Jurnal
Teknologi

Full Paper

THE ASSESSMENT OF THE VARIATION OF PHYSICO-CHEMICAL SOURCES FOR DRINKING WATER QUALITY USING CHEMOMETRICS: A CASE STUDY AT WATER TREATMENT PLANTS IN KLANG VALLEY

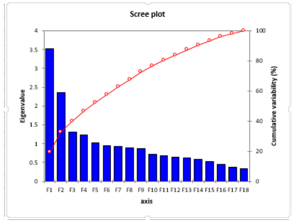
Article history
Received
4 March 2015
Received in revised form
14 June 2016
Accepted
18 October 2016

*Corresponding author
hafizanj@gmail.com

Manutha Appa Rwoo[a], Hafizan Juahir[a*], Nor Malisa Roslan[b], Mohd Ekhwan Toriman[a], Azizah Endut[a], Azman Azid[a], Che Noraini Che Hasnam[a], Ahmad Shakir Mohd Saudi[a], Mohd Khairul Amri Kamarudin[a], Sharmila Appa Rwoo[c], Ahmad Dasuki Mustafa[a]

[a]East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Kuala Terengganu, Terengganu, Malaysia
[b]Jabatan Kesihatan Negeri Selangor, Wisma Sunway Mas, Jalan Persiaran Kayangan, Seksyen 9, 40100 Shah Alam, Selangor, Malaysia
[c]Faculty of Education, Universiti Malaya, 50603, Malaysia

Graphical abstract



Abstract

This case study characterizes the drinking water quality by using the multivariate technique. The spatial variation of the physico-chemical and heavy metals parameters toxicity with the drinking water quality based on 28 water treatment plants in Selangor, Malaysia from 2009 to 2012 was evaluated. The objectives of this study are to analyze the physio-chemical activities and heavy metals activities in the collected drinking water samples from the treatment plants, and to detect the source of pollution for the most revealing parameters. The discriminant analysis (DA) and the principal component analysis (PCA) are the chemometric techniques used to investigate the spatial variation of the most significant physico-chemical and heavy metal parameters of the drinking water samples. The classification matrix accuracy for standard mode of DA, forward stepwise and backward stepwise for the physico-chemical and heavy metal parameters are excellent. PCA highlighted 13 significant parameters out of 18 physico-chemical water quality parameters and 14 significant parameters out of 16 heavy metal parameters. PCA was carried out to identify the origin and source of pollution of each water quality parameters. For that reason, this study proves that chemometric method is the principle way to explain the characteristic of the drinking water quality.

Keywords: Chemometric techniques, discriminant analysis, principal component analysis, drinking water quality

Abstrak

Kaji Selidik ini mengklasifikasikan kualiti air minuman dengan menggunakan teknik "multivariate". Kajian keragaman ruang parameter-parameter fizikokimia dan logam berat ini adalah berdasarkan 28 loji rawatan air di Selangor, Malaysia dari tahun 2009-2012. Objektif – objektif kajian ini adalah untuk menganalisis aktiviti-aktiviti parameter fizikokimia dan logam berat di dalam sampel air minuman yang diambil dari loji-loji rawatan air dan bagi mengenal pasti punca-punca pencemaran bagi parameter-parameter yang lebih menonjol. Analisis Diskriminan (DA) dan Analisis Komponen Utama (PCA) adalah teknik kemometrik yang digunakan untuk menyelidik keragaman ruang bagi parameter-parameter fizikokimia dan logam berat yang paling signifikan. Matrik Korelasi Ketepatan PCA telah menggariskan 13 daripada 18 parameter fizikokimia dan 14 daripada 16 parameter logam berat. PCA dijalankan untuk mengenalpasti sumber dan punca pencemaran bagi setiap parameter air minuman. Oleh yang demikian, kajian ini membuktikan bahawa kemometrik adalah cara yang efektif untuk menjelaskan ciri-ciri kualiti air minuman.

Kata kunci: Teknik Kemometrik, matrik korelasi, analisis diskriminan, analisis komponen utama, kualiti air minuman

1.0 INTRODUCTION

Water, the supporter of life, is also the bringer of death if there is a massive deterioration in its quality. Water is an essential nutrient for every living organism and forms as much as 90% of bodily fluids besides regulating the body temperature through perspiration. [1]. Recently, there are several water epidemics impact of water borne diseases caused by the poor water treatment system such as cholera and dysentery. Such epidemics impacts a big health risk to humans. The presence of pathogens in the water has caused more than 2 million deaths annually in nations suffering from poverty [2].The quality of drinking water for human consumption has been a growing concern for decades [3, 4]. In Malaysia, the Drinking Water Quality Surveillance Unit of the Ministry of Health Malaysia has prepared a set of guidelines for safe and portable water supply. This was performed under the guidance of experts from the World Health Organization (WHO) with a panel from the Public Works Department (PWD), Department of Chemistry (DOC) and Department of Environment (DOE). These panels were involved in the surveillance of drinking water quality after the National Guidelines for Drinking Water Quality 1983 was published. Even with the presence of modern treated water networks, there are still outbreaks of contamination at several stages of drinking water production. In other circumstances, raw water can be contaminated by an excess of pollutants. Such contamination is attributed to natural and anthropogenic factors [5, 6]. Furthermore, the quality of drinking water may also be compromised when there are certain process disruptions and calamities at the treatment plants [7]. As a result, the consumer's water distribution system may not be safe enough to sustain the high standards of the final product as established at the water treatment plant [8, 9]. Those pollution incidents need to be dealt with effectively and promptly to prevent harm to human health. Thus, an effective action has to be taken as soon as possible. Prevention should always come first because the degradation in water quality will multiply rapidly from the initial point to consumers [10]. Therefore, it is vital to ensure the water treatment and distribution network is well maintained [11, 12]. A comprehensive risk analysis is also dynamic to detect the critical points along the production stages, so that risks are lessened [13, 14, 15, 16].

