

Application of Surface Water Quality Classification Models Using Principal Components Analysis and Cluster Analysis

Mohamed Hamed*

Civil Engineering Department, Canadian International College (CIC), Cairo, Egypt

Abstract

Water quality monitoring has one of the highest priorities in surface water protection policy. Many techniques and methods focus in analyzing the concealing parameters that determine the variance of observed water quality of various source points. A considerable proportion of them mainly depend on statistical methods, multivariate statistical techniques in particular.

In the present study, the use of multivariate techniques is required to reduce the large variables number of Nile River water quality upstream Cairo Drinking Water Plants (CDWPs) and determination of relationships among them for easy and robust evaluation. By means of multivariate statistics of principal components analysis (PCA), Fuzzy C-Means (FCM) and K-means algorithm for clustering analysis, this study attempted to determine the major dominant factors responsible for the variations of Nile River water quality upstream Cairo Drinking Water Plants (CDWPs). Furthermore, cluster analysis classified 21 sampling stations into three clusters based on similarities of water quality features.

The result of PCA shows that 6 principal components contain the key variables and account for 75.82% of total variance of the study area surface water quality and the dominant water quality parameters were: Conductivity, Iron, Biological Oxygen Demand (BOD), Total Coliform (TC), Ammonia (NH_3), and pH.

However, the results from both of FCM clustering and K-means algorithm, based on the dominant parameters concentrations, determined 3 cluster groups and produced cluster centers (prototypes). Based on clustering classification, a noted water quality deteriorating as the cluster number increased from one to three, thus the cluster grouping can be used to identify the physical, chemical and biological processes creating the variations in the water quality parameters.

Keywords: Surface water; Principal component analysis; Cluster analysis

Introduction

The Nile constitutes the essential source of life in Egypt; it provides people with their fresh water needs. It is an essential factor of production and vital for agriculture, transport, tourism and henceforth the socio-economic development of the country. However, the Nile has become, to a great extent, adversely affected by human activities. On the other hand, industrial waste discharge, leakage of sewage by urban agglomeration and agricultural runoff contributes to the Nile contamination [1].

Surface water quality deterioration at the intakes of Cairo water treatment plants along River Nile due to increasing level of some pollutants concentration above the guidelines paid the attention of public concern and may cause health hazards. Thus, the need for better management of Cairo treatment plants water sources quality is becoming essential.

Multivariate statistical techniques can be used to characterize and evaluate surface water quality; they are useful in verifying temporal and spatial variations caused by natural and anthropogenic factors linked to seasonality. Multivariate analysis of variance determines if there are any significant differences between several groups of multivariate data.

Principal component analysis includes correlated variables with the purpose of reducing the numbers of variables and explaining the same amount of variance with fewer variables (principal components).

Fuzzy C-means (FCM) can be achieved through more careful and informed initialization based on data content. By carefully selecting the cluster centers in a way which disperses the initial cluster centers through the data space, the resulting FCM approach samples starting cluster centers during the initialization phase. The cluster centers are

well spread in the input space, resulting in both faster convergence times and higher quality solutions.

K-means can be used for cauterizing monitoring stations with similar water quality characteristics. K-means cluster analysis is a divisive clustering method with k number of groups set a priori to analysis [2]. Once the number of clusters is set as an input and cluster centroids are initialized, observations are added iteratively to the most similar cluster, whose centroid is then recalculated until all of the observations are grouped [3].

Study area

Cairo, located on the Nile River south of the Mediterranean Sea, just upstream of the point where the river widens into the Delta. Cairo has an area of 353 km² with an average reach length along the river about 50 km (from 900 to 950 km referenced to Aswan High Dam). Figure 1 illustrates the study area layout which covers Cairo Governorate along the Nile, bounded by El Saff town (Giza Governorate) at 877 km from the South and El Kanater town (Qalubia Governorate) at 953 km from the North. The study scope will focus on the upstream of drinking water plants located in Cairo Governorate along Nile River (Tibeen, Kafr Elw, North Helwan, Maadi, Fostat, El Roda and Rod Farg).

