Sains Malaysiana 50(3)(2021): 571-593
http://dx.doi.org/10.17576/jsm-2021-5003-02

Seasonal Effects on Spatial Variations of Surface Water Quality in a Tropical River Receiving Anthropogenic Influences

(Kesan Bermusim ke atas Variasi Ruang Kualiti Permukaan Air di Sungai Tropika yang Menerima Pengaruh Antropogen)

TENGKU NILAM BAIZURA TENGKU IBRAHIM, FARIDAH OTHMAN*, NOOR ZALINA MAHMOOD & TAHER ABUNAMA

ABSTRACT

This study investigates the seasonal and spatial water quality patterns along a tropical river that continuously receives various pollution sources. Multivariate analysis was used to study the spatial and temporal variations of the water quality parameters and to determine the origin of the pollution sources. Three regions (low, moderate, and high pollution levels) were determined based on cluster analysis. The stepwise DA mode proposed six parameters (pH, EC, COD, NO_3, TC, and Fe) with 75% correct assignations as the most significant water quality parameters to present the spatial variations. In the temporal discrimination, forward stepwise mode analysis showed eight parameters (EC, TUR, BOD, COD, AN, NO_3, Cu, and Cr) with 92% correct assignations, while five parameters (EC, AN, Al, Cu, and Cr) affording 89% correct assignations in backward stepwise mode analysis. Principal component analysis and factor analysis were used to investigate the origins of each water quality parameter based on the three clustered regions and successfully yielded eight latent factors loadings for each period that significantly identified the pollution sources and types along the river. The pollution sources for moderate and high pollution level areas are anthropogenic sources (landfill, industrial activities, and sewage discharge). Agricultural runoff is the main pollution source for the low pollution level areas. This study has shown classifications of river water quality based on seasonal and spatial criteria.

Keywords: Multivariate analysis; pollutants; spatial and seasonal variation; water quality

ABSTRAK

Penyelidikan ini mengkaji corak kualiti air bermusim dan ruang di sepanjang sungai tropika menerima pelbagai sumber pencemaran. Analisis multivariat digunakan untuk mengkaji variasi ruang dan temporal parameter kualiti air dan mengenal pasti sumber pencemaran. Tiga kumpulan (tahap pencemaran rendah, sederhana dan tinggi) ditentukan berdasarkan analisis kelompok. Mod DA langkah demi langkah mencadangkan enam parameter (pH, EC, COD, NO_3, TC dan Fe) dengan 75% penetapan yang betul sebagai parameter kualiti air yang paling signifikan untuk menunjukkan variasi ruang. Dalam diskriminasi temporal, analisis mod bertahap maju menunjukkan lapan parameter (EC, TUR, BOD, COD, AN, NO_3, Cu dan Cr) dengan 92% penetapan yang betul, sementara lima parameter (EC, AN, Al, Cu dan Cr) memberikan 89% penugasan yang betul dalam analisis mod bertahap mundur. Analisis komponen utama dan analisis faktor digunakan untuk mengkaji asal-usul setiap parameter kualiti air berdasarkan ketiga-tiga kelompok. Sumber pencemaran untuk kawasan paras pencemaran yang sederhana dan tinggi adalah sumber antropogen (tapak pelupusan, aktiviti industri, pelepasan kumbahan). Larian air pertanian adalah sumber pencemaran utama bagi kawasan paras pencemaran yang rendah. Kajian ini telah mendedahkan pengelasan kualiti air sungai berdasarkan kriteria bermusim dan ruang.

Kata kunci: Analisis multivariat; bahan cemar; kualiti air; variasi ruang dan bermusim

INTRODUCTION

Malaysia has experienced population growth and industrial development as a developing country. High-rise commercial buildings have become commonplace and the number of housing developments is skyrocketing at an unthinkable rate. These developments come at a price (Alssgeer et al. 2018; Bian et al. 2019). Pollution from unregulated man-made activities on most rivers in Malaysia is due to the lack of policy enforcement. Not only are the systems insufficiently designed and underfunded, but regulatory and management aspects remain shaky (Elfithri et al. 2011). Due to these, the government has agreed to review policies such as the Environment Act 1974, the Selangor Water Management Authority Enactment 1999, and the Water Services Industry Act 2006 to enforce higher penalties to the polluters (Ahmed et al. 2018). The focus of water resource management has always been on meeting growing water demands without properly considering the need to assure water quality, and maintain ecosystems and biodiversity.

Changes in land use due to deforestation, agriculture, and industrial and residential development have triggered major environmental effects such as increase in food production, decline in forest resources, regulation of climate conditions and air quality, spread of infectious diseases as well as deterioration of fresh water quality in many river systems. Continuous land changes can increase the sources of pollution in river systems (Yan et al. 2019). Various contaminants and heavy metals enter rivers from point sources (sewage treatment plants, landfills, animal farms, and factory effluents) and nonpoint inputs of natural and anthropogenic pollution origins (agricultural activities and surface runoff) (Ogwueleka 2014). In urban and industrial areas, the pollution loading on river basins combines more than one contaminant, such as nutrients, organic pollutants, and heavy metals (Mousa et al. 2018). All these can affect not only river to aquatic ecosystem, but also the provinces that use water as a domestic supply (Horn et al. 2017). At the same time, seasonal variations have different characteristics of water quality across different seasons. Thus, it is crucial to analyse water quality regularly and to describe the spatial and seasonal changes of water quality via frequent monitoring for competent environmental management (Zhang et al. 2017).

Selangor is considered as the fastest developing state in Malaysia (Abunama et al. 2018). Consequently, as many rivers in Selangor are heavily utilised to fulfil numerous developmental needs, their water quality status have changed. The quality degradation of such crucial water resources and their ecosystems directly affects the country's industry, agriculture, and living quality, and may result in long term economic losses (Le et al. 2017). According to the report by the Department of Environment (DOE 2017), out of 477 rivers in Malaysia, 46% were categorized as clean, 43% were categorized as slightly polluted and 11% as polluted. This classification is based on Water Quality Index (WQI), which was proposed by the DOE and this index is being practised in Malaysia for more than 25 years. The WQI is a single parameter that gives the overall status of the river water quality. It ranges from 0-100, and is calculated based on six parameters, namely dissolved oxygen (DO), biochemical oxygen demand (BOD, chemical oxygen demand (COD), total suspended solids (TSS), ammoniacal nitrogen (AN) and pH (DOE 2015; Idris et al. 2003; Othman et al. 2012). Based on the WQI readings, the river can generally be classified into 3 categories: Clean: WQI 81-100, Slightly polluted: WQI 60-80, and Polluted: WQI 0-59.

A more comprehensive National Water Quality Standards (NWQS) based on individual parameters are given in Table 1. It can be classified into 6 river classes (Class 1 - Class 5) according to the parameters stated in the table. As a mean of comparison, the US EPA water quality standards are tabulated in Table 2. Furthermore, 60% of rivers are used for domestic, crop production and industrial activities (Liu & Zou 2012). These are serious matters for decision makers who manage water resources as they need to conserve river water quality for future generations (Georgiva et al. 2013).

TABLE 1. National Water Quality Standards (NWQS) of Malaysia (DOE 2015)

	NWQS CLASSES					
Parameter	I	IIA	IIB	III	IV	V
AN (mg/L)	0.1	0.3	0.3	0.9	2.7	>2.7
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
pH	6.5 - 8.5	6 - 9	6 – 9	5 - 9	<3	<1
EC (µS/cm)	1000	1000	-	-	6000	-
TDS (ppm)	500	1000	-	-	4000	-
TSS (mg/L)	25	50	50	150	300	300
Turbidity (NTU)	5	50	50	-	-	-
TC (CFU/100 mL)	100	5000	5000	50000	50000	>50000
NO_3 (mg/L)		7	7	-	5	
PO_4		-	-	-	-	
Fe (µg/L)		1	1	1	5	
Al (µg/L)		-	-	0.06	0.5	
Mn (µg/L)	Natural levels or absent	0.1	0.1	0.02		Levels above IV
Cu (µg/L)		0.02	0.02	-	-	
Cr (µg/L)		0.05	1.4	0.1		
Zn (µg/L)		5	5	0.4	2	
Cd (µg/L)		0.01	0.01	0.01	0.01	
Pb (µg/L)		0.05	0.05	0.02	5	

TABLE 2. U.S. EPA water quality standards (USEPA 2001)

EU Directive or National (Ministerial Regulations)	Concentration
AN (mg/L)	0.2 - 4
BOD (mg/L)	5 - 7
COD (mg/L)	40
DO (mg/L)	5 - 9
pH	5.5 - 9
EC (μS/cm)	1,000
TDS (ppm)	-
TSS (mg/L)	50
Turbidity (NTU)	-
TC (CFU/100 mL)	5,000 - 100,000
NO_3 (mg/L)	50
PO_4	0.5 - 0.7
Fe (μg/L)	0.2 - 2
Al (μg/L)	200
Mn (μg/L)	0.05 - 1
Cu (μg/L)	0.05 - 1
Cr (μg/L)	0.05
Zn (μg/L)	3 - 5
Cd (μg/L)	0.005
Pb (μg/L)	0.05

Knowledge in obtaining water quality status as well as in identifying pollution sources needs to be prioritised in the implementation of effective and sustainable water management by continuously monitoring, assessing, and applying appropriate control and mitigation measures (VishnuRadhan et al. 2017). It is crucial to obtain baseline information on the spatial seasonal attributes and variations in river water quality (Elias et al. 2018). As river water quality is measured using numerous water quality parameters with different units on different time scales, it is necessary to apply environmental analyses to explore the rich information behind the collected datasets pertaining to water quality status and behaviour.

In water quality management, it is important to determine each water quality parameter to acquire collective water quality information, as it can provide succinct information on overall environmental conditions. The water quality indicators of important and influential variables are designed to give a single number to the water quality of a source based on a system that translates its existing concentrations into a single number in a sample (Wu et al. 2018). Various techniques can be utilised to study river water quality attributes and variations. In this study, the water quality parameters and heavy metals were classified based on the National Water Quality Standards (NWQS) for Malaysia by the Department of Environment to ensure that this study meets local standards. The United States Environmental Protection Agency (US EPA) standards were also used throughout the study period so that the outcome of this study can be used as a guide for new studies especially for those countries with similar seasonal events. Tables 1 and 2 show the water quality standards.

Large scale datasets from river water quality monitoring programmes require statistical and even trend analysis for extensive interpretation and determination of the pollution sources and variations. Multivariate analysis (MVA) includes different statistical exploratory techniques. Various recent studies in Malaysia and abroad have already established these techniques as suitable for analysing extensive scale of water quality data (Mavukkandy et al. 2014). MVA can assist with the statistical interpretation of complicated data matrices and account for the differences and similarities to better explain the spatial-seasonal variances of water quality. MVA is also capable of identifying the potential factors that affect such variances, such as water pollution sources. Water quality problems often vary from one region to another, even within a single country, from one year to another and from one season to another (Abunama et al. 2019). This is the first comprehensive study on water quality. No study has been carried out so far on seasonal effects and spatial variations of Sungai Sembilang. Furthermore, the outcomes of the research can provide a benchmark for exploring methods to protect human health and the ecosystem as well as an environmental reference for water pollution control of Sungai Sembilang and other tropical rivers that experience similar seasonal events.

