Spatial Assessment of Selangor, Malaysia Water Treatment Plant Performance Using Chemometric Technique

H. M. Zolkipli[1], H. Juahir[1*], G. Adiana[1*], N. Zainuddin[2], A. B. H. M. Maliki[3], M. E. Toriman[4], M. Mokhtar[5], R. Elfitri[4]

[1]*East Coast Environmental Research Institute, Universiti Sultan ZainalAbidin, Gong Badak Campus, 21300 Kuala Nerus, Terengganu, Malaysia*
[2]*Ministry of Health, Engineering Services Division Department, Aras 3-7, Block E3, Kompleks E, Presint 1, Federal Government Administrative Centre, 62590 Putrajaya, Malaysia*
[3]*Faculty of Applied Society Science, Universiti Sultan ZainalAbidin, Gong Badak Campus, 21300 Kuala Nerus, Terengganu, Malaysia*
[4]*Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*
[5]*Institute for Environment and Development, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*
*Corresponding author E-mail: hafizanjuahir@unisza.edu.my, adiana.ghazali@gmail.com

Abstract

This study aims to identify the most significant parameters in drinking water quality, spatial disparities of treated water (TW) and performance of water treatment plant (WTP) in Selangor. Physico- chemical (PCPs), Inorganic (IPs), Heavy metal and organic (HMOPs) and pesticide (PPs) were selected as parameters to discriminate the source of WTP pollutant. Chemometric technique such as principle component analysis (PCA), one-way analysis of variance (ANOVA) and discriminant analysis (DA) was applied to validate the performance of water treatment plant. PCA identified the most significant parameters which are highlighted six out of eight parameters for PCPs, six out of twelve parameters for IPs, nine out of sixteen parameters for HMOPs and all seventh parameters for PP. ANOVA for distinguish two categories region in WTP and showed both of PCPs and IPs had significant differences due to their concentration ($p < 0.5$) and HMOPs suggested fifth of significant differences within regions ($p < 0.05$). PP doesn't give any significant differences ($p > 0.05$). DA was suggested PCPs, IPs and HMOPs in good performance (76.96%, 91.90% and 93.27%) except PP (50.43%). We can conclude that this chemometric technique can expose which area of WTP need to be properly maintains their performance to produce high quality of drinking water.

Keywords: Drinking water quality; Water treatment plant; Principal component analysis; One-way analysis of variance; Discriminant analysis.

1. Introduction

Water treatment plants (WTPs) in Selangor and was studied from chemical, biological, bacteriological, organic pollutants, and heavy metals aspects to determine water treatment plants performance, formed water quality and the engineering aspect [1]. WTPs is a place where water is treated to remove the sediment, bacteria and other contamination. This treatment is any process that improves the quality of water to make it more acceptable before distribute to the specific end-uses. There are many uses such as for drinking, water supply to the industrial, irrigation system, maintenances of river flow, water recreation or many other uses, together with being safely refunded to the environment. In this study, the end use is for drinking water or to use for cookery.

Acceptable of drinking water quality (DWQ) is when water through the WTP with the correct sequence flow process (raw water resources, screening, aeration, flocculation, sedimentation, filtration, disinfection, pH adjustment, fluoridation, clean water tank, water quality monitoring and water distribution system) and the most important thing is the drinking water parameters values should meet the permissible limit from drinking water quality standard (DWQS). Once the WTP process flow and DWQS

achieve, maintenance and analysis should be continuously maintaining to detailed assessment of plant performance and description of any compulsory changes [1]. In order to deliver a healthy lifestyle, National drinking water quality surveillance programmed (NDWQSP) was implemented in 1983 by Malaysian Ministry of health. The main objective of this program is to achieve the standard of health by confirming that safe drinking water is provided to the residents that fulfills with the required standards, therefore reducing the occurrence of water-borne diseases or intoxication associated with poor quality public water supplies. Various water quality (WQ) parameters have been measured and monitored throughout the Malaysian water supply system. The selected WQ parameters were classified into four groups namely Group 1: Physico-chemical parameters, Group 2: Inorganic parameter, Group 3: Heavy metals and organic parameter and Group 4: Pesticides. One of routine activity that was carried out by Engineering Services Divisions in this program is inspection of water treatment plants with high violations. The relevant departments should take necessary remedial actions to improve the water quality situation [2].

Millions of WQ data were collected and dealing with huge amount of the dataset is a thought-provoking task towards a better understanding of data exhibition since NDWQSP was implemented. Therefore, the application of chemometric technique is the good practice that can be practical in order to

investigate the performance of WTPs each group. In the chemistry and environmental science field, Chemometric technique is a statistical method has been practiced [3]. Principal component analysis (PCA), cluster analysis (CA) and discriminant analysis (DA), as well as factor analysis (FA) are extensively used for data analysis via chemometric techniques [4-5]. Multiuse of chemometric technique let the decrease of difficulty sum of dataset for the presentable data elaboration [6-7]. Present study has been conducted to identify the most significant parameters in drinking water quality and the spatial disparities of treated water (TW) to distinguish the parameters among these two regions and at last to investigate the performance of water treatment plants (WTPs) in Selangor which are take remedial action at certain WTPs that was suggested from chemometric technique analyzation.