***Corresponding author:** Mohamed Hamed, Assistant Professor, Civil Engineering Department, Canadian International College, Cairo, Egypt, Tel: +201118905720; E-mail: moha_hamed@cic-cairo.com; mohamedahmedreda@yahoo.com

Received March 24, 2019; **Accepted** May 03, 2019; **Published** May 10, 2019

Citation: Hamed M (2019) Application of Surface Water Quality Classification Models Using Principal Components Analysis and Cluster Analysis. Irrigat Drainage Sys Eng 8: 231. doi: [10.4172/2168-9768.1000231](https://doi.org/10.4172/2168-9768.1000231)

Copyright: © 2019 Hamed M. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

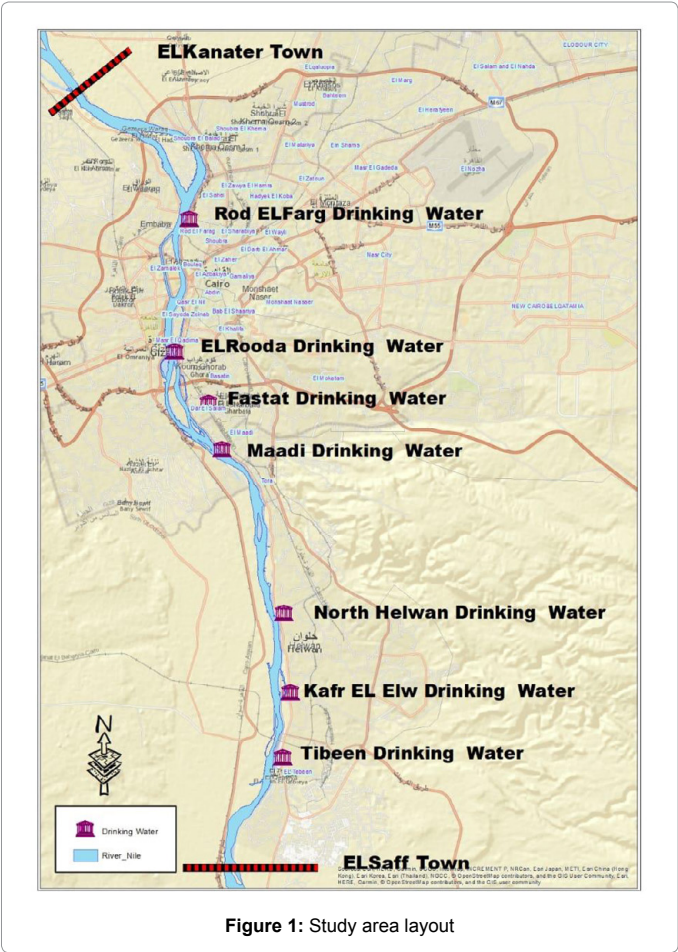


Figure 1: Study area layout

Cairo drinking water plants (CDWPs)

Cairo water company (CWC), a subsidiary of the holding company of water and wastewater; produces potable water with an amount reaches to 6 million m³/day used by inhabitants of Greater Cairo. This is done through 13 Cairo drinking water plants (Tibeen, Kafr Elw, North Helwan, Maadi, Fostat, El Roda, Rod El Farg, Amerea, Mostrod, El Marg, El Obour, El Asher and Shubra el Khiema) distributed in Greater Cairo. Table 1 shows the annual average raw water, treated water and sludge and washing water for Greater Cairo drinking water plants [4]. From Table 1 and according to the study scope which focus on seven CDWPs located on Cairo Governorate along Nile River only (Tibeen, Kafr Elw, North Helwan, Maadi, Fostat, El Roda and Rod Farg).

Materials and Methods

Data requirements

Surface water samples were collected from various sampling locations of rivers, canal, drains and industrial pollution sources of the study area. The analyses of water samples were carried on twenty water quality parameters according to the standard methods for the examination of water and wastewater for twelve consequence months during two years (2017 and 2018) to show the effect of the spatial and temporal variation.

These water quality parameters included: pH, turbidity, electric conductivity (EC), total hardness, total dissolved solids (TDS), total alkalinity, sulfates, chlorides, ammonia (NH₃), nitrates (NO₃), nitrites

Treated water production (m ³ /day)	Raw water (m ³ /day)	Surface water source	Plant intake geographic location	Drinking water plant
155649	178608	River Nile	Cairo	Tibeen
70728	78238			Kafr Elw
283539	321003			North Helwan
161772	209179			Maadi
1046974	1114381			Fostat
164625	323216			El Roda
720908	819695			Rod El Farg
389853	404226	Ismailia Canal	Cairo	Amerea
1155899	1281328			Mostrod
526232	650000			El Marg
790000	860000		New Cities	El Obour
500000	600000			El Asher
358091	379146	Sharkawia Canal	Qalubia	Shubra el Khiema

Table 1: CDWPs surface water source, annual average raw water and treated water.

