromium element with the value of r = 0.332 at the significant level of p value less than 0.05 (Table 3(i)). While for the next six month, the arsenic element are continue to affect the iron element with negative value of r = -0.326 at the significant level of p value less than 0.05 (Table 3(ii)). Generally, river contamination due to arsenic element is caused by feces through agriculture and livestock activities; chromium element is affected by municipal and domestic waste through urban runoff; as well as iron element is due to industrial wastes and effluents such as steelmaking industries [11, 27]. During dry season, the contamination in river is suspected to originate from agriculture and livestock activities, as well as municipal and domestic wastes through point and non-point source pollution [11, 13]. Despite that, wet season had cause the water volume in the river to increase, including the foreign substances that transport from elsewhere through surface runoff in the basin. In other words, non-point source pollution might consider having greater impact than point source pollution to cause contamination in the Melaka River basin. Hence, the existence of arsenic plus iron elements in the river would indirectly cause more phytoplankton, microorganism, and organic matter, as well as metal processing wastes to toxic the aquatic organism. Continuous exposure of both elements in Melaka River water quality could cause a hazard to the aquatic biota.

Lastly, total coliform has positive correlation with Escherichia coliform for the year of 2016, whereby the both variables have the value of r = 0.527 at the significant level of p value less than 0.01 in dry season, and the value of r = 0.474 at the significant level of p value less than 0.01 in wet season, respectively. On the other hands, pH variable are negatively correlated with *Escherichia* coliform (r = -0.383, $p<0.01$) and total coliform (r = -0.323, $p<0.05$) only during the rainy day or wet season. The presence of coliform in the water is due to the activities that involved with feces in the agriculture, livestock, and sewage treatment plant within the river basin area. During field sampling, the livestock activities refer to chicken, cow, goat and pig that concentrated at Kampung Kelemak, Kampung Sungai Petai, and several areas in Kampung Panchor, are farmed in small scale within the Melaka River basin. Without proper management, the livestock wastes (especially from warm-blooded animal) to spread widely through direct exposure to the soil or direct input into the river [13, 26]. It will lead to non-point source pollution and increases the coliform through spreading of microorganism in the river. These circumstances, especially when wet season arrived, will enhance the phytoplankton activities through contamination of coliform which flow directly and/or indirectly into the river. It is proved by the pH parameter that significantly correlated with the coliform parameter, which has the ability to detect the existence of foreign concentration in the river water quality. Simultaneously, major practices of fertiliser usage in agriculture activities by farmers which focuses in Kampung Sungai Petai sub-basin and Kampung Panchor sub-basin, might enhance the biological parameters through surface runoff into the riverine ecosystem.

Source Identification of Monitored Variables

The correlation matrix is performed before the outcome is included into PCA to understand the relationships of water quality variable in monitored stations, as well as to determine the water quality characteristic that being affected by the potential sources of contamination. The number of PCs is determined using the criteria of Kaiser [24]. The basic data structure will provide the result for the PCs number based on scree plot graphs [28]. The outcome from scree plots is the eigenvalues that being sorted from large to small before it can be used as PC number. In Fig. 2a (i) and Fig. 2b (i), both scree plots represent the fifth eigenvalues having significantly to curve before remaining as a horizontal line. Hence, only five components were retained for further determination. Then, to proceed with the eigenvalues in the PCs, only the value is greater than 1 will be extracted, while the value lower than 1 will be removed (Table 4). Based on Kim and Mueller [29], the value less than 1 will be considered as low significant. Therefore, the selected five components of PCs for dry (total of variance = 73%) and wet (total of variance = 77%) season is above the requirement value of 1 (Kaiser Normalization) in the water dataset. Lastly, according to Liu et al (2003), the PC loadings can only be determined based on the loading values of >0.75 as strong corresponding, 0.75-0.50 as moderate corresponding, and 0.50-0.30 as weak corresponding respectively. The result can be shown in Table 4.

