J. Mater. Environ. Sci. 7 (12) (2016) 4403-4410 *Sabri et al.*
ISSN : 2028-2508
CODEN: JMESC

Evaluation of Hydrochemistry Variation in Water Quality of Cempaka Lake, Malaysia using Multivariate Statistical Analysis

N. A. S. M. Sabri[1], M. P. Abdullah[1,2*], S. Mat[3], R. Elfithri[4], W. M.A. W. M. Khalik[1]
[1]*School of Chemical Sciences and Food Technology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia*
[2]*Centre for Water Research and Analysis (ALIR), Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia*
[3]*Solar Energy Research Institute (SERI), Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia*
[4]*Institut for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia*

Received 07 Apr 2016, Revised 13Aug 2016, Accepted 25 Aug 2016
**Corresponding author. E-mail: mpauzi@ukm.edu.my (M.P. Abdullah); Phone: +60 3 89215447; Fax: +60 3 89215401*

Abstract
Hydrochemistry variation in water quality of Cempaka Lake was successfully evaluated. Data sets were tested subject to multivariate statistical analysis. Seven sampling points were selected and monitoring work was conducted monthly (May 2013–May 2014). The physicochemical parameters, namely temperature, pH, conductivity, dissolved oxygen, suspended solid, ammonia nitrogen, chemical oxygen demand, biochemical oxygen demand and phosphate were measured. Based on cluster analysis, seven stations were grouped into two clusters at $(D_{link}/D_{max}) \times 100 < 30$. Four principle components explained about 87.90 % of the total variance in the water data sets from eigenvalue > 1. Classification matrices were assigned correctly of 85.70 % and 92.90 % for spatial and temporal variations respectively. Variables namely dissolved oxygen, biochemical oxygen demand, chemical oxygen demand, ammoniacal nitrogen and phosphate were recognized to have significant influence on the deterioration of water quality in Cempaka Lake.

Keywords: Chemometric, Man-made lake, Physico-chemical analysis, Water pollution

1. Introduction
Lake is one of the aquatic ecosystems that exist naturally or is man-made, which is also essential for living organisms. It may serve as a recreational park, hydro energy or irrigation supply, food source and other daily water use of anthropogenic activities. Besides river and marine ecosystem, pollution of water in lake is becoming a rising environmental concern. Pollution of lake water is associated to excessive nutrient content, toxic metals, pesticide residue and other trace pollutants. Agricultural, industrial and urban activities are considered to be major sources of the emergence of toxic chemicals and nutrients to lake ecosystems [1]. Homogeneity or inhomogeneity of pollutant contributed in the water chemistry of lake varies depending on local characteristics. For shallow lake, it is characterized by a high degree of heterogeneity in spatial and temporal changes [2].
In fact, interpretation of complex data from monitoring works is difficult and complicated thus requires proper tools to extract crucial fact with minimal loss of original information [3, 4]. Multivariate statistical analysis offers a useful tool to evaluate the variation of water quality. It is also well-documented to evaluate ecological status, pattern recognition of species, and valuable tool for reliable management of water resources as well as rapid solutions on pollution problems [5 – 8]. Previous studies have demonstrated that multivariate statistical analysis

was widely used in the evaluation of water quality in lakes, such as Baiyangdian Lake and Dianchi Lake, China [9, 10], Hyderabad Lake, India [11], Naivasha Lake, Kenya [12], Manchar Lake, Pakistan [13], Uluabat Lake and Seyfe Lake, Turkey [1, 14] as well as Chini Lake, Curtin Lake and HarapanLake in Malaysia [15 – 17].
Cempaka Lake (or called Tasik Cempaka) is situated in the state of Selangor, west coast of Peninsular Malaysia. The origin of this lake is part of a tributary of Sungai Ayer Itam, Bangi, which was extended to be a man-made lake since 1999. It serves as a public park for the residents of Bandar Baru Bangi for recreational activities. This lake has received a variety of pollution sources mainly from changes in land use, municipal discharge, household activities and small scale agricultural activity in the upper region of Ayer Itam River. According to Elzwayie et al. [18], sewage from domestic discharge become main non-point source of pollution in this area. Meanwhile a recent study by Gasim et al. [19] conclude that deterioration of Cempaka Lake's water due to urban activities and nutrition content was increased in last decade. The water quality was dropped to class II dan III based on national water quality guidelines. Previous study led by Taweel et al. revealed that food source (tilapia fish) collected from this lake has risk contamination toward copper, lead and arsenic metals [20].
In the present study, water quality data sets obtained during 13 months of monitoring program were interpreted subject to different multivariate statistical analyses. Latent information was extracted relevant to address site similarities, identify pollution source and discriminate variables responsible for spatial and temporal variations. This study was important in order to provide good lake management practice for local authority and promoting good sustainability concept for Bangi residential area. To fulfill this aspect, identify major contribution of pollution must priory identified. Variation based on temporal change (wet and dry) also discussed in this study.