(NO₂), phosphate, iron, manganese, calcium, magnesium, aluminum, biological oxygen demand (BOD), chemical oxygen demand (COD), dissolved oxygen(DO) and total coliform (TC).

Methods

- The methods consisted of four main components as follows:
1. Perform the principal component analysis
 2. Develop the dominant water quality parameters
 3. Develop the optimum number of clusters by using Fuzzy C-Means (FCM)
 4. Apply K-means algorithm technique to produce the generalized characteristics of clusters using the dominant parameters normalized data.

Principal component analysis: Principal component analysis (PCA) is mainly applied for the removal of data noise by the reduction of their dimensionality [5]. PCA searches new abstract orthogonal principal components (eigenvectors) which explain most of the data variation in a new coordinate system. Each principal component (PC) is a linear combination of the original variables and describes a different source of variation.

PC_i=w₁x₁+w₂x₂+...+ w_nx_n (1)

where x_i and w_i are the original variable and the component weight, respectively.

Before the computation, the testing data were standardized in order to avoid misclassifications arising from different orders of magnitude of tested variables. Therefore, the original data were meaning (average) centered and scaled by the standard deviations.

Procedural steps of the PCA [6] are:

- Number of components equal to number of variables is generated
- The number of components to retain is determined
- Components are rotated (rotations is a linear transformation of the solution to make interpretation easier)
- Rotated solution is interpreted

Dominant water quality parameters: In this study, to determine the main dominant water quality parameter, varimax rotation used as an effective orthogonal rotation method that minimized the number of variables that have high loading on each factor. The varimax coefficient having correlation greater than 0.75 are considered as strong and indicate high proportion of its variance explained by the factor, between 0.50 and 0.75 is considered as moderate loading while 0.30-0.50 as weak significant factor loading, indicating much of that attribute's variance remains unexplained and it is less important [7].

Fuzzy C-means clustering (FCM) analysis: FCM applied for clustering the raw data into several categories using the selected operators without respect to any predetermined criteria in relation to each category. Most of the rules designed for FCM are based on the proper search for centroids or representative objects around which all observations will be clustered on a minimum basis [8,9].

FCM seeks to minimize the following objective function, C, made up of cluster memberships and distances [10].

In fuzzy clustering, the following coefficients must be determined:

i. Dunn's partition coefficient may be normalized so that it varies from 0 (completely fuzzy) to 1 (hard cluster). The normalized version is:

$$Fc(U) = \frac{F(U) - (1/K)}{1 - (1/K)}$$

ii. Another partition coefficient, given in Kaufman (1990) is:

$$D(U) = \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^N (h_{ik} - m_{ik})^2$$

iii. The normalized version of this equation is:

$$Dc(U) = \frac{D(U)}{1 - (1/K)}$$

$F_c(U)$ and $D_c(U)$ together give a good indication of an optimum number of clusters. We should choose K so that maximize the value of $F_c(U)$ and minimizes $D_c(U)$ [10].

K-means algorithm: K-means is a simple and efficient algorithm. It divides n observations into given K clusters and each observation belongs to cluster with nearest mean.

It uses the sum of square error criteria. The cluster pattern is assigned when sum of square error is minimum. The sum of square error equation (SSE) for K-means is given by Kaufman and Rousseeuw [10]:

$$SSE = \sum_{c_j} \sum_{x \in c_j} \|x - m_i\|^2$$

where m_i is the mean of the i^{th} cluster and $x \in C_i$ is a pattern assigned to that cluster. The K-means clustering has advantage over other methods as it can be used to assign new cases to the existing clusters.

Results and Discussion

Descriptive statistics

Table 2 shows the details of descriptive statistics for the water quality variables measured in two years.

Principal component analysis

The calculated principal components loadings, eigenvalues, total variance and cumulative variance are shown in Table 3, while the scree plot of the eigenvalues of observed components is depicted in Figure 2.