The objectives are to apply MVA to determine the spatial-seasonal patterns and variations of the river water quality and identify the key water quality parameters responsible for the spatial-seasonal variances in river water quality due to the effects of point source pollutions. Unlike the simple identification and normal plotting/visualising of the pollution sources across the river, the presented methods in the article systemically and statistically identified the pollution variations and pointed out which parameters should be further monitored and investigated that have the high pollution loadings. Rather than seasonally classifying the water quality variations and spatially grouping the water sampling locations, this study analysis identified the most significant water quality parameters that caused the pollution loadings in each category and season. This analysis can have an important role in river water quality management by lowering the number of water parameters and focusing on the monitoring stations affected. In order to define the pollution sources and types along the river, PCA (Principle Component Analysis) and FA (Factor Analysis) were used to label the latent loading factors of the water quality datasets during the study period. This type of method is well established for assessing water quality. However, as far as the authors are aware, there is limited information of the seasonal and temporal analysis of the water quality using this method for a tropical river, particularly within the study area. Therefore, the outcomes form this study will give more information and can be used to assist watershed management decision-making in accomplishing the objectives of water quality.

MATERIALS AND METHODS

STUDY AREA AND MONITORING STATIONS

Sungai Sembilang, located in a tropical region, is a tributary of Sungai Selangor which runs through Kuala

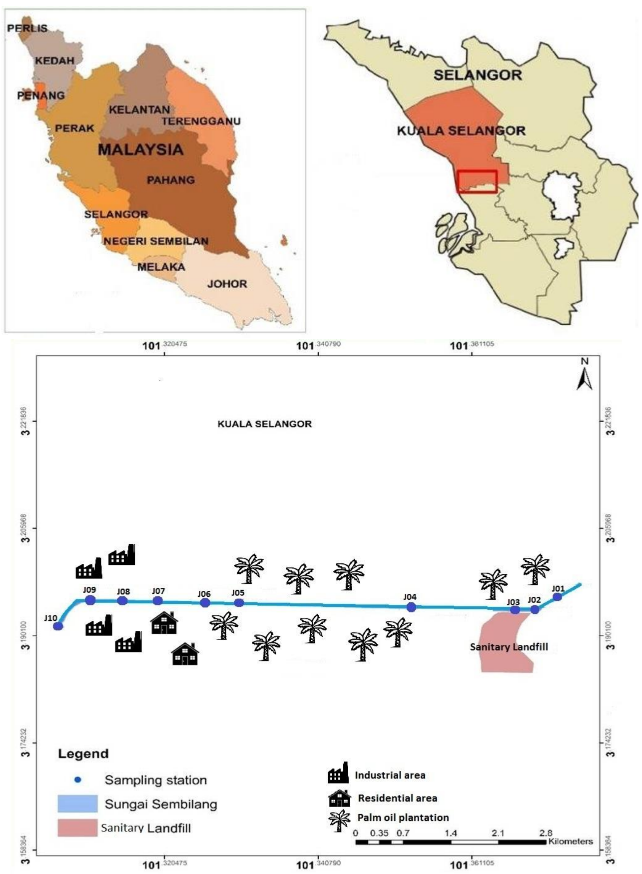


FIGURE 1. Study area and motoring stations locations

Selangor (Figure 1). The river flows through palm oil plantations from east to west towards the sea shore. It is facing real challenges due to effluents from a sanitary landfill and different pollution sources such as industrial factories, palm oil plantations, and small residential areas along the downstream of the river. At the same time, this river is also a major source of income for local people, namely fisheries and aquaculture activities. Pantai Remis, located downstream of the river, is also a major tourist destination in Kuala Selangor, Selangor. Malaysia faces two monsoon seasons, the Southwest and Northeast Monsoon seasons, with an average annual rainfall of about 2,155 mm. The significance of this study relies on the high demand of conducting assessment of one of the most important attributes of Sungai Selangor. This includes introducing a comprehensive analysis for case studies in the tropical regions (Figure 1). Among the selected stations, two stations are situated upstream of the river, one within the landfill site, and the rest are located downstream of the landfill site.

WATER SAMPLING

Sampling has been done for 10 times, started from May 2015 to September 2016 for 20 parameters for each monitoring stations; J01, J02, J03, J04, J05, J06, J07, J08, J09, and J10. Water sampling, samples preservation, in-suite measurements and laboratory tests were performed according to standard methods of examining water and wastewater (APHA 1988). The *in-situ* water quality parameters measured included pH, total dissolved solid (TDS), dissolved oxygen (DO), and turbidity. As for other parameters such as 5-day biochemical oxygen demand (BOD5), chemical oxygen demand (COD), ammonia nitrogen (AN), total suspended solid (TSS), nitrate (NO3), phosphate (PO4), total coliform (TC), and heavy metals, they were measured in the laboratory. After collecting the water samples, they were analysed. Table 3 summarises the 20 different water quality parameters as well as the applied analytical methods for each parameter.

TABLE 3. List of the measured water quality parameters and the applied analytical techniques

Category	Parameters	Units	Analytical methods
In-situ tests	pH		pH meter
	Electric Conductivity (EC)	µS/cm	Multiple meter
	Total Dissolved Solid (TDS)	ppm	
	Dissolved Oxygen (DO)	mg/L	DO Meter
	Turbidity (TUR)	NTU	Turbidity meter
Laboratory water quality tests	Total Suspended Solid (TSS)	mg/L	Standard method (APHA)
	Biochemical Oxygen Demand (BOD_5)		
	Chemical Oxygen Demand (COD)		
	Ammoniacal Nitrogen (AN)		
	Nitrate (NO_3)		
	Phosphate (PO_4)		
	Total coliform (TC)	CFU/100 mL	
Heavy metals	Iron (Fe)	µg/L	Inductively Coupled Plasma Atomic Emission Spectrometry (ICP-OES)
	Aluminum (Al)		
	Manganese (Mn)		
	Copper (Cu)		
	Chromium (Cr)		
	Zinc (Zn)		
	Cadmium (Cd)		
	Lead (Pb)		

DATA PREPARATION AND DESCRIPTIVE STATISTICS

The normality distribution of each parameter was checked prior to conducting analysis. It was tested using the Shapiro-Wilk's test and Q-Q plots. The natural logarithmic transformation was carried out where violations of normality assumptions occurred (Garson 2012; Hair et al. 1988; Van Ael et al. 2014; Voyles et al. 2012). The main descriptive statistics of the sampling results are listed in Table 4, which include the mean, standard error (SE), and standard deviation (SD), compared against Malaysia's National Water Quality Standards (NWQS) (DOE 2015). This is to ensure that this study meets local standards so that the outcome can be used as a guide for new studies especially for those countries with similar season conditions.

MULTIVARIATE EXPLORATORY TECHNIQUES

Multivariate statistical analysis indicates to multiple advanced tools for examining the connection between multiple variables consisting of more than one dependent variable (result or phenomenon of interest), more than one independent variable (predictor), or both (Hair et al. 1998). A number of different statistical techniques are available to perform multivariate analysis which vary with the study dataset type and key research questions. For the Sungai Sembilang water quality monitoring programme, Cluster analysis (CA), Discriminant analysis (DA), Principle component analysis (PCA), and Factor analysis (FA) were conducted.

Cluster Analysis (CA) aims to categorise different variables (e.g. sampling stations and monitoring frequencies) that have more similar characteristics than those in other clusters into separate groups or clusters. CA was implemented to classify the spatial-seasonal patterns and similarities in the river water quality by grouping the sampling sites and frequencies into different clusters based on water quality characteristics. HCA was performed using Ward's method by means of squared Euclidean distance as a similarity scale. The linkage distance was reported as (Dlink/Dmax)*100, which is the quotient between the linkage distance (Dlink) of one case over the maximal linkage distance (Dmax), and multiplied by 100 to standardise the distances (Hair et al. 1998).

Discriminant analysis (DA) is a method of defining the differences among the clusters pre-determined by CA. DA provides discriminant functions (DFs) for each group, which can be illustrated in the following equation:

$$f(Gi) = k_i + \sum_{j=1}^{n} \left(w_{ij} p_{ij}\right) \qquad (1)$$

where i is the number of groups (G); k_i is the constant inherent in each group; n is the number of parameters used to classify a set of data into a given group; w_j is the weight coefficient; assigned by DA to a given selected parameter (p_j). To determine the performance of DA, it provides a statistical classification table or prediction matrix that shows the correct and incorrect estimations.

If the DA prediction is effective, high correct percentages in the classification table will be yielded. After HCA, DA was conducted on the spatial-seasonal variations in river water quality. DA was performed on each clustered data matrix using standard, forward stepwise, and backward stepwise modes in constructing the DFs to evaluate both the spatial and seasonal variations in river water quality.

Principle component analysis (PCA) is a linear conversion of the original variables into uncorrelated variables projected on a new coordinate system. The projected variables are called principal components (PCs) (equation 2) (Hair et al. 1998). The first coordinate is called the first PC and holds the greatest variances among variables. The second greatest variances lie in the second PC followed by the third coordinate 'PC'. PCA defines the most meaningful variables capable of presenting the whole datasets and provides indices of the variation type in the analysed data.

$$z_{ij} = a_{i1} x_{1j} + a_{i2} x_{2j} + a_{i3}\, x_{3j} + \ldots + a_{im} x_{mj} \qquad (2)$$

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

Factor analysis (FA) can be defined as a data reduction technique that can reduce a large number of variables into a smaller number of factors. FA helps uncover the hidden patterns among variables and assists in clustering the highly interrelated variables into factors. This can be carried out by rotating the axis defined by PCA into new variables, called varifactors (VFs), as show in (3).

$$z_{ji} = a_{f1} f_{1i} + a_{f2} f_{2i} + a_{f3} f_{3i} + \ldots + a_{fm} f_{mi} + e_{fi} \qquad (3)$$

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

The resulting PCs from PCA were subjected to 'raw varimax rotation' in order to generate the VFs. Lastly, a small number of variables defined by FA were able to account for approximately the same information provided by the original datasets.

Results and Discussion

SPATIAL AND SEASONAL SIMILARITIES AND GROUPING USING CLUSTER ANALYSIS

The spatial application of the hierarchical cluster analysis (CA) on the monitoring stations helped to detect similarities between three main groups with respect to river water quality. The dendrogram in Figure 2 illustrates the three groups of ten monitoring stations for (Dlink/Dmax)*100 < 40. These clusters, namely Group A, Group B, and Group C were classified depending on the water quality parameters shown in Table 3. These clusters of monitoring stations indicate that each cluster has a water quality of its own which is different from the other clusters.