2.Experimental

2.1.*Sample collection and laboratory analysis*

Seven sampling stations were chosen based on possible different anthropogenic sources. Water samples were taken monthly starting in May 2013 until May 2014. Sampling location and coordinates were determined by the Global Positioning System (GPS) as presented in Table 1 and Figure 1 respectively. Water samples were collected on surface level (< 0.5 meter) and stored in 1liter high density polyethylene and glass bottles. Samples were then placed in a cold box filled up with ice packs before being transferred to laboratory for further analysis. Parameters, namely pH, temperature, conductivity and dissolved oxygen were measured in-situ using YSI 550 multiprobe sensor. Chemical parameters, namely ammoniacal nitrogen, biochemical oxygen demand, chemical oxygen demand, suspended solids and phosphate were measured. Standard methods were used in this chemical study analysis, which include ammoniacal nitrogen (salicylate method), biochemical oxygen demand (incubation method as BOD_5), chemical oxygen demand (reactor digestion and colorimetric determination), suspended solids (gravimetric method) and phosphate (ascorbic acid method). Colorimetric determination of ammoniacal nitrogen, chemical oxygen demand, phosphate was completed using HACH DR 2400.

Table 1: The coordinate of sampling stations

Station	Latitude	Longitude
S1	2°57.427' N	100°44.289 E
S2	2°57.401' N	101°44.926 E
S3	2°57.448' N	101°45.477 E
S4	2°57.530' N	101°45.520 E
S5	2°57.516' N	101°45.582 E
S6	2°57.528' N	101°44.584 E
S7	2°57.635' N	101°45.591 E

2.2. *Statistical analysis*

All data interpretation and analysis were conducted using statistical software Minitab Version 17 (Minitab Inc., State College, USA). Water quality data sets were tested subject to different multivariate statistical analyses namely cluster analysis (CA), principle component analysis with factor analysis (PCA/FA) and discriminant analysis (DA). Agglomerative hierarchical cluster analysis using Ward method with squared Euclidean distance was applied to distinguish between clusters (site similarities). Identifying the source of pollution was recognized using principle component analysis through Eigen decomposition method. Factor analysis (Varimax method with Kaiser normalization) was used to explicate the latent variables rendered in the data sets. The factor loadings were marked as strong (VF > 0.75), moderate (0.75 < VF > 0.5) and weak (< 0.50) corresponding to absolute varifactors values [21, 22]. Discriminant analysis (standard algorithms mode) was used to construct best discriminant functions within temporal and spatial variation during the period of study.

