



ANALYSIS OF SURFACE WATER POLLUTION IN THE KINTA RIVER USING MULTIVARIATE TECHNIQUE

(Penilaian Pencemaran Air Permukaan di Sungai Kinta Menggunakan Teknik Multivariat)

Hamza Ahmad Isiyaka and Hafizan Juahir*

East Coast Environmental Research Institute (ESERI)
University Sultan Zainal Abidin,
Gong Badak Campus, 21300 Kuala Terengganu, Terengganu. Malaysia.

*Corresponding author: hafizanjuahir@unisza.edu.my

Received: 14 April 2015; Accepted: 9 July 2015

Abstract

This study aims to investigate the spatial variation in the characteristics of water quality monitoring sites, identify the most significant parameters and the major possible sources of pollution, and apportion the source category in the Kinta River. 31 parameters collected from eight monitoring sites for eight years (2006-2013) were employed. The eight monitoring stations were spatially grouped into three independent clusters in a dendrogram. A drastic reduction in the number of monitored parameters from 31 to eight and nine significant parameters ($P < 0.05$) was achieved using the forward stepwise and backward stepwise discriminate analysis (DA). Principal component analysis (PCA) accounted for more than 76% in the total variance and attributes the source of pollution to anthropogenic and natural processes. The source apportionment using a combined multiple linear regression and principal component scores indicates that 41% of the total pollution load is from rock weathering and untreated waste water, 26% from waste discharge, 24% from surface runoff and 7% from faecal waste. This study proposes a reduction in the number of monitoring stations and parameters for a cost effective and time management in the monitoring processes and multivariate technique can provide a simple representation of complex and dynamic water quality characteristics.

Keywords: multivariate techniques, source apportionment, multiple linear regression, principal component analysis; Kinta river; water pollution

Abstrak

Kajian ini bertujuan untuk menyiasat variasi bagi ruang dalam mengenalpasti ciri-ciri di stesen pemantauan kualiti air, mengenalpasti parameter yang paling penting dan sumber utama kemungkinan terhadap pencemaran dan membahagi sumber kategori di Sungai Kinta. 31 parameter yang dikumpul dari lapan stesen pemantauan selama lapan tahun (2006-2013) telah digunapakai dalam kajian ini. Lapan stesen pengawasan telah dibahagikan kepada tiga kelompok bebas dalam bentuk *dendrogram*. Pengurangan mendadak bilangan parameter yang dipantau dari 31 parameter kepada lapan dan sembilan parameter penting ($P < 0.05$) telah dicapai dengan menggunakan kaedah langkah demi langkah ke hadapan dan langkah demi langkah ke belakang melalui analisis pembezaan (AP). Analisis komponen utama (AKU) menyumbang lebih daripada 76% dalam jumlah varians dan sifat-sifat punca pencemaran kepada proses antropogenik dan semula jadi. Pembahagian sumber menggunakan pelbagai regresi linear gabungan dan skor komponen utama menunjukkan bahawa 41% daripada jumlah beban pencemaran adalah daripada luluhawa batu dan air sisa yang tidak dirawat, 26% daripada pelepasan sisa, 24% daripada air larian permukaan dan 7% daripada sisa najis. Kajian ini mencadangkan pengurangan dalam bilangan stesen pemantauan dan parameter untuk pengurusan kos yang berkesan dan masa dalam proses pemantauan dan teknik multivariat boleh menyediakan perwakilan yang mudah untuk ciri-ciri kualiti air yang kompleks dan dinamik.

Kata kunci: teknik multivariate, sumber pembahagian; regresi linear, analisis komponen utama; Sungai Kinta; pencemaran air

Introduction

Harnessing the availability of qualitative water and minimizing its vulnerability of been exposed to pollution is a challenging task for most societies in the developing nations. Water security and conservation requires the provision and protection of an acceptable quality and quantity of clean water [1]. Water quality is a term used to classify the status of the chemical, physical and biological characteristics of water based on its uses [2]. Nevertheless, the degeneration of water quality affects the health status of local inhabitants, biodiversity and the ecosystem balance in general [3,4].

River basin serves as the major source of water resources for domestic, agricultural practices, efficient inland transport system and industrial use [5,6]. Even though it is a renewable resource, little effort has been channeled towards its conservation and pollution risk assessment [7]. A river basin is an interconnected system of main river course and its tributaries [5]. River water quality can easily be degraded by pollutants induced by anthropogenic and natural processes [8]. This is because most fresh waters are converted to the endpoint of effluent discharge from industrial and domestic sewage [9]. Whatever the case, the rapid economic development, change in land use pattern to industrial, agricultural practices and concentration of settlement along the river course makes river basin susceptible to pollution [10,11,12]. A report by The Environmental Protection Agency (EPA) indicates that about one-third of the global surface water are unsafe for numerous uses and pose a serious threat to the general wellbeing of man and his environment. This assertion has created awareness on ensuring global water quality [13]. Under the 8th and 9th Malaysia plan, several efforts have been made since 2001 to assess the rate at which water is been polluted [14].

However, an effective management and understanding of the chemical, biological and hydro-morphological pattern of river requires a fundamental application of robust statistical technique [15] for efficient pollution control and water quality management [16]. In order to understand the economic, environmental and social impact of water pollution, it is imperative to model the major possible sources and percentage contribution of water pollutants [17]. The application of pattern recognition and environmetric modelling techniques (multivariate statistics) is a sophisticated method for understanding water quality characteristics [18]. This technique have been applied by [4,8,9,18 – 28] to model different environmental issues. In this study, multivariate techniques such as CA, DA, PCA and MLR were used to model the complex and dynamic characterisation in the level of pollution in the Kinta River. CA was applied as an unsupervised pattern recognition technique in order to classify the water quality monitoring sites into interrelated groups. The most discriminating parameters with a drastic reduction in the number of significant variables were identified using DA. The sources of pollution were identified using PCA. The source category apportionment was modelled using MLR.

The objectives of this study are; to determine the spatial characteristics in the similarities of water quality monitoring processes in the Kinta River; to identify the most statistical significant parameters and possible sources of water pollution; and to apportion the mass source contribution for each source category within the study area.