Group A corresponded to less polluted sites, which included the first two upstream sites (J01 and J02), as well as midstream stations (J06, J07, and J08) that are further downstream of the landfill and palm oil plantation areas. The stations J03, J04, and J05 were clustered in Group B, which exhibited a higher pollution level than Group A. In this group, three sites are located close to the landfill and palm oil plantation areas. The last two downstream stations J09 and J10 were included in Group C, which also showed higher pollution levels. This is probably due to the pollutants coming from nearby industrial factories and residential areas as well as their close proximity to the estuarine area, which receives discharge from the industrial area.

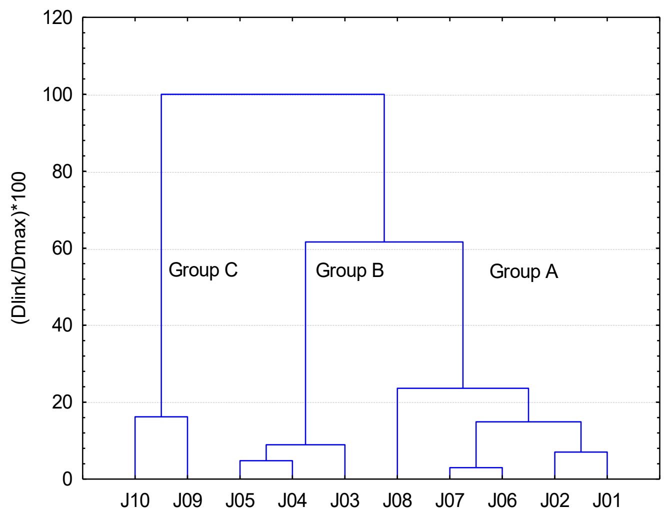


FIGURE 2. Dendrogram of spatial CA

TABLE 4. Mean, SE and SD of Sungai Sembilang water quality parameters throughout the study period

Parameter	NWQS Class I-V	Statistical parameters	J01	J02	J03	J04	J05	J06	J07	J08	J09	J10
pH	6.5-9	Mean	3.84	3.52	4.76	5.55	5.51	5.73	6.22	6.45	6.54	6.57
		SE	0.27	0.17	0.46	0.45	0.46	0.40	0.39	0.40	0.35	0.41
		SD	0.85	0.53	1.46	1.42	1.45	1.26	1.24	1.27	1.12	1.31
EC (μS/cm)	1000-6000	Mean	459.1	512.1	1,218	1,171	1,074	1,310	1,418	2,494	4,479	6,943
		SE	83.3	89.4	205.1	243.6	198.4	273.3	285.8	982.8	2,358	2,346
		SD	263.3	282.7	648.6	770.3	627.4	864.4	903.8	3,108	7,456	7,419
TDS (ppm)	500-4000	Mean	293.1	334.4	786.4	717.6	627.1	800.6	829.8	1,973	3,775	3,400
		SE	55.1	59.9	130.9	138.1	95.7	152.2	138.4	655.9	1,875	1,321
		SD	174.3	189.5	413.8	436.8	302.6	481.2	437.8	2,074	5,929	4,177

DO (mg/L)	7-3	Mean	4.20	4.17	3.88	3.21	3.70	4.26	4.21	3.38	3.76	4.32
		SE	0.77	0.65	0.69	0.55	0.54	0.62	0.59	0.39	0.62	0.79
		SD	2.44	2.06	2.19	1.73	1.71	1.96	1.87	1.23	1.96	2.50
Turbidity (NTU)	5-50	Mean	6.14	7.62	35.91	36.82	37.45	74.58	41.16	44.17	58.62	47.36
		SE	1.65	1.64	18.89	8.78	9.53	17.95	11.03	14.63	13.90	6.65
		SD	5.21	5.20	59.75	27.75	30.15	56.76	34.89	46.27	43.94	21.02
TSS (mg/L)	25-300	Mean	7.50	10.20	33.40	41.90	36.60	28.59	27.30	33.00	46.20	54.00
		SE	3.41	2.72	17.72	14.93	9.61	8.22	10.41	7.49	14.33	13.04
		SD	10.79	8.61	56.02	47.20	30.39	25.98	32.92	23.69	45.30	41.23
BOD (mg/L)	1-12	Mean	0.34	2.08	10.73	7.53	4.77	5.24	6.22	8.69	7.16	8.20
		SE	0.18	0.69	2.03	1.73	1.06	1.11	2.39	2.45	1.89	2.14
		SD	0.57	2.19	6.42	5.48	3.34	3.52	7.55	7.75	5.99	6.77
COD (mg/L)	10-100	Mean	19.10	44.45	134.47	90.61	84.25	86.45	79.94	91.88	74.36	59.69
		SE	4.80	7.39	33.81	14.53	22.10	19.01	15.82	18.95	13.85	9.53
		SD	15.19	23.37	106.91	45.96	69.90	60.11	50.04	59.91	43.79	30.14
AN (mg/L)	0.1-2.7	Mean	0.75	1.23	11.78	15.14	15.30	11.32	15.56	13.89	11.94	8.30
		SE	0.26	0.12	5.20	4.76	5.12	3.11	4.57	4.00	2.75	1.63
		SD	0.83	0.39	16.44	15.05	16.18	9.82	14.45	12.66	8.69	5.16
NO_3 (mg/L)	-	Mean	1.19	5.08	75.85	71.36	53.73	36.49	31.28	99.36	43.46	33.73
		SE	0.51	2.15	35.50	31.88	21.82	11.78	8.82	42.26	20.11	13.29
Parameter	NWQS Class I-IV	Statistical Parameters	J01	J02	J03	J04	J05	J06	J07	J08	J09	J10
		SD	1.62	6.80	112.27	100.82	69.01	37.24	27.88	133.65	63.59	42.03
PO4 (mg/L)	-	Mean	0.02	0.32	0.28	0.42	0.14	0.57	2.66	10.33	4.41	4.53
		SE	0.01	0.09	0.12	0.11	0.05	0.18	1.85	3.67	2.05	2.04
		SD	0.03	0.29	0.39	0.36	0.17	0.58	5.86	11.61	6.47	6.46
TC (CFU/100 mL)	100-50,000	Mean	2,100	2,350	10,875	2,750	3,825	9,450	4,562.5	1,063	4,650	5,038
		SE	787	919	3,345	864	727.7	1,361	594.4	327	612	2,165
		SD	2,489	2,907	10,577	2,733	2,301	4,305	1,880	1,035	1,935	6,847
Fe (µg/L)	1000-5000	Mean	1,667	2,736	4,623	4,533	3,991	2,822	3,016	2,493	2,289	1,927
		SE	196	344	431	828	730	720	1,069	881	800	694
		SD	620	1,089	1,363	2,619	2,309	2,278	3,379	2,785	2,531	2,194
Al (µg/L)	60-500	Mean	4,911	5,084	6,492	7,352	7,883	4,237	4,642	3,027	2,769	3,149
		SE	743	956	1079	1703	1665	906	1747	864	971	1326
		SD	2,229	2,867	3,236	5,110	4,993	2,718	5,241	2,590	2,911	3,977
Mn (µg/L)	100-200	Mean	497	568	663	792	813	804	793	584	705	559
		SE	57	102	74	118	107	104	138	157	154	118
		SD	180	323	233	372	339	330	438	497	487	373
Cu (µg/L)	20-200	Mean	31.6	22.9	27.6	26.1	40.9	24.4	30.7	22.1	21.3	19.8
		SE	14.0	9.6	11.7	10.5	10.0	10.3	12.1	10.3	9.3	8.9
		SD	44.2	30.4	37.0	33.3	31.7	32.4	38.3	32.7	29.4	28.1
Cr (µg/L)	50-100	Mean	5.41	3.53	12.01	7.70	9.32	5.78	7.50	4.69	7.08	5.55
		SE	1.34	1.07	3.73	1.65	1.63	1.37	2.01	1.12	1.84	1.14
		SD	4.23	3.40	11.78	5.23	5.15	4.34	6.34	3.56	5.83	3.59

Zn (µg/L)	5000-2000	Mean	132.6	194.1	639.7	172.7	528.8	184.9	228.7	276.5	266.7	259.7
		SE	20.1	64.0	331.9	29.9	195.5	26.6	23.9	37.6	32.0	50.3
		SD	63.7	202.5	1,050	94.4	618.2	84.2	75.5	118.8	101.3	159.1
Cd (µg/L)	10	Mean	1.43	1.47	1.58	1.47	6.49	1.46	1.43	1.67	1.47	1.44
		SE	0.71	0.74	0.70	0.68	3.08	0.67	0.65	0.65	0.59	0.57
		SD	2.24	2.33	2.22	2.14	9.74	2.12	2.06	2.05	1.86	1.81
Pb (µg/L)	50-5000	Mean	4.24	3.34	5.67	3.96	12.46	5.28	3.60	2.93	1.97	2.60
		SE	1.74	1.64	1.48	1.73	3.11	2.13	1.71	1.67	1.22	1.45
		SD	5.51	5.19	4.68	5.48	9.83	6.74	5.39	5.28	3.85	4.59

On the other hand, the seasonal hierarchical clustering divided the water quality datasets representing one year and a half (May 2015 to September 2016) into two seasons (dry season and wet season), as shown in the dendrogram in Figure 3. At distance (Dlink/Dmax)*100 < 80, the differences among these clusters/periods were significant. Dry season included mostly samplings during the southwest monsoon (SWM) season (October to March) while wet season comprised samplings from the northeast monsoon (NEM) season (May to September). During SWM, the area experiences heavy rain, so the water pollution is lower than the samplings from NEM. The river water quality is also affected by the river's hydrological conditions apart from the seasonal variations. Previous study showed that high volume of inflow following heavy rainfall promotes mixing and disturbs stratification in the river. The low and high precipitation during dry and wet seasons in a tropical country like Malaysia can greatly change the water quality of the river. High precipitation during the wet season can reduce the pollutant concentration by dilution and deteriorate the water quality of the river due to higher surface runoff from anthropogenic activities (Wang et al. 2016). Hierarchical CA provides a decent technique for categorising surface water in the study area and makes it viable to have a better future monitoring strategy that can decrease the number of monitoring periods and sites. For future studies, it is proposed that sampling rates are carried out in a span of multiple years to observe significant changes that may result from pollutant sources along the river.