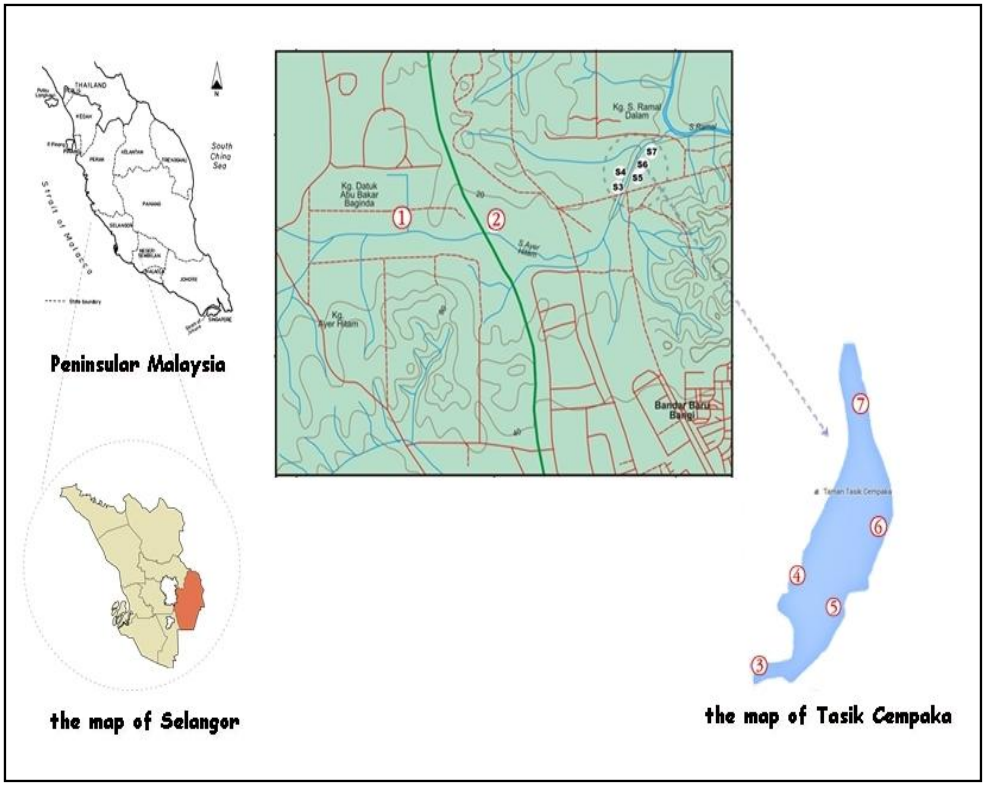

Figure 1: The map location of sampling stations

3. Results and discussion

3.1. *Water chemistry*

The basic description of range and mean concentration of each parameter is presented in Table 2. The mean value of pH was measured slightly acidic (6.88) but is considered normal for tropical freshwater, which is usually reported to be in the range of 6.5 – 8 [23]. Mean concentrations of chemical oxygen demand and suspended solid were measured to be varied widely in the range of 24.07 – 30.30 mgL^{-1} and 25.76 – 45.16 mgL^{-1} respectively. Nutrient content (phosphate) was recorded at low level concentration with a mean value of 0.48 mgL^{-1}. The kurtosis and skewness values ranged from – 0.87 to 1.57 and –1.93 to 2.40 respectively, indicating that all the data were in a normal distribution. The mean value of ammoniacal nitrogen was categorized in class IV, dissolved oxygen and chemical oxygen demand in class III, suspended solids and biochemical oxygen demand in class II. Other parameters are remaining in class I [24]. Class III and IV was considered polluted which require treatment process.

Table 2: The mean and range concentrations of nine physico-chemical parameters

Variables	Unit	Range	Mean (SD)	Skewness	Kurtosis
pH		6.79 – 7.01	6.88 (0.07)	0.06	-1.75
Dissolved oxygen	mgL^{-1}	2.58 – 3.82	3.05 (0.40)	1.08	1.61
Conductivity	μscm^{-1}	20.82 – 22.57	21.71 (0.59)	0.11	-0.25
Temperature	°C	24.87 – 26.31	25.71 (0.56)	-0.76	-1.21
Biochemical Oxygen Demand	mgL^{-1}	3.41 – 3.56	3.49 (0.05)	-0.18	-1.61
Chemical Oxygen Demand	mgL^{-1}	24.07 – 30.30	26.02 (2.15)	1.57	2.45
Suspended Solids	mgL^{-1}	25.76 – 45.16	34.93 (5.06)	-0.87	0.67
Ammoniacal Nitrogen	mgL^{-1}	2.71 – 3.68	3.07 (0.35)	0.95	-0.07
Phosphate	mgL^{-1}	0.30 – 0.68	0.48 (0.14)	0.28	-1.93

