www.czasopisma.pan.pl PAN www.journals.pan.pl

Archives of Environmental Protection
Vol. 44 no. 4 pp. 111–122

PL ISSN 2083-4772
DOI 10.24425/aep.2018.124575

An identification source of variation on the water quality pattern in the Malacca River basin using chemometric approach

Ang K. Hua

Universiti Putra Malaysia

Corresponding author's e-mail: angkeanhua@yahoo.com

Keywords: hierarchical cluster analysis; discriminant analysis; principal component analysis; multiple linear regression analysis.

Abstract: The Malacca River basin experienced river water pollution which caused a major deterioration to the ecosystems and environmental health. This study is carried out to assess the water quality data and identify the pattern of water pollution sources in the study area, and also to develop a predictive performance of water quality in the Malacca River basin. A chemometric approach using a combination of HCA, DA, PCA, and MLR, was applied into twenty water quality variables from nine sampling stations that were collected from January until December of 2015 in the river basin. HCA pointed out three clusters, namely Cluster 1 (C1) with low pollution source, Cluster 2 (C2) with moderate pollution source, and Cluster 3 (C3) with high pollution source. In the DA analysis, the results showed 21 variables, 12 variables, and 9 variables for standard mode, forward stepwise mode, and backward stepwise mode, respectively. Meanwhile, the PCA indicated that the main source of pollutants is detected from residential, industrial, commercial, agricultural, animal livestock, as well as forest land. Among the three models developed from MLR analysis, C3 with a high pollution source is detected to be the most suitable model to be used for the prediction of Water Quality Index in the Malacca River basin. This study proposed for an effective river water quality management by having new water quality monitoring network to be designed for more practical use in order to reduce time and effort, as well as cost saving purposes.

Introduction

Rivers have been water resources for centuries. It became valuable resources to the economic development of a country. The rivers play an important role in providing a variety of goods and services to the aquatic ecosystems that are worth billions of dollars (Huang et al. 2017, Kostecki et al. 2017, Li et al. 2017, Barbier et al. 2011). Without any warning, the rapid urbanization within the river basin has caused river water pollution that has received many inputs from both natural and anthropogenic origin (Hua et al. 2016, Al-Badaii et al. 2016, Lim et al. 2013). Based on the Report from the Department of Environmental Malaysia (2012), the main contributions of river pollution in Malaysia are mainly from manufacturing industries, domestic sewage and/or livestock farming, urban settlements, agricultural runoff, as well as improper earthworks and land clearing activities. In other words, the continuous monitoring of river water quality indicates that the clean water is 278 (59%), the slightly polluted is 161 (34%), and the polluted is 34 (7%) for a total of 473 rivers that were monitored (DOE, 2012). The report on the river water quality assessment also listed the Malacca River as one of the basin areas that would be exposed to water pollution due to rapid development which has occurred in the Malacca State. Since the Malacca State was awarded the UNESCO World Heritage site as a historical tourism center in 2007 (UNESCO Official Portal 2007), there is no doubt that urbanisation development has doubled as compared to the past 10 years (Hua 2017, Rosli et al. 2015, Daneshmand et al. 2011). Therefore, frequent water assessment and monitoring is essential to prevent the occurrence of extreme contamination in the river.

To assess the amount of contamination on the water quality, the monitoring data on a wide range of physical, chemical and biological parameters is required. Nevertheless, a large number of available data will cause difficulty in analysing the water quality. Hence, the advantages of specific statistical methods are able to benefit by obtaining the meaningful results. The most common method applied to analysing the data, is known as the chemometric technique analysis, which is involved with the hierarchical cluster analysis (HCA), the discriminant analysis (DA), the principal component analysis (PCA), as well as multiple linear regression analysis (MLR) (Lim et al. 2013, Mustapha and Aris 2011). Chemometric techniques have the ability to explore the assessment of water quality datasets and interpret the complex data matrices to better understand the identification of possible sources that influence the water systems (Mustapha and Aris 2011). Therefore, these methods proved to be priceless tools for developing suitable plans for the efficient management of the river water quality monitoring network (Al-Badaii et al. 2016, Mustapha and Aris 2011).

The objective of this study is to illustrate the view of water quality in the Malacca River basin by recognising the pollution source, identify the most significant water quality variables in this study, and to develop a predictive performance of water quality along the study area.

Methods & materials

Study area

Situated between 2°13'12.80"N to 2°24'11.66"N and 102° 9'9.98"E to 102°18'44.69"E is the Malacca River Basin. Before reaching the Straits of Malacca, is a main river that runs from Alor Gajah district to the central district of Malacca. Positioned along the river there is one reservoir, known as the Durian Tunggal Reservoir, with a catchment of 20 km². This reservoir serves as the primary source of water supply to all the inhabitants in Malacca.

The rapid growth in the population of Malacca has brought about developments which are beyond control and management, namely housing, sewage, transportation, as well as a critical need for water supply (Hua et al. 2016, Rosli et al. 2015). Based on the viewpoint for land use in respect of the Malacca River basin, most of the inhabitants are mostly concentrated in the city centre; which extends about 10km to the west, 10 km to the east and 20 km to the north (Hua 2017). The local authorities of Malacca are of the view that there is an urgent need to protect the quality of the river water from being contaminated continuously.

The river basin measuring about 670 km and covering 80 km length of the Malacca River drifts into the state. The river encompasses 13 subbasins of watershed, namely Kampung Ampang Batu Gadek subbasin, Kampung Balai subbasin, Kampung Batu Berendam subbasin, Kampung Buloh China subbasin, Kampung Cheng subbasin, Kampung Gadek subbasin, Kampung Harmoni BelimbingDalam subbasin, Kampung Kelemak subbasin, Kampung Panchor subbasin, Kampung Pulau subbasin, Kampung Sungai Petai subbasin, Kampung Tanah Merah subbasin, and Kampung Tualang subbasin. Out of the 13 subbasins, only 7 subbasins with 9 sampling stations along the river were chosen (Figure 1).