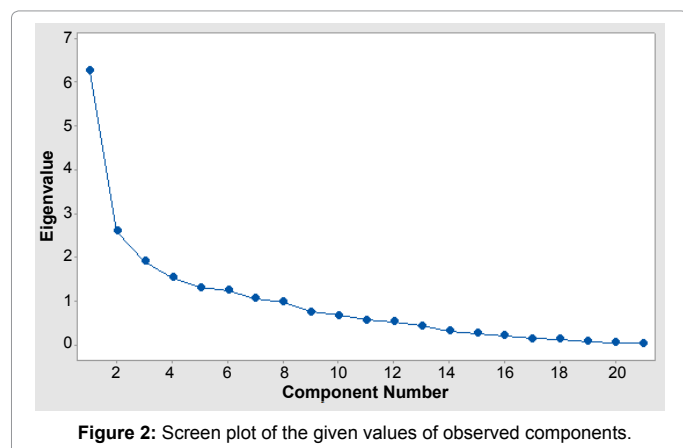
Variables	Mean	Standard Deviation	Minimum	Maximum
pH	8.291	0.041	8.2	8.45
Turbidity	7.751	2.294	2.81	13.96
EC	420.622	59.143	338	584
Total hardness	128.623	12.188	116	164
TDS	278.514	40.173	176.5	385.44
Total alkalinity	151.42	7.131	138	170
Sulphates	24.679	8.608	11.77	48.28
Chlorides	25.226	7.461	11.52	51
Ammonia (NH ₃)	0.15	0.09	0.01	0.47
Nitrates (NO ₃)	0.336	0.311	0.01	1.22
Nitrites (NO ₂)	0.035	0.051	0.01	0.49
Iron	0.208	0.185	0.003	0.69
Manganese	0.016	0.017	0.002	0.064
Calcium	31.47	2.696	22.8	38.4
Magnesium	11.934	1.778	0.03	17.28
Aluminium	0.052	0.037	0.001	0.14
BOD	3.851	0.202	3.37	4.12
COD	17.013	0.177	16.58	17.47
DO	7.473	0.153	2	7.99
TC	20106.33	3327.82	14000	24000

Table 2: Mean and standard deviation of surface water quality parameters.

Variables	PC1	PC2	PC3	PC4	PC5	PC6
PH	0.241	0.038	-0.085	-0.385	-0.054	0.701
Turbidity	-0.124	0.455	-0.065	0.135	0.6	0.046
EC	0.902	0.134	-0.091	0.165	0.105	0.05
Total hardness	0.887	0.149	0.046	-0.124	0.121	0.059
TDS	0.889	0.196	-0.082	0.154	0.131	0.059
Total alkalinity	0.65	-0.529	-0.132	0.29	0.078	0.169
Sulfates	0.833	-0.115	-0.278	0.071	0.036	-0.016
Chlorides	0.818	-0.454	-0.106	0.147	0.038	-0.001
Ammonia (NH ₃)	-0.137	-0.001	0.043	-0.13	0.853	-0.043
Nitrates (NO ₃)	-0.674	-0.122	-0.083	0.17	-0.43	-0.205
Nitrites (NO ₂)	-0.336	-0.183	-0.134	0.362	0.028	0.256
Iron	-0.038	0.879	-0.075	-0.134	0.158	0.116
Manganese	0.088	0.819	0.007	0.09	0.038	-0.018
Calcium	0.726	0.536	0.003	-0.151	0.088	0.075
Magnesium	0.811	-0.24	0.087	-0.079	0.075	0.005
Aluminum	-0.332	0.133	0.042	0.353	0.286	0.589
BOD	0.099	0.045	0.938	0.065	0.009	0.065
COD	0.096	-0.099	0.926	0.086	0.017	-0.144
DO	0.353	-0.002	0.025	-0.127	0.062	0.666
TC	0.066	0.003	-0.167	0.76	0.1	0.021
% Variability	31.48	13.29	9.73	6.55	7.36	7.4
Cumulative percentage	31.48	44.78	54.51	61.06	68.42	75.82

Table 3: Principal component analysis after varimax rotation.

