Asian Journal of Water, Environment and Pollution, Vol. 13, No. 2 (2016), pp. 39–48.
DOI 10.3233/AJW-160015

Water Quality Assessment with Varied Lake Depths by Using Multivariate Statistical Approach

Abdul Jalil[1,3,4*], Li Yiping[1,4], Ijaz Ahmad[2] and Khalida Khan[3]

[1]College of Environment, Hohai University, Nanjing, 210098, P.R. China
[2]College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing, 210098, P.R. China
[3]Center for Integrated Mountain Research, University of the Punjab, Lahore, Pakistan
[4]Key Laboratory of Integrated Regulation and Resource Development on Shallow Lakes, Ministry of Education
Hohai University, Nanjing, 210098, P.R. China
✉ jalil_ahmed21@yahoo.com

Received November 28, 2015; revised and accepted March 21, 2016

Abstract: Lakes depth is a most important component to evaluate the impacts on water quality scenarios. Present study discusses the impacts of depth on water quality by using multivariate statistical analysis. This well established phenomenon of correlation between lake depth and water is firstly proved by using multivariate statistical techniques. Depth of Rawal Lake was divided into three groups of surface, middle and bottom to analyze impacts between these stages on water quality. There were sixteen parameters (physico-chemical, bacteriological and metals) analyzed for which samples were collected and analyzed from a fresh water Rawal lake for four seasons in 2012-2013. The statistical correlation was developed between the water quality parameters and between the layers, by using multivariate scatterplot, cluster analysis and discriminant analysis. Results of these statistical techniques revealed a strong correlation (positive or negative) among most of the water quality parameters. Aluminum was found to be with medium variability and temperature at high variability in cluster analysis. Statistically significant correlation was found between the two dimensions of canonical discriminant functions with canonical correlation of 0.957 and 0.586. Therefore, these statistical analyses validated the high impact of depth on different water quality parameters.

Key words: Water quality, statistical analysis, lake depth, discriminant analysis, cluster analysis.

Introduction

In terms of dynamic systems, lakes are considered as highly heterogenic with respect to space and time. The interaction among the components of water quality can be predicted but are more complex at variable depth of the lakes. Due to this complex nature of interactions in lake waters, it is very difficult to find the measurements especially at variable depths by using multivariate statistical techniques and mathematical modelling. Rawal lake is one of the important lake in capital city of Islamabad with approximately 8.8 km^2 area. Observed maximum depth of lake is about 31 m. Kurrang river, Nullah of Shahdara and Noorpur stream are the major contributors of water in Rawal lake. The total documented catchment area of the Rawal lake is approximately 268 km^2 (Aftab, 2010).

The use of statistical analysis tools has become very important tool for the water quality, its classification into different categories and to identify the potential sources/causes of pollutants (Ghumman, 2011). Among them multivariate statistical analyses such as scatter plot, cluster analysis and discriminant analysis are one of the recent adaptations to reduce and classify data

*Corresponding Author

and has wide applications in mineralogy (St Seymour et al., 2004), geochemistry (Papatheodorou et al., 1999; Papatheodorou et al., 2002a), marine geophysics (Papatheodorou et al., 2002b), and hydrochemistry (Voudouris et al., 1997; Lambrakis et al., 2004) etc.

Water quality monitoring and assessment by using multivariate statistical has now become important tool because of the need to get reduced data to analyze and for decision making (Vega et al., 1998; Helena et al., 2000; Lambrakis et al., 2004).

Characterization and evaluation of surface waters by using multivariate statistical tools has become important intervention as described by Reisenhofer et al. (1995), Miller et al. (1997), De Ceballos et al. (1998), Momen et al. (1999), Perona et al. (1999), Lau and Lane (2002), Simeonov et al. (2003) and Yu et al. (2003), for monitoring the quality of groundwater mentioned by Vengosh and Keren (1996), Suk and Lee (1999), Helena et al. (2000), Lambrakis et al. (2004) and Panagopoulos et al. (2006). These are also important for providing novel information about spatial and temporal variations due to natural and human influences and are linked to dynamics of the lake due to mixing of waters in different seasons. Water bodies having distinct boundaries like shallow lakes are most extensively modelled systems. There are many mathematical models which have been used for modelling shallow lakes. Jørgensen (1976) initially used differential equation model. Zhang et al. (2003) developed structurally dynamic models, and artificial neural network models were developed by Recknagel et al. (1997).

Similarly, generalized logistic models by Tan and Beklioglu (2005) have been used to investigate physio-chemical relationships to draw interactions among biotic factors and physio-chemical factors and processes. Multivariate statistical techniques in shallow waters were not being consistently used. De Ceballos et al. (1998) were users of factor analysis for defining parameters set to characterize the state of salinity in three lentic water bodies of Brazil. Similarly, Momen et al. (1999) applied cluster and discriminant analysis for classifying lake nitrogen in 24 lakes of New York. Lau and Lane (2002) applied factor analysis for studying role of NOx, P, Si and zooplankton in a shallow lake.

In present study three different multivariate techniques (multivariate scatter plot, Cluster analysis and discriminant analysis) were applied to analyze the water quality data (physico-chemical, bacteriological and metals) of Rawal Lake at variable depths to predict the concentrations on the discussed levels.