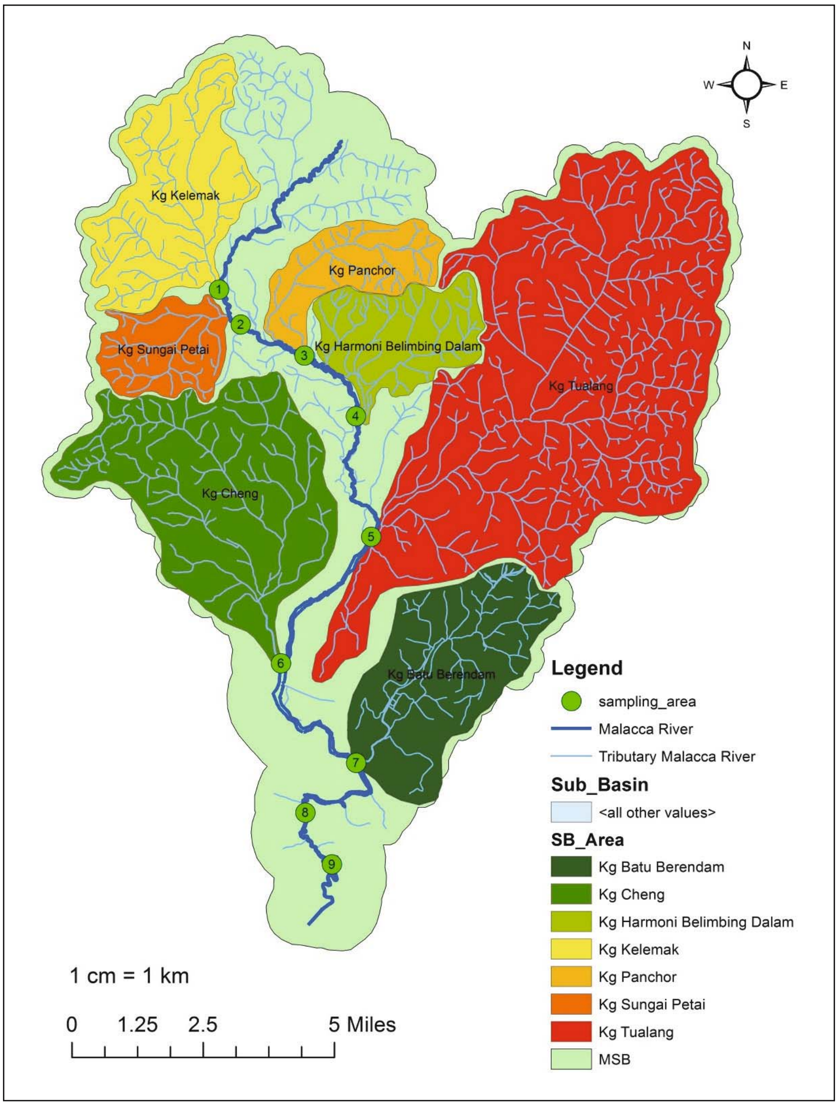

Fig. 1. Seven subbasin with nine sampling stations along Malacca River Basin

Data collection and water quality analysis

Table 1 illustrates the locations of the 9 sampling stations which were documented using the GPS system. Commencing January to December 2015, water quality samples were gathered every month. The data on the quality of the river water consists of the physic-chemical parameters: pH, temperature, electrical conductivity (EC), salinity, turbidity, total suspended solids (TSS), total dissolved solids (TDS), dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), ammoniacial nitrogen (NH_3N), trace elements (including mercury, cadmium, chromium, arsenic, zinc, lead, and iron), and biological parameters (which include Escherichia coliform and total coliform).

In-situ measurements were carried out to measure (1) pH using a SevenGo Duo pro pH meter (Mettler Toledo AG), (2) turbidity using a portable turbidity meter (Handled Turbidimeter Hach 2100), and (3) using a multi-parameter probe (Orion Star Series Portable Meter) to measure the temperature, EC, DS, salinity and DO. An analysis on NH_3N was conducted, simultaneously, using the Hach Method 8038 spectrophotometer at a specific wavelength, while COD was measured using the APHA 5220B open reflux technique, BOD using APHA 5210B, and TSS using the APHA 2540D method. Applying the membrane filtration method based on APHA 9221B, both the E-coli and total coliform were also examined. In the trace metal analysis, 500mL of water sample was sieved through 0.45 µm Whatman filter paper and acidified with nitric acid (HNO_3) to pH lower than 2, and then examined using inductive-coupled plasma-mass spectrometry (ICP-MS, ELAN DRC-e, Perkin Elmer).

Quality assurance and quality control

Prior to conducting any laboratory analysis, it is vital that the laboratory apparatus and polyethylene bottles be thoroughly washed using 5% (v/v) of nitric acid and soaked overnight to remove any contaminants and remains of cleaning reagent (APHA 2005). For the purpose of BOD analysis, BOD bottles were covered with aluminium foil. The collected samples of river water were then preserved using 1% (v/v) nitric acid (HNO3) for trace metals and then examined within one month. Every sample had to be examined three times before the mean value was calculated. To obtain the accuracy of each parameter measured with less than 20%, a standard deviation (SD) was applied. Before carrying out any analysis, all the probe meters and instruments had to be calibrated first. In most cases, for the purpose of reducing any matrix interference during analysis, the standards and blanks were handled in the same manner as the representative river water samples.

The accuracy of ICP-MS performance is based on the diluting preparation using ICP Multi-Element Mixed Standard III (Perkin Elmer) into concentration with the same acid mixture used for sample dissolution. The recovery of samples for all the targeted elements have complied with the standard requirements (90–110%).

Data analysis

All the river water quality data will be scrutinized using the Statistical Package for Social Sciences, version 23 (SPSS v.23) for chemometric techniques using HCA, DA, PCA and MLR.

Hierarchical cluster analysis

In order to cluster the different objects, with similarities and associations, into one group, HCA was employed which involves the following three ways (Boyacioglu and Boyacioglu 2017, Baharuddin et al. 2014, Lim et al. 2013, Gazza et al. 2012, Mustapha and Aris 2011):

(1) Ward's method – the variance analysis is employed to gauge the distance between clusters with a reduced sum of square (SS) for any two clusters that are formed at each step,
(2) Squared Euclidean Distance – creates similarity between two samples and a distance which can be represented by differences between analytical values from the samples,
(3) Dendrogram – the results from this diagram show the group with high similarity and small distances between clusters; while the dissimilarity between the groups represented by the maximum of all possible distances between clusters.

HCA can be defined by Eq. (1);

$$d(x,y) = \sum_{m-1}^{p} (x_m - y_m) \tag{1}$$

where *d(x.y)* is the Euclidean distance between two samples represented by x_m and y_m, and p is the dimensional space of the variables (Bierman et al. 2011). This study applies HCA to examine the grouping of the sample sites (spatial).

Table 1. Geographical Coordinate of nine (9) sampling stations details in Malacca River Basin, Malaysia

Sampling Station	Malacca River Subbasin	Latitutde	Longitude
1	Kampung Kelemak	2°21'57.09"N	102°13'7.15"E
2	Kampung Sungai Petai	2°21'23.72"N	102°13'29.07"E
3	Kampung Panchor	2°20'54.60"N	102°14'31.82"E
4	Kampung Harmoni Belimbing Dalam	2°19'43.97"N	102°15'28.67"E
5	Kampung Tualang	2°17'59.11"N	102°15'45.88"E
6	Kampung Cheng	2°15'45.82"N	102°14'10.56"E
7	Kampung Batu Berendam	2°14'4.10"N	102°15'24.80"E
8	–	2°13'23.06"N	102°14'34.03"E
9	–	2°12'29.45"N	102°15'5.21"E

Discriminant analysis

DA is a statistical tool capable of discriminating between two or more groups or clusters by introducing a discriminant function (DF) for each group (Boyacioglu and Boyacioglu 2017, Lim et al. 2013, Gazza et al. 2012, Mustapha and Aris 2011). DA can be defined by Eq. (2);

$$f(G_i) = k_i + \sum_{j=1}^{n} W_{ij}P_{ij} \tag{2}$$

where i is the number of groups (G), k_i is the constant inherent to each group, n is the number of parameter employed to categorize a set of data into a specific group, and w_{ij} is the weight coefficient allocated by DF analysis (DFA) to a specified parameter (P_{ij}). For this study, DA was employed to explain whether the mean of variables differ within the groups and the variables will be utilized to predict the group pattern. Depending on the grouping of HCA results, the raw data are examined using DA which are involved with standard, forward stepwise, and backward stepwise modes to develop the DFs in assessing the spatial variations of river water quality. This study assigned the stations (spatial) as dependent variables (which is referred to as grouping), and all parameters are independent variables.

Principal component analysis

PCA is a statistical technique with the ability to provide information on most significant parameters due to spatial and temporal variations that explains the whole data set by excluding less significant parameters with a minimum loss of original information (Boyacioglu and Boyacioglu 2017, Baharuddin et al. 2014, Lim et al. 2013, Gazza et al. 2012, Mustapha and Aris 2011). PCA can be defined by Eq. (3);

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{1j} + \cdots + a_{im}x_{mj} \tag{3}$$

where z is the component score, a is the component loading, x is the measured value of the variable, i is the component number, j is the sample number, and m is the total number of variables. The procedures used in PCA are (1) the hypothesis in an original data group then reduced to dominant components or factors (source of variation) that influence the observed data variance; and (2) the whole data set extracted through eigenvalues and eigenvectors from the square matrix produced by multiplying the data matrix (Lim et al. 2013, Mustapha and Aris 2011). Eigenvalues which are greater than 1 are considered significant to perform a new group of variables, namely varimax factors (VFs). VF coefficients that have a correlation greater than 0.75 are considered 'strong', 0.75 to 0.50 as 'moderate', and 0.50 to 0.30 as 'weak' (Mustapha and Aris 2011). In this study, PCA was applied to the normalized data set (21 variables) separately based on different spatial regions obtained from the HCA techniques.