3.2. *Site similarities*

All sampling stations were grouped into two clusters, namely cluster I and cluster II at $(D_{link}/D_{max}) \times 100 < 30$. The dendrogram of cluster analysis as rendered by the Ward's method is depicted in Figure 2. Cluster I represents three stations (S1, S2, S3) while cluster II comprises of four stations (S4, S5, S6, S7). The similarity among stations in cluster I corresponds to the upper region of lake, which is located at Ayer Hitam River. For cluster II, site similarities were largely influenced by pollution background occurring in the lake's water. Higher concentrations of chemical oxygen demand, suspended solids and phosphate were found in cluster I while cluster II was dominated by high concentration of biochemical oxygen demand, ammoniacal nitrogen and low measurement of dissolved oxygen. Ammoniacal nitrogen was known as main source of pollution came from sewage discharge. High residential area in vicinity area and local village at upsteam water flow was expected as root cause of this problem. High ammoniacal nitrogen content will lowering availability of dissolved oxygen since bacteria was consumed more for degradation process.

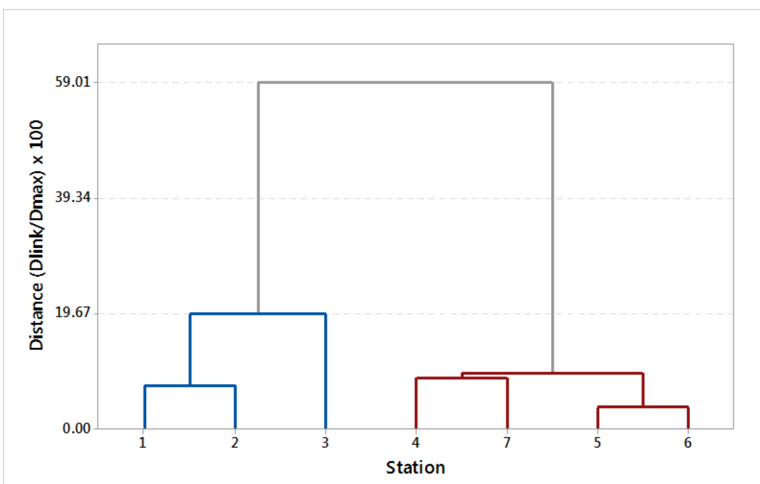

Figure 2: Dendrogram of clustering the site similarities

3.3. *Source identification*

Principle components show pronounced change in the scree plot after 4 eigenvalues (3.42, 1.97, 1.37 and 1.13). Four principle components rendered explained about 87.90 % of the total variance in the water quality data sets as presented in Table 3. The first component had strong loading on dissolved oxygen, biochemical oxygen demand and phosphate with 26.90 % of the total variance data sets. Thus, VF1 is represented by nutrient pollution.

Domestic wastewater, particularly those containing detergents, industrial effluents and fertilizer runoff were linked to elevated levels of phosphates in surface waters [1]. In this case, higher concentration was measured at S1 (0.51 – 0.68 mgL^{-1}), which affirms that the occurrence of nutrient content is associated to agricultural activity occurred at upstream area. The second component was associated to strong negative loading on conductivity, suspended solids and pH value with 24.30 % of the total variance data sets. The VF2 could be interpreted as erosion pollution. Dredging activities was observed during period of study. Higher concentration of suspended solids was recorded at SI (33.63 – 45.16 mgL^{-1}) and S2 (39.50 – 42.65 mgL^{-1}), which indirectly enhance the number of ionic species that enter the lake. The third component accountedfor 20.80 % of the total variance data sets with strong loading on chemical oxygen demand and ammoniacal nitrogen. Thus, VF3 points to the source of organic and inorganic matter from anthropogenic input. The fourthcomponent had negative loading on temperature measurement with 15.90 % of the total variance data sets. This varifactor, VF4, could be interpreted as physical chemistry of the lake. Although the measurement of watertemperature observed in this study seems to be varied, it was believed to be due to time of sampling rather than seasonal effect.

