

# Characterization of Surface Water Quality along Ismailia Canal, Nile River, Egypt

Mohamed Ahmed Reda Hamed, Assistant Professor, Civil Engineering Department,  
Canadian International Colleague (CIC), El Sheikh Zayed, Giza, Egypt.  
Email:moha\_hamed@cic-cairo.com

## ABSTRACT

Ismailia Canal, one of the main branches of the Nile River in Egypt, is considered as one of the most important irrigation and drinking water source for Ismailia, Port Said and Suez governorates. The canal received industrial, municipal and agricultural waste water which caused deterioration in its water quality. To determine the spatial variability of Ismailia canal water quality and identify the sources of pollution that presently affects the canal water quality, the study scope of work divided into three main parts. In the first part, the assessment of water quality data was monitored at thirty different sampling station along the canal, over the period of two years (2017, 2018) using 30 physicochemical and biological water quality variables and using multivariate statistics of principal components analysis (PCA) to interpret, then analyzing the concealed variables that determine the variance of observed water quality of various source points. In the second part the major dominant factors responsible for canal water quality variations is driven. In the third part, K-means algorithm was used for cluster characterization analysis.

The result of PCA shows that 8 principal components contain the key variables and account for 87.34% of total variance of the canal water quality and the dominant water quality parameters were: Lead (Pb), Total Phosphorus(TP), Ammonia(NH<sub>3</sub>), Turbidity, Fecal Coliform (FC), Iron (Fe) and Aluminum (AL).

However, the results from K-Means Algorithm for clustering analysis, based on the dominant parameters concentrations, determined 5 cluster groups and produced cluster centers (prototypes). Based on clustering classification, a noted water quality deteriorating as the cluster number increased from 1 to 5, thus the cluster grouping can be used to identify the physical, chemical and biological processes creating the variations in the canal water quality parameters.

This study provides an insight into the various statistical models, when water quality monitoring data is combined with spatial data for characterizing spatial and temporal trends, indicating their important potential for decreasing the costs associated with monitoring. This can also very useful to international water resources authorities for the control and management of pollution and better protection of surface water quality.

**Keyword:** Nile River, Ismailia Canal, PCA, K-Means Algorithm.

## 1. Introduction

Ismailia Canal transports 5.000.000 m<sup>3</sup> per day of fresh water from River Nile in North Cairo at El-Mazalat square (mouth of canal) to Ismailia, Port Said and El-Suez

governorates, (Ibrahim et al., 2009). Unfortunately, it is note that many factories are constructed on the two banks of the canal, discharging their wastes into the canal water; deteriorate canal. The canal water quality is varied and depends on the location of the source and the state environmental protection in a given area. Therefore, the canal water quality may be determined by physical and chemical analysis (El-Sayed, 2008).

The application of multivariable statistical methods offers a better understanding of water quality for interpreting the complicated data sets. These methods have many applications in different environmental studies. They have been presented as appropriate tools in water quality assessment, identification of pollution sources/factors and understanding temporal/spatial variations in water quality for Effective River water quality management, (Shrestha, et al, 2007).

Principal component analysis (PCA) is a distinctive statistical technique that includes correlated variables with the purpose of reducing the numbers of variables and explaining the same amount of variance with fewer variables (Wulder 2007).

Cluster analysis is an exploratory data analysis method, whose primary purpose is to assemble objects based on the characteristics they possess, which can be independently applied for pattern recognition. K-Means Algorithm for cluster analysis is a divisive clustering method with K number of groups set a priori to analysis. Once the number of clusters is determined, cluster centroids are initialized and observations are added iteratively to the most similar cluster whose centroid is then recalculated until all of the observations are grouped (Davis, 2002).

