Pertanika J. Sci. & Technol. 31 (1): 495 - 510 (2023)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Water Quality Assessment and Characterization of Rivers in Pasir Gudang, Johor via Multivariate Statistical Techniques

Muhammad Syafiq Mohamad Desa[1], Mohd Aeddy Sulaiman[2] and Shantakumari Rajan[1]*

[1]Centre of Environmental Health & Safety, Faculty of Health Sciences, Universiti Teknologi MARA, UiTM Cawangan Selangor, 42300 Puncak Alam, Selangor, Malaysia
[2]Environmental and Safety Unit, Department of Public Health and Environmental, Level 15 Menara Aqabah, Majlis Bandaraya Pasir Gudang, Jalan Bandar, 81700 Pasir Gudang, Johor, Malaysia

ABSTRACT

In Pasir Gudang, an accelerated industry-based economy has caused a tremendous increase and diversity of water contamination. The application of multivariate statistical techniques can identify factors that influence water systems and is a valuable tool for managing water resources. Therefore, this study presents spatial evaluation and the elucidation of inordinate complex data for 32 parameters from 25 sampling points spanning 20 rivers across Pasir Gudang, summing up to 1500 observations between 2015-2019. Hierarchical cluster analysis with the K-means method grouped the rivers into two main clusters, i.e., proportionately low polluted rivers for Cluster 1 (C1) and high polluted rivers for Cluster 2 (C2), based on the similitude of water quality profiles. The discriminant analysis applied to the cluster resulted in a data reduction from 32 to 7 parameters (Cl, Cd, S, OG, temperature, BOD, and pH) with a 99.5% correct categorization in spatial analysis. Hence, element complexity was reduced to a few criteria accountable for large water quality differences between C1 and C2. The principal component analysis produced 6 and 7 principal components after rotation for C1 and C2, respectively, where total variance was 62.48% and 66.85%. In addition, several sub-clusters were identified; two from C1 and three from C2, based on the principal contributing components. These results show that the functionality of multivariate techniques can be effectively used to identify spatial water characteristics and pollution sources. The outcomes of this study may benefit legislators in managing rivers within Pasir Gudang.

Keywords: Cluster analysis, discriminant analysis, water quality assessment

ARTICLE INFO

Article history:
Received: 15 March 2022
Accepted: 21 June 2022
Published: 09 November 2022

DOI: https://doi.org/10.47836/pjst.31.1.29

E-mail addresses:
muhdsyafiqdesa@gmail.com (Muhammad Syafiq Mohamad Desa)
aeddy@mbpg.gov.my (Mohd Aeddy Sulaiman)
shanta@uitm.edu.my (Shantakumari Rajan)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Developing countries worldwide are facing alarming environmental issues; one of the significant issues is the decline of surface water quality due to pollution. Rivers are amongst the most vulnerable surface water bodies because of their capacity as reservoirs for all surface contaminants. Moreover, due to rapid development, the surface water quality of the river has deteriorated over time owing to its role in carrying off discharge from municipal and industrial wastewater with additional run-off from the agricultural sector. The main functions of rivers are common as a supply for drinking water, irrigation, and various other industrial purposes in addition to recreation.

Since the early 1980s, Malaysia has rapidly transformed from an agriculture-based economy to an industrial-based economy. Industrialization has provided many benefits in terms of economic growth in Malaysia, but these have occurred at the expense of the environment (Samsudin et al., 2017), as a great number of anthropogenic activities end up in water bodies. Overall, river water pollution in Malaysia generally is surging as the percentage of clean rivers has decreased to 47% in 2016 compared to 58% in the previous year, and polluted rivers increased from 7% to 10% in 2016 (Department of Environment, 2016). Ultimately, a recent case happened in March 2019 that had a high impact on Malaysian citizens; the case of pollution in the Kim Kim River involved untreated factory waste being illegally and directly discharged into the river body (Academy of Sciences Malaysia, 2019). As a result, several toxic gases, such as methane xylene, acrylonitrile and toluene, were emitted following the interaction of the chemicals concerned with water and air, causing this crisis to be unique as the water pollution subsequently caused air pollution (Chai, 2020). As a result, pollutants were removed from the 1.5 km stretch of Sungai Kim Kim, where chemicals were expected to cost approximately RM 6.5 million (The Straits Times, 2019).

Alkarkhi et al. (2008) stated that industrialization is the main factor in ecosystems' pollution. Unsustainable factories are the real culprit as pollutants are discharged into rivers and lakes with minimal treatment. Water demand in Malaysia for the year 2020 was approximately 53% for domestic and industrial usage, while the remaining 47% for agricultural activities. It is predicted that by 2050, water demand in Malaysia will increase by 103% for domestic, industrial, and agriculture (Ishak & Ismail, 2020). Malaysia may clearly face a huge problem in water supply if there is a lack of comprehension of the water data assessment and practical decisions in water quality management. Improving water resource capacity and maintaining excellent standards can specifically mitigate health and water supply issues. However, a water quality issue in the country is related to catchment protection, pollution and flooding control management, and legislation weaknesses. Generally, a holistic cognition of water quality monitoring and analysis is needed to curb the current situation from deteriorating further.