Materials and Methods

Rawal Lake is a small recreational lake in Islamabad (capital of Pakistan) situated at Korang River. It is located at latitude of 33°42′ N, longitude 73°07′ E, and altitude of 1,800 m. Its water is being used for domestic purpose in Islamabad and Rawalpindi (Iqbal et al., 2013a). Recent investigations are the proof of water contamination due to freshly deposited sediment from upper catchment of the study area (Iqbal et al., 2013b).

Present study provides the descriptive analysis of water quality parameters at three levels of lake water (surface, middle and bottom). Sampling provides the true representation of study to avoid the misleading of research conducted. Samples were collected in 2012-2013 in four seasons for estimating water quality and their impacts at variable depths of the lake. There were 14 points selected to collect water samples according to the depth scenarios. pH, EC, DO, TDS and temperature were measured in-situ. EC and temperature were measured with portable conductivity meter (Model HI 8633, M/S HANNA Instruments). pH was measured with portable pH meter (Model PS-19 M/S Corning, Canada). Other physico-chemical and bacteriological measurements were conducted in Pakistan Council for Research in Water Resources (PCRWR) laboratory.

Statistical and Assessment Methods

Before analysis, data was standardized at 0 mean and 1 standard deviation. The data was tested for normality by using Kolmogorov-Smirnov (K-S) and Shapiro-Wilk (S-W) tests. Homogeneity of variance was tested for multivariate evaluation of data by using One-Way ANOVA Levene's statistics test and significance was measured within and between groups. All data was log-transformed for stabilizing the variance and to get normal distribution. Multivariate scatter plot was drawn to establish the relationship between individual parameter of the discussed three groups and to estimate the effect of individual parameter on other ones (Figure 2). Hierarchical agglomerative cluster analysis was conducted by means of the Ward's method using squared Euclidian distances as measure of the similarity to divide the lake into three groups (surface, middle and bottom) and to display the relationship between them in the form of a combine cluster (Figure 3) (Hussain et al., 2008). Similarly, discriminant analysis was conducted to measure the stand alone group by using centroid of each group (Figure 4).

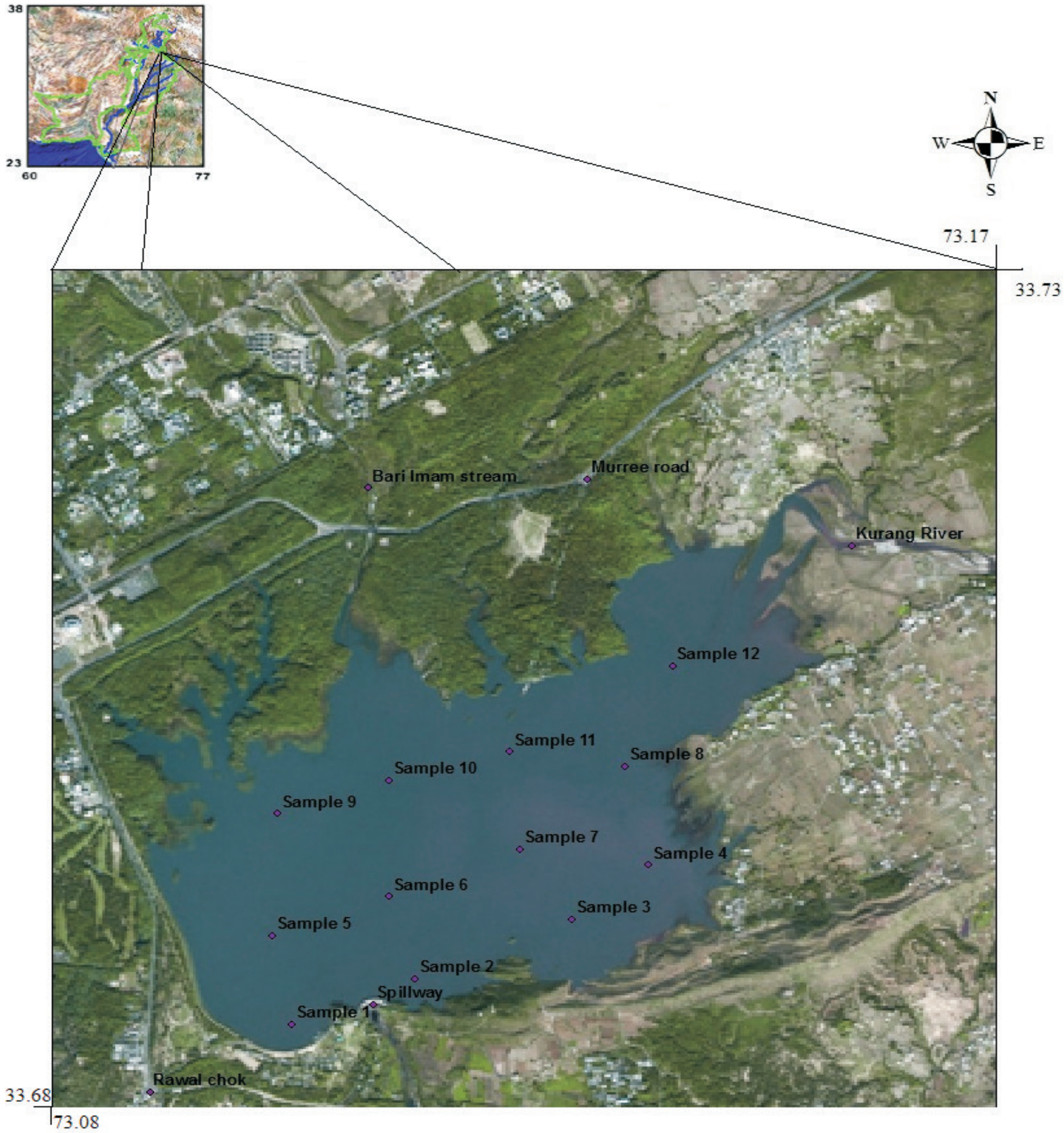

Figure 1: Location of sampling points in the study area of Rawal Lake, Islamabad, Pakistan by using 2.5 m, Spot satellite imagery.