In this study, the spatial analysis was done to reveal the most significant variables in the drinking water quality. However, poring through the huge data collection to identify the sources of pollutants (point source and non-point source) was quite a challenge. A comprehensive and correct data interpretation method is required to interpret the huge and complex raw data [17]. The research technique used in this study is called Chemometric. Chemometric is a branch of chemistry related to the analysis of chemical data (extracting information of data) and ensuring that experimental data contain maximum information (the designs of environment) [18]. Chemometric technique has been applied for the assessment of water quality from the natural reservoir to the tap of the consumer [19]. The application of such mathematical tools can aid in enhancing the safety of consumers in food quality[20]. The growing importance of multivariate statistical analyses such as discriminant analysis (DA) and principal component analysis (PCA) in environmental studies has been very helpful in measurements and monitorings [21]. Chemometric comprises of multivariate statistical modelling and data treatment [22]. Chemometric is also known as environmetric and also aids in evaluating multifaceted relationships with a wide scope of environmental applications namely agriculture, forestry, ecology, and environmental. [23, 24, 25] Chemometric has been widely used in analysing environmental data [24]. Chemometrics is also classified as the best way to avoid misinterpretation of a huge environmental monitoring data [26]. This technique is always applied in exploratory data analysis tools for the classification [27, 28] of samples (observations) or sampling stations and the identification of pollution sources [23, 29]. Chemometrics has also been practised to describe and analyse the water quality by validating spatial and temporal variations contributed by natural and anthropogenic factors [30, 31]. The usage of chemometrics such as PCA and DA were explored as means of differentiating the quality of drinking waters originating from the 28 water treatment plants in Selangor. The two multivariate techniques PCA and DA were applied in this study to achieve the objectives of the presented research. The first objective of this study is to analyze the physiochemical activities and heavy metals activities in the collected drinking water samples from the treatment plants. The next objective is to detect the source of pollution for the most revealing parameters.

2.0 METHODOLOGY

2.1 Study Area

Selangor is a state located in the west coast of Peninsular Malaysia and covers 8000 square kilometres extended along the west coast of Peninsular Malaysia at the northern outlet of the Straits of Malacca. The geographical coordinates of this area are 3° 31' 11.5068'' N and 101° 32' 17.2176'' E. The advantageous geographic position and rich natural resources have made Selangor the most prosperous state in Malaysia. The raw water data were collected from 28 water treatment plants which are responsible for the treated water in Selangor, Malaysia. The evaluation of the physico-chemical levels and heavy metals levels of the drinking water quality in the water treatment plants is based on the secondary data from year 2009 to 2013. The locations of the water treatment plants in Selangor are shown in Figure 1.

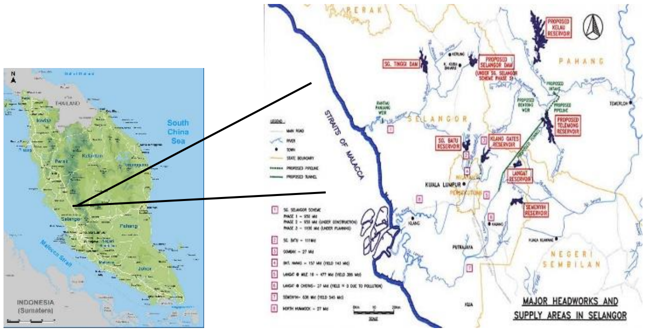


Figure 1 The Locations of Water Treatment Plants in Selangor

2.2 Data Collection

The drinking water quality data were obtained from the Public Health Division in the Selangor Department of Health (JKNS) Malaysia. The Sungai Selangor Water Supply Phase 3 (SSP3) Rasa water treatment plant, equipped with a total treatment capacity of 250 Million Litres Perday (MLD) serves the area in the northern region of Selangor. The other treatment plant, namely SSP3 Bukit Badong with the capacity to treat drinking water up to 800 MLD caters to the Klang Valley area. The Sungai Selangor Dam is a crucial part of the SSP3 in providing the additional water supply for 2 million residents and industries in Selangor and Klang Valley. All water treatment plants were identified based on the availability of data starting from January 2009 to December 2013. The data are classified under two types of water sample which are 'raw water' and 'cleaned water'.

The 34 water quality variables used in this study were categorized as physical-chemical parameters and heavy metal parameters. There are 18 physical-chemical parameters, namely; total coliform (TC) (cfu), *E. coli* (mg/l), pH, turbidity (NTU), colour (TCU), temperature (°C), total dissolved solid (mg/l), chloride (Cl) (mg/l), ammonia (NH3N) (mg/l), nitrate (NO3N) (mg/l), Iron (Fe) (mg/l), fluoride (FL) (mg/l), hardness (mg/), aluminium (Al) (mg/l), manganese (Mn) (mg/l), chemical oxygen demand (COD) (mg/l), biochemical oxygen demand (BOD) (mg/l) and free residual Cl (mg/l). There are 16 heavy metal parameters, namely; mercury (Hg) (mg/l), cadmium (Cd) (mg/l), arsenic (As) (mg/l), lead (Pb) (mg/l), chromium (Cr) (mg/l), copper (Cu) (mg/l), zinc (Zn) (mg/l), Sodium (Na) (mg/l), sulphate (SO4) (mg/l), selenium (Se) (mg/l), argentums (Ag) (mg/l), magnesium (Mg) (mg/l), chloroform(CHCl3) (mg/l),bromoform (CHBr3) (mg/l), dibromochloromethane (CHBr2Cl) (mg/l), and trihalomethane (CHCl2Br)(mg/l). The secondary data on physicochemical levels of the drinking water quality of the water treatment plants were analysed.

2.3 Data Analysis

The purpose of PCA and DA using XLSTAT 2014 in this study is to distinguish between drinking waters originating from the 28 water treatment plants (WTPs) in the Selangor state. These multivariate methods predict the origin of the pollutants from the water sources in order to curb problems originating from WTPs.

