INSTITUT FÜR
UMWELTWISSENSCHAFTEN
LANDAU

UNIVERSITÄT
KOBLENZ · LANDAU

Module: ACP 1 Water Analysis

Water Analysis SS-2020

An abstract

on

WATER QUALITY ASSESSMENT OF SELECTED WATER STREAMS IN LANDAU WITH WATER QUALITY INDEX AND MULTIVARIATE STATISTICS

Institute of Environmental Sciences

University Koblenz-Landau

Campus Landau

Fortstraße 7

76829 Landau

Submitted By

Md. Dider Hossain

Matriculation Number: 218203108

Email: hoss8144@uni-landau.de

Date of Submission: 31.12.2020

Supervisor:

Dr. Dörte Diehl

Dr. Clara Mendoza-Lera

1. Introduction: River act like a reservoir that holds the industrial, municipal wastewater and water that comes from nearby lands which is the source of water for human consumption (Khaledian et al. 2018). River water polluted by agrochemical and industrial pollutants is one of the biggest environmental problems in the world (Lam et al. 2012). Water quality assessment helps us to better understand the river water quality and to detect the sources of pollution (Hua et al. 2016). The study was conducted on the Queich river at Offenbach, Landau (49°12'07.1"N 8°11'23.4"E). The river Queich is located in the Rhineland Palatinate state that is mostly covered by sandstone and flowing within the upper Rhine valley. (Mbaka et. al. 2014) The objective of this study is to determine the water quality conditions. Water quality parameters including electrical conductivity (EC), total hardness (KH), pH, calcium (Ca^{2+}), potassium (K^{+}), sodium (Na^{+}), sulphate (SO_4^{2-}), magnesium (Mg^{2+}), phosphate (PO_4^{3-}), biochemical oxygen demand (BOD), dissolve oxygen (DO), nitrate (NO_3^-), acid neutralization capacity (ANC) was measured. Water quality index (WQI), cluster analysis (CA), principal component analysis (PCA) help to find out the changes of water quality (Khaledian et al. 2018). WQI is an important tool that gives a unique rating to finalize water quality status in a single value (Tyagi et al. 2013). CA is an important analyzing tool that classified different objects into groups on the basis of Euclidean distance. The resulting clusters show high similarity within the cluster's member and heterogeneity between the clusters (Salah et al. 2012). PCA analysis was conducted to determine the main water quality parameters that have an influence on the variation of water quality (Mustapha & Abdu 2012). To evaluate the water quality assessment, I asked the following question, (Q1) what is the water quality status of the sampling points? (Q2) which parameters have an influence on the water quality variations? (Q3) Does the effluent water have an influence on the water quality of the river? (Q4) To what extent physical and chemical variables are correlated with each other? To effectively monitor the quality of water and finding important variables that have influenced water quality changes, many researchers used CA and PCA (Facchinelli et al. 2001, Khaledian et al. 2018, Wu & Kuo 2012, Bhat et al. 2014). With these findings, I hypothesized that CA and PCA may be the effective tools to evaluate the quality of water and finding out the significant important variables for water quality changes (H_1).

2. Materials and Method: Ten water samples from three separate streams (river, effluent, river-effluent mixture) were collected in plastic bottles. Each sample was collected from downstream to upstream. Physical or chemical parameters such as pH, EC, DO were determined at sampling points and other chemical parameters such as BOD, KH. Ca^{2+}, Mg^{2+}, SO_4^{2-}, PO_4^{3-}were determined in the lab. All lab analysis had done following the directive of the 'Water Analysis-Manual' arranged by Diehl & Mendoza-Lera (2020). According to Brown et al. (1972), WQI value was determined for each stream to identify water quality status. For statistical analysis, data imputation had been carrying out to replace missing data with substitute value. Shapiro-Wilk's normalization test was performed to check the normality of data and non-normal data were log or Tukey transformed to get more accurate assumptions of normality. Tukey HSD test had been performing to identify the significant difference between each sampling stream. Normalization had been done before CA and PCA which scaling all of our variables to have unit variance. CA was performed to classify all of our water sampling points into groups according to the similar chemical composition that using all of the variances of our whole data set. Hierarchical clustering had performed that was determined by the complete linkage method as well as Euclidean distance was calculated as a function of similarity coefficient. (Khaledian et al. 2018) Finally, PCA was performed to find out the relationship between water quality variables and to detect significant variables that have effect on water quality variations (Mustapha & Abdu 2012). All statistical analysis was conducted by R Studio^{R} (Version 1.3.1093).

