Journal of Environmental Science and Technology

ISSN 1994-7887

science alert

ANSInet
an open access publisher
http://ansinet.com

OPEN ACCESS Journal of Environmental Science and Technology

ISSN 1994-7887
DOI: 10.3923/jest.2016.277.284

Research Article
Multivariate Chemometric Approach on the Surface Water Quality in Langat Upstream Tributaries, Peninsular Malaysia

[1,4]Mohammad Zahirul Haque, [1]Sahibin Abd Rahim, [3]Md. Pauzi Abdullah, [5]Ahmad Fuad Embi, [2]Rahmah Elfithri, [1]Tukimat Lihan, [3]W.M.A. Wan Mohd Khalik, [6]Md Firoz Khan and [2]Mazlin Mokhtar

[1]School of Environmental and Natural Resources Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600 Selangor, Malaysia
[2]Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, Bangi 43600 Selangor, Malaysia
[3]School of Chemical Sciences and Food Technology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600 Selangor, Malaysia
[4]Department of Bangladesh Forest, Bono Bhaban, Agargaon, Dhaka 1207, Bangladesh
[5]Department of Hydrology, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi 43600 Selangor, Malaysia
[6]Centre for Tropical Climate Change System, Institute of Climate Change, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia

Abstract
The small tributaries to upstream Langat of Peninsular Malaysia play an important role to water quality in downstream. This study was carried out to investigate the indicator pollution and identify the potential sources of pollutants using multivariate chemometric techniques. Sampling campaign was conducted on monthly basis from January-June, 2015, duly interval dry and rainy seasons at six stations. Hierarchical cluster analysis (HACA) was employed on temporal and spatial dataset. Temporal dataset were grouped into two clusters on the basis of rainfall before collecting samples; the months of January, March and June formed one cluster and February, April and May appeared in the other. Spatial dataset were grouped into three clusters namely less polluted, medium polluted and polluted sites. Factor Analysis (FA) and Principal Component Analysis (PCA) were applied to identify the significant sources of pollutants, which resulted in five latent factors amounting to 73.0% of the total variance in data sets. Varifactors obtained from factor analysis indicate that the parameters responsible for water quality variations are related to physicochemical parameters and nutrients from both nonpoint and point sources. The nonpoint sources include plantation area, weathering of sedimentary rock and natural vegetation and point sources include mainly domestic wastewater. Thus, this study illustrates the water quality assessment, identification of pollution factors and temporal/spatial variations in water quality for the surface water of upstream tributaries to implement effective river water quality management with multivariate statistical techniques for analysis and interpretation of complex data sets.

Key words: Physicochemical water quality, cluster analysis, factor analysis, principal component analysis

Received: December 10, 2015 Accepted: January 20, 2016 Published: April 15, 2016

Citation: Mohammad Zahirul Haque, Sahibin Abd Rahim, Md. Pauzi Abdullah, Ahmad Fuad Embi, Rahmah Elfithri, Tukimat Lihan, W.M.A. Wan Mohd Khalik, Md Firoz Khan and Mazlin Mokhtar, 2016. Multivariate chemometric approach on the surface water quality in langat upstream tributaries, peninsular Malaysia. J. Environ. Sci. Technol., 9: 277-284.

Corresponding Author: Md. Pauzi Abdullah, School of Chemical Sciences and Food Technology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600 Selangor, Malaysia Tel: +603-89215447 Fax: +603-8921 5410

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Primarily natural processes as well as anthropogenic activities are the inducement for surface water pollution. Among others precipitation intensity, weathering processes and sediment transport are the major natural process for surface water quality deterioration. On the other hand, anthropogenic activities that denote landscape changes through urban development and industrial and tillage practices often cause degradation of water quality, physical habitat and biological integrity of lotic system (Davis *et al*., 2003; Casatti *et al*., 2006). Continuous degradation of water resources in the upstream catchment is accountable for the pollution of mainstream (Shields *et al*., 2008). The upstream assimilates and transports the effluent of domestic and industrial wastewater and runoff from the watershed downward (Varol and Sen, 2009). This anthropogenic discharge constitutes the constant polluting source, however, the surface runoff is a seasonal phenomenon in the basin mainly affected by climatic factors. Many factors are susceptible to the spatial and temporal variation of surface water characteristics. Generally rivers and tributaries are highly heterogeneous in various spatial scales (Fausch *et al*., 2002). This heterogeneity is developed due to micro environmental conditions, which impacts physical characteristics, while scale of temporal variability of surface water varies by the function of river or tributary, which mainly depends on the chemical parameters of interest (Poff and Ward, 1990). This variation may also takes place by some hydrologic inputs that originate from precipitation, direct overland flow, subsurface flow through shallow soils, drainage from aquifers and stream processes. Stream processes include dilution, metal release and adsorption from sediments (Devito *et al*., 1996).