All the statistical analysis was performed by using IBM SPSS Statistics, version 20 (SPSS Inc., Chicago, IL, USA).

Results and Discussion

Current study was conducted to measure the interaction of physico-chemical, bacteriological parameters and metals at three levels, Firstly, between water quality parameters, secondly, between three layers (surface, middle and bottom) and thirdly, at overall status of all parameters at those levels.

These interactions were measured by drawing those scenarios of compiled data using above discussed statistical analysis.

Interactions of Contaminant (Physico-chemical, Bacteriological and Trace Metals)

It could be observed from the scatter plot (Figure 2) that fit line of temperature shows that it has weak positive

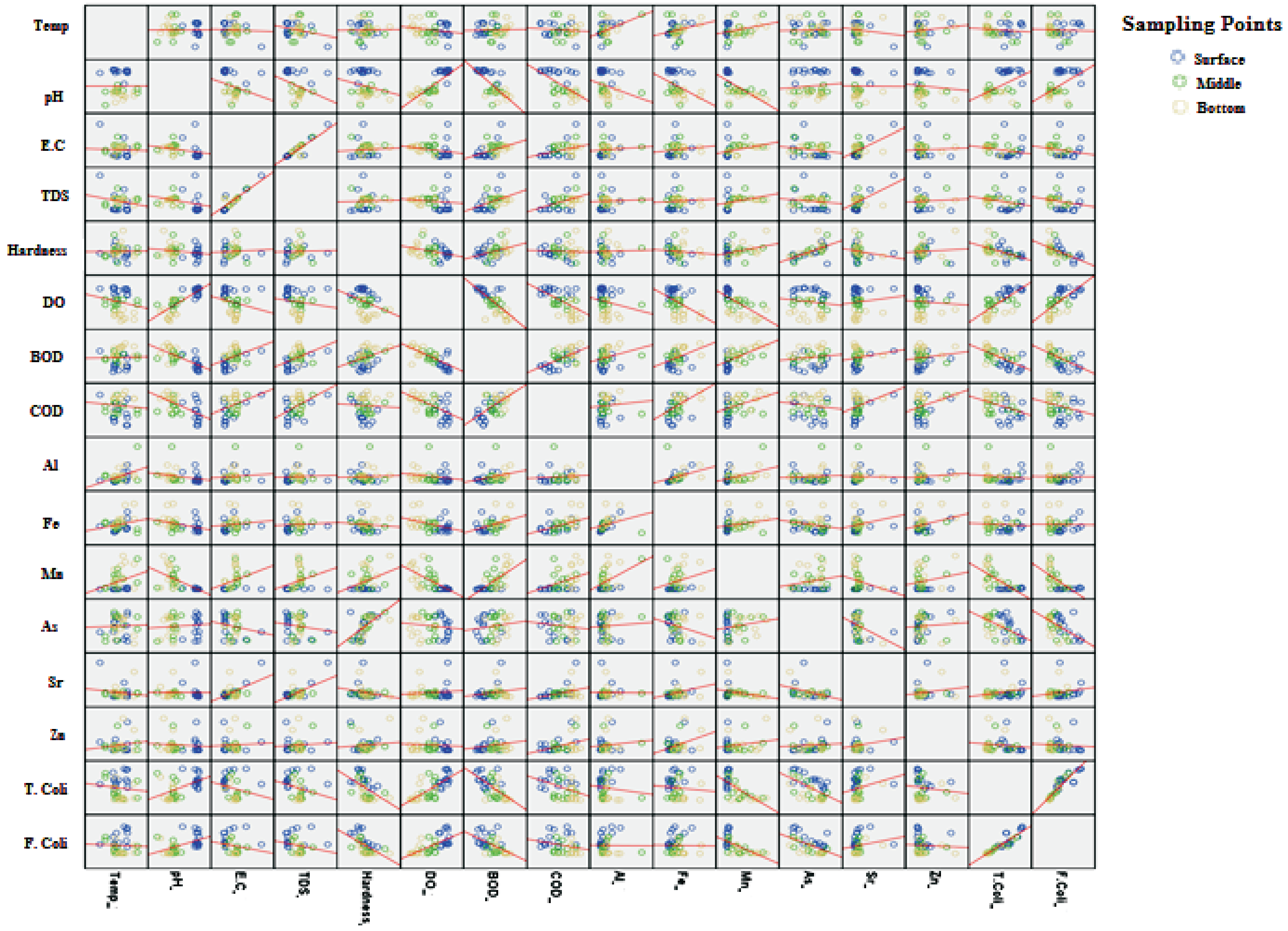

Figure 2: Scatterplot of physico-chemical, metals and bacteriological contaminants of Rawal Lake.