2.4 Discriminant Analysis

DA was applied to spot the substantial variability among two or more naturally existing groups. If discriminant analysis is operational for a set of data, the classification table of correct and incorrect approximations will yield a high correct percentage. DA works on raw data and builds a discriminant function for each group as in Eq. (1):

$$f(G_i)=k_i+\sum_{k=0}^{n} w_{ij}p_{ij} \tag{2}$$

where i represents the number of groups (G) [34, 35]. Next, ki is the constant inherent to each group. The *n* is the number of parameters used to classify a set of data into a given group and wj is the weight coefficient, assigned by DA to a given selected parameter (pj).

In this study, DA was used to study the spatial variation between the different classes of drinking water such as raw water and cleaned water. The drinking water classes were used as dependent variables, while the physico-chemical parameters and heavy metal parameters as the independent variables.

Wilks' lambda is a test statistic used to investigate whether there are variations between the means of identified groups of subjects on a combination of dependent variables [36]. Wilks' lambda test was performed in this study to test whether the mean score of two groups, raw water and cleaned water, is the same across physico-chemical parameters and heavy metal parameters simultaneously.

2.5 Principal Component Analysis

Analyzing variables, one by one, from a vast trove of numbers has resulted in an inadequate amount of data since the correlation between variables were not measured [37]. The variables were known as principal components, which are linear combination of the original variables. PCA generates a group of new orthogonal variables with a linear arrangement of the original variables known as principal components (PC). The new axes lie along the directions of highest variance [38]. The PCA techniques extract the eigenvalues and eigenvectors of the covariance matrix of original variables and provides a clear view about the relationship of a big number of variables with adequate details [39]. PC provides facts on the significant parameters that describes the total data set affording data reduction without losing the original sources [39, 30]. The PC can be expressed as Equation 3:

$$zij = ai1x\ ij + ai2x2j + \ldots + aimxmj \tag{3}$$

where z is the component score, a is the component loading, x is the measured value of the variable, i is the component number, j is the sample number, and m is the total number of variables.

If the PCs generated by PCA cannot be analyzed, they will be rotated by varimax rotation. Varimax rotations are considered significant if applied to the PCs with eigenvalues more than one [40]. Varimax rotation generates a new group of variables called varimax factors (VFs). The varimax rotations attain the same numbers of varimax factors as the variables in accordance with general features and may comprise unobservable, hypothetical, and latent variables [41]. The VF coefficients with a correlation superior than 0.75 are known as "strong", 0.75-0.50 as "moderate"; and 0.50-0.30 as "weak" significant factor loadings [42].

PC was performed for interpretation by detecting the latent factors that influence each classes (Raw Water, and Cleaned Water). For this study, PCA was applied to the standardized data sets (34 variables) for two different classes.

3.0 RESULTS AND DISCUSSION

3.1 Spatial Variation of Physicochemical Parameters

The DA was applied to study the spatial variation of two drinking water groups in Selangor. The accuracy of spatial classification using standard, forward stepwise and backward stepwise mode for the physico-chemical parameters were 99.88%, 99.85% and 99.85%, respectively. The standard DA mode has 18 discriminant variables (Residual Cl, Fl, BOD, TC, E-Coli, Turbidity, Colour, pH, Cl, NH3-N, NO3-N, Fe, Temperature, TDS, hardness, Al, Mn and COD). The DA forward stepwise has showed 14 discriminant variables (Residual Cl, Fl, BOD, *E. coli,* pH, Temperature, TDS, Cl, NO3-N, Fe, hardness, Al, Mn, and COD). The DA backward stepwise has revealed 14 discriminant variables (E-Coli, pH, Temperature, TDS, Cl, NO3-N, Residual Cl, Fe, hardness, Al, Mn, BOD and COD). This shows that the classification matrix accuracy was excellent. Table 1 shows the classification matrix by DA for spatial variations of physicochemical parameters in the drinking water in Selangor.

The Wilk's Lambda test for standard mode gave a Lambda value of 0.195 and $p < 0.0001$. The null hypothesis (H0) reveals that the means of vectors of the two classes are equal. The alternative hypothesis on the other hand reveals that at least one of the means vector is different from another. Since the computed p-value is lower than the significance level alpha=0.95, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha [24]. The risk to reject the null hypothesis H0 while it is true is lower than 0.01%. Thus, the two clusters are different from each other.

Table 1 Classification matrix by DA for spatial variations of physico-chemical parameters in the drinking water in Selangor

	Regions assigned by DA		
Sampling Regions	**% Correct**	**Raw**	**Treated**
Standard DA mode			
Raw	99.88	1657	2
Treated	99.88	2	1671
Total	99.88	1659	1673
Forward Stepwise Mode			
Raw	99.82	1656	3
Treated	99.88	2	1671
Total	99.85	1658	1674
Backward Stepwise Mode			
Raw	99.82	1656	3
Treated	99.88	2	1671
Total	99.85	1658	1674

The accuracy of spatial classification using standard mode of DA, forward stepwise and backward stepwise for the heavy metals parameters were 100%, respectively. The standard DA mode has 15 discriminant variables (Hg, As, Pb, Zn, Cr, Na, SO4, Se, CHBr3, and CHBr2Cl). The DA forward stepwise mode and backward stepwise revealed seven discriminant variables (As, Cr, Na, SO4, Se, CHCl3, CHBr2Cl). This shows that the classification matrix accuracy is excellent. Table 2 shows the classification matrix by DA for spatial variations of heavy metal parameters in the drinking water in Selangor.

The Wilk's Lambda test for standard mode gave a Lambda value of 0.075 and $p < 0.0001$. The null hypothesis H0 reveals that the means of vectors of the 2 classes are equal. The alternative hypothesis on the other hand reveals that at least one of the means vector is different from another. Since the computed p-value is lower than the significance level alpha=0.95, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha [24]. The risk to reject the null hypothesis H0 while it is true is lower than 0.01%. Thus the 2 clusters are different from each other.

Table 2 Classification matrix by DA for spatial variations of heavy metal parameters in Selangor.