2. Methodology

2.1. Location of Sampling Station

The sampling task was carried out in Selangor and Kuala Lumpur Water Treatment Plants (WTPs). There are thirty of WTPs in Selangor and three WTPs in Kuala Lumpur which are point at untreated water (UTW) and treated water (TW). Foto 1 showed sampling point in red colour words that is intake sampling station equal to UTW and treatment plant outlet station refer to TW in the range of WTP. The Selangor and Kuala Lumpur were located in the west region of the Malaysia Peninsular are in a residential and commercial area with is high density of population.

Fig. 1: Typical drinking water supply system in Malaysia

2.2. Drinking Water Quality Data

Drinking water quality data was obtained from Malaysian Ministry of Health at the selected sampling stations for this study. Five years data from 2012 until 2016 was recorded base on group parameters with each group sampling frequency. PCPs group data was recorded per week, IOPs group noted per month and group of HMOPs and PPs does sampling task every three month. In PCPs group include microbiological parameters that is total coliform and E. coli and physical parameters such as colour, pH, residual chlorine, temperature and conductivity. IOPs group with twelve parameters are total dissolve solids, chloride (Cl-), nitrogen ammonia (NH_3), nitrogen nitrate (NO_3), iron (Fe), fluoride (F-), hardness, aluminium (Al), manganese (Mn), chemical oxygen demand (COD), biological oxygen demand (BOD) and total organic carbon (TOC). Another group is HMOPs that was included eighteen parameters (mercury (Hg), cadmium (Cd), arsenic (As), cyanide (Cn), lead (Pb), chromium (Cr), copper (Cu), zinc (Zn), sodium (Na), sulphate (SO_4), selenium (Se), argentum (Ag), magnesium (Mg), mineral oil, chloroform ($CHCl_3$), bromoform ($CHBr_3$), dibromochloromethane

($CHBr_2Cl$) and bromodichloromethane ($CHBrCl_2$)). Last group is PPs that include seven parameters there are aldrin/dieldrin, DDT, heptachlor and heptachlor epoxide (H & He), methoxychlor, lindane, chlordane and endosulfan.

2.3. Pre-Treatment Data

Running the data analysis using chemometric technique need to clean up first in order to do the data observation. All missing data are required to be replaced is the data blank while removed the data were typing error and data which contains alphabet and symbol. It is a compulsory that, the pre-treatment data was done prior to the chemometric technique analysis. The highest of missing data are come from PCPs with 98.21% from 53, 784 observations, follow by IOPs group with 69.93% of missing data from 39, 396 observations, HMOPs group with 51.41% of missing data from the overall observation data (19, 232) and last group is PPs with a very small missing data (0.33%) from the 8, 113 observations. Nearest neighbour method is univariate methods selected were simple and straightforward which is the endpoints of the gaps are used as evaluations for all the missing values as shown in Equation (1):

$$Y = Y1 \text{ if } x \leq x1 + (x2 - x1)/2 \text{ or} \qquad (1)$$
$$Y = Y1 \text{ if } x > x1 + (x2 - x1)/2 \text{ or}$$

where y is the interpolant, x is time point of the interpolant, y1 and x1 are the coordinates of the starting [8].

2.4. Principle Component Analysis

Reducing the dimensionality of the data sets with used the PCA analysis allowed the identification of an association between variables [9]. Three major steps were involved in PCA firstly to produce new variables, the adjustment of dimensions need have equal weights in the analysis by autoscaling the data. In other words, the mean is equal to zero and the standard deviation is equal to the unit. Secondly is identifying the eigenvalues and their corresponding eigenvectors by calculation of the covariance matrix and thirdly for a small proportion of the variation in data sets, the elimination of components that account only [10, 29]. Present study, only varimax factors with values more than 0.70 will be interpreted. Variables with loadings greater than 0.7 are considered strong, less than 0.7 to 0.5 are moderate and lower than 0.5 are considered a weak variable [11-12]. Therefore, present study uses this PCA analysis to identify the most significant parameters each group of drinking water quality.

2.5. One-Way Analysis of Variance

One-way ANOVA is a method of testing differences between more than two treatments. Identifying specific differences between pairs of treatments are described by multiple comparison procedures and orthogonal contrasts. In others word call self-determining samples t-test which is the test statistic is calculated by dividing the variance between the sample means by the standard error of the variance. The standard error is the variability between the samples is compared with the variability within the samples [13, 31]. The equation of the variance is given by the following:

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \qquad (2)$$

where $\sum_{i=1}^{n}(x_i - \bar{x})^2$ is the sum of squares and n − 1 is degree of freedom.

One-way ANOVA use to discriminate the parameters between two sampling regions (UTW and TW).