Materials and Methods

Study Area

Kinta River is located in Perak which is the second largest state in Malaysia in terms of land mass (21,006 km²). Ipoh is the capital of Perak situated in the Kinta Valley [29]. The River is divided into upstream, middle stream and downstream with eight water quality monitoring sites covering the entire river network pattern. The upstream region is monitored by two stations (2PK22 and 2PK24), the middle course is assigned with three stations (2PK25, 2PK34, 2PK59) and the downstream of the river is stationed with three monitoring sites (2PK19, 2PK33, 2PK60). Kinta River is a primary source of water for the inhabitants in Ipoh and Perak (Malaysia). It has 8 tributaries and gains its source from Gunung Korbu in Ulu Kinta Perak and stretches to about 2420km². The capacity of the river was boosted by constructing a dam to supply about 639,000 m³ of water daily until 2020 [29]. The basic land use pattern within the catchment area ranges from residential, industrial, mining, oil and rubber palm plantations and logging [30]. The location of the study area is described in Table 1 and Figure 1.

Table 1. Location of the study area based on the distance, altitude and coordinates

Stations	Distance (KM)		Altitude	Coordinates	
	From source	Between stations		X	Y
			1599	518256.14	367777.98
2PK22	155	15.5	264	516816.52	352365.48
2PK24	224	6.9	50	512701.07	346869.20
2PK25	339	11.5	34	502746.06	341721.84
2PK59	375	3.6	28	499164.57	340907.99
2PK34	439	6.4	23	492824.41	338308.55
2PK33	576	13.7	11	479622.81	343084.89
2PK60	648	7.2	9	472475.77	342025.07
2PK19	77.1	12.3	7	462757.07	336011.65

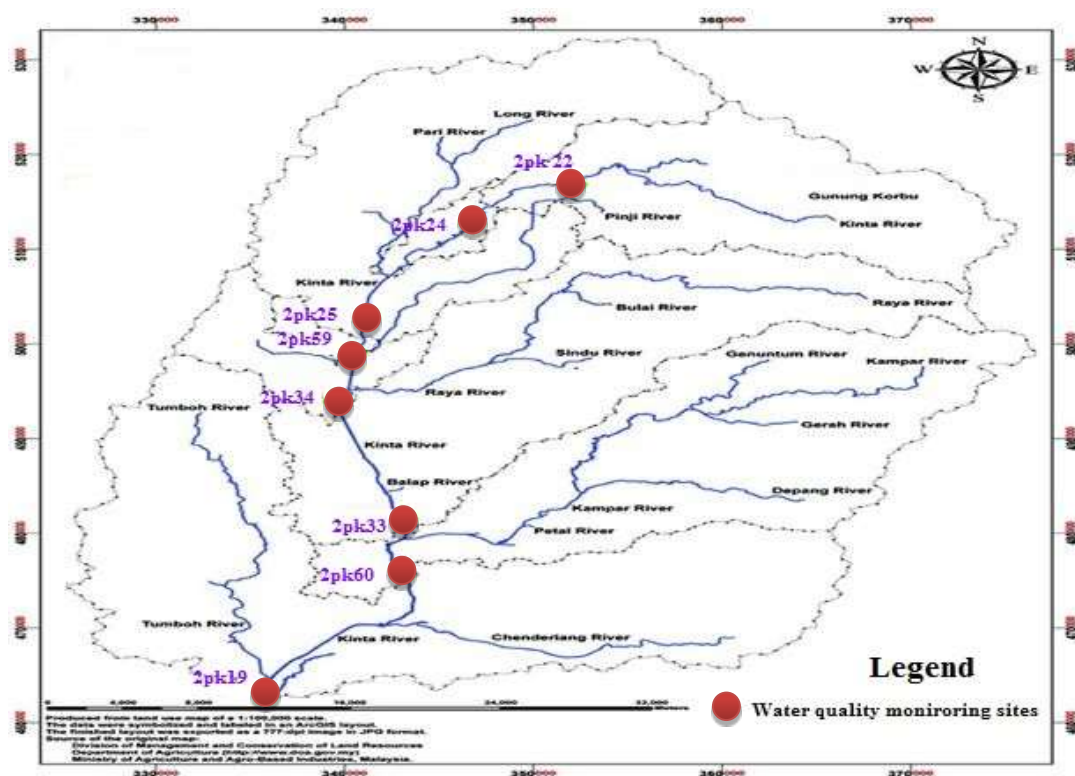


Figure 1. Description of the study area showing water quality monitoring sites in the Kinta River

Data Collection

Thirty one recent physicochemical water quality parameters monitored over a period of eight years (2006- 2013) were sourced from the Department of Environment Malaysia. The monitored parameters comprises of dissolved oxygen (DO); biochemical oxygen demand (BOD); chemical oxygen demand (COD); suspended solid (SS); pH; ammonia nitrogen ($\text{NH}_3\text{-NL}$); NH_4F , temperature ($^{\circ}\text{C}$), COND, salinity (SAL), turbidity (TUR), dissolved solid

(DS) , total solid (TS) nitrate nitrogen (NO₃-N), chloride (Cl), phosphate phosphorous (PO₄-P), arsenic (AS), mercury (Hg), cadmium (Cd), chromium (Cr), lead (Pb), Zinc (Zn), calcium (Ca), iron (Fe), potassium (K), magnesium (Mg), sodium (Na), oil and grease (OG), methylene blue active substances (MBAS), E-coli and Coliform.

Cluster Analysis (CA)

CA is an unsupervised pattern recognition technique used to aggregate a similar group of entities that portrays a strong internal (within-class) homogeneity and strong external (between classes) heterogeneity [18]. These spatial classes of water quality sampling sites can be illustrated using a dendrogram that measures the degree of risk homogeneity through Ward's method and Euclidean distance [31]. A dendrogram also highlights a reduction in the dimensionality and complexity of the data sets. CA has the ability to extract the pattern of surface water quality that can be used as a guide for future sampling [32].

Discriminant Analysis (DA)

DA is a supervised technique that can identify the most statistically significant parameters that best discriminate between natural occurring groups. It is used to construct new discriminate functions (DFs) from an existing group in order to explain the spatial characterization of surface water pollution [19,25]. DFs are calculated using equation 1:

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

where i = the number of group G ; k_i = constant inherent to each group; n = the number of parameters used to classify a set of data into a given group; w_j = the weight coefficient assigned by discriminant function analysis (DFA) to a given parameter P_j .