In this study, the PC1 in dry season with total variance of 20.69% to indicate the positive loadings on total dissolved solid, salinity, and electrical conductivity. Based on the result, it is expected land clearing, livestock and agriculture activities [1, 8] are carried out within the basin area. Meanwhile, PC2 with total variance of 17.73% to resulted the positive loading of chemical oxygen demand, while PC3 with the total variance of 14.34% to provide the negative loading of dissolved oxygen. The existence of chemical oxygen demand is usually related to municipal sewage discharge, sewage treatment plants, and industrial wastes [8, 11, 13]; and dissolved oxygen is associated to agriculture activities and forest field which is due to dissolved organic matter that required large number of oxygen in river water [1]. Then, PC4 represent by the turbidity and total suspended solid as positive loadings with total variance of 11.60%. This can be link to soil erosion, especially involved with physical modification of river such as dredging and channelization could cause disruption in the river [8, 13]. Lastly, PC5 with total variance of 8.86% to resulted in positive loading of iron in the Melaka River. The iron in river water can be attributed with industrial effluents such as electroplating [11, 13].

Wet season indicates total of variance in PC 1 are 28.29% to provide the positive loadings on total dissolved solid, electrical conductivity, and salinity. Similar to dry season, the pollutant source in Melaka River is suspected from land clearing, livestock and agriculture activities. Next, PC 2 with positive loadings on biological oxygen demand, chemical oxygen demand, and ammoniacal nitrogen had provided the result with total of variance is 17.31% in the Melaka River. Anthropogenic activities of industrial effluents, municipal and domestic wastes, sewage treatment

Table 4. Varimax rotation PCs for water quality data within Melaka River basin.

Variables (Unit)	Dry Season					Wet Season				
	PC 1	PC 2	PC 3	PC 4	PC 5	PC 1	PC 2	PC 3	PC 4	PC 5
Turbidity (NTU)	-.176	-.173	-.268	**.878**	.114	-.276	-.212	**-.879**	-.001	-.056
Total Dissolved Solid (mg/L)	**.945**	.161	.013	-.065	.006	**.946**	-.105	.216	-.108	-.025
Electrical Conductivity (uS)	**.895**	-.007	.182	-.103	.019	**.945**	-.107	.208	-.111	-.014
Salinity (ppt)	**.935**	-.015	.182	-.094	-.045	**.944**	-.115	.219	-.121	-.030
Temperature (ºC)	.042	.064	-.084	-.598	.385	-.055	.158	.144	.172	-.048
Dissolved Oxygen (mg/L)	-.180	.174	**-.804**	.158	.258	-.011	-.685	-.190	-.344	.367
Biological Oxygen demand (mg/L)	.192	.585	.382	-.040	.404	-.023	**.926**	.035	.016	.038
Chemical Oxygen Demand (mg/L)	.065	**.899**	.147	.020	-.087	-.098	**.780**	.151	-.079	.230
Total Suspended Solid (mg/L)	-.131	-.119	-.240	**.898**	.113	-.299	-.289	**-.854**	.071	-.022
Acidity/Alkalinity (pH)	.246	.696	-.356	.046	.070	.655	-.208	.087	-.308	.473
Ammociacal Nitrogen (mg/L)	.144	.713	.408	-.335	-.234	-.124	**.770**	.260	.232	-.105
E-coli (cfu/100ml)	.067	.126	.715	-.098	-.042	-.095	.353	-.192	.694	-.043
Coliform (cfu/100ml)	.217	.154	.673	-.060	-.223	-.280	.177	-.098	**.816**	.612
Arsenic (mg/L)	.179	.742	.154	-.104	.002	.473	.292	.189	.262	.159
Mercury (mg/L)	.282	.068	.113	.014	-.675	-.069	.228	-.268	-.651	.014
Zinc (mg/L)	-.144	.133	.631	-.119	.349	-.047	.193	.506	.182	-.600
Iron (mg/L)	.309	-.017	-.109	.048	**.762**	-.062	-.125	.317	.060	**.768**
Initial Eigenvalue	3.93	3.37	2.72	2.20	1.68	5.09	3.12	2.35	1.74	1.49
% of Variance	20.69	17.73	14.34	11.60	8.86	28.29	17.31	13.03	9.69	8.27
Cumulative %	20.69	38.42	52.76	64.36	73.22	28.29	45.59	58.62	68.31	76.58

* The bold value are factor loadings above 0.75 that were taken after Varimax rotation was performed.