## 2. Study area

The study area includes the Ismailia canal with 128 km long , 1-3m depth and about 30-70m width extended from the River Nile, north of Cairo, and runs directly to the east of Ismailia governorate. At Ismailia it bifurcates into two branches as: one to the north to supply port-Said governorate (90km long) and the second to the south to Suez governorate (about 80km long). The water canal discharge is 433.56 m<sup>3</sup>/sec. The total area surrounded the canal is about 108.200 fedden and transported 5.000.000 m<sup>3</sup> per day for drinking, industrial and irrigation purposes, (El-Hadad, 2005 and El-Sayed, 2008). Figure (1) illustrates Ismailia canal location.



Figure (1) Ismailia canal location

### 3. Materials and Methods

#### 3.1 Data Requirements

Surface water samples were collected from thirty sampling locations of Ismailia canal. The analyses of water samples were carried on thirty water quality parameters according to the standard methods for the examination of water and wastewater (APHA, 2012) fourteen consecutive months during two years (2017, 2018) to show the effect of the spatial and temporal variation. Figure (2) illustrates the selected sampling stations for Ismailia canal.

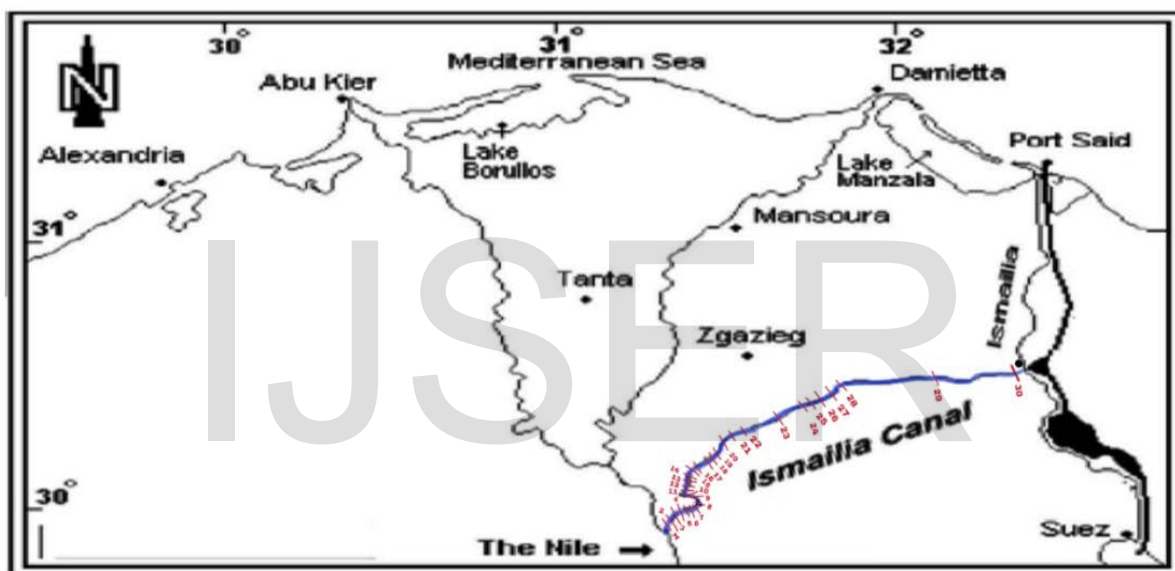


Figure (2) Selected sampling stations of Ismailia canal

Generally, More sample locations are needed near to Cairo governorate than Ismailia governorate due to the existing industrial pollution sources which potentially affects and deteriorates the water quality of the near to Cairo, (Geriesh et al., 2008). In case of sampling sites locations at any existing pollution source, the determination of their locations is mainly based on taking two samples at each pollution location: the first, along the canal and just before pollution source. The second, along the canal at 200 meter after the pollution source to ensure a complete mixing of pollution water source discharge with canal water, (Fischer, et al., 1979).