Multiple linear regressions

MLR method is suitable for investigating the relationship between independent and dependent variables through formation of linear equation on observed data and giving a percentage on each parameter of the river water quality (Hamid et al. 2016, Lim et al. 2013, Mustapha and Aris 2011). This study adopts the MLR method to justify the relationship between water quality parameter (the most significant within the 21 variables) with total water quality index (WQI) data. The MLR model can be defined by Eq. (4);

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{p-1}x_{p-1} + \varepsilon \tag{4}$$

where Y is the response variable, p-1 is the explanatory variable for $x_1, x_2, \ldots, x_{p-1}$ with p is the parameter (regression coefficient) of β_0, β_1, β_2, …, β_{p-1}, and ε is the error associated with the regression. In determining the best fitting linear regression equation, the coefficient of determination (R^2), adjusted coefficient of determination (Adjusted R^2), and root mean square error (RMSE). The value of R^2 furnishes information on how well the model performs on the external data; Adjusted R^2 is considered all possible number of variables; and RMSE measures the residual error and the mean difference between observed and modeled value of WQI (Hamid et al. 2016, Lim et al. 2013, Mustapha and Aris 2011). Generally, the higher R^2 value (which is near to 1) will be considered as the best linear model (Hamid et al. 2016).

Results and discussion

The characteristic of river water quality data

Table 2 indicates the results of the mean and standard deviation values of the Malacca River water quality for physic-chemical (include trace elements) and biological parameter data for year 2015. The river water quality of physical parameter shows that salinity (S1 to S3 and S7), electrical conductivity (S1 and S7), total dissolved solid (S1) and turbidity (S3) are suspected in class 5; while class 4 is detected in S8 and S9 in turbidity; and class 3 resulted in total suspended solid (S1, S3 to S7, and S9), turbidity (S1 and S5), and total dissolved solid (S7), as well as other stations that have class 2 and class 1 (Table 3). In other words, the order of contamination concentration from high to low is salinity > turbidity > electrical conductivity > total suspended solid > total dissolved solid, which has highlighted that land clearing for agricultural and animal husbandry activities have occurred in the Malacca River basin. Meanwhile, the chemical parameter of river water quality shows that NH_3N (S1 to S3 and S7 to S8) and BOD (S2 and S7 to S9) have resulted in class 4; while class 3 is detected in COD (S1 to S3 and S7 to S8), BOD (S1 and S3 to S6), DO (S1 to S3 and S7) and NH_3N (S4 to S6 and S9), and the rest of stations is in class 2 and class 1 (Table 3). Higher contamination concentration is NH_3N, followed by BOD, COD, and DO. Lastly, trace elements, pH, and temperature are detected in class 1; while majority of the biological parameters is suspected in class 5. Overall, the result shows that residential, commercial, animal husbandry, as well as industrial activities are carried out along the Malacca River.

Spatial classification based on water quality parameters

The analysis of HCA indicates that 3 clusters were formed from 9 sampling stations (Figure 2). Cluster 1 (C1) consists of stations 1, 2, and 3, while cluster 2 (C2) consists of stations

Table 2. Mean and standard deviation values of water quality data along Malacca River in 2015

Category	Unit		S1	S2	S3	S4	S5	S6	S7	S8	S9
pH	–	Mean	6.497	6.433	6.459	6.456	6.545	6.490	6.635	6.398	6.331
		SD	0.336	0.280	0.311	0.430	0.213	0.312	0.355	0.339	0.469
Temp	°C	Mean	27.191	26.892	26.883	26.633	26.592	26.858	27.617	27.492	28.333
		SD	0.872	1.134	0.810	0.967	0.723	1.013	0.826	1.058	0.716
Sal	%	Mean	21.036	9.388	4.003	0.508	0.065	0.054	7.000	0.308	0.060
		SD	9.776	3.504	3.323	0.436	0.028	0.018	8.707	0.271	0.041
EC	µS/cm	Mean	16330.217	1403.733	1950.717	280.425	218.367	109.600	8173.900	1069.333	1093.833
		SD	12329.039	1067.370	1366.128	154.037	98.446	29.718	11118.647	459.168	630.463
TSS	mg/L	Mean	51.083	38.750	59.833	116.167	97.667	75.083	50.750	44.750	67.000
		SD	15.808	11.624	16.525	97.404	65.003	63.977	50.713	13.844	20.684
TDS	mg/L	Mean	10444.608	1095.967	1241.167	588.250	454.833	610.000	3360.250	720.583	271.500
		SD	7745.212	592.667	709.869	315.438	234.016	290.743	5452.005	450.054	126.174
Tur	NTU	Mean	116.700	73.608	583.567	99.858	121.175	84.850	63.617	297.767	209.708
		SD	67.180	29.591	494.931	70.092	65.260	30.013	47.167	128.606	276.701
BOD	mg/L	Mean	5.083	6.500	4.500	5.167	5.333	5.167	6.500	9.250	6.583
		SD	1.929	2.780	0.905	1.267	1.073	0.718	1.314	2.094	2.843
COD	mg/L	Mean	34.167	47.167	33.250	24.000	23.917	24.333	27.584	46.250	24.667
		SD	11.900	27.425	11.450	11.763	7.255	5.297	8.712	16.966	6.624
DO	mg/L	Mean	4.113	4.160	3.886	5.017	5.606	5.116	3.347	5.541	5.526
		SD	0.880	1.530	1.461	1.422	0.971	1.006	1.238	2.191	0.863
NH3N	mg/L	Mean	0.992	1.081	1.417	0.771	0.521	0.676	2.385	1.760	0.646
		SD	0.467	0.502	1.074	0.211	0.279	0.406	1.701	0.781	0.195
As	mg/L	Mean	0.003	0.004	0.002	0.003	0.002	0.002	0.003	0.004	0.004
		SD	0.004	0.005	0.002	0.001	0.001	0.001	0.001	0.001	0.002
Hg	mg/L	Mean	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD
		SD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD
Cd	mg/L	Mean	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		SD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD
Cr	mg/L	Mean	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.003	0.001
		SD	0.002	0.004	0.002	<LOD	<LOD	<LOD	<LOD	0.004	<LOD
Pb	mg/L	Mean	0.010	0.010	0.010	0.001	0.010	0.010	0.010	0.010	0.010
		SD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD
Zn	mg/L	Mean	0.058	0.050	0.057	0.010	0.048	0.048	0.043	0.051	0.057
		SD	0.023	0.020	0.022	<LOD	0.021	0.019	0.021	0.015	0.022
Fe	mg/L	Mean	0.482	0.400	0.563	0.053	0.878	1.141	0.357	0.786	0.890
		SD	0.412	0.355	0.371	0.028	0.439	1.033	0.284	0.597	0.131
Total Coliform	count/100mL	Mean	698366.67	584866.67	701375.00	1005583.33	571808.33	656116.67	155958.33	117783.33	161558.33
		SD	510510.49	198044.61	448855.58	865227.40	340474.95	244295.76	107181.10	74326.53	94999.84
Ecoli	count/100mL	Mean	71275.00	63275.00	40358.33	17317.33	7634.17	20766.67	13745.00	14829.17	11455.83
		SD	33664.63	32261.69	43663.93	22922.74	5087.25	35458.43	6812.92	6701.57	7260.65

