Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

ASSESSMENT OF DRINKING WATER QUALITY USING PRINCIPAL COMPONENT ANALYSIS AND PARTIAL LEAST SQUARE DISCRIMINANT ANALYSIS: A CASE STUDY AT WATER TREATMENT PLANTS, SELANGOR

M. A. Rwoo[1], H. Juahir[1,*], N. M. Roslan[2], A. Endut[1], M. K. A. Kamarudin[1] and M. A. Amran[1]

[1]East Coast Environmental Research Institute, Universiti Sultan ZainalAbidin, Gong Badak Campus, 21300 Kuala Nerus, Terengganu, Malaysia

[2]Selangor State Health Department, Wisma Sunway Mas, 40100 Shah Alam, Selangor, Malaysia

Published online: 08 August 2017

ABSTRACT

This study characterizes the drinking water quality on 28 water treatment plants in Selangor from 2009 to 2012 using multivariate techniques. The objectives of this study are to analyze the quality of collected drinking water and to detect the source of pollution for the most revealing parameters. The Partial Least Square Discriminant Analysis (PLS-DA) model showed a high correlation matrix of analysis for physicochemical quality of two types of water with 99.43% significant value. The classification matrix accuracy of the principal component analysis (PCA) highlighted 13 significant physico-chemical water quality parameters and 14 significant heavy metal parameters. PCA was carried out to identify the origin and source of pollution of each water quality parameters. Therefore, this study proves that chemometric method is the principle way to characterize the drinking water quality.

Keywords: partial least square, discriminant analysis; principal component analysis; drinking water quality.

Author Correspondence, e-mail: hafizanjuahir@unisza.edu.my

doi: http://dx.doi.org/10.4314/jfas.v9i2s.12

1. INTRODUCTION

The provisions of safe water supply and sufficient sanitation for sustaining good health are among the basic human rights [1]. Up to 90% of every human weight consist of water. Water serves a number of essential functions in human bodies [1]. The low quality of drinking water has directly affected the economy of the countries facing poverty. The widespread of problem due to water-borne diseases is becoming a serious global concern [2-3]. A set of guidelines for safe and portable water supply in Malaysia has been provided by the Drinking Water Quality Surveillance Unit of the Ministry of Health Malaysia. 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). After the National Guidelines for Drinking Water Quality 1983 was published the panels mentioned above were involved in the surveillance of drinking water quality. Although with the existence of modern treated water networks, there are occurance of deformities during certain stages of the production. In several cases the raw water could be polluted with a large number of contaminants due to natural and anthropogenic factors [4-5]. Certain process disruptions and calamities at the treatment plants could affect the drinking water quality [6]. As a consequence, the standards of water quality throughout the consumer's water distribution system are not at par with the final product attained at the water treatment plant [7-8]. Effective remedial measures has to be applied to avert further impact on human health. Immediate prevention steps should be taken to curb the degradation in water quality from the initial point to consumers [9]. Therefore, a good maintenance of water treatment plants and distribution network is very important [10-11]. A comprehensive risk analysis is needed in detecting the critical points along the production stages, so that risks are reduced [12-13].

In this study, the spatial analysis was done to reveal the most significant variables in the drinking water quality. A comprehensive and precise data interpretation method is required to analyze the large and complex raw data [14]. The technique applied in this research 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) [15]. Chemometric technique has been applied for the analysis of water quality from the natural reservoir to the tap of the consumer [16]. The application of such mathematical tools can help to promote the safety of consumers in food quality [17]. The growing importance of multivariate statistical analyses such as partial least squares discriminant analysis (PLS-DA) and principal component analysis (PCA) in environmental studies has been very useful in measurements and monitoring [18]. Chemometric consist of multivariate statistical modelling and data treatment [19]. 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 [20-22]. Chemometric has been diversely used in analyzing environmental data [21]. Chemometrics is also classified as the best way to avoid misinterpretation of a large environmental monitoring data [23]. This technique is always applied in exploratory data analysis tools for the classification of samples [24-25].

2. METHODOLOGY

Selangor is a state located in the west coast of Peninsular Malaysia and covers 8000 square kilometers 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 Fig. 1.