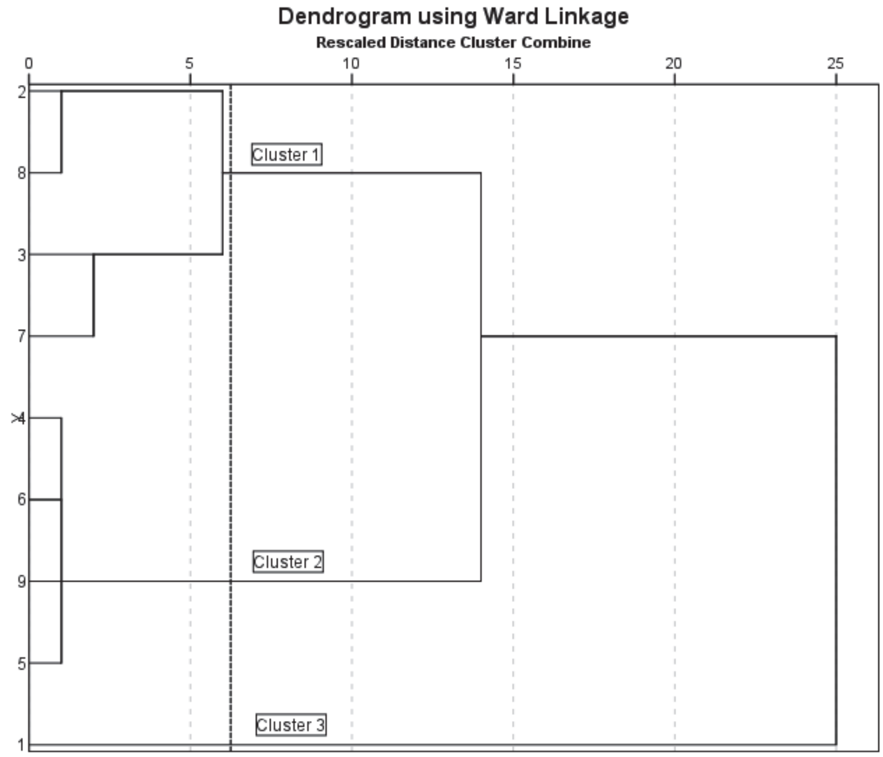

Fig. 3 (i). Dendogram of water quality monitoring stations clusters using Ward Linkage method in CA for January to June of 2016.