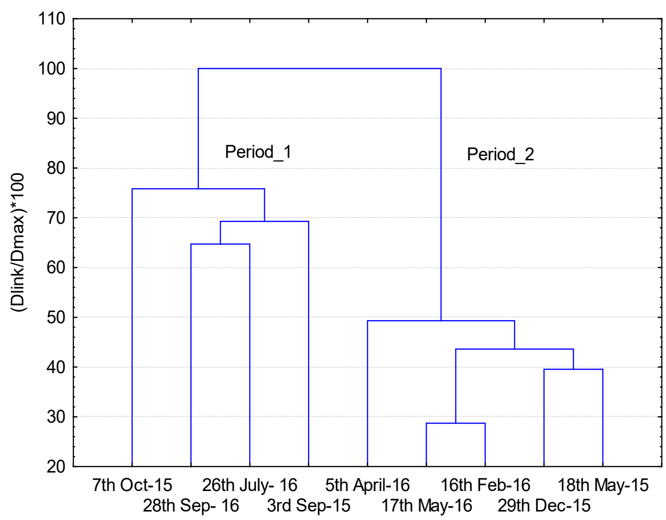


FIGURE 3. Dendrogram of temporal CA

SPATIAL AND SEASONAL VARIATIONS USING DISCRIMINANT ANALYSIS

The spatial variations between the three monitoring station groups were assessed using discriminant analysis (DA) after dividing the whole data sets to Group A, B, and C obtained through the spatial CA. The spatial DA was implemented using standard and stepwise modes in order to testify for the significance of the discriminant functions (DFs) and to define the most significant variables that created the differences among clusters. The Wilks' lambda and chi-square are values estimated for the DFs to ensure that the spatial DA is applicable (Table 5). The Wilks' lambda and chi-square values of the DFs in both standard and stepwise DA modes varied from 0.223 to 0.718 and from 31.3 to 131.2, respectively. The p-values were less than 0.001, which shows that the spatial DA is accurate and acceptable.

TABLE 5. Wilks lambda and Chi-square tests and p-values in the standard and stepwise spatial DA modes

Modes	DF	R	Wilks lambda	Chi-square	p-value
Standard DA mode	1	.780	.223	131.172	.000
	2	.655	.571	49.025	.000
Stepwise DA mode	1	.634	.430	79.768	.000
	2	.531	.718	31.276	.000

The discriminant functions (DFs) calculated with both standard and stepwise DA modes are listed in Tables 6 and 7, respectively. The standard DA mode yielded DFs with approximately 90%, 77%, and 60% correct prediction rates for the Groups A, B, and C, respectively. Overall, the total correct percentage was 80%. Hence, the most significant parameters via the standard DA mode were pH, DO, BOD, AN, and the heavy metals (Cr, Cd, Cu, and Pb). However, by applying the stepwise mode, the DA reduction produced DFs with 75% correct assignment by using 6 discriminant parameters. The suggested DFs in this mode belonged to parameters pH, EC, COD, NO_3, TC, and Fe. These DFs were the most significant water quality parameters to describe and account for ¾ of the expected spatial variations among the three groups.

TABLE 6. Classification matrix for the spatial DA modes

Modes	Cluster	Percent correct %	Cluster assigned by DA		
			Group A	Group B	Group C
Standard DA mode	Group A	90.0	45.0	7.0	8.0
	Group B	76.7	3.0	23.0	0.0
	Group C	60.0	2.0	0.0	12.0
	Total	80.0	50.0	30.0	20.0
Stepwise DA mode	Group A	76.0	43.0	10.0	12.0
	Group B	83.3	5.0	19.0	1.0
	Group C	60.0	2.0	1.0	7.0
	Total	75.0	50.0	30.0	20.0

TABLE 7. Discriminant functions (DFs) coefficients for the spatial DA modes

Parameter	Standard DA mode			Stepwise DA mode		
	Group A	Group B	Group C	Group A	Group B	Group C
pH	4.656	4.543	5.719	2.588	2.200	3.321
EC	.000	-.001	.000	.000	.000	.000
TDS	.000	.001	.000			
DO	3.091	2.841	2.746			
TUR	-.009	-.026	-.015			
TSS	-.007	.015	.021			
BOD	.144	.214	.121			
COD	.003	.013	-.018	-.001	.011	-.009
AN	-.206	-.209	-.296			
NO3	-.015	.010	-.022	-.008	.007	-.012
PO4	-.013	-.275	-.045			
TC	.003	.006	.006	.001	.003	.004
Fe	-.001	.000	-.001	.001	.001	.001
Al	.000	.001	.000			
Mn	.012	.008	.015			
Cu	.131	.073	.108			
Cr	.249	.433	.428			
Zn	.002	.003	.004			
Cd	-.204	-.130	-.146			
Pb	.120	.361	.066			
(Constant)	-24.368	-30.154	-32.142	-7.968	-11.070	-14.465

Figure 4 shows the mean ±0.95 confidence interval plots of the 6 discriminating parameters nominated by the spatial DA. The lower pH and EC levels at Group A and B stations were due to leachate whereby in Group C, the pH and EC increased to 6.57 and 6943 µS/cm, respectively, which added to the industrial sewage effluents from the last two monitoring stations. The pH falls within Class I according to NWQS and under US EPA (USEPA 2001) standards (i.e. 6-9). As the pH levels move away from this range, they can stress aquatic organisms and reduce hatching and survival rates. The solubility and toxicity of chemicals and heavy metals in water can also be affected by pH (Sarkar et al. 2007). Studies showed that the pH level for Sungai Sembilang is acidic due to the oxidation activity of pyrite (FeS_2) in the river soil. The acidity is also caused by the nitrification process in leachate treatment. Moreover, the higher pH reading in Group C is caused by the palm oil plantation near the landfill where NPK fertiliser is used (Psaltopoulos et al. 2017).

In contrast, the EC values are higher than both NWQS and US EPA standards (USEPA 2001). Conductivity has no direct effects on human health. Variances in EC values are based on factors such as agricultural and industrial activities and land use, which impact the mineral content and thus, the water's EC.

High concentrations of COD (Figure 4(c)) and NO_3 (Figure 4(d)) were observed at the Group B sites. This can be attributed to the treated leachate effluents from the nearby landfill as well as the use of fertilisers that contain nitrogenous elements in palm oil plantation areas at these sites. The COD level was greater than 50 mg/L except for Group A due to the higher rate of oxygen consumption from water. The COD values of water samples ranged between minimum 19.10 mg/L in Group A and maximum 134.47 mg/L in Group C. Meanwhile, higher NO_3 (31.28-99.36 mg/L) concentrations were found downstream of the river due to fertiliser runoff from the nearby palm oil plantation. Both parameters exceed the NWQS for Class

V (Table 1) and the US EPA standards (USEPA 2001). Stevenson and Rollins (2017) also reported high nitrate content in the river due to plantation activity which contributes nutrients to the body of water.

Nitrate enrichment in rivers can lead to increased algae and macrophytes growth, reduced biodiversity, and odour problems (Stevenson & Rollins 2017). In addition, the high values of TC (Figure 4(e)) and Fe (Figure 4(f)) in this group (Group C) are related to biological and heavy metal pollutions mainly due to treated effluents caused by young leachate. Based on landfill age, the leachate is generally classified as young or stabilised (5-10 years old) with low pH values and high organic matter content and biodegradability of heavy metals (Corsino et al. 2020; Stefania et al. 2018). The TC values varied from 1063 CFU/100 mL in Group A to 10875 CFU/100 mL in Group C, which fall within class III according to NWQS. They also do not meet the US EPA standards (USEPA 2001). TC is widely measured and used as indicators of the presence of pathogenic microbes that pose a threat to people, animals, and aquatic life (Chatanga et al. 2019). The Fe values of surface water of all groups were > 1000 µg/L and they are within the recommended limit of the NWQS Class IV (i.e. 1000-5000 µg/L). However, the values exceed the US EPA standards (USEPA 2001). The impact of Fe contamination can minimise the occurrence and diversity of numerous aquatic organisms, such as fish. High Fe concentrations along with their precipitation in aquatic ecosystems do have harmful effects on the behaviour, reproduction, and survival of aquatic animals (Edokpayi et al. 2016).

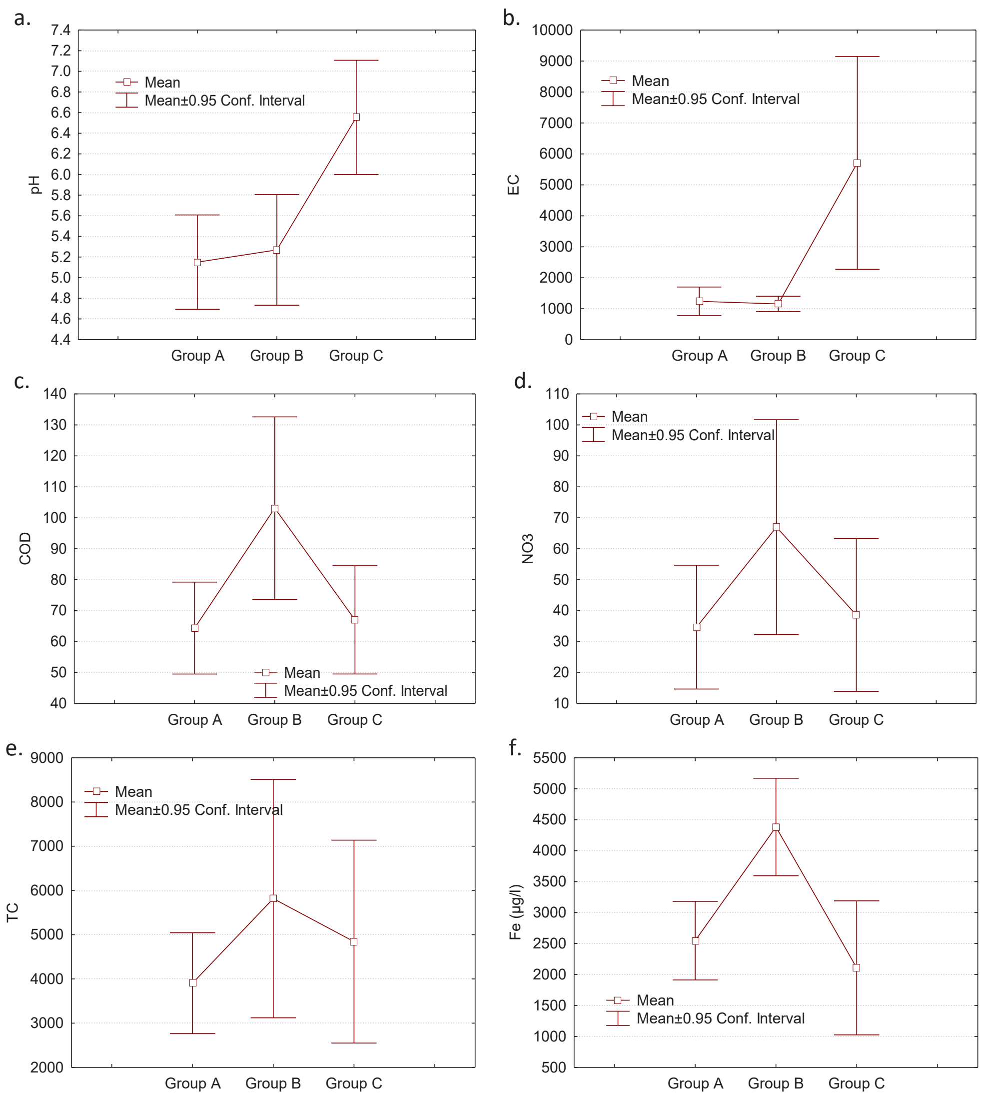


FIGURE 4. Mean ±0.95 confidence interval plots for the 6 parameters (pH, EC, COD, NO_3, TC and Fe) recognized with the spatial DA

The seasonal DA was conducted using standard, and both forward and backward stepwise modes. Forward and backward stepwise modes gave a good sequence and showed correlation coefficients (R) of 0.831, 0.826, and 0.777, respectively (Table 8). For the two clustered seasons, the Wilks' lambda and the Chi-square for the DFs were ranged from 0.310 to 0.396 and from 88.6 to 105.7 for the three modes of DA. Table 8 shows that the p-values of less than 0.001 indicated that the seasonal DA is reliable, as the DFs were correlated with seasonal variations among the two clusters.