2.6. Principle Component Analysis

DA is a technique for classifying objects by a set of independent variables into predefined classes [14]. For example, present study to identify the performance of WTPs either in a good or poor treatment and suggested which plants should be under monitor. On the basic application. This method enables the grouping or clustering of the observations based on input variables or the variables known as predictors. This technique construct set of a linear function of the predictors, known as the discriminant functions (DF) and the equation as follow:

$$L = b_1X_1 + b_2X_2 + \ldots\ldots\ldots\ldots + b_nX_n + C \tag{3}$$

where, b is the discriminant coefficient, x is the input variables or predictors and C is a constant. Further, Discriminant analysis (DA) functionality aids in determining the best cluster. The applied DA on the original dataset offers similar discriminant ability towards the original dataset with or without standardization in constructing the discriminant factor (DF) based on the following Equation (4) [6]:

$$f(G_i) = k_i + \sum_{j=1}^{n} w_{ij} \cdot p_{ij} \tag{4}$$

where i denotes the number of groups (G), k_i is the constant coherent to each group, n is the number of parameter used to classify a set of data into given group and w_{ij} is the weight coefficient assigned by discriminant factor (DF) to a given parameter (p_{ij}) [4, 5, 7, 30].

3. Results and Discussion

3.1. Drinking Water Quality Significant Parameters

3.1.1. Physico-Chemical Parameter (PCP)

This first independent variable physico-chemical parameters (PCP) as shown in Table 1 showed three PCs were attained with the total sum of cumulative approximately 64.56%. Only six of significant parameters (TC, E. coli, turbidity, colour, pH and residual chlorine) respectively out of eight. D1 described that the higher of total variance with 26.92% that is loads two strong positive factors loading which are Total coliform (0.8786) and E. coli (0.9040). Both of these polluted became as a biological drinking water indicator and part of drinking water regulation. This indicator conceivable to investigation drinking water directly for E. coli and total coliforms inexpensively and simply [15]. Second PCs is D2 displayed the total variance is 22.27% and also give two of positive strong factor loading which are turbidity and colour with value 0.7328 and 0.7664. Turbidity and colour had a relationship both of them in the parameters control in coagulation or flocculation process in the WTP. Therefore, to remove them aluminium sulphate provided to rise in the effectiveness of coagulation or flocculation process [16]. D3 is the last PCs was described with the value of variance is 15.37% and put off the negative and positive factors loading for pH (-0.7616) and residual chlorine (0.7640). Residual chlorine became the most important parameter because of reducing the bacterial in drinking water. The increasing of the residual chlorine, showed the decreasing of distribution of bacteria [17]. The permissible limit of pH is around 6.5 until 9.0 maximum level in MDWQS. This six's significant parameters were used for the next One- way ANOVA analysis.

Table 1: Factor loading for selected PCP parameters in drinking water.

PCP	D1	D2	D3
TC	0.8786	0.0809	0.0634
E. coli	0.9040	0.1211	0.1080
Turbidity	0.5208	0.7328	-0.0526
Color	-0.0571	0.7664	0.0470
PH	0.0534	-0.2449	-0.7616
Residual chlorine	0.2830	-0.1832	0.7640
Temperature	-0.0987	0.4352	0.1863
Conductivity	0.4439	0.5940	0.1036
Eigenvalue	2.7629	1.2759	1.1259
Variability (%)	26.9166	22.2671	15.3739
Cumulative %	26.9166	49.1836	64.5575

3.1.2. Inorganic Parameters (IOP)

Fig. 1 (b) and Table 2 highlight that six out of the twelve drinking water variables satisfy the 0.70 factor loading. These variables are Al, Mn, Cl, hardness, COD and BOD. These contaminants are then classified as the major contributing pollutants at the selected monitoring stations in Selangor WTP. The total sum of cumulative is around 17.35% with three varimax factors (VFs) which are the value of the variance for D1 is 18.30%, D2 is 19.58% and the last D3 is 17.35%. The principle components D1 are Al (0.7527) and Mn (0.7655) were the strong positive factor loading more than 0.70. Two of them Al and Mn are critical and non- critical metals which is a major cause for in Alzheimer's disease and Parkinsonian disturbances [18]. Chloride (Cl-) and hardness give the positive strong factor loading in D2 VFs. The percent of variance in this VFs is 19.58% and choose two pollutants (Cl-, 0.8147 and hardness, 0.8311) give very strong factors loading. Cl- was originates from huge of sources in the environment and hardness caused by dissolved calcium and magnesium in drinking water. Together of them depending on the alkalinity of the water and became the main reason to the corrosion of metals in the distribution system [19]. Last of VF is D3 with the total variance 17.35% and also have two significant components are COD and BOD. The value of positive strong loading for both of them are 0.7931 and 0.7275. The relationship of them is measure the number of organic compound in water and the relation oxygen-reduction effect of a waste pollutant. The best of about these two pollutants is both have been commonly adopted as a measure of pollution effect [20]. Thus, for further analysis process only stayed to use these six most significant parameters.

Table 2: Factor loading for selected IOP parameters in drinking water.

