Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 5 Issue 11, November - 2018

A Chemical Assessment of Water Quality of the Detroit River Using Multivariate Analysis

Farinaz Bozorgnia and Herman Tang
School of Engineering Technology
Eastern Michigan University
Ypsilanti, MI, United States
fbozorgnia@gmail.com, htang2@emich.edu

***Abstract*— It has become excessively important to design rapid and efficient methods of analyzing the quality of drinking water. Previous studies have shown that multivariate analyses are an effective tool to study water quality. However, no study has specifically looked at how seasonal variations can affect the biochemical qualities of water including PH, temperature, turbidity, potassium, etc. In this Study, the effect of seasonal variations in three treatment stations in the Detroit River (Michigan, US) during four seasons were analyzed for 12 parameters using multivariate statistical techniques. Cluster analysis grouped the 12 months into four clusters of similar water quality. The results of the analysis indicate that one significant contributor to water quality in one month amongst any one parameter did not have the same load factor across others months. Also, a parameter that can be significant in contribution to water quality variations in the Detroit River in one month may not be as significant in a subsequent month. Based on this study, there was a not significant variation in the chemical qualities of the water during the selected months and parameters for this study.**

Keywords—Factor analysis; Cluster analysis; Principal component analysis; Water pollution

I. Introduction

Water quality is essential in balancing our ecosystem and it is vital to human life. Water as a natural resource is vital in any forms such as marine, deep-water wetlands, fresh waters such as rivers and lakes and last not the least, groundwater environments. All the above sources are important in terms of environmental, social and economic values, therefore maintaining acceptable and high quality water is necessity. Physical, chemical and biological properties commonly define water quality and based on its quality and properties, it can be used for variety of purposes such as human consumption, irrigation and industrial process. Human activities such as agricultural, industrial and urban activities in addition to natural process such as soil erosion, weathering fluctuations and precipitation rate affects the water quality (Zhao and Cui 2008. This human activities and natural process leads to variation in water quality, which also can add excess nutrients and toxic chemicals to aquatic ecosystem. High concentration of toxic chemicals and biologically available nutrients can lead to diverse problems and would be harmful to the living organisms and human lives. This is why monitoring water quality is extremely important. In the United States, an emphasis is placed on monitoring for compliance with the Clean Water Act and Safe Drinking Water Act, which are administered by the U.S. Environmental Protection Agency (EPA). The Detroit River is one of the main water sources in state of Michigan. The water drawn from the Detroit River and Lake Huron are used for a variety of purposes. The Great Lakes Water Authority (GLWA) is regional water and sewer authority, which operates five treatment plants, that treats water drawn from the Detroit River and Lake Huron. There is a common assumption the water drawn from the Detroit River and Lake Huron, which provides water to City of the Detroit and its suburbs, which has been treated, is safe for environment and human consumption. By examining the reality behind this assumption, it will be better able to reassures individuals and population.

Currently, there are many studies researching the water quality before any treatment in different seasons vs. after water treatment in water plants. There is a lack of research on quality of water after treatment process. Having high quality water treatment program to efficiently deliver water is the basis of a strong quality of life that leads to reduce the risk of many water contaminations. Drinking water, irrigation, fishery and energy production are important multi-usage components of lakes and rivers. Having access to safe quality water is essential to human beings and the ecosystem. There are many factors, which influence water quality such as natural process and anthropogenic activities. Natural process, including precipitation rate, weathering processes and sediment transport, largely determines the water quality. On the other hand, anthropogenic activities including urban development and expansion, and industrial and agricultural practices are considered being major sources of the addition of chemicals and nutrients to aquatic ecosystem [1, 2]. Maintaining a well- balanced ecosystem is essential for beneficial interaction of living things and the environment, which water quality obviously plays a critical role in this relationship [1]. There are many ways to evaluate water quality. The Environmental Protection Agency (EPA) defines water quality as, “Water quality standards are provisions of state, territorial, authorized tribal or federal law approved by EPA that describe the desired condition of a waterbody or the level of protection or mandate how the desired condition will be expressed or established for such waters in the future. These standards form a legal basis for controlling pollution entering the waters of the United States from a variety of sources (e.g., industrial facilities, wastewater treatment plants, and storm sewers)".