Water quality monitoring programs are conducted by several government agencies such as the Department of Environment, Ministry of Health and Local Authorities. These observing projects regularly appraise and estimate physicochemical boundaries with limit values suggested by national or worldwide bodies (Bhuiyan et al., 2011). These programs will eventually produce large complex data sets over time, requiring efficient analysis to be fully understood. Therefore, applying appropriate statistical methodology when analyzing water quality data is essential to draw valid conclusions and provide useful advice in water management (Uddin et al., 2021). Nevertheless, data analysis using multivariate techniques has been proven to be one of the legitimate methods. Multivariate techniques can analyze complex water data sets in a straightforward way for better interpretation. Many elements in multivariate statistical methods such as Cluster Analysis (CA), Principal Component Analysis (PCA), and Discriminant (DA) help understanding the complex data sets, such as those created by long-term water quality monitoring programs, allowing a better view of the temporal and spatial variations in water quality (Shrestha & Kazama, 2007). Studies have shown that multivariate statistical analysis is useful for assessing spatial water quality variations in a river. For example, a multivariate analysis of the Kinta River's water quality parameters in Perak revealed rock weathering along the riverbanks and untreated wastewater as the main pollution source (Isiyaka & Juahir, 2015). This analysis output provided valuable information for decision-making in river management.

In this study, collaboration with the Pasir Gudang City Council (PGCC) was developed to understand better water surface analysis across rivers in areas under the PGCC administration. These rivers are being monitored as Pasir Gudang is a hub for many industry types such as chemical-based (oleochemical, plastic production, petrochemicals), heavy engineering, logistics and small and medium enterprises. Therefore, multivariate statistical methods can classify the rivers into clusters based on the type and concentration of pollutants present. Identifying clustering patterns is based on identifying similarities between the rivers and would enable enhanced management of this region. Therefore, this study focused on applying multivariate techniques such as CA, PCA, and DA to determine the major pollutants affecting the water quality of the rivers in this locality and to identify the presence of river clusters according to their water quality characteristics.

MATERIALS AND METHODS

Study Area

Pasir Gudang is a significant industrial and port city located in the southwestern part of Johor Bahru district, Johor, Malaysia (Abdullah et al., 2012). There are 25 sampling points scattered across the Pasir Gudang region (Table 1). Most of the sampling points focus on downstream areas covering 20 different rivers. Some of these rivers are the main water sources for the Sultan Iskandar reservoir, which is used for residential and industrial water supply and irrigation.

Water Quality Data

The archive data set obtained from PGCC covers water quality data from 2015-2019, where water samples were collected monthly by PGCC personnel from each sampling point. All sampling bottles were acid-washed, cleaned, and dried before use. The samples were placed in an ice box and transported to the laboratory for further analysis. A certified laboratory carried out all analyses except temperature, measured in situ. The data set comprises 32 parameters summarised in Table 2 and the analytical methods used (APHA, 2016).

Data Analysis

The data sets were analyzed using Statistical Package for the Social Sciences (SPSS) version 26. Cluster analysis (CA) was used to classify rivers into clusters based on their similarities (exposure level pattern). It will give optimal grouping and combine hierarchical and K-means clustering. The hierarchical method applied Ward's method

Table 1
Sampling points and rivers

Sampling point	River
P1	Masai (A)
P2	Masai (B)
P3	Masai (C)
P4	Kim Kim
P5	Kopok
P6	Kong Kong
P7	Tiga
P8	Perapat
P9	Cupak
P10	Hujung
P11	Tengah
P12	Laloh
P13	Tukang Batu (A)
P14	Tukang Batu (B)
P15	Perembi
P16	Buluh
P17	Selangkah
P18	Kopok Baru
P19	Serai
P20	Redan
P21	Air Puteh
P22	Tiram (A)
P23	Tiram (B)
P24	Tiram (C)
P25	Penderam

Table 2
Water quality parameters and analytical methods

No.	Parameter	Analytical Method
1	Silver	APHA 3111 B
2	Aluminum	APHA 3111 D
3	Ammoniacal-Nitrogen (AN)	APHA 4500 – NH_3 C/ NH_3 B
4	Arsenic	APHA 3114 B
5	Boron	APHA 4500 – B B
6	Barium	APHA 3111 D
7	Biochemical Oxygen Demand (BOD)	APHA 5210 B/ APHA 4500- OG
8	Cadmium	APHA 3111 B
9	Cyanide	APHA 4500 – CN C&E and Merek Method 14429
10	Chemical Oxygen Demand (COD)	APHA 5220 C
11	Chromium (III)	APHA 3500 - Cr B
12	Chromium (IV)	APHA 3500 – Cr B

Table 2 *(continue)*

No.	Parameter	Analytical Method
13	Copper	APHA 3111 B
14	Fluoride	APHA 4500 - F D
15	Iron	APHA 3111 B
16	Formaldehyde	Macherey Nagel Nano color Formaldehyde
17	Mercury	APHA 3112 B
18	Manganese	APHA 3111 B
19	Nickel	APHA 3111 B
20	Oil & Grease (OG)	APHA 5520 B
21	Lead	APHA 3111 B
22	pH	APHA 4500 – H^{+}B
23	Phenol	APHA 5530 B & APHA 5530 D
24	Selenium	APHA 3114 B
25	Tin	APHA 3114 B
26	Total Suspended Solids (TSS)	APHA 2540 D
27	Zinc	APHA 3111 B
28	Chloride	APHA 4500 - CI B
29	Chlorine	APHA 4500 – CI B
30	Sulfide	APHA 4500 -S^{2-}F
31	Temperature	In-situ
32	Color	APHA 2120 F

with squared Euclidean distances as a similarity measure. The data sets were treated by -1 to 1 scale transformation. At the same time, K-means can portray each cluster's mean exposure toward water parameters predetermined by the hierarchical method. Combining hierarchical and K-means results are used to identify river clusters and interpreted into a useful grouping. Parameters that did not have a stored data value for the observation and were recorded as not detected were set at zero during the data cleaning stage before the analysis.

