American-Eurasian J. Agric. & Environ. Sci., 4 (2): 258-265, 2008
ISSN 1818-6769

The Use of Chemometrics Analysis as a Cost-effective Tool in Sustainable Utilisation of Water Resources in the Langat River Catchment

[1]Hafizan Juahir, [2]T. Mohd Ekhwan, [3]Sharifuddin M. Zain,
[4]Mazlin B. Mokhtar, [5]Zaihan Jalaludin, [6]Ijan Khushaida M. Jan

[1]Department of Environmental Science, Faculty of Environmental Study,
Universiti Putera Malaysia, 43000 Serdang
[2]School of Social, Development and Environment, FSKK,
Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
[3,6]Chemistry Department, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
[4]Institute for Environment and Development (LESTARI),
Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
[5]Forest Research Institute Malaysia, 52109 Kepong, Kuala Lumpur, Malaysia

Abstract: Malaysia was adopted the concept of sustainable development as mentioned in the National Documents of the 8th Malaysian Plan and OPP3. This calls for environmental studies within the context of sustainable science and governance. This study details the application of chemometrics in environmental chemistry for sustainable utilization of resources in the Langat Basin, Selangor, Malaysia. We hope to demonstrate in this work the importance of historical data, if they are available, in planning sampling strategies to achieve desired research objectives. To achieve the objectives, this study highlights the possibility of determining the optimum number of sampling stations, which in turn would reduce cost and time of sampling. The seasonally dependent water quality data of Langat River was investigated during the period of December 2001 to May 2002. Monthly water samples were collected from four different stations. Concentrations of nitrate, sulfate, phosphate, lead, cadmium, iron, zinc and copper were determined. Dissolved oxygen (DO), biological oxygen demand (BOD), chemical oxygen demand (COD), temperature, pH, total suspended solids (TSS), ammoniacal nitrogen (AN) and conductivity were measured insitu. Chemometric treatments using cluster, principal component analysis and factorial design were employed where data were characterized as function of season and sampling sites, thus, enabling significant discriminating factors to be discovered. Results showed that at a chord distance of 75.25 the cluster gave two groups of sampling plot. Group I consists of 6 sampling stations while Group II consists of 14 sampling stations. The two clusters are discussed in terms of the difference in data variability.

Key words: Chemometrics · Principles component analysis · Cluster analysis · Factorial design

INTRODUCTION

Environmental data is complex and depends on unpredictable factors that are usually characterized by their high variability. The main origins of this variability are geogenic, hydrological, meteorological and also anthropogenic (such as different emitters and dischargers) [1]. Due to the non-linear nature of environmental data, analyzing these data may be tricky. The multivariate nature of these data together with their complex interrelation requires that multivariate data analysis techniques be employed in order to decipher any structure within the data. In this study, the application of chemometrics methods was used to determine the number of sampling sites which appear significantly different to each other. This work is motivated by the fact that an

Corresponding Author: Dr. M.Ekhwan Toriman, School of Social, Development and Environmental Studies, FSSK Universiti Kebangsaan Malaysia, 43600 Bangi Selangor Malaysia

understanding of the nature of sampling sites would help in reducing the number of redundant sites, thus reduce cost and time.

The data collected in this study was obtained from four different sampling plots which provide in total of 17 sampling sites altogether. The study was conducted by a researcher to measure the impacts of palm oil plantation activities to the water quality in the Langat River Basin. The selected plots are Kg. Bukit Dugang, Kg. Jenderam, Bukit Changgang and Labohan Dagang which located along the river basin. In this study, the sampling plots were selected based on the economic needs of two districts involved in this study area (Kuala Langat and Sepang Districts). The main economic activities for both districts are agriculture and industry with palm oil plantation as the main agricultural activity [2]. Unfortunately, the study was conducted without proper sampling design and the selected plots were not statistically identified. It is well known that much have been studied on water quality by many researchers for the Langat River Basin. Therefore, secondary data from previous studies can be used to obtain additional information to help us in designing new research approach at the Langat River Basin. The abundance of secondary data motivated us to use chemometric methods in order that proper sampling design can be obtained.

Chemometrics can be defined as "a chemical discipline that uses mathematics, statistics and formal logic (a) to design or select optimal experimental procedures; (b) to provide maximum relevant chemical data; and (c) to obtain knowledge about chemical systems". Chemometric methods have been used for the classification and comparison of different samples [3]. Some examples of the use of chemometrics are as a multicriteria decision-making [4], chemometric investigation of variable and site correlations [5], determination of correlation of chemical and sensory data in drinking waters by factors analysis [6]. The chemometric applications in evaluating environmental data has been demonstrated in several publications [7,8,9].