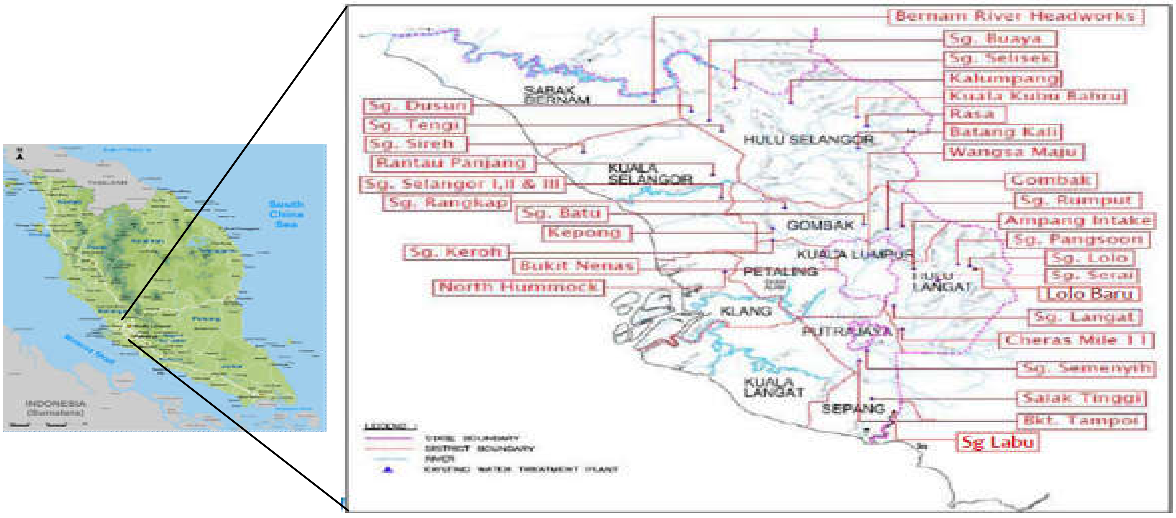

Fig.1. The locations of water treatment plants in Selangor, Malaysia

2.1. 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 liters per day (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 biochemical parameters and halogen and heavy metal parameters. There are 18 biochemical parameters namely total coliform (TC) (cfu), E. coli (mg/l), pH, turbidity (NTU), colour (TCU), temperature (°C), total dissolved solid (TDS) (mg/l), chloride (Cl) (mg/l), ammonia (NH3N) (mg/l), nitrate (NO3N) (mg/l), fluoride (F) (mg/l), hardness (mg/), chemical oxygen demand (COD) (mg/l), biochemical oxygen demand (BOD) (mg/l) and free residual Cl (mg/l).

There are 16 halogen and 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 ($CHBr_2Cl$) (mg/l) and trihalomethane ($CHCl_2Br$)(mg/l). The secondary data on biochemical and halogen and heavy metals levels of the drinking water quality of the water treatment plants were analyzed.

2.2. Data Pre-Treatment

A total of 55,440 data observations (34 variables x 1680 data set) were analyzed in this analysis. Only a small number of missing data were detected in the whole data. The total number of missing data in the data observations was very small (~3%) from the overall data. In order to facilitate the analyzing data, the nearest neighbour method was applied, which based on the endpoints of the gaps using Equation (1):

$$y = y1 \text{ if } x \leq x1 + [(x2 - x1)/2] \text{ or}$$
$$y = y2 \text{ if } x > x1 + [(x2 - x1)/2] \quad (1)$$

where y is the interpolant, x is the time point of the interpolant, y1 and x1 are the coordinates of the starting point of the gap, y2 and x2 are the endpoints of the gap.

2.3. Data Analysis

Descriptive statistics were calculated by using Excel 2013 (Microsoft Office). The multivariate statistical analysis such as PCA and PLS-DA were performed by using the XLSTAT 2014.The purpose of PCA and PLS-DA in this study is to distinguish the 34 drinking water parameters 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. 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 [26]. 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 [27]. 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 [28]. PC provides facts on the significant parameters that describes the total data set affording data reduction without losing the original sources [28-29]. The PC can be expressed as Equation (2):

$$zij = ai1x\ ij + ai2x2j + \ldots + aimxmj \quad (2)$$

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

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 [29]. 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 [30]. 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 [31]. PCA was performed for interpretation by detecting the latent factors that influences the drinking water quality in Selangor.