Discriminant analysis (DA) was used to calculate the mathematical weights for each score of variables (water quality parameters) and reflects it to scores on each variable (32 parameters), where it can differentiate clusters obtained from CA. This analysis can produce a discriminant function (DF) depending on the number of groups/clusters from CA (Singh et al., 2005), whereby the DF is made of parameters that can distinguish between clusters in this study. The DA was applied using a stepwise method where all variables are reviewed and evaluated to determine which one contributes most to the discrimination between groups.

Principal component analysis (PCA) is a dimensionality reduction method that uses fewer variables to explain most of the variation in the original data and converts many highly correlated variables into independent or unrelated variables (Li et al., 2017). Principal components (PC) provide information on the most meaningful parameters. Hence, the PCs can unveil more hidden vital parameters related to possible pollution sources based on spatial aspects. Only parameters with loading factors of more than 0.5 meet the requirements to clarify each cluster's water quality variation pattern. Principally, varimax rotations were applied to PCs with eigenvalues of more than one. Factors with eigenvalues > 1 explained total variation in the data than individual water quality variables. PCA was carried out separately for each cluster to describe the covariance relationship among variables for each cluster individually.

RESULTS AND DISCUSSION

Site Similarity and Clustering

Phenol, cyanide, chlorine and selenium concentration for Masai (A), Masai (B), Masai (C), Kim Kim, Kopok, Kong Kong, Tiga, Prapat, Cupak, Hujung, Tengah, Laloh, Tukang Batu (A), Tukang Batu (B), Perembi, Buluh, Selangkah, Kopok Baru, Serai, Redan, Air Puteh, Tiram (A), Tiram (B), Tiram (C), and Penderam Rivers were below detection limits in all samples between the years 2015-2019. Most water quality parameters are classified in groups I to III according to the National Water Quality Standards for Hg, As, Sn, Ag, Cl_2, Phenol, Cd, Pb, Zn, Al, S, CN, Cr^3, Cr^6, Ni, Fe, and Ba. Class I to III are considered fit for being biologically viable and have economic value. Various physico-chemical parameters, such as F, Cu, BOD, and AN, are in class IV. At the same time, only COD falls under class V. Water within class IV, and V is limited to irrigation use only (class IV) or has no economic value or other benefits (class V). The mean values of water quality parameters was arranged in the order, Cl > COD > TSS > BOD > pH > AN > F > Fe > B > Zn > OG > Mn > S > Cu > Pb > As > Ni > Sn > Cr^{6+} > Al > Cd > Formaldehyde > Cr^{3+} > Ag > Ba > Hg.

The spatial variation of water quality mainly depends on the different water quality parameters. The clustering pattern showed that the data set could be divided into two common groups where the dendrogram derived yielded two main clusters (Figure 1). The confirmed clusters subjected to K-means analysis yielded Cluster 2 as a considerably bigger cluster than Cluster 1. There was sufficient internal homogeneity within the clusters and external heterogeneity between the clusters to show 2 main clusters as the stations in these groups have similar and natural backgrounds of the water quality characteristics (McKenna, 2003). The analysis outcome grouped 11 rivers into cluster 1, consisting of Kim Kim, Kopok, Perapat, Kong Kong, Cupak, Hujung, Tengah, Laloh, Air Puteh, and Penderam River, while the remaining 14 rivers comprised cluster 2 (Tiram A, B, C, Redan, Masai A, B, C, Tukang Batu A, B, Perembi, Buluh, Selangkah, Kopok Baru, and Serai River).

The differences in water quality characteristics between clusters 1 and 2 show a clear and distinct line for most parameters such as BOD, COD, TSS, AN, Cr^{6+}, As, Mn, Ni, Sn, Zn, B, Fe, F, OG, and Color (Table 3). These parameters are highly affiliated with anthropogenic discharges related to industrialization, commercialization and residential activities (Khatri & Tyagi, 2015), becoming important factors influencing surface water quality (Hussain et al., 2008). According to Phiri et al. (2005), indiscriminate disposal of municipal solid waste in river systems cause an increase in BOD, COD, TDS, TSS, and toxic metals such as Cd, Cr, Ni and Pb. Degraded streams and rivers that drain urbanized landscapes often have higher nutrient loads and contaminant concentrations, as well as altered stream morphology and reduced biodiversity (Meyer et al., 2005). In contrast, cluster 1 is highly correlated with water quality parameters such as B, Cl^-, and S suggesting a strong correlation with natural factors such as weathering of the parent rock. Cl^-, S, and B can be classified as conductivity ions commonly present when surface runoff is in solid form (Isiyaka & Juahir, 2015). The cluster mapping in Figure 2 clearly illustrates that cluster 1