This study was carried out to fulfill these objectives, namely (i) to apply chemometrics in recognizing patterns in the sampling data (ii) to evaluate and interpret river pollution data (iii) to encourage the use of secondary data to help scientists and researchers in designing better approaches to future studies and (iv) to understand how computer and software technologies can

MATERIALS AND METHODS

Study Site: Langat River Basin is formed by three main rivers which are Langat River, Semenyih River and Labu River. At the length of about 125.6 km, the rivers flow across states of Negeri Sembilan and Selangor. Langat River is one of the most important raw water resources for drinking water and other activities such as recreation, industrial uses, fishery and agriculture. In this area, agriculture is the main activity which covers 53.1% of the area, while 3.6% are for commercial purposes. Palm oil plantation takes 20,993 ha from the area and another 13,574 ha is covered by rubber plantation.

Up to 17 sampling sites were selected to cover the study (Fig. 1). To select the location of the sampling stations, the conventional method based on economic activities are taken into consideration. The sampling stations were divided into four plots which are plots one and two namely Kampung Bukit Dugang and Kampung Jenderam consisting of five sampling stations located in the Sepang District. Plots 3 and 4, namely Bukit Changgang and Labohan Dagang are located in Kuala Langat District which consists of four and three sampling stations, respectively (Table 1).

Sampling: A total of 102 water samples were collected from each plot consisting of 17 sampling stations during the site visit between December 2001 and May 2002. The sampling dates were divided into two weather conditions; three sampling days in dry weather season (10th January 2002, 19th February 2002 and 15th May 2002) and another three days during the rainy season (26th December 2001, 3rd March 2002 and 13th April 2002). Table 2 indicates the stations which were sampled during each site visit. 16 physico-chemical parameters were determined; water temperature (°C), pH, TSS, DO, BOD, COD, conductivity, nitrate, sulfate, phosphate, lead, cadmium, iron, zinc, manganese, potassium, calcium, magnesium and copper (Table 3).

Analytical Procedures: DO, temperature, pH and conductivity were measured in situ. Ammoniacal nitrogen, phosphate, nitrate, sulfate were determined using HACH Kit (Models FF-2 and FF-1A) and 8038 Spectrophotmeter HACH DR 2000. Heavy metals (Pb, Fe, Zn, Cu and Cd), BOD and TSS were analyzed according to methods of American Public Health Association. While COD was determined using HACH Kit (Models 8000),

Table 1: Locations of plots and sampling stations

			Coordinate	Area description	
District	Study area (plot No.)	Station No.	Latitude	Longitude	
Sepang	Kampung Bukit Dugang (Plot 1)	1.1	101°43.387'	02°53.778'	• Surrounded by palm oil
		1.2	101°43.282'	02°53.904'	plantation
		1.3	101°43.262'	02°53.818'	• Orang Asli village
		1.4	101°43.088'	02°53.760'	• Sand mining (st. 1.4 & 1.5)
		1.5	101°42.925'	02°53.633'	
	Kampung Jenderam (Plot 2)	2.1	101°43.853'	02°52.036'	• Surrounded by palm oil plantation
		2.2	101°43.523'	02°52.177'	• Village
		2.3	101°43.208'	02°52.430'	
		2.4	101°42.795'	02°52.841'	
		2.5	101°42.571'	02°53.013'	
Kuala	Bukit Changgang (Plot 3)	3.1	101°39.079'	02°49.156'	• Surrounded by palm oil plantation
Langat		3.2	101°38.590'	02°48.806'	• Village
		3.3	101°38.564'	02°48.823'	
		3.4	101°38.500'	02°48.787'	
	Labohan Dagang (Plot 4)	4.1	101°36.990'	02°47.510'	• Surrounded by palm oil plantation
		4.2	101°36.964'	02°47.520'	• Village
		4.3	101°36.853'	02°47.454'	• Wetland (st. 4.3)