2.5. Partial Least Square Discriminant Analysis (PLS-DA) for Classifying Groups and Discriminating Problems

Partial least squares regression is a robust multivariate regression technique that supports a large number of analyses. Besides providing a rapid outline of the main systematic types of dissimilarities in the data from multifaceted systems, partial least squares regression aids to detect errors in the input data. PLSR is appropriate for selectivity enhancements of analytical devices [32]. Therefore, PLSR could be practiced as a multivariate calibration of one dependent variable against numerous independent variables. PLSR relates two data matrices, X and Y through a linear multivariate model. PLSR also models the structure of X and Y. The usefulness of PLSR derived from its tendency to analyze data with many noisy, collinear and in certain incomplete variables in both X and [33]. PLS regression used by Partial Least Square Discriminant Analysis (PLS-DA) functions in decreasing the dimension, redundancy and discrimination problems besides finding the latent variables [34].

PLS-DA were formed when the existence of X (dimensional I x J) as a centered matrix of predictive drinking water variables of the calibration set. Next, if g is a vector of I integer values which codes the qualitative groups to convey the group number related with the observation I and this g will signify the total number of qualitative groups [32]. The general primary model of multivariate PLS stated in Equation (3) and (4):

$$X = TPT + \mathrm{E} \quad (3)$$

$$Y = UQT + F \quad (4)$$

where X is n x p matrix of responses while Y is matrix of qualitative comprising the groups of drinking water stations. The application of PLS-DA in this drinking water quality study includes the drinking water quality variables. Advanced estimation of drinking water quality data is able to be attained by categorizing the interpretations run by PLS-DA. The reduction of redundant that contributes to disruptions and effects the non-precise values are portrayed by the percentage of corrects tested by PLS-DA.

3. RESULTS AND DISCUSSION

3.1. Source Identification of the Monitoring Area

Fig. 2 highlights all the 12 out of 18 biochemical 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 biochemical parameters are revealed in Table 3.