IOP	D1	D2	D3
TDS	0.4186	0.3132	0.0221
Cl-	-0.0319	0.8147	0.1445
NH3N	0.4011	0.2815	0.4185
NO3N	0.1891	0.6320	-0.1549
Fe	0.6513	-0.1306	0.3138
Fl	-0.3237	0.5675	-0.3333
Hardness	0.1571	0.8311	0.0206
Aluminum	0.7527	-0.0409	0.1164
Mangan	0.7655	0.2571	0.1474
COD	0.3055	-0.0207	0.7931
BOD	-0.0204	0.0883	0.7275
TOC	0.1553	-0.0563	0.6770
Eigenvalue	3.2008	2.3356	1.0922
Variability (%)	18.3070	19.5790	17.3533
Cumulative %	18.3070	37.8860	55.2393

3.1.3. Heavy Metal and Organic Parameters (HMOP)

The result after varimax rotation using PCA analysis presented as shown in Table 3. There are five of VFs of HMOP with the total percent of cumulative around 53.67%. First VFs or D1 give the higher of variance among the components is 19.37% followed by D2 is 11.03%, D3 is 9.15%, D4 is 7.60% and D5 is 6.52%. All of fives VFs, positive strong factor loading was entertained with each their two significant parameters. For D1 the most significant drinking water parameters was set is CHBr2Cl and CHBrCl2. These two pollutants were present from Drinking water disinfection by-products (DBPs). Both of them are an accidental significance of using chemical disinfectants to destroy damaging pathogens in water [21]. The positive strong factor loading for D2 are Arsenic/ As (0.8090) and Lead/ Pb (0.7344). The exposure of

As in drinking water shown to cause cancer in humans through consumption of drinking-water. The evidence from the epidemiological studies that drinking from the high contaminant of As causally related development of cancer at several sites, particularly skin, bladder and lung. Compare to Pb, rarely present in tap water is primarily from household plumbing systems containing lead in pipes, solder, fittings [20]. PCs D3 showed Mercury (Hg) and Argentum (Ag) were most significant parameters. D4 and D5 was located Chromium (Cr), Cuprum (Cu) and Zinc (Zn) as their strong factor loading. For Cu and Cr, two of them metals are natural components in soil and the most common heavy metals contaminants [22, 32] while Zn is metals are micronutrients needed for plant growth [23]. In this HMOP out of nine from sixteen parameters should be taking as consideration in the future sampling task.

Table 3: Factor loading for selected HMOP parameters in drinking water.

HMOP	D1	D2	D3	D4	D5
Hg	-0.0324	0.0507	0.7030	0.1035	-0.2739
Cd	0.0232	-0.0311	0.5682	-0.0681	0.2958
As	-0.2063	0.8090	0.0320	0.0613	-0.0203
Pb	-0.1404	0.7344	0.0792	0.0903	-0.0180
Cr	0.0331	0.1054	0.0415	0.7672	-0.0203
Cu	0.0469	0.0206	-0.0256	0.7468	0.1025
Zn	-0.0554	0.0008	0.0143	0.0721	0.8045
Na	0.5679	0.3794	0.0072	-0.1444	0.0815
SO4	0.5611	0.2267	-0.0349	-0.0533	0.1529
Se	0.0241	0.1033	0.1611	0.0189	-0.4128
Ag	0.0394	0.0431	0.7426	-0.0349	0.0953
Mg	0.4934	0.4571	-0.2225	0.0091	0.0457
Chloroform	0.6194	-0.1222	0.0518	-0.0272	-0.0607
Bromoform	0.6823	-0.2835	0.0776	0.0479	-0.0559
$CHBr_2Cl$	0.7835	-0.1332	0.0038	0.0832	-0.0437
$CHBrCl_2$	0.8265	-0.1614	-0.0124	0.0419	-0.0414
Eigenvalue	3.1309	1.7698	1.4677	1.1853	1.0337
Variability (%)	19.3716	11.0292	9.1498	7.5991	6.5211
Cumulative %	19.3716	30.4008	39.5506	47.1497	53.6708

3.1.4. Pesticide Parameters (PP)

Table 4 described only two VFs with Eigen value more than one (> 1). The percent of the variance for D1 higher with the value is 56.34% and for D2 is 20.97% witch the total sum of cumulative is 77.30%. Aldrin, DDT, H & He, Methox, Lindane, Chlordane and Endosulfan are components in pesticides are usually used in agricultural production to control pests, diseases, weeds and other plant pathogens [24]. Seven parameters showed the positive strong factor loading and all of them are the most significant parameters which are should be under more observation.

Table 4: Factor loading for selected PP parameters in drinking water.

PP	D1	D2
Aldrin/Dieldrin	0.8454	0.0268
DDT	0.8533	0.0319
H & He	0.9667	0.0444
Metho	-0.0779	0.8669
Lindane	0.1668	0.8427
Chlordane	0.9587	0.0436
Endosulfan	0.7831	0.0232
Eigenvalue	3.9639	1.4474
Variability (%)	56.3358	20.9690
Cumulative %	56.3358	77.3048

3.2. Spatial Disparities of Treated Water (TW)

The most significant parameters each group which is analyzed from PCA analysis was continue used in this study. Six out of eight the most significant parameters for PCP followed by IP also give six important parameters was monitored from twelve parameters, HMOP highlight nine out of sixteen significant parameters and lastly all the PP parameters became the most

important parameters. One-way Analysis of Variance (One- way ANOVA) was used to analyze this further study by each group.

3.2.1. Physico-Chemical Parameters

Table 5 (A) present the differences of two categories sample firstly treated water (TW) and secondly untreated water (UTW) for PCP. There are six parameters TC, E. coli, turbidity, colour, pH and residual chlorine) showed the significant differences due to their concentration obtained among UTW and TW with the variances of LS means (p < 0.05). Five of pollutants give the reducing numbers means of samples from UTW to TW only pH was increased the number mean of sample which is 6.9 for UTW and 7.305 for TW. The increase of the pH means decrease the concentration of lead produces from the corrosion process in plumbing system [25]. So, for this PCP study all parameters had been variances among the UTW and TW categories.