Table 2: Sampling plots showing samples taken during dry and wet days

		Sampling date					
Plot	Station	a	b	c	d	e	f
I	1.1	cloudy	cloudy	dry	overcast	overcast	overcast
	1.2	cloudy	dry	dry	overcast	overcast	clear
	1.3	cloudy	dry	dry	overcast	overcast	clear
	1.4	overcast	dry	dry	overcast	overcast	clear
	1.5	overcast	dry	dry	overcast	overcast	clear
II	2.1	overcast	dry	dry	overcast	overcast	dry
	2.2	overcast	dry	dry	overcast	overcast	dry
	2.3	overcast	dry	dry	overcast	overcast	dry
	2.4	overcast	dry	dry	overcast	overcast	dry
	2.5	overcast	dry	dry	overcast	overcast	dry
III	3.1	overcast	dry	dry	overcast	overcast	dry
	3.2	overcast	dry	dry	overcast	overcast	dry
	3.3	overcast	dry	dry	overcast	overcast	dry
	3.4	overcast	dry	dry	overcast	clear	dry
IV	4.1	overcast	dry	dry	overcast	clear	dry
	4.2	overcast	dry	dry	overcast	clear	dry
	4.3	overcast	dry	dry	overcast	clear	dry

(a) 26 December 2001, (b) 10 January 2002, (c) 19 February 2002, (d) 3 March 2002, (e) 13 April 2002 and (f) 15 May 2002

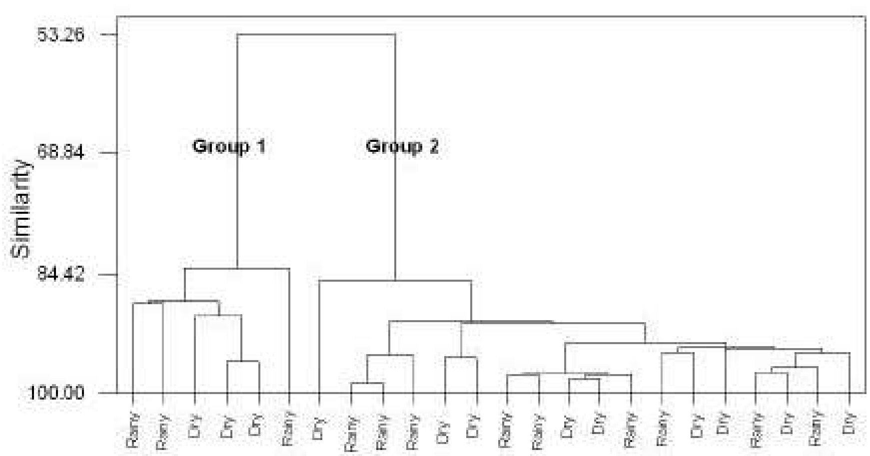

Fig. 1: Seasonal dendogram calculated by the Ward method for the variables of Table 2. The four sampling plots with six sampling periods

Thermoreactor Model-Eco 16 Thermoreactor Velp Scientifica.

Statistical Procedures: In principal components analysis (PCA), eigenanalysis of the experimental data was performed to extract principal components (PCs) of the measured data, using two selection criteria: the scree plot test and corrected average eigenvalue. For hierarchical cluster analysis (CA), the squared Euclidean distance between normalized data was used to measure similarities between samples. Both average linkage between groups and Ward's method were applied to standardized data and the results obtained were represented in a dendogram. The design of experiment (DOE) method was employed to identify the interaction between the seasonal effected to the water quality parameter.

Statistical analysis was carried out by using both Datalab for Teach/Me software [10], Minitab 13.0 and Excel for Windows software packages.

RESULTS AND DISCUSSION

Table 3 reports the data obtained for the samples collected. The data set comprises of 24 samples which comes from four different plots which consists of 17 sampling sites. Plot one and two consist of five sampling sites each. Plot three consists of four sampling sites and plot four consists of three sampling sites. The samples were collected in six different sampling days. For each of the 24 samples, 16 features have been evaluated.

Cluster Analysis: Cluster analysis is one of the method that was applied in unsupervised pattern recognition [11,12,13]. In this study, cluster analysis was applied to search for clusters due to different sampling days or different sampling sites by using water quality variables or features. The agglomerative hierarchical cluster analysis according to Ward (1963) [14] was applied to detect multivariate similarities between sampling sites in different sampling plots for different sampling days. From Fig. 1, it is observed that separation between group 1 and group 2 are clearly not due to seasonal changes. Differences in the feature values (water quality parameters) where probably due to seasonal changes were distributed over the whole area of sampling plots. It does not form the basis of the separation observed in the objects (sampling sites).