Water resources are massively affected by the gradual increase of population. They release contaminants in rivers and tributaries cause massive degradation of fresh-water wetland (Cespedes *et al*., 2006).

Analyzing and monitoring of water quality by interpreting temporal and spatial variations in the basin is an essential task to control water pollution (Ning and Chang, 2004).

The application of different multivariate statistical techniques, such as Cluster Analysis (CA), Principal Component Analysis (PCA) and Factor Analysis (FA), help in the interpretation of complex data matrices of the water quality in easy way (Singh *et al*., 2004). It also helps in identifying of possible and significant sources of pollution that affects water quality and suggests for a valuable tool of appropriate water resources management and a solution for pollution problems as well (Alberto *et al*., 2001). Multivariate statistical techniques have been proved useful in analyzing temporal and spatial variations and therefore are applied to characterize and evaluate surface water quality (Simeonov *et al*., 2003).

In the present study, data sets obtained during six months of monitoring program were analyzed by several multivariate statistical techniques (PCA, FA and CA) to evaluate the sources of pollution in the surface water of three tributaries in Langat sub basin. This help in reducing the frequency of temporal and spatial monitoring numbers physicochemical characters and quick mitigating measures for pollution controlling.

MATERIALS AND METHODS

Description of the study area: The study area comprises three tributaries namely Tasik Kejuruteraan, Alur Ilmu and Puripujanga in the Universiti Kebangsaan Malaysia (UKM) catchment, the upstream basin of Langat river. The lengths of these tributaries measured using ArcGIS 10.2 tools are of 1596.60, 2098.75 and 1770.20 m, respectively (Fig. 1) and geographically bounded by 2°55′22″-2°55′30″N latitudes and 101°46′18″-101°46′23″E longitude. Six sampling stations (S1, S2, S3, S4, S5 and S6) were monitored in the tributaries (Table 1).

Minimum and maximum altitudes of the basin are 20 and 108 m, while the mean is the 56 ± 19 m above the mean sea level. The average annual precipitation recorded from the UKM rain station is 2197.20 m. The land use of UKM catchment comprises crop land 12.90%, forest land 38.49%, grassland 10.89%, mangroves 0.29%, wetlands 1.66%, settlements 34.74% and others 1.04%. The average sediment loaded in the water from basin is approximately 3600 t year[1].

Table 1: Geographical description of the study area

		Coordinates			
Station	Tributary	Latitude	Longitude	Altitude (M)	Description
S 1	Tasik Kejuruteraan	$2^0$55′26.88″N	$101^0$46′17.35″E	31.40	UKM Lake
S2		$2^0$55′26.88″N	$101^0$46′16.43″E	46.04	Anthropogenic source
S3		$2^0$55′26.88″N	$101^0$46′14.99″E	66.46	Vegetation cover
S4	Alur Ilmu	$2^0$55′26.88″N	$101^0$46′38.30″E	34.45	Out-let
S5	Puripujanga	$2^0$55′26.88″N	$101^0$46′59.64″E	35.37	Water during dry season
S6	Alur Ilmu	$2^0$55′26.88″N	$101^0$46′53.51″E	44.51	University cafeteria