	Regions assigned by DA		
Sampling Regions	**% Correct**	**Raw**	**Treated**
Standard DA mode			
Raw	100.00	516	0
Treated	100.00	0	479
Total	100.00	516	479
Forward Stepwise Mode			
Raw	100.00	516	0
Treated	100.00	0	479
Total	100.00	516	479
Backward Stepwise Mode			
Raw	100.00	516	0
Treated	100.00	0	479
Total	100.00	516	479

3.2 Source Identification of the Monitoring Area.

Figure 2 highlights all the 12 out of 18 physico-chemical parameters quality variables used in this study that satisfies the 0.50 factor loadings threshold. These pollutants are then classified as the most contributing pollutants in the selected monitoring stations in Selangor. For the qualitative evaluation of clustering behaviour PCA with varimax rotation for physico chemical parameters are revealed in Table 3.

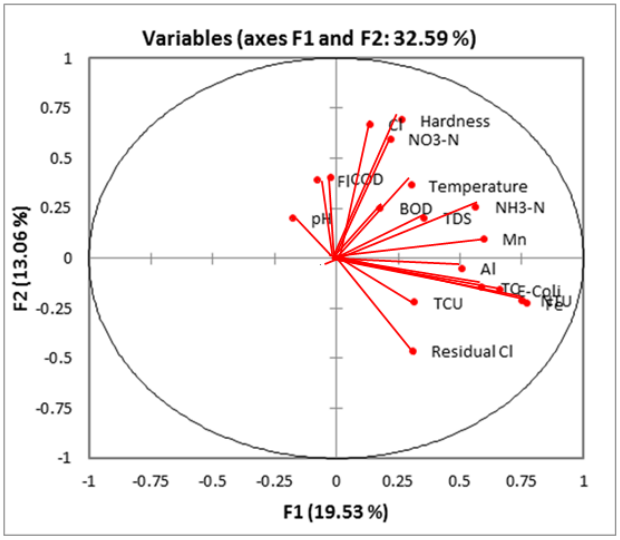


Figure 2 Factor loading plot after varimax rotation for physico-chemical parameters

The PCA of these data indicate their association and grouping with seven factors in drinking water. The total cumulative for the seven factors in drinking water was 62.86%. VF1 contributed 13.21% to the total variance with a high loading on TC (0.85), *E. coli* (0.85) and Turbidity (0.60) as shown in Table 3. TC and *E. coli* levels are indicated due to the presence of microorganisms. Frequently, these pathogens derived from water contaminated with human waste [22].

Table 3 Factor loading for selected physico-chemical parameter in drinking water

	V1	V2	V3	V4	V5	V6	V7
TC	**0.851**	0.018	0.026	-0.025	-0.029	-0.019	0.019
E. coli	**0.847**	0.026	-0.036	0.045	0.015	-0.039	0.094
NTU	**0.602**	0.023	-0.125	**0.500**	0.203	-0.008	0.010
TCU	0.055	-0.124	0.040	**0.745**	0.088	0.058	-0.169
pH	-0.051	0.059	0.028	-0.015	-0.014	**0.954**	0.042
Residual Cl	0.141	-0.109	-0.669	0.236	-0.061	-0.084	-0.011
Temperature	0.056	0.436	0.275	0.436	-0.156	-0.209	-0.013
TDS	-0.057	0.355	-0.149	0.462	-0.127	-0.157	0.269
Cl	-0.018	**0.653**	0.212	-0.035	-0.010	0.070	0.103
NH3-N	0.272	0.444	-0.290	0.087	0.181	-0.063	0.372
NO3-N	0.013	**0.749**	-0.049	-0.065	0.063	0.059	-0.091
Fe	0.497	0.041	-0.243	0.482	0.329	-0.185	0.058
Fl	-0.162	0.147	0.352	-0.221	**0.653**	0.013	-0.044
Hardness	0.055	**0.747**	0.134	0.008	0.058	0.081	0.071
Al	0.196	-0.027	-0.073	0.288	**0.736**	-0.017	0.106
Mn	0.384	0.306	-0.207	0.114	0.342	-0.127	0.088
COD	0.014	0.077	**0.740**	0.139	0.043	0.003	0.149
BOD	0.073	0.017	0.089	-0.062	0.038	0.055	**0.908**
Variability(%)	13.207	12.304	8.344	9.334	7.402	5.905	6.365
Cumulative(%)	13.207	25.511	33.855	43.189	50.591	56.496	62.861

VF2 accounts 12.30% of the total variance has a high loading on Cl (0.65), NO3-N (0.75) and hardness (0.75) as shown in Table 3. The sources of NO3-N is eutrophication of the dissolved nitrate can easily leach into surface and groundwater to be a major contaminant [41]. Cl in water is originated from natural sources, sewage, urban runoff, and industrial wastewater [43].

VF3 accounts 8.34% of the total variance has a high loading on COD (0.74). The COD level indicates the presence of biological waste such as dead leaves and animal waste [44].

VF4 accounts 9.33% of the total variance has a high loading on Turbidity (0.50) and Colour (0.75). The occurence of Turbidity is due to the presence of mineral salts compounds. The high loadings of turbidity and colour related to the discharge from urban development areas that involves the clearing of lands [45], soil erosion due to surface runoffs [46] and agricultural runoff [47].

VF5 accounts 7.40% of the total variance was highly correlated to Fl (0.65) and Al (0.74). The high level of Fl in drinking water could contribute to the utmost prevalence of dental fluorosis [48]. The Al level indicates the usage of Aluminium Sulphate (AlSO4) as coagulants to reduce the organic matter, colour, turbidity, and microorganism level in untreated waters [2].

The VF6 accounts 5.91% of the total variance was highly correlated to pH (0.95) resulted from the breakdown of water treatment plants [43].

The VF7 accounts 6.37% of the total variance was highly positively correlated to only BOD. This can be explained by the impact of biological contaminants from point sources such as discharge from wastewater treatment plants, domestic wastewater, and industrial effluents [44].