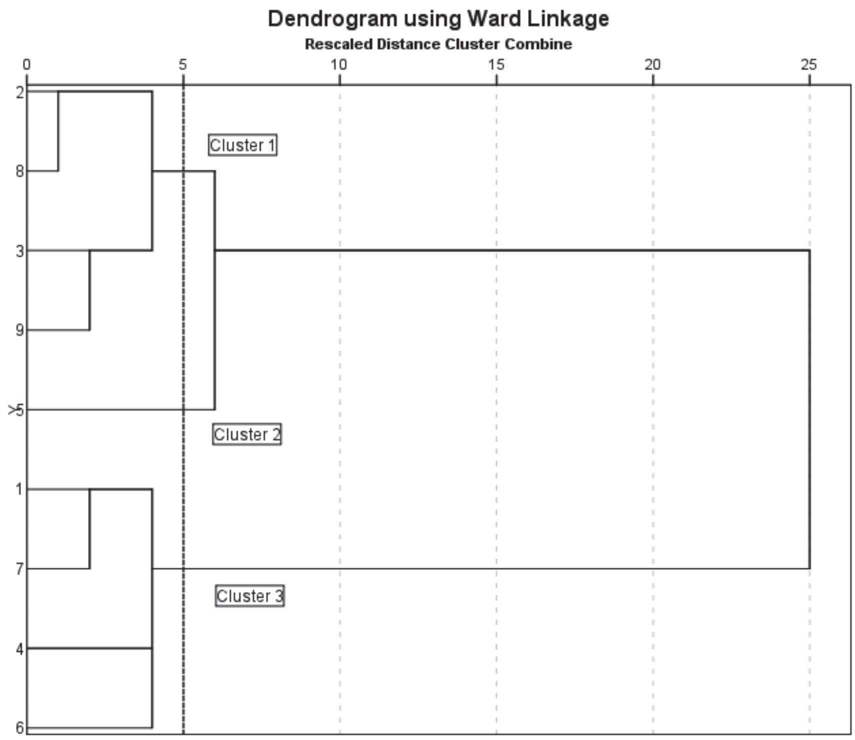

Fig. 3 (ii). Dendogram of water quality monitoring stations clusters using Ward Linkage method in CA for July to December of 2016.

plants, commercial, and agriculture runoff, are assumed to emerge from point and non-point source pollution and contaminate the river basin. As highlighted in correlation analysis, phytoplankton and microorganism, as well as organic matter from various sources of human activities had indirectly activated the biological process, which consumed large amount of chemical oxygen demand and biochemical oxygen demand to undergo the anaerobic fermentation process [13, 26]. Without notifying, continuous input of contamination during rainy day will enhance the process in river. This led to the production of ammoniacal nitrogen [13, 26]. Continuing with PC 3, the result shows total variance of 13.03% to provide the result of negative loadings on turbidity and total suspended solid. Hydrological modification through dredging, water diversion, and channelization, is suspected to cause disruption in the river [8, 13]. Lastly, the PC 4 and PC 5 result in the total of variance with 9.69% and 8.27% to produce positive loading on total coliform and iron, respectively. Through surface runoff, the existence of coliform in the river is expected to come from domestic waste and fertilizer (animal waste) used in agriculture activities; while iron element can be attributed with industrial effluents in the river basin [11, 13].

Spatial and Temporal Variation in River Water Quality

Cluster analysis (CA) is used to classify the monitored stations in similarity based on the similarities of water quality which referred to the spatial and temporal variation in the basin area. Ward's method is applied into dry and wet season, and the result is performed using dendogram as shown in Fig.e 3. The cluster is grouped using the equation of (D_{max}/D_{link}) for dry and wet season. The (D_{max}/D_{link}) value for dry season is 6.25 (Fig. 3(i)), as well as wet season is 5 (Fig. 3(ii)), which provided the results as three clusters for both seasons.

In dry season, the cluster 1 is included in the station S2, S3, S7, and S8; while cluster 2 is composed by the station S4, S5, S6, and S9; and the cluster 3 is involved with only the station S1 (Fig. 3(i)). Based on the calculation of water quality index (WQI) in Table 2(i), the result indicate cluster 3 as polluted, cluster 2 as slightly polluted, and cluster 1 as clean condition in the Melaka River. The main activities involved in S1 are agriculture and livestock, which resulted as polluted condition within the river basin area. This area is referred to Kampung Kelemak sub-basin, whereby it is located at urban area and upstream river. Hence, the contamination in river within the S1 area might probably due to the feces that originated from livestock wastes and also it used as fertiliser in agriculture activities. These circumstances may lead to non-point source pollution to be included the biological, physical and chemical parameter as the result for contamination in the river. Meanwhile, slightly polluted is assumed to occur in S2 (Kampung Sungai Petai sub-basin), S3 (Kampung Panchor sub-basin), S7 (Kampung Batu Berendam sub-basin) and S8, in the Melaka River basin. As expected that several construction activities, land clearing as well as modification on the river's physical characteristic would contribute to the contamination which link to physical parameter in the basin.