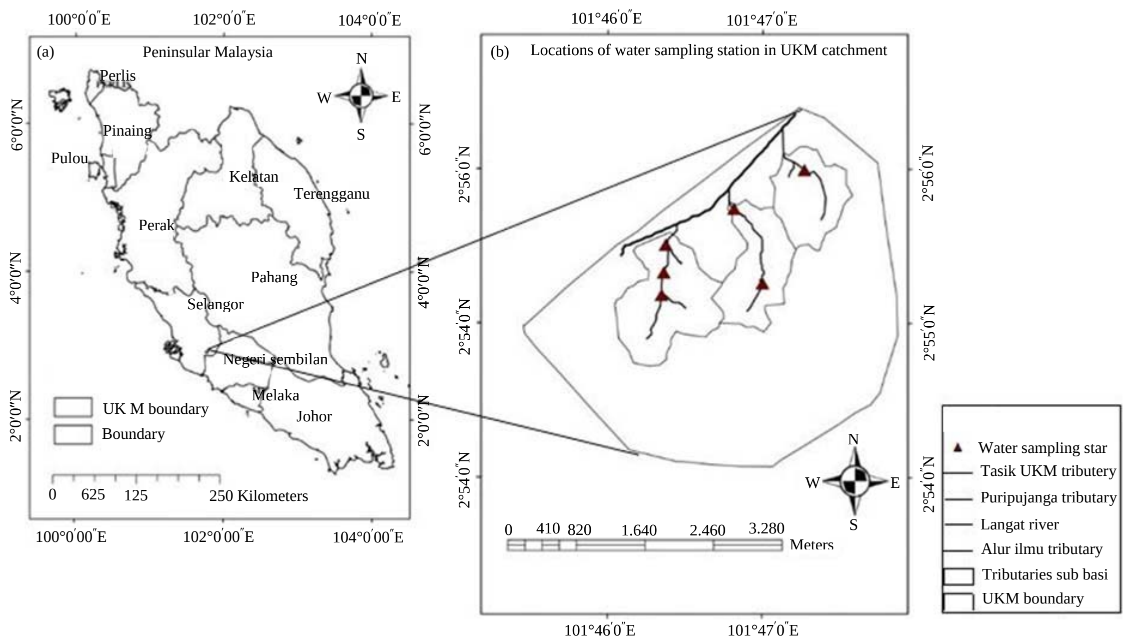

Fig. 1: (a-b) Location map of the study area

Variables and analytical procedures: Thirteen physicochemical variables were selected to monitor from six sampling stations. These were temperature (Temp), pH, Electrical Conductivity (EC), salinity, Dissolved Oxygen (DO), Turbidity, Total Dissolved Solids (TDS), Total Suspended Solids (TSS), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), ammoniacal nitrogen (NH_3-N), nitrate nitrogen (NO_3-N) and phosphate (PO_4-3). The monitoring was done monthly from January-June, 2015 for each station. All surface water samples were collected at a 0.5 m depth by using 1 L pre-cleaned polyethylene and glass bottles. Samples were then placed in ice filled cool box prior to transfer to ALIR, UKM laboratory.

Temperature, pH, EC, DO, TDS and salinity were measured *in situ* using a multi-probe sensor instrument (YSI 550, USA). Other parameters were analyzed in the laboratory using standard methods namely BOD (5 days incubation), COD (reactor digestion and colorimetric determination), TSS (gravimetric method), turbidity, (Absorptometric method), NH_3-N (salicylate method), NO_3-N (Cadmium reduction method and PO_4-3 (ascorbic acid method).

Water Quality Index (WQI) was determined by Malaysia Department of Environment formula based on six parameters as given by the following equation (Abidin *et al*., 2015):

$$\text{WQI} = 0.22\times\text{SIDO}+0.19 \text{ X SIBOD}+0.16\times\text{SICOD}+ 0.15\times\text{SIAN}+ 0.16\times\text{SISS}+0.12\times\text{SipH}$$

where, SI is the sub-index of each parameter. WQI was then used to classify river segment based on National Water Quality Standard, (NWQS) Malaysia, which categorized water quality into five classes namely, class I (WQI>92.7), class II (WQI 76.6-92.7), class III (WQI 51.9-76.5), class IV (WQI 31.0-51.9) and class V (WQI<31.0) based on beneficial use of the water.

Multivariate statistical techniques: At the beginning the dataset was log normally distributed using the Kolmogorov-Smirnov (K-S) statistics (Lampariello, 2000). According to the K-S test, all the variables were log normally distributed with 95% or higher con dence excepting turbidity, TSS NO3-N and PO_4^{-3}, which were found non normal with the same confidence level. Similarly, to examine the suitability of the data for principal component analysis and factor analysis, Kaiser-Meyer-Olkin (KMO) and Bartlett's Sphericity tests were performed (Chen *et al*., 2003; Shrestha and Kazama, 2007). High value (close to 1) generally indicates that principal component/factor analysis may be useful. However, in this study KMO value was 0.56. The signi cance level ($p<0.05$) indicates that there are signi cant relationships among variables. Thus, the dataset is fully suitable to be feed with PCA and FA analysis. Prior to the PCA and FA procedures, the dataset was subjected to normalization with the following formula:

$$N = \frac{CV - MC}{SD}$$

where, N is normalized data, CV the concentration of the each variable, MC the mean concentration of each variable and SD is the standard deviation of the data variable.

Cluster analysis: Cluster analysis states a group of multivariate techniques, whose initial aim is to accumulate objects based on their characteristics. Cluster analysis classi es the objects in such a way that each object is similar to the others in that cluster with respect to a predetermined selection criterion. The most common approach is Hierarchical Agglomerative Clustering Analysis (HACA), which provides intuitive similarity relationships between any one sample and the entire data set and is typically illustrated by a dendrogram (Ketchen and Shook, 1996). In this study, HACA was performed on the normalized data set by means of the Ward's method, using Euclidean distances as a measure of similarity.