TABLE 8. Wilks lambda and Chi-square tests and the p-values of the temporal DA modes

Modes	DFs	R	Wilks lambda	Chi-square	p-value
Standard DA mode	1	.831	.310	103.012	.000
Forward stepwise DA mode	1	.826	.317	105.715	.000
Backward stepwise DA mode	1	.777	.396	88.578	.000

Both standard and forward DA modes were showed to a correct percentage of 92% of the cases in the two seasons (Table 9). Hence, the most significant DFs were EC, TUR, BOD, COD, AN, and NO_3, as well as heavy metals such as Cu and Cr as shown in Table 9. Meanwhile, in the backward stepwise mode, the DA produced DFs with 89% correct assignment using only five discriminant parameters, namely EC, AN, Al, Cu, and Cr (Table 10). Therefore, the seasonal stepwise DA modes have suggested that parameters such as EC, TUR, BOD, COD, AN, and NO_3, as well as heavy metals (Al, Cu and Cr) as the most vital parameters to distinguish between the two seasons and account for the most predicted seasonal variations in the river water quality.

TABLE 9. Classification matrix for the temporal DA modes

Modes	Cluster	Percent correct %	Cluster assigned by DA	
			Period_1	Period_2
Standard mode	Period_1	93	56.0	4.0
	Period_2	90	4.0	36.0
	Total	92	60	40
Forward stepwise mode	Period_1	93	56.0	4.0
	Period_2	90	4.0	36.0
	Total	92	60	40
Backward stepwise mode	Period_1	92	55.0	5.0
	Period_2	85	6.0	34.0
	Total	89	61	39

TABLE 10. Discriminant function's (DF's) coefficients for the temporal DA modes

Parameter	Standard DA mode		Forward stepwise DA mode		Backward stepwise DA mode	
	Period_1	Period_2	Period_1	Period_2	Period_1	Period_2
pH	4.700	4.585				
EC	.000	.000	.001	.000	.001	.000
TDS	.000	.000				
DO	2.999	3.239				
TUR	-.027	.004	-.020	.011		
TSS	-.026	-.021				
BOD	.334	.176	.312	.137		
COD	.022	.005	.033	.015		
AN	-.104	-.198	.022	-.069	.111	.012
NO3	-.021	-.011	.005	.014		
PO4	-.040	-.083				
TC	.000	.000				
Fe	-.001	-.001	-.001	.000		
Al	.001	.000	.000	.000	.001	.000
Mn	.013	.010	.011	.008		
Cu	.249	.116	.150	.013	.084	-.009
Cr	-.157	.268	-.261	.143	-.041	.255
Zn	.002	.002				
Cd	-.352	-.197	.061	.227		
Pb	.112	.123				
(Constant)	-29.784	-24.083	-10.755	-4.875	-6.494	-2.334

Figures 5 and 6 show the mean ±0.95 confidence interval plots of the discriminant parameters selected by the seasonal DA. The electrical conductivity (EC) is due to the dissolved salts present in the water. EC values of the river water in both dry and wet seasons are within the NWQS and US EPA standards which are 6000 and 1000 µS/cm, respectively (Sarkar et al. 2007). The average concentration of electrical conductivity (Figure 5(a)) was higher in the dry season (2725 µS/cm) compared to the wet season (1181 µS/cm) due to the dilution impact of high rainfall and runoff. The average BOD concentrations (Figure 5(c)) followed the same trend with higher values observed in the dry season. The lower concentrations of BOD recorded in wet season (4.94 mg/L) compared to dry season (6.865 mg/L) is a result of higher concentrations of organic substances. The maximum BOD value to maintain the presence of aquatic organisms must be 6 mg/L for NWQS and 5 mg/L for US EPA standards (USEPA 2001). It is important to keep the BOD below these values for aquatic organisms such as fishes, molluscs, crustaceans, and other organisms to survive. As the values of BOD have exceeded the maximum requirement for healthy ecosystem, it is important for the environmental planner and management to incorporate the best management practices and to upgrade this river for sustainability and livelihood of the local people.

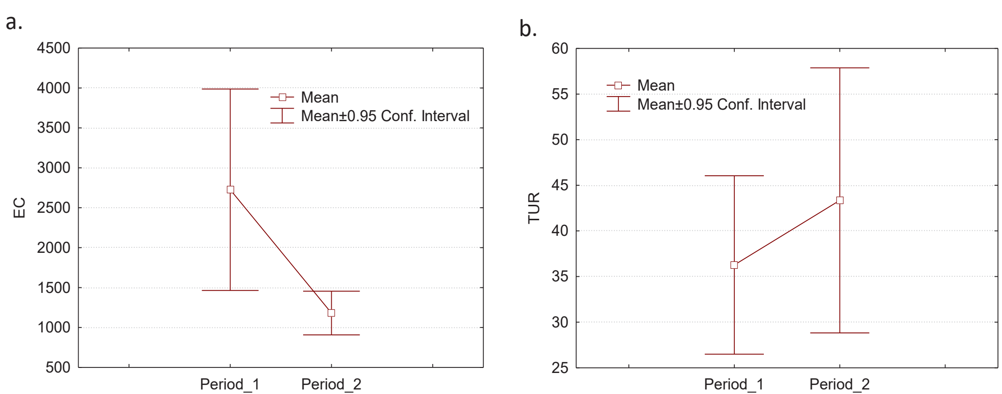
Due to the runoff effects, the TUR levels were higher in the wet season as shown in Figure 5(b). Urban land cover often corresponds with increased loads of suspended solids and decreased water quality (Cunha et al. 2016). The differences in TUR levels were recorded during dry season (36.27 NTU) and wet season (43.35 NTU). The values fall within the acceptable limit of NWQS for Malaysian rivers and are categorised as Class II. Higher TUR can negatively affect fish and other aquatic life by reducing food supplies, degrading spawning beds, and affecting gill functions. This can also decrease the river's aesthetic quality and adversely impact recreation and tourism (Matta 2015).

The COD levels (Figure 5(d)) slightly differed between the two seasons and were higher in the wet season. This is due to pollutant runoff from surrounding areas. Matta (2015) demonstrated the same results where the overall water quality of Rwizi catchment in Uganda was low, particularly during the dry season. The values are within the NWQS, Class IV but exceed the US EPA standards (USEPA 2001) (i.e. 100 and 40 mg/L, respectively). This could be related to the leaching and transports of natural and domestic sewage, and agricultural and industrial pollutants (Ojok et al. 2017).

The primary source of AN is fertilisers that are used for agricultural activities within the study area. AN and NO_3 in Figure 5(e) and 5(f) showed an inverse pattern in the two seasons. In the dry season, the AN concentration was higher (12.70 mg/L) than that in the wet season (7.28 mg/L). Based on NWQS and US EPA standards (2001), the value of AN along this river is at the maximum level (0.2 mg/L). In contrast, the NO_3 level were higher in the wet season (77.77 mg/L) while lower concentration of NO_3 was recorded in the dry season with 23.40 mg/L. Nitrogen is the most widely used nutrient for agricultural production. In addition to fertilisers, nitrogen naturally occurs in the soil in organic forms decomposing plant materials, animal residues, and domestic sewage. The concentration of unionised ammonia (<0.2 mg/L) can affect some species of fish. Eutrophication and methemoglobinemia will occur at high nitrate concentrations (Barakat et al. 2016). During the study period, NO_3 concentrations are higher than the limits of NWQS and US EPA standards which are 5 and 50 mg/L, respectively (USEPA 2001).

As for heavy metals, the main three parameters were Al, Cu, and Cr. Figure 6 shows a similar trend for the selected metals by the DA. In dry season, the heavy metal concentrations were higher than in the rainy season because of the dilution effect. The values for Al are above the limits of both NWQS and US EPA standards (USEPA 2001) (i.e. 60-500 µg/L and 200 µg/L, respectively) and can be classified under Class V (NWQS). High levels of Al may impact the ability of certain species to regulate ions, such as salts, and may constrain respiratory functions, including breathing. Al can build up on the surface of the gill of fish, resulting in respiratory dysfunction and eventually death (Cano-Rocabayera et al. 2019). This is particularly alarming because fishing activities are a major source of income to the locals. If Al high concentration continues, fishing activities can be affected in the study area. As for Cu and Cr, the average values along the river are within the permissible standards of US EPA (USEPA 2001). Based on NWQS, Cu can be classified under Class III while Cr is classified under Class I. Anthropogenic sources of Cu include agriculture, and pesticide use. Cu is an essential nutrient at low concentrations, but at higher concentrations, it is harmful to aquatic organisms that can adversely affect survival, growth, reproduction, and mortality (Jaishankar et al. 2014).

Results have shown that the stepwise DA can generate DFs and recognise the most significant variables in seasonal and spatial variations. Hence, DA has pointed out a few parameters responsible for large variations in water quality that could decrease the number of sampling parameters. The results of spatial and seasonal CA have also supported the trends of discriminant parameters in water quality. A combination of low discharge and high residence times of water with constant discharge from agriculture, industry, and urban areas as well as diffuse polluted waters can expedite the decline of water quality. Previous studies reported higher concentration of water quality parameters in dry season. Increased pollution values and anthropogenic drainage water depletions have caused critical situation in the water quality of Sungai Sembilang downstream. Extreme weather events such as severe droughts, rainfall, and floods can adversely impact the water supply system. Mitigation techniques are needed to prevent decline in water quality and damage to public health. Our results have suggested that seasonal variations clearly affect the river water quality.