3. Results: To determine the significant differences between each treatment, Tukey HSD test was performed. The test shows that NO_3^-(0.037), SO_4^{2-}(5.7×10^{-5}), Na^+(0.0001), Ca^{2+}(0.001), pH (0.0007), EC (0.0001), PO_4^{3-}(0.020) and DO (0.0002) significantly increased with significant p-value ($p<0.05$) when effluent water mixing with river water. From the water quality index value, it is clear that river water quality status is good with WQI value 33.43 as well as the water quality status of effluent and mixing water samples are excellent with the WQI value 17.54 and 20.86 respectively (**Table 1**). **Figure 1** represents the graphical summary of cluster analysis that was carried out with a normalized dataset using a complete linkage method to calculate Euclidean distance as a function of similarity among the water quality parameters for all water sampling sites. The dendrogram classified all of the sampling point into three statistically significant cluster. Cluster 1 contains River1 and River2, cluster 2 comprises Effluent1 and Effluent 2, cluster 3 involve Mixing1-Mixing6 sampling site respectively.

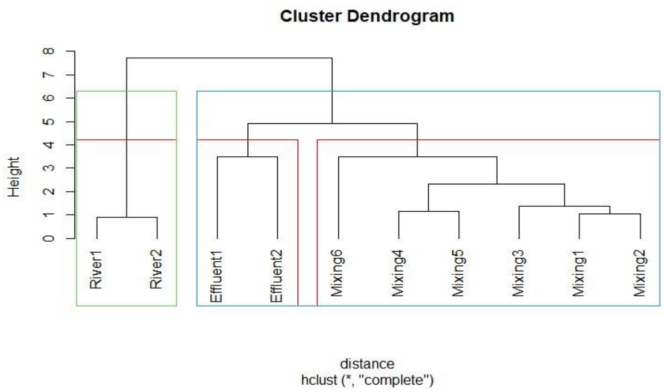

Figure 1: Cluster dendrogram for water sampling site

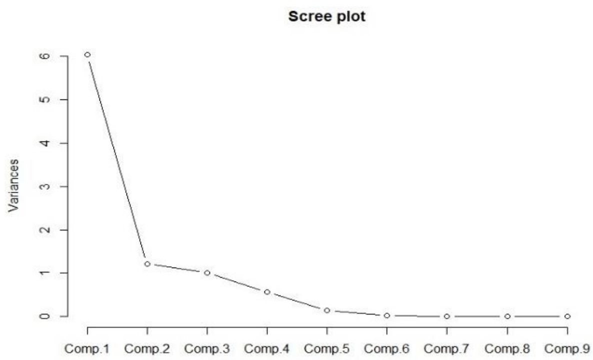

Figure 2: Scree plot for principal components

On the other hand, beyond the Euclidean distance of 6, sampling sites can be classified into two clusters. Cluster1 contains with River1 and River2, cluster2 comprises Effluent1, Effluent2 and Mixing1-Mixing6 sampling sites respectively. **Figure 2** shows principal components in terms of variances. From this graph, it is clear that first three PCs accounted most of the variability of our datasets, beyond that drop in variability within two components is not much significant. **Table 4** shows the summary of principal component analysis that consists of parameters loading, eigenvalues, amount of variability that explained by each component and cumulative variance. The results depend on the loadings of parameter for PC1, PC2 and PC3. PC1 describes 66.95% variance of our datasets with eigen values (6.025) that has less positive loading on EC, SO_4^{2-}, Na^+, Ca^{2+} and less negative loading on pH and PO_4^{3-}. PC2 accounts 13.47% variability of the whole data with eigen value (1.21)