During wet season arrived, three clusters had been spotted within basin area, namely cluster 1 of S2, S3, S8, and S9; cluster 2 of S5; as well as cluster 3 of S1, S4, S6, and S7 (Fig. 3(ii)). The WQI indicate cluster 1 as polluted, cluster 3 as slightly polluted, and cluster 2 as clean river (Table 2(ii)). The rainy will enhance surface runoff and increase foreign substances to transport into the river. Therefore, the wastes from land clearing, agriculture, and livestock activities, which spotted carried out in S1 basin will be transported into the river and its tributaries before the contamination is captured in S2 and S3. Almost similar circumstances to the above, majority industrial activities concentrated in S7 (Kampung Batu Berendam sub-basin) are expected to discharges the industrial effluents into the river before the pollution is flow and spotted at S8 and S9. Simultaneously, highly built-up area in urban region had indirectly caused an increment of municipal and domestic wastes into S8 and S9 of the Melaka River basin. This might increase the result of point and non-point source pollution in the river. Fortunately, only S5 monitored station of Kampung Tualang sub-basin area is still remain as clean river in study area. Therefore, the result proved that wet season enhances the contamination through increases surface runoff directly and/or indirectly into the Melaka River.

Conclusion

This study was carried out to determine the pollutant source based water quality data in Melaka River basin and its tributaries. Multivariate statistical techniques such as Pearson correlation analysis, principal component analysis, and cluster analysis, were applied into river water quality data to verify and identify the spatial based on seasonal variations. Referring to biological, chemical and trace metal elements, and biological parameter, monitored station indicate some variables is exceeded the average concentration of recommended based the National Water Quality Standards from DOE, Malaysia. Therefore, the river water quality is suspected to be affected by the pollutants in river catchment area. By using PCA approach, this technique benefits in determinig the sources of pollution which degrade the river water quality. The pollutant sources in river water quality are associated with weathering of mineral and soil (PC 1 and PC 4), as well as municipal, domestic and urban waste discharge (PC 2, PC 3, and PC5). Cluster analysis is used to group the similarity and differentiate of spatial and temporal variation based on monitoring stations. The outcome stated that only minor affect from non-point sources pollution is included to contaminate the river, whereby the water quality is suspected to control by surface runoff and wastewater discharges. Meanwhile, wet season highlighted the organic contamination with relatively high levels of contamination detected in the river. The contamination is related to municipal and domestic waste, as well as urban wastewater discharges (PC 2, PC 4, and PC 5) in basin area, which mainly detected in the S2, S3, S8, and S9. Only Kampung Tualang sub-basin (S5) remains clean within the basin. The water quality in Melaka River basin during wet season is expected to contaminate from point and non-point source pollution, and it is controlled by precipitation, surface runoff, mixed of natural, and wastewater discharges.

Overall, multivariate statistical techniques in large dataset to assess will provide important information in better understanding the surface water and able to assist in making decision for action plans. This study concludes that the management of industrial effluents as well as municipal and domestic wastes is needed to reduce the accumulation of contamination in river, and able to reduce the degradation towards environment. By suggesting municipal and industrial wastewater with proper treatment, improvement in agricultural practices, and landfills of municipal solid waste, this might be achieved to minimize the impact to the environmental degradation.

Acknowledgment

The authors would like to thank to Research Management Centre (PPP), Universiti Malaysia Sabah (UMS) for financial support and fully funding the publication of this article journal; continued by the Malaysian Department of 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 Melaka State.

Conflict of Interest

The authors declare no conflict of interest.