On the other hand, Fig. 2 shows that if the separation is grouped according to sampling plots, the separation shows clear discrimination of Labohan Dagang and the other sites. It can be seen that Labohan Dagang (Group 1) sampling plot at similarity level 75.25 (dash line in Fig. 2.) is very different from the others. In this study the other three sampling plots which merge at similarity level 75.25 (Bukit Changgang, Kampung Jenderam and Kampung Bukit Dugang) forms a single group (Group 2).

The two groups of samples from plots 4 (Group 1) and plots 1, 2 and 3 (Group 2) join at a lower level of similarity in the sampling plot dendogram (Fig. 2) compared to the seasonal dendogram (Fig. 1). This demonstrates that, from a hierarchical point of view, the

Table 3: Experimental data

	Variable															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Sampling site	pH	Temp.	Cond. (µS/cm)	TSS (mg/L)	DO (mg/L)	BOD (mg/L)	COD (mg/L)	AN (mg/L)	PO (mg/L)	NO (mg/L)	SO (mg/L)	Pb (mg/L)	Cd (mg/L)	Fe (mg/L)	Zn (mg/L)	Cu (mg/L)
Kg. Bukit Dugang (26/12/2001)	5.8	30.0	69	65.4	3.0	5.44	21	1.57	0.16	3.1	0.8	0.54	0.01	2.80	0.04	32.56
Kg. Jenderam (26/12/2001)	3.5	27.0	126	2.8	1.5	3.74	18	1.57	0.14	0.9	6.6	0.26	0.01	2.20	0.08	2.01
Bukit Changgang (26/12/2001)	5.9	28.0	67	186.3	4.7	6.20	9	1.32	0.08	1.3	0.6	0.37	0.01	0.09	0.02	2.46
Labohan Dagang (26/12/2001)	5.8	29.0	96	815.3	3.6	5.44	45	0.57	0.04	6.3	138.9	0.55	0.02	2.20	0.08	2.00
Kg. Bukit Dugang (10/01/2002)	5.8	32.0	74	10.6	4.3	2.00	9	1.60	1.50	2.6	3.0	1.65	0.15	2.44	2.28	2.92
Kg. Jenderam (10/01/2002)	5.2	24.5	211	1.6	1.2	0.45	6	2.41	0.85	0.8	15.9	3.42	0.44	1.46	2.04	2.41
Bukit Changgang (10/01/2002)	5.3	29.6	189	283.7	4.2	1.32	24	1.34	0.11	2.8	20.6	2.73	0.14	3.80	2.24	3.31
Labohan Dagang (10/01/2002)	5.6	30.0	175	746.9	1.7	0.68	10	0.87	0.03	5.7	102.6	1.11	0.16	0.38	1.67	2.05
Kg. Bukit Dugang (19/02/2002)	5.5	31.0	76	95.4	4.2	2.51	8	1.24	1.94	1.4	2.0	3.85	0.25	2.59	2.19	71.95
Kg. Jenderam (19/02/2002)	6.3	28.1	255	0.1	0.3	0.10	1	2.22	0.96	0.7	13.0	4.28	0.45	1.61	1.88	2.38
Bukit Changgang (19/02/2002)	5.4	32.9	215	119.9	5.0	1.17	2	1.71	0.12	3.9	25.0	2.57	0.13	5.87	1.96	1.44
Labohan Dagang (19/02/2002)	5.5	30.5	290	724.3	0.6	0.01	27	1.44	0.01	3.9	44.0	1.79	0.13	0.62	2.23	1.62
Kg. Bukit Dugang (3/03/2002)	5.7	30.5	29	158.9	4.2	1.28	7	0.60	0.01	0.9	7.0	8.27	0.67	1.92	3.96	0.49
Kg. Jenderam (3/03/2002)	4.7	28.2	105	0.1	1.2	1.63	25	1.95	0.04	0.8	7.0	6.85	0.36	0.81	3.6	0.19
Bukit Changgang (3/03/2002)	4.2	29.2	153	147.6	1.1	1.14	0	1.84	0.01	1.4	27.0	3.57	0.69	3.47	3.42	0.26
Labohan Dagang (3/03/2002)	5.1	29.1	74	951.4	3.4	0.29	10	2.04	0.01	1.1	31.0	2.84	0.18	0.16	5.89	0.12
Kg. Bukit Dugang (13/04/2002)	5.8	29.4	76	188.1	2.3	0.50	8	0.50	0.14	1.1	5.0	4.45	0.39	1.27	3.41	0.12
Kg. Jenderam (13/04/2002)	5.2	29.6	106	0.2	2.1	0.43	1	1.89	0.26	1	1.0	2.58	0.18	1.18	6.87	0.13
Bukit Changgang (13/04/2002)	5.9	29.8	132	123.5	3.6	0.99	2	1.89	0.01	1.5	32.0	2.39	0.43	3.21	3.14	0.03
Labohan Dagang (13/04/2002)	5.1	29.9	92	795.7	4.0	0.67	26	1.99	0.01	1.2	29.0	3.81	0.1	0.14	7.21	0.18
Kg. Bukit Dugang (15/05/2002)	6.6	27.8	163	133.5	6.1	1.74	2	1.84	0.38	1.2	9.0	1.09	0.09	2.27	4.54	0.16
Kg. Jenderam (15/05/2002)	6.7	31.2	85	0.3	4.6	0.35	4	0.23	0.25	0.8	5.0	6.74	0.16	1.09	3.4	0.28
Bukit Changgang (15/05/2002)	6.3	32.4	104	85.3	5.1	1.21	1	1.23	0.00	1.2	18.0	5.54	0.6	3.49	4.39	0.22
Labohan Dagang (15/05/2002)	4.6	30.3	263	734.7	4.7	0.43	7	2.41	0.02	1.5	63.0	3.79	0.01	0.15	1.79	0.43