Principal Component Analysis (PCA)

PCA is a powerful pattern recognition technique used in identifying the major possible sources of pollution [18,33,34]. It can be used to transform original observed variables to new orthogonal, uncorrelated variables called principal components (PCs) preserved as a linear combination of the original variables [20]. PCA reduces the dimensionality of large data sets without loss of the original variable [35].

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

where z is the component score, a is 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.

Furthermore, the varimax rotation of the PCs generated by PCA with eigenvalues greater than one is imperative to simplify its complexity for proper interpretation [19]. The varimax rotation is used to obtain new set of variables called varimax factors (VFs) [36,37] using factor analysis method (FA). FA can extract variables with strong positive loading that can account for high variability. According to [38] the strength of a factor loading is classified into strong (> 0.75), moderate (0.75-0.50) and weak (0.50-0.30). FA is calculated using equation 3:

$$z_{ij} = a_{f1} f_{1i} + a_{f2} f_{2i} + \dots + a_{fm} f_{mi} + e_{fi} \quad (3)$$

where z is the measured value of a 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, j is the variable number, and m is the total number of factors.

Multiple Linear Regression Model (MLR)

MLR is a statistical technique that is used to predict the variability that exists between the dependent and independent variable [20,27,39,40]. Thus, the regress model can be represented as [41].

$$Y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad Eq \quad (4)$$

where $i = 1 \dots n$, β_0 , β_1 and β_k are regression coefficient, x_1 and x_k are independent variables and ε is error associated with the regression.

However, the percentage mass contribution of each identified possible source category was modelled using a combined principal component scores (PCS) after varimax rotation and MLR. The leave-one-out cross-validation technique was introduced to the model in order to examine the response in the network error if an input variable is removed at a time [42]. However, the deterioration in the network performance if one principal component score is eliminated determines the efficacy and significance of the variable in the overall network performance [43]. Invariably, the coefficient of determination “ R^2 ”, adjusted coefficient of determination “Adjusted R^2 ” and Root mean square error “RMSE” were used to test the goodness of fit in the model performance [39,40].

Results and Discussion

Descriptive statistics

The descriptive statistics of the observed parameters summarizes the entire data sets for a better in order to simplify the observations. However, Table 2 below comprises of the total number of observations in the data set, the minimum and maximum values, media, mean, variance and standard deviation.

Table 2. Summary statistics of the observed parameters from 2006-2013

Parameters	Observations	Minimum	Maximum	Mean	Std. deviation
DO	198	0.280	9.440	4.923	2.054
BOD	198	1.000	54.000	4.227	4.445
COD	198	3.000	108.000	23.682	14.781
SS	198	0.500	1340.000	100.841	141.249
pH	198	4.540	8.290	7.020	0.462
NH3-NL	198	0.005	3.940	0.343	0.591
NH4F	198	0.005	2.640	0.446	0.326
TEMP °C	198	23.480	32.930	27.894	1.821
COND	198	6.000	327.000	118.682	67.546
SAL	198	0.010	0.150	0.055	0.032
TUR	198	1.000	675.100	93.897	107.957
DS	198	11.000	180.000	60.364	33.371
TS	198	17.000	1402.000	161.882	143.920
NO ₃	198	0.005	5.580	0.584	0.612
Cl	198	0.500	25.000	6.306	4.834
PO4	198	0.005	0.400	0.058	0.069
As	198	0.001	0.049	0.010	0.008
Hg	198	0.000	0.000	0.000	0.000
Cd	198	0.001	0.001	0.001	0.000
Cr	198	0.001	0.023	0.003	0.005
Pb	198	0.005	0.005	0.005	0.000
Zn	198	0.005	0.090	0.022	0.019
Ca	198	0.380	37.200	12.587	7.739
Fe	198	0.005	2.050	0.194	0.242
K	198	0.050	13.400	3.152	1.997
Mg	198	0.050	7.500	1.821	1.035
Na	198	0.050	23.100	5.677	3.567
OG	198	0.500	2.000	0.540	0.184
MBAS	198	0.025	0.025	0.025	0.000
<i>E-coli</i>	198	96.000	500000.000	29371.281	54546.213
Coliform	198	4100.000	1700000.000	175869.697	248608.836

Spatial Classification in The Similarities of Water Quality Monitoring Sites

The eight water quality monitoring sites were unsupervisedly and spatially classified based on their level of similarities and differences in the characteristics of pollutant concentration in the study area. The resultant output gave a dendrogram comprising of three distinct independent clusters as displayed in Figure 2.

Cluster 1 is classified as a low pollution area (LPA) comprising of two clusters (2pk33 and 2pk22) located in both the upper course and lower course of the Kinta River. The two stations have a strong identity in terms of their similarities in the level of pollution. Cluster 2 comprises an integration of three water quality monitoring sites (2pk60, 2pk19 and 2pk24) with a strong similarities in their pollution level. This cluster represents a moderate pollution area (MPA). Two stations 2pk60 and 2pk19 are located in the lower course of the River while 2pk24 is situated at the upper course. The third cluster (2pk59, 2pk34 and 2pk25) is classified as the high pollution area (HPA).

However, from the dendrogram in Figure 2 it is crystal clear that one station in each cluster can give a relative spatial representation of the water quality monitoring processes of the Kinta River. This reveal the redundancy nature of some monitoring sites that in turn increase cost of equipment installation, cost of labour and time wastage. Based on the findings, it is advisable to reduce and relocate station 2pk33, 2pk60 and 2pk24 as well as 2pk25 and 2pk34 to other Rivers that have insufficient monitoring equipment since 2pk22, 2pk59 and 2pk19 in the three clusters represent the upper, middle and lower courses of the Kinta River. This action will not only reduce the overhead cost but also time management during the monitoring processes.