(Tur = Turbidity; TDS = Total Dissolved Solids; Con = Electrical Conductivity; Sal = Salinity; Temp = Temperature; DO = Dissolved Oxygen; BOD = Biochemical Oxygen Demand; COD = Chemical Oxygen Demand; TSS = Total Suspended Solids; pH = Acidic or Basic water; NH_3N = Ammoniacal Nitrogen; E coli = *Escherichia* Coliform; Coli = Coliform; As = Arsenic; Hg = Mercury; Cd = Cadmium; Cr = Chromium; Pb = Lead; Zn = Zinc; Fe = Iron; SD = Standard Deviation; S1 to S9 = Station 1 to Station 9; <LOD = < Limit of Detection; Very Good Condition=Blue; Good Condition=Green; Moderate Condition=Yellow; Weak Condition=Orange; Bad Condition=Red)

Table 3. National Water Quality Standards for Malaysia (Source: DOE (Malaysia) report, 2012)

Category	Unit	Class					
		I	IIA	IIB	III	IV	V
pH	–	6.5–8.5	6–9	6–9	5–9	5–9	–
Temp	°C	–	Normal + 2°C	–	Normal + 2°C	–	–
Sal	%	0.5	1	–	–	2	–
EC	μS/cm	1000	1000	–	–	6000	–
TSS	mg/L	25	50	50	150	300	300
DS	mg/L	500	1000	–	–	4000	–
Tur	NTU	5	50	50	–	–	–
BOD	mg/L	1	3	3	6	12	>12
COD	mg/L	10	25	25	50	100	>100
DO	mg/L	7	5–7	5–7	3–5	< 3	< 1
NH3N	mg/L	0.1	0.3	0.3	0.9	2.7	> 2.7
As	mg/L	–	0.05	0.05	0.4 (0.05)	0.1	–
Hg	mg/L	–	0.001	0.001	0.004(0.0001)	0.002	–
Cd	mg/L	–	0.01	0.01	0.01 (0.001)	0.01	–
Cr	mg/L	–	0.05	0.05	1.4 (0.05)	0.1	–
Pb	mg/L	–	0.05	0.05	0.02 (0.01)	5	–
Zn	mg/L	–	1	1	3.4	0.8	–
Fe	mg/L	–	1	1	1	1 (leaf) 5 (others)	–
Total Coliform	count/100 mL	100	5000	5000	5000 (20000)	5000 (20000)	> 50000
Ecoli	count/100 mL	10	5000	5000	50000	50000	> 50000

(Tur means Turbidity; DS means Dissolved Solid; TDS means Total Dissolved Solid; EC means Electrical Conductivity; Sal means Salinity; Temp means Temperature; DO means Dissolved Oxygen; BOD means Biological Oxygen Demand; COD means Chemical Oxygen Demand; TSS means Total Suspended Solids; pH means Acidic or Basic water; NH_3N means Ammoniacal Nitrogen; E coli means *Escherichia* Coliform; Coli means Coliform; As means Arsenic; Hg means Mercury; Cd means Cadmium; Cr means Chromium; Pb means Lead; Zn means Zinc; Fe means Iron)

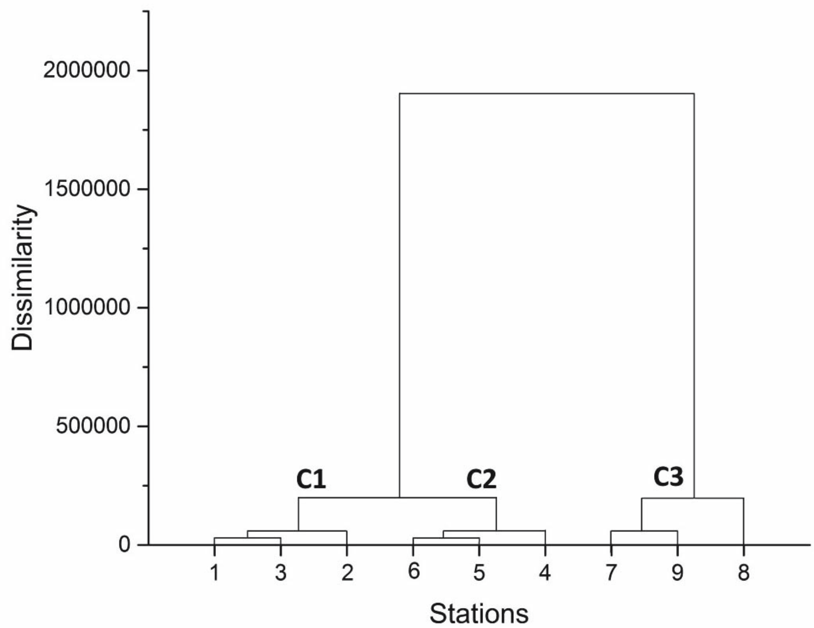

Fig. 2. HCA using Ward linkage method to generate dendogram

4, 5, and 6, as well as cluster 3 (C3) consists of stations 7, 8, and 9. Based on the National Water Quality Standard Malaysia, the C1 are having the average value of WQI is 75 to resulted as a low pollution source (LPS), while C2 having an average value of WQI is 61 to react as a moderate pollution source (MPS), and C3 as a high pollution source (HPS) with the average value of WQI is 52 (Table 3). In other words, the river basin that is constituted by LPS are Kampung Kelemak subbasin, Kampung Sungai Petai subbasin, and Kampung Panchor subbasin, and it is likely to occur in rural areas; while MPS are Kampung Harmoni Belimbing Dalam subbasin, Kampung Tualang subbasin, and Kampung Cheng subbasin, which occur in the sub-urban area, and HPS are Kampung Batu Berendam subbasin, which occur in the urban area. So, the HCA technique has proved that the ability to reduce the monitored stations, especially by suggesting that the category of water quality is based on the entire region, are beneficial to improving the monitoring network in the future.

Discriminant analysis based on spatial variation

A further analysis by applying the DA method based on the clustering provided from HCA is referred to as C1, C2, and C3. In the DA techniques, the methods involved three modes, which are referred to as standard, forward stepwise, and backward stepwise. The analysis of DA shows that the accuracy of spatial variation for standard mode, forward stepwise mode, and backward stepwise mode is 98.62% with 21 variables, 95.55% with 12 variables as well as 9 variables, respectively (Table 4). In this study, null hypothesis (H_0) stated that at least one of the mean vectors is different from the others, while an alternative hypothesis (H_a) stated that the mean vectors of the three classes are equal. Simultaneously, the *p*-value is lower than the significant level of alpha (0.05), and the null hypothesis (H_0) will be rejected by accepting the alternative hypothesis (H_a). Since the Pillai's Trace test for standard, forward and backward provided the result of 1.612, 1.355, and 1.547 respectively, are above than 0.01%; as well as the *p*-value resulted is 0.001 which is lower than 0.05% and it is true to reject the H_0, whereby indicating the three classes are having the same mean vectors. Therefore, 9 variables (which includes temperature, turbidity, salinity, BOD, COD, As, Fe, total coliform and Escherichia coliform) of the river water quality were selected which showed high spatial variations (with the most significant p-value less than 0.05) for the backward stepwise mode which were applied into the box and whisker plots for further discussion (Figure 3).

Identification source of variation

PCA was used to determine the pattern of water quality variables and identify the factor based on the discovery regions of C1, C2, and C3. The results in Table 5 indicated that seven VFs were obtained in the three regions with the eigenvalues greater than 1. The total variance for C1, C2, and C3 regions was 75.08%, 68.58%, and 74.20% respectively.