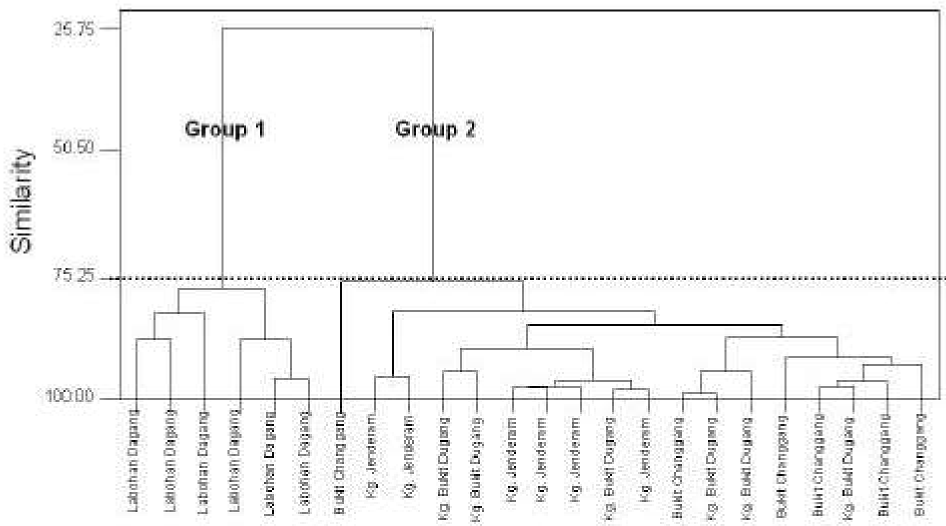

Fig. 2: Sampling plot dendogram clearly separating Labohan Dagang and the other plots