The results of principal components analysis illustrated in Table 3 and Figure 2 of Cattell scree plot show that of the 20 components, only 6 had extracted eigenvalues over 1 [11]. This is based on Chatfield and Collin [12] assumption which stated that components with an eigenvalue of less than 1 should be eliminated. The extracted 6 components were subsequently rotated according to varimax rotation in order to make interpretation easier and fundamental significance of extracted components to the water quality status of the selected study period. The result of rotation revealed further, the percentages of the total variances of the 6 extracted components when added account for 75.82% (that is their cumulative variance) of the total variance of



the observed variables. This indicates that the variance of the observed variables had been accounted for by these 6 extracted components.

As it is obvious, the first principal component (PC1), accounts for 31.48% from total variation, can be called as an indication of salt component because it is mainly saturated with conductivity, hardness (including calcium). PC1 accounts show a strong loading on EC (0.902), TDS (0.889), total hardness (0.887), sulfates (0.883), chlorides (0.881), magnesium (0.811), while moderate loading on calcium (0.726), nitrates (0.674) and total alkalinity (0.65). Electric conductivity (EC) measurements indicate the presence of dissolved salts and electrolytic contaminants, but it gives no information about specific ion compositions [13]. There was a strong positive correlation between TDS and EC values which revealed positively strong correlation to each other ($r=+0.99$), so the study results were in accordance with Toufeek and Korium [14].

The second principal component (PC2), accounts for 13.29%, is associated with strong loading on iron (0.879) and manganese (0.819), while moderate loading on calcium (0.536). The concentration of iron and manganese recorded higher attribute due to the intense of human activities and industrial effluents from iron and steel companies.

The third principal component (PC3) described 9.73% of the total variance had a strong positive loading on BOD (0.938) and COD (0.926). These factors loading explained the effects of organic pollution and reflect strong influence of anthropogenic activities in the area, probably from domestic waste and industrial waste. High BOD and COD levels in the study area are related to the existence of high bacterial load and organic matters as well as relative high temperatures which enhance the enumeration of bacteria. However all results of study area water samples were higher than the permissible limit guidelines (COD should not exceed 10 mg/L) according to Egyptian National water quality standards, Law 48/1982 regarding the protection of the River Nile and waterways from pollution.

Out of the total variance, 6.55% is explained by the fourth principal component (PC4), is mainly carried by TC with a positive strong correlation (0.76) that is indicators for water contamination. The high counts of total coliform might be due to pollution by industrial activities discharging their wastes to the Nile water in Cairo [15]. All results of Nile water samples were higher than the permissible limit guidelines (TC should not exceed 5000 cfu/100 mL) according to Tebbutt [16]. The study results also agree with Rabeh [17].

Additionally, 7.36% of the total variance of water quality is exhibited by NH_3 with a strong positive loading under the fifth principal

component (PC5). NH_3 is closely related to the organic matter contents of the sediment and this high amount of nutrients might also result from the application of manure in agricultural activities [18].

The six principal component (PC6), with 7.4% of the total variance, consists mainly of pH (0.701) and DO (0.701) with a moderate loading. This factor resulted due to the anaerobic conditions in the river from the strong loading of dissolved organic matter which leads in the formation of organic acids. pH value has an effect on the biological, chemical reactions, as well as it controls the metal ion solubility and thus it affects the natural aquatic life. The study results were in accordant with Toufeek and Korium [14].

Based on the component loadings, the variables are grouped accordingly with their designated components as follows:

- Component 1: EC, TDS, total hardness, sulfates, chlorides, magnesium, calcium, nitrates and total alkalinity.
- Component 2: Iron and manganese.
- Component 3: BOD and COD.
- Component 4: TC.
- Component 5: NH_3 .
- Component 6: pH and DO.

Dominant water quality parameters

The dominant parameters identified by the PCA are: EC, iron, BOD, TC, NH_3 and pH (Table 3). The previous discussion indicated that most of measured water quality parameters such as EC, TDS, total hardness, different major ions and total alkalinity, loaded with positive values, and they have strong effects on PC1. EC has the maximum strong loading value in PC1. Thus, EC is considered as a dominant parameter.

The iron is considered as the next dominant water quality parameter as it is loaded strong in PC2 with the highest positive values. Also, the concentrations iron in the Nile water causes the exceedance of the drinking water guidelines, particularly at the anthropogenic impact points, where, iron is regulated by secondary drinking water contaminant that may cause offensive taste, odor, color corrosion or staining problems.