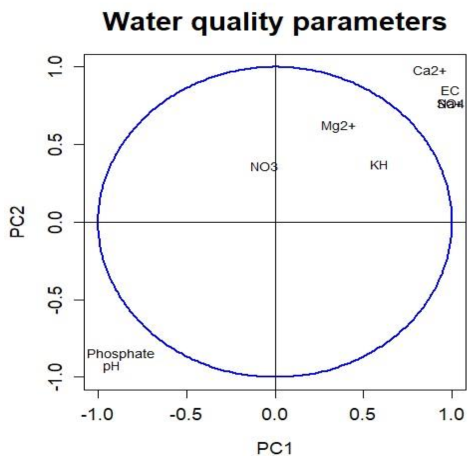

Figure 3: Correlation of original variables with PC1 and PC2

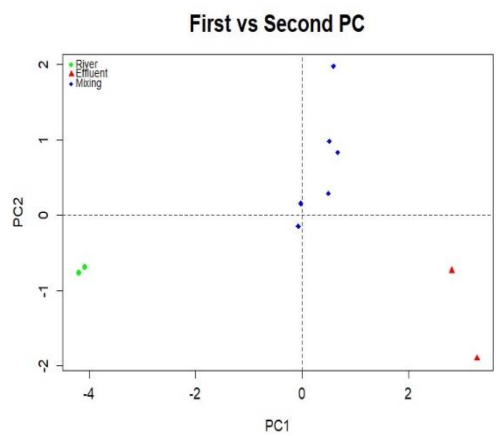

Figure 4: PC1 vs PC2 for all sampling sites

and has less positive loading on KH, SO_4^{2-}, Na^+, pH, PO_4^{3-}, EC and strong negative loading on NO_3^- as well as less negative loading on Mg^{2+} and Ca^{2+}. **Figure 3** shows pH, PO_4^{3-}, Ca^{2+}, Na^+, SO_4^{2-}, EC corresponded to correlations near the periphery of the circle that means these variables are well explained by the first two pc components where pH, PO_4^{3-} is negatively correlated and Ca^{2+}, Na^+, SO_4^{2-}, EC are positively correlated with each other. The correlation matrix of variables for PC1 and PC2 has shown in **Table 5.** The variance percentage of KH and NO_3^- variables explained by the first two PCs is relatively small (**Table 4 & Table 5**). From the component rotation matrix (**Table 5**), the most important contributors for PC1 are pH and EC variables. The second PC is best described by the NO_3^- and sum of hardness, SO_4^{2-}, Na^+, pH, EC and PO_4^{3-} variables. If we look simultaneously at **Figure 3** and **Figure 4** that shows that the effluent and mixing sampling locations are roughly characterized by large values of Ca^{2+}, Na^+, SO_4^{2-}, EC variables and smaller values of pH, PO_4^{3-} variables.

4. Discussion: To visualize the actual effects and to answer our research question Q3, boxplots have been creating for all variables (**Figure 8**). If we consider the river as control and looking at Tukey HSD test value and **Figure 8** simultaneously, it has cleared that NO_3^-, SO_4^{2-}, Na^+, Ca^{2+}, EC significantly increased as well as pH, PO_4^{3-}, DO significantly decreased when the river water was mixed with effluent water. The average pH value of the mixing water sample was 7.23 which means that the mixing water sample was slightly acidic. The amount of nitrogen and PO_4^{3-} can be increased for several reasons such as the higher quantity of organic matter decomposition, disposal of industrial waste, water runoff from surrounding agricultural field which have NO_3^- or phosphate-containing fertilizer. (Hamil et al. 2018, Zhao & Cui 2009) According to Hamil et al. (2018), high quantity of NO_3^- or PO_4^{3-} presence in water indicating high organic pollution and eutrophication in water. On the other hand, higher values of EC indicate the high amount of salt presenting in the water due to geochemical reactions which caused by human activities, the high degree of anthropogenic activities for example disposal of waste, agricultural runoff (Khaledian et al. 2018, Hamil et al. 2018). Also, EC can be increased for natural changes. The higher amount of Na^+ in water can come from water treatment and the high amount of SO_4^{2-} presence in water can be caused by human activities, natural pollution. (Khaledian et al., 2018) According to Shukla et al. (2013), the low quantity of DO is present in the water when the free O_2 removed by bacteria through respiration and the organic matter decomposition in water is high.