3.2.2. Inorganic Parameters

One- way ANOVA compares the mean of two samples with six dependent variables. Chloride, hardness, aluminium, manganese, COD and BOD was statically presented significant different among UTW and TW which is P value less than 0.05 (p < 0.05) as shown in Table 5 (B). The increase of contamination of Cl- (3.378 UTW – 7.019 TW) through the treatment it's because of chlorine or chloride is used during disinfection process however concentrations in excess of about 250 mg/litre of Cl- can give rise to obvious taste in water [26]. Thus, the level of water hardness (12.121 UTW, 20.553 TW) determined by magnesium (Mg) and calcium (Ca) contents. These two components were important sources from water along with diet [27].

3.2.3. Heavy Metal and Organic Parameters

Nine out of sixteen the most significant of HMOP was mentioned by PCA analysis use for this study. As shown in Table 5 (C) mentioned, only five parameters were statistically showed significant different among them (p < 0.05). There are mercury (Hg), arsenic (As), iron (Pb), dibromochloromethane (CHBr2Cl) and bromodichloromethane (CHBrCl2). Out of four (chromium (Cr), Cuprum (Cu), zinc (Zn) and argentum (Ag)) classified as a no significant different (p > 0.05) with two categories UTW and TW samples. Therefore, if the samples no variances, so that pattern of both UTW and TW was similar. However, this value still under maximum acceptable limit under DWQS. These two pollutant CHBr2Cl and CHBrCl2 give the increase value of contaminant after treatment process. It is because both of the pollutant DBPs (Trihalomethanes), which is Cl is used to disinfect drinking water [28]. That is why the increasing of the value both of them.

3.2.4. Pesticide Parameters

Table 5 (D) was elaborated all of seven PP were not significantly difference in pre- treatment (UTW) and post treatment (TW) samples with p value more than 0.05 (p > 0.05). Trend of water quality pattern quite similar, either UTW or TW samples for PP. Therefore, WTP need to additional technical ways for removal the PP respectively and follow the SOP of maintenance system. Even though all these parameters didn't have any changes from UTW through TW, the value of these pollutants meet the permissible maximum limit of DWQS.

Table 5: Analysis of variances between treated and untreated water for PCP, IP, HMOP and PP parameters.

(A) PCP	LS means		Pr > F	Significant
	UTW	TW		
TC	7288.394 a	0.104 b	0.0000	Yes
E. coli	2921.820 a	0.016 b	0.0000	Yes
TURBIDITY	68.452 a	1.195 b	0.0000	Yes

TCU	37.506 a	2.921 b	0.0000	Yes
PH	6.900 b	7.305 a	0.0000	Yes
Residual chlorine	7.294 a	1.830 b	0.0000	Yes
(B) IP				
Cl	3.387 b	7.019 a	0.0000	Yes
Hardness	12.121 b	20.553 a	0.0000	Yes
Aluminum	0.245 a	0.085 b	0.0000	Yes
Mangan	0.046 a	0.024 b	0.0000	Yes
COD	7.829 a	0.839 b	0.0000	Yes
BOD	2.433 a	0.327 b	0.0000	Yes
(C) HMOP				
Hg	0.001 a	0.001 b	0.0096	Yes
As	0.004 a	0.001 b	0.0000	Yes
Iron	0.002 a	0.001 b	0.0000	Yes
Cr	0.001 a	0.001 a	0.5079	No
Cu	0.004 a	0.004 a	0.7441	No
Zn	0.016 a	0.018 a	0.7728	No
Ag	0.002 a	0.002 a	0.8565	No
Dibromochloromethane	0.000 b	0.002 a	0.0000	Yes
Bromodichloromethane	0.000 b	0.003 a	0.0000	Yes
(D) PP				
Aldrin/Dieldrin	0.003 a	0.003 a	0.7490	No
DDT	0.026 a	0.026 a	0.9668	No
H & He	0.006 a	0.006 a	0.9960	No
Methoxychlor	0.021 a	0.021 a	1.0000	No
Lindane	0.012 a	0.012 a	0.9834	No
Chlordane	0.053 a	0.053 a	0.9849	No
Endosulfan	0.008 a	0.008 a	0.9916	No

3.3. Water Treatment Plant (WTP) Performance

3.3.1. Physico-Chemical Parameters

Spatial classification using standard, forward stepwise and backward stepwise mode DFA give the correctness which are 76.96% (six discriminant variables), 76.82% (four discriminant variables) and 76.97% (six discriminant variables) respectively (Table 6). The significant variables that exist from forward stepwise are TC, Turbidity, pH and Residual Chlorine. In terms of their spatial distribution, these parameters have high difference was classified.

Table 6: Classification matrix for DA of spatial variations in Selangor and Kuala Lumpur Federal Territory WTP for PCP.