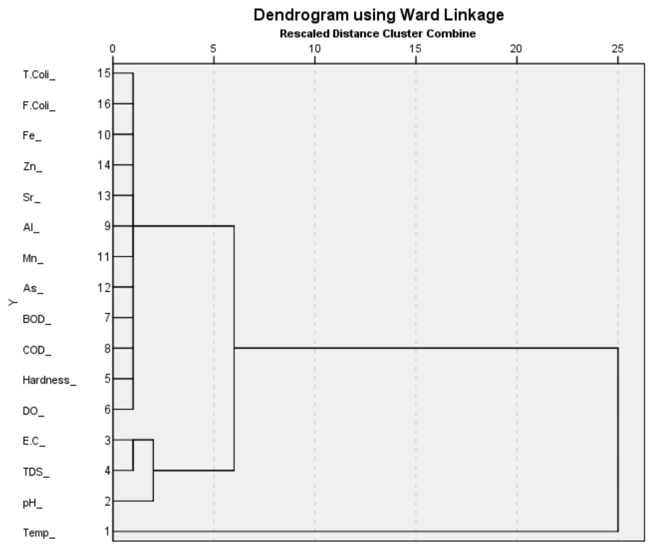

Figure 3: Dendrogram based on agglomerative hierarchical cluster (Ward linkage) of physico-chemical, metals and bacteriological contaminants of Rawal Lake.

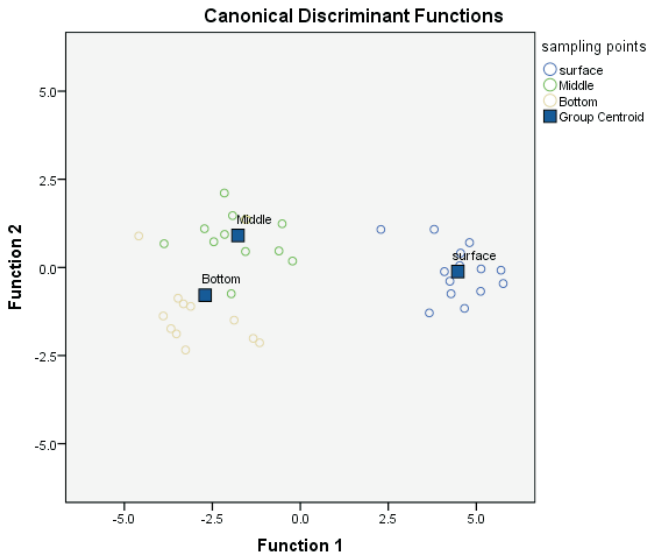

Figure 4: Canonical discriminant function (three layers) of Rawal Lake water quality parameters.

correlation with the metals (Zn, Mn, Fe and Al) and weak negative correlation with TDS, DO, COD, Sr, total coliform and faecal coliform. Weak or no correlation of temperature was found with pH, EC, hardness and BOD. Negative correlation of temperature with most of the other parameters might be explained by coupling effects of increase and decrease of temperature during wet and dry season.

pH has strong positive correlation with DO and weak positive correlation with total and faecal coliforms. pH has weak correlation with EC, TDS, BOD, COD, Al, Fe and Mn. There was no linear correlation of pH found with temperature, hardness, As, Sr and Zn.

Strong positive correlation was found between EC and TDS at all levels (surface, middle and bottom) of lake water. There was moderate positive correlation of EC with COD, BOD, and Sr. Weak positive correlation of EC was found with Zn, Mn, Fe and Al. pH, DO, As, total coliform and faecal coliforms found to have weak negative correlation with EC. Temperature and hardness have no linear correlation with EC.

TDS has weak positive linear correlation with BOD, COD, Mn and Sr and weak negative correlation with temperature, pH, DO, As, total coliform and faecal coliforms. There was no correlation found between TDS and hardness, Al, Fe and Zn.

Hardness and As found to have strong linear correlation, whereas hardness has strong negative correlation with faecal and total coliforms. Hardness has weak negative correlation with pH, DO, Fe and Sr. There is no correlation found between hardness and temperature, EC, TDS, Al, Fe, COD and Zn.

DO has weak positive linear correlation with pH, faecal coliform and total coliforms. BOD and DO have strong negative correlation with each other. DO have weak negative correlation with COD, Fe and Mn. There is weak or no correlation of DO found with temperature, EC, TDS, hardness, Al, As and Zn.

There is strong linear correlation of BOD with COD whereas BOD has weak positive correlation with EC, TDS, hardness, Al, Fe, Mn and As. BOD has strong negative correlation with pH, DO, faecal coliform and total coliforms. COD has weak positive linear correlation with EC, TDS, Fe, Mn, Sr and Zn. COD has weak negative correlation with pH, DO, As, total coliform and faecal coliforms. All the other parameters of water quality of Rawal Lake have vice versa effects compared to above discussed parameters as shown in Figure 2.

Cluster Analysis

Cluster analysis technique was adopted to classify the lake into certain groups according to variability of depth. The dendrogram of the lake obtained by Ward's method is shown in Figure 3. The results indicated that lake water quality parameters can be classified into two statistically significant groups. Cluster I (three clusters group) indicated the low variability of all 14 parameters according to depth of the lake. cluster II (four clusters group) showed the medium variability of parameters; especially Al and Cluster III (five clusters group) indicated high variability of temperature.