Table 3: The varifactors of varimax rotated loading results

Variable	**Communality**	**VF1**	**VF2**	**VF3**	**VF4**
pH	0.885	0.137	**-0.804**	-0.001	-0.468
Dissolved Oxygen	0.777	**0.842**	0.054	0.058	-0.251
Conductivity	0.852	-0.041	**-0.880**	0.241	0.131
Temperature	0.937	0.071	0.083	-0.072	**-0.959**
Biochemical Oxygen Demand	0.932	**0.949**	-0.107	-0.047	0.133
Chemical Oxygen Demand	0.926	-0.026	-0.060	**0.950**	0.136
Suspended Solids	0.837	0.178	**-0.771**	0.173	0.426
Ammoniacal Nitrogen	0.877	0.311	-0.319	**0.822**	-0.048
Phosphate	0.893	**0.813**	-0.199	0.437	-0.004
Variance		26.90	24.30	20.80	15.90
Cumulative Variance		26.90	51.20	72.00	87.90

Bold value indicate strong loading

3.4. *Temporal and spatial variation*

The final result of discriminant analysis showed that the classification matrix of spatial variation was assigned correctness with 85.70 % using only 8 discriminant variables. The classification matrix and discriminant functions for spatial variation obtained from standard mode are depicted in Tables 4 and 5 respectively. In this study, variables, namely pH, dissolved oxygen, temperature, biochemical oxygen demand, chemical oxygen demand, suspended solids and phosphate were recognized as latent factors to distinguish pollution occurring between clusters. Cluster I was remarkably affected by the elevated concentration of chemical oxygen demand, suspended solids and phosphate. The upstream sites (station S1 –S3) show that the degradation of water quality is higher in comparison to sites located downstream (S4–S7). The water quality index was calculated and the status of each station was stated as the following: S1 (58), S2(59), S3(59), S4(63), S5 (60), S6 (61) and S7 (61). Deterioration of water quality in Cluster II was dominated by higher presence of ammoniacal nitrogen and higher consumption of oxygen in biological microorganism activity (biochemical oxygen demand). During this study, higher level of ammoniacal nitrogen was recorded at station S7 (3.20 – 3.42 mgL^{-1}), close to drainage channel flow from local settlements.

Table 4: Classification matrix of discriminant analysis for spatial variation

Spatial	Correct (%)	Region Assign by Discriminant Analysis	
		Cluster I	Cluster II
Cluster I	66.70	26	0
Cluster II	100	13	52
Total	85.70	39	52

Table 5: Discriminant functions of latent variables for spatial variation

Variable	Cluster I	Cluster II
pH	29.52	27.09
Dissolved Oxygen	-2.51	-3.20
Temperature	0.57	1.09
Biochemical Oxygen Demand	43.38	45.58
Chemical Oxygen Demand	1.77	1.76
Suspended Solids	-0.17	-0.18
Ammoniacal Nitrogen	-4.62	-4.90
Phosphate	-19.23	-20.67
Constant	-189.03	-188.90

In this study, the period of observation was grouped into two periods (May 2013 – December 2013) and (January 2014–May 2014) as period I (wet) and II (dry), respectively. The classification matrix and discriminant functions for temporal variation obtained from standard mode are depicted in Tables 6 and 7 respectively. The classification matrix of discriminant analysis was assigned correctness with 92.90 % using 5 variables. Temporal variation in water quality of Cempaka Lake was influenced by the concentration of dissolved oxygen, biochemical oxygen demand, chemical oxygen demand, ammoniacal nitrogen and phosphate.

Table 6: Classification matrix of discriminant analysis for temporal variation

Temporal	Correct (%)	Period Assign by Discriminant Analysis	
		Period I	Period II
Period I	100	49	7
Period II	85.70	0	42
Total	92.90	49	49

Table 7: Discriminant functions of latent variables for temporal variation

Variable	Period I	Period II
Dissolved Oxygen	-9.81	-9.48
Biochemical Oxygen Demand	202.50	198.46
Chemical Oxygen Demand	6.75	6.59
Ammoniacal Nitrogen	-14.90	-14.25
Phosphate	-113.10	-110.55
Constant	-375.90	-361.80