(1) Cluster 1 Region

The C1 regions, which were detected as low pollution source, indicated that the VF1 has 16.64% of the total variance to produce strong positive significant loadings of DS, EC and NH_3N, as well as moderate positive loadings of TS and salinity. Based on the result, it shows that the occurrence of contaminations can be connected with the erosion of riverbanks due to dredging activities that happened within the Malacca River. Moreover, the existence of salinity and NH_3N is suspected from agricultural runoff (Aris et al. 2013) and animal husbandry activities to cause pollution in the river (Mustapha and Abdu 2012). Specifically, agricultural activities are involved with pesticide usage in oil palm-rubber plantations, as well as animal farms of chicken, cow and goat which are carried out along the river, and could result in the non-point source pollution that happens through surface water flows entering the nearby sub-basin. Meanwhile, in VF2 it shows the total variance of 12.37% to produce strong positive significant loadings of BOD and COD. This represents the influence of organic pollutants (Simeonov et al. 2003), but probably is interpreted to represent the non-point source

Table 4. Classification matrix of DA for spatial variation in Malacca River Basin

Sampling Regions	% Correct	Regions assigned by the DA		
		Cluster 1	Cluster 2	Cluster 3
Standard Stepwise				
Cluster 1	94.29%	12	9	9
Cluster 2	97.14%	9	15	11
Cluster 3	99.19%	9	12	14
Total	98.62%	30	36	34
Forward Stepwise				
Cluster 1	91.43%	15	11	9
Cluster 2	82.86%	10	16	9
Cluster 3	98.89%	10	11	14
Total	95.55%	35	38	32
Backward Stepwise				
Cluster 1	94.29%	14	11	10
Cluster 2	88.57%	10	13	12
Cluster 3	98.76%	11	10	14
Total	95.55%	35	34	36

pollution that came from the agricultural activities and forest area (Juahir et al. 2011). The VF3 indicates a total variance of 10.60% to result to the strong negative loading of turbidity and moderate negative loading of the total suspended solid. In other words, the occurrence of both contaminations can be related to soil erosion that is involved with human interruption activities towards the hydrologic modifications such as dredging, water diversions, and channelization to cause disruption in the river (Daneshmand et al. 2011).

On the other hand, the result shows a moderate positive loading of the total coliform to produce 9.53% of total variance in VF4, while a strong positive loading of DO to have 8.94% of total variance in VF5. In this case, the existence of total coliform can result from the discharge into the river through the surface runoff of domestic waste and fertilizer used in the agricultural activities. However, Papaioannou et al. (2010), stated that contamination involved with total coliform is probably due to direct input by a warm blooded animals (such as animal farm activities) or through the soil that flow into the river. The DO contamination refers to the high level of dissolved organic matter that has consumed large amount of oxygen, which was detected to come from agricultural activities and forest areas. Lastly, the VF6 and VF7 indicate a strong positive loading of As and Zn to provide a total variance of 8.74% and 8.26%, respectively. The pollution from As is suspected to be from agricultural land, while Zn is from village houses that have zinc roofs. The Zn contamination in the river could happen due to the houses and buildings uses metallic roofs, which mobilize into atmosphere and waterways when in contact with acid rain or smog (Juahir et al. 2011).

(2) Cluster 2 Region

A moderate pollution source region is suspected to result in VF1 having a total variance of 13.04% to provide strong positive loadings of DS, TS, and total coliform, compared to a moderate positive loading of temperature. This contamination is detected from anthropogenic activities, which originates from point source and non-point source pollution sources (Aris et al. 2013, Juahir et al. 2011). Meanwhile, VF2 indicated 12.25% of total variance to provide strong positive loadings of turbidity and salinity, as well as moderate positive loadings of electrical conductivity and total suspended solid. As explained before, the contamination of turbidity, electrical conductivity and total suspended solid came from human activities involved

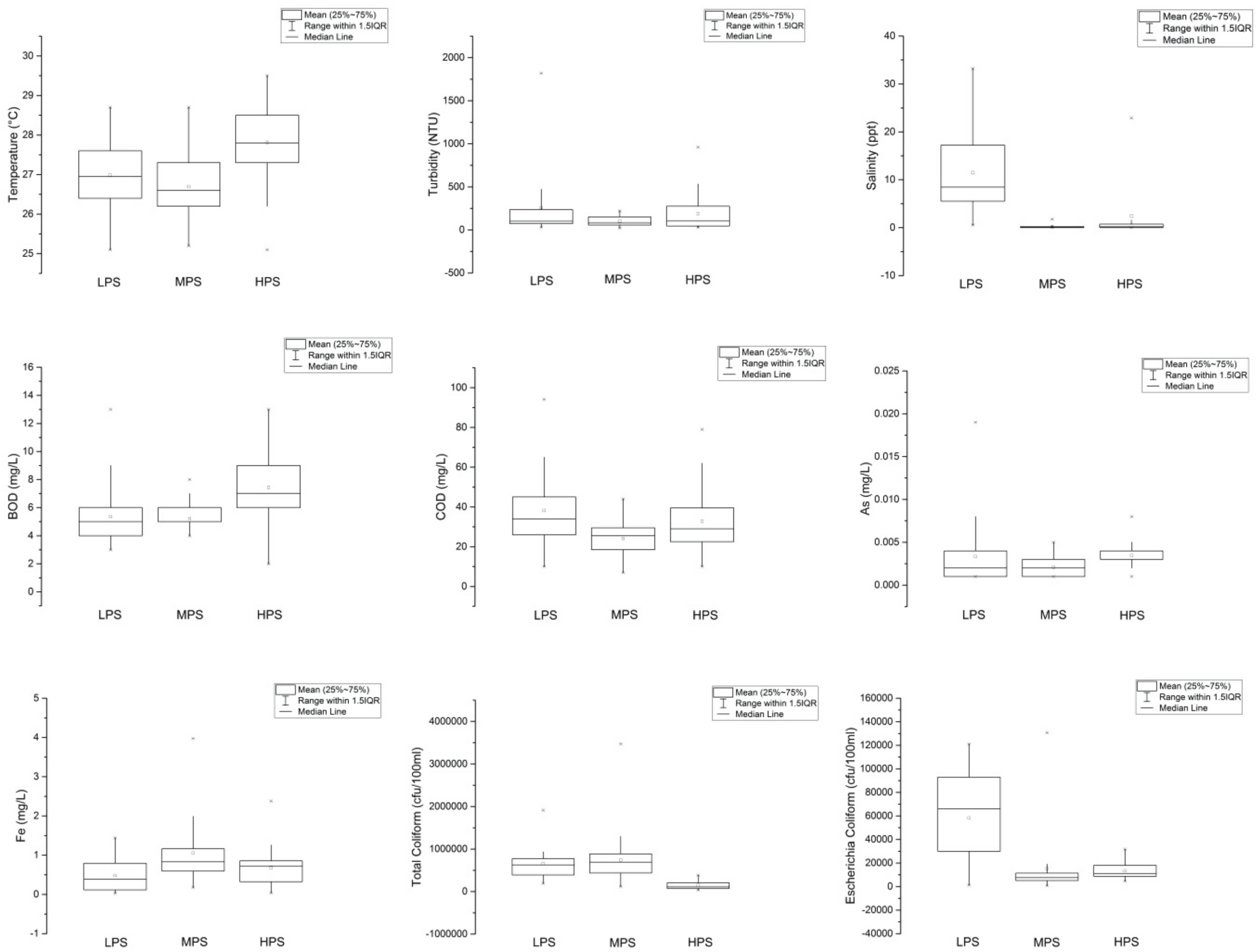

Fig. 3. Box and whisker plot for temperature, turbidity, salinity, BOD, COD, As, Fe, total coliform and Escherichia coliform, that generated from backward stepwise mode in DA of river water quality in Malacca River Basin

Table 5. Varimax rotation PCs for water quality data based on three clusters in Malacca River basin