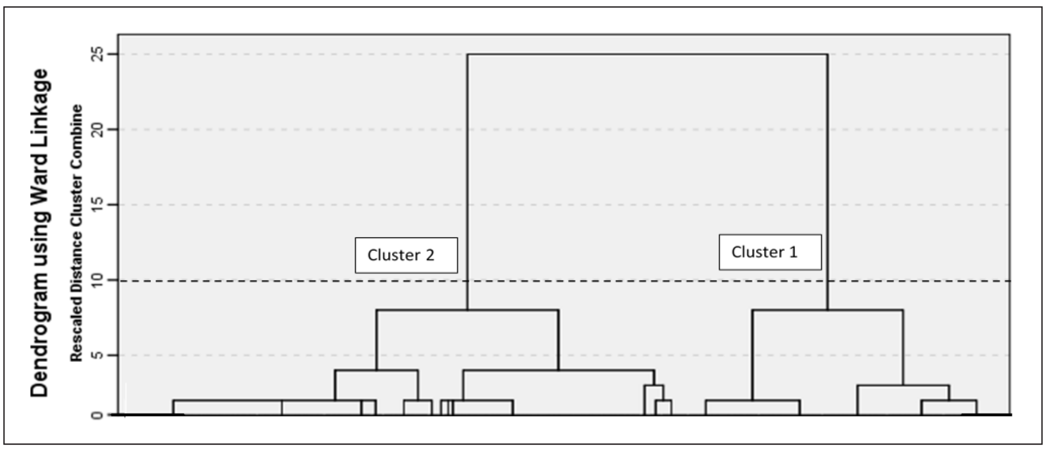

Figure 1. Dendrogram of river clusters

Table 3
Final cluster centers of water quality parameters

	Cluster 1	Cluster 2
pH	7.53	6.91
Temperature (°C)	28.32	28.54
BOD (mg/L)	5.77	10.60
COD (mg/L)	103.46	139.68
TSS (mg/L)	17.63	25.50
AN (mg/L)	0.65	3.09
Hg (mg/L)	ND	ND
Cd (mg/L)	ND	ND
Cr^{3+} (mg/L)	ND	ND
Cr^{6+} (mg/L)	ND	0.01
As (mg/L)	ND	0.03

Table 3 *(continue)*

	Cluster 1	Cluster 2
Pb (mg/L)	0.03	0.03
Cu (mg/L)	0.04	0.04
Mn (mg/L)	0.05	0.08
Ni (mg/L)	0.01	0.03
Sn (mg/L)	ND	0.01
Zn (mg/L)	0.16	0.24
B (mg/L)	0.32	0.26
Fe (mg/L)	0.40	0.48
Ag (mg/L)	ND	ND
Al (mg/L)	ND	ND
Se (mg/L)	ND	ND
Ba (mg/L)	ND	ND
F (mg/L)	0.80	0.87
Cl (mg/L)	15871	1416
Cl_2 (mg/L)	ND	ND
S (mg/L)	0.08	0.03
OG (mg/L)	0.10	0.15
Formaldehyde (mg/L)	ND	ND
Phenol (mg/L)	ND	ND
CN (mg/L)	ND	ND
Color (ADM)	16.5	29.1

Number of cases in each cluster

Cluster	1	560.0
	2	940.0
Valid		1500.0
Missing		0.0

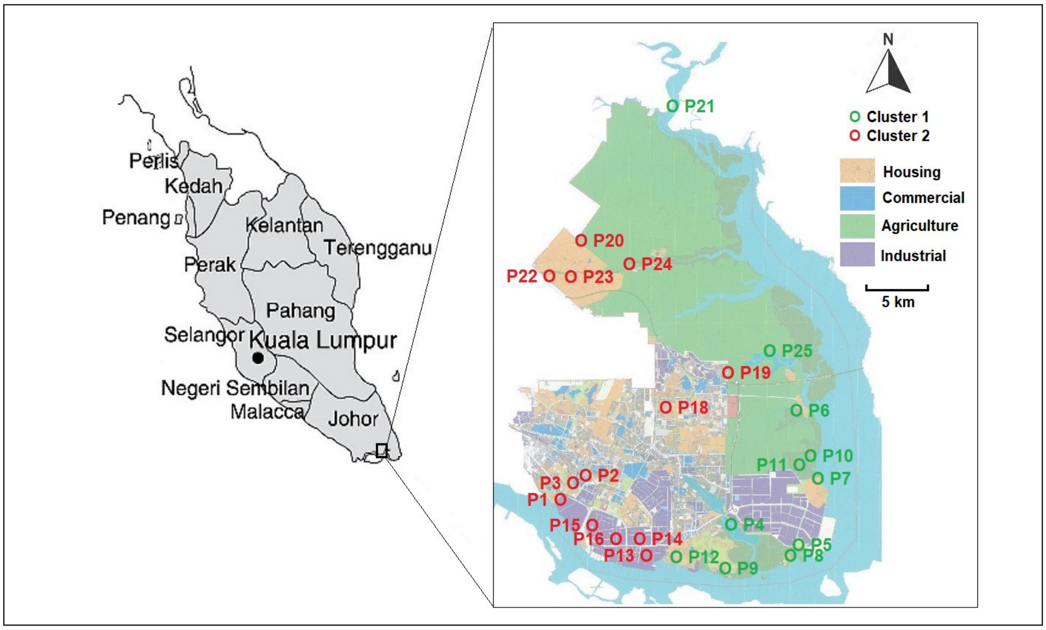

Figure 2. The geographic locations of river clusters overlay the area's general land use area
Source. GeoJohor

(green) focuses more on rivers associated with low agricultural activities and undeveloped land. In contrast, cluster 2 (red) is highly connected with pollution loading from urban human activities. Studies have shown that changes in landscape patterns induced by human activities have major impacts on river conditions (Allan, 2004; Bhat et al., 2006; Hopkins, 2009). Cluster analysis has provided a practical grouping system for the rivers under PGCC surveillance, which can be used to construct gilt-edged spatial observation at a lower cost.