Discriminant Analysis

There are two discriminant dimensions formed using the data and results show that both are statistically significant (The canonical correlations for the dimensions one and two are 0.957 and 0.586). As Figure 4 shows that bottom and middle groups of water quality parameters tend to be at the more negative end of the dimension 1 compared to the surface group of water quality parameters which is at highest positive end of dimension 1. On the dimension 2, bottom group is again at negative side but middle group is at positive side of the dimension 2. Surface group do not have significant discriminant variability towards dimension 2. The counterpart data for the selected physico-chemical, bacteriological and metals related to the water samples during all seasons contains 95.4% of variance in function 1 of the canonical discriminant analysis and having 10.94 eigenvalue.

Similarly, function 2 contains 4.6% of variance and having 0.523 eigenvalue between water quality sample groups. Thus, the results of discriminant analysis clearly defined the percent of variance between three groups along function 1 as shown in Table 1 and Figure 4. Surface values have significant difference with middle and bottom. So, it can be seen from the results that

Table 1: Eigenvalues of canonical functions

Function	*Eigenvalue*	*% of variance*	*Cumulative %*	*Canonical correlation*
1	10.944[a]	95.4	95.4	0.957
2	0.523[a]	4.6	100.0	0.586

middle and bottom water quality are more similar to each other than that of surface water quality.

Variance Effects of Individual Parameters at Three Depth Levels

Effect of Depth on Physico-chemical Parameters
At normal temperature and pressure, physico-chemical properties of water were found to have been variable due to submerged aquatic plants at bottom and have small variation in the middle zone quality as compared to surface physico-chemical water quality. Low availability of sunlight is also one of the major causes of differential physico-chemical properties. Figure 5 shows the variance of physical parameters. High frequency of pH values (8.3) at surface and 7.5 at middle were found which indicate the high pH levels at surface. Similarly, temperature, TDS, EC, DO, BOD and COD also showed similar results and found most variance in surface layer. Hardness distribution of variance was very low as in each layer there was a similar trend found.

Effect of Depth on Metals Concentration
Metals concentrations were highest at bottom, lower at middle and lowest at surface of the lake water. Physical properties of metals are strong reason of their concentrations at these levels in the lake water. Discussed metals can be found in stable form when they make the compounds with other elements. Therefore it can be concluded that those contaminants and metals have nearly equally distributed at bottom layer as can be seen in Figures 5 and 6.

Effect of Depth on Bacteriological Contaminants
The bacteriological contamination was found highest at surface and lower at middle of the lake depth. Bacteriological contamination was lowest at bottom at same temperature and pressure which is due to high BOD, COD (BOD, COD has strong negative correlation with coliforms) and low DO (weak positive correlation with coliforms) concentrations in the water of the lake at bottom as compared to the concentrations at middle and surface. Largest amount of total and faecal coliform were found at surface level, than middle and bottom layers as can be seen in Figure 7.

Therefore, the results of discriminant analysis and that of graphical illustrations of individual physical, chemical and bacteriological parameters are compatible and graphical illustrations show the same overall trend of more surface water quality effects than middle and bottom layers.

Conclusion

The multivariate statistical analysis techniques (multivariate scatterplot, cluster analysis and discriminant analysis all together) are firstly adopted in the present study to find out the correlation to process the different water quality parameters. Effect of depth on water quality parameters has been analyzed by using multivariate scatterplot, cluster analysis and discriminant analysis. Multivariate scatterplot has drawn the important correlation (positive and negative) among water quality parameters and has shown a significant reliability of most of the parameters on each other. There were two significantly different groups developed in cluster analysis showing the medium variability of Al, and high variability of temperature as compared to other water quality parameters. Discriminant analysis showed the significant correlation among three layers (surface, middle and bottom) of lake water quality parameters. Therefore, it is concluded that there is strong correlation of water quality parameters with each other and lake depth has significant impact in terms of variation of water quality parameters of the lake. The level and significance of impacts increases or decreases by the physical phenomenal effects on lake such as watershed conditions, flooding, seasonal variation of water levels, turbidity, winds/tides causing mixing of water, solar radiations penetration, aquatic plants, fishes (amount and types), etc. Present study is evident that resuspension of sediments and nutrients from the bottom of the lake are also major source of contamination in the waters of the lake. Therefore, future studies will be needed to focus on the pattern and scenarios of sediment/nutrient resuspension in the shallow lakes according to the hydrodynamic conditions through the development of direct linkage with water depth.

Acknowledgements

Present study was conducted by Center for Integrated Mountain Research, University of the Punjab, Pakistan. Center for Integrated Mountain Research (CIMR) arranged field observations and samples were tested in PCRWR laboratory. Therefore, we are thankful to the Muhammad Ashraf from PCRWR for his help for testing the samples and we would also like to thank Muhammad Yaseen from Pakistan Agricultural Research Council (PARC) for guiding and facilitating us during field observations.