Sampling regions	% correct	Regions assign by DA	
		TW	UTW
Standard DA mode (6 variables)			
TW	98.57	3313	48
UTW	55.34	1501	1860
Total	76.96	4814	1908
Forward stepwise mode (4 variables)			
TW	98.57	3313	48
UTW	55.07	1510	1851
Total	76.82	4823	1899
Backward stepwise mode (6 variables)			
TW	98.57	3313	48
UTW	55.34	1501	1860
Total	76.97	4814	1908

Thus, backward stepwise included E. coli and colour as the six parameters to have a high spatial variation. Box and whisker plots of some PCP parameters are shown in Figure 1. These result highlight that the performance of WTP is good but in the meantime, there are 48 of WTP from the observation need to be improve because the contamination of pollutant was increased from TW became UTW.
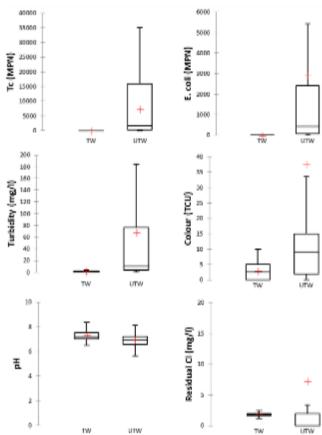

Fig. 1: Box and whisker plots of some PCP parameters separated by spatial DA

3.3.2. Inorganic Parameters

Based on finding given by standard DA mode, the correct specification of the spatial variation of IOP is 91.90% which is very convincing value (Table 7). So, further analysis of DA using forward stepwise and backward stepwise mode will be accepted in order to recognize the most significant water quality parameters which theatres an important part in discriminating the UTW and TW categories. By forward stepwise and backward stepwise mode highlight that both of these value same with standard DA mode is 91.90%. Sixth significant parameters (Cl-, hardness, Al, Mn, COD and BOD) showed high differences was analyzed from the backward stepwise mode.

Table 7: Classification matrix for DA of spatial variations in Selangor and Kuala Lumpur Federal Territory WTP for IOP

Sampling regions	% correct	Regions assign by DA	
		TW	UTW
Standard DA mode (6 variables)			
TW	94.70	1554	87
UTW	89.09	179	1462
Total	91.90	1733	1549
Forward stepwise mode (6 variables)			
TW	94.70	1554	87
UTW	89.09	179	1462
Total	91.90	1733	1549
Backward stepwise mode (6 variables)			
TW	94.70	1554	87
UTW	89.09	179	1462
Total	91.90	1733	1549

Thus, for more cleared with these variances among sixth variables, box and whisker plot shown as shown in Figure 2. Although this WTP give very good performance with the convince value 91.9% of distinguish variables, to be in mind that there are 87 spatial observation should be under maintenance.
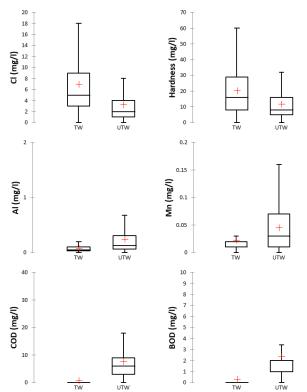
Fig. 2: Box and whisker plots of some IOP parameters separated by spatial DA

3.2.3. Heavy Metal and Organic Parameters

The percent correction of parameters distinguishes given by standard DA mode was very convince value which is 93.27% with five variables (Hg, As, Pb, BOD and COD). Further analysis with forward and backward stepwise DA mode for identify the most significant drinking water quality variables. The value of both them are 94.14% and give fourth discriminant parameters that is Hg, As, BOD and COD.

Table 8: Classification matrix for DA of spatial variations in Selangor and Kuala Lumpur Federal Territory WTP for HMOP.

Sampling regions	% correct	Regions assign by DA	
		TW	UTW
Standard DA mode (5 variables)			
TW	89.69	513	59
UTW	96.85	18	554
Total	93.27	531	613
Forward stepwise mode (4 variables)			
TW	91.43	523	49
UTW	96.85	18	554
Total	94.14	541	603
Backward stepwise mode (4 variables)			
TW	91.43	523	49
UTW	96.85	18	554
Total	94.14	541	603

The variance of these four variables was presented on box and whisker plot with their mean variance each parameter (Figure 3). In the sampling regions TW, DA was discriminated region TW is 513 spatial observation from 531 observations. The rest of TW (59) observation, DA assign as UTW. So, for precautionary measures to make WTP in high performance these 59-spatial variations were observed should under properly preservation.
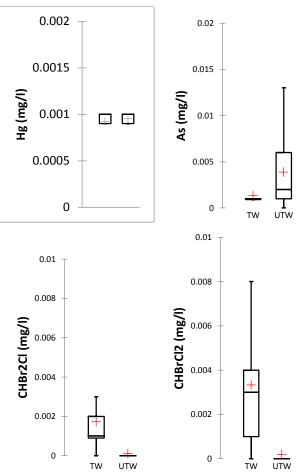

Fig. 3: Box and whisker plots of some HMOP parameters separated by spatial DA

3.2.4. Pesticide Parameters

The standard DA mode failed to discriminate all the data obtained from two station categories due to the spatial variation of pesticide patterns (Table 9). All the data obtained having a similar pattern. Based on the results, the total percent correct classification given by the standard DA mode is lower which only 50.43%. The pesticide pollutant patterns obtained from both station category TW and UTW have achieved almost 50% correct classification with the total number of observed data are 665 and 493 respectively. The output shows highly similar patterns ($p > 0.05$) with the data obtained from both station category. Based on the results found, the stepwise forward and backward could not be proceeded. Thus, it strongly confirms that the overall water quality pattern for the UTW and TW water are similar to each other (similar in chemical characteristics) in other words not good performance. Therefore, almost all the WTP for this group compulsory to properly maintain to produce the high quality of drinking water to end users.