VAR	Cluster 1							Cluster 2							Cluster 3						
	VF1	VF2	VF3	VF4	VF5	VF6	VF7	VF1	VF2	VF3	VF4	VF5	VF6	VF7	VF1	VF2	VF3	VF4	VF5	VF6	VF7
Tur	-.153	.058	**-.830**	.022	-.016	-.024	-.058	-.065	**.868**	.041	-.020	-.126	-.133	.094	-.263	.474	**-.733**	.093	-.367	-.260	-.068
DS	**.939**	-.010	.048	0.24	.091	-.025	.086	**.843**	.092	.101	.025	-.095	.104	.059	**.551**	-.133	-.103	-.099	-.180	.308	-.091
TS	**.667**	.380	.029	-.027	-.353	.086	-.264	**.794**	-.055	-.099	.219	.196	.059	.134	**.924**	-.096	-.038	-.006	-.012	.084	.006
EC	**.899**	-.016	.047	-.080	.038	.038	-.023	.415	**.577**	-.085	.066	.348	.249	-.220	**.903**	-.108	.041	-.006	.066	.100	-.028
Sal	**.715**	-.032	.121	-.068	.335	.272	.176	-.074	**.792**	-.098	-.075	-.075	.312	.162	**.772**	.037	.017	-.133	.073	-.204	-.201
Temp	.292	.237	.393	-.208	.145	.250	.355	**-.598**	.112	-.119	.448	.025	.239	.328	-.021	.199	.039	-.102	**.788**	-.043	-.079
DO	.028	.162	-.019	-.193	**.809**	.116	-.022	-.115	.115	**.718**	.059	-.048	-.122	.083	-.220	.356	-.270	**.525**	.342	-.207	-.191
BOD	.084	**.757**	-.215	.014	.044	.234	-.029	.180	-.048	.402	**.771**	.119	-.241	.443	-.184	-.182	.023	**.756**	-.168	-.265	.241
COD	-.358	**.860**	.374	.364	.260	.486	-.039	-.169	-.498	**.601**	.034	-.113	-.071	.022	.056	.241	.161	**.851**	-.014	.070	-.011
TSS	-.030	.480	**-.517**	-.076	-.257	-.306	.066	.365	**.666**	.086	.111	-.174	.203	-.032	-.395	-.014	**.634**	-.205	.076	.005	.346
PH	.155	-.073	.193	.027	-.151	-.195	-.253	-.045	-.146	-.130	-.086	**-.817**	.185	-.098	.173	**-.528**	.170	.077	-.232	-.167	-.115
NH_3N	**.857**	-.158	-.091	-.086	-.050	-.034	.035	.166	.147	.169	**-.795**	.203	.205	.131	.002	-.229	.008	.267	**.892**	.091	-.062
Ecoli	.205	-.174	.008	.094	-.024	.435	-.201	.147	.239	-.183	-.120	**.720**	-.101	.073	.436	**.757**	.107	.452	-.096	.399	-.216
TColi	.023	.018	-.067	**.627**	.252	.077	.122	**.783**	-.408	-.054	-.060	-.205	.122	-.191	.098	**.781**	-.297	.004	.267	.749	-.071
As	.131	-.060	-.072	-.111	-.033	**.865**	-.041	-.174	.075	.194	-.466	-.061	-.053	.177	-.066	**-.557**	-.201	.339	.148	-.466	.064
Hg	-.115	.303	-.049	.054	.083	.231	.027	-.005	.080	-.138	.091	-.155	-.263	**-.654**	-.040	-.077	-.087	-.047	.154	**.864**	.151
Cr	.012	-.138	.072	-.015	-.098	-.153	.172	-.023	.002	.321	-.194	.289	.132	.094	-.093	.119	-.059	-.043	-.046	-.214	**-.776**
Zn	.285	-.202	-.365	.180	.163	-.052	**.875**	-.006	.295	.057	.070	-.045	**.771**	.202	-.259	.201	.004	.051	.024	**.844**	.053
Fe	.255	.337	.455	-.320	.207	-.338	-.402	.190	-.034	.082	-.102	-.102	**.804**	-.043	.047	.129	-.270	.084	-.057	.073	**.798**
IE	3.161	2.349	2.015	1.811	1.698	1.660	1.570	2.478	2.330	1.870	1.859	1.594	1.507	1.393	2.894	2.212	2.116	2.072	1.735	1.713	1.355
%V	16.64	12.37	10.60	9.53	8.94	8.74	8.26	13.04	12.26	9.84	9.78	8.39	7.93	7.33	15.23	11.64	11.14	10.91	9.13	9.02	7.13
C%	16.64	29.01	39.61	49.14	58.08	66.81	75.08	13.04	25.30	35.14	44.93	53.32	61.25	68.58	15.23	26.87	38.01	48.91	58.05	67.07	74.20

*Factor loadings above 0.5 were taken after the Varimax rotation is performed.

(VAR=Variables; VF=Varimax Factors; IE=Initial Eigenvalue; %V=Percentage of Variance; C%=Cumulative Percentage; Tur=Turbidity; DS=Dissolved Solid; TS=Total Solid; EC=Electrical Conductivity; Sal=Salinity; Tem=Temperature; DO=Dissolved Oxygen; BOD=Biochemical Oxygen Demand; COD=Chemical Oxygen Demand; TSS=Total Suspended Solid; PH=Acidity or Alkalinity; NH_3N=Ammoniacal Nitrogen; Ecoli=Escherichia Coliform; TColi=Total Coliform; As=Arsenic; Hg=Mercury; Cr=Chromium; Zn=Zinc; Fe=Iron.)

with dredging and channelization towards the hydrologic modification. Salinity is suspected to come from pesticide used in agricultural activities and animal livestock from farming activities. In VF3 and VF4, the results indicated that the total variance of 9.84% and 9.78% provide moderate positive loadings of DO and COD, as well as strong positive loadings of BOD and NH_3N, respectively. Generally, DO contamination comes from agricultural land, while COD contamination is related to the discharge of municipal and industrial waste, whereas BOD and NH_3N are originated from wastewater treatment plants, domestic wastewater, and industrial effluents. According to Rosli et al. (2015), the existence of BOD and NH_3N is contributed by pollution loading from the livestock activities (e.g. cow and goat farms) that have contributed 197.8 kg/day of BOD and 852.7 kg/day of NH_3N.

Meanwhile, VF 5 resulted to 8.39% of the total variance to provide a strong positive loading of pH and moderate positive loading of Escherichia coliform (*E. coli*). The existence of *E. coli* contamination in the river shows that the pollution is related to municipal wastes, oxidation ponds, and animal husbandry, which consumes a large amount of oxygen to undergo the anaerobic fermentation process to produce ammonia and organic acids. The hydrolysis process of acidic material could lead to a decrease in the pH river water values (Aris et al. 2013). Lastly, VF6 and VF7 indicated that the total variance of 7.93% and 7.33% produced a strong positive loading of Zn and Fe, as well as a moderate negative loading of Hg. In particular, Zn contamination is suspected to come from the houses and buildings that have zinc roofs, while Fe contamination is possibly generated from industrial activities such as electroplating, and Hg contamination is likely to relate to the plastic waste from chemical industries (Hua et al. 2016, Juahir et al. 2011, Papaioannou et al. 2010).

(3) Cluster 3 Region

C3 with a high pollution source region indicates that VF1 has a strong positive loading of TS, salinity and electrical conductivity, and a moderate positive loading of DS which resulted in 15.23% of the total variance. As explained previously, the existence of physical parameter contaminations is connected with the erosion of riverbanks due to dredging in the river, and the salinity pollution is suspected to have come from agricultural runoff. In VF2, the result indicates the total variance of 11.64% which produces strong positive loadings of *E. coli* and total coliform, and moderate negative loadings of pH and As. The pollution involved with *E. coli*, total coliform and pH are suspected to have come from municipal wastes, wastewater treatment plants and animal husbandry activities, that were carried out within the sub-basin (Gazzaz et al. 2012), together with the As pollution which results from agricultural activities. Meanwhile, VF3 has 11.14% of the total variance to produce a moderate positive loading of total suspended solid and moderate negative loading of turbidity, which was also detected to have originated from the hydrologic modifications such as dredging, water diversions, and channelization activities.