These water quality parameters included: pH, Turbidity, Electric Conductivity(EC), Total hardness, Total suspended solids (TSS), Total dissolved solids (TDS), Total alkalinity, Carbonate ( $\text{CO}_3$ ), Bicarbonate ( $\text{HCO}_3$ ) Chlorides, Ammonia ( $\text{NH}_3$ ), Nitrates ( $\text{NO}_3$ ), Nitrites( $\text{NO}_2$ ), Phosphorus ( $\text{P}_04$ ), Total Phosphorus (TP), Silica ( $\text{Si}_04$ ), Potassium (K), Sodium(Na) Iron (Fe), Aluminum (Al), Manganese(Mg), Lead(Pb), Zinc (Zn), Cadmium(Cd), Total organic carbon(TOC), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Dissolved Oxygen(DO), Fecal Coliform

(FC), Total Coliform(TC). Table (1) illustrates Ismailia canal surface water sampling

No	Sampling Station	Latitude	Longitude
1	El-Mazalat- Mouth of Ismailia Canal	30° 06' 30"	31° 15' 10"
2	Orabi Bridge	30° 06' 33"	31° 15' 17"
3	About Bridge	30° 06' 35"	31° 15' 33"
4	North Cairo electricity company	30° 06' 37"	31° 16' 42"
5	Al Ameria DWP	30° 06' 41"	31° 15' 46"
6	Al Amiria sewage plant	30° 06' 38"	31° 15' 52"
7	Al Ameria Bridge	30° 06' 55"	31° 15' 55"
8	El Delta company for iron and steel	30° 07' 11"	31° 15' 58"
9	El-Cablata company for electrical wire	30° 07' 24"	31° 16' 16"
10	Nile company for oil and detergents	30° 07' 39"	31° 16' 28"
11	Mostour Bridge	30° 07' 47"	31° 16' 39"
12	Egyptian starch and glucose company	30° 08' 21"	31° 16' 48"
13	Petroleum pipeline company	30° 08' 58"	31° 16' 59"
14	Shiny company for ceramics production	30° 09' 22"	31° 17' 17"
15	Misr Petroleum Company	30° 09' 43"	31° 17' 25"

IJSER

location details.

16	Mostour DWP	30° 09' 55"	31° 17' 36"
17	Ring Road	30° 10' 09"	31° 18' 20"
18	Ezbet Eoes	30° 11' 11"	31° 22' 53"
19	Sariakos	30° 12' 46"	31° 26' 12"
20	Kafr Hamza gas station	30° 13' 52"	31° 29' 33"
21	Abu Zaabaal fertilizer Company	30° 14' 58"	31° 31' 58"
22	Aluminum Sulfate Company	30° 16' 46"	31° 33' 06"
23	Ezbet El Kasemia	30° 16' 49"	31° 23' 07"
24	Belbeis	30° 22' 57"	31° 34' 33"
25	Ezbet Makena	30° 25' 13"	31° 36' 35"
26	Ezbet Moftah	30° 27' 24"	31° 38' 41"
27	El Wadi Drainage	30° 29' 33"	31° 40' 30"
28	El Abbasa	30° 32' 04"	31° 42' 35"
29	El Tel El Kebeer	30° 32' 47"	31° 52' 09"
30	El-Ismailia	30° 34' 05"	32° 14' 06"

Table (1) Detailed surface water sampling location of Ismailia canal.

### 3.2 Methods

The methods consisted of three main parts as follows:-

- 1) Assessment of water quality data and perform the principal component analysis
- 2) Develop the canal dominant water quality parameters
- 3) Apply K-means algorithm technique to produce the generalized characteristics of clusters using the dominant parameters normalized data

#### 3.2.1 Principal component analysis

Principal component analysis (PCA) is mainly applied for the removal of data noise by the reduction of their dimensionality (Jolliffe, 2002). 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.

$$PC_i = w_1X_1 + w_2X_2 + \dots + w_nX_n \quad (1)$$

Where  $x_i$  and  $w_i$  are the original variable and the component weight, respectively. The principal component weights are used as measures of the correlation between the variables and the principal components. The largest or first PC is oriented in the direction of largest variation of the original variables and passes through the center of the data. The second largest PC lies in the direction of the next largest variation, passes through the center of the data and is orthogonal to the first PC. The third largest PC is directed towards the next largest variance, goes through the data center and is orthogonal to the first and second PCs, and so forth.