For cluster analysis, the scree plot (clusters number vs within group of sums of the square) shows that a drop in variability or within the group of sums of squares among the first three clusters is very large (**Figure 6)**. It means that we should go for lower number of clusters, maybe two or three, beyond that the results are not very significant. This is why we considered three or two cluster for our datasets. **Figure 5** shows the average Silhouette width (S_i) value that indicates whether the cluster formation had been good or not. If the S_i value is high, the members in the cluster are closure to each other. S_i (0.82) value of cluster 1 (River 1 & River 2) is very higher than cluster 2 (0.14) and cluster 3 (0.49). So, it can be said that effluent and mixing sampling point water qualities are likely to be same (**Figure1**). This finding is validated by water quality index values. According to Brown et al. (1972), WQI value had calculated by considering the WHO standard for water quality as a standard value (**Table 1, Table 2**). The result shows that river water quality is good with WQI value 33.43 and effluent water, mixing water are excellent with WQI value 17.54 and 20.86 respectively. With these results, the answer for research question Q1 can be interpreted.

PCA had performed to determine those water quality parameters that have significant influence on water quality variations. Component coefficient or factor loading have classified as strong (>0.75), moderate (0.75-0.50), less (<0.4) that relating to actual loading (Zhao & Cui 2009). Bi-plot of PC1 vs PC2 for water quality variables (**Figure 7**) shows that most influential variables for PC1 is Ca^{2+}, EC, SO_4^{2-}, Na^+ and for PC2 the most influential variable is NO_3^-. From correlation matrix (**Table 5**), it is found that Ca^{2+}, EC, SO_4^{2-}, Na^+ have strong positive correlation, KH and Mg^{2+} have less positive correlation as well as pH and PO_4^{3-} have strong negative correlation. **Figure 3** and **Figure 4** shows that the effluent and mixing water have large values of Ca^{2+}, Na^+, SO_4^{2-}, EC as well as smaller values of pH, PO_4^{3-} and river samples have large values of pH, PO_4^{3-}. The cluster means for all water quality variables shows the same result (**Table 3**). For example, the river water in cluster 1 has a higher value of pH (1.710) and PO_4^{3-}(1.733) whereas the effluent water in cluster 2 seems to have a lower-than-average value of pH (-0.960) and PO_4^{3-}(-0.883). These averages indicate which variables are really playing an important role in characterizing the clusters. With these results, we find the answers for research question Q2 and Q4 that were about the most influential variables for water quality variables and correlation among the water quality parameters. In a nutshell, we can also accept our hypothesis H_1 which describes that the PCA and CA are an effective tool to describe the water quality variations.

4. Conclusion: Overall, we can say that CA and PCA can effectively characterize water quality status. CA classified all of the water sampling point into three cluster, river water in cluster1 was largely characterized by pH, PO_4^{3-}. Effluent water in cluster2 has higher value of KH, EC, SO_4^{2-}, Ca^{2+}, Mg^{2+}, Na^+ and mixing water in cluster 3 has higher value of NO_3^- (**Table 3**). WQI values determine the actual water quality status where river water is describing as in good condition as well as effluent and mixing water samples are describing as in excellent condition. PCA determined that NO_3^-, Ca^{2+}, Na^+, EC, SO_4^{2-}, KH are the significant variables for changing the water quality. And Tukey HSD test shows that NO_3^-, SO_4^{2-}, Na^+, Ca^{2+}, pH, EC, PO_4^{3-} and DO was significantly increased when river water comes into the contact with the effluent water.

References

Bhat, S. A., Meraj, G., Yaseen, S., & Pandit, A. K. (2014). Statistical Assessment of Water Quality Parameters for Pollution Source Identification in Sukhnag Stream: An Inflow Stream of Lake Wular (Ramsar Site), Kashmir Himalaya. Journal of Ecosystems, 2014, 1–18. doi:10.1155/2014/898054

Brown RM, McClelland NI, Deininger RA, O'Connor MF (1972) A water quality index—crashing the psychological barrier. In: Indicators of environmental quality

Diehl, D. & Mendoza-Lera, C. (2020). Water analysis course manual (unpublished).

