



## CHEMOMETRIC ANALYSIS OF SELECTED ORGANIC CONTAMINANTS IN SURFACE WATER OF LANGAT RIVER BASIN

(Analisis Kemometrik Bahan Cemar Organik Terpilih dalam Air Permukaan Lembangan Sungai Langat)

Mohamad Rafaie Mohamed Zubir , Rozita Osman, Norashikin Saim\*

Faculty of Applied Sciences,  
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

\*Corresponding author: noras691@salam.uitm.edu.my

Received: 9 December 2014; Accepted: 9 October 2015

### Abstract

Chemometric techniques namely hierarchical agglomerative cluster analysis (HACA), discriminant analysis (DA), principal component analysis (PCA) and factor analysis (FA) were applied to the distribution of selected organic contaminants (polycyclic aromatic hydrocarbons (PAHs), sterols, pesticides (chlorpyrifos), and phenol) to assess the potential of using these organic contaminants as chemical markers in Langat River Basin. Water samples were collected from February 2012 to January 2013 on a monthly basis for nine monitoring sites along Langat River Basin. HACA was able to classify the sampling sites into three clusters which can be correlated to the level of contamination (low, moderate and high contamination sites). DA was used to discriminate the sources of contamination using the selected organic contaminants and relate to the existing DOE local activities groupings. Forward and backward stepwise DA was able to discriminate two and five organic contaminants variables, respectively, from the original 13 selected variables. The five significant variables identified using backward stepwise DA were fluorene, pyrene, stigmastanol, stigmasterol and phenol. PCA and FA (varimax functionality) were used to identify the possible sources of each organic contaminant based on the inventory of local activities. Five principal components were obtained with 66.5% of the total variation. Result from FA indicated that PAHs (pyrene, fluorene, acenaphthene, benzo[a]anthracene) originated from industrial activity and socio-economic activities; while sterols (coprostanol, stigmastanol and stigmasterol) were associated to domestic sewage and local socio-economic activities. The occurrence of chlorpyrifos was correlated to agricultural activities, urban and domestic discharges. This study showed that the application of chemometrics on the distribution of selected organic contaminants was able to trace the sources of contamination in surface water.

**Keywords:** chemometrics, organic contaminants, chemical markers, Langat River

### Abstrak

Teknik kemometriks iaitu Analisis Kelompok Agglomeratif Berhierarki (HACA), Analisis Diskriminan (DA), Analisis Komponen Utama (PCA) dan Analisis Faktor (FA) telah diaplikasi ke atas taburan bahan pencemar organik terpilih (aromatik hidrokarbon polisiklik (PAH), sterol, racun perosak (klorpirifos), dan fenol) untuk menilai potensi penggunaan bahan pencemar organik ini sebagai bahan pengesanan di Lembangan Sungai Langat. Sampel air diambil dari bulan Februari 2012 hingga Januari 2013 pada setiap bulan di sembilan stesen pemantauan sepanjang Lembangan Sungai Langat. HACA dapat mengelaskan kawasan persampelan kepada tiga kelompok yang boleh dikaitkan dengan tahap pencemaran (pencemaran rendah, pencemaran sederhana dan pencemaran tinggi). DA telah digunakan untuk membezakan sumber – sumber pencemaran dengan menggunakan bahan – bahan organik terpilih dan mengaitkannya dengan aktiviti – aktiviti setempat berdasarkan DOE. Analisis Diskriminan Langkah Hadapan dan Belakang masing – masing dapat mendiskriminasikan dua dan lima pembolehubah bahan pencemar organik daripada 13 pembolehubah asal yang dipilih. Lima pembolehubah yang penting adalah fluorena, pyrena, stigmastanol, stigmasterol dan fenol. PCA dan FA telah digunakan untuk mengenalpasti sumber – sumber bagi setiap bahan pencemar organik berdasarkan inventori aktiviti setempat. Lima komponen utama telah diperolehi dengan 66.5 % daripada jumlah variasi. Hasil

daripada FA menunjukkan PAH (pirena, flurena, asenaphtena, benzo[a]antrasena) berpunca dari kegiatan industri dan aktiviti sosio-ekonomi; manakala sterol (coprostanol, stigmastanol dan stigmasterol) daripada kumbahan domestik dan aktiviti – aktiviti sosio-ekonomi setempat. Kehadiran klorpirifos telah dikaitkan dengan aktiviti pertanian, bandar dan pelepasan domestik. Kajian ini menunjukkan bahawa teknik kemometriks yang diaplikasikan ke atas taburan bahan pencemar organik terpilih mampu mengesan punca pencemaran dalam air permukaan.

**Kata kunci:** kemometriks, bahan pencemar organik, penanda kimia, Sungai Langat

### Introduction

Malaysia is a developing country whereby industrialisation and urbanisation becomes the main programme. Due to the rapid growth of industrialisation and urbanisation, Malaysia is facing a number of problems especially water pollution is brought about by discharges of all kinds of effluents, particularly from domestic sewage, industries, and construction sites during dry and wet deposition from the atmosphere. As a result, a wide range of organic contaminants are recognized as environmentally relevant for their potential adverse effects on human and ecosystem health [1]. Thus, organic contaminants have become one of the rising issues in the field of environmental research. Environmental monitoring studies usually resulted in huge amounts of physical parameters and chemical concentrations collected in various geographical sites, time periods or environmental compartments (surface water, air, sediments and biota) [2].