#### 3.2.2 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 (Reghunath, et al 2002).

### 3.2.3 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 (Kanungo et al. 2002):

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

Where  $m_i$  is the mean of the  $i^{th}$  cluster and  $x \in C_i$  is a pattern assigned to that cluster.

The method of initializing the clusters influences the final cluster solution. For each trial, randomly assigns each point to a cluster. This configuration is optimized using the k-means algorithm. Trying several random starting configurations will greatly increase the probability of finding the global optimum solution for a particular number of clusters. The goodness-of-fit criterion used to compare various cluster configurations is based on the within-cluster sum of squares,  $WSS_K$ , where:

$$WSS_K = \left( \frac{NP}{NP - m} \right) \sum_{k=1}^K \sum_{i=1}^P \sum_{j=1}^{n_k} (1 - \delta_{ijk}) (z_{ij} - c_{ik})^2$$

Where  $c_{ik}$  is the average (center) value of the  $i^{th}$  variable in the  $k^{th}$  cluster. The percent of variation ( $PV_k$ ) is defined as:

$$PV_K = 100 \frac{WSS_K}{WSS_1}$$

## 4. Results and Discussion

### 4.1 Descriptive Statistics

Basic statistics were carried out in order to give initial information about the water quality data. Table (2) shows the details of descriptive statistics for the water quality variables measured in two years.

Table (2) Ismailia canal surface water quality parameters

Variables	Mean	Standard Deviation
PH	8.271	0.274
Turbidity	8.874	1.759
EC	453.297	24.316
Total Hardness	144.047	14.036
TSS	138.743	29.242
TDS	292.853	28.386
Total Alkalinity	154.411	13.504
Carbonate	9.934	1.816
Bicarbonate	175.453	18.230
Chlorides	34.872	6.932
Ammonia(NH3)	0.245	0.083

Nitrates(NO3)	0.354	0.107
Nitrites(NO2)	0.061	0.009
Phosphorus	0.150	0.064
Total Phosphorus	0.317	0.115
Silica	6.471	0.862
Potassium	8.807	2.588
Sodium	28.401	7.557
Iron	0.675	0.259
Aluminum	3.821	2.958
Manganese	0.229	0.094
Lead	0.024	0.009
Zinc	0.039	0.016
Cadmium	0.0009	0.00019
TOC	3.769	0.240
BOD	4.034	0.633
COD	9.172	3.852
DO	7.367	0.417
FC	382.000	76.000
TC	18446.000	1468.000