The quality of water is identified in terms of its physical, chemical and biological parameters. The majority of researchers monitored all or a combination of the following quality parameters based on their sampling sites and data requirements. The parameters include, but not limited, to electrical conductivity (EC), total solids (TS), total suspended solids (TSS), total dissolved solids (TDS), pH, dissolved oxygen (DO), temperature (TEMP), turbidity (T), chemical oxygen demand (COD), biochemical oxygen demand (BOD), ammonia-nitrogen (NH+4-N), nitrate (NO3), nitrate nitrogen (NO-3-N), total phosphorus (TP), sulphate (SO-3.4), iron (Fe), copper (Cu), lead (Pb), manganese (Mn), arsenic (As) total coliform (TC) and fecal coliform (FC) [1, 2, 3, 4, 5].

It is important to evaluate and be able to interpret the complex data sets such as those created by long-term water quality monitoring programs. Using multivariate statistical techniques in the assessment of water quality has been practical for the interpretation of complex data. These methods include cluster analysis (CA), factor analysis (FA), and principal component analysis (PCA), which help to identify important components and factors accounting for most of the variance of a system [1]. In addition to the above methods, discriminant analysis (DA) was also used by many researches [6]. According to previous researchers, applications of multivariate statistical methods have been widely used in order to extract meaningful information from data set. Iscen et al. assessed the surface water quality of Lake Uluabat located in west part of Turkey during the period of 2004-2005 and similarly Ayla Bilgin looked at the impact of 24 water parameters measured semiannually from four sites between 2011 and 2013 in Coruh Basin, Turkey [7]. They both applied CA to delineate and interpret the data driven from monitoring sites. Sherestha and Kazama examined large and complex water quality data set of the Fuji river basin in Japan during 8 years (1995-2002) of monitoring 12 parameters at 13 different sites (14976 observations). Multivariate statistical techniques such as CA and DA were used to identify the significant parameters and optimize the monitoring network.

The aim of this study is to analyze the ten water parameters in three treatments plants of the Detroit River for seasons of summer and fall of 2015 and winter and spring of 2016. The obtained data set is subjected to multivariate statistical methods such as CA, PCA, and FA to evaluate information about similarities and dissimilarities between treatment stations in order to understand water quality variables for seasonal variations in river water quality as well as the effect of pollution sources on the water quality parameters of the Detroit River Basin. The purpose of this study is to assess water quality in three treatments operational plants of the Detroit River in four seasons of summer and fall of 2015 and winter and spring of 2016. This study examined data obtained from three treatment plans; Lake Huron (LH), Springwells (SPW) and Belle Isle (BII) intake. The data was analyzed in order to evaluate similarities and dissimilarities of water quality in the above treatment stations via using multivariate statistical methods such as cluster analysis (CA), Principal component analysis (PCA), and factor analysis (FA).

II. MATERIALS AND METHODS

A. *Study Area*

The Detroit River is located in Wayne County and southeast region of Michigan. It originates from Lake St. Clair and flows for 24 nautical miles (44 Km; 28 mi) to Lake Erie as a channel in the great lakes system and forms part of the border between Canada and the United States. It lies between longitude of 83°05'34" NAD27 and Latitude of 42°17'53". The Detroit River has served an important role and is a major source of recreational and economic activity in addition to being a source of water supply to the city. The Great Lakes Water Authority (GLWA), which launched in January 2016, began the management of regional water and wastewater, which plays an important role in monitoring the quality of drinking water. The Great Lakes Water Authority operates five water treatment plants that treat water drawn from Lake Huron and the Detroit River to meet Safe Drinking Water Act requirements.

Fig. 1. The Detroit River map

B. *Data Collection And Analytical Methods*

The data was obtained from the Great Lakes Water Authority located on E. Jefferson, Detroit, Michigan. The reports contained 31 parameters which 12 of them being selected for the purpose of this study. The selected parameters include: turbidity, total solids, sodium, potassium, chloride, phosphorus, total hardness, total alkalinity, chemical oxygen demand, dissolved oxygen, nitrate nitrogen, pH and temperature. The units and method of analysis of these parameters have been represented in Table 1.