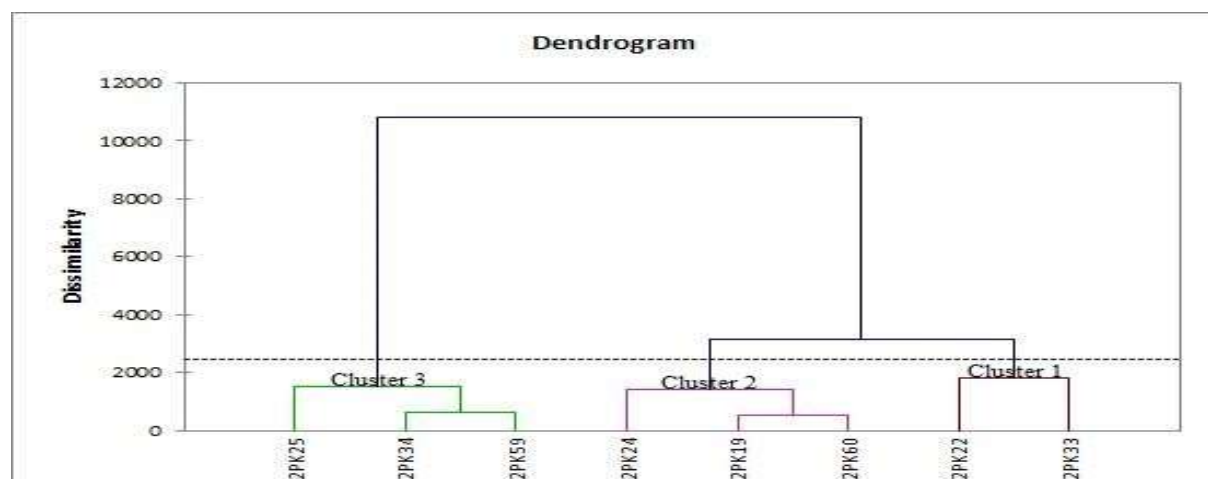


Figure 2. Dendrogram of spatial similarities in the monitoring sites

Spatial Identification of the Most Discriminating Parameters

The 36 water quality parameters monitored in the Kinta River represents a complex relationship that makes it tedious to pinpoint the most discriminating parameters affecting the water quality. Based on the clusters developed by HACA, the most significant parameters were determined using standard mode, forward stepwise and backward stepwise DA. To achieve this all the observed parameters were introduced as the independent variables while the three clusters were used as the dependent variables. The result for standard DA gave a reduction in the number of observations indicating that only 15 parameters ($P < 0.0001$) discriminate best with a strong assignation of 87.37% in the undimensional test of equality of the mean of class. Using the forward stepwise DA, only 8 parameters discriminate best with a P -value < 0.0001 and correct assignation of 84.34%. This indicates a significant decrease in the number of observed parameters. Backward stepwise DA gave a correct assignation of 86.87% showing that only 9 parameters discriminate best with a p -value < 0.0001 .

Identification of Major Pollution Sources in The Kinta River

The compositional pattern in the characteristics of the examined parameters and the major possible sources of pollution were identified using PCA/FA. Nine varimax factors rotated to an eigenvalue greater than one (>1) were obtained with a cumulative variance greater than 76% for the entire factors. In order to identify the source of pollution only factors with strong loading (> 0.75) capable of explaining the variation in the water quality pattern were selected for interpretation as highlighted in Table 3 and Figure 3. Furthermore, Figure 3b describes the scree plot diagram where the cut-off points for strong factors with eigenvalue greater than one are selected for interpretation.

Table 3. Identification of the most discriminating parameters based on clusters

Sampling Sites	Regions			
	Cluster 1	Cluster 2	Cluster 3	% correct
Standard mode 15 parameters				
Cluster 1	47	6	1	87.04%
Cluster 2	6	64	2	88.89%
Cluster 3	0	10	62	86.11%
Total	53	80	65	87.37%
Forward stepwise 8 parameters				
Cluster 1	47	6	1	87.04%
Cluster 2	8	60	4	83.33%
Cluster 3	0	12	60	83.33%
Total	55	78	65	84.34%
Backward stepwise 9 parameters				
Cluster 1	47	7	0	87.04%
Cluster 2	7	63	2	87.50%
Cluster 3	0	10	62	86.11%
Total	54	80	64	86.87%