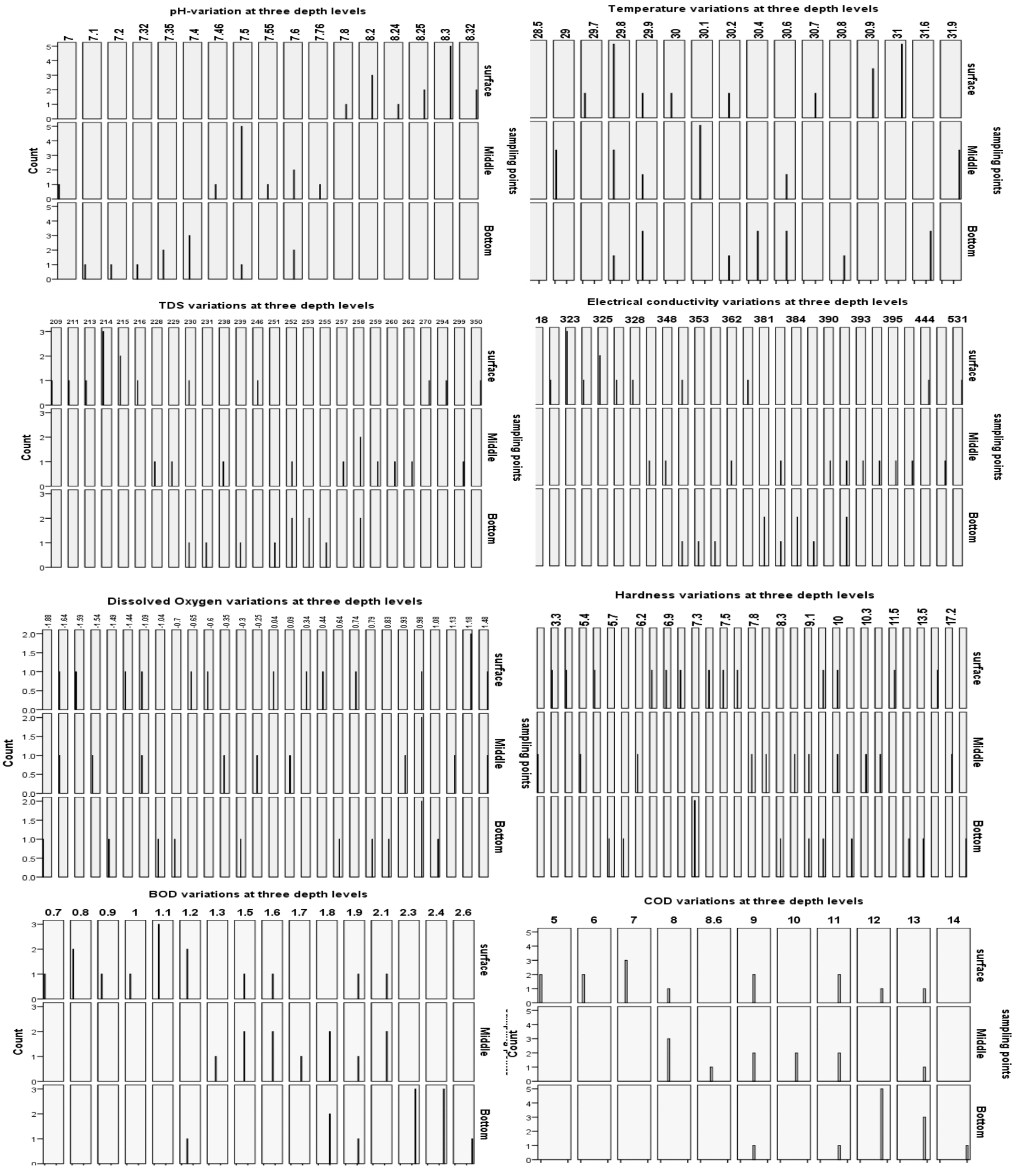

Figure 5: Variation distribution graphs of physico-chemical parameters at varied lake depths.