TABLE I. WATER QUALITY PARAMETERS ABBREVIATIONS AND REFERENCE METHODS USED FOR WATER SAMPLES OF THE DETROIT RIVER

Parameter	Abbr.	Ref.	Ed.
pH	pH	SM 4500-h+ b	20
Sodium	Na	SM 3111 b	18
Total hardness as $caco_3$	TH	SM 2340 c	20
Chemical oxygen demand	COD	SM 4500-p e	20
Alkalinity as caco	T-alk	SM 2320 b	20
Chloride	Cl	SM 4500-cl $^-$ b	18
Nitrate	NO_3	SM 4500-no3 e	18
Potassium	K	SM 3111 b	18
Total dissolved solids 180 °C	TDS	SM 2540 c	18
Turbidity	TURB	SM 2130 b	20
Dissolved oxygen	DO	SM 4500-no3 e	18
Temperature	T	Thermometer	20

In the current investigation, water quality data were collected from three water treatments plants operated by GLWA in four different seasons (summer and fall of 2015; winter and spring of 2016). All samples were collected during natural activities. Few parameters were measured on-site such as pH, temperature and dissolved oxygen using a thermometer and pH and DO meter, respectively. All GLWA water samples were collected and analyzed according to Standard Methods Examination of Water and/or Waste Water or U.S. EPA protocols in accordance with Michigan Department of Environmental Quality guidelines.

C. *Statistical Methods*

1) *Cluster Analysis*

Statistical analysis was performed on the data sets using MATLAB, Minitab, and Microsoft Office Excel. To describe the water quality based on months, CA that is a group of multivariate techniques was used. CA primary purpose is to assemble objects based on the characteristics they possess (Adams, 1998; Einax and Soldt, 1999). CA classifies and grouping objects together that share similar values. One of the most common clustering is Hierarchical (agglomerative) clustering which was used in this study. Hierarchical clustering combines objects in hierarchical sequence portrayed a tree called a dendrogram. The dendrogram is a tree diagram that illustrates a visual summary of the clustering process of the groups and their proximity, which reduces the length of the original data (SAS Institute, 2015). In order to better understand the leading factors and to visualize the summary of intra-relationship among variations parameters, cluster analysis was used in this investigation.

2) *Principal component analysis/factor analysis (PCA/FA)*

In order to identify important parameters related to measurements of water quality, principal components analysis and factors analysis was conducted in order to reduce factor sizes and efficiently analyze the data. Based on the outcomes of factor analysis, we were able to demonstrate the most significant parameters in each station. One of prerequisites of this analysis is the normality of data. There are different approaches to investigate this property in data. In this study, the Kolmogorov-Smirnov (K-S) statistics test is used to confirm the normal distribution of data. The obtained results with a 95% or higher confidence interval, indicating normal distribution in most of the parameters. There were a few parameters lacking this property, which were modified by logarithm transformations.

III. RESULTS AND DISCUSSION

A. *Spatial similarity with CA*

Hierarchical clustering analysis was applied to extract similar months within a year in water quality parameters in the three treatment plant sites. The obtained dendrograms shown in Lake Huron (Figure 2), Springwells (Figure 3) Bell Isle Intake (Figure 4) demonstrate the months when quality of water in each treatment plant is similar. Considering (D_Link/D_max) _100 < 50 in LH clusters consist; {{1,12}, {2,3}, {4,5}, {6,8,9,11,7}, {10}}, in SW consist {{1,12,2}, {3,4,5,6}, {7,9,11,8,10}}, and in BII consist {{1,2}, {3,5,4,6}, {4,5}, {7,8,9,12}, {10}}. These results indicate some differences and similarity of months among three treatment plants which may refer to different sources of pollution for water at these sites. During the month of October, all three plants (LH, SW, BII) were significantly different in compositions in comparison to the rest of the year.