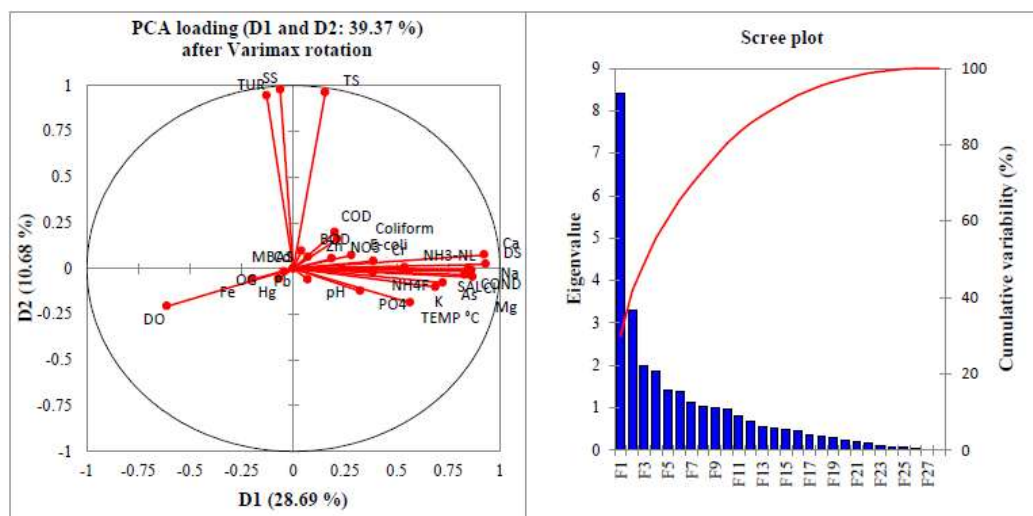


Figure 3. (a) PCA loading and (b) scree plot diagram after varimax rotation

Table 4. Factor loading after varimax rotation

Variables	VF1	VF2	VF3	VF4	VF5	VF6	VF7	VF8	VF9
DO	-0.612	-0.207	-0.224	0.004	0.227	0.226	-0.064	0.088	-0.166
BOD	0.039	0.095	0.173	0.017	0.090	-0.050	0.863	-0.097	0.110
COD	0.202	0.198	0.034	0.123	-0.221	0.094	0.816	0.014	-0.093
SS	-0.063	0.977	0.009	0.001	-0.026	-0.031	0.071	-0.016	-0.002
pH	0.070	-0.061	0.084	-0.085	0.014	0.869	-0.001	-0.019	0.113
NH3-NL	0.538	0.006	-0.166	-0.063	0.465	0.220	0.180	0.167	-0.103
NH4F	0.386	-0.022	-0.162	-0.059	0.405	-0.132	0.142	0.269	-0.153
TEMP °C	0.566	-0.188	0.031	-0.084	-0.137	-0.281	-0.123	-0.099	0.391
COND	0.846	-0.016	0.186	-0.103	0.035	-0.087	-0.001	0.077	0.217
SAL	0.848	-0.002	0.185	-0.119	0.050	-0.071	0.016	0.080	0.214
TUR	-0.129	0.944	0.107	-0.024	-0.054	0.014	0.047	-0.028	-0.006
DS	0.931	0.023	0.020	0.007	0.068	0.010	0.097	0.047	0.103
TS	0.154	0.962	0.018	0.013	0.004	-0.025	0.087	-0.005	0.030
NO3	0.185	0.053	-0.191	0.055	0.088	0.199	0.058	-0.020	0.772
Cl	0.829	-0.035	0.079	0.038	0.124	0.016	0.057	-0.031	-0.120
PO4	0.324	-0.124	0.117	0.110	0.708	0.126	-0.126	-0.078	0.165
As	0.724	-0.079	0.210	0.111	-0.075	-0.056	-0.034	0.096	0.125
Hg	-0.072	-0.058	0.035	-0.028	-0.032	0.004	-0.083	0.872	-0.012
Cd	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cr	0.387	0.040	-0.065	0.189	-0.593	0.452	0.173	0.095	0.030
Pb	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Zn	0.071	0.063	-0.130	0.778	-0.090	-0.078	0.212	0.173	0.267
Ca	0.925	0.075	0.055	-0.096	-0.006	0.068	0.024	-0.118	0.018
Fe	-0.197	-0.058	0.064	0.840	0.057	-0.006	-0.031	-0.172	-0.163
K	0.688	-0.102	0.179	0.220	0.048	0.169	0.215	-0.173	-0.260
Mg	0.864	-0.010	-0.044	-0.005	-0.090	0.076	0.024	-0.131	0.024
Na	0.868	-0.047	-0.046	-0.001	0.138	0.219	0.100	0.050	-0.171
OG	-0.046	-0.017	0.683	-0.045	-0.085	0.206	0.025	0.048	0.044
MBAS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
E-coli	0.282	0.072	0.832	0.007	0.056	0.092	0.127	0.007	-0.066
Coliform	0.213	0.159	0.775	-0.027	0.051	-0.140	0.109	-0.003	-0.146
Eigenvalue	8.396	3.314	1.988	1.849	1.427	1.374	1.130	1.033	1.004
Variability (%)	29.985	11.834	7.098	6.604	5.095	4.906	4.035	3.690	3.586
Cumulative (%)	29.985	41.819	48.917	55.521	60.617	65.523	69.558	73.248	76.835

The first varimax factor (VF1) explains 29.9% of the total variance with a strong positive loading for COND (0.846), SAL (0.848), DS (0.931), Cl (0.829), Ca (0.925), Mg (0.864) and Na (0.868). The composition of these factors is characterised with chemical parameters rich in mineral components from untreated wastewater discharge and weathering of existing parent rock [8,19]. Ca, Na, Mg affect the level of water hardness [19] thereby accelerating the cases of kidney stones and heart diseases [44]. COD SAL and DS allow high electrical flow due to high concentration of ions. Conductivity ions originate from dissolve salt and inorganic materials such as chloride, sulphides, alkalis and carbonate compounds. However, water with high salinity is likely to be conductive in nature. Cl encourages corrosion when in contact with metal ions thereby produce high concentration of metal in drinking water with a salty taste [45]. Ca occurs naturally in water since its primary origin is from parent rock. The high concentration (Ca) can lead to colorectal cancer, obesity, kidney stone, stroke and hypertension [46]. The strong loading for Mg can encourage low blood pressure, slow breathing, unconsciousness and death [47].

VF2 have strong positive loading for SS (0.977), TUR (0.944), and TS (0.962) that explains 11.8% of the total variance in the data sets. The composition pattern in the source of these pollutants originates from surface runoff in the solid form from field with high concentration of soil and waste disposal activities [19]. High TUR reflect the presence of silt, clay, organic matter and other microorganism [48] capable of altering the level of water clarity that limits light penetration as well as affect aquatic life. TUR can also encourage the existence of virus and bacteria parasites that can cause diarrhoea and headache [48].

VF3 have strong positive loading for *E-coli* (0.832) and coliform (0.775) explaining 7.1% of the total variance. The source of these pollutants is from faecal waste. VF4 exhibit a strong loading for Zn (0.778) and Fe (0.840) that accounts for 6.6% of the total variance in the datasets. These pollutants are likely to originate from industrial waste water discharge triggered by anthropogenic point source activities [49]. Metals originating from the discharged industrial waste and weathering of parent rocks are the major sources of Fe [48]. The presence of Zn in water within the permissible limit provides organisms the basic for physiological and metabolic processes. Its excesses affect human health [50].

VF6 represent a strong positive loading (0.869) for pH. High concentration of pH denotes the presence of alkaline that originates from erosion of river banks and weathering of existing parent materials rich in carbonates and bicarbonates limestone rocks. The presence of alkaline encourages the growth of algae, changes the water appearance to a greener colour [51,52]

VF7 accounts for 4% of the total variance with a strong positive loading for BOD (0.863) and COD (0.816). These factors represent organic pollutants from point sources such as waste water discharge from treatment plants, domestic waste water and discharge from industrial effluent [4]. VF8 (Hg, 0.872) represent pollution discharge from industrial and mining activities [53] explaining 3.7% of the total variance in the data sets.