In order to reveal relevant patterns and sources of variation in these large environmental datasets, chemometric analysis has been proposed [2,4]. Massart et al. [3] defined chemometrics as a chemical discipline that uses mathematics, statistics and formal logic to design or select optimal experimental procedures; to provide maximum relevant chemical data and to obtain knowledge about chemical systems. Chemometric techniques is very useful in tracing the sources and fate of the contaminants in the environment. Application of different pattern recognition techniques to reduce the complexity of large datasets have been proven to give a better interpretation and understanding of pollution source. Chemometric techniques covered three main areas of interest in environmental studies which are quantitative analysis, monitoring for environmental quality assessment and modeling or prediction of toxicology effects [2]. Commonly used chemometric techniques used for clustering are the hierarchical agglomerative cluster analysis (HACA) and principal components analysis (PCA) with factor analysis (FA). The three techniques of chemometric analysis were often used to interpret the data usually the data obtained had multivariate nature also several of variables were correlated [9].

Many studies have used chemometric techniques to identify the sources and distribution of organic (such as PAHs and sterols) or inorganic (such as heavy metal) contaminations in environmental compartments such as water, air and sediment [4–8]. The application of chemometric techniques would be helpful in the interpretation of complex datasets and gives a better understanding into the sources of contamination. Previous study by Osman et al. [4] reported the application of this techniques provide a more objective interpretation of the results and able to apply organic contaminants as a chemical markers. Thirteen organic contaminants were analyzed which include polycyclic aromatic hydrocarbons (PAHs), sterols, pesticide (chlorpyrifos) and phenol to evaluate as they are potential chemical marker in the Langat River Basin. The dataset covers the period from February 2012 to January 2013 and the water samples were taken at nine monitoring sites provided by Department of Environment (DOE), Malaysia. The applications of chemometric techniques in this study will be able to give valuable information on the distribution and source of contamination of organic contaminants. This approach may reveal the possible sources of organic contaminants along Langat River Basin, Malaysia.

### Materials and Methods

#### Chemicals and reagents

Solvents of gas chromatography (GC) grade or higher (methanol, dichloromethane, n-hexane and acetone) were purchased from Merck (Darmstadt, Germany). Hydrochloric acid (fuming 37% pro analysis) was obtained from Merck (Darmstadt, Germany). Individual standards PAHs; acenaphthene and pyrene were purchased from Dr. Ehrenstorfer GmbH (Augsburg, Germany) and acenaphthylene, fluorene, benzo[a]anthracene and benzo[a]pyrene were obtained from Supelco (Bellefonte, USA). Individual standard of sterols; 5 $\beta$ -cholestan-3- $\beta$ -ol (coprostanol), 5-cholesten-3 $\beta$ -ol (cholesterol), stigmasterol and stigmastanol were purchased from Sigma-Aldrich (Steinheim,

Germany) while  $\beta$ -sitosterol was purchased from Supelco (Bellefonte, USA). Phenol was purchased from Supelco (Bellefonte, USA). Chlorpyrifos (PESTANAL® 98%) and N,O bis(trimethylsilyl)trifluoroacetamide + trimethylchlorosilane (TMCS), (99:1) were purchased from Sigma-Aldrich (Steinheim, Germany). SPE cartridges, C<sub>18</sub> (EC) (1000 mg) and LiChrolut (EN 200 mg) were obtained from International sorbent technology (IST) (Mid Glamorgan UK) and Merck (Darmstadt, Germany) respectively.

### Instrumentation and equipments

Water samples were extracted using solid phase extraction (SPE) using SPE Vacuum Manifold (Supelco, St. Louis, USA). Water samples were filtered prior to SPE extraction by using 0.45  $\mu$ m glass fiber filter (Whatman International Ltd Maidstone, England). The pH of sample was measured using Metler Toledo AG, Analytical, (Switzerland) pH meter. Organic contaminants were analysed using gas chromatography (HP6890 series II, Agilent Technologies Palo Alto, CA, USA) with flame ionisation detector (GC-FID) (detection of PAHs sterols), gas chromatography (7890A Agilent Technologies) with electron capture detector (GC-ECD) (detection of chlorpyrifos), gas chromatography (Agilent GC6890/5973MSD) with mass spectrometry detector (GC-MSD) (detection of phenol).

### Monitoring sites

Nine monitoring sites (Table 1 and Figure 1) were selected from a list (24 sites) of monitoring sites of the Department of Environmental (DOE), Malaysia. Sampling of water samples was conducted along Langat River Basin from February 2012 to January 2013. The selected monitoring sites represented the sites that experienced domestic sewage discharges, socio-economic and industrial activities.