The BOD is considered as the third dominant water quality parameter as it is loaded strong in PC3 with highest value (0.938). These two parameters (BOD and COD) may have a strong relationship with each other, particularly the discharge of industrial and agricultural effluents containing a large amount of organic matter [19].

The TC is considered as the fourth dominant water quality parameter as it is loaded strong in PC4 with highest value (0.760).

The ammonia is considered as the fifth dominant water quality parameter as it is loaded strong in PC5 with highest value (0.853). Ammonia may result from fertilizers that are present in soil and it is relatively easily oxidized to nitrite and finally to nitrate [20] and it possesses a serious threat to public health.

The pH is considered as the third dominant water quality parameter as it is loaded strong in PC6 with highest value (0.701). pH value has an effect on the biological, chemical reactions, as well as it controls the metal ion solubility and thus it affects the natural aquatic life. Moreover pH could control the pathogenic microorganism growth [21].

Cluster analysis

Optimum number of clusters: FCM applied to determine the optimum number of clusters (k) that maximize the value of Fc (U) and minimizes Dc (U) [10]. Table 4 illustrates the values of Fc (U) and Dc (U) with the corresponding number of clusters. FCM results illustrated in the Table 4, it noticed that the optimum number of clusters for the study area is three clusters which satisfies the above conditions.

Clusters characteristics: According to the optimum number of clusters which determined by using FCM in the previous step, K-means algorithm applied to produce the generalized clusters characteristics using the dominant parameters. After finding medians of clusters, the clusters are developed by assigning each object of dataset to the nearest medians of the clusters. The dissimilarities from each of the objects in the dataset from these centers of the clusters are determined using Euclidean distance. Cluster Centers are selected on the basis of the minimum distance. Silhouette is used for interpretation and validation of clusters [10]. Table 5 and Figure 3 illustrate generalized characteristics mean values and the six dominant parameters mean values of the three clusters respectively. It is obvious from Table 5 and Figure 3 for the K-means algorithm results, as the cluster number changed from 1 to 3, the value of the six dominant parameters and the water quality deterioration increased [22-27].

DWPs and monitoring stations clusters allocation: According to the K-means algorithm generalized clusters characteristics results, the allocation for CDWPs and monitoring stations clusters were developed. The output of the cluster characteristics analysis is dispensed

Number of clusters (K)	Fc (U)	Dc (U)
2	0.3333	0.9998
3	0.25	1.0122
4	0.2	1.0131
5	0.1667	1.0134
6	0.1429	1.0136
7	0.125	1.0137
8	0.3333	0.9998

Table 4: Results of optimum number determination

Dominant Parameter	Cluster 1	Cluster 2	Cluster 3
pH	8.303	8.34	8.401
Turbidity	8.052	8.5382	10.537
EC	414.611	502.888	539.111
Total hardness	148.444	153.2	164.666
TDS	334.976	363.307	398.026
Total alkalinity	143.222	150.622	163.777
Sulphates	23.484	25.683	27.358
Chlorides	19.777	20.3	24.777
Ammonia	0.225	0.12	0.228
Nitrates	0.497	0.531	0.684
Nitrites	0.118	0.126	0.217
Iron	0.12	0.155	0.617
Manganese	0.006	0.014	0.039
Calcium	34.022	34.302	34.666
Magnesium	10.854	11.141	11.3
Aluminium	0.069	0.043	0.068
BOD	3.792	3.852	3.944
COD	16.975	17.006	17.95
DO	7.219	7.402	7.563
TC	18266.67	19922.22	21666.67

Table 5: Clusters centers mean values.

in dendrogram (Figure 4). Dendrogram gives the picture of the clusters describing the spatial variation in the water quality and the grouped monitoring stations, CDWPs of each cluster.

Based on the results of cluster analysis, stations and CDWPs grouped under each cluster in Figure 4, it was concluded that: The first cluster, mainly located in the upstream of the study area with less polluted (LP) stations, included the stations from 878 to 868 and three DWPs (Tibeen, Kafr Elw and North Helwan). The changes in water quality in this cluster were mainly due to the agricultural drainage water mixed with partially treated or untreated domestic wastewater, industrial wastewater and wastewater from these three drinking water plants sludge disposal.