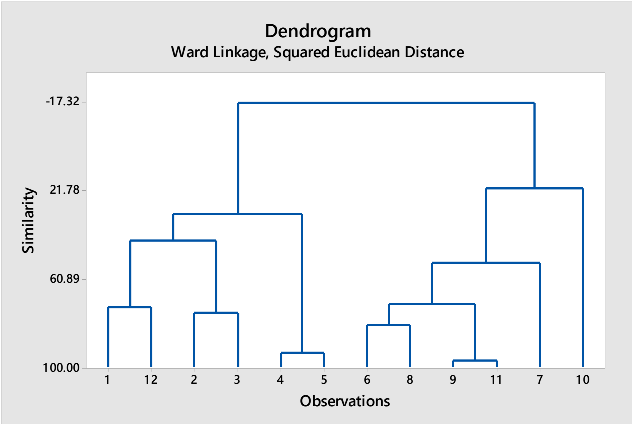

Fig. 2. Dendrogram of cluster analysis for LH station throughout the year (LH: Lake Huron)

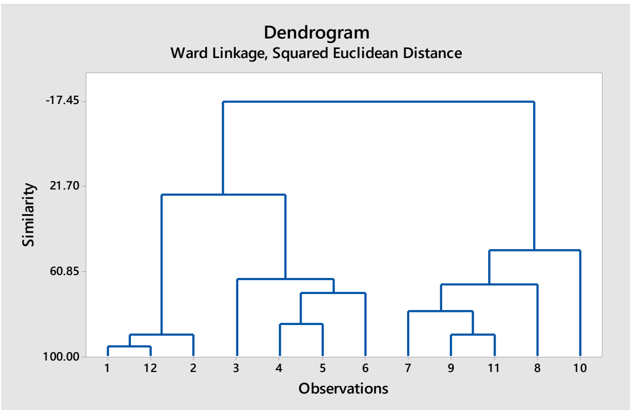

Fig. 3. Dendrogram of cluster analysis for SW station throughout the year (SW: Springwells)

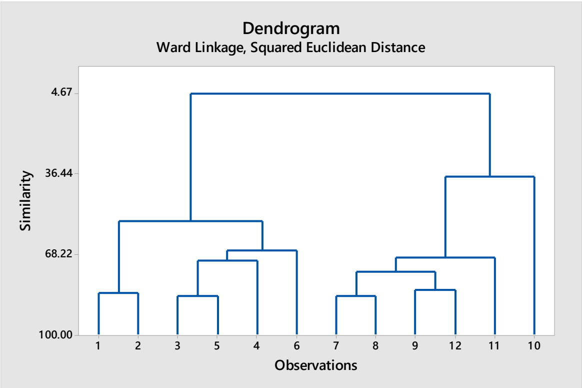

Fig. 4. Dendrogram of cluster analysis for BII station throughout the year (BII: Bell Isle Intake)

B. Most important parameters

PCA is a dimension-reduction technique, which the primary purpose is to drive a small number of linear combinations (principal components) to explain the variance of a large data set of intercorrelated variables with a smaller set of independent variables (Iscen et al. (2007); Kim et al. 2005; SAS Institute, 2015). In order to identify important water quality parameters, PCA was executed on 12 variables for the 12 months of the three-treatment stations, Lake Huron, Springwells, and Belle Isle intake. Figure 5 shows the results of Kolmogorov-Smirnov (K-S) statistical test on data of Lake Huron. This test confirms the normal distribution of data and thus compatibility of them with PCA analysis. Each principal component is calculated by taking a linear combination of the eigenvalue of the correlation matrix with the variables. The factors with the highest eigenvalue are the most significant. Eigenvalues of 1.0 or greater are considered significant. Variable loadings and explained variance are presented for Lake Huron (Table 2), Springwells (Table 3), and Belle Isle intake (Table 4) where strong loading values have been highlighted.