References

1. HUA A.K. Analytical and Detection Sources of Pollution Based Environmetric Techniques in Malacca River, Malaysia, Applied Ecology and Environmental Research, **15** (1), 485, **2017**.
2. ISMAIL A., TORIMAN M.E., JUAHIR H., ZAIN S.M., HABIR N.L.A., RETNAM A., KAMARUDDIN M.K.A., UMAR R., AZID A. Spatial assessment and source identification of heavy metals pollution in surface water using several chemometric techniques, Marine pollution bulletin, **106** (1), 292, **2016**.
3. KAMARUDDIN A.F., TORIMAN M.E., JUAHIR H., ZAIN S.M., RAHMAN M.N.A., KAMARUDIN M.K.A., AZID A. Spatial characterization and identification sources of pollution using multivariate analysis at Terengganu river basin, Malaysia, Jurnal Teknologi, **77** (1), 269, **2015**.
4. DASHORA M., KUMAR A., KUMAR S., KUMAR P., KUMAR A., SINGH C.K. Geochemical assessment of groundwater in a desertic region of India using chemometric analysis and entropy water quality index (EWQI), Natural Hazards, **112**, 747, **2022**.
5. MAI Y., ZHAO X., HUANG G. Temporal and spatial variability of water quality in an urban wetland and the effects of season and rainfall: a case study in the Daguan Wetland, China, Environmental Monitoring and Assessment, **194** (5), 1, **2022**.
6. ZHANG Z.M., ZHANG F., DU J.L., CHEN D.C. Surface Water Quality Assessment and Contamination Source Identification Using Multivariate Statistical Techniques: A Case Study of the Nanxi River in the Taihu Watershed, China, Water, **14** (5), 778, **2022**.
7. PANSERIYA H.Z., GOSAI H.B., VALA A.K., GAVALI D.J., DAVE B.P. Assessment of surface water of Gulf of Kachchh, west coast of India: A chemometric approach, Marine Pollution Bulletin, **170**, 112589, **2021**.
8. HUA A.K., KUSIN F.M., PRAVEENA S.M. Spatial Variation Assessment of River Water Quality Using Environmetric Techniques, Polish Journal of Environmental Studies, **25** (6), **2016**.
9. MOHAMED I., OTHMAN F., IBRAHIM A.I., ALAA-ELDIN M.E., YUNUS R.M. Assessment of water quality parameters using multivariate analysis for Klang River basin, Malaysia, Environmental monitoring and assessment, **187** (1), 4182, **2015**.
10. CHO Y.C., CHOI H., LEE M.G., KIM S.H., IM J.K. Identification and Apportionment of Potential Pollution Sources Using Multivariate Statistical Techniques and APCS-MLR Model to Assess Surface Water Quality in Imjin River Watershed, South Korea, Water, **14** (5), 793, **2022**.
11. SUKANYA S., SABU J. Water quality assessment using environmetrics and pollution indices in a tropical river, Kerala, SW coast of India, Current World Environment, **15** (1), 11, **2020**.
12. SHAN V., SINGH S.K., HARITASH A.K. Evaluation of water quality and potential metal contamination in ecologically important Bhindawas bird sanctuary, India, Applied Water Science, **11** (1), 1-9, **2021**.
13. MASTHURAH A., JUAHIR H., MOHD ZANURI N.B. Case study Malaysia: Spatial water quality assessment of Juru, Kuantan and Johor River Basins using environmetric techniques, Survey in Fisheries Sciences, **7** (2), 19, **2021**.
14. LOI J.X., CHUA A.S.M., RABUNI M.F., TAN C.K., LAI S.H., TAKEMURA Y., SYUTSUBO K. Water quality assessment and pollution threat to safe water supply for three river basins in Malaysia. Science of The Total Environment, **832**, 155067, **2022**.
15. LORENZ M., NGUYEN H.Q., LE T.D.H., ZEUNERT S., DANG D.H., LE Q.D., LE H., MEON G. Discovering Water Quality Changes and Patterns of the Endangered Thi Vai Estuary in Southern Vietnam through Trend and Multivariate Analysis, Water, **13** (10), 1330, **2021**.
16. LIMA D.R., GOMES A.A., LIMA E.C., UMPIERRES C.S., THUE P.S., PANZENHAGEN J.C., DOTTO G.L., EL-CHAGHABY G.A., DE ALENCAR W.S. Evaluation of efficiency and selectivity in the sorption process assisted by chemometric approaches: removal of emerging contaminants from water, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, **218**, 366, **2019**.
17. FU D., WU X., CHEN Y., YI Z. Spatial variation and source apportionment of surface water pollution in the Tuo River, China, using multivariate statistical techniques, Environmental Monitoring and Assessment, **192** (12), 1, **2020**.