VF9 have a strong positive loading for NO_3 that account for 3.6% of the total variance in the data sets. The source of this pollutant comes from agricultural related activities such as fertilizer application, pesticides and organic manure [54]. Application of nitrogen fertilizer by farmers to boost farm yield may undergo nitrification processes [49].

Source Apportionment using A Combined PCS/MLR

The mass contribution of each source category based on the leave-one-out cross-validation method is displayed in Figure 4. Each category comprises a combination of factors that explains the source apportionment of pollution in the study area.

Based on the findings, rock weathering and untreated waste water from point source and non-point sources contributes about 41% of the total pollution load in the Kinta River. The composition of these parameters affects the mineralogical content of the river. Point source pollution of waste discharged from industrial effluent, treatment plants and domestic sewage contributes 26% of the pollution load. The next most significant source apportioned accounted for 24% explaining an influence from natural processes triggered by surface runoff of soil and waste

materials. Faecal waste contributes 7% in the total pollution load, followed by 1% from erosion and weathering processes. The remaining category ranges from 0.1% to 0.2%.

This method reveals the most significant variables that need to be retained and those that are redundant during subsequent monitoring processes. It also provides a simple explanation in the source classification of pollution in the Kinta River that can be used as a base by government and other stakeholders to develop policies and strategies to limit pollutant discharge and improve water quality for numerous uses.

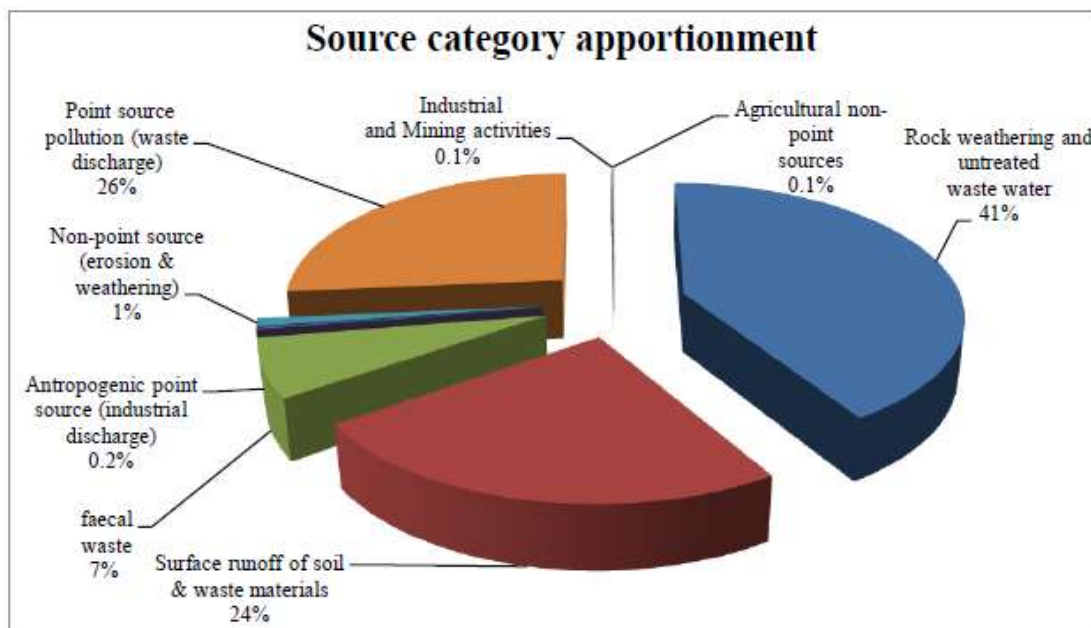


Figure 4. Pie chart for mass source contribution of water pollution

Conclusion

In this study, a reduction in the number of monitoring sites and observed parameters in the Kinta River were proposed based on the multivariate findings using HACA, DA, PCA and MLR. Spatial classification of the eight water quality monitoring station into three clusters was achieved using HACA. This classification reveals the level of similarities and redundancy in the characteristics of the monitoring processes. However, the findings for HACA indicates that three stations (2Pk22, 2Pk59 and 2Pk19) located in the upstream, middle and lower course can give a relative spatial representation of the water quality monitoring processes in the Kinta River. This will reduce cost and time of monitoring unnecessary sites. The 31 physicochemical parameters were also reduced to 8 most significant parameters with a P-value <0.0001 and a correct assignation of 84.34% using the forward stepwise DA. In the subsequent monitoring processes, it is suggested that these 8 parameters can provide a spatial representation of the water quality status in the Kinta River. This will also reduce the cost and time spent in observing the remaining 23 parameters. The result for PCA revealed that anthropogenic activities (industrial, treatment plants and domestic waste discharge, mining activities, fertilizer and pesticides) and natural processes (weathering of parent rock, surface runoff and erosion of river banks) constitutes the major possible sources of pollution in the study area. Furthermore, the mass source category apportionment was achieved using a combined MLR and principal component scores. In this regard, rock weathering and untreated waste water contributes 41% of the total pollution load, 26% by point source waste discharge, 24% by surface runoff, 7% faecal waste, erosion and weathering 1% and the remaining ranges from 0.1-0.2%. Even though water quality parameters exhibit a complex characteristic, multivariate technique can provide a simple explanation for a proper policy implementation by government and stakeholders involve in water quality management.

Acknowledgement

The authors wish to extend their sincere appreciation to the Department of Environment Malaysia (DOE) and the East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin Malaysia for their continuous support.