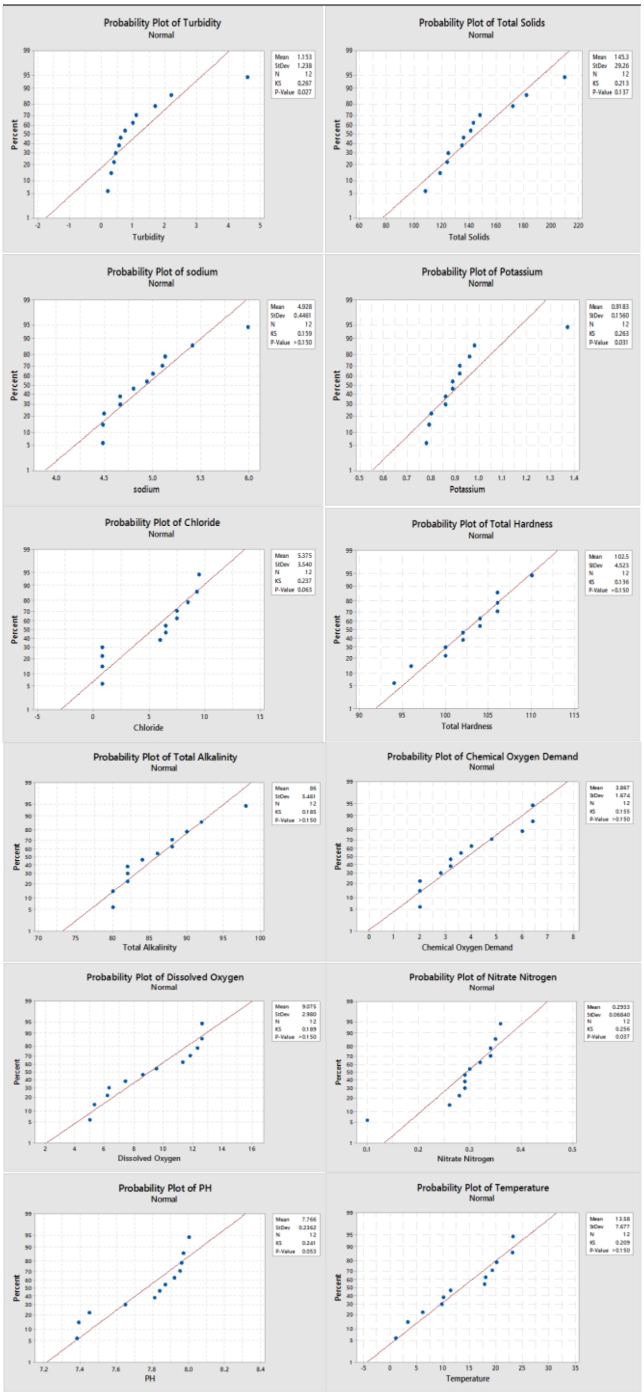

Fig. 5. The results of normality tests for LH over 1 year period (LH: Lake Huron)

Factor analysis in LH station demonstrates six significant parameters as the highest eigenvalue contributing the most covering total variance. Extraction method: Principal component analysis; Rotation method: Varimax with Kaiser normalization with nine iteration convergence. (Table 2).

Factor 1: Chloride, Dissolved oxygen and temperature, Factor 2: Nitrate and Turbidity, Factor 3: PH and Alkalinity, Factor 4: Potassium and Chemical Oxygen Demand, Factor 5: Total Hardness and Total Solids and Factor 6: Sodium has the strongest correlation. The 6 factors of PCA/FA include totally more than 94.329% of the total variance in each season respecting water quality in LH station.

Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 5 Issue 11, November - 2018

Additionally, the factor loading value is convergent in 9 iterations.

TABLE II. FACTOR LOADING FOR LH STATION

LH	Component					
	1	2	3	4	5	6
TURB	0.080	0.970	-0.127	-0.017	0.016	-0.076
TDS	-0.289	-0.248	0.160	-0.026	0.887	0.086
Na	0.044	0.082	-0.038	0.021	0.121	0.985
K	0.250	0.390	0.079	-0.824	0.139	-0.102
Cl	0.912	0.092	0.177	0.113	-0.031	0.214
TH	0.358	0.305	-0.109	0.138	0.806	0.105
T-alk	0.109	-0.148	0.889	0.289	0.199	-0.053
COD	-0.039	0.318	0.338	0.817	0.285	-0.083
DO	0.824	0.119	-0.439	-0.219	0.134	-0.038
NO_3	-0.118	-0.927	0.179	0.055	0.026	-0.215
PH	-0.321	-0.183	0.860	-0.096	-0.109	-0.002
T	-0.800	-0.075	0.199	0.408	0.093	0.161