In VF4 and VF5, the result shows that the total number of variance is 10.91% and 9.13% to have factor loadings of DO, BOD, COD, NH_3N and temperature. This pollution is considered as chemical parameter contamination which is involved with anthropogenic activities which were suspected to have come from point-source pollution such as sewage treatment plants, domestic wastewater, industrial effluents, and municipal sewage (Juahir et al. 2011). Lastly, VF6 and VF7 have 9.02% and 7.13% of total variance to produce strong positive loadings of Hg and Zn, as well as moderate positive loading of Fe and moderate negative loading of Cr. Hg and Fe contaminations are suspected to have come from industrial wastes that are involved with chemical plastic waste and electroplating activities, while Zn contamination originated from villages using zinc roofs and Cr contamination is related with the urban storm runoff. In other words, the principal components 6 and 7 are subjected to the point source pollution that was discharged directly into the Malacca River.

Multiple Linear Regressions (MLR) of the Malacca River Water Quality Index (WQI)

Generally, MLR modeling is used to identify the variables behavior in the linear least-square fitting process and to determine possibly every trace element source (Hamid et al. 2016, Mustapha and Abdu 2012, Mustapha and Aris 2011). In this study, the source of apportionment of river water pollutant parameter is used to determine the potential of total water quality index (WQI). In other words, three models will be developed by using API value as a dependent variable, while the independent variable will be based on the water quality parameter from C1 (5 variables), C2 (6 variables), and C3 (6 variables).

A better coefficient result in the MLR model is dependent on the R^2, Adjusted R^2, and RMSE value, which is important to be used in C1, C2, and C3. The results of R2, Adjusted R2, and RMSE value in C1 are 0.867, 0.708, and 0.447, followed by C2 with 0.897, 0.774, and 0.393; as well as C3 with 0.930, 0.856, and 0.314, respectively. The equations of R^2, Adjusted R^2, and RMSE are shown in Eq. 5i to 5iii;

Cluster 1 (5 variables)
WQI = 3.261 + 392.9 (Total Dissolved Solid) + 644.0 (Electrical Conductivity) + 26.93 (Salinity) – 19.92 (Temperature) + 186.0 (Iron)

$$[R^2 = 0.867;\ \text{Adjusted } R^2 = 0.708;\ \text{RMSE} = .447] \quad (5i)$$

Cluster 2 (6 variables)
WQI = 4.358 + 20.06 (Turbidity) – 442.6 (Total Dissolved Solid) – 21.41 (Electrical Conductivity) + 434.4 (Salinity) + 23.92 (Temperature) + 217.3 (Zinc)

$$[R^2 = 0.897;\ \text{Adjusted } R^2 = 0.774;\ \text{RMSE} = .393] \quad (5ii)$$

Cluster 3 (6 variables)
WQI = 3.812 + 429.7 (Salinity) + 17.75 (Dissolved Oxygen) + 473.8 (Ammoniacal Nitrogen) + 19.23 (Total Coliform) + 33.01 (Iron) + 291.6 (Mercury)

$$[R^2 = 0.934;\ \text{Adjusted } R^2 = 0.856;\ \text{RMSE} = .314] \quad (5iii)$$

According to the result of the equation presented in 5i to 5iii, the highest coefficient of determination (R^2) is from C3 with 0.934 for salinity, dissolved oxygen, ammoniacal nitrogen, total coliform, iron, and mercury, followed by C2 with 0.897 for turbidity, salinity, temperature, zinc, as well as negative for total dissolved solid and electrical conductivity; and the lowest is C1 with 0.867 for total dissolved solid, electrical conductivity, salinity, iron, and negative for temperature. Based on the result, C3 has been selected as the best model due to the R^2 value is closest to 1 and smallest RMSE when compared with the other parameters. The main reason for this matter is because the model is performed with the RMSE value are smaller among the others and the R^2 value is closest to 1 (Mustapha and Abdu 2012, Mustapha and Aris 2011).

Meanwhile, Figure 4 shows the observed residual analysis and the predicted total WQI using C1, C2, and C3. The results indicate that the model of standard residual for C1, C2, and C3 have differences in the range of -2.1 to 1.9, -1.8 to 1.1, and -1.6 to 1.9, while the standard predicted value ranges between -1.9 to 1.4, -1.9 to 1.3, and -1.2 to 1.4, respectively. In other words, the results show the deficiency of the model for standardized residual and standardized predicted value in C1, C2, and C3. The main objective of the scatter plot diagram is to prove that C3 model is suitable to be used for the total WQI prediction, because the model provided results which have a great difference in the predicted WQI and the calculated WQI.

Conclusion

This study concluded that the spatial variation on the water quality pattern in the Malacca River Basin was successfully studied using the chemometric approach such as HCA, DA, PCA,