Table 1. Langat River Basin monitoring sites

River	Station No	Coordinate		Location
		Longitude	Latitude	
Langat	1L01	E 101° 47.030'	N 02° 57.835'	West country Land
Batang Labu	1L02	E 101° 45.014'	N 02° 47.241'	Near Salak Tinggi
Semenyih	1L03	E 101° 50.870'	N 02° 56.772'	Semenyih Town
Batang Benar	1L04	E 101° 49.457'	N 02° 50.898'	Batang Benar village
Rinching	1L05	E 101° 51.765'	N 02° 55.122'	Rinching Ulu village
Balak	1L06	E 101° 46.065'	N 02° 00.934'	Simpang Balak Bridge
Beranang	1L07	E 101° 49.737'	N 02° 53.297'	Bridge at Kampung Kuala Pajam
Buan	1L08	E 101° 47.371'	N 02° 51.874'	Nilai Utara R&R area
Batang Nilai	1L09	E 101° 48.653'	N 02° 48.844'	Infront of Hualon

### Extraction of organic contaminants using tandem solid phase extraction (SPE)

Extraction of PAHs, sterols, chlorpyrifos and phenol were done using tandem solid phase extraction procedure [10]. Two cartridges namely, C<sub>18</sub> and LiChrolut EN were used to extract selected organic contaminants from water samples. PAHs, sterols and chlorpyrifos were extracted using non-polar sorbent (C<sub>18</sub>) while phenol extracted using polymeric sorbent (LiChrolut EN). The SPE cartridges (C<sub>18</sub> and LiChrolut EN) were conditioned separately. The C<sub>18</sub> cartridge (1000 mg) was conditioned by using 10 mL of methanol followed by 6 mL of deionized water without applying vacuum while LiChrolut EN cartridge (200 mg) was conditioned using 10 mL of acetone followed with 3 mL methanol and 3 mL of acidified deionized water (pH 2) and also without applying vacuum. For sample loading, both cartridges C<sub>18</sub> and LiChrolut EN were arranged in tandem whereby C<sub>18</sub> cartridge was placed on top of the LiChrolut EN sorbent. Then, one litre (1 L) of water sample was loaded into the column using vacuum manifold at a flow rate of 6 mL min<sup>-1</sup>. During the process, the column was not allowed to dry. After sample loading, the two

cartridges were separated. Elution of interference and elution of analytes were done separately. Each cartridge was vacuum dried for 30 min before the removal of interference matrix. For C<sub>18</sub> cartridge, interferences were eluted by using 10 mL deionized water, while for LiChrolut EN cartridge the interferences were removed by using 5 mL 40% aqueous methanol. Both cartridges were vacuum dried. For C<sub>18</sub> cartridge, the compounds of interest were eluted by gravity. First elution using 6 mL of n-hexane eluted PAHs and chlorpyrifos while second elution using 2×3 mL dichloromethane eluted sterols. For LiChrolut EN cartridge, phenol was eluted by gravity using 2×3 mL dichloromethane. Nitrogen was used to blow down both extract to a final volume of 1 mL for gas chromatographic analysis.