Moreover, four significant parameters are the highest eigenvalue contributing the most covering total variance in SW station (Table 3) Extraction method: Principal component analysis; Rotation method: Varimax with Kaiser normalization with seven iteration convergence. Factor 1: Turbidity, Sodium, Potassium, Alkalinity and Nitrate, Factor 2: Dissolved Oxygen, PH and Temperature, Factor 3: Total Solids, Total Hardness, Factor 4: Chloride and Chemical Oxygen Demand, have the strongest correlation. The 4 factors of PCA/FA include totally more than 81.22% of the total variance in each season respecting water quality in SW station. In addition, the factor loading value is convergent in seven iterations.

Regarding BII station, five significant parameters had the highest eigenvalue, with the highest total variance in that station. Extraction method: Principal component analysis; Rotation method: Varimax with Kaiser normalization with ten iteration convergence (Table 4). Factor 1: Turbidity, Total Hardness and Potassium, Factor 2: Alkalinity, PH and Nitrate, Factor 3: Total Solids and Chemical Oxygen Demand, Factor 4: Chloride, Dissolved Oxygen and Temperature, Factor 5: Alkalinity and Sodium have the strongest correlation. The five factors of PCA/FA include totally more than %86.658 of the total variance in each season respecting water quality in BLL station. The factor loading value is convergent in ten iterations.

TABLE III. FACTOR LOADING FOR SW STATION

SW	Component			
	1	2	3	4
TURB	0.575	0.557	0.276	0.312
TDS	0.454	-0.015	0.843	0.000
Na	0.882	-0.098	0.374	-0.137
K	0.841	0.127	0.168	0.331
Cl	0.408	-0.049	0.350	0.584
TH	-0.032	0.186	0.797	0.118
T-alk	0.707	0.340	-0.196	0.109
COD	-0.043	0.116	0.018	-0.946
DO	0.350	0.754	0.380	-0.071
NO_3	0.689	0.132	0.584	0.150
PH	0.014	-0.818	-0.060	0.036
T	-0.074	-0.960	0.054	0.118

In order to evaluate variations in surface water quality of Detroit River totally, principal components analysis/factors analysis was conducted on 12 factors under all three different sampling stations (LH, SW, BII) in four seasons. In more detail, the underlying factor analysis in the three stations (LH, SW, BII) demonstrated the four significant parameters as the highest eigenvalue contributing the most covering total variance. Factor 1: PH, Dissolved Oxygen and Temperature, Factor 2: Turbidity, Chloride and Potassium, Factor3: Nitrate, Sodium, Total Solids and Total Hardness, Factor 4: Alkalinity and Chemical Oxygen Demand have the strongest correlation. According to acquired information, it is possible to analyze and explain four parameters instead of 12. The 4 factors of PCA/FA include totally more than %66.497 of the total variance in the above 3 stations.

TABLE IV. FACTOR LOADING FOR BII STATION

BII	Component				
	1	2	3	4	5
TURB	0.890	0.093	-0.314	-0.221	-0.067
TDS	-0.339	0.210	0.855	-0.116	0.163
Na	0.120	0.174	-0.108	-0.052	0.927
K	0.916	0.109	-0.142	-0.118	0.087
Cl	0.136	0.186	-0.456	0.743	0.104
TH	0.791	0.401	0.322	0.103	0.190
T-alk	-0.030	-0.608	-0.061	-0.156	0.656
COD	0.066	-0.051	0.816	0.062	-0.316
DO	0.219	0.461	-0.032	-0.749	0.337
NO_3	0.146	0.845	0.039	-0.028	0.103
PH	-0.227	-0.729	-0.002	0.384	0.004
T	-0.384	-0.260	0.366	0.749	-0.129

The five most prominent water quality parameters that can contribute to evaluate variations in the three treatment plants are signalized in table 5. Total Dissolved Solids, Chloride, Dissolved Oxygen, PH and Temperature are the parameters with the strongest factor loadings contributing to water quality.

TABLE V. THE FIVE PARAMETERS WITH STRONG FACTOR LOADINGS IN EACH STATION.