Facchinelli, A., Sacchi, E., & Mallen, L. (2001). Multivariate statistical and GIS-based approach to identify heavy metals sources in soils. Journal of Environ Pollution 114(3), 313–324. doi:10.1016/s0269-7491(00)00243-8

Hamil, S., Arab, S., Chaffai, A., Baha, M., & Arab, A. (2018). Assessment of surface water quality using multivariate statistical analysis techniques: a case study from Ghrib dam, Algeria. Arabian Journal of Geosciences, 11(23), 754(1-14). doi:10.1007/s12517-018-4102-5

Hua, A. K., Kusin, F. M., & Praveena, S. M. (2016). Spatial Variation Assessment of River Water Quality Using Environmetric Techniques. Polish Journal of Environmental Studies, 25(6), 2411-2421. DOI: 10.15244/pjoes/64082

Khaledian, Y., Ebrahimi, S., Natesan, U., Basatnia, N., Nejad, B. B., Bagmohammadi, H., & Zeraatpisheh, M. (2018). Assessment of Water Quality Using Multivariate Statistical Analysis in the Gharaso River, Northern Iran. Urban Ecology, Water Quality and Climate Change, Volume 84, 227–253. doi:10.1007/978-3-319-74494-0_18

Lam, Q.D., Schmalz, B., & Fohrer, N. (2012). Assessing the spatial and temporal variations of water quality in lowland areas, Northern Germany., 438-439(none), 137–147. doi:10.1016/j.jhydrol.2012.03.011

Mbaka, J. G., Somlai, C., Köpfer, D., Maeck, A., Lorke, A., & Schäfer, R. B. (2014). *Methane-Derived Carbon in the Benthic Food Web in Stream Impoundments. PLoS ONE, 9(10), e111392.* doi:10.1371/journal.pone.0111392

Mustapha, A., & Abdu, A. (2012). Application of Principal Component Analysis & Multiple Regression Models in Surface Water Quality Assessment. Journal of Environment and Earth Science, 2(2), 2224-3216.

Salah, E. A. M., Turki, A. M., & Al-Othman, E. M. (2012). Assessment of Water Quality of Euphrates River Using Cluster Analysis. Journal of Environmental Protection, 3, 1629-1633. DOI: 10.4236/jep.2012.312180

Shukla, D., Bhadresha, K., Jain, N.K., & Modi, H. A. (2013). Physicochemical Analysis of Water from Various Sources and Their Comparative Studies. IOSR Journal of Environmental Science, 5(3), 89-92. ISSN: 2319-2399

Tyagi, S., Bhavtosh Sharma, B., Singh, P., & Dobhal, R. (2013). Water Quality Assessment in Terms of Water Quality Index. American Journal of Water Resources, 1(3), 34-38. DOI:10.12691/ajwr-1-3-3

Wu, E. M., & Kuo, S. (2012). Applying a Multivariate Statistical Analysis Model to Evaluate the Water Quality of a Watershed. Water Environment Research, 84(12), 2075–2085. doi:10.2175/106143012X13415215906979

Zhao, Z., & Cui, F (2009). Multivariate statistical analysis for the surface water quality of the Luan River, China. Journal of Zhejiang University SCIENCE A, 10(1), 142-148. ISSN 1862-1775

Appendix

Table 1: Water quality status (Brown et al. 1972)

Site	Index value	WQI range	Status
River	33.43	26-50	Good
Effluent	17.54	0-25	Excellent
Mixing	20.86	0-25	Excellent

Table 2: Water quality standard for our variables according to WHO

Parameters	WHO Standard
pH	8.5
TH (mmol/L)	500
EC (ms/m)	300
Sulphate (mg/L)	250
Nitrate (mg/L)	50
Calcium (mg/L)	200
Magnesium (mg/L)	150
Sodium (mg/L)	60
Phosphate (mg/L)	1