Table 9: Classification matrix for DA of spatial variations in Selangor and Kuala Lumpur Federal Territory WTP for PP.

Sampling regions	% correct	Regions assign by DA	
		TW	UTW
Standard DA mode			
TW	57.86	335	224
UTW	43.01	330	249
Total	50.43	665	493

Box and whisker plot showed all seventh variables have the same variance weather in UTW and TW (Figure 4).
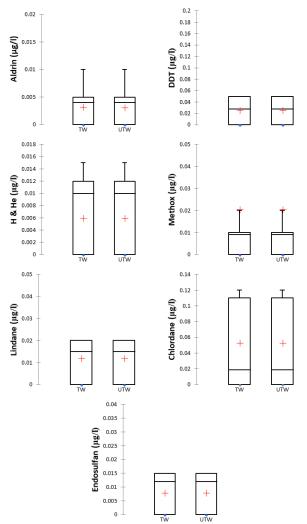
Fig. 4: Box and whisker plots of some HMOP parameters separated by spatial DA

4. Conclusion

NDWQSP have recorded tones, multidimensional and complicated data set, which requires chemometric techniques for data analysis and explanation of the original information. Present study was used three of the chemometric methods (PCA, One-way ANOVA and DA) for analyze the complex of data through pre-treatment data until data was analyzed to produce the good result. The most significant of parameters in each group (PCPs, IOPs, HMOPs and PPs) was identified using PCA with confident level 95% which are only selected parameters should be monitored in the future. Thus, the most significant parameters become as variables for the next technique that is one-way ANOVA. In this analysis, one- way ANOVA was successful distinguish pollutants from two category of sampling stations that is UTW and TW in all group parameters except certain of HMOPs and all of PPs. DA was validated the performance of WTPs in each of group parameters. Out of three group parameters (PCPs, IOPs and HMOPs) was suggested WTPs in a good performance with small numbers of WTPs observation should under maintenance. Chemometric approach is the best analysis technique for any relevant agencies (Ministry of Health (MOH), Kementerian Tenaga, Teknologi Hijau dan Air Malaysia (KeTTHA), Suruhanjaya Perkhidmatan Air Negara (SPAN), Water Services Department (WSD)/Water Board (WB) and other Water Purveyors, Department of Chemistry (DOC), Dewan Bandaraya Kuala Lumpur (DBKL) and Local Authorities all over the country) which are collaboration and co-operation to make

NDWQSP more efficient and effective implementation. Advantages of this approach is a lot of useful information was generated from a million of data set thus will reduce the cost and time of sampling management. So, in the future only the most significant parameters in each group should under observe in the sampling task and certain of WTPs will keep on eyes their maintenance for good in performance.

Acknowledgement

References

[1] Eldib, M. A., & Elbayoumy, M. A. A. (2003, April). Evaluation of A Water Treatment Plant Performance-Case Study. In Seventh International Water Technology Conference, pp. 1-3.

[2] National Drinking Water Quality surveillance Programme (NDWQSP) Annual Report Year 2016

[3] Christensen, J.H., Tomasi, G. 2007. A multivariate approach to oil hydrocarbon fingerprinting and spill source identification. Oil Spill Environmental Forensics XII 293–325.

[4] Juahir, H., Zain, S.M., Yusoff,M.K., Hanidza, T.I.T., Armi, A.S.M., Toriman,M.E., Mokhtar, M.2011. Spatial water quality assessment of Langat River Basin (Malaysia) usingenvironmetric techniques. Environmental Monitoring Assessment 173:625–641.

[5] Ismail, A., Toriman, M. E., Juahir, H., Md. Kassim, A., Md Zain, S., Ahmad, W. K. W., Fah, W. K., Retnam, A., Abdul Zali, M., Mokhtar, M. and Yusri, M. A. 2016a. Chemometric techniques in oil classification from oil spill fingerprinting. Marine Pollution Bulletin (article in press).

[6] Kannel, P.R., Lee, S., Kanel, S.R., Khan, S.P. 2007. Chemometric application in classificationand assessment of monitoring locations of an urban river system. Analytica Chimica Acta582, 390–399.

[7] Ismail, A., Toriman, M. E., Juahir, H., Md Zain, S., Abdul Habir, N. L., Retnam, A.,Kamaruddin, M. K. A., Umar, R. and Azid, A. 2016b. Spatial assessment and source identification of heavymetals pollution in surface waterusing several chemometric techniques. Marine Pollution Bulletin (article in press).

[8] Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., & Kolehmainen, M. (2004). Methods for imputation of missing values in air quality data sets. Atmospheric Environment, 38(18), 2895-2907.

[9] Osman, R., Saim, N., Juahir, H., & Abdullah, M. P. (2012). Chemometric application in identifying sources of organic contaminants in Langat river basin. Environmental monitoring and assessment, 184(2), 1001-1014.