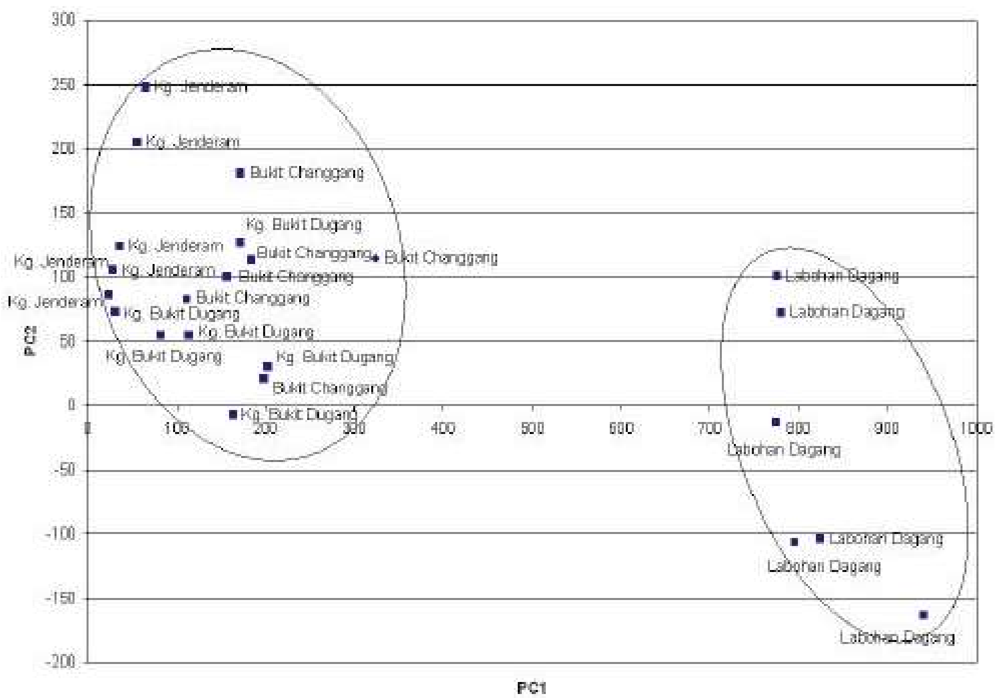

Fig. 3: Principal component analysis for four sampling plots (with six sampling periods)

difference between the two separated groups (1 and 2) is larger in the sampling plot dendogram (Fig. 2) compared to the seasonal dendogram (Fig. 1). This is an indication that separation of sampling plot should be used as a more significant factor in forming the basis of choosing sampling sites in order to study the effects of palm oil plantation on water quality. Searching for seasonal dependencies based on the conventionally chosen sampling sites is consequently perhaps an ineffective exercise which involves high cost and much time being wasted.

Principal Component Analysis: Table 4 shows the variance explained by the principal components obtained in a principal component analysis (PCA). It clearly shows that most of the data variance is explained in the first two

Table 4: Variances of PCA for the first six PCs

PC	Variance (%)	Total
1	92.70	92.70
2	6.76	99.46
3	0.26	99.72
4	0.17	99.88
5	0.07	99.96
6	0.04	99.99

Table 5: ANOVA: Two factor with replication

Summary	Overcast	Dry	Total
A			
Count	3	3	6
Sum	2562.4	2205.9	4768.3
Average	854.1333	735.3	794.7167
Variance	7191.643	127.96	7164.25
B			
Count	3	3	6
Sum	416	239.5	655.5
Average	138.6667	79.83333	109.25
Variance	3853.243	3957.843	4162.843
Total			
Count	6	6	
Sum	2978.4	2445.4	
Average	496.4	407.5667	
Variance	157985.7	130525.3	

PCs (99.46%). This result is in agreement with the observed highly redundant information caused by the presence of several variables with high covariance.

Figure 3 shows the scores of the objects (sampling sites) in a space spanned by PC1 and PC2. The loadings of each feature (water quality variables) are shown for PC1 in Fig. 4. In Fig. 3, the scores plot clearly shows two linearly separable clusters. The cluster on the right is formed by sampling sites in the Labohan Dagang plot while the rest of the sampling stations in the three sampling plots (Kampung Bukit Dugang, Kampung Jenderam and Bukit Changgang) form the other cluster. This further confirms, via visual inspection, the dendograms obtained from the hierarchical analysis results. It can be remarked from the values of the loadings of the features for PC1 (92.70%) (Fig. 4) that the difference between the two groups of sampling plots (Group 1 and 2) is mainly due to the total suspended solid (TSS) (variable 4). Suspended solid parameter is related to the natural erosion from the forest and agriculture area [15]. The second important variable is the conductivity (variable 3) which is due to high concentration of inorganic compounds in the water sample. This

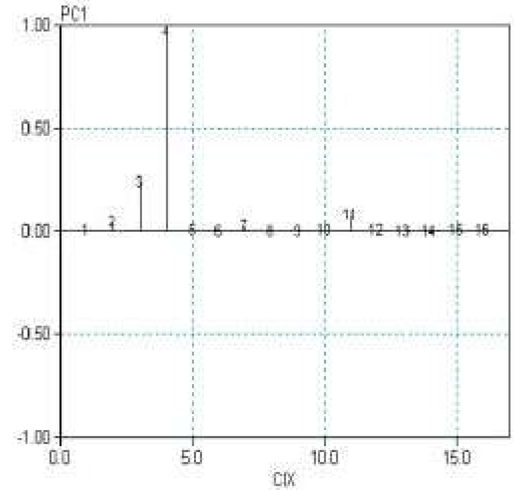

Fig. 4: Plot of PC1 loadings

observation would form the second part of our study – relating the SS and conductivity difference in Labohan Dagang to palm oil plantation, if any, or relating them to other, as yet, unknown activities near the sampling sites in the plots of study.