Figure 3 highlights all the 14 out of 16 heavy metal parameters quality variables used in this study, which satisfy the 0.50 factor loadings threshold. These pollutants are then classified as the most contributed pollutants in the selected monitoring stations in Malaysia. For the qualitative evaluation of clustering behaviour, PCA with varimax rotation for heavy metal parameters are revealed in Table 4.

For the heavy metal parameters, the VF1 accounts 13.40% of the total variance, was highly correlated to Se, $CHBr_3$ and $CHCl_3$. The concentrations of Se raise at high and low pH as a result of exchange into compounds of greater solubility in water [49]. The presence of $CHCl_3$ in drinking-water is via direct pollution of the source and the formation from naturally existing organic compounds during chlorination [26]. The $CHBr_3$ is detected in chlorinated drinking-water as a result of the reaction between chlorine, which is added during water treatment, and natural organic substances in the existence of bromide ion [50]. Exposure to $CHBr_3$ and $CHCl_3$ could be detrimental to health [51].

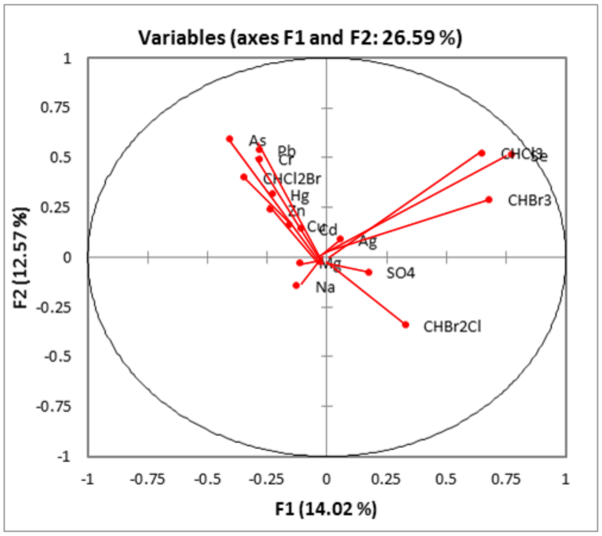


Figure 3 Factor loading plot after varimax rotation for heavy metals

The VF2 accounts 11.67% of the total variance, was highly correlated to As, Pb and CHCl2Br. The mobilization of As is dependent upon factors such as agricultural insecticides, larvicides, herbicides and wood preservatives, anthropogenic activities, weathering conditions, and redox conditions of water and soil. Approximately, 80% of the arsenic produced by humans is released to the water bodies of the environment in the form of impurities in pesticides [52]. Normally, Pb exists in suspended form in the river [53]. and erosion of natural deposits [54]. CHCl2Br are mobile in soils [54]. In addition to anthropogenic emission, CHCl2Br are formed due to the chlorination process as a disinfectant in municipal water supply systems in Malaysia [55].

VF3 accounts 8.06% of the total variance, was highly correlated to SO4 and CHBr2Cl. SO4 is originated from fertilizers, chemicals, dyes, glass, paper, soaps, textiles, fungicides, insecticides, astringents and emetics sulphate [54].The chlorination method applied at water treatment plants has resulted in the formation of chlorinated disinfection by-products (DBPs) such as CHBr2Cl [52].

VF4 accounts 8.28% of the total variance, was highly correlated to Hg and Cd. Hg contamination arises mainly from its industrial uses, such as the production of Hg cell batteries and Hg discharge lamps in the city area [41].The usage of pesticides or irrigation purpose contributes to the existence of Hg [55].

Higher level of cadmium may be found in the water near industrial areas or hazardous waste sites throughout the environment. The Environmental Protection Agency (EPA) has discovered that a lifetime exposure to cadmium has high tendency to cause health deficiencies like nausea, vomiting, diarrhoea, muscle cramps, salivation, sensory disturbances, liver injury, convulsions, shock and renal failure instantly [56].

VF5 accounts 7.59% of the total variance, was highly correlated to Cr and Cu. Usually, chromium concentrations in water are very low. Chromium is exposed to water bodies through anthropogenic sources such as electroplating factories, leather tanneries and textile manufacturing facilities [55]. The natural total chromium content in water is approximately 0.5-2 ppb [57]. Since, there is no heavy industry involving chromium in Malaysia, therefore chromium occurs in combination with other elements as chromium salts which dissolves in water [58]. The reduction process of chromium VI via organic matter produces chromium III. The presence of acid pH, alkaline pH or high-carbonate waters in distribution system often elevates the copper concentrations in drinking-water [58].

VF6 accounts 7.28% of the total variance, was highly correlated to Ag. Ag enters to the environment through anthropogenic activities such as manufacturing, household waste, agriculture, sewage, mining, and motor vehicle emissions [59, 60, 61, 62].

VF7 accounts 7.67% of the total variance, was highly correlated to Ag and Mg. Fertilizer usage and cattle feeds allows Mg to enter the environment. Decomposition of calcium and magnesium aluminosilicates and dissolution of limestone at high concentration causes Mg to be released to the environment [45].

Table 4 Factor loading for selected heavy metal parameters in drinking water

	V1	V2	V3	V4	V5	V6	V7
Hg	0.015	0.110	-0.225	**0.703**	0.062	-0.176	-0.002
Cd	-0.022	-0.078	0.072	**0.802**	0.008	0.097	-0.066
As	0.006	**0.770**	-0.060	0.036	0.073	-0.041	0.059
Pb	0.040	**0.686**	0.030	-0.088	0.119	0.427	0.056
Cr	0.027	0.365	-0.034	0.101	**0.664**	-0.021	-0.049
Cu	-0.050	-0.038	0.048	-0.010	**0.846**	0.005	0.069
Zn	-0.062	0.420	0.438	0.236	-0.044	-0.029	0.138
Na	-0.124	0.067	0.333	-0.011	-0.100	0.170	**0.691**
SO4	0.083	-0.009	**0.749**	-0.017	0.039	-0.134	0.014
Se	**0.933**	0.009	0.018	-0.029	-0.024	0.073	-0.021
Ag	0.033	0.037	-0.036	-0.007	-0.017	**0.892**	-0.010
Mg	0.007	-0.005	-0.132	-0.051	0.108	-0.100	**0.824**
CHCl3	**0.846**	0.093	-0.005	-0.010	-0.042	-0.110	-0.038
CHBr3	**0.721**	-0.190	0.069	0.053	0.059	0.102	-0.010
CHBr2Cl	0.040	-0.303	**0.567**	-0.324	0.057	0.177	-0.015
CHCl2Br	-0.068	**0.576**	-0.132	0.006	-0.010	-0.163	-0.170
Variability(%)	13.398	11.673	8.061	8.280	7.592	7.279	7.666
Cumulative(%)	13.398	25.070	33.131	41.412	49.004	56.283	63.949