IJSER

#### 4.2 Principal Component Analysis

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

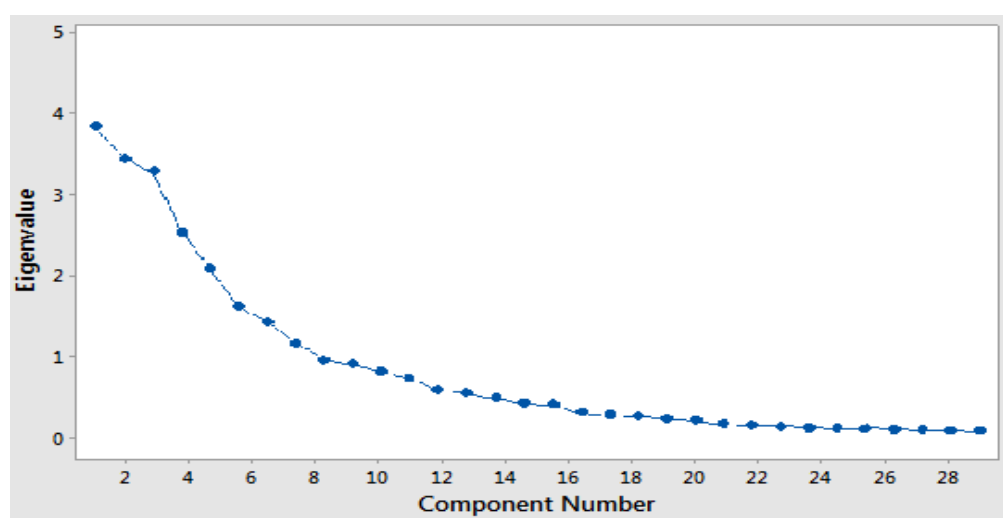


Figure (2) Scree plot of the eigenvalues of observed components

Table (3) Principal component analysis after varimax rotation

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
PH	0.2364	0.0873	0.5887	0.4468	0.1420	0.032	0.3011	0.3499
Turbidity	0.0504	0.0522	0.2584	0.8592	0.0994	0.1256	0.0134	0.0895
EC	0.4369	0.1689	0.0931	0.1530	0.12103	0.7804	0.0395	0.1054
Total Hardness	0.0248	0.3490	0.0458	0.5527	0.3352	0.7519	0.5951	0.0698
TSS	0.1342	0.2357	0.1428	0.2135	0.2180	0.6575	0.2838	0.1529
TDS	0.1297	0.2875	0.2340	0.4110	0.4434	0.7556	0.2201	0.0907
Total Alkalinity	0.7041	0.2637	0.02060	0.4110	0.2572	0.0336	0.1206	0.2989
Carbonate	0.3208	0.0374	0.80030	-0.1138	0.0775	0.2597	0.2774	0.1226
Bicarbonate	0.0976	0.0232	0.0559	0.0503	0.7901	0.2641	-0.0287	0.1939
Chlorides	0.1105	0.4947	-0.1560	0.0140	0.1184	0.2291	-0.1211	0.1086
Ammonia(NH3)	0.1117	0.0738	0.9039	0.0646	0.0758	0.0741	0.2415	0.0376
Nitrates(NO3)	0.7124	0.3402	0.3228	-0.001	0.0437	0.2229	0.1709	0.2296
Nitrites(NO2)	0.0300	0.0825	0.0555	-0.1757	0.4071	0.2443	0.4068	0.2314
Phosphorus	0.0866	0.9520	0.0174	0.0008	0.0321	0.0865	0.0503	-0.0438
TP	0.0247	0.9582	0.13251	-0.0013	0.0246	0.1723	0.0254	-0.0448
Silica	0.6146	0.1707	0.6507	-0.0416	0.1403	0.1059	0.0796	-0.1880
Potassium	0.4894	0.0817	-0.3044	0.6960	0.0707	0.2459	0.0711	-0.0268
Sodium	0.2277	0.0384	-0.5912	0.2011	0.0586	0.6581	0.0755	0.0042
Iron	0.2282	0.0681	0.0684	-0.0985	0.1106	0.0446	0.8621	0.2538
Aluminum	0.4027	0.0493	0.0491	0.2416	0.0483	0.0175	0.5306	0.8418
Manganese	0.0294	0.0597	-0.0674	0.6245	0.0122	0.1416	0.7783	0.3514
Lead	0.8754	0.1471	-0.2118	0.0478	0.2901	0.0962	0.0287	0.2522
Zinc	0.5434	0.1599	-0.1863	0.0131	0.0338	0.6650	0.2165	0.2136
Cadmium	0.3619	0.1904	0.2365	0.6245	0.0744	0.1610	0.4721	0.1959
TOC	0.3827	0.2287	0.1516	0.4350	0.1864	0.0308	0.5857	0.2458
BOD	0.1620	0.5917	0.0606	0.0798	0.6075	0.1999	0.0121	0.3883
COD	0.2443	0.0531	0.6034	0.0931	0.1056	0.3922	0.0436	0.2855
DO	0.2495	0.2371	0.6877	0.4373	0.2064	0.0176	0.0717	0.2596
FC	0.0290	0.1837	0.0234	-0.3159	0.8379	0.0984	0.1340	0.0525
TC	0.8753	0.1846	0.00037	-0.0195	0.7484	0.0510	0.1547	0.2334
% Variability	15.49	12.62	12.99	12.88	8.61	9.59	9.25	5.91
Cumulative percentage	15.49	28.11	41.09	53.97	62.58	72.18	81.43	87.43

The results of principal components analysis illustrated in Table (3) and Figure (2) of show that of the 30components, only 8 had extracted eigenvalues over 1. This is based



on Chatfield and Collin (1980) assumption which stated that components with an eigenvalue of less than 1 should be eliminated. The extracted 8 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 8 extracted components when added account for 87.43% (that is their cumulative variance) of the total variance of the observed variables.