Identification of Discriminating Parameters and Pollution Sources

Based on clusters generated from the CA, water quality parameters were analyzed in the stepwise method using discriminate analysis. The DA identified 7 parameters that can discriminate against the naturally occurring cluster. Since there are two clusters, only one Differentiate Function (DF) appeared in DA. Wilk's Lambda test showed that DF is statistically significant, p *<0.05*. The same value, also parallel with Wilk's Lambda test, is the canonical value at 0.925 near the value of 1, indicating it is very strong, indicating that the clustering result of CA was reliable, proving that the water quality had significant spatial variation. The DF produced can be labeled as a natural factor that can measure score using Equation 1. The main contributing parameters to the equation are Cl^-, Cd, OG, BOD, Temperature, S and pH, which suggests that these parameters are important for differentiation among the clusters. The equation can determine the cluster membership of new cases using cluster centroids. In this study, the centroids of cluster 1 and cluster 2 were 3.148 and -1.875, respectively. Cl^- has the highest contribution value in differentiating the two river clusters with a value of 1.023, as there was a much larger divergence between average Cl^- values between clusters. In comparison, other parameters contributed less to explaining the variation between the two clusters. The relative contribution of all water quality parameters in differentiating the two clusters can be arranged in the order, Cl^- > Cd > OG > BOD > Temperature > S > pH.

$$DF = 1.023(Cl) + 0.132(Cd) + 0.096(OG) - 0.085(BOD) - 0.083(Temperature) + 0.075(S) - 0.066(pH) \quad (1)$$

The classification matrix showed that 99.5% of the cases were correctly classified (Table 4). The result from DA showed significant differences between the two clusters, which are expressed as one discriminant function.

Principle component analysis was applied to the water quality data for both clusters to identify the spatial sources of pollution and constitution patterns. In order to pin down the source of pollution, only loading factors of >0.5 qualify to clarify the variation of the water quality pattern of each cluster; hence only 17 water quality parameters were measured out of 32 in cluster 1. The data scored 0.703 in the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy test, which was considered middling (Kim & Muller, 1987). Six Principal Components (PCs) were obtained with eigenvalues larger than 1, and the total percentage variance was explained as 62.48%. While for cluster 2, there were 18 parameters used in PCA with a KMO test value of 0.692, which was considered mediocre (Kim & Muller, 1987). The total PCs obtained were seven, with a total percentage of variance explained at 66.85%. The water quality parameters excluded in analysis with values <0.5 were Mn, Fe, Ag, Sn and Hg for cluster 1 and S, Cd, Mn, Sn, Formaldehyde, Ag, Al, and Cu for cluster 2, respectively.

6 PCs explain the variability of river water quality from 2015-2019 for cluster 1 (Table 5). Accumulation of 6 factors explains about 62.48% of total variance after rotation. These 6 factors can unfold the pattern in the characteristics of the examined parameters and point out major possible sources of pollution. In cluster 1, PC1, with 17.84% of the total variance, has strong positive loadings on Pb, B, and F but moderate loadings for Cd. Pb and Cd are normal components in industry effluent (Rosli et al., 2015), while B and F can be generated from urban runoff from aquaculture and

Table 4
Percentage of correctly classified cluster variables

		Predicted Group Membership		
	Cluster K-means	1	2	Total
Count	1	556	4	560
	2	4	936	940
%	1	99.3	0.7	100.0
	2	0.4	99.6	100.0

a. 99.5% of original grouped cases correctly classified

Table 5
Principle components loading of Cluster 1

	PC1	PC2	PC3	PC4	PC5	PC6
Pb (mg/L)	0.882					
F (mg/L)	0.831					
B (mg/L)	0.792					
Cd (mg/L)	0.704					
COD (mg/L)		0.879				
BOD (mg/L)		0.836				
TSS (mg/L)		0.652				
OG (mg/L)			0.768			
Color (ADM)			0.677			
Temperature (^oC)			0.518			
Cu (mg/L)				0.718		
pH				-0.559		
AN (mg/L)					0.857	
Zn (mg/L)					0.591	
S (mg/L)						0.763
Ni (mg/L)						0.676

residential activities. The PC2 with 12.75 % variance shows strong COD, BOD, and TSS loading. These parameters represent anthropogenic input in most cases, where industrial effluent or residential discharge can contribute to elevated levels of either or both BOD and COD. The positive loading on TSS can be from soil erosion or earthworks. It can happen when there is significant interference from human activities toward rivers, such as water dredging, diversion and channelization (Hua et al., 2016).

In PC3, which contributes 10.007% of the variance, the components have strong positive loading for OG with medium loading for color and temperature. Oil contaminated wastewater comes from various sources such as the petrochemical industry, oil refinery, crude oil production, compressor condensates, metal processing, lubricant and cooling agents, vegetable oil industries, food processing industries, car washing and restaurants (Lan et al., 2009). The largest source of oily wastewater produced is during the oil extraction processes in most oil mills and the mill effluents such as palm oil mill effluent. Additionally, residential sources of OG are from kitchen greywater, which has been reported as the highest contributor to oil and grease in domestic greywater (Alade et al., 2011). PC4 has a positive loading of Cu and negative loading of pH, showing an inverse proportional relationship contributing 7.525% of the variance. The concentration of Cu is below the standard level set within the National Water Quality Standards of Malaysia, with a mean of 0.0255 mg/L. The low values of Cu indicate that there is no significant source of pollution, and the presence of this element is due to the weathering of rocks (Hussain et al., 2017) and soil (Kumar et al., 2013), where leaching has contributed to Cu presence in the water.