Stations	Parameters with Strong Factor Loadings
LH	Turb,Na,Cl,T-alk,NO_3
SW	TDS,Na,K,COD,T
BII	Turb,TDS,Na,K,NO_3
Total	TDS,Cl,DO,PH,T

IV. Conclusions

In the present study, the quality of water drawn from the Detroit River in three treatment stations was evaluated using different multivariate statistical techniques including cluster analysis, principal component analysis and factor analysis. Cluster analysis grouped 12 months of collected water data into three clusters of similar water quality. Using multivariate statistical techniques is practical to simplify complex dataset to simplify the results. Based on the information obtained from the data analysis, a parameter that can be significant in contribution to water quality variations in the Detroit River in one month is not as significant in a subsequent month.

Changes in weather properties in late October and the sudden drop in water temperature could lead to the conclusion that as water temperatures decrease, the density increases which influences several physical and chemical properties of the water. The natural, inorganic and organic parameters were the most significant factors contributing to water quality variations in the selected treatment stations. The data that resulted from the chosen parameters demonstrated that the chemical qualities of the water did not illustrate a significant variation throughout the selected months. Furthermore, when selecting water quality parameters for implementing environmental treatment plans, the seasonal variation may not be considered for the assessment of water quality. Finally, our data analysis and multivariate analysis provides an optimal sampling strategy to assist and gather complex chemical water properties to predict variations and ultimately guide water quality monitoring plans in the Detroit River. Although the research has reached its goals, there were some limitations. Due to the limited time, this research was conducted only on one year's worth of data. Chemical parameters have only been studied in this research. For further research projects, there is an ample amount of historical data available. The biological parameters and their effects on the water quality can be researched in future studies.

References

[1] Filik Iscen, C., Emiroglu, Ö., Ilhan, S., Arslan, N., Yilmaz, V., & Ahiska, S. (2008). Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake, Turkey. Environmental monitoring and assessment, 144(1), 269-276.

[2] Muangthong, S., & Shrestha, S. (2015). Assessment of surface water quality using multivariate statistical techniques: case study of the Nampong River and Songkhram River, Thailand. *Environmental monitoring and assessment*, *187*(9), 548.

[3] Li, D., Huang, D., Guo, C., & Guo, X. (2015). Multivariate statistical analysis of temporal–spatial variations in water quality of a constructed wetland purification system in a typical park in Beijing, China. *Environmental monitoring and assessment*, *187*(1), 4219.

[4] Pejman, A. H., Bidhendi, G. N., Karbassi, A. R., Mehrdadi, N., & Bidhendi, M. E. (2009). Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques. International Journal of Environmental Science & Technology, 6(3), 467-476.

[5] Zhou, F., Liu, Y., & Guo, H. (2007). Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern NewTerritories, Hong Kong. *Environmental Monitoring and Assessment*, *132*(1), 1-13.

[6] Easton, Z. M., Gérard-Marchant, P., Walter, M. T., Petrovic, A. M., & Steenhuis, T. S. (2007). Identifying dissolved phosphorus source areas and predicting transport from an urban watershed using distributed hydrologic modeling. *Water Resources Research*, *43*(11).

[7] Bilgin, A. (2015). An assessment of water quality in the Coruh Basin (Turkey) using multivariate statistical techniques. *Environmental monitoring and assessment*, *187*(11), 721.

[8] Adams, M. J. (1998). 12 The principles of multivariate data analysis. *Analytical methods of food authentication*, 308.

[9] Detroit River." *American Rivers*. American Rivers, n.d. Web. 14 Apr. 2017. <https://www.americanrivers.org/river/detroit-river/>.

[10] Kim, J. H., Kim, R. H., Lee, J., Cheong, T. J., Yum, B. W., & Chang, H. W. (2005) Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea. *Hydrological Processes*, *19*(6), 1261-1276.

[11] In *Water Quality Standards Handbook*. Retrieved from https://www.epa.gov/wqs-tech/water- quality- standards handbook

[12] Institute, SAS. *JMP 12 Multivariate Methods*. SAS Institute, © 2015. Books24x7. Web. Apr.21,2017, http://common.books24x7.com.ezproxy.emich.edu/toc.aspx?bookid=93029