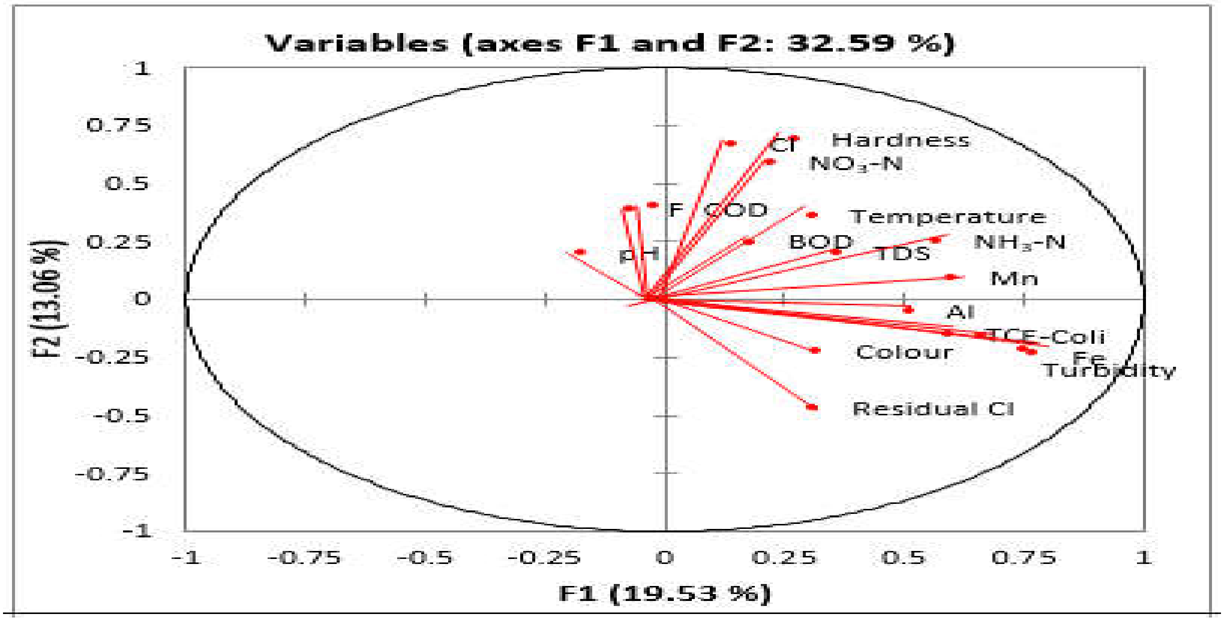

Fig.2. Factor loading plot after varimax rotation for biochemical parameters

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 [19]. 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 [30]. Cl in water is originated from natural sources, sewage, urban runoff, and industrial wastewater [35]. 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. VF4 accounts 9.33% of the total variance has a high loading on Turbidity (0.50) and Colour (0.75). The occurrence 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, soil erosion due to surface runoffs and agricultural runoff [36]. VF5 accounts 7.40% of the total variance was highly correlated to F (0.65). The high level of Fl in drinking water could be due to the existence of fluoride bearing minerals in the area [37]. The Al level indicates the usage of Aluminium Sulphate ($AlSO_4$) as coagulants to reduce the organic matter, colour, turbidity and microorganism level in untreated waters [38]. The high level of Flouride in drinking water could contribute to the utmost prevalence of dental fluorosis [37].

The VF6 accounts 5.91% of the total variance was highly correlated to pH (0.95) resulted from the breakdown of water treatment plants. 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 [37].

Fig. 3 highlights all the 14 out of 16 halogen and 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 halogen and heavy metal parameters are revealed in Table 2. For the halogen and heavy metal parameters, the VF1 accounts 13.40% of the total variance was highly correlated to Se, CHBr3 and CHCl3. The concentrations of Se raise at high and low pH as a result of exchange into compounds of greater solubility in water [39]. The presence of CHCl3 in drinking-water is via direct pollution of the source and the formation from naturally existing organic compounds during chlorination. The CHBr3 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. Exposure to CHBr3 and CHCl3 could be detrimental to health [35]. 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. The water retention in the pipelines causes the leached Pb to dissolved [39]. CHCl2Br are mobile in soils. In addition to anthropogenic emission, CHCl2Br are formed due to the chlorination process as a disinfectant in municipal water supply systems in Malaysia [35].

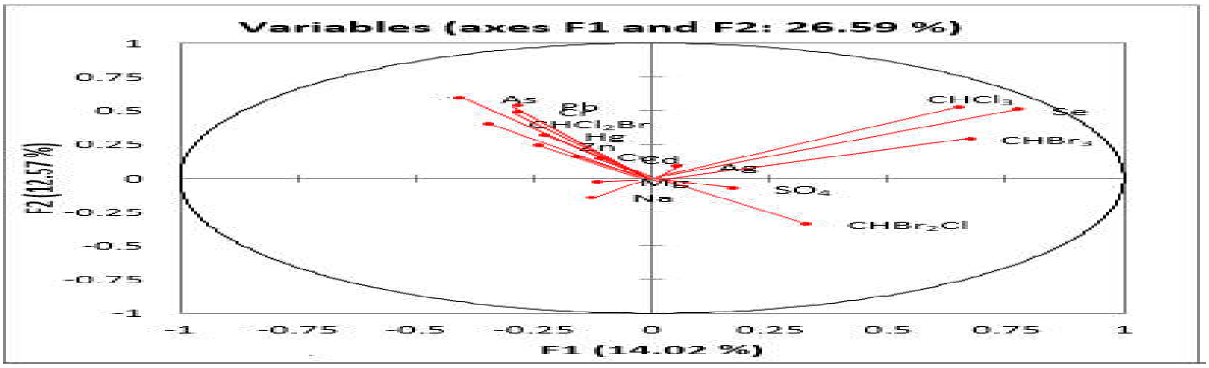

Fig.3. Factor loading plot after varimax rotation for halogen and heavy metals

Table 1. Factor loading for selected biochemical 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
Turbidity	0.602	0.023	-0.125	0.500	0.203	-0.008	0.010
Colour	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 C1	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
C1	-0.018	0.653	0.212	-0.035	-0.010	0.070	0.103
NH_3-N	0.272	0.444	-0.290	0.087	0.181	-0.063	0.372
NO_3-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
F	-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
A1	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

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 [37]. The chlorination method applied at water treatment plants has resulted in the formation of chlorinated disinfection by-products (DBPs) such as CHBr2Cl [35].