PC5, with a variance of 7.317%, has a strong loading of AN and Zn. Municipal, industrial, and agricultural activities generate AN discharge into environmental resources. Therefore, these parameters are attributed to products from anthropogenic activities with urban impacts. Meanwhile, PC6 has the lowest percent of variance with 7.044%, strong positive loading of S and moderate loading of Ni. The mean concentration of Ni (0.0063 mg/L) is below the standard of the NWQS, whereas sulfur concentration has a value of 0.0764 mg/L and is considered above class II and III. The source of nickel is possibly due to weathering of rocks, soil erosion, and vegetation as natural occurrences like metamorphic rocks with high Ni can add up the element in the water profile (Al-Badaii et al., 2016). In contrast, S content is highly likely derived from surface runoff from agricultural activities and soil erosion (Bhuiyan et al., 2011). Overall, for cluster 1, it can be concluded that two sub-clusters exist; sub-cluster A (P4, P5, P8, P9 & P12) associated with anthropogenic activities and sub-cluster B (P6, P7, P19, P11, P21 & P25) associated with nature factors (Figure 3).

7 PCs explain the variability of river water quality for cluster 2, whereby the accumulation of 7 factors explains 66.85% of total variance after rotation (Table 6). PC1 specified 16.38% variance after rotation with strong positive loading of As, Cr^{6+}, and Pb. These parameters are derived mainly from industrial activities such as industrial processes,

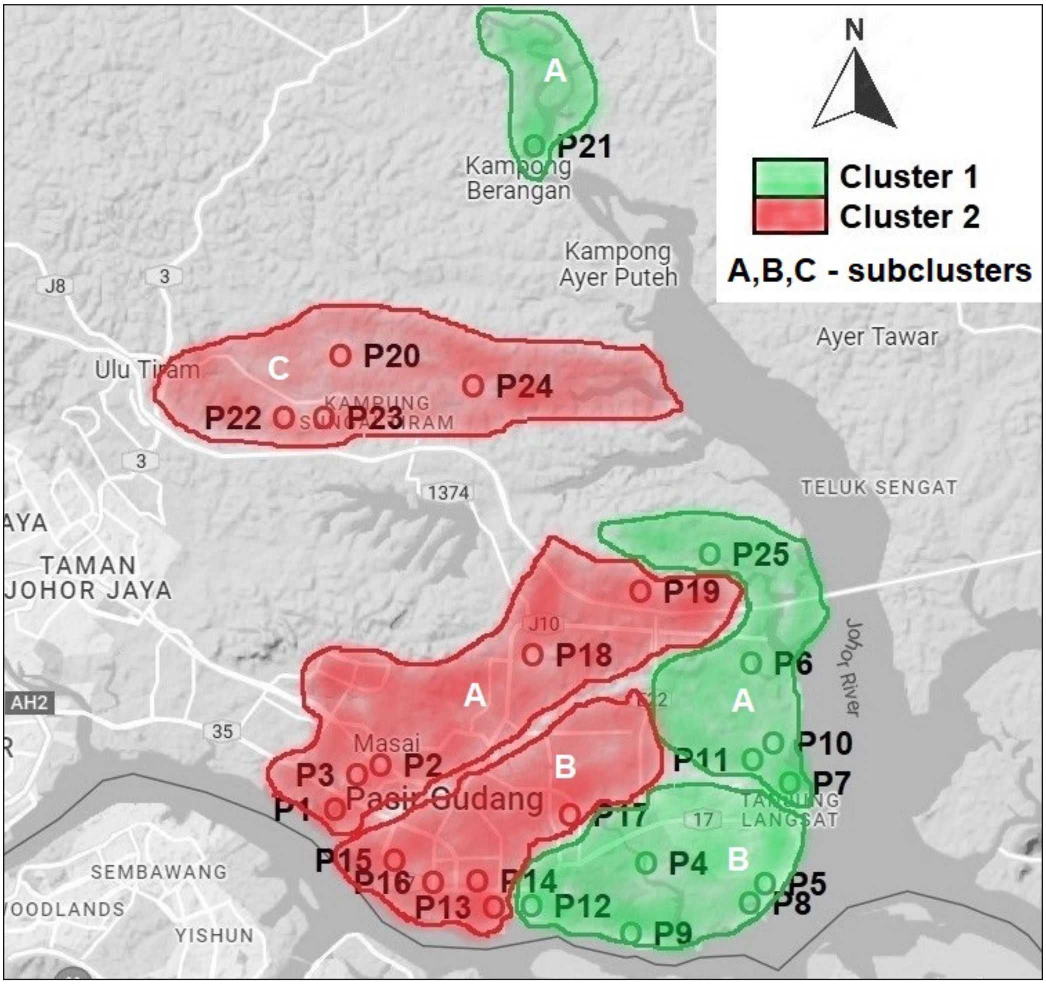

Figure 3. Sub-clustering of rivers within the main cluster pattern

including mining, burning of fossil fuels, metal waste recycling activities, and cement manufacturing (Zhang et al., 2010). For PC2, the percent of the variance is 12.594, with a positive loading for B, COD, BOD, and temperature. The relationship between B, COD and BOD are directly proportional; however, they are inversely proportional to temperature. Boron is widely used in metal processing for containers of high-temperature reactions and electrodes. The loading of COD and BOD can signify the combination of effluent from industries, sewage, commercial area, and small and medium enterprises (Halder & Islam, 2015).