Principal Component Analysis (PCA) and Factor Analysis (FA): The PCA is designed to transform the original variables into new and uncorrelated variables between the component factors but highly correlated to one factor (axes), called the principal components, which are linear combinations of the original variables. The factors explain the maximum of the variance of the dataset. Mathematically PCA can be defined with the following equation:

$$z_{if} = \alpha_{1i} x_{1j} + \alpha_{2i} x_{2j} + \dots \dots \dots \dots \dots \dots + \alpha_{ni} x_{nj}$$

where, z is component score, a is component loading, x is the measured value of the variable, i is the component number, j is the sample number and n is the total number of variables. The main purpose of FA is to reduce the contribution of less signi cant variables to simplify even more of the data structure coming from PCA. Principal Component (PC) is a linear combination of observable water quality variables, whereas, varifactors (VF) can include unobservable, hypothetical, latent variables (Wang *et al*., 2013).

RESULTS

The summary statistics of the six months data set on tributaries water quality were summarized in Table 2. The variables such as temperature, pH, TSS, turbidity and BOD were recorded the highest level of mean at station 1 and pH, EC, salinity, DO and TDS were recorded the lowest was at station 6.

Cluster analysis on the basis of temporal data: Figure 2 demonstrates the results of cluster analysis using the temporal data. This analysis produced two clusters. Cluster 1 was separated by the months of January, March and June, where data collection was performed in the dry weather. Cluster 2 included the months February, April and May, that closely corresponded to the sample collection after rainfall.

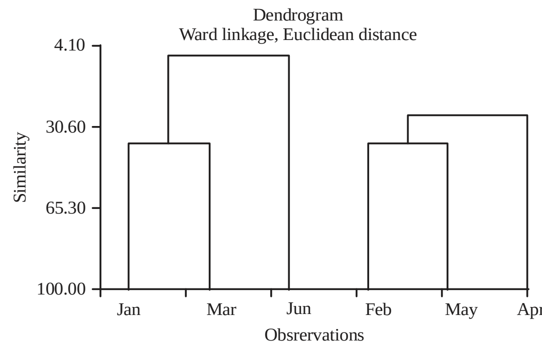

Fig. 2: Dendrogram showing temporal cluster

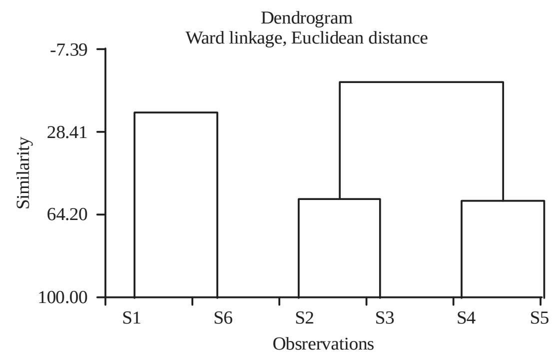

Fig. 3: Dendrogram showing spatial clustering of six stations

Cluster analysis on the basis of spatial data: Figure 3 represents the output of the CA using spatial data. The CA using spatial dataset generated three groups. Group 1, 2 and 3 consisted of stations S1 and S6, S2 and S3 and S4 and S5 respectively.

Source apportionment of pollutant related to water quality: Principal component analysis was performed using the water quality dataset. The dataset was normalized before feed into PCA method. The PCA produces five factor components, which the eigenvalues greater than 1 and explain 73% of the total variability. Eigenvalues of 1.0 or more then that are consi dered signi cant as suggested by Yidana *et al.* (2008). The Scree plot identified the number of PCs to be retained in order to comprehend the underlying data structure (Helena *et al*., 2000).

Table 2: Mean, standard deviation and maximum and minimum values of water quality parameters