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 [40]. The usage of pesticides or irrigation purpose contributes to the existence of [41].

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, diarrhea, muscle cramps, salivation, sensory disturbances, liver injury, convulsions, shock and renal failure instantly [30].

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 [30].

The natural total chromium content in water is approximately 0.5-2 ppb. 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. 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 [30].

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 [43-47].

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 [48].

Table 2. Factor loading for selected halogen and 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**
SO_4	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**
$CHCl_3$	**0.846**	0.093	-0.005	-0.010	-0.042	-0.110	-0.038
$CHBr_3$	**0.721**	-0.190	0.069	0.053	0.059	0.102	-0.010
$CHBr_2Cl$	0.040	-0.303	**0.567**	-0.324	0.057	0.177	-0.015
$CHCl_2Br$	-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

3.2. PLS-DA

PLS-DA is the supervised statistical approach that can provide the variety and detailed variation assessment of biochemical sources for drinking water quality in Klang Valley. There are two major types of water quality present in this study which are Raw Water (Type-RAW) and Treated Clean Water (Type-TPO). Fig. 4 shows the halogen and heavy metal content in two types of water. The Type-RAW show that the high concentration of biochemical content are BOD, TDS, COD, Fe and Cl. Then, for Type-TPO the biochemical content detected are pH, Hardness, NO3-N and Temperature which are higher in concentration. While NTU, E. coli, Al, Mn and TC were detected moderately at this two types of water quality.