In PC3, valued at 10.834% of the variance, the composition showed two strong positive loadings for OG and color. The sources for these parameters are the same as for PC3 for cluster 1. PC4 scored 8.130% of the variance with strong positive loading for Cl^- and moderate loading for pH. Additionally, chloride, which is only commonly high during rainy seasons due to weathering, may become abnormally high due to sand mining activities along rivers. Meanwhile, for PC5, it scored a percentage of variance at 6.626 alongside strong positive loading for Ni and moderate loading for Fe. Commonly both elements can be sourced from natural and anthropogenic factors; however, in this study, the influence is more due to anthropogenic sources based on concentration levels.

PC6 contributed 6.122 % of the variance and consisted of strong positive loading for Cr^{3+} and moderate loading for F only. Natural sources of chromium are rare and are mainly associated with industrial effluent derived from the production of corrosion inhibitors and pigments (Oliveira, 2012). As for F, this element is commonly present in water, and in this study, the mean concentration is below NWQS only at 0.87 mg/L. Finally, the last PC7 has a percent variance of 5.899 scores and is strongly associated with Zn, whereby sources are related to mining, industries, and residential land use. Therefore, cluster 2 can be divided into three main sub-clusters that are all contributed by anthropogenic activities; sub-cluster A (P1, P2, P3, P18 & P19), sub-cluster B (P13, P14, P15, P16 & P17), and sub-cluster C (P20, P22, P23 & P24) (Figure 3).

Table 6
Principle components loading of Cluster 2

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
As (mg/L)	0.994						
Cr^{6+} (mg/L)	0.994						
Pb (mg/L)	0.987						
B (mg/L)		0.726					
COD (mg/L)		0.711					
BOD (mg/L)		0.705	0.515				
Temperature (°C)		-0.593					
OG (mg/L)			0.768				
Color (ADM)			0.766				
Cl^- (mg/L)				0.774			
pH				0.716			
Ni (mg/L)					0.796		
Fe (mg/L)					0.508		
Cr^{3+} (mg/L)						0.774	
F (mg/L)						0.638	
Zn (mg/L)							0.928

CONCLUSION

Multivariate techniques (cluster analysis, discriminant analysis, and principal component analysis) were successfully applied to appraise the spatial variation of water quality across rivers under PGCC surveillance. CA grouped the sampling stations into two main clusters based on similar water characteristics, representing low and high anthropogenic activities. DA revealed significant data reduction, giving seven parameters (Cl, Cd, S, OG, temperature, BOD, and pH) with 99.5 % correct assignation. The PCA extracted principal components for cluster 1, in which the major sources of pollution are a mix of anthropogenic and natural factors, while cluster 2 was affected by anthropogenic activities only. Therefore, multivariate techniques are important in environmental management for authorities and decision-makers to develop an optimal strategy for reducing sampling stations. Based on these findings, authorities can design optimal sampling stations and reduce experimental analysis costs. Additionally, these methods are important to avoid misinterpreting environmental monitoring data contamination due to uncertainties. A comprehensive approach to the land use pattern in tandem with onsite assessment is needed to decipher the sources of pollution and discharges into the rivers. These techniques can assist by providing an enhanced view of selecting which rivers should be focused on for remediation.

ACKNOWLEDGEMENTS

The authors wish to express their sincere recognition to the Pasir Gudang City Council for providing raw data.

REFERENCES

Abdullah, J., Ahmad, Z., Shah, R. N. H. R. A., & Anor, N. (2012). Port city development and quality of life in Pasir Gudang Port, Johor, Malaysia. *Procedia-Social and Behavioral Sciences*, 35, 556-563. https://doi.org/10.1016/j.sbspro.2012.02.122

Academy of Sciences Malaysia. (2019). *ASM Report: Lessons from Sungai Kim Kim, Pasir Gudang. Academy of Sciences Malaysia*. https://www.akademisains.gov.my/asm-publication/lessons-from-sungai-kim-kim-pasir-gudang/

Alade, A. O., Jameel, A. T., Muyubi, S. A., Karim, M. I. A., & Alam, M. Z. (2011). Removal of oil and grease as emerging pollutants of concern (EPC) in wastewater stream. *IIUM Engineering Journal, 12*(4), 161-169.

Al-Badaii, F., Halim, A. A., & Shuhaimi-Othman, M. (2016). Evaluation of dissolved heavy metals in water of the Sungai Semenyih (Peninsular Malaysia) using environmetric methods. *Sains Malaysiana, 45*(6), 841-852.

Alkarkhi, A. F. M., Ahmad, A., & Easa, A. M. (2008). Assessment of surface water quality of selected estuaries of Malaysia: Multivariate statistical techniques. *The Environmentalist, 29*(3), 255-262.

Allan, J. D. (2004). Landscapes and riverscapes: The influence of land use on stream ecosystems. *Annual Review of Ecology, Evolution, and Systematics, 35*, 257-284.

APHA. (2005) *Standard methods for the examination of water and wastewater* (21st Ed.). American Public Health Association/American Water Works Association/Water Environment Federation. https://www.standardmethods.org/doi/book/10.2105/SMWW.2882

Bhat, S., Jacobs, J. M., Hatfield, K., & Prenger, J. (2006). Relationships between stream water chemistry and military land use in forested watersheds in Fort Benning, Georgia. *Ecological Indicators, 6*(2), 458-466. https://doi.org/10.1016/j.ecolind.2005.06.005

Bhuiyan, M. A., Rakib, M. A., Dampare, S. B., Ganyaglo, S., & Suzuki, S. (2011). Surface water quality assessment in the central part of Bangladesh using multivariate analysis. *KSCE Journal of Civil Engineering, 15*(6), 995-1003.