	S1				S2				S3			
Variable	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Temp	30.03	1.95	27.43	32.91	28.83	1.15	27.55	31.01	28.35	0.57	27.78	29.42
pH	7.31	0.44	6.91	7.83	7.15	0.15	7.00	7.38	7.12	0.10	6.95	7.22
EC	135.60	30.19	117.00	188.00	185.67	45.73	136.00	244.00	191.67	49.79	122.00	251.00
DO	3.03	3.16	1.08	8.62	3.94	2.53	0.40	6.54	3.37	2.69	0.14	6.99
Turb	25.43	26.00	0.00	69.00	10.81	7.99	4.06	22.53	6.65	2.69	3.80	10.03
Salinity	0.07	0.02	0.05	0.09	0.09	0.02	0.06	0.11	0.09	0.02	0.06	0.12
TDS	96.00	21.13	77.00	122.00	125.50	42.50	80.00	195.00	124.67	32.34	79.00	163.00
TSS	65.20	29.20	39.00	106.00	25.33	10.60	12.00	37.00	22.00	7.16	8.00	28.00
COD	15.33	2.88	12.67	19.33	16.50	6.19	10.00	25.33	20.28	8.60	12.33	33.00
BOD	2.40	1.40	0.50	3.70	1.68	1.05	0.40	3.00	2.23	0.94	0.70	3.30
NH_3-N	2.57	1.78	0.00	4.80	3.45	1.46	1.85	5.27	4.58	3.65	0.10	9.23
NO_3-N	0.22	0.18	0.03	0.47	0.24	0.17	0.01	0.51	0.69	0.65	0.01	1.34
PO_4-3	4.15	2.97	0.00	8.38	2.52	3.45	0.69	9.40	2.89	3.28	0.77	8.45
	S4				**S5**				**S6**			
Temp	27.40	3.10	21.21	29.75	28.18	0.72	27.33	29.49	29.05	4.14	26.86	37.47
pH	7.16	0.09	7.05	7.27	6.81	0.19	6.49	7.05	6.57	0.34	6.02	7.09
EC	149.67	30.66	90.00	174.00	170.00	52.17	107.00	225.00	54.50	10.71	35.00	65.00
DO	4.54	2.44	0.17	6.75	3.44	1.78	0.25	5.41	1.63	1.44	0.11	4.17
Turb.	18.25	29.53	1.70	78.10	23.64	33.19	6.15	91.10	20.76	21.82	2.28	59.50
Salinity	0.07	0.02	0.04	0.08	0.08	0.02	0.05	0.11	0.02	0.01	0.01	0.03
TDS	97.33	19.64	59.00	113.00	104.17	33.45	70.00	147.00	34.33	5.96	23.00	40.00
TSS	43.67	29.13	14.00	90.00	34.00	18.37	16.00	68.00	38.17	15.01	23.00	57.00
COD	7.08	6.33	2.37	19.33	8.28	6.73	1.33	19.00	11.72	13.15	1.33	32.00
BOD	1.73	1.44	0.60	4.40	1.72	1.37	0.20	3.40	1.78	1.36	0.50	3.80
NH_3-N	1.95	1.49	0.12	4.17	2.53	1.92	0.13	5.88	4.68	1.56	2.56	6.73
NO_3-N	0.43	0.37	0.10	1.15	0.61	0.67	0.00	1.60	0.25	0.15	0.03	0.42
PO_4-3	3.81	3.64	1.35	9.57	5.29	5.81	1.37	15.25	4.54	4.07	1.23	10.35

EC: Electric conductivity, DO: Dissolved oxygen, Turb: Turbidity, TDS: Total dissolved solids, TSS: Total sispended solids, COD: Chemical oxygen demand, BOD: Biological oxygen demand, NH_3-N: Ammonica; nitrogen, NO_3N: Nitrateni Trogen and PO_3: Phosphate

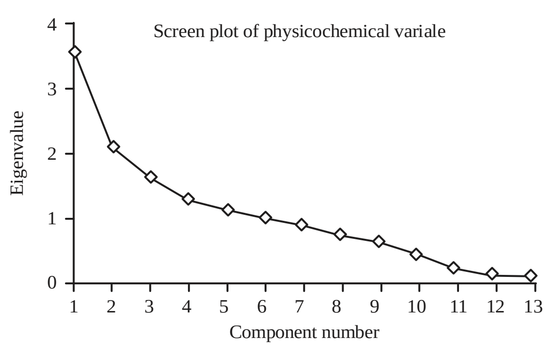

Fig. 4: Scree plot of physicochemical variables

In the present study, the scree plot (Fig. 4), showed a pronounced change of slope after the fifth Eigenvalue. Equal numbers of VFs were obtained for parameters through FA performed on the PCs. Corresponding VFs, variable loadings and the variance explained are presented in Table 3 (Liu *et al*., 2003), which classi ed the factor loadings as "Strong," "Moderate" or "Weak," corresponding to absolute loading values of >0.75, 0.75-0.50 and 0.50-0.30, respectively.

The first principal component had variance 3.5752 (equal to the largest eigenvalue) and accounted for 0.275 (27.5%) of the total variation in the data. The second principal component accounts for 0.158 (16%) of the total data variation. Five chosen factors are now interpre Table 4.

DISCUSSION

Hierarchical cluster analysis (HACA) reveals that the classified groups were generated based on the similar characteristic features and natural backgrounds that are affected by the sources. Group A (stations S1 and S6) contains anthropogenic source of pollution mainly through soil erosion from the upper barren sites with loose soil at near Kolej Zaba and FST new building of UKM campus. Anand *et al*. (2003) also indicated that effects of soil erosion from the land, which was converted into barren slope by removal of plant litter causes water pollution in the stream. The other point sources such as domestic effluents mostly from university laboratory, cafeteria and student hostels and runoff from the fern garden are significant. The stations in this group are situated at the down of the tributaries. The water of these sites remains cloudy due to high content of TSS and turbidity till sediment is deposited and trapped in the lake (S1) and tributary Alur Ilmu (S6)