The second cluster, comprised only the three DWPs (Maadi, Fostat and El Roda) with moderate pollution (MP), is mainly affected by the cumulative pollution from the previous cluster in additional to the wastewater from the three mentioned drinking water plants sludge disposal. The common feature of these sites was relatively high dominant parameters concentrations compared to the first cluster.

The third cluster located in the downstream of the study area,

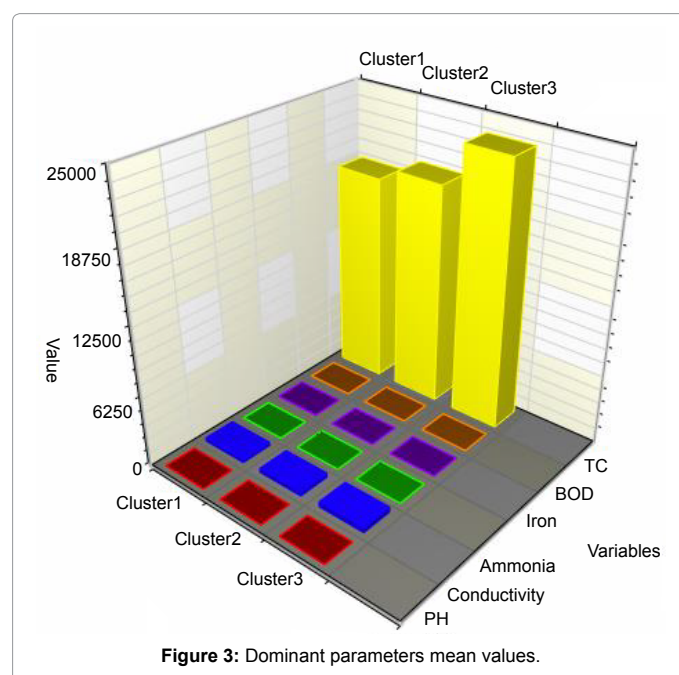


Figure 3: Dominant parameters mean values.

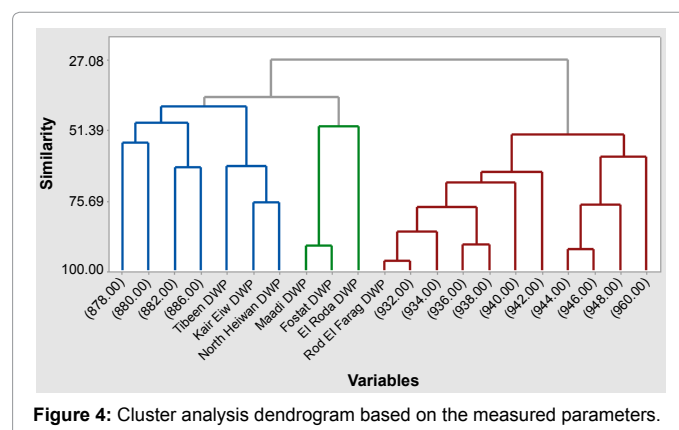


Figure 4: Cluster analysis dendrogram based on the measured parameters.

included Rod El Farag DWP and the different stations from 932 to 950 along the river, had the highest pollution level (HP). These stations are distinguished from other stations concerning the level of pollution and have the most distance from other stations.

Conclusions

This study presents the application of multivariate statistical techniques to evaluate the water quality upstream Cairo drinking water plants along Nile River. The paper outcomes can be beneficial for:

- Understand quality of source waters (i.e., lakes, rivers, and other water bodies) that supply drinking water to big and small communities in any region of the world,
- Apply the study methodology on the monthly, seasonal or yearly water quality sampling data to identify major principal component analysis (PCA) and extract dominant parameters, and
- Allocation of clusters to source waters might be helpful to understand the effect of natural processes, pollution types, and seasonal changes on the water quality of source waters.