References

1. Mustapha, A., Aris, A. Z., Ramli, M. F. and Juahir, H. (2012). Spatial-temporal variation of surface water quality in the downstream region of the Jakara River, north-western Nigeria: A statistical approach. *Journal of Environmental Science and Health Part A* 47(11): 1551-1560.
2. Khalil, B., Ouarda, T. B. M. J. and St-Hilaire, A. (2011). Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *Journal of Hydrology* 405(3): 277-287.
3. Zhang, Q., Li, Z., Zeng, G., Li, J., Fang, Y., Yuan, Q. and Ye, F. (2009). Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: a case study of Xiangjiang watershed, China. *Environmental Monitoring and Assessment* 152(1-4): 123-131.
4. Satheeshkumar, P. and Khan, A. B. (2012). Identification of mangrove water quality by multivariate statistical analysis methods in Pondicherry coast, India. *Environmental Monitoring and Assessment* 184(6): 3761-3774.
5. Vieira, J., Fonseca, A., Vilar, V. J. P., Boaventura, R. A. R. and Botelho, C. M. S. (2012). Water quality in Lis River. Portugal. *Environmental Monitoring and Assessment* 184(12): 7125-7140.
6. Ogwueleka, T. C. (2015). Use of multivariate statistical techniques for the evaluation of temporal and spatial variations in water quality of the Kaduna River, Nigeria. *Environmental Monitoring and Assessment* 187(3): 1-17.
7. Usman, U. N., Toriman, M. E., Juahir, H., Abdullahi, M. G., Rabi, A. A. and Isiyaka, H. (2014). Assessment of Groundwater Quality Using Multivariate Statistical Techniques in Terengganu. *Science and Technology* 4(3): 42-49.
8. Mustapha, A., Aris, A. Z., Juahir, H., Ramli, M. F. and Kura, N. U. (2013). River water quality assessment using environmentric techniques: case study of Jakara River Basin. *Environmental Science and Pollution Research* 20(8): 5630-5644.
9. Wang, Y. B., Liu, C. W., Liao, P. Y. and Lee, J. J. (2014). Spatial pattern assessment of river water quality: implications of reducing the number of monitoring stations and chemical parameters. *Environmental Monitoring and Assessment*, 186(3): 1781-1792.
10. Mustapha, A., Aris, A. Z., Juahir, H., Ramli, M. F. and Kura, N. U. (2013). River water quality assessment using environmentric techniques: case study of Jakara River Basin. *Environmental Science and Pollution Research* 20(8): 5630-5644.
11. Zhang, Q., Li, Z., Zeng, G., Li, J., Fang, Y., Yuan, Q. and Ye, F. (2009). Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: a case study of Xiangjiang watershed, China. *Environmental Monitoring and Assessment* 152(1-4): 123-131.
12. Mustapha, A., Aris, A. Z., Ramli, M. F. and Juahir, H. (2012). Spatial-temporal variation of surface water quality in the downstream region of the Jakara River, north-western Nigeria: A statistical approach. *Journal of Environmental Science and Health Part A*, 47(11): 1551-1560.
13. Mustapha, A., Aris, A. Z., Yusoff, F. M., Zakaria, M. P., Ramli, M. F., Abdullah, A. M. and Narany, T. S. (2014). Statistical Approach in Determining the Spatial Changes of Surface Water Quality at the Upper Course of Kano River, Nigeria. *Water Quality, Exposure and Health* 6(3): 127-142.
14. Department of Environment Malaysia (DOE) (2009) Malaysia Environmental Quality Report, Ministry of Science, Technology and Environment, Kuala Lumpur.
15. Shrestha, S. and Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software* 22(4): 464-475.
16. Fan, X., Cui, B., Zhao, H., Zhang, Z. and Zhang, H. (2010). Assessment of river water quality in Pearl River Delta using multivariate statistical techniques. *Procedia Environmental Sciences* 2: 1220-1234.
17. Tobiszewski, M., Tsakovski, S., Simeonov, V. and Namieśnik, J. (2010). Surface water quality assessment by the use of combination of multivariate statistical classification and expert information. *Chemosphere* 80(7): 740-746.

18. Al-Odaini, N. A., Zakaria, M. P., Zali, M. A., Juahir, H., Yaziz, M. I. and Surif, S. (2012). Application of chemometrics in understanding the spatial distribution of human pharmaceuticals in surface water. *Environmental Monitoring and Assessment* 184(11): 6735-6748.
19. Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. T., Armi, A. M., Toriman, M. E. and Mokhtar, M. (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environmental Monitoring and Assessment* 173(1-4): 625-641.
20. Deepulal, P. M., Sujatha, C. H. and George, R. (2012). Chemometric study on the trace metal accumulation in the sediments of the Cochin Estuary—Southwest coast of India. *Environmental Monitoring and Assessment* 184(10): 6261-6279.
21. Wahid, N. B. A., Latif, M. T. and Suratman, S. (2013). Composition and source apportionment of surfactants in atmospheric aerosols of urban and semi-urban areas in Malaysia. *Chemosphere* 91(11): 1508-1516.
22. Aris, A. Z., Praveena, S. M., Isa, N. M., Lim, W.Y., Juahir, H., Yusoff, M. K. and Mustapha, A. (2013). Application of environmetric methods to surface water quality assessment of Langkawi Geopark (Malaysia). *Environmental Forensics* 14(3): 230–239.
23. Oyeyiola, A. O., Davidson, C. M., Olayinka, K. O., Oluseyi, T. O. and Alo, B. I. (2013). Multivariate analysis of potentially toxic metals in sediments of a tropical coastal lagoon. *Environmental Monitoring and Assessment* 185(3): 2167-2177.
24. Isiyaka, H., Juahir, H., Toriman, M. E., Gasim, B. M., Azid, A., Amri, M. K., Ibrahim, A., Usman, U. N., Rano, A. R. A. and Garba, M. A. (2014). Spatial Assessment of Air Pollution Index Using Environ Metric Modeling Techniques. *Advances in Environmental Biology* 8(24): 244-256.
25. Pati, S., Dash, M. K., Mukherjee, C. K., Dash, B. and Pokhrel, S. (2014). Assessment of water quality using multivariate statistical techniques in the coastal region of Visakhapatnam, India. *Environmental Monitoring and Assessment* 186(10): 6385-6402.
26. Latif, M. T., Dominick, D., Ahamad, F., Khan, M. F., Juneng, L., Hamzah, F. M. and Nadzir, M. S. M. (2014). Long term assessment of air quality from a background station on the Malaysian Peninsula. *Science of The Total Environment* 482: 336-348.
27. Mustaffa, N. I. H., Latif, M. T., Ali, M. M. and Khan, M. F. (2014). Source apportionment of surfactants in marine aerosols at different locations along the Malacca Straits. *Environmental Science and Pollution Research* 21(10): 6590-6602.
28. Jaafar, S. A., Latif, M. T., Chian, C. W., Han, W. S., Wahid, N. B. A., Razak, I. S., Khan, M. F. and Tahir, N. M. (2014). Surfactants in the sea-surface microlayer and atmospheric aerosol around the southern region of Peninsular Malaysia. *Marine Pollution Bulletin* 84(1): 35-43.
29. Ghani, A., Zakaria, N. A., Kiat, C.C., Ariffin, J., Abu Hasan, Z., Abdul Gaffar, A. B., 2007. Revised equations for manning's coefficient for sand-bed rivers. *International Journal of River Basin Management* 5 (4): 329–346.
30. Zali, M. A., Retnam, A., Juahir, H., Zain, S. M., Kasim, M. F., Abdullah, B. and Saadudin, S. B. (2011). Sensitivity analysis for water quality index (WQI) prediction for Kinta River, Malaysia. *World Applied Sciences Journal*: 60 – 65.
31. Lau, J., Hung, W.T. and Cheung, C.S. (2009). Interpretation of air quality in relation to monitoring station's surrounding. *Atmospheric Environmetric* 43: 769-777.
32. Singh, K. P., Malik, A., Mohan, D. and Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water Research* 38(18): 3980-3992.
33. Farmaki, E. G., Thomaidis, N. S., Simeonov, V. and Efstathiou, C. E. (2012). A comparative chemometric study for water quality expertise of the Athenian water reservoirs. *Environmental Monitoring and Assessment* 184(12): 7635-7652.
34. Azid, A., Juahir, H., Toriman, M. E., Kamarudin, M. K. A., Saudi, A. S. M., Hasnam, C. N. C. and Yamin, M. (2014). Prediction of the Level of Air Pollution Using Principal Component Analysis and Artificial Neural Network Techniques: a Case Study in Malaysia. *Water, Air, & Soil Pollution* 225(8): 2063 – 2077.
35. Zhang, Q., Li, Z., Zeng, G., Li, J., Fang, Y., Yuan, Q. and Ye, F. (2009). Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: a case study of Xiangjiang watershed, China. *Environmental Monitoring and Assessment* 152(1-4): 123-131.