Table 3: Factor loadings of the variables using principal component analysis

Variables	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Temp	0.106	-0.132	0.632	-0.179	0.057	0.463
pH	0.583	0.271	0.151	-0.515	-0.109	0.713
EC	0.943	0.183	-0.105	0.115	-0.020	0.948
DO	0.461	-0.592	-0.033	-0.452	0.127	0.785
Turbidity	-0.249	-0.065	0.022	-0.412	-0.629	0.633
Salinity	0.950	0.168	-0.131	0.107	-0.054	0.962
TDS	0.938	0.223	-0.095	0.094	-0.100	0.957
TSS	-0.327	0.445	-0.255	-0.489	0.121	0.624
COD	0.071	0.745	0.183	-0.154	0.275	0.693
BOD	0.088	-0.001	0.819	-0.179	0.015	0.711
NH_3-N	-0.333	0.705	0.185	0.128	0.219	0.707
NO_3-N	0.201	-0.501	0.131	-0.007	0.615	0.687
PO_4-3	-0.082	0.001	-0.535	-0.462	0.324	0.612
Variance	3.5752	2.0591	1.5683	1.2300	1.0619	9.4946
Var (%)	27.0	16.0	12.0	9.0	8.0	73.0

EC: Electric conductivity, DO: Dissolved oxygen, Turb: Turbidity, TDS: Total dissolved solids, TSS: Total sispended solids, COD: Chemical oxygen demand, BOD: Biological oxygen demand, NH_3-N: Ammonica; nitrogen, NO_3N: Nitrateni Trogen, PO_3: Phosphate

Table 4: Factor loading of physicochemical variables to surface waters using factor analysis

Variable	VF1	VF2	VF3	VF4	VF5	Communality
Temp	0.013	0.059	-0.668	-0.101	-0.058	0.463
pH	-0.628	0.001	-0.393	0.338	0.223	0.713
EC	-0.957	0.059	0.031	-0.089	-0.143	0.948
DO	-0.252	0.733	-0.260	0.293	-0.172	0.785
Turbidity	0.187	0.199	-0.107	0.098	0.733	0.633
Salinity	-0.965	0.087	0.052	-0.089	-0.114	0.962
TDS	-0.971	0.039	0.023	-0.098	-0.055	0.957
TSS	0.159	-0.351	0.094	0.647	0.217	0.624
COD	-0.239	-0.709	-0.210	0.294	-0.051	0.693
BOD	0.006	-0.090	-0.823	-0.163	0.007	0.711
NH_3-N	0.150	-0.824	-0.044	0.058	-0.033	0.707
NO_3-N	0.038	0.372	-0.228	0.082	-0.699	0.687
PO_4-3	0.051	0.121	0.281	0. 710	-0.106	0.612
Variance	3. 3907	2.0592	1.5463	1.2923	1.2061	9.4946
Var (%)	26.0	15.0	12.0	10.0	9.0	73.0

Among ve Varifactors (VFs) (Table 4), VF1 explained 26.1% of total variance, had moderate (0.75-0.50) or strong loadings (>0.70) on pH (0.624), EC (-0.957), salinity (-0.965) and TDS (-0.971). The VF2 described 15.8% of the total variance, had strong positive loadings on DO (0.733), COD (-0.709), NH_3-N (-0.824). The VF3 (11.09% of the total variance) had strong negative loadings on Temp (-0.668) and BOD (-0.823). The VF4 (9.9% of the total variance) had strong positive loadings on TSS (0.647) and PO_4-3(0.710). The VF5 stated the lowest variance (9.3%), had strong positive loadings on turbidity (0.733) and negative loading on NO3-N (-0.699)

respectively. Lane and Sheridan (2002) indicated the similar findings, where they calculated that suspended sediment loads were 0.78 t at upstream and 2.77 t at downstream and also stated that rainfall and runoff simulation exposed the principal sediment sources, to be a fill slope that contributed coarse bed load material through rill erosion and unprotected toe scour. Group B comprises of stations S2 and S3 corresponds to relatively less polluted sites as these stations are situated at less anthropogenic disturbed upstream sites of the tributary Tasik Kejuruteraan. Sweeney *et al*. (2004), revealed the similar findings that forest buffers preventing nonpoint source pollutants from entering small streams, they also enhance the in-stream processing of both nonpoint and point source pollutants, thereby reducing their impact on downstream rivers and estuaries. On the other hand, Emelko *et al*. (2011), contradicted with the statement mentioning that large scale natural disturbances from severe insect infestation rather than human interference may cause water quality deterioration in the water channel. However, these stations receive pollutants mainly from nonpoint sources, i.e., surface runoff from vegetation covered area of the catchment. Group C formed by the stations S4 and S5 receives pollution from both point and non-point sources like domestic and laboratory effluents and basin runoff but with less quantity. The station S4 contents less sediment because sediment is artificially trapped at upper area to S4. Again cluster formed due to temporal variation of dataset reveals that the temporal patterns were on the basis of sample collected in rainy weather or dry weather (Lane and Sheridan, 2002).