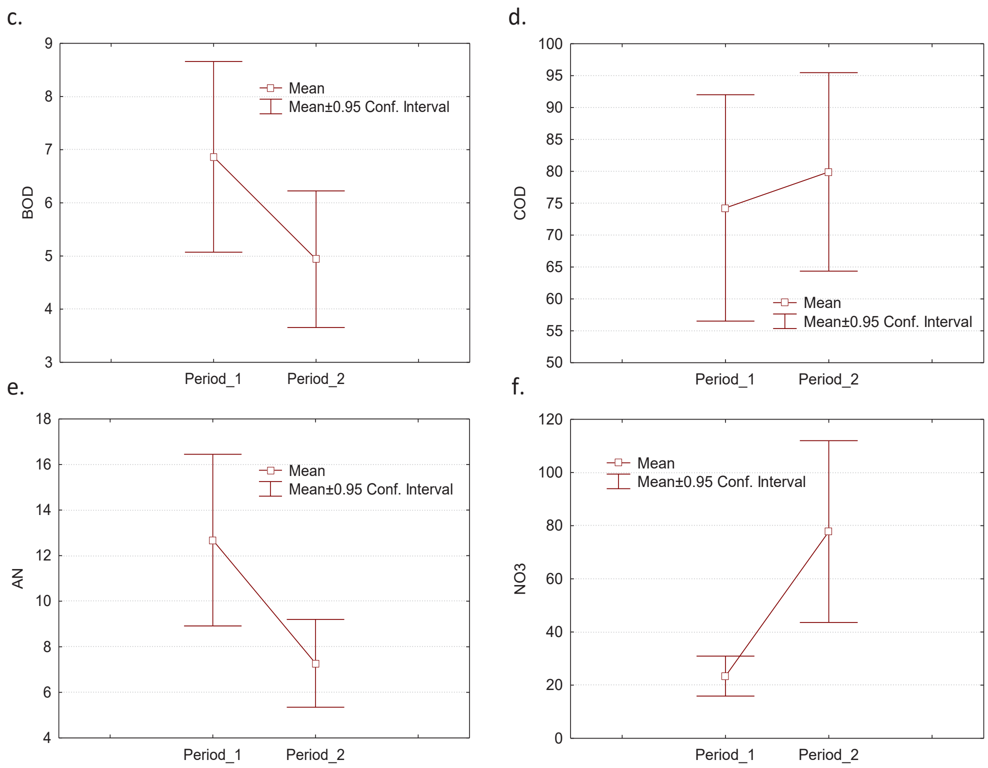


FIGURE 5. Mean ±0.95 confidence interval plots for the water quality parameters selected by the temporal DA (EC, TUR, BOD, COD, AN and NO_3)

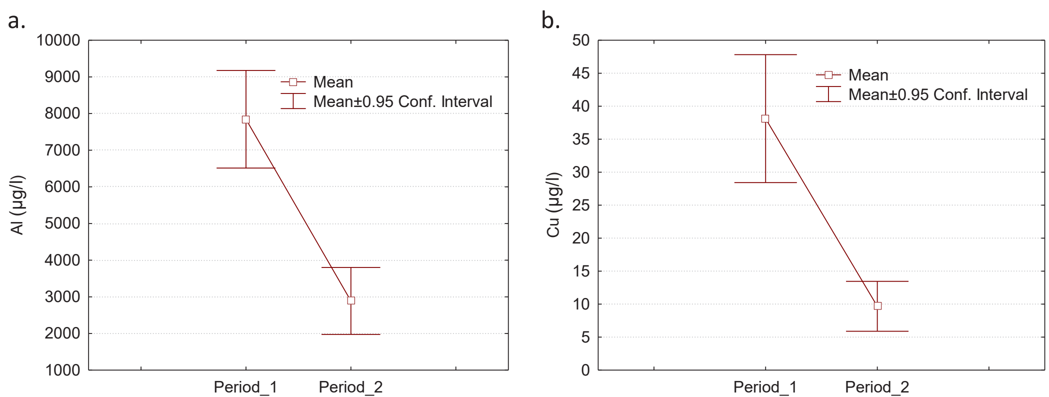


FIGURE 6. Mean ±0.95 confidence interval plots of some selected heavy metals with the temporal DA (Al and Cu)

POLLUTION SOURCE IDENTIFICATION USING PRINCIPLE COMPONENT ANALYSIS (PCA) AND FACTOR ANALYSIS (FA)

PCA/FA was used to examine the differences within the clustered groups or seasons, and to determine the latent factors. As shown in the scree plot of eigenvalues in Figure 7, eigenvalues equal to or greater than 1 are considered significant. The analysis yielded six varifactor (VF) loadings or principal components (PCs) for each season.

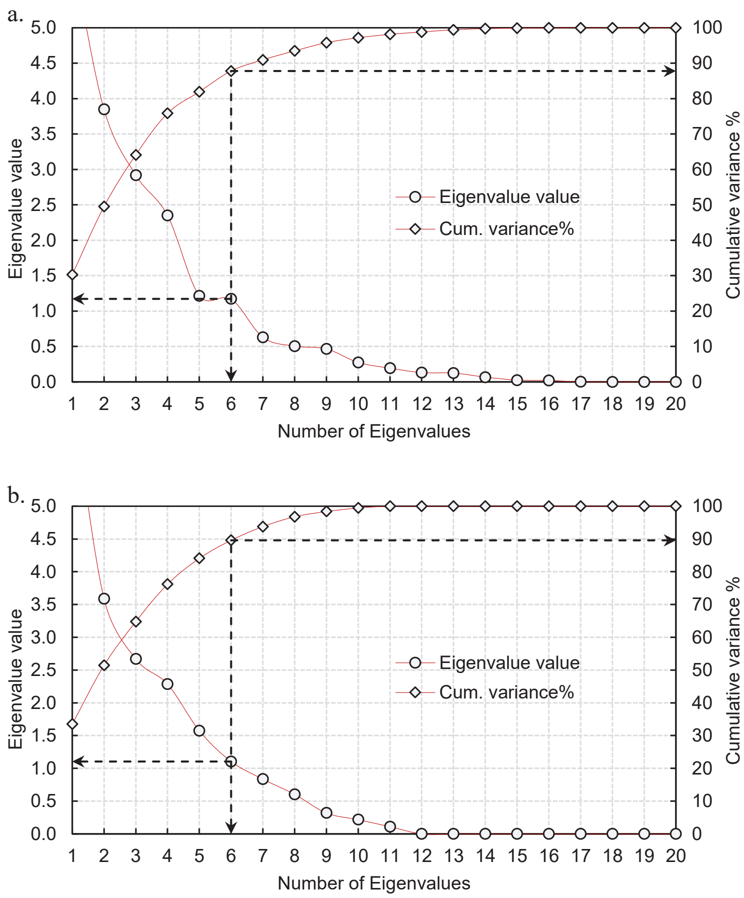


FIGURE 7. Screen plot of eigenvalues in group B a. Period_1 b. Period_2

These factors explained 87.8% and 89.6% of the total variances of water quality in dry season and wet season, respectively.

Table 11 summarises the varifactor (VF) loadings, eigenvalues, percentage of variances, and cumulative variance contributions in percent for both seasons in group B Factor loading can be categorised as strong for loading value >0.75 and <-0.75. Meanwhile, the categorisation is moderate for values between ±0.75 and ±0.5, and weak for the range of ±0.5 to ±0.3. In dry season, VF1 explained 30.7% of the total variances in Group B stations which had strong loadings in heavy metals and moderate negative loadings in DO, and positive moderate loadings in BOD and PO_4. VF1 represents oxygen depletion, and organic and heavy metal pollutions due to treated landfill leachate discharge in the river. VF2 (19.2% of the total variance in the group) had high loadings in pH, TDS, NO_3, and Fe, and mild loadings in NO_3 and Cu. VF3 (14.6% of the total variance) had high positive loadings in TUR, TSS, and BOD. The last three VFs, represented by 11.75%, 6.08%, and 5.86% of the total data variances, had high positive loadings in COD, AN, and some heavy metals, and high negative loadings in TC. The first VF explained 33.49% of the data variance in the second sampling period. It had strong positive loadings in TSS, AN, and heavy metals. VF2 (17.9%) also showed strong positive loading values in EC, TDS, and NO_3. VF3 (13.3%) showed negative loadings in pH and strong positive of some heavy metals i.e. Fe and Al. The last three VFs, however, indicated various pollutant types as shown in Table 12.

TABLE 11. Loading factors of the water quality parameters in Group B for the two periods

Parameter	Period_1						Period_2					
	VF1	VF2	VF3	VF4	VF5	VF6	VF1	VF2	VF3	VF4	VF5	VF6
pH	.760	.106	.449	-.212	.290	.202	.258	.310	-.843	-.095	.141	.196
EC	.147	.842	.416	.152	-.041	-.047	-.132	.942	-.242	-.090	.042	.013
TDS	-.045	.899	-.224	.155	.023	-.094	-.128	.947	-.221	-.102	.039	.067
DO	-.555	.318	-.169	-.158	-.463	-.326	-.130	.279	-.033	.139	.312	.857
TUR	-.001	-.063	.951	.034	-.134	.008	.367	-.368	.308	-.145	.675	.162
TSS	.001	-.066	.896	.052	.325	.079	.857	-.143	.063	-.061	.095	-.155
BOD	.554	.002	.538	.134	.362	-.255	-.153	-.299	.125	-.216	-.888	-.078
COD	.255	.279	.323	.787	.094	-.105	.398	-.123	.176	.065	-.809	-.005
AN	.013	-.025	.035	.336	.798	.151	.766	-.195	-.364	-.013	.216	.326
NO3	.005	.687	.343	.046	.608	.047	-.254	.873	-.264	.051	.224	.177
PO4	.573	-.240	.080	-.025	.408	-.134	.163	.234	-.174	-.636	.102	.648
TC	-.118	.172	-.044	-.134	-.130	-.904	.003	-.390	-.220	-.784	.256	-.181
Fe	.006	-.873	.237	.065	.154	.069	.029	-.455	.736	.152	-.199	-.123
Al	-.877	-.117	.001	-.333	-.062	.038	.357	-.338	.778	.149	.144	-.211
Mn	-.926	.011	-.037	-.174	-.056	.100	.055	.128	.406	.222	.167	-.853
Cu	.509	-.654	.020	.250	-.009	.457	.778	-.127	.293	.164	-.161	-.349
Cr	.258	-.318	.080	.810	.347	.057	.256	-.093	.004	.780	.124	-.269
Zn	-.194	.155	-.152	.878	.080	.248	.766	-.102	.214	.342	-.412	-.116
Cd	.863	-.027	-.056	-.150	-.152	.301	.725	-.096	-.244	.364	.044	.231
Pb	.728	-.009	-.160	-.070	-.434	.331	.240	-.387	.139	.663	.188	.030
Eigenvalue	6.05	3.85	2.92	2.35	1.22	1.17	6.70	3.59	2.67	2.29	1.57	1.10
Percentage of variance	30.27	19.24	14.60	11.75	6.08	5.86	33.49	17.94	13.34	11.43	7.87	5.52
Cumulative %	30.27	49.51	64.11	75.86	81.94	87.80	33.49	51.42	64.77	76.19	84.06	89.58

Similarly, the previous procedure was applied on the two seasons for Group A and Group C to summarise and identify the pollution's pattern and seasonality (Tables 13 & 14). Table 11 lists the results of the pollution sources and features based only on the first four loading factors. Overall, the results have indicated that there are different pollution types along the river during the study period, which include organic pollutants, nutrients, phosphorous, turbidity, salinity, fecal pollutant, heavy metals, and natural pollutant. For Group A sites located along the upper river stream, the main effects observed were oxygen depletion, salinity, nutrients, phosphorous, and heavy metals. The effects were due to nearby agricultural activities mainly palm oil plantation and domestic activities from small residential areas. Mousa et al. (2018) reported that non-point source is primarily from agricultural activities where atmospheric deposition is a major source of nutrient pollution which highlights ammonia nitrogen. Meanwhile, industrial and domestic waste water lead to organic pollution (BOD and COD). Lastly, natural pollution is mainly affected by the meteorological variations (temperature and DO) (McKinley et al. 2019). This can be supported by the fact that as water temperature increases in the river, biological activities of aquatic organisms strengthen, hence, the concentration of oxygen consumption rises. More oxygen also dissolves in cooler water (Dobsa et al. 2014).