Design of Experiments

Factoral Designs: These experimental designs have been classified under the name of factorial designs, because they evaluate the effects of two or more factors simultaneously [16]. To interpret the results, by testing whether there is an interaction effects between factor I (sampling station) and factor II (weather condition). If the interaction effect is significant, one must be cautious in the interpretation of any significant main effects. On the other hand, if the interaction effect is not significant, the focus should be on the main effects-potential differences in sampling station and potential differences in weather condition (factor II).

At the 0.05 level of significance to determine whether there is evidence of an interaction, the decision rule is to reject the null hypothesis of no interaction between sampling station and weather condition if the calculated F exceeds 5.32 (Table 6), the upper-tail critical value from the F distribution with 1 degree of freedom in the numerator and 18 degree of freedom in the denominator. Because $F = 0.71 < F_u = 5.32$, or, from Table 5 and 6, because the p-value $= 0.42 > 0.05$, we do not reject null hypothesis and we conclude that there is insufficient evidence of an interaction between sampling station and weather condition. The focus is now on the main effects.

Table 6: ANOVA

Source of variation	SS	df	MS	F	P-value	F crit
Sample	1409594.00	1	1409594.00	372.6449	5.38E-08	5.317655
Columns	23674.08	1	23674.08	6.25856	0.036844	5.317655
Interaction	2700.00	1	2700.00	0.713781	0.422737	5.317655
Within	30261.38	8	3782.673			
Total	1466229.00	11				

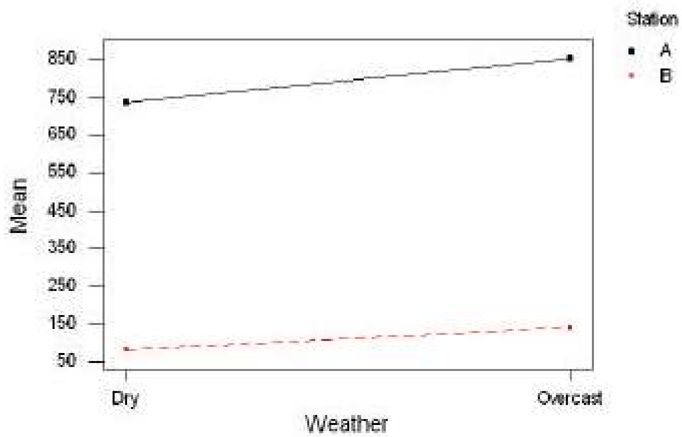

Fig. 5: Interaction plot – data means for TSS

In testing at the 0.05 level of significance for a difference between the two sampling station (factor A), the decision rule is to reject the null hypothesis if the calculated F value exceeds 5.32, the upper-tail critical value from the F distribution with 1 degree of freedom in the numerator and 18 degree of freedom in the denominator. Because $F = 372.65 > F_u = 5.32$, or, from Table 6, because the p-value = 0.00 < 0.05, we reject null hypothesis and conclude that there is evidence of a difference between the two sampling station in term of the average amount of TSS. Sampling station A (Labohan Dagang) is more TSS was observed (an average of 854.13 mg/L) than sampling station B (Kg. Jenderam, Bukit Changgang and Kg Bukit Dugang) (an average of 138.67 mg/L).

In terms of the factors in this study, if there were no interaction between sampling stations and weather condition, any difference between sampling station A and B would be the same under conditions of dry season as it is under conditions of overcast season. In the Table 5 and 6, for dry season, station A is 655.47 mg/L above station B (735.30 compare to 79.83); for overcast season, Station A is 715.46 mg/L above station B (854.13 compare to 138.67). The concept of interaction can be illustrated graphically by plotting the average values for each sampling station for each weather condition obtained from Table 5 and 6. From Fig. 5, we note that the difference between station A and B is larger for overcast season than for dry season. In the analysis, it's clearly shown that the test for the interaction found to be insignificant. Therefore the difference between the sampling stations for each weather condition is considered to be a sample effect or due to chance.

CONCLUSION

In conclusion, this study demonstrates that simple chemometrics treatments are able to draw out from raw data, information that would enable us to more effectively determine the "right" sampling sites for a particular objective, in order to reduce cost and time. In the case of the data obtained in the study, in order to determine the effects of palm oil plantation to water quality in the future, the researcher can determine the sampling sites in a more effective manner; relating the objective of the study to the types of sites to be chosen for sampling purposes. However data are needed for chemometrics analysis for future in process. Without historical data chemo metrics study would deem useless.