4.0 CONCLUSION

Based on the prolonged observation of the concentration of the physico-chemical levels and heavy metals in the water distributed to the water supply in Selangor, the chemometric statistical technique helped to provide significant input on the spatial variability of a large and multifaceted drinking water quality data. Drinking water quality monitoring programs have generated huge, multidimensional and complex data set that requires chemometric techniques for data analysis and interpretation of the underlying information. In this study, the standard DA mode has 18 discriminant variables (Residual Cl, Fl, BOD, TC, *E. coli*, Turbidity, Colour, pH, Cl, NH3-N, NO3-N, Fe, Temperature, TDS, hardness, Al, Mn and COD). The DA forward stepwise has showed 14 discriminant variables (Residual Cl, Fl, BOD,E-Coli, pH, Temperature, TDS, Cl, NO3-N, Fe, hardness, Al, Mn, and COD). The DA backward stepwise has revealed 14 discriminant variables (E-Coli, pH, Temperature, TDS, Cl, NO3-N, Residual Cl, Fe, hardness, Al, Mn, BOD and COD). We applied the method of PCA to identify the pollution sources for drinking water quality variation in Selangor. The seven VFs generated by rotated PCA for the physico-chemical parameters indicated that the parameters of TC, E-Coli, Turbidity, COD, Colour, Fl, Al, pH and BOD are responsible for drinking water quality variations, which mainly from microorganisms, non-point source pollution, biological waste, mineral salts, acidity of water conditions and anthropogenic compounds. The seven VFs generated by rotated PCA for the heavy metal parameters indicated that the parameters of Se, $CHBr_3$, $CHCl_3$, As, Pb, $CHCl_2Br$, SO_4, $CHBr_2Cl$, Hg, Cd, Cr, Ag and Mg are responsible for drinking water quality variations, which mainly from the changes in pH, naturally existing from the organic compounds during chlorination, natural organic substances in the existence of bromide ion, weathering of rocks and minerals, anthropogenic compounds, corrosion of household plumbing systems, erosion of natural deposits, fertilizers, chemicals, dyes, glass, paper, soaps, textiles, fungicides, insecticides astringents and emetics sulphate, industrial uses, domestic waste and anthropogenic sources. Thus, this study indicated that for the future and effective management of the Malaysian drinking water quality, efforts should be placed as a priority in controlling point and non-point pollution sources by ensuring the human activities are in compliance with the environmental laws and legislations set by the Selangor Department of Environment (DOE) as well as the Ministry of Health Malaysia (MOH). Moreover, this research findings may serve as a reference for other related studies carried out in the future.

Acknowledgement

First and foremost, the authors would like to thank the Jabatan Kesihatan Negeri Selangor for providing us with the secondary data and valuable source of information.

References

[1] Baloch, M. K., Jan, I. and Ashour, S. T. 2000. Effect Of Septic Tank Effluents On Quality Of Ground Water. *Pakistan Journal Of Food Science*.10: 31-34.

[2] WHO. 2011. *Guidelines For Drinking-Water Quality*. 4th Edition. Vol. 1. Geneva.

[3] Rizak, S. and Hrudey, S. 2008. Drinking-water Safety-Challenges for Community Managed Systems. *Journal of Water Health*. 6: 33-41.

[4] Zhou, X., Liu, Y., Calvert, L., Munoz, C.G.W. Otim-Nape, D., Robinson, J. and Harrison, B. D. 1997. Evidence that DNA-A of a Geminivirus Associated With Severe Cassava Mosaic Disease In Uganda Has Arisen By Interspecific Recombination. *Journal of General Virology*. 78: 2101-2111.

[5] Brookes, J. D., Antenucci, J., Hipsey, M., Burch, M. D. N. Ashbolt, J. and Ferguson, C. 2004. Fate And Transport Of Pathogens In Lakes And Reservoirs. *Environ International*. 4(5): 741-759.

[6] Prudham, S. 2004. Poisoning The Well: Neoliberalism And The Contamination Of Municipal Water In Walterton, Ontario. *Geoforum*. 35(3): 343-359.

[7] Winston, G. and Leventhal. A. 2008. Unintentional Drinking-Water Contamination Events Of Unknown Origin:Surrogate For Terrorism Preparedness. *Journal of Water Health* 6 *(Suppl. 1)*. 11-19.

[8] Jakopanec, I., Borgen, K., Vold, L., Lund, H., Forseth, T., Hannula, R. and Nygård, K. 2008. A Large Waterborne Outbreak Of Campylobacteriosis In Norway: The Need To Focus On Distribution System Safety. *BMC Infect*. 128.10.1186/1471-2334-8-128.

[9] Van de Poel, E., O'Donnell, O. and Van Doorslaer, E. 2007. Are Urban Children Really Healthier? Evidence From 47developing Countries. *Social Science & Medicine*. 65(10): 1986-2003.

[10] Mahmud, S. G., Shamsuddin, S. A. J., Ahmed, M. F. Davison, A., Deere, D. and Howard, G. 2007. Development And Implementation Of Water Safety Plans For Small Water Supplies In Bangladesh: Benefits And Lessons Learned. *Journal of Water Health*. 5(4): 585-597.