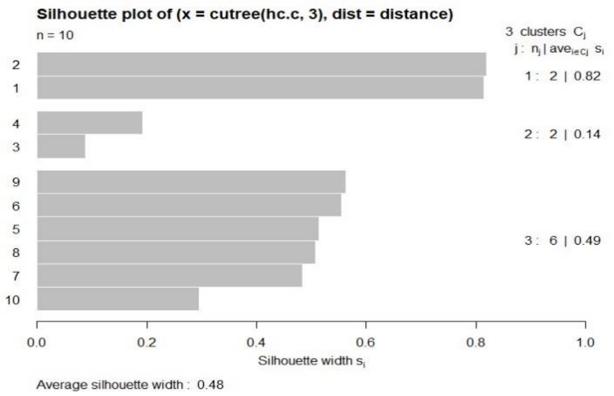

Figure 5: Silhouette values for three clusters

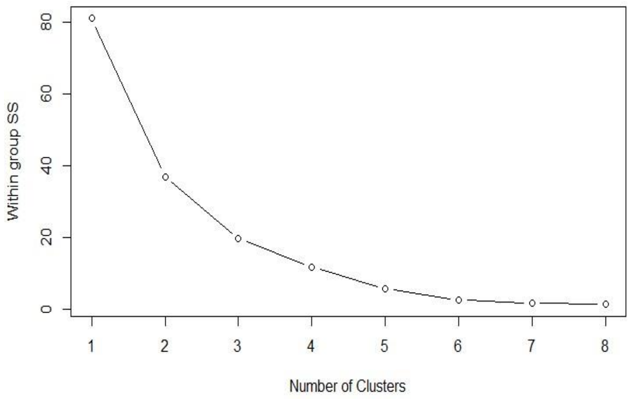

Figure 6: Scree plot of clusters number vs withing group SS

Table 3: Within cluster means for water quality variables

Parameters	Cluster1 (River 1&2)	Cluster2 (Effluent 1&2)	Cluster3 (Mixing 1-6)
pH	1.710	-0.960	-0.249
KH	-0.876	0.865	0.003
EC	-1.594	1.314	0.093
Sulphate	-1.425	1.536	-0.037
Nitrate	-0.707	-1.131	0.612
Calcium	-1.701	0.944	0.252
Magnesium	-0.968	0.414	0.184
Sodium	-1.431	1.519	-0.029
Phosphate	1.733	-0.883	-0.283

Table 4: Principal components loading for water quality parameters

Parameters	PC1	PC2	PC3
Hardness	0.243	**0.321**	**0.353**
Nitrate	0.023	**-0.816**	**0.405**
Sulphate	**0.389**	0.208	0.051
Sodium	**0.389**	0.197	0.071
Calcium	**0.379**	-0.238	-0.207
Magnesium	**0.192**	-0.195	**-0.800**
pH	**-0.388**	**0.191**	-0.077
EC	**0.399**	0.063	0.098
Phosphate	**-0.379**	**0.106**	-0.049
Eigen values	6.02	1.21	1.01
SD	2.45	1.10	1.00
Variance (%)	66.95	13.47	11.28
CV (%)	66.95	80.42	91.71

*SD = Standard deviation
*CV = Cumulative variance

Table 5: Correlation and component rotation matrix for variables

	Correlation Matrix		**Component Rotation Matrix**	
Parameters	**PC1**	**PC2**	**PC1**	**PC2**
Hardness	0.585	0.376	0.243	**-0.321**
Nitrate	-0.604	0.364	0.023	**0.816**
Sulphate	0.994	0.766	0.389	**-0.208**
Sodium	0.994	0.767	0.389	**-0.197**
Calcium	0.873	0.982	0.379	0.238
Magnesium	0.363	0.616	0.192	0.195
pH	-0.921	-0.930	**-0.388**	**-0.191**
EC	0.991	0.854	**0.399**	**-0.063**
Phosphate	-0.874	-0.849	-0.379	**-0.106**