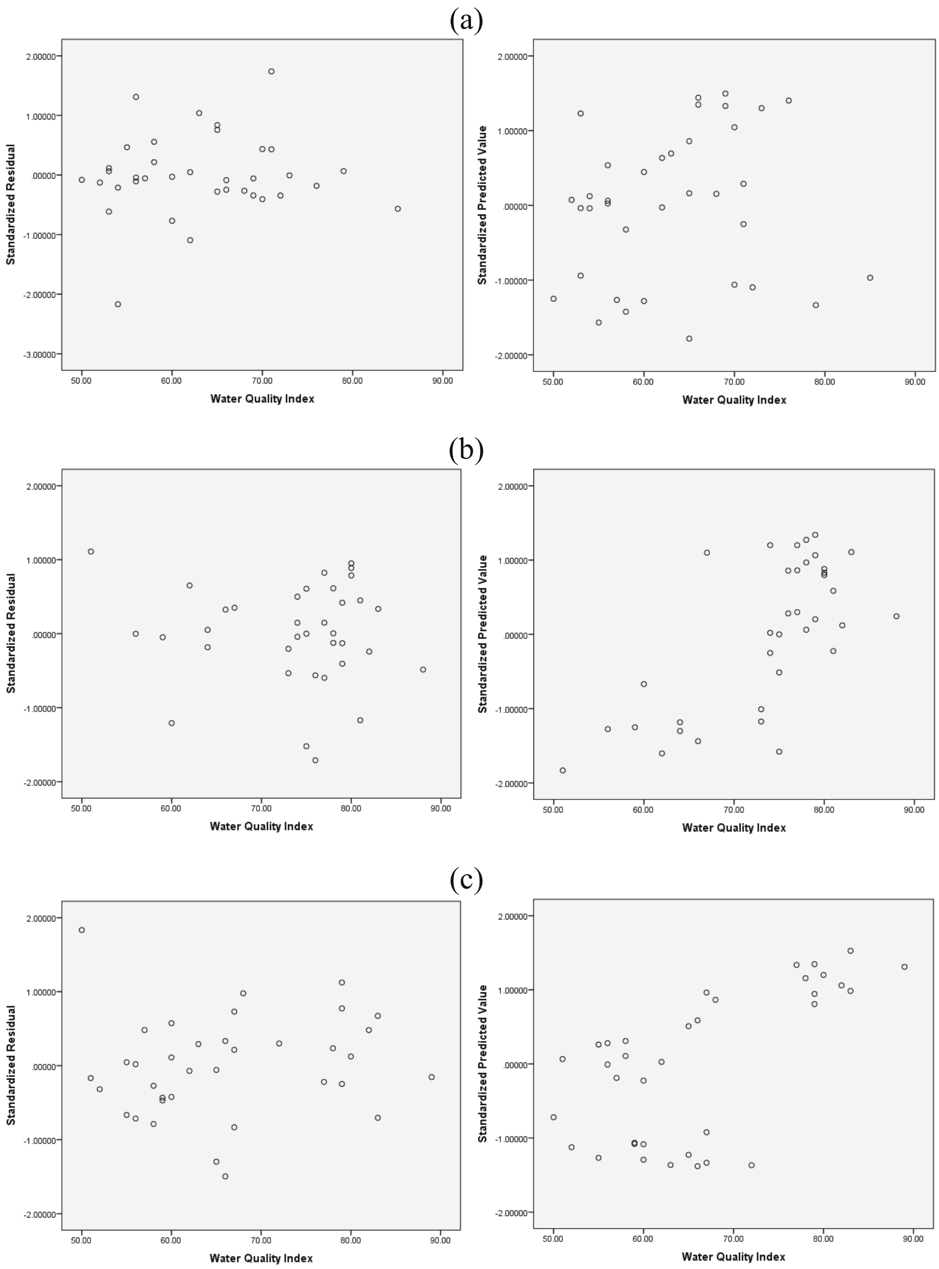

Fig. 4. Scatter plot diagram of standard residuals and standard predicted value for (a) Cluster 1, (b) Cluster 2, and (c) Cluster 3

and MLR. HCA successfully grouped the 9 sampling stations of 20 river water quality variables into three significant clusters, namely cluster 1 (C1), cluster 2 (C2), and cluster 3 (C3). In other words, HCA had benefited the monitoring network approach by reducing the quantity of monitoring stations in the Malacca River basin. Meanwhile, the cluster from HCA is applied into DA, which confirms that the standard mode, forward stepwise mode, and backward stepwise mode for the accuracy are 98.62% and 95.55% respectively. It also confirms the backward stepwise mode was selected with 9 variables of temperature, turbidity, salinity, BOD, COD, Arsenic, Iron, total coliform and Escherichia coliform. In PCA, seven CFs were detected in C1, C2, and C3 regions, with a total variance of 75.08%, 68.58%, and 74.20% respectively. The sources of variations detected in this study are residential activities, industrial activities, commercial activities, agriculture activities, animal livestock, as well as forest land. MLR analysis is carried out to determine the variability of proposed equation to predict the values of the total WQI. The resulted R^2 value is strong due to the high significance at p value with smaller than 0.05 when compared to the developed three models. The highest R^2 value is C3 with 0.934, followed by C2 with 0.897, and C1 with 0.867. The most suitable model to be used for total WQI prediction is the C3 model due to the result which has a great difference between the predicted WQI and the calculated WQI. Apart from identifying the pollution sources and understanding the variations of water quality data in the Malacca River basin, this study also suggests that the effectiveness of the river water quality can be managed by having a new water quality monitoring network that is required to be designed for more practical use which reduces time and effort, as well as cost saving purposes.

References

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), pp. 841–852.

American Public Health Association (APHA) (2005). Standard Methods for the Examination of Water and Wastewater. 21st ed. Washington: American Water Works Association, Water Environment Federation.

Aris, A.Z., Lim, W. Y., Praveena, S.M., Yusoff, M.K., Ramli, M.F. & Juahir, H. (2013). Water quality status of selected rivers in Kota Marudu, Sabah, Malaysia and its suitability for usage, *Sains Malaysiana*, 43(3), pp. 377–388.

Baharuddin, N., Nor'ashikin, S.A.I.M. & Zain, S.M. (2014). Characterization of spatial patterns in river water quality using chemometric techniques, *Sains Malaysiana*, 43(9), pp. 1355–1362.

Barbier, E.B., Hacker, S.D., Kennedy, C., Koch, E.W., Stier, A.C. & Silliman, B.R. (2011). The value of estuarine and coastal ecosystem services, *Ecological Monographs*, 81(2), pp. 169–193.

Bierman, P., Lewis, M., Ostendorf, B. & Tanner, J. (2011). A review of methods for analyzing spatial and temporal patterns in coastal water quality, *Ecological Indicators*, *11*(1), pp. 103–114.

Boyacioglu, H. & Boyacioglu, H. (2017). Application of environmetric methods to investigate control factors on water quality, *Archives of Environmental Protection*, 43(3), pp. 17–23.

Daneshmand, S., Huat, B.B., Moayedi, H. & Ali, T.A.M. (2011). Study on water quality parameters of linggi and melaka rivers catchments in Malaysia, *Engineering Journal*, 15(4), 41

Department of Environment Malaysia (DOE) (2012). *Malaysia Environmental Quality Report 2012*, Ministry of Natural Resources and Environment.

Gazzaz, N.M., Yusoff, M.K., Ramli, M.F., Aris, A.Z. & Juahir, H. (2012). Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques, *Marine Pollution Bulletin*, *64*(4), pp. 688–698.

Hamid, A., Bhat, S.A., Bhat, S.U. & Jehangir, A. (2016). Environmetric techniques in water quality assessment and monitoring: a case study, *Environmental Earth Sciences*, *75*(4), pp. 1–13.

Hua, A.K. (2017). Land use land cover changes in detection of water quality: A study based on remote sensing and multivariate statistics, *Journal of Environmental and Public Health*, 2017.

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), pp. 2411–2421.

Huang, Y., Zhang, D., Xu, Z., Yuan, S., Li, Y. & Wang, L. (2017). Effect of overlying water pH, dissolved oxygen and temperature on heavy metal release from river sediments under laboratory conditions, *Archives of Environmental Protection*, 43(2), pp. 28–36.

Juahir, H., Zain, S.M., Yusoff, M.K., Hanidza, T.T., Armi, A.M., Toriman, M.E. & Mokhtar, M. (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques, *Environmental Monitoring and Assessment*, 173 (1–4), pp. 625–641.

Kostecki, M., Tytła, M., Kernert, J. & Stahl, K. (2017). Temporal and spatial variability in concentrations of phosphorus species under thermal pollution conditions of a dam reservoir – the Rybnik Reservoir case study, *Archives of Environmental Protection*, *43*(3), pp. 42–52.

Li, L., Shen, X. & Jiang, M. (2017). Change characteristics of DSi and nutrition structure at the Yangtze River Estuary after Three Gorges Project impounding and their ecological effect, *Archives of Environmental Protection*, 43(2), pp. 74–79.

Lim, W.Y., Aris, A.Z. & Praveena, S.M. (2013). Application of the chemometric approach to evaluate the spatial variation of water chemistry and the identification of the sources of pollution in Langat River, Malaysia, *Arabian Journal of Geosciences*, 6(12), pp. 4891–4901.

Mustapha, A. & Abdu, A. (2012). Application of principal component analysis & multiple regression models in surface water quality assessment, *Journal of Environment and Earth Science*, 2(2), pp. 16–23.

Mustapha, A. & Aris, A.Z. (2011). Spatial aspect of surface water quality using chemometric Analysis, *Journal of Applied Sciences in Environmental Sanitation*, 6(4), pp. 411–426.

Papaioannou, A., Mavridou, A., Hadjichristodoulou, C., Papastergiou, P., Pappa, O., Dovriki, E. & Rigas, I. (2010). Application of multivariate statistical methods for groundwater physicochemical and biological quality assessment in the context of public health, *Environmental Monitoring and Assessment*, 170(1), pp. 87–97.

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), pp. 13–18.

Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. & Kouimtzis, T. (2003). Assessment of the surface water quality in Northern Greece, *Water Research*, 37(17), pp. 4119–4124.

UNESCO Official Portal (2007). Melaka and George Town, Historic Cities of the Straits of Malacca (http://whc.unesco.org/en/list/1223/(14.08.2018)).