[11] Huck, P. M. and Coffey, B. M. 2004. The Importance Of Robustness In Drinking-Water Systems. *Journal of Toxicology and Environmental Health*. *Part A*. 67: 1581-1590.

[12] Yanga, B., Uchidab, M., Kimc, H. M., Zhanga, X. and Kokubo, T. 2004. Preparation Of Bioactive Titanium Metal Via Anodic Oxidation Treatment. *Biomaterials*. 25: 1003-1010.

[13] Damikouka, I. and Tzia. K. C. 2007. Application of HACCP Principles In Drinking Water Treatment. *Desalination*. 210: 138-145.

[14] Hrudey, S. E., Hrudey, E. J. and Pollard, S. J. T. 2006. Risk Management For Assuring Safe Drinking Water. *Environ. International*. 32: 948-957

[15] Pollard, S. J. T., Estrutt, J. E., MacGillivra, B. H., Hamilton, P.D. and Hrudey, S. E. 2004. Risk Analysis And Management In The Water 28 Utility Sector-A Review Of Drivers, Tools And Techniques. Trans IChemE Part B: Process Saf. *Environ. Protect*. 82. 29(B6): 453-462.

[16] Yu, G., Sun, D. and Zheng, Y. 2007. Health Effects Of Exposure To Natural Arsenic In Groundwater And Coal In China: An Overview Of Occurrence. *Environmental Health Perspectives*. 115(4): 636-642.

[17] Chapman, M. S. and Verma. I. M. 1996. Transcriptional Activation By BRCA1. *Nature*. 382: 678-679.

[18] Rännar, S., Geladi, P., Lindgren, F. and Wold, S. 1995. A PLS Kernel Algorithm For Data Sets With Many Variables And Few Objects. Part II: Cross-Validation, Missing Data And Examples. *Journal of Chemometrics*. 9(6): 459-470.

[19] Astel, A., Tsakovski, S., Barbier, P. and Simenov, V. 2007. Multivariate Classification And Modelling In Surface Water Pollution Estimation. *Water Resources*. 41: 4566-4578.

[20] Georgieva, M., Ivana, N., Kiril, M., Nikolina, Y., Jasenka, G. K. and Zelimir, K. 2013. Application Of NIR Spectroscopy And Chemometrics In Quality Control Of Wild Berry Fruit Extracts During Storage. *Croation Journal of Food Technoloy. Biotechnology a Nutrition*. 8: 67-73.

[21] Osei, J., Nyame, F. K., Armah, T. K., Osae, S. K. Dampare, S. B., Fianko, J. R., Adomako, D. and Bentil, N. 2012. Application Of Multivariate Analysis For Identification Of Pollution Sources In The Densu Delta Wetland In The Vicinity Of A Landfill Site In Ghana. *J Journal of Water Resource and Protection*. 2(12): 1020-1029. doi:10.4236/jwarp.2010.212122.

[22] Revenga, C. and Mock, C. 2000. Dirty Water: Pollution Problems Persist. A Pilot Analysis of Global Ecosystems: Freshwater Systems, Earth Trends. World Resource Institute, Washington. 156p.

[23] 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 and Software*. 22: 464-475.

[24] Juahir, H., Zain, M.K., Yusoff, M. K., Hanidza, T. I. T., Armi, M. A. S., Toriman, E. and Mokthar, M. 2010. Spatial Water Quality Assessment Of Langat River Basin (Malaysia) Using Environmetric Techniques. *Environmental Monitoring Assess*. 173: 625-641.

[25] Feret, J. B., Francois, C., Gitelson, A., Asner, G. P., Barry, K. M., Panigada, C., Richardson, A. D. and Jacquemoud, S. Optimizing Spectral Indices And Chemometric Analysis Of Leaf Chemical Properties Using Radiative Transfer Modelling. *Remote Sensing Of Environment*. 115: 2742-2750.

[26] Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. and Kouimtzis, T. 2003. Asessment Of The Surface Water Quality In Northern Greece. *Water Research*. 37: 4119-4124.

[27] Brodnjak-Voncing, D., Dobcnik, D., Novic, M and Zupan, J. 2002. Chemometric Characterization Of The Quality Of River Water. *Anal Chim Acta*. 462: 87-100.

[28] Kowalkowski, T., Zbytniewski, R. and Szpeina, J. 2006. Application Chemometrics In River Water Classification. *Wat Res*. 40: 744-752.

[29] Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C. De Jong, S., Lewi, P. J. and J. Smeyers-Verbeke. 1997. *Handbook of Chemometrics and Qualimetrics: Part A. Data Handling in Science and Technology 20A*. Elsevier. 355-356.

[30] Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M. and Fernandez, L. 2000. Temporal Evaluation Of Groundwater Composition In An Alluvial Aquifer (Pisuerga River, Spain) By Principal Component Analysis. *Water Research*. 34: 807-816.

[31] Singh, K. P., Malik, A., Mohan, D., Sinha, S and Singh, V. K. Chemometric Data Analysis Of Pollutants In Wastewater – A Case Study. *Analytica Chimica Acta*. 532(1): 15-25.

[32] Azid, A., Juahir, H., Latiff, M.T., Zain, S. M. and Osman, M. R. 2013. Source Apportionment of Air Pollution: A Case Study in Malaysia. *Journal of Environmental Protection*. 4(12A).

[33] Bierman, P., Lewis, M., Ostendrof, B. and Tanner, J. 2009. A Review Of Methods For Analyzing Spatial And Temporal Patterns In Coastal Water Quality. *Ecological Indicators*. 103-114.

[34] Johnson, R. A. and Wichern, D. W. 1992. *Applied Multivariate Statistical Analysis.* Third Editon. Prentice Hall, Englewood Cliffs, New Jersey. 642 p.

[35] Wunderline, D. A., Diaz, M. D., Ame, V., Pesce, S., Hued, A, and Bistoni, M. 2000. Pattern Recognition Techniques From The Evaluation Of Spatial And Temporal Variations In Water Quality. A Case Study: Suquia River Basin (Cordoba-Argentina). *Water Research*. 35(12): 2881-2894.