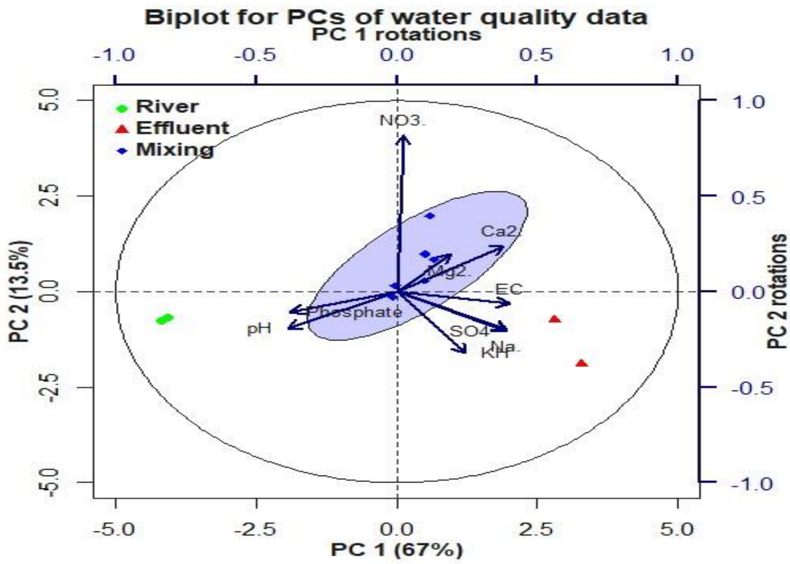

Figure 7: PC1 vs PC2 biplot for water quality parameters

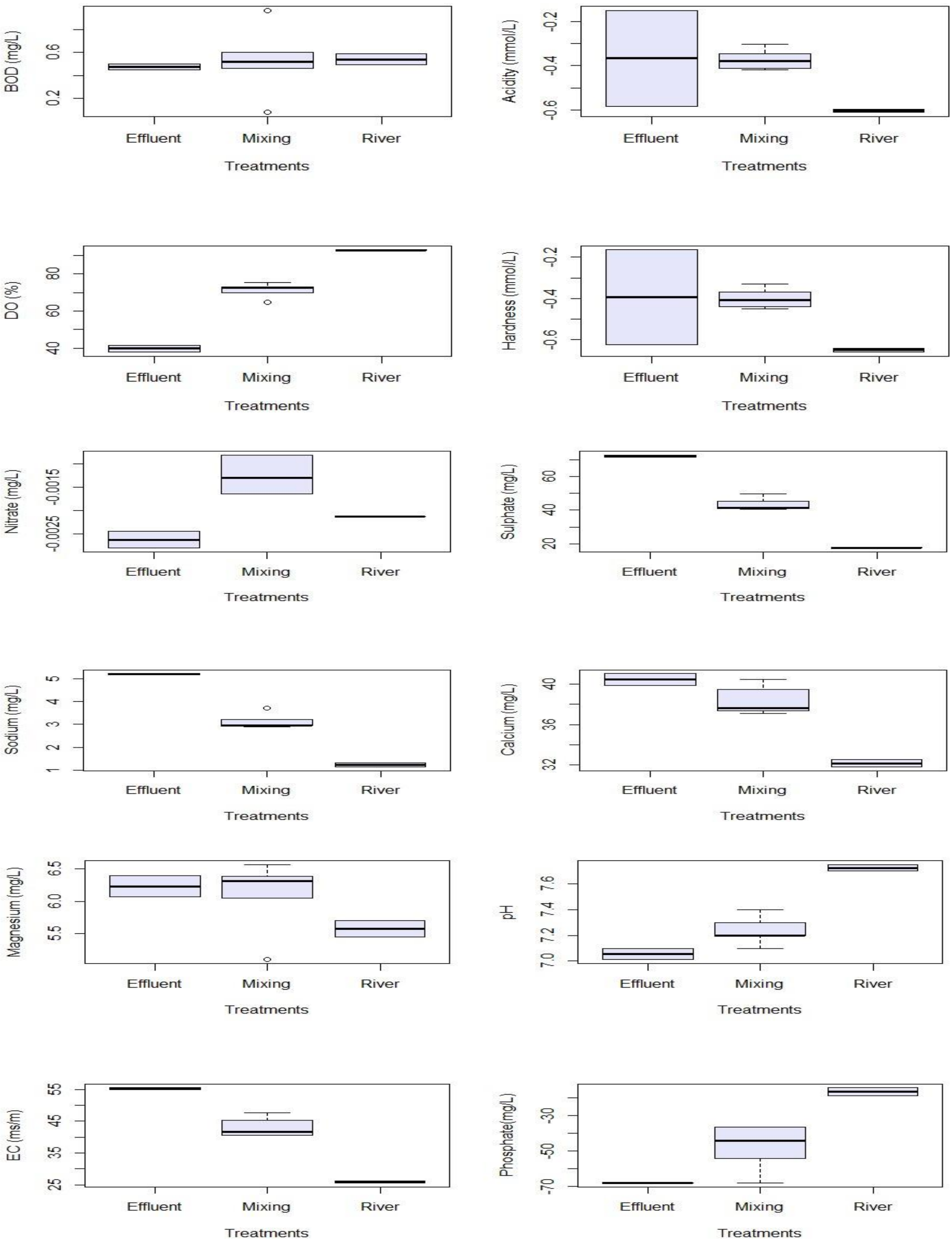

Figure 8: Boxplot of water quality variables for different treatment