18. SAALIDONG B.M., ARAM S.A., OTU S., LARTEY P.O. Examining the dynamics of the relationship between water pH and other water quality parameters in ground and surface water systems, PloS one, **17** (1), e0262117, **2022**.
19. NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (NOAA) Malacca Climate, Retrieved from http://tgftp.nws.noaa.gov/weather/current/WMKM.html, **2018**.
20. HUA A.K. Applied Chemometric Approach in Identification Sources of Air Quality Pattern in Selangor, Malaysia, Sains Malaysiana, **47** (3), 471, **2018**.
21. SAMSUDIN M.S., AZID A., RAZIK M.A., ZAUDI M.A., SHAHARUDIN S.M. Source of apportionment of Air Quality Parameters at Federal Port of Malaysia with Emphasis on Ship Emission. In IOP Conference Series: Earth and Environmental Science, **810** (1), 012052, **2021**.
22. UDDIN M.G., NASH S., OLBERT A.I. A review of water quality index models and their use for assessing surface water quality, Ecological Indicators, **122**, 107218, **2021**.
23. MFONKA Z., KPOUMIE A., NGOUH A.N., MOUNCHEROU O.F., NSANGOU D., RAKOTONDRABE F., TAKOUNJOU A.F., ZAMMOURI M., NGOUPAYOU J.R.N., NDJIGUI P.D. Water Quality Assessment in the Bamoun Plateau, Western-Cameroon: Hydrogeochemical Modelling and Multivariate Statistical Analysis Approach, Journal of Water Resource and Protection, **13** (2), 112, **2021**.
24. SHRESTHA N. Factor analysis as a tool for survey analysis, American Journal of Applied Mathematics and Statistics, **9** (1), 4, **2021**.
25. GYIMAH R.A.A., GYAMFI C., ANORNU G.K., KARIKARI A.Y., TSYAWO F.W. Multivariate statistical analysis of water quality of the Densu River, Ghana, International Journal of River Basin Management, **19** (2), 189, **2021**.
26. MATTA G., KUMAR A., NAYAK A., KUMAR P. Appraisal of spatial-temporal variation and pollution source estimation of Ganga River system through pollution indices and environmetrics in Upper Ganga basin, Applied Water Science, **12** (3), 1, **2022**.
27. ISLAM M.S., KHALID Z.B., GABAR S.M., YAHAYA F.M. Heavy metals pollution sources of the surface water of the Tunggak and Balok river in the Gebeng industrial area, Pahang, Malaysia, International Journal of Energy and Water Resources, 1-8, **2022**.
28. NATH A.V., SELVAM S., REGHUNATH R., JESURAJA K. Groundwater quality assessment based on groundwater pollution index using Geographic Information System at Thettiyar watershed, Thiruvananthapuram district, Kerala, India, Arabian Journal of Geosciences, **14** (7), 1, **2021**.
29. GORETZKO D., PHAM T.T.H., BUHNER M. Exploratory factor analysis: Current use, methodological developments and recommendations for good practice, Current Psychology, **40** (7), 3510, **2021**.
30. AMERICAN PUBLIC HEALTH ASSOCIATION (APHA). Standard methods for the examination of water and wastewater (21st Ed.), Washington, DC: **2005**.