Chai, L. G. (2020). The river water quality before and during the Movement Control Order (MCO) in Malaysia. *Case Studies in Chemical and Environmental Engineering*, *2*, Article 100027. https://doi.org/10.1016/j.cscee.2020.100027

Department of Environment. (2016). River Water Quality. *Malaysia Environmental Quality Report.* https://enviro2.doe.gov.my/ekmc/digital-content/laporan-kualiti-alam-sekeliling-2016-2/

Halder, J. N., & Islam, M. N. (2015). Water pollution and its impact on human health. *Journal of Environment and Human, 2*(1), 36-46.

Hopkins, R. L. (2009). Use of landscape pattern metrics and multiscale data in aquatic species distribution models: A case study of a freshwater mussel. *Landscape Ecology, 24*(7), 943-955. https://doi.org/10.1007/s10980-009-9373-5

Hua, A. K., Kusin, F. M., & Praveena, S. M. (2016). Spatial variation assessment of river water quality using environmetric techniques. *Polish Journal of Environmental Studies, 25*(6), 2411-2421. https://doi.org/10.15244/pjoes/64082

Hussain, J., Husain, I., Arif, M., & Guota, N. (2017) Studies on heavy metal contamination in Godavari River basin. *Applied Water Science, 7*(8), 4539-4548. https://doi.org/10.1007/s13201-017-0607-4

Hussain, M., Ahmed, S. M., & Abderrahman, W. (2008). Cluster analysis and quality assessment of logged water at an irrigation project, eastern Saudi Arabia. *Journal of Environmental Management, 86*(1), 297-307. https://doi.org/10.1016/j.jenvman.2006.12.007

Ishak, A. M., & Ismail, M. A. (2020). Tool for water resources management. *Journal of Water Resources Management, 34***,** 447-462.

Isiyaka, H. A., & Juahir, H. (2015). Analysis of surface water pollution in the Kinta River using multivariate technique. *Malaysian Journal of Analytical Sciences, 19*(5), 1019-1031.

Khatri, N., & Tyagi, S. (2015). Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas. *Frontiers in Life Science, 8*(1), 23-39. https://doi.org/10.1080/21553769.2014.933716

Kim, J. O., & Mueller, C. W. (1987). *Introduction to factor analysis: What it is and how to do it.* Sage Publications.

Kumar, R. N., Solanki, R., & Kumar, J. N. (2013). Seasonal variation in heavy metal contamination in water and sediments of river Sabarmati and Kharicut canal at Ahmedabad, Gujarat. *Environmental Monitoring and Assessment, 185*(1), 359-368.

Lan, W. U., Gang, G. E., & Jinbao, W. A. N. (2009). Biodegradation of oil wastewater by free and immobilized Yarrowia lipolytica W29. *Journal of Environmental Sciences, 21*, 237-242.

Li, H., Liu, J., Liu, R., Xiong, N., Wu, K., & Kim, T. (2017). A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis. *Sensors, 17*(8), Article 1792. https://doi.org/10.3390/s17081792

Meyer, J. L., Paul, M. J., & Taulbee, W. K. (2005). Stream ecosystem function in urbanizing landscapes. *Journal of the North American Benthological Society, 24*(3), 602-612.

Oliveira, H. (2012). Chromium as an environmental pollutant: Insights on induced plant toxicity. *Journal of Botany, 2012*, Article 375843. https://doi.org/10.1155/2012/375843

Phiri, O., Mumba, P., Moyo, B. H. Z., & Kadewa, W. (2005). Assessment of the impact of industrial effluents on water quality of receiving rivers in urban areas of Malawi. *International Journal of Environmental Science & Technology, 2*(3), 237-244. https://doi.org/10.1007/BF03325882

Rosli, S. N., Aris, A. Z., & Majid, N. M. (2015). Spatial variation assessment of Malacca River water quality using multivariate statistical analysis. *Malaysian Applied Biology*, *44*(1), 13-18.

Samsudin, M. S., Azid, A., Khalit, S. I., Saudi, A. S. M., & Zaudi, M. A. (2017). River water quality assessment using APCS-MLR and statistical process control in Johor River basin, Malaysia. *International Journal of Advanced and Applied Sciences, 4*(8), 84-97. https://doi.org/10.21833/ijaas.2017.08.013

Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River Basin, Japan. *Environmental Modelling and Software*, *22*, 464-475. http://dx.doi.org/10.1016/j.envsoft.2006.02.001

Singh, K. P., Malik, A., Singh, V. K., Mohan, D., & Sinha, S. (2005). Chemometric analysis of groundwater quality data of alluvial aquifer of Gangetic plain, North India. *Analytica Chimica Acta, 550*, 82-91. https://doi.org/10.1016/j.aca.2005.06.056

The Straits Times. (2019, March 21). Cleaning up toxic river Sungai Kim Kim in Pasir Gudang to cost S$2.16 million. *The Straits Times*. https://www.straitstimes.com/asia/se-asia/cleaning-up-toxic-river-sungai-kim-kim-in-pasir-gudang-to-cost-s216-million

Uddin, M. G., Nash, S., & Olbert, A. I. (2021). A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators, 122*, Article 107218. https://doi.org/10.1016/j.ecolind.2020.107218

Zhang, H., Cui, B., Xiao, R., & Zhao, H. (2010). Heavy metals in water, soils and plants in riparian wetlands in the Pearl River Estuary, South China. *Procedia Environmental Sciences, 2*, 1344-1354. https://doi.org/10.1016/j.proenv.2010.10.145