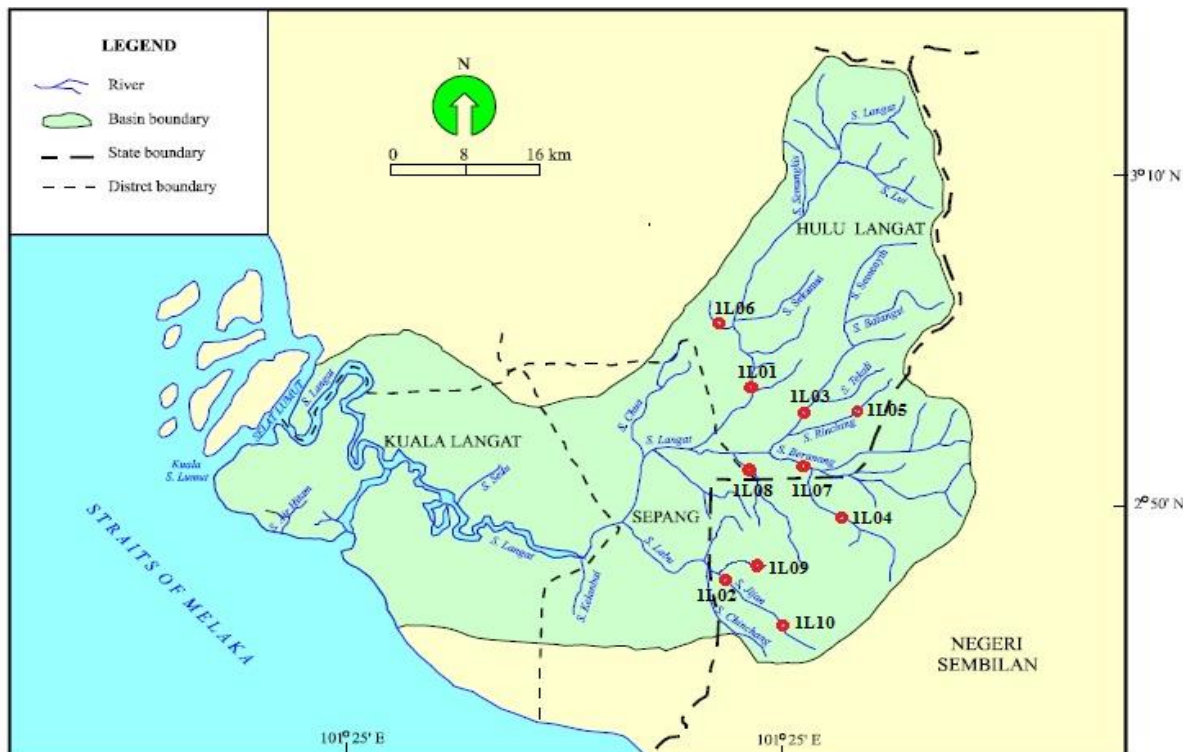


Figure 1. Monitoring sites in Langat River Basin (Source: Juahir *et al.* [13])

### Chromatographic analysis

Gas chromatography with flame ionization detector (GC-FID) was used for the separation and quantification of PAHs and sterols. Chlorpyrifos and phenol were analysed using gas chromatography with electron capture detector (GC-ECD) and chromatography with mass spectrometry detector (GC-MSD), respectively [10].

### GC-FID analysis

Separation and identification of PAHs and sterols was performed using gas chromatography with splitless injection and flame ionization detector. A 30 m × 0.25 mm id × 0.25 µm film thickness HP5-MS capillary column (Agilent technologies) was used to achieve separation of PAHs and sterols with the following temperature program: initial temperature, 50°C; held for 2 min; increased by 18°C min<sup>-1</sup> to 250°C, increased by 10°C min<sup>-1</sup> to 310°C; held for 11 min. The detector temperature was set at 310°C. PAHs quantification was carried out using a five-point calibration plot containing 1.00, 2.00, 3.00, 4.00 and 5.00 mg/L PAH standard mixtures. Sterols quantification was carried out using five-point calibration plot containing 1.00, 2.00, 3.00, 4.00 and 5.00 mg/L sterols standard. Prior to injection, sterols were derivatised by treatment with BSTFA: TMCS, (99:1) at 60°C for one hour.

### **GC-ECD analysis**

Separation of chlorpyrifos was achieved using gas chromatograph equipped with  $^{63}\text{Ni}$  electron capture detectors (GC-ECD). A  $30\text{ m} \times 0.25\text{ mm id} \times 0.25\text{ }\mu\text{m}$  film thickness HP5-MS capillary column (Agilent technologies) was used for the quantitative analysis of chlorpyrifos. Injection port and detector temperatures were set at  $250^{\circ}\text{C}$ . The injection volume was  $1\text{ }\mu\text{L}$ , with 2 min splitless period following the injection. The ECD detector utilized pure  $\text{N}_2$  (>99.999%) and the carrier gas and make-up gas was at a controlled constant velocity of  $60\text{ mL min}^{-1}$ . The temperature program was set to  $150^{\circ}\text{C}$  for 1 min, then increased by  $25^{\circ}\text{C min}^{-1}$  to  $260^{\circ}\text{C}$  for 8 min. Compounds were identified based on the retention time of the standards and quantified by external standard calibration. Chlorpyrifos quantification was carried out using five-point calibration plot containing 1.00, 2.00, 3.00, 4.00 and 5.00 mg/L chlorpyrifos standard.

### **GC-MSD analysis**

Analysis of phenol was performed with gas chromatography with mass spectrometry detector (GC-MSD). Phenol was separated using a cross-linked fused silica capillary column HP5-MS ( $30\text{ m} \times 0.25\text{ mm id} \times 0.25\text{ }\mu\text{m}$ ). The head pressure of the carrier gas (high purity helium) was 50 kPa. The temperature programmed was set at an initial temperature of  $50^{\circ}\text{C}$ , followed by an increased by  $10^{\circ}\text{C min}^{-1}$  to  $200^{\circ}\text{C}$  and held for 15 min. The MS detector was operated in the full scan using the electron impact (EI) ionization mode at 70 eV, scanning from 50-550 mass units at  $0.45\text{ s scan}^{-1}$ . Phenol was quantified by single point external standard calibration. Prior to injection, phenol were derivatised by treatment with BSTFA: TMCS, (99:1) at  $60^{\circ}\text{C}$  for one hour.

### **Chemometric analysis**

XLSTAT2013 software package was employed in the chemometric application using hierarchical agglomerative cluster analysis (HACA), discriminant analysis (DA) and principal component analysis (PCA) with the factor analysis (FA). These techniques were chosen to interpret and evaluate the organic contaminants dataset along Langat River Basin.