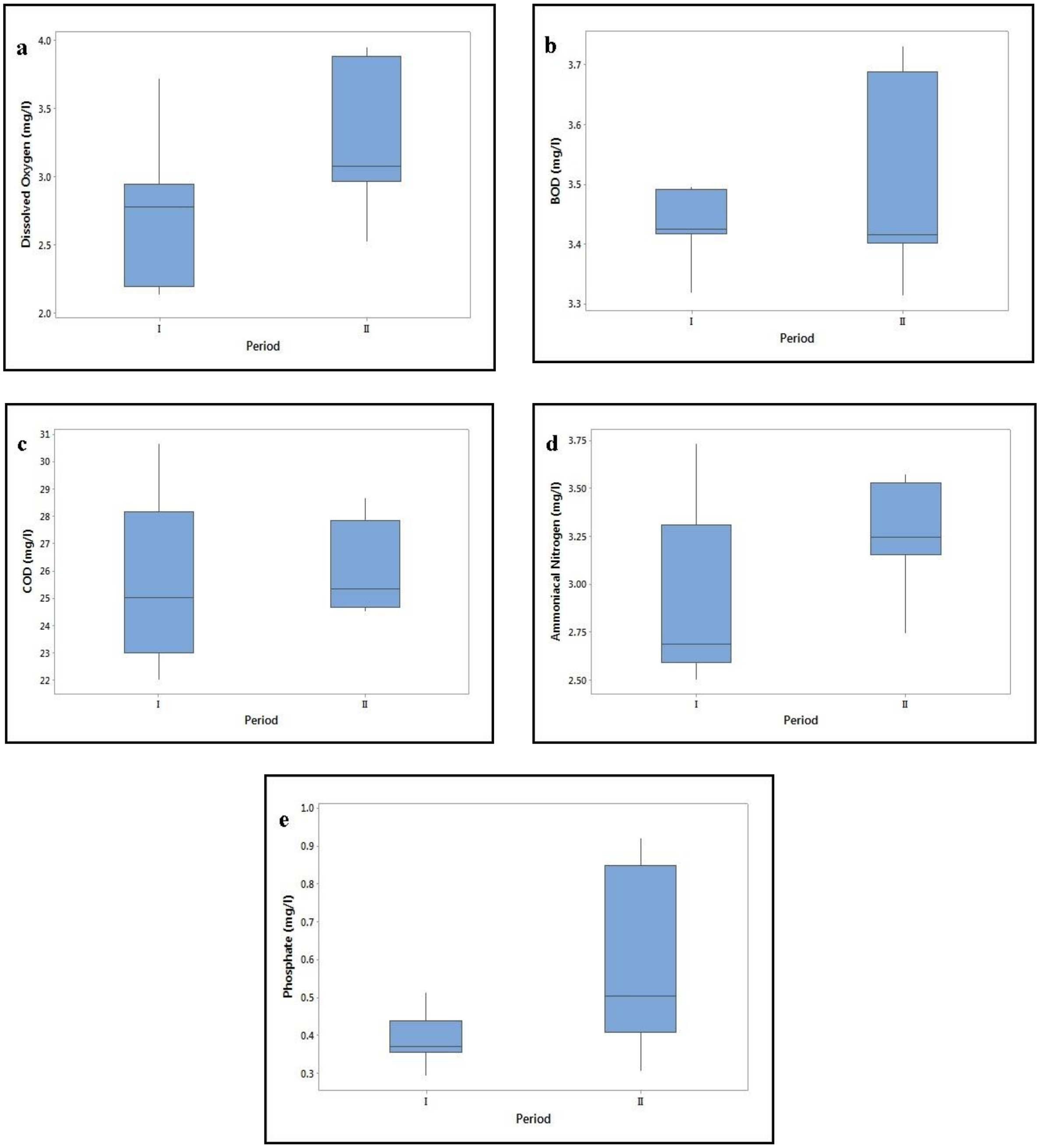

Figure 3: Box and whisker plots showing temporal variation of a) dissolved oxygen, b) biochemical oxygen demand, c) chemical oxygen demand, d) ammoniacal nitrogen and e) phosphate

Box and whisker plots of selected variables showing temporal variations are illustrated in Figure 3. The mean concentrations of all variables except pH and chemical oxygen demand were observed lower during period I (wet). Dilution effect during wet period plays a significant role in reducing the concentration of some variables like ammoniacal nitrogen, suspended solid and phosphate, which are good for the restoration of water quality. In case of chemical oxygen demand, soil erosion problem which bring in organic matter as particulate form will flow entering the lake, thus generated the increment of pollutant concentration.