References

1. Abd El-Daiem AS (2011) Water Quality Management in Egypt. *J Water Res Dev* 27: 181-202.
2. Akume D, Weber GW (2002) Cluster Algorithms: Theory and Methods. *Journal of Computational Technologies* 7: 15-27.
3. Davis JC (2002) *Statistics and Data Analysis in Geology* (3rd edn), John Wiley and Sons, Inc., New York.
4. Cairo Drinking Water Company (2018) Central laboratory annual technical report.
5. Jolliffe IT (2002) *Principal component analysis* (2nd edn), Springer-Verlag, New York.
6. Suhr D (2005) Principal component analysis vs exploratory factor analysis (paper 203-30). In: *Proceedings of the thirtieth annual SAS® users group international conference 2005* 203, p: 30.
7. Reghunath R, Murthy STR, Raghavan BR (2002) The utility of multivariate statistical techniques in hydrogeochemical studies. *Water Resaerch* 36: 2437-2442.
8. Selim SZ (1984) Soft clustering of multi-dimensional data: A semi-fuzzy approach. *Pattern Recognition* 17: 559-568.
9. Trauwaert E, Kaufman L, Rousseeuw P (1991) Fuzzy clustering algorithms based on the maximum likelihood principle. *Fuzzy Sets and Systems* 42: 213-227.
10. Kaufman L, Rousseeuw PJ (1990) *Finding Groups in Data-An Introduction to Cluster Analysis*. John Wiley & Sons Inc., New York.
11. Cattell RD (1966) The scree test for the number of factors. *Multivariate Behav Res* 1: 245-276.
12. Chatfield C, Collin AJ (1980) *Introduction to Multivariate Analysis*. Chapman and Hall in Association with Methuen, Inc., 733 Third Avenue, New York.
13. Adekunle L, Adetunji M, Gbadebo A (2007) Assesment of ground water quality in a typical rural settlement in south Nigeria. *Int J Environ Res Public Health* 4: 307-318.
14. Toufeek MA, Korium MA (2009) Physico-chemical characteristics of water quality in Lake Nasser water. *Global J Environ Res* 3: 141-148.
15. Saleh AR (2009) Bacteria and viruses in the Nile. *Monographiae Biologicae* 89: 407-429.
16. Tebbutt THY (1998) *Principles of Water Quality Control*, (5th edn), Butterworth-Heinemann, UK.
17. Rabeh SA (2001) Ecological studies on nitrogen cycle bacteria in Lake Manzalah. *Egypt J Aquat Biol Fish* 5: 263-282.
18. Terceiro P, Lobo-Ferreira JP, Leita TE (2008) Analysis of water quality and governance issues Albufeira Do Alqueva. Communication presented at the 9th Water-Water Congress: Challenges of today, demands of tomorrow. Southampton, United Kingdom.
19. Goher ME, Hassan AM, Abdel-Moniem IA, Fahmy AH, El-Sayed SM (2014) Evaluation of surface water quality and heavy metal indices of Ismailia Canal, Nile River, Egypt. *Egypt J Aquat Res* 40: 225-233.
20. Karavoltzos S, Sakellar A, Mihopoulos N, Dassenakis M, Scoullou MJ (2008) Evaluation of the quality of drinking water in regions of Greece. *Desalination* 224: 317-329.
21. Zamxaka M, Pironcheva G, Muyima NYO (2004) Microbiological and physico-chemical assessment of the quality of domestic water sources in selected rural communities of the Eastern Cape Province, South Africa *Water Sa*: 30(3): 333-340.
22. Clare AA, Dianne FJ, Stuart LS (2007) Effect of overlying water pH, dissolved oxygen, salinity and sediment disturbances on metal release and sequestration from metal contaminated marine sediments. *Chemosphere* 69:1428-1437.
23. Manly BFJ, Alberto JAN (2005) *Multivariate Statistical Methods: A Primer* (4th edn), Chapman & Hall/CRC, Boca Raton, p: 253.
24. McKenna JE (2003) An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. *Environmental Modelling and Software* 18: 205-220.
25. Panda UC, Sundaray SK, Rath P, Nayak BB, Bhatta D (2006) Application of Factor and Cluster Analysis for Characterization of River and Estuarine Water Systems-A Case Study: Mahanadi River (India). *J Hydrol* 331: 434-445.
26. Yu S, Shang J, Zhao J, Guo H (2003) Factor analysis and dynamics of water quality of the Songhua River Northeast China. *Water Air and Soil Pollut* 144: 159-169.
27. Zeng X, Rasmussen TC (2005) Multivariate statistical characterization of water quality in lake Lanier, Georgia, USA. *J Environ Qual* 34: 1980-1991.