## **Results and Discussion**

### **Hierarchical agglomerative cluster analysis (HACA)**

HACA is a common method to classify variables or cases (observations/samples) into clusters with high homogeneity level within class and high heterogeneity level between classes [5,12]. HACA was performed on the selected organic contaminants data set to evaluate spatial variation among the nine monitoring sites. This analysis resulted in the grouping of sampling stations into three clusters as shown in Figure 2. Cluster 1 formed by five monitoring sites (1L01, 1L02, 1L03, 1L04 and 1L05) represents moderate contamination sites. 1L01, 1L03 and 1L04 monitoring sites located near city area which is experiencing urbanization over the past 15 years as several towns have grown and developed [13]. Possible sources of contaminations of these three monitoring sites are mainly from domestic discharges, and urban discharges. 1L02 and 1L05 monitoring sites located at rural areas but experiencing contamination from socio-economic activities and domestic sewage discharges. Cluster 2, formed by monitoring site (1L09) represents high contamination site located near the city and industrial area. This site is situated in a tributary of Batang Nilai, which carries high pollution flowing through the industrial areas [4]. Cluster 3 represents low contamination sites consisted of three monitoring sites (1L06, 1L07 and 1L08). These sites are located away from municipal discharges and industrial activities and correspond to socio-economic activities except 1L06 monitoring which is experiencing contamination from industrial activities and domestic sewage discharges. This analysis shows that HACA is able to classify the site based on the level of pollution and different possible sources of organic contaminant as the clusters in these groups have similar characteristic. Spatial variations of organic contaminants in water were further evaluated by using discriminant analysis.

### **Discriminant analysis (DA)**

DA was used to discriminate the sources of contamination using the selected organic contaminants and relate the sources to the existing DOE inventory. Figure 3 shows the plot of discriminant functions obtained from water samples and three group's sources of contamination formed. These groups were related to the inventory of local activities provided by DOE, Malaysia and the contamination of Langat River can divide into three sources: domestic sewage/urban discharges, industrial activities and socio-economic activities.

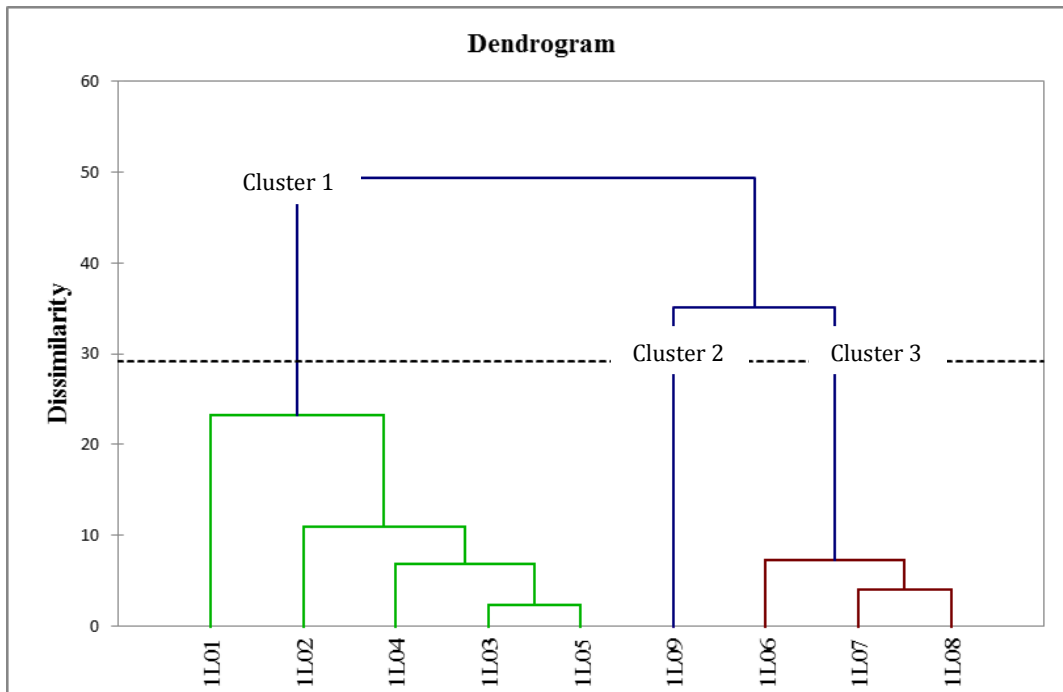


Figure 2. Dendrogram of clustering of monitoring sites according to organic contaminants analysed in water samples of Langat River Basin using Wards' method