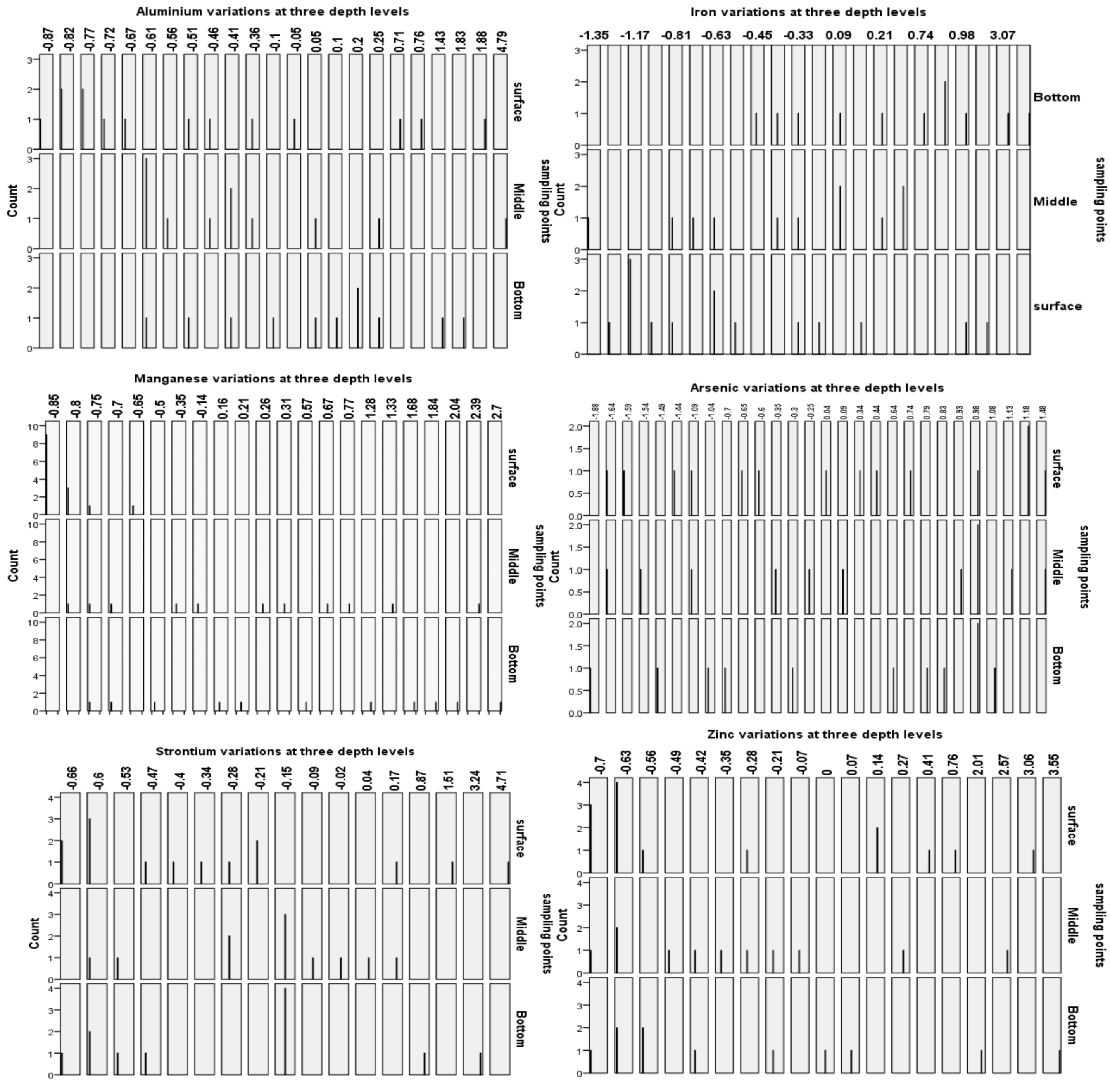

Figure 6: Variation distribution graphs of metals at varied lake depths.

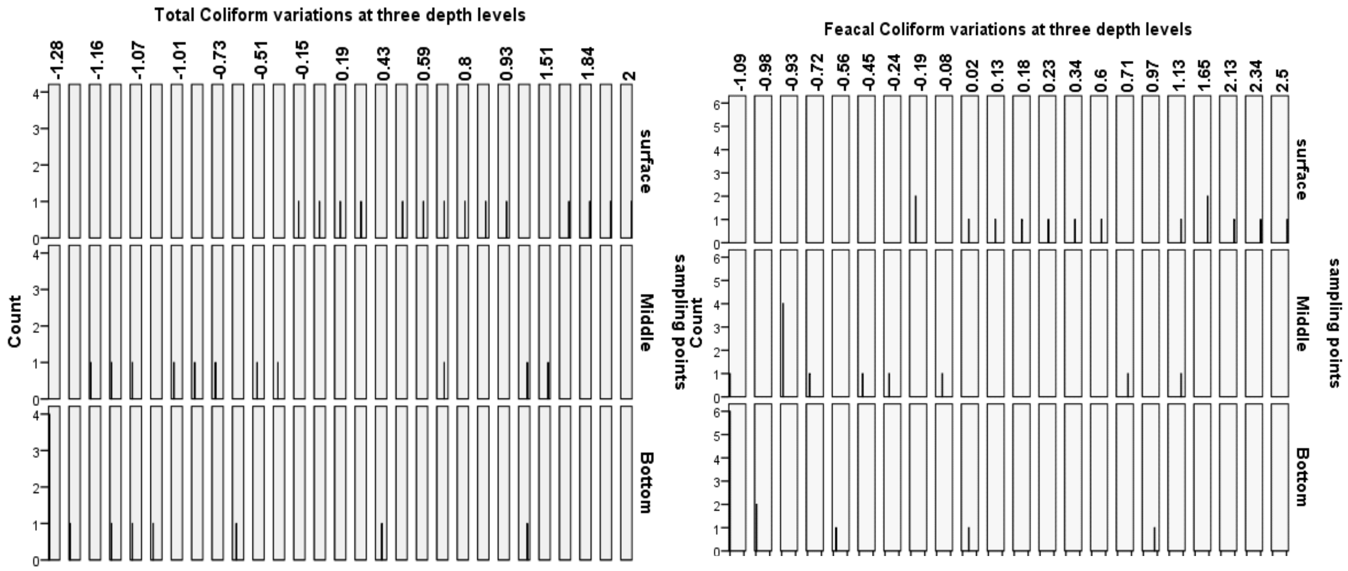

Figure 7: Variation distribution graphs of bacteriological parameters at varied lake depths.

References

Aftab, N. (2010). Haphazard colonies polluting Rawal Lake. *Daily Times*, Monday, March 01.

De Ceballos, B., König, A. and J. De Oliveira (1998). Dam reservoir eutrophication: A simplified technique for a fast diagnosis of environmental degradation. *Water Research*, **32:** 3477-3483.