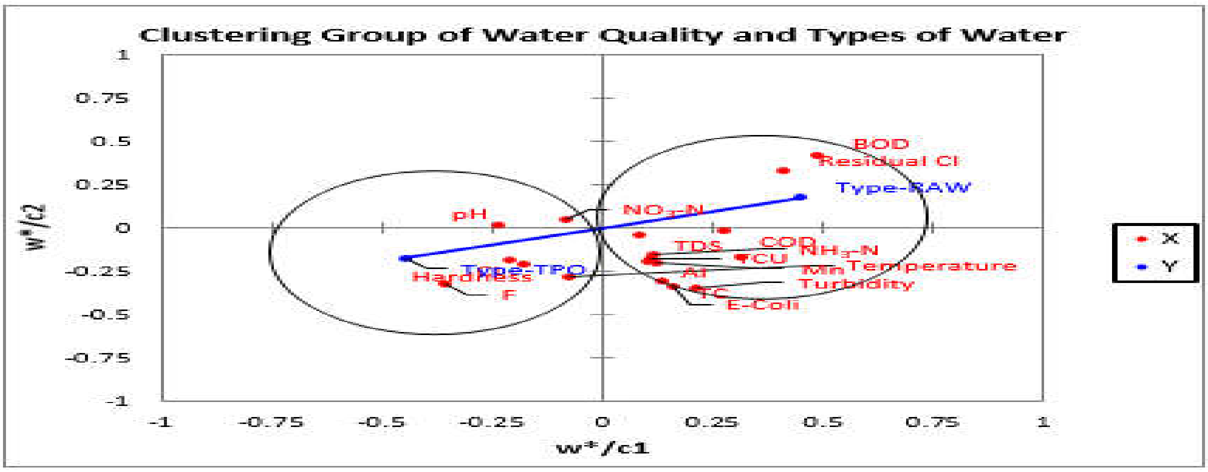

Fig.4. PLS-DA clustering water quality parameter and types of water

This shows the effectiveness of treatment process in reducing the biochemical contain from Type-RAW which are the source of intake water to be treated before supplied into consumers. The biochemical that highly detected by output at Type-TPO such as pH, Fl, Temperature and Hardness due to the chemical reaction used the reagent or catalyst for treated the raw water and preserved in water contained and pipelines. The reasons pH and Fl higher in Type-TPO in this region were to prevent the contamination and microorganism's growth while flowed into the pipelines and maintain the water quality supplied. Table 3 illustrates the high correlation matrix of analysis involved this two types of water biochemical quality. The percentages about 99.43% were the significant value to evaluate and monitor the quality of water for future monitoring activities due to the environmental changes and human activities that can affect the biochemical quality of water. The need of more concerned for Type-RAW (99.58%) for detection of concentration value for each water parameter for preparing the exact amount of reagent and catalyst used in treatment process in reducing and eliminating its harmful biochemical contains.

Table 3. Correlation matrix for type-raw and type-tpo

from \ to	Type-RAW	Type-TPO	Total	% correct
Type-RAW	1651	7	1658	99.58%
Type-TPO	12	1661	1673	99.28%
Total	1663	1668	3331	99.43%

4. CONCLUSION

Based on the prolonged observations of the concentration of the biochemical levels and halogens 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. The application of PLS-DA in this drinking water study has managed to distinguish the drinking water parameters. The PCA managed to reveal that the TC, E-Coli, Turbidity, COD, Colour, F, 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 statistical analysis shown a strong basis for the identification and classification of various sources of contamination and the correlation between the drinking water parameters and drinking water quality. Finally, it is concluded 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.

5. ACKNOWLEDGEMENTS

First and foremost, the authors would like to thank the Selangor State Health Department for providing us with the secondary data and valuable source of information

6. REFERENCES

[1] Baloch M K, Jan I, Ashour S T. Effect of septic tank effluents on quality of ground water. Pakistan Journal of Food Sciences, 2000, 10(3-4):31-34

[2] Rizak S, Hrudey S. Drinking-water safety-Challenges for community managed systems. Journal of Water Health, 2008, 6(S1):33-41

[3] Zhou X, Liu Y, Calvert L, Munoz C, Otim-Nape G W, Robinson D J, Harrison B D, Evidence that DNA-A of a Geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. Journal of General Virology, 1997, 78(8):2101-2111

[4] Brookes J D, Antenucci J, Hipsey M, Burch M D, Ashbolt N J, Ferguson C. Fate and transport of pathogens in lakes and reservoirs. Environment International, 2004, 30(5):741-759

[5] Prudham S. Poisoning the well: Neoliberalism and the contamination of municipal water Walterton, Ontario. Geoforum, 2004, 35(3):343-359

[6] Winston G, Leventhal A. Unintentional drinking-water contamination events of unknown origin: surrogate for terrorism preparedness. Journal of Water Health, 2008, 6(S1):11-19

[7] Jakopanec I, Borgen K, Vold L, Lund H, Forseth T, Hannula R, Nygård K. A large waterborne outbreak of campylobacteriosis in Norway: The need to focus on distribution system safety. BMC Infectious Diseases, 2008, 8(1):1-11

[8] Van de Poel E, O'Donnell O, Van Doorslaer E. Are urban children really healthier? Evidence from 47 developing countries. Social Science and Medicine, 2007, 65(10):1986-2003

[9] Mahmud S, Shamsuddin S A J, Ahmed M F, Davison A, Deere D, Howard G. Development and implementation of water safety plans for small water supplies in Bangladesh: benefits and lessons learned. Journal of Water and Health, 2007, 5(4):585-597

[10] Huck P M, Coffey B M. The importance of robustness in drinking water. Journal of Toxicology and Environmental Health, Part A, 2004, 67(20-22):1581-1590

[11] Yang B, Uchida M, Kim H M, Zhang X, Kokubo T. Preparation of bioactive titanium metal via anodic oxidation treatment. Biomaterials, 2004, 25(6):1003-1010

[12] Damikouka I, Katsiri A, Tzia C. Application of HACCP principles in drinking water treatment. Desalination, 2007, 210(1-3):138-145