36. Brūmelis, G., Lapiņa, L., Nikodemus, O. and Tabors, G. (2000). Use of an artificial model of monitoring data to aid interpretation of principal component analysis. *Environmental Modelling & Software* 15(8): 755-763.
37. Love, D., Hallbauer, D., Amos, A. and Hranova, R. (2004). Factor analysis as a tool in groundwater quality management: two southern African case studies. *Physics and Chemistry of the Earth, Parts A/B/C* 29(15): 1135-1143.
38. Liu, C.W., Lin, K. H. and Kuo, Y. M. (2003). Application of factor analysis in the assessment of groundwater quality in a black foot disease area in Taiwan. *Science of Total Environment* 313: 77–89.
39. Chatterjee, S. and Price, B. (1999) Regression Analysis by Example. third ed. Wiley, Chichester.
40. Petrie, A. and Sabin, C. (2000). Medical Statistics at a Glance. Blackwell Science, Oxford.
41. Kovač-Andrić, E., Brana, J. and Gvozdić, V. (2009). Impact of meteorological factors on ozone concentrations modelled by time series analysis and multivariate statistical methods. *Ecological Informatics* 4(2): 117-122.
42. Gazzaz, N. M., Yusoff, M. K., Aris, A. Z., Juahir, H. and Ramli, M. F. (2012). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollution Bulletin* 64(11): 2409-2420.
43. Pastor-Bárcenas, O., Soria-Olivas, E., Martín-Guerrero, J. D., Camps-Valls, G., Carrasco-Rodríguez, J. L. and del Valle-Tascón, S. (2005). Unbiased sensitivity analysis and pruning techniques in neural networks for surface ozone modelling. *Ecological Modelling* 182(2): 149-158.
44. Napacho, Z. A. and Manyele, S. V. (2010). Quality assessment of drinking water in Temeke District (part II): Characterization of chemical parameters. *African Journal of Environmental Science and Technology* 4(11): 775-789.
45. Kanmani, S. and Gandhimathi, R. (2013). Investigation of physicochemical characteristics and heavy metal distribution profile in groundwater system around the open dump site. *Applied Water Science* 3(2): 387-399.
46. WHO. (2011). Hardness in drinking water: background document for preparation of WHO guidelines for drinking water quality. World Health Organization, Geneva.
47. Busse, M. (2013). *Sign and symptoms of too much magnesium*. [http:// www.livestrong.com/article/379016-signs-and-symptoms-of-toomuch- magnesium](http://www.livestrong.com/article/379016-signs-and-symptoms-of-toomuch-magnesium)
48. Seth, R., Mohan, M., Dobhal, R., Gupta, V. K., Singh, P., Singh, R. and Gupta, S. (2014). Application of Chemometric Techniques in the Assessment of Groundwater Quality of Udham Singh Nagar, Uttarakhand, India. *Water Quality, Exposure and Health* 6(4): 199-216.
49. Chen, H., Teng, Y., Yue, W. and Song, L. (2013). Characterization and source apportionment of water pollution in Jinjiang River, China. *Environmental Monitoring and Assessment* 185(11): 9639-9650.
50. Myers, S. A., Nield, A. and Myers, M. (2012). Zinc transporters, mechanisms of action and therapeutic utility: implications for type 2 diabetes mellitus. *Journal of Nutrition and Metabolism*: 1-14.
51. Minnesota Pollution Control Agency. (2007). Phosphorus: Sources, Forms, Impact on Water Quality
52. Kumar, A., Bisht, B. S., Joshi, V. D., Singh, A. K. and Talwar, A. (2010). Physical, chemical and bacteriological study of water from rivers of Uttarakhand. *Journal of Human Ecology* 32(3): 169-173.
53. Yang, L., Linyu, X. U. and Shun, L. (2009). Water quality analysis of the Songhua River Basin using multivariate techniques. *Journal of Water Resource and Protection* 1(02): 110-121.
54. McFarland, A. M., & Hauck, S. L. (1999). Relating agricultural land uses to in-stream stormwater quality. *Journal of Environmental Quality* 28(2): 836–844.