The factor analysis shows that varifactor 1 (VF1) contains hydro-physical variables (pH, EC, salinity and TDS) and geological components of soils (Zhou *et al*., 2012). The significant factor loading of pH, EC, salinity and TDS to this factor can be considered a result of cation-exchange processes at soil-water interface (Guo and Wang, 2004). The VF2 represents the effects of chemical oxidation contributed by nonpoint pollution, which is originated from artificial cultivated area such as fern garden, experimental plots in the university campus where planters use phosphate and nitrogen fertilizers and later ephemeral tributaries receive through surface runoff (Carpenter *et al*., 1998). However, NH_3-N used for experiments also comes from some point sources like laboratory and sewerage (Meyer-Reil and Koster, 2000). The VF3 explains the deoxygenation of water (George and Heaney, 1978). The VF4 expresses the nonpoint sources of PO_4^{-3} and TSS that occurs with soil erosion and water runoff from the anthropogenic disturbed area of the basin. This factor can be identified as algal availability factor (Uusitalo *et al*., 2001). The VF5 is revealed as organic matter factor. Nitrate nitrogen availability is due to numerous sources, such as geologic deposits, natural organic matter decomposition and domestic waste runoff (Ruiz-Fernandez *et al*., 2002).

The PCA was employed and explained about 73% of the variance of the dataset. Five factors were extracted, which are identified as the contribution from the physical factors, chemical factors, de-oxygenation, nonpoint sources of nutrient from soil erosion. No variables were sensitive to factor component 4.

CONCLUSION

Different multivariate statistical methods were used to assess temporal and spatial variations in surface water quality of three tributaries to Langat sub basin. Cluster analysis successfully examined the existence of three clusters from six sampling stations and two temporal clusters separately for their differing water quality. According to principal component analysis followed by factor analysis loading resulted the five varifactors extracted, which indicated the anthropogenic activities mainly affecting the tributaries to Langat upstream include run-off, domestic and laboratory discharge, the temporal and spatial similarities and groupings could facilitate the design of an optimal future monitoring strategy that could decrease monitoring frequency, the number of sampling stations and the corresponding costs. The factor analysis/principle component analysis helped extract and identify the factors/sources responsible for variations in upstream tributaries water quality at sampling sites. Varifactors obtained from factor analysis indicate that the parameters responsible for water quality variations are mainly related to physicochemical variables and nutrients in the basin.

ACKNOWLEDGEMENT

The university research project-AP 2014-010 of Universiti Kebangsaan Malaysia (UKM) is gratefully acknowledged for supporting this research.

REFERENCES

Abidin, I.Z., H. Juahir, A. Azid, A.D. Mustafa and F. Azaman, 2015. Application of excel-VBA for computation of water quality index and air pollutant index. Malaysian J. Anal. Sci., 19: 1056-1064.

Alberto, W.D., D.M. del Pilar, A.M. Valeria, P.S. Fabiana, H.A. Cecilia and B.M. de los Angeles, 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia River Basin (Cordoba-Argentina). Water Res., 35: 2881-2894.

Anand, M., K.M. Ma, A. Okonski, S. Levin and D. Mccreath, 2003. Characterising biocomplexity and soil microbial dynamics along a smelter-damaged landscape gradient. Sci. Total Environ., 311: 247-259.

Carpenter, S.R., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley and V.H. Smith, 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol. Applic., 8: 559-568.

Casatti, L., F. Langeani and C.P. Ferreira, 2006. Effects of physical habitat degradation on the stream fish assemblage structure In a pasture region. Environ. Manage., 38: 974-982.

Cespedes, R., S. Lacorte, A. Ginebreda and D. Barcelo, 2006. Chemical monitoring and occurrence of alkylphenols, alkylphenol ethoxylates, alcohol ethoxylates, phthalates and benzothiazoles in sewage treatment plants and receiving waters along the Ter River basin (Catalonia, N. E. Spain). Anal. Bioanal. Chem., 385: 992-1000.

Chen, M.Y., E.K. Wang, R.J. Yang and Y.M. Liou, 2003. Adolescent health promotion scale: Development and psychometric testing. Pubic Health Nurs., 20: 104-110.

Davis, N.M., V. Weaver, K. Parks and M.J. Lydy, 2003. An assessment of water quality, physical habitat and biological integrity of an urban stream in Wichita, Kansas, prior to restoration improvements (phase I). Arch. Environ. Contam. Toxicol., 44: 351-359.