[13] Pollard S J T, EStrutt J E, MacGillivra B H P, Hamilton D, Hrudey S E. Risk analysis and management in the water 28 utility sector-A review of drivers, tools and techniques. Process Safety and Environmental Protection, 2004, 82(6):453-462

[14] Yu G, Sun D, Zheng Y. Health effects of exposure to natural arsenic in groundwater and coal in China: An overview of occurrence. Environmental Health Perspectives, 2007, 115(4):636-642

[15] Chapman M S, Verma I M. Transcriptional activation by BRCA1. Nature, 1996, 382(6593):678-679

[16] Rännar S, Geladi P, Lindgren F, Wold S. A PLS kernel algorithm for data sets with many variables and few objects. Part II: Cross-validation, missing data and examples. Journal of Chemometrics, 1995, 9(6):459-470

[17] Astel A, Tsakovski S, Barbieri P, Simeonov V. Comparison of self-organizing map classification approach with cluster and principal components analysis for large environmental data sets. Water Research, 2007, 41(19):4566-4578

[18] Georgieva M, Ivana N, Kiril M, Nikolina Y, Jasenka G K, Zelimir K. Application of NIR spectroscopy and chemometrics in quality control of wild berry fruit extracts during storage. Croatian Journal of Food Technology, Biotechnology and Nutrition, 2013, 8(3-4):67-73

[19] Osei J, Nyame F K, Armah T K, Osae S K, Dampare S B, Fianko J R, Adomako D, Bentil N. Application of multivariate analysis for identification of pollution sources in the Densu Delta Wetland in the vicinity of a landfill site in Ghana. Journal of Water Resource and Protection, 2012, 2(12):1020-1029

[20] Revenga C, Mock G. Dirty water: Pollution problems persist. Pilot analysis of global ecosystems: Freshwater systems. Washington: World Resource Institute, 2000

[21] Shrestha S, Kazama F. Assesment of surface water quality using multivariate statistical techniques: a case study of the Fuji River Basin, Japan. Environmental Modelling and Software, 2007, 22(4):464-475

[22] Juahir H, Zain S M, Yusoff M K, Hanidza T T, Armi A M, Toriman M E, Mokhtar M. Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. Environmental Monitoring and Assessment, 2010, 173(1-4):625-641

[23] Feret J B, Francois C, Gitelson A, Asner G P, Barry K M, Panigada C, Richardson A D, Jacquemoud S. Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling. Remote Sensing of Environment, 2011, 115(10):2742-275

[24] Simeonov V, Stratis J A, Samara C, Zachariadis G, Voutsa D, Anthemidis A, Sofoniou M, Kouimtzis T. Assessment of the surface water quality in Northern Greece. Water Research, 2003, 37(17):4119-4124

[25] Brodnjak-Voncina D, Dobcnik D, Novic M, Zupan J. Chemometrics characterisation of the quality of river water. Analytica Chimica Acta, 2002, 462(1):87-100

[26] Choudhary P, Rawtani, Vishwakarma M. Comparative study of drinking water quality parameters of three manmade reservoirs i.e.Kolar, Kaliasote and Kerwa Dam. Current World Environment, 2011, 6(1):145-149

[27] Massart D. L., Vandeginste B. G., Buydens L. M., Lewi P. J., Smeyers-Verbeke J., Jong S. D. Handbook of chemometrics and qualimetrics: Part A. Amsterdam: Elsevier Science Inc., 1997

[28] Tadesse D, Desta A, Geyid A, Girma W, Fisseha S, Schmoll O. Rapid assessment of drinking-water quality in the Federal Democratic Republic of Ethiopia: Country report of the pilot project implementation in 2004-2005. Geneva: World Health Organization/United Nations Children's Fund, 2010

[29] Jae-On K., Mueller C. W. Introduction to factor analysis: What it is and how to do it. California: Sage Publications, 1987

[30] Alsaffar M S, Suhaimi J M, Ahmad K N. Evaluation of heavy metals in surface water of major rivers in Penang, Malaysia. International Journal of Environmental Sciences, 2016, 6(5):657-669