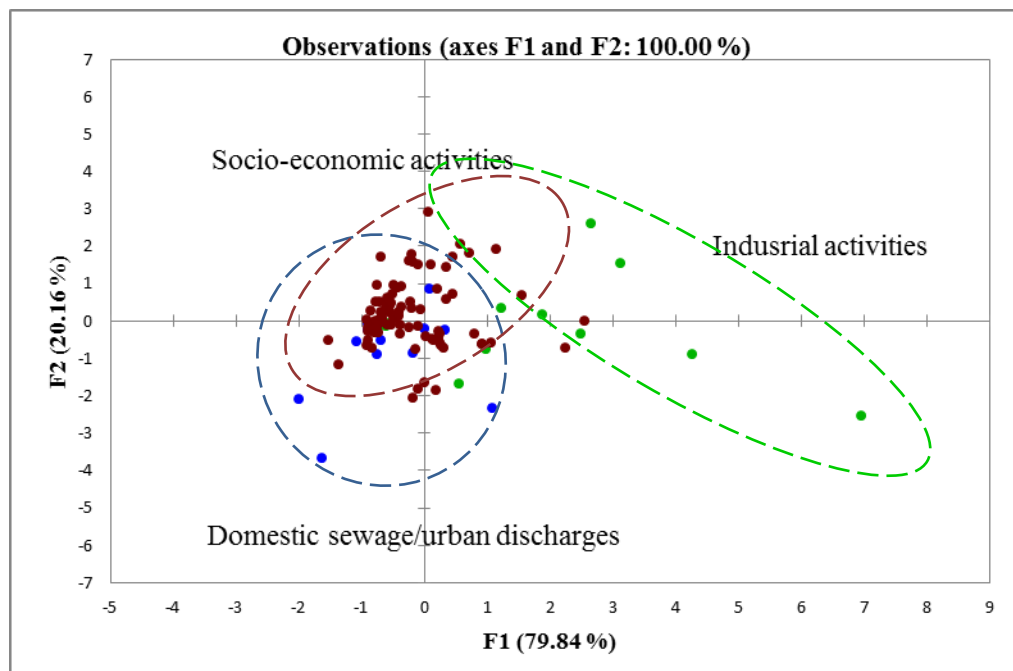


Figure 3. Plot of discriminant functions of water samples datasets

The three groups were treated as dependent variables, while the 13 selected organic contaminants were treated as independent variables. DA was carried out via three modes which are standard, forward stepwise and backward stepwise modes. Classification matrices (CMs) and discriminants functions (DFs) obtained from standard, forward stepwise and backward stepwise mode were tabulated in Table 2 and Table 3, respectively. The standard mode DA was able to discriminate the contamination source to 83.33 % correctly classified with 13 discriminant variables. The forward stepwise mode yielded 80.56 % correctly classified using only two discriminant variables (stigmastanol and stigmasterol), while the percent of cases correctly classified in backward stepwise mode also 79.63 % using only five discriminant variables (fluorene, pyrene, stigmastanol, stigmasterol and phenol). From the result of DA, the five organic contaminants based on backward stepwise DA were chosen as the significant variables to discriminate the sources contamination between the monitoring sites. Fluorene and pyrene were detected at relatively high concentration in the water samples. They were found high at socio-economic activities and industrial area. These contaminants represent the gasoline, coal and vehicular combustion [8,14,15]. Stigmastanol and stigmasterol were detected near domestic sewage area and urban discharges. Both compounds corresponded to faecal contamination and much related to the livestock farming activities [6,11]. Phenol clearly showed high concentration in wastewater of industrial area. This contaminant was associated with industrial activity. Industrial sources of this contaminant were from oil refineries, coal gasification sites and petrochemical units [16].

Table 2. Classification matrix of DA for all data measurements in water samples (n=108) from Langat River

Pollution source	% Correct	Source assigned by DA		
		Domestic sewage/Urban discharges	Industrial activities	Socio-economic activities
<b>Standard DA mode</b>				
Domestic sewage/Urban discharges	16.67	2	0	10
Industrial activities	50.00	0	6	6
Socio-economic activities	97.62	0	2	82
<b>Total</b>	<b>83.33</b>	<b>2</b>	<b>8</b>	<b>98</b>
<b>Forward DA mode</b>				
Domestic sewage/Urban discharges	0.00	0	0	12
Industrial activities	50.00	0	6	6
Socio-economic activities	96.43	0	3	81
<b>Total</b>	<b>80.56</b>	<b>0</b>	<b>9</b>	<b>99</b>
<b>Backward DA mode</b>				
Domestic sewage/Urban discharges	16.67	2	0	10
Industrial activities	50.00	0	6	6
Socio-economic activities	92.86	3	3	78
<b>Total</b>	<b>79.63</b>	<b>5</b>	<b>9</b>	<b>94</b>

Table 3. Classification function coefficients for discriminant analysis of organic contaminants in water samples from Langat River Basin (n=108)

Parameters	Standard mode			Forward stepwise mode			Backward stepwise mode		
	DS	IA	SA	DS	IA	SA	DS	IA	SA
Acenaphthylene	-2.345	-4.271	-2.752						
Acenaphthene	22.860	24.500	25.177						
Fluorene	-0.048	-2.395	-1.847				2.350	0.007	0.602
Pyrene	4.508	7.124	6.470				0.530	2.471	1.947
BaA	0.945	1.858	0.630						
BaP	-8.206	-10.950	-10.153						
Coprostanol	-12.844	-13.134	-11.928						
Cholesterol	-6.200	-5.168	-6.001						
Stigmasterol	2.303	5.586	2.541	1.765	5.293	2.461	2.438	6.065	2.928
Stigmastanol	62.518	81.888	65.450	42.867	57.942	43.832	45.460	62.664	46.856
β-sitosterol	0.000	0.000	0.000						
Chlorpyrifos	142.991	144.654	145.780						
Phenol	0.967	1.603	0.378				2.042	2.456	1.387
Constant	-38.678	-50.968	-38.471	-13.102	-23.000	-11.760	-15.637	-26.483	-13.644