[36] Polit, D. F. 1996. *Data Analysis and Statistics for Nursing Research*. Appleton and Lange. 320-321.

[37] Choudhary, R., Rawtani, P. and Vishwakarma, M. 2011. Comparative Study of Drinking Water Quality Parameters of three Manmade Reservoirs i.e Kolar, Kaliasote and Kerwa Dam. *Current World Environment*. 6(1): 145-149.

[38] Scatena, F. N., 2000. Drinking Water from Forests and Grasslands. *A Synthesis of the Scientific Literature*. 246p.

[39] Tadesse, D., Desta, A., Geyid, A., Girma, W., Fisseha, S. and Schmoll, O. 2010. Rapid Assessment Of Drinking Water Quality In The Federal Democratic Republic Of Ethiophia: Country Report Of The Pilot Project Implementation In 2004-2005. Geneva: World Health Organization & United Nations Children's Fund. 28-235.

[40] Kim, J. O. and Mueller, C. W. 1978. *Introduction To Factor Analysis: What It Is And How To Do It*. Newbury Park: Sage University Press.

[41] Vega, M., Pardo, R., Barrado, E. and Deban, L. 1998. Assessment Of Seasonal And Polluting Effects On The Quality Of River Water By Exploratory Data Analysis. *Jurnal of Sustainable Development*. 32: 3581-3592.

[42] Liu, C. W., Lin, K. H. and Kuo, Y. M. 2003. Application Of Factor Analysis In The Assessment Of Ground Water Quality In A Blackfoot Disease Area In Taiwan. *The Science of the Total Environment*. 313: 77-89.

[43] WHO. 2004. *Guidelines For Drinking-Water Quality*. 3rd edition. Vol 1.

[44] IARC On The Evaluation Of Carcinogenic Risks To Humans. *Chlorinated Drinking-water; Chlorination By- products; Some Other Halogenated Compounds; Cobalt and Cobalt Compounds*. Volume 52. World Health Organization International Agency For Research On Cancer.

[45] U.S. Geological Survey (USGS). 2007. Water Quality In The Upper Anacostia River, Maryland: Continuous And Discrete Monitoring With Simulations To Estimate Concentrations And Yields, 2003–05. Scientific Investigation Report 2007-5142.

[46] Goonetilleke, A., Thomas, E., Ginn, S., & Gilbert, D. 2005. Understanding The Role Of Land Use In Urban Stormwater Quality Management. *Journal of Environmental Management.* 74: 31-42.

[47] Schlosser, I. J. and Karr, J. R. 1981. Water Quality In Agricultural Watershed: Impact Of Riparian Vegetation During Base Flow. *Water Resources Bulletin*. 17: 233-240.

[48] Brijesh, M. K., Sunil, K. G. and Alok, S. 2014. Human Health Risk Analysis From Disinfection By-Products (Dbps) In Drinking And Bathing Water Of Some Indian Cities. *Journal of Environmental Health Science and Engineering*. 73.

[49] World Health Organization. 2011. *Guidelines For Drinking-Water Quality*. 4nd ed. Vol. 2.

[50] Leland, H. V. and McNurney, J. M. 1979. Lead Transport In A River Ecosystem. *Proc. Int. Conf. Transport of Persistent Chemicals in Aquatic Ecosystems III*. 17-23.

[51] Sukiman, S. and A. P. Md. 1993. Chemical Quality Of Malaysian Drinking Water Sources. *Drinking Water Health Aspects*. Edited by Jangi, M. S. Bangi: Universiti Kebangsaan Malaysia.

[52] Abdullah, J. 2014. Assessment of Potential Risks from Trihalomethanes in Water Supply at Alexandria Governorate. *J Pollut Eff Cont.* 2: 119. doi:10.4172/2375-4397.1000119.

[53] Abdul Aziz, H., Omran, A. and Zakaria, W. R. 2010. H_2O_2 Oxidation of Pre-Coagulated Semi Aerobic Leachate. *International Journal Environmental Research*. 4(2): 209-216.

[54] EPA. 1992. *Guidance for Nuisance Chemicals*. 92-001. Environmental Protection Agency, United States.

[55] Prasad, D. A. and Mishra, S. 2008. Hexavalent Chromium (VI): Environment Pollutant And Health Hazard. *Journal of Environmental Research a Development.* 2(3): 386-392.

[56] World Health Organization. 1996. *Guidelines For Drinking-Water Quality*. 2nd ed. Vol. 2.

[57] Soylak, M., Divrikii, U., Saracoglu, S. and Elci, L. 1998. Membrane Filteration – Atomic Absorption Spectrometry Combination for Copper, Cobalt, Cadmium, Lead and Chromium in Environmental Sample. *Environmental Monitoring and Assessment*. 127: 1-3.

[58] US EPA. 1995. Effect of pH, DIC, Orthophosphate Sulfate On Drinking Water Cuprosolvency. *US Environmental Protection Agency, Office of Research and Development* (EPA/600/R-95/085).

[59] Department of Environment. 2009. *Ministry of Natural Resources and Environment, Malaysia.* ISSN:0127-6433.

[60] Zulkifli, S. Z., Mohamat-Yusuff, F., Arai, T., Ismail, A. and Miyazaki, N. 2010. An Assessment Of Selected Trace Elements In Intertidal Surface Sediments Collected From The Peninsular Malaysia. *Environmental Monitoring and Assessment.* 169: 457-72.

[61] Yap, C. K. and Pang, B. H. 2011. Assessment of Cu, Pb and Zn Contamination In Sediment Of North Western Peninsular Malaysia By Using Sediment Quality Values Different Geochemical Indices. *Environmental Monitoring and Assessment*. 183: 23-39.

[62] Rwoo, M. A., Juahir, H., Azid, A., Sharif, S. M., Roslan, N. M. Zain, S. M. and Toriman, M. E. 2014. *Spatial Variations of Drinking Water Quality Monitoring in Water Treatment Plant Using Environmetric Techniques*. Springer Singapore. 325-329.