Devito, K.J., A. Hill and N. Roulet, 1996. Groundwater-surface water interactions in headwater forested wetlands of the Canadian shield. J. Hydrol., 181: 127-147.

Emelko, M.B., U. Silins, K.D. Bladon and M. Stone, 2011. Implications of land disturbance on drinking water treatability in a changing climate: Demonstrating the need for source water supply and protection strategies. Water Res., 45: 461-472.

Fausch, K.D., C.E. Torgersen, C.V. Baxter and H.W. Li, 2002. Landscapes to riverscapes: Bridging the gap between research and conservation of stream fishes. BioScience, 52: 483-498.

George, D.G. and S.I. Heaney, 1978. Factors influencing the spatial distribution of phytoplankton in a small productive lake. J. Ecol., 66: 133-155.

Guo, H. and Y. Wang, 2004. Hydrogeochemical processes in shallow quaternary aquifers from the Northern part of the Datong Basin, China. Applied Geochem., 19: 19-27.

Helena, B., R. Pardo, M. Vega, E. Barrado, J.M. Fernandez and L. Fernandez, 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. Water Res., 34: 807-816.

Ketchen, D.J. and C.L. Shook, 1996. The application of cluster analysis in strategic management research: An analysis and critique. Strat. Manage. J., 17: 441-458.

Lampariello, F., 2000. On the use of the Kolmogorov-Smirnov statistical test for immunofluorescence histogram comparison. Cytometry, 39: 179-188.

Lane, P.N. and G.J. Sheridan, 2002. Impact of an unsealed forest road stream crossing: Water quality and sediment sources. Hydrol. Processes, 16: 2599-2612.

Liu, C.W., K.H. Lin and Y.M. Kuo, 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. Sci. Total Environ., 313: 77-89.

Meyer-Reil, L.A. and M. Koster, 2000. Eutrophication of marine waters: Effects on benthic microbial communities. Mar. Pollut. Bull., 41: 255-263.

Ning, S.K. and N.B. Chang, 2004. Optimal expansion of water quality monitoring network by fuzzy optimization approach. Environ. Monitor. Assess., 91: 145-170.

Poff, N.L. and J.V. Ward, 1990. Physical habitat template of lotic systems: Recovery in the context of historical pattern of spatiotemporal heterogeneity. Environ. Manage., 14: 629-645.

Ruiz-Fernandez, A., C. Hillaire-Marcel, B. Ghaleb, M. Soto-Jimenez and F. Paez-Osuna, 2002. Recent sedimentary history of anthropogenic impacts on the Culiacan River Estuary, Northwestern Mexico: Geochemical evidence from organic matter and nutrients. Environ. Pollut., 118: 365-377.

Shields, C.A., L.E. Band, N. Law, P.M. Groffman and S.S. Kaushal *et al*., 2008. Streamflow distribution of non-point source nitrogen export from urban-rural catchments in the Chesapeake Bay watershed. Water Resour. Res., Vol. 44, No. 9. 10.1029/2007WR006360

Shrestha, S. and F. Kazama, 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River Basin, Japan. Environ. Modell. Software, 22: 464-475.

Simeonov, V., J.A. Stratis, C. Samara, G. Zachariadis and D. Voutsa *et al*., 2003. Assessment of the surface water quality in Northern Greece. Water Res., 37: 4119-4124.

Singh, K.P., A. Malik, D. Mohan and S. Sinha, 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)-A case study. Water Res., 38: 3980-3992.

Sweeney, B.W., T.L. Bott, J.K. Jackson, L.A. Kaplan and J.D. Newbold *et al*., 2004. Riparian deforestation, stream narrowing and loss of stream ecosystem services. Proc. Natl. Acad. Sci. USA., 101: 14132-14137.

Uusitalo, R., E. Turtola, T. Kauppila and T. Lilja, 2001. Particulate phosphorus and sediment in surface runoff and drainflow from clayey soils. J. Environ. Qual., 30: 589-595.

Varol, M. and B. Sen, 2009. Assessment of surface water quality using multivariate statistical techniques: A case study of Behrimaz Stream, Turkey. Environ. Monitor. Assess., 159: 543-553.

Wang, Y., P. Wang, Y. Bai, Z. Tian and J. Li et al., 2013. Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua River Harbin region, China. J. Hydro-Environ. Res., 7: 30-40.

Yidana, S.M., D. Ophori and B. Banoeng-Yakubo, 2008. A multivariate statistical analysis of surface water chemistry data-The Ankobra Basin, Ghana. J. Environ. Manage., 86: 80-87.

Zhou, Z., G. Zhang, M. Yan and J. Wang, 2012. Spatial variability of the shallow groundwater level and its chemistry characteristics in the low plain around the Bohai Sea, North China. Environ. Monitor. Assess., 18: 3697-3710.