DS: Domestic Sewage/Urban discharge, IA: Industrial activities SA: Socio-economic activities BaA=Benzo[a]anthracene; BaP=Benzo[a]pyrene

### Principal component analysis (PCA) with factor analysis (FA)

PCA was performed to identify the possible sources of organic contaminants in Langat River Basin. It provides information on the most significant variables, which explains the whole data set by excluding the less significant variables with minimum loss of original information [17]. Five PCs with eigenvalues larger than one were obtained through the analysis. In reference to eigenvalues, only the PCs with eigenvalues greater than one are considered to carry out the FA. Equal numbers of varifactors (VFs) were obtained via factor analysis performed on the PCs and comprised of 66.5% total variance. Table 4 shows the corresponding VFs, variables loadings and explained variance. From the four VFs, VF1 explains 19.9 % from the total variance, showing strong positive loading of coprostanol and moderate loading of stigmastanol. These contaminants were detected from sampling sites associated with domestic sewage, urban discharge and socio-economic activities. Coprostanol and stigmastanol were indicators for faecal contamination [11]. Coprostanol especially is one of general marker of anthropogenic faecal contamination [18]. VF2 explains 16.8% of total variance was dominated by strong positive loading of fluorene and pyrene whereas VF3 accounted for 11.5 % of the total variance, was predominated by moderate loadings of acenaphthene and benzo[a]anthracene. Inputs of the contaminants in VF2 and VF3 can be associated with industrial activities and urban activities along Langat River Basin. As reported by Li et al. [14, 19], the possible sources of pyrene and fluorene were from anthropogenic activities such as combustion. Acenaphthene and benzo[a]anthracene were associated with vehicle exhaust [8]. VF4 and VF5 accounted for 9.9 % and 8.5 % of the total variance, and showing strong positive loading of chlorpyrifos and stigmasterol, respectively. The presence of chlorpyrifos much related to contamination effects of socio-economic activity (agricultural activities such oil plantation) within Langat River Basin [20] whereas presence of stigmasterol usually indicator of faecal contamination and much related to the livestock farming activities along the river basin.



Table 4. Loadings of organic contaminants after varimax rotation (PC extracted 5 factors) of water samples from Langat River Basin

Parameters	VF1	VF2	VF3	VF4	VF5
Acenaphthylene	-0.382	0.235	0.525	0.426	0.248
Acenaphthene	0.120	-0.124	<b>0.713</b>	0.043	-0.207
Fluorene	-0.037	<b>0.779</b>	0.309	0.073	-0.011
Pyrene	-0.019	<b>0.900</b>	-0.077	-0.051	0.039
Benzo[a]anthracene	0.329	0.131	<b>0.740</b>	-0.106	0.173
Benzo[a]pyrene	0.444	0.496	-0.197	0.404	-0.181
Coprostanol	<b>0.815</b>	0.011	0.214	-0.047	0.135
Cholesterol	0.344	0.002	0.230	0.257	0.624
Stigmasterol	0.094	0.026	-0.087	0.012	<b>0.784</b>
Stigmastanol	<b>0.688</b>	-0.058	0.079	0.077	0.160
$\beta$ -sitosterol	0.000	0.000	0.000	0.000	0.000
Chlorpyrifos	0.011	-0.011	0.004	<b>0.892</b>	0.055
Phenol	0.088	0.397	-0.202	0.243	-0.436
Eigenvalue	2.390	2.012	1.377	1.189	1.014
Variability (%)	19.915	16.765	11.471	9.909	8.452
Cumulative (%)	19.915	36.681	48.152	58.060	66.513

(Strong loadings > 0.75 are shown in bold; moderate loadings 0.5-0.75 are in italic bold)

### Conclusion

The application of chemometric techniques provides valuable information in assessing the variation of organic contaminants in Langat River Basin. The sources of specific organic contaminants can be explained by correlating the groupings obtained through chemometric approach with local activities identified by DOE. HACA classified the sampling sites into three clusters which can be correlated to the sources of contamination. Forward and backward stepwise DA was able to discriminate two (stigmastanol and stigmasterol) and five (fluorene, pyrene, stigmastanol, stigmasterol and phenol) organic contaminants variables, respectively, from the original 13 selected variables. Result from PCA and FA, shows that the sources of contamination within the Langat River Basin were influenced by industrial, urban and socio-economic activities, represented by selected PAHs (pyrene, fluorene, acenaphthene and benzo[a]anthracene). Sources from domestic and urban discharges were represented by sterols (coprostanol, stigmastanol and stigmasterol). This study shows that the application of chemometric techniques can reveal meaningful information from the datasets of organic contaminant.