In this study, seasonal variation was found not to be the main separation factor. Thus, the initial sampling strategy used in order to study the effects of palm oil plantation as well as looking for seasonal changes at different sampling sites proves to be ineffective. The sampling sites chosen in plots 1, 2 and 3 prove to be redundant in this study and should be reassessed to give a more optimum number of sampling stations. The separation of sampling plots due to suspended solids and conductivity, if these were historically available for the studied area, should have been the significant factors to be taken into consideration in designing the initial sampling strategy. The abundance of historical data should be taken advantage of in designing these new sampling strategies. The use of chemometric methods, for example, should be encouraged in the analysis of the data that would bring about new information which will prove to be useful in reducing cost and time of sampling. The application of cluster analysis, followed by principal component analysis as a classification method, as demonstrated in this study, helps to separate differently polluted river sections and would help tremendously in future river pollution monitoring projects.

ACKNOWLEDGMENT

The Institute of Environment and Development (LESTARI), Universiti Kebangsaan Malaysia and Chemistry Department, Faculty of Science, University Malaya.

REFERENCES

1. Einax, J.W., D. Truckenbrodt and O. Kampe, 1998. River Pollution Data Interpreted of Chemometrics Methods. Microchem. J., 58: 315-324.
2. Pihak Berkuasa Perancang Tempatan Kuala Langat, Sepang and Putrajaya, 2000. Rancangan Struktur. Kuala Lumpur.
3. Massart, D.I., B.G.M. Vandeginste, L.M.C. Buydens, S.De Long, P.J. Lewis and J. Smeyers-Verbeke, 1997. Handbook of Chemometrics and Qualimetrics. Part A. Elsevier, Amsterdam.
4. Khalil, W.A., A. Goonetillke, S. Kokot and S. Carrol, 2003. Use of chemometrics methods and multicriteria decision-making for site selection for sustainable on-site sewage effluent disposal. Analytica Chimica Acta, 506: 41-56.
5. Abollina, O., M. Aceto, C.L. Gioia, C. Sarzanini and E. Mentasti, 2001. Spatial and seasonal variations of major, minor and trace elements in Antartic seawater. Chemometric investigation of variable and site correlations. Adv. Environ. Res., 6: 29-43.
6. Meng, A.K. and I.H. Suffet, 1997. Environ. Sci. Technol., 31: 337-345.
7. Mendiguchia, C., C. Moreno, D.M. Galindo-Riano and M. Garcia-Vargas, 2004. Using chemometric tools to assess anthropogenic effects in river water. A case study: Guadalquivir River (Spain). Analytica Chimica Acta, 515: 143-149.
8. Brodnjak-Voncina, D., D. Dobcnik, M. Novic and J. Zupan, 2002. Chemometrics characterization of the quality of river water. Analytica Chimica Acta, 462: 87-100.
9. Marengo, E., M.C. Gennaro, D. Giacosa, C. Abrigo, G. Saini and M.T. Avignone, 1995. How chemometrics can helpfully assist in evaluating environmental data Lagoon water. Analytica Chimica Acta, 317: 53-63.
10. Lohninger, H., 1999. Teach/Me, SDL-Software Development Lohninger Teach/Me DataLab 2.002. Springer. Berlin.
11. Forina, M., S. Lanteri and R. Todeschini, 1998. Chemometrics for Sampling and Analysis: Theory and Environmental Applications. Trend. Anal. Chemist., 3: 122-132.
12. Einax, J.W., D. Truckenbrodt and O. Kampe, 1998. River Pollution Data Interpreted by Means of Chemometric Methods. Microchem. J., 58: 315-324.
13. Brereton, R.G., 2002. Chemometrics: Data Analysis for the Laboratory and Chemical Plant, John Wiley and Son Ltd., West Sussex, England.
14. Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. American Stat. Assoc., 58: 236-244.
15. Mohd, E.T. and K. Osman, 2008. Use of InforWork RS in modeling the impact of urbanization on sediment yield in Cameron Highlands. Proceeding. 10th Annual Institute of Engineers, Malaysia. Universiti Kebangsaan Malaysia Publisher, pp: 21-30.
16. Levine, M.D., P.P. Ramsey and R.K. Smidt, 2001. Applied Statistics For Engineers And Scientists: Using Microsoft Excel and MINITAB. Int. Edition, Prentice-Hall Int. Inc.