[31] Liu C W, Lin K H, Kuo Y M. Application of factor analysis in the assessment of ground water quality in a Blackfoot disease area in Taiwan. Science of the Total Environment, 2003, 313(1):77-89

[32] Martens H, Martens M. Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). Food Quality and Preference, 2000, 11(1):5-16

[33] Wold S, Sjostrom M, Eriksson L. PLS-regression: A basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 2001, 58(2):109-130

[34] Barker M, Rayens W. Partial least square for discrimination. Journal of Chemometrics, 2003, 17(3):166-173

[35] Mohammad A, Miri M, Ebrahimi A, Khorsandi H, Nemati S. Monitoring of THMs concentration in Isfahan water distribution system and zoning by GIS, a case study in the Centre of Iran. Iranian Journal of Health, Safety and Environment, 2016, 3(1):421-427

[36] Naubi I, Zardari N H, Shirazi S M, Ibrahim N F, Baloo L. Effectiveness of water quality index for monitoring Malaysian river water quality. Polish Journal of Environmental Studies, 2016, 25(1):231-239

[37] Alagumuthu G, Rajan M. Chemometric studies of water quality parameters of Sankarankovil block of Tirunelveli, Tamilnadu. Journal of Environmental Biology, 31(5):581-586

[38] Shah A, Khan M A, Kanwal N, Bernstein R. Assessment of safety of drinking water in tank district;an empirical study of water-borne diseases in rural Khyber Paktunkhwa, Pakistan. International Journal of Environmental Sciences, 2016, 6(4):418-428

[39] Hashim H, Yusop H M. Drinking water quality of water vending machines in Raja, Batu Pahat, Johor. In IOP Conference Series: Materials Science and Engineering, 2016, pp. 1-8

[40] Vega M, Pardo R, Barrado E, Deban L. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research, 1998, 32(12):3581-3592

[41] Prasad D A, Mishra S. Hexavalent Chromium (VI): Environment pollutant and health hazard. Journal of Environmental Research and Development, 2008, 2(3):386-392

[42] Nguyen N T, Hoang H A. Vulnerability of rice production in Mekong River delta under impacts from floods, salinity and climate change. International Journal on Advanced Science, Engineering and Information Technology, 2015, 5(4):272-279

[43] Zulkifli S Z, Mohamat-Yusuff F, Arai T, Ismail A N, Miyazaki N. An assessment of selected trace elements in intertidal surface sediments collected from the Peninsular Malaysia. Environmental Monitoring and Assessment, 2010, 169(1):457- 472

[44] Abdullahi M G, Kamarudin M K, Toriman M E, Gasim M B, Endut A, Garba I. Assessment of natural groundwater recharge in Terengganu, Malaysia. International Journal on Advanced Science, Engineering and Information Technology, 2016, 6(5):781-786

[45] Yap C K, Pang B H. Assessment of Cu, Pb, and Zn contamination in sediment of north western Peninsular Malaysia by using sediment quality values and different geochemical indices. Environmental Monitoring and Assessment, 2011, 183(1):23-39

[46] Wahab N A, Kamarudin M K, Gasim M B, Umar R, Ata F M, Sulaiman N H. Assessment of total suspended sediment and bed sediment grains in upstream areas of Lata Berangin, Terengganu. International Journal on Advanced Science, Engineering and Information Technology, 2016, 6(5):757-763

[47] Miller C V, Gutiérrez-Magness A L, Majedi B L, Foster G D. Water quality in the Upper Anacostia River, Maryland: Continuous and discrete monitoring with simulations to estimate concentrations and yields. Scientific Investigations Report 2007-5142, Virginia Geological Survey (US), 2007

[48] Lee M R, Mohamad A F. Assessment of Bekok River water quality status and its suitability for supporting the different uses: A review. International Journal on Advanced Science, Engineering and Information Technology, 2011, 1(6):672-674

How to cite this article:
M. A. Rwoo, H. Juahir, N. M. Roslan, A. Endut, M. K. A. Kamarudin, M. A. Amran.Assessment of drinking water quality using principal component analysis and partial least square discriminant analysis: A case study at water treatment plants, Selangor. J. Fundam. Appl. Sci., 2017, 9(2S), *157-173*.