### Acknowledgement

The authors would like to acknowledge financial support obtained from Organisation for the Prohibition of Chemical Weapons (OPCW), Netherlands and Research Management Institute (RMI) Universiti Teknologi MARA (UiTM).

### References

- Burkhardt, M. R., ReVello, R. C., Smith, S. G. and Zaugg, S. D. (2005). Pressurized liquid extraction using water/isopropanol coupled with solid-phase extraction cleanup for industrial and anthropogenic waste-indicator compounds in sediment. *Analytica Chimica Acta*, 534(1): 89 – 100.
- Mas, S., de Juan, A., Tauler, R., Olivieri, A. C. and Escandar, G. M. (2010). Application of chemometric methods to environmental analysis of organic pollutants: A review. *Talanta*, 80(3), 1052 – 1067.

3. Massart, D. L., Vandeginste, B., Buydens, L., De Jong, S., Lewi, P. and Smeyers-Verbeke, J. (1997). *Handbook of chemometrics and qualimetrics: Part A*: Elsevier Science Pub Co.
4. Osman, R., Saim, N., Juahir, H. and Abdullah, M. P. (2012). Chemometric application in identifying sources of organic contaminants in Langat river basin. *Environmental Monitoring and Assessment*, 184(2):1001 – 1014.
5. 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.
6. Saim, N., Osman, R., Sari Abg Spian, D. R., Jaafar, M. Z., Juahir, H., Abdullah, M. P. and Ghani, F. A. (2009). Chemometric approach to validating faecal sterols as source tracer for faecal contamination in water. *Water Research*, 43(20): 5023 – 5030.
7. Dominick, D., Juahir, H., Latif, M. T., Zain, S. M. and Aris, A. Z. (2012). Spatial assessment of air quality patterns in Malaysia using multivariate analysis. *Atmospheric Environment*, 60: 172 – 181.
8. Retnam, A., Zakaria, M. P., Juahir, H., Aris, A. Z., Zali, M. A. and Kasim, M. F. (2013). Chemometric techniques in distribution, characterisation and source apportionment of polycyclic aromatic hydrocarbons (PAHS) in aquaculture sediments in Malaysia. *Marine Pollution Bulletin*, 69(1-2): 55 – 66.
9. Singh, K. P., Malik, A., Mohan, D., Sinha, S. and Singh, V. K. (2005). Chemometric data analysis of pollutants in wastewater – a case study. *Analytica Chimica Acta*, 532(1): 15 – 25.
10. Osman, R., Saim, N. and Abdullah, M. P. (2009). Simultaneously extraction of organic compounds with wide polarity range in water using solid phase extraction technique. *Research Journal of Chemistry and Environment*, 13(3): 7 – 18.
11. Osman, R., Saim, N. and Abdullah, M. P. (2010). Organic Contaminants In Soil/Sediment As A Tracer For Pollution Sources. *Chemical Sciences Journal*, 2010: 1 – 10.
12. McKenna, J. (2003). An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. *Environmental Modelling & Software*, 18(3): 205 – 220.
13. Juahir, H. (2008). Water Quality Data Analysis and Modeling at Langat River Basin. *PhD dissertation, Universiti Putra Malaysia*.
14. Li, J., Zhang, G., Li, X. D., Qi, S. H., Liu, G. Q. and Peng, X. Z. (2006). Source seasonality of polycyclic aromatic hydrocarbons (PAHs) in a subtropical city, Guangzhou, South China. *Science of Total Environment*, 355(1-3): 145 – 155.
15. Shen, Q., Wang, K. Y., Zhang, W., Zhang, S. C. and Wang, X. J. (2009). Characterization and sources of PAHs in an urban river system in Beijing, China. *Environmental Geochemistry and Health*, 31(4): 453 – 462.
16. Dabrowski, A., Podkoscielny, P., Hubicki, Z. and Barczak, M. (2005). Adsorption of phenolic compounds by activated carbon--a critical review. *Chemosphere*, 58(8): 1049 – 1070.
17. Kannel, P. R., Lee, S., Kanel, S. R. and Khan, S. P. (2007). Chemometric application in classification and assessment of monitoring locations of an urban river system. *Analytica Chimica Acta*, 582(2): 390 – 399.
18. Shah, V. G., Dunstan, R. H., Geary, P. M., Coombes, P., Roberts, T. K. and Von Nagy-Felsobuki, E. (2007). Evaluating potential applications of faecal sterols in distinguishing sources of faecal contamination from mixed faecal samples. *Water Research*, 41(16): 3691 – 3700.
19. Li, W. H., Tian, Y. Z., Shi, G. L., Guo, C. S., Li, X. and Feng, Y. C. (2012). Concentrations and sources of PAHs in surface sediments of the Fenhe reservoir and watershed, China. *Ecotoxicology Environment Safety*, 75(1): 198 – 206.
20. Juahir, H., Zain, S., Khan, R., Yusoff, M., Mokhtar, M. and Toriman, M. (2009). Using chemometrics in assessing Langat River water quality and designing a cost-effective water sampling strategy. *Maejo International Journal of Science and Technology*, 3(1): 26 – 42.