Conclusions

The present study successfully provides latent information about site similarities, pollution source background, spatial and temporal variations in the water quality of Cempaka Lake. Spatial heterogeneity was explained through cluster analysis, represented by cluster I and II. Four sources responsible for variation in water quality, namely nutrient (phosphate), erosion or runoff (suspended solid), organic and inorganic loads (chemical oxygen demand and ammoniacal nitrogen) and atmospheric influence (temperature) were identified through principle component and factor analysis. Five variables were weighed in discriminating of spatial and temporal variations. To mitigate this issue from continuing to worsen, the concept of integrated water resource and lake management should be implemented in this area.

Acknowledgments-This study has been funded by Pharmaniaga research grant (PHUM-2013-04). Authors are also thankful to Mr. Ikhsan Idris (Universiti Kebangsaan Malaysia) for valuable assistance during sampling activities.

References

1. Filik Iscen C., Ilhan S., Arslan N., Yilmaz V., Ahiska, S., *Environ. Monit. Assess*. 144 (2008) 269.
2. Papatheodorou G., Demopoulou G., Lambrakis N., *Ecol. Modell.* 193 (2006) 759.
3. Shrestha S., Kazama, F., *Environ. Model. Softw.* 22 (2007) 464.
4. Razmkhah H., Abrishamchi A., Torkian A., *J. Environ. Manage.* 91 (2009) 852.
5. Wunderlin A.D., Díaz M.P., Amé M.V., Pesce F.S., Hued A.C., Bistoni M. Á., *Water Res.* 35 (2001) 2881.
6. Simeonov V., Stratis J.A., Samara C., Zachariadis G., Voutsa D., Anthemidis A., Sofoniou M., Kouimtzis T., *Water Res.* 37 (2003) 4119.
7. Singh K.P., Malik A., Mohan D., Sinha S., *Water Res.* 38 (2004) 3980.
8. Ouyang Y., *Water Res.* 39 (2005) 2621.
9. Yang Y.H., Zhou F., Guo H.C., Sheng H., Liu H., Dao X., He C. J., *Environ. Monit. Assess.* 170 (2010) 407.
10. Zhao Y., Xia X.H., Yang Z.F., Wang F., *Procedia Environ. Sci.* 13 (2012) 1213.
11. Kumar A.S., Reddy A.M., Srinivas L., Reddy P.M., *Environ. Pollut.* 4 (2012) 14.
12. Ndungu J., Augustijn D.C.M., Hulscher S.J.M.H., Fulanda B., Kitaka N., Mathooko J.M., *Mar. Freshw. Res.* 66 (2015) 177.
13. Kazi T.G., Arain M.B., Jamali M.K., Jalbani N., Afridi H.I., Sarfraz R.A., Baig J.A., Shah A.Q., *Ecotoxicol. Environ. Saf.* 72 (2009) 301.
14. Kiymaz S., Karadavut U., *Tar. Bir. Der.* 20 (2014) 152.
15. Prasanna M.V, Praveena S.M., Chidambaram S., Nagarajan R., Elayaraja A., *Environ. Earth Sci.* 67 (2001) 1987.
16. Ahmad A.K., Shuhaimi O.M., Lim E.C., Aziz Z.A., *Sains Malaysiana* 42 (2013) 587.
17. Salih S.S., Alkarkhi A.F., Lalung J., Ismail N., *World Appl. Sci. J.* 26 (2013) 75.
18. Elzwayie, A., El-shafie, A., Yaseen, Z.M., Afan, H.A., Allawi, M.F., *Neural Comp. App.* (2016) 1.
19. Gasim, M.B., Toriman, M.E., Muftah, S., Barggig, A., Aziz, N.A.A., Azaman, F., Muhamad, H., *Malay. J. Anal. Sci.* 19 (2015) 1391.
20. Taweel A., Shuhaimi-Othman M., Ahmad A.K., *Biologia (Bratisl).* 68 (2013) 983.
21. Liu C.W., Lin K.H., Kuo Y.M., *Sci. Total Environ.* 313 (2003) 77.
22. Yerel S., Ankara H., *J. Geol. Soc. India* 79 (2012) 89.
23. Hamzah A., Hattasrul Y., *Proceedings of Taal2007: The 12th World Lake Conference*. (2007) 189.
24. Khalik, W.M.A.W.M., Abdullah, M.P., Amerudin, N.A., Padli, N., *J. Mater. Environ. Sci* 4 (2013) 488.