TABLE 12. Pollution source identification based on the significant VFs for each period

Periods	VF1	VF2	VF3	VF4	VF5	VF6
Group A						
Period_1	Oxygen depletion Heavy metals	Salinity Nutrients Phosphorous	Turbidity Natural pollution	Heavy metals	Organic pollution Nutrients	Fecal pollution
Period_2	Salinity Phosphorous Natural pollution	Nutrients Natural pollution	Oxygen depletion Turbidity Heavy metals	Heavy metals	Turbidity Organic pollution Heavy metals	Oxygen depletion Phosphorous
Group B						
Period_1	Oxygen depletion Organic Pollution Phosphorous Heavy metals	Salinity Nutrients Heavy metals	Turbidity Organic pollution	Organic pollution Heavy metals	Nutrients	Fecal pollution
Period_2	Turbidity Organic pollution Heavy metals	Salinity Nutrients	Salinity Heavy metals	Phosphorous Heavy metals Fecal pollution	Turbidity Organic pollution	Oxygen depletion Phosphorous
Group C						
Period_1	Salinity Oxygen depletion Heavy metals	Salinity Heavy metals	Organic pollution Nutrients Fecal pollution	Turbidity Organic pollution	Turbidity Organic pollution	Organic pollution
Period_2	Salinity Organic pollution Heavy metals	Organic pollution Nutrients Phosphorous Heavy metals	Turbidity	Oxygen depletion Nutrients Organic pollution	Oxygen depletion Organic pollution Nutrients	Fecal pollution

Meanwhile, at group B sites, the VFs indicated were organic pollution, turbidity, nutrients, phosphorous, coliform contamination, and heavy metal pollution. These can be explained by the effects of treated leachate effluents which receive solid waste from various sources as well as fertilisers from palm oil plantation along the river (Ebrahimi et al. 2017; Hajigholizadeh & Melesse 2017; Zhang et al. 2018). At the last two monitoring sites, Group C, high salinity and oxygen depletion were observed in VF1 during the two monitoring periods due to the sewage effluents from nearby industrial factories. Different types of pollution along the river indicate that the river has various physical and chemical characteristics based on various natural and anthropogenic factors. The spatial differences in dry season suggest that water quality problems in Group B and C are worse than Group A. Thus, this region should be given more attention. This enhances the applicability of decision-makers to obtain optimal solutions for the final decision-making scheme taking into account the technical and economic feasibility of the levels of pollutant treatment.

TABLE 13. Loading factors of the water quality parameters on the significant VFs in Group A for the two seasons

Parameter	Dry Season				Wet Season			
	VF1	VF2	VF3	VF4	VF1	VF2	VF3	VF4
pH	0.760	0.106	0.449	-0.212	0.258	0.310	-0.843	-0.095
EC	0.147	0.842	0.416	0.152	-0.132	0.942	-0.242	-0.090
TDS	-0.045	0.899	-0.224	0.155	-0.128	0.947	-0.221	-0.102
DO	-0.555	0.318	-0.169	-0.158	-0.130	0.279	-0.033	0.139
TUR	-0.001	-0.063	0.951	0.034	0.367	-0.368	0.308	-0.145
TSS	0.001	-0.066	0.896	0.052	0.857	-0.143	0.063	-0.061
BOD	0.554	0.002	0.538	0.134	-0.153	-0.299	0.125	-0.216
COD	0.255	0.279	0.323	0.787	0.398	-0.123	0.176	0.065
AN	0.013	-0.025	0.035	0.336	0.766	-0.195	-0.364	-0.013
NO3	0.005	0.687	0.343	0.046	-0.254	0.873	-0.264	0.051
PO4	0.573	-0.240	0.080	-0.025	0.163	0.234	-0.174	-0.636
TC	-0.118	0.172	-0.044	-0.134	0.003	-0.390	-0.220	-0.784
Fe	0.006	-0.873	0.237	0.065	0.029	-0.455	0.736	0.152
Al	-0.877	-0.117	0.001	-0.333	0.357	-0.338	0.778	0.149
Mn	-0.926	0.011	-0.037	-0.174	0.055	0.128	0.406	0.222
Cu	0.509	-0.654	0.020	0.250	0.778	-0.127	0.293	0.164
Cr	0.258	-0.318	0.080	0.810	0.256	-0.093	0.004	0.780
Zn	-0.194	0.155	-0.152	0.878	0.766	-0.102	0.214	0.342
Cd	0.863	-0.027	-0.056	-0.150	0.725	-0.096	-0.244	0.364
Pb	0.728	-0.009	-0.160	-0.070	0.240	-0.387	0.139	0.663
Eigenvalue	6.05	3.85	2.92	2.35	6.70	3.59	2.67	2.29
% of Variance	30.27	19.24	14.60	11.75	33.49	17.94	13.34	11.43
Cumulative %	30.27	49.51	64.11	75.86	33.49	51.42	64.77	76.19

TABLE 14. Loading factors of the water quality parameters on the significant VFs in Group C for the two seasons

Parameter	Dry Season				Wet Season			
	VF1	VF2	VF3	VF4	VF1	VF2	VF3	VF4
pH	-0.028	0.857	0.267	0.256	-0.824	0.242	-0.294	0.276
EC	-0.883	0.332	0.024	0.059	-0.860	0.402	-0.074	-0.116
TDS	-0.899	0.248	-0.253	0.020	-0.858	0.407	-0.080	-0.106
DO	-0.826	0.175	0.069	0.189	0.217	-0.074	0.009	0.835
TUR	0.001	0.254	0.029	0.898	0.174	0.025	0.971	0.008
TSS	-0.207	0.004	0.268	0.888	0.023	0.170	0.951	-0.102
BOD	0.082	0.337	0.672	0.530	-0.892	0.368	0.003	-0.127
COD	-0.185	0.315	-0.017	-0.103	-0.206	0.552	-0.010	0.675
AN	0.276	-0.233	0.522	0.556	0.407	-0.693	0.206	-0.438

NO3	-0.112	-0.178	0.927	0.140	-0.055	0.063	-0.181	0.883
PO4	-0.157	0.059	0.786	0.469	-0.614	0.699	0.064	0.318
TC	-0.289	0.127	0.801	-0.091	0.208	0.056	0.106	0.055
Fe	0.924	0.075	-0.334	-0.055	0.649	-0.099	0.567	-0.314
Al	0.564	-0.669	-0.318	0.121	0.753	-0.112	0.442	-0.262
Mn	0.213	-0.915	-0.008	-0.259	0.910	-0.347	0.184	0.007
Cu	0.900	0.220	-0.340	-0.013	0.573	0.033	0.558	-0.306
Cr	0.932	0.284	-0.047	0.162	-0.193	0.895	0.111	0.063
Zn	-0.269	-0.795	0.395	-0.104	0.687	0.257	-0.282	0.211
Cd	0.699	0.669	-0.069	-0.192	-0.311	0.810	0.258	-0.244
Pb	0.682	0.657	-0.138	-0.230	0.901	-0.077	0.114	-0.112
Eigenvalue	7.19	4.65	3.74	1.70	9.73	3.80	2.66	1.46
% of Variance	35.95	23.23	18.72	8.48	48.65	19.01	13.30	7.31
Cumulative %	35.95	59.18	77.90	86.38	48.65	67.67	80.97	88.27

CONCLUSION

In this research, multivariable statistical approaches have been applied to assess the spatial-seasonal differences and identify the pollution sources of Sungai Sembilang in terms of water quality. This river continuously receives effluents from the sanitary landfill, factories, industrial, and residential areas as well as sewage from land activities along the river. Using CA analysis, this study has categorized Sungai Sembilang into three different categories namely; lower, moderately and highly pollution level. The water quality parameters with high loading factors that cause the variations were exactly identified as a percent of the total variation. Thus, the study answers the question, as percent how much each pollution loading is responsible and causes the variation in the rivers' water quality. In the stepwise DA analysis shows 6 parameters (pH, EC, COD, NO_3, TC, and Fe) and correctly assigned about 75% of the total variance between the 3 groups.

While, for the temporal variations analysis, the forward and backward stepwise DA modes selected twelve and five parameters, and with 92% and 89% correct percent in each, respectively. These parameters included EC, TUR, BOD, COD, AN and NO3, and heavy metals such as Fe, Al, and Cu. PCA/FA analysis yielded eight latent factors loadings for each period that significantly identified the pollution sources and types along the river. The results also showed that, based on NWQS, the water quality upstream of the river is acceptable, becomes polluted near the landfill, improves slightly along the remaining length of the river before decreasing again as the river meets the industrial and residential areas. This shows that different sources of pollution contribute to the water quality degradation which can be analyzed using principle component analysis and factor analysis. The results can be used by authorities and decision makers in planning and managing the pollution sources, locating the pollution sources and wastewater treatment, and employing the best management practices to improve river water quality to ensure a healthy river ecosystem and sustainable river management. The study results could help as a feedback for re-shaping the monitoring process by focusing on dominant parameters, reducing the monitoring costs by eliminating the un-necessary sampling location where there were no variations and conducting less water tests (especially in heavy metals tests as they are expensive). In addition, this study can be used as a guide to new studies and water quality management studies such as water management for aquafarming, water security, ecological impact assessment, and sustainable water quality management. Furthermore, the results of this study can also serve as a preliminary guide and knowledge for assessment and evaluation of other tropical rivers having the similar characteristics and perhaps some variation in pollutant sources.

ACKNOWLEDGEMENTS

The authors would like to thank the relevant authorities and companies who have contributed and assisted throughout the study period. This research was funded by University of Malaya Research Grant (RP017C-15SUS, GPF070A-2018). We are most grateful and would like to thank the reviewers for their valuable suggestions, which have led to substantial improvement of the article.

REFERENCES

Abunama, T., Othman, F., Ansari, M. & El-Shafie, A. 2019. Leachate generation rate modeling using artificial intelligence algorithms aided by input optimization method for an MSW landfill. *Environmental Science and Pollution Research* 26(4): 3368-3381.

Abunama, T., Othman, F. & Younes, M.K. 2018. Predicting sanitary landfill leachate generation in humid regions using ANFIS modeling. *Environmental Monitoring and Assessment* 190(10): 597.