Ghumman, A.R. (2011). Assessment of water quality of Rawal Lake by long-time monitoring. *Environmental Monitoring and Assessment*, **180:** 115-126.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J.M. and L. Fernandez (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research*, **34:** 807-816.

Hussain, M., Ahmed, S.M. and W. Abderrahman (2008). Cluster analysis and quality assessment of logged water at an irrigation project, eastern Saudi Arabia. *Journal of Environmental Management*, **86:** 297-307.

Iqbal, J., Shah, M.H. and G. Akhter (2013a). Characterization, source apportionment and health risk assessment of trace metals in freshwater Rawal Lake, Pakistan. *Journal of Geochemical Exploration*, **125:** 94-101.

Iqbal, J., Tirmizi, S.A. and M.H. Shah (2013b). Statistical apportionment and risk assessment of selected metals in sediments from Rawal Lake (Pakistan). *Environmental Monitoring and Assessment*, **185:** 729-743.

Jørgensen, S.E. (1976). A eutrophication model for a lake. *Ecological Modelling*, **2:** 147-165.

Lambrakis, N., Antonakos, A. and G. Panagopoulos (2004). The use of multicomponent statistical analysis in hydrogeological environmental research. *Water Research*, **38:** 1862-1872.

Lau, S. and S. Lane (2002). Biological and chemical factors influencing shallow lake eutrophication: A long-term study. *Science of the Total Environment*, **288:** 167-181.

Miller, C.V., Denis, J.M., Ator, S.W. and J.W. Brakebill (1997). Nutrients in streams during baseflow in selected environmental settings of the Potomac River Basin. Wiley Online Library.

Momen, B., Zehr, J.P., Boylen, C.W. and J.W. Sutherland (1999). Determinants of summer nitrate concentration in a set of Adirondack lakes, New York. *Water, Air, and Soil Pollution*, **111:** 19-28.

Panagopoulos, G., Antonakos, A. and N. Lambrakis (2006). Optimization of the DRASTIC method for groundwater vulnerability assessment via the use of simple statistical methods and GIS. *Hydrogeology Journal*, **14:** 894-911.

Papatheodorou, G., Hotos, G., Geraga, M., Avramidou, D. and T. Vorinakis (2002a). Heavy metal concentrations in sediments of Klisova Lagoon (southeast Mesolonghi-Aetolikon Lagoon Complex), W. Greece. *Fresenius Environmental Bulletin*, **11:** 951-956.

Papatheodorou, G., Lyberis, E. and G. Ferentinos (1999). Use of factor analysis to study the distribution of metalliferous bauxitic tailings in the seabed of the Gulf of Corinth, Greece. *Natural Resources Research*, **8:** 277-286.

Papatheodorou, G., Mitsis, C., Christodoulou, D. and G. Ferentinos (2002b). A multivariate statistical approach to the investigation of pockmarks growth and activity. An example from a pockmark field in the Gulf of Patras (W Greece). Proceedimgs of the Eighth Annual Conference of the IAMG, 2002.

Perona, E., Bonilla, I. and P. Mateo (1999). Spatial and temporal changes in water quality in a Spanish river. *Science of the Total Environment*, **241:** 75-90.

Recknagel, F., French, M., Harkonen, P. and K.-I. Yabunaka (1997). Artificial neural network approach for modelling and prediction of algal blooms. *Ecological Modelling*, **96:** 11-28.

Reisenhofer, E., Picciotto, A. and D. Li (1995). A factor analysis approach to the study of the eutrophication of a shallow, temperate lake (San Daniele, North Eastern Italy). *Analytica Chimica Acta*, **306:** 99-106.

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

St Seymour, K., Christanis, K., Bouzinos, A., Papazisimou, S., Papatheodorou, G., Moran, E. and G. Dénès (2004). Tephrostratigraphy and tephrochronology in the Philippi peat basin, Macedonia, Northern Hellas (Greece). *Quaternary International*, **121:** 53-65.

Suk, H. and K.K. Lee (1999). Characterization of a ground water hydrochemical system through multivariate analysis: Clustering into ground water zones. *Groundwater*, **37:** 358-366.

Tan, C.O. and M. Beklioglu (2005). Catastrophic-like shifts in shallow Turkish lakes: A modeling approach. *Ecological Modelling*, **183:** 425-434.

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

Vengosh, A. and R. Keren (1996). Chemical modifications of groundwater contaminated by recharge of treated sewage effluent. *Journal of Contaminant Hydrology*, **23:** 347-360.

Voudouris, K., Lambrakis, N., Papatheothorou, G. and P. Daskalaki (1997). An application of factor analysis for the study of the hydrogeological conditions in Plio-Pleistocene aquifers of NW Achaia (NW Peloponnesus, Greece). *Mathematical Geology*, **29:** 43-59.

Yu, S., Shang, J., Zhao, J. and H. Guo (2003). Factor analysis and dynamics of water quality of the Songhua River, Northeast China. *Water, Air, and Soil Pollution*, **144:** 159-169.

Zhang, J., Jørgensen, S.E., Tan, C.O. and M. Beklioglu (2003). A structurally dynamic modelling—Lake Mogan, Turkey as a case study. *Ecological Modelling*, **164:** 103-120.