[10] Vialle, C., Sablayrolles, C., Lovera, M., Jacob, S., Huau, M. C., & Montréjaud-Vignoles, M. (2011). Monitoring of water quality from roof runoff: Interpretation using multivariate analysis. Water research, 45(12), 3765-3775.

[11] Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. I. T., Armi, A. S. M., Toriman, M. E., Mokhtar, M. 2010. Environmental Monitoring and Assessment. 173(1-4): 625-641.

[12] Kamaruddin, A. F., Toriman, M. E., Juahir, H., Zain, S. M., Rahman, M. N. A., Kamarudin, M. K. A., & Azid, A. (2015). Spatial characterization and identification sources of pollution using multivariate analysis at Terengganu river basin, Malaysia. Jurnal Teknologi, 77(1), 269-273.

[13] Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 9: one-way analysis of variance. Critical care, 8(2), 130.

[14] Morrison, D. G. (2011). Discriminant analysis and predictive classification. Marketing Classics Press.

[15] Edberg, S. C. L., Rice, E. W., Karlin, R. J., & Allen, M. J. (2000). Escherichia coli: the best biological drinking water indicator for public health protection. Journal of Applied Microbiology, 88(S1).

[16] Cardoso Valverde, K., Konradt Moraes, L. C., Carvalho Bongiovani, M., Pereira Camacho, F., & Bergamasco, R. (2013). Coagulation diagram using the Moringa oleifera Lam and the aluminium sulphate, aiming the removal of color and turbidity of water. *Acta Scientiarum. Technology*, *35*(3).

[17] Mirzaei, N., Ghaffari, H. R., Karimyan, K., Moghadam, F. M., Javid, A., & Sharafi, K. (2015). Survey of effective parameters

(water sources, seasonal variation and residual chlorine) on presence of thermotolerant coliforms bacteria in different drinking water resources. International Journal of Pharmacy and Technology, 7(3), 9680-9689.

[18] Zatta, P., Lucchini, R., van Rensburg, S. J., & Taylor, A. (2003). The role of metals in neurodegenerative processes: aluminum, manganese, and zinc. Brain research bulletin, 62(1), 15-28.

[19] Edition, F. (2011). Guidelines for drinking-water quality. WHO chronicle, 38(4), 104-8.

[20] Lokhande, R. S., Singare, P. U., & Pimple, D. S. (2011). Pollution in water of Kasardi River flowing along Taloja industrial area of Mumbai, India. World Environment, 1(1), 6-13.

[21] Richardson, S. D., & Postigo, C. (2011). Drinking water disinfection by-products. In Emerging organic contaminants and human health (pp. 93-137). Springer, Berlin, Heidelberg.

[22] M. M. Lasat, "Phytoextraction of metals from contaminated soil: a review of plant/soil/metal interaction and assessment of pertinent agronomic issues," Journal of Hazardous Substance Research, vol. 2, no. 5, pp. 1–25, 2000.

[23] Gaur and A. Adholeya, "Prospects of arbuscular mycorrhizal fungi in phytoremediation of heavy metal contaminated soils," Current Science, vol. 86, no. 4, pp. 528–534, 2004.

[24] Damalas, C. A., & Eleftherohorinos, I. G. (2011). Pesticide exposure, safety issues, and risk assessment indicators. International journal of environmental research and public health, 8(5), 1402-1419.

[25] Kim, E. J., Herrera, J. E., Huggins, D., Braam, J., & Koshowski, S. (2011). Effect of pH on the concentrations of lead and trace contaminants in drinking water: A combined batch, pipe loop and sentinel home study. Water research, 45(9), 2763-2774.

[26] Guidelines for drinking-water quality, 2nd ed. Vol. 2. Health criteria and other supporting information, World Health Organization, Geneva, 1996.

[27] Kanadhia, K. C., Ramavataram, D. V. S. S., Nilakhe, S. P. D., & Patel, S. (2014). A study of water hardness and the prevalence of hypomagnesaemia and hypocalcaemia in healthy subjects of Surat district (Gujarat). Magnesium research, 27(4), 165-174.

[28] Xie, Y. (2016). Disinfection byproducts in drinking water: Formation, analysis, and control. CRC Press.

[29] Ismail, A., Toriman, M. E., Juahir, H., Zain, S. M., Habir, N. L. A., Retnam, A., ... & Azid, A. (2016). Spatial assessment and source identification of heavy metals pollution in surface water using several chemometric techniques. Marine pollution bulletin, 106(1), 292-300.

[30] Al-Odaini, N. A., Zakaria, M. P., Zali, M. A., Juahir, H., Yaziz, M. I., & Surif, S. (2012). Application of chemometrics in understanding the spatial distribution of human pharmaceuticals in surface water. Environmental monitoring and assessment, 184(11), 6735-6748.

[31] Juahir, H., Zain, S. M., Aris, A. Z., Yusof, M. K., Samah, M. A. A., & Mokhtar, M. (2010). Hydrological trend analysis due to land use changes at Langat River Basin. Environment Asia, 3(2010), 20-31.

[32] Aris, A. Z., Abdullah, M. H., Praveena, S. M., Yusoff, M. K., & Juahir, H. (2010). Extenuation of saline solutes in shallow aquifer of a small tropical island: a case study of Manukan Island, North Borneo. Environment Asia, 3, 84-92.