Ahmed, M.F., Mokhtar, M., Alam, L., Ta, G.C., Ern, L.K. & Khalid, R.M. 2018. Recognition of local authority for better management of drinking water at the Langat River Basin, Malaysia. *International Journal of Engineering & Technology* 7(3.30): 148-154.

Alssgeer, H.M.A., Gasim, M.B., Hanafiah, M.M., Abdulhadi, E.R.A. & Azid, A. 2018. GIS-based analysis of water quality deterioration in the Nerus River, Kuala Terengganu, Malaysia. *Desalination and Water Treat* 112: 334-343.

APHA. 1988. *Standard Methods for Examination of Water and Waste Water.* 17th ed. Washington DC: American Public Health Association.

Barakat, A., El Baghdadi, M., Rais, J., Aghezzaf, B. & Slassi, M. 2016. Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. *International Soil and Water Conservation Research* 4(4): 284-292.

Bian, Z., Liu, L. & Ding, S. 2019. Correlation between spatial-temporal variation in landscape patterns and surface water quality: A case study in the Yi River Watershed, China. *Applied Sciences* 9(6): 1053.

Cano-Rocabayera, O., de Sostoa, A., Padros, F., Cardenas, L. & Maceda-Veiga, A. 2019. Ecologically relevant biomarkers reveal that chronic effects of nitrate depend on sex and life stage in the invasive fish *Gambusia holbrooki*. *PLoS ONE* 14(1): e0211389.

Chatanga, P., Ntuli, V., Mugomeri, E., Keketsi, T. & Chikowore, N.V. 2019. Situational analysis of physico-chemical, biochemical and microbiological quality of water along Mohokare River, Lesotho. *The Egyptian Journal of Aquatic Research* 45(1): 45-51.

Corsino, S.F., Capodici, M., Di Trapani, D., Torregrossa, M. & Viviani, G. 2020. Assessment of landfill leachate biodegradability and treatability by means of allochthonous and autochthonous biomasses. *New Biotechnology* 55: 91-97.

Cunha, D.G.F., Sabogal-Paz, L.P. & Dodds, W.K. 2016. Land use influence on raw surface water quality and treatment costs for drinking supply in São Paulo State (Brazil). E*cological Engineering* 94: 516-524.

Department of Environmental. 2017. Environmental Quality Report.

Department of Environmental. 2015. Environmental Quality Report.

Dobsa, J., Meznaric, V., Tompic, T., Legen, S. & Zeman, S. 2014. Evaluation of spatial and temporal variation in water contamination along Croation Highways by multivariate exploratory analysis. *Water Air Soil Pollution* 186(10): 6867-6878.

Ebrahimi, M., Gerber, E.L. & Rockaway, T.D. 2017. Temporal performance assessment of wastewater treatment plants by using multivariate statistical analysis. *Journal of Environmental Management* 193: 234-246.

Edokpayi, J., Odiyo, J., Popoola, O. & Msagati, T. 2016. Assessment of trace metals contamination of surface water and sediment: A case study of Mvudi River, South Africa. *Sustainability* 8(2): 135.

Elfithri, R., Toriman, M.E., Mokhtar, M.B. & Juahir, H. 2011. Perspectives and initiatives on integrated river basin management in Malaysia: A review. *The Social Sciences* 6(2): 169-176.

Elias, M.S., Ibrahim, S., Samuding, K., Ab Rahman, S., Wo, Y.M. & Daung, J.A.D. 2018. Multivariate analysis for source identification of pollution in sediment of Linggi River, Malaysia. *Environmental Monitoring and Assessment* 190(4): 257.

Garson, G.D. 2012. *Testing Statistical Assumptions.* Asheboro, NC: Statistical Associates Publishing.

Georgieva, N., Yaneva, Z. & Kostadinova, G. 2013. Analyses and assessment of the spatial and temporal distribution of nitrogen compounds in surface waters. *Water and Environment Journal* 27(2): 187-196.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. & Tatham, R.L. 1998. *Multivariate Data Analysis*. Volume 5. Upper Saddle River, NJ: Prentice Hall. pp. 207-219.

Hajigholizadeh, M. & Melesse, A.M. 2017. Assortment and spatiotemporal analysis of surface water quality using cluster and discriminant analyses. *Catena* 151: 247-258.

Horn, A.H., Torres, I.C., Ribeiro, E.V. & Junior, A.P.M. 2017. Relationship between metal water concentration and anthropogenic pressures in a Tropical Watershed, Brazil. *Geochimica Brasiliensis* 30(2): 158.

Idris, A.B., Mamun, A.A., Amin, M., Soom, M., Noor, W. & Azmin, W. 2003. Review of water quality standards and practices in Malaysia. *Pollution Research* 22(2): 145-155.

Jaishankar, M., Tseten, T., Anbalagan, N., Mathew, B.B. & Beeregowda, K.N. 2014. Toxicity, mechanism and health effects of some heavy metals. *Interdisciplinary Toxicology* 7(2): 60-72.

Le, T.T.H., Zeunert, S., Lorenz, M. & Meon, G. 2017. Multivariate statistical assessment of a polluted river under nitrification inhibition in the tropics. *Environmental Science and Pollution Research* 24(15): 13845-13862.

Liu, D. & Zou, Z. 2012. Water quality evaluation based on improved fuzzy matter-element method. *Journal of Environmental Sciences* 24(7): 1210-1216.

Matta, G. 2015. Evaluation and prediction of deviation in physic-chemical characteristics of River Ganga. *International Journal of Advancements in Research and Technology* 4(6): 14-30.

Mavukkandy, M.O., Karmakar, S. & Harikumar, P. S. 2014. Assessment and rationalization of water quality monitoring network: A multivariate statistical approach to the Kabbini River (India). *Environmental Science and Pollution Research* 21(17): 10045-10066.

McKinley, K., McLellan, I., Gagné, F. & Quinn, B. 2019. The toxicity of potentially toxic elements (Cu, Fe, Mn, Zn and Ni) to the cnidarian *Hydra attenuata* at environmentally relevant concentrations. *Science of The Total Environment* 665: 848-854.

Mousa, I.E., Emara, I.K., Farfour, S.A. & Eldourghamy, A.S. 2018. Microbial profile and its changing rates of Lake Burullus, Egypt as wastewater receiving body. *Water and Environment Journal* 32(1): 67-74.

Ogwueleka, T.C. 2014. Assessment of the water quality and identification of pollution sources of Kaduna River in Niger State (Nigeria) using exploratory data analysis. *Water and Environment Journal* 28(1): 31-37.

Ojok, W., Wasswa, J. & Ntambi, E. 2017. Assessment of seasonal variation in water quality in River Rwizi using multivariate statistical techniques, Mbarara Municipality. Uganda. *Journal Water Resource Protection* 9(1): 83-97.

Othman, F., Alaa Eldin, M.E. & Mohamed, I. 2012. Trend analysis of a tropical urban river water quality in Malaysia. *Journal of Environmental Monitoring* 14: 3164.

Psaltopoulos, D., Wade, A.J., Skuras, D., Kernan, M., Tyllianakis, E. & Erlandsson, M. 2017. False positive and false negative errors in the design and implementation of agri-environmental policies: A case study on water quality and agricultural nutrients. *Science of The Total Environment* 575: 1087-1099.

Sarkar, S.K., Saha, M., Takada, H., Bhattacharya, A., Mishra, P. & Bhattacharya, B. 2007. Water quality management in the lower stratch of the River Ganges, East Coast of India: An approach through environmental education. *Journal of Cleaner Production* 15: 1559-1567.

Stevenson, R.J. & Rollins, S.L. 2017. Ecological assessment with benthic algae. In *Methods in Stream Ecology.* New York: Academic Press. pp. 277-292.

Stefania, B., Calabrò, P.S., Rosa, G. & Moraci, N. 2018. Selective removal of heavy metals from landfill leachate by reactive granular filters. *Science of The Total Environment* 644: 335-341.

USEPA. 2001. *Parameters of Water Quality: Interpretation and Standards. Environmental Protection Agency*. Johnstown Castle, Co. Wexford, Ireland.

Van Ael, E., Belpaire, C., Breine, J., Geeraerts, C., Van Thuyne, G., Eulaers, I. & Bervoets, L. 2014. Are persistent organic pollutants and metals in eel muscle predictive for the ecological water quality? *Environmental Pollution* 186: 165-171.

VishnuRadhan, R., Zainudin, Z., Sreekanth, G.B., Dhiman, R., Salleh, M.N. & Vethamony, P. 2017. Temporal water quality response in an urban river: A case study in Peninsular Malaysia. *Applied Water Science* 7(2): 923-933.

Voyles, J., Vredenburg, V.T., Tunstall, T.S., Parker, J.M., Briggs, C.J. & Rosenblum, E.B. 2012. Pathophysiology in mountain yellow-legged frogs *(Rana muscosa)* during a chytridiomycosis outbreak. *PLoS ONE* 7(4): 353-374.

Wang, Z., Meador, J.P. & Leung, K.M. 2016. Metal toxicity to freshwater organisms as a function of pH: A meta-analysis. *Chemosphere* 144: 1544-1552.

Wu, Z., Wang, X., Chen, Y., Cai, Y. & Deng, J. 2018. Assessing river water quality using water quality index in Lake Taihu Basin, China. *Science of The Total Environment* 612: 914-922.

Yan, J., Xu, Z., Yu, Y., Xu, H. & Gao, K. 2019. Application of a hybrid optimized BP network model to estimate water quality parameters of Beihai Lake in Beijing. *Applied Sciences* 9(9): 1863.

Zhang, L., Zou, Z. & Shan, W. 2017. Development of a method for comprehensive water quality forecasting and its application in Miyun reservoir of Beijing, China. *Journal of Environmental Sciences* 56: 240-246.

Zhang, Y., Xu, M., Li, X., Qi, J., Zhang, Q., Guo, J., Yu, L. & Zhao, R. 2018. Hydrochemical characteristics and multivariate statistical analysis of natural water system: A case study in Kangding County, Southwestern China. *Water* 10(1): 80.

Tengku Nilam Baizura Tengku Ibrahim & Noor Zalina Mahmood
Department of Environmental Management
Institute of Biological Sciences
Faculty of Science
University of Malaya
50603 Kuala Lumpur, Federal Territory
Malaysia

Tengku Nilam Baizura Tengku Ibrahim
Department of Environmental Health
Faculty of Health Sciences
MAHSA University
Jln SP 2, Bandar Saujana Putra
42610 Jenjarom, Selangor Darul Ehsan
Malaysia

Faridah Othman* & Taher Abunama
Department of Civil Engineering
Faculty of Engineering
University of Malaya
50603 Kuala Lumpur, Federal Territory
Malaysia

Taher Abunama
Institute of Water and Wastewater Technology
Durban University of Technology
Musgrave, Berea
South Africa

*Corresponding author; email: faridahothman@um.edu.my

Received: 30 December 2019
Accepted: 8 August 2020