As it is obvious, the first principal component (PC1), with 15.49% of the total variance, consists mainly of Pb (0.8754) with strong loading and also both NO<sub>3</sub> (0.7124) and Total alkalinity (0.7041) with a moderate loading. Lead is one of the most toxic heavy metals having unknown biochemical benefits. Pb shows a strong pollution effects at all canal sampling stations for drinking and aquatic life utilization.

Out of the total variance, 12.62% is explained by the second principal component (PC2), is mainly carried by TP (0.9582) and Phosphorus (0.9520) with strong loading. The high effect of TP at most of Ismailia canal sites may be related to extensive agricultural activities in the region and the release of domestic and industrial sewage, (Abdel-Satar, 2005).

Additionally, 12.99% of the total variance of water quality is exhibited by NH<sub>3</sub> with a strong positive loading (0.9039) under the third principal component (PC3). The third principal component had a moderate loading with both of DO and COD (0.6877), (0.6034). NH<sub>3</sub> 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 (Terceiro et al, 2008).

The fourth principal component (PC4) described 12.88% of the total variance had a strong positive loading on turbidity (0.8592). Turbidity in the canal is resulted from the presence of suspended particles such as silt, plankton, clay, organic matter, and other microscopic or decomposers organisms.

Out of the total variance, 8.61% is explained by the fifth principal component (PC5), is mainly carried by fecal coliform with a positive strong correlation (0.8379) and total coliform (0.7484) that is indicators for water contamination. This may reveals that one of the fecal coliform sources could be related to human recreational activities. All TC results of Ismailia canal samples were higher than the permissible limit guidelines (TC should not exceed 5000 cfu/100 ml) according to Tebbutt (1998). The high counts of total coliform might be due to pollution by industrial activities discharging their wastes to the canal.

The principal component (PC6), accounts for 9.59% from total variation, can represent an indication of salt component because it is mainly saturated with Electric conductivity, Total dissolved solids, Total hardness .PC6 accounts show a strong loading on EC (0.7804), TDS (0.7556) and Total Hardness (0.7519).Electric Conductivity(EC) measurements indicate the presence of dissolved salts and electrolytic contaminants, but it gives no information about specific ion compositions, (Adekunle et al., 2007).

The principal component (PC7), accounts for 9.25%, is associated with strong loading on Iron (0.8621) and Manganese (0.7783). The concentration of Iron and Manganese

recorded higher attribute due to the intense of human activities and industrial effluents from for iron and steel companies.

The eights principal components (PC8), with 5.91% of the total variance, had a strong positive loading on Aluminum (0.8418). Ismailia Canal suffers from obviously different contamination grades with the measured metals; AL exhibits serious effect according to drinking and aquatic life criteria, (Abd El-Hady, et al., 2012)

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

- Component 1: Lead (Pb), Nitrates and Total alkalinity.
- Component 2: Total Phosphorus (TP) and Phosphorus.
- Component 3: NH<sub>3</sub>, DO and COD.
- Component 4: Turbidity.
- Component 5: Fecal Coliform (FC), Total Coliform (TC).
- Component 6: EC, TDS and Total Hardness.
- Component 7: Fe and Mg.
- Component 8: AL.

### **4.3 Dominant water quality parameters**

According to the previous discussion, the dominant parameters identified by the PCA are: Lead, Total Phosphorus, Ammonia, Turbidity, Fecal Coliform, Electric Conductivity, Iron and Aluminum. The previous discussion indicated that most of measured water quality parameters such as Lead (Pb), Nitrates and Total alkalinity, loaded with positive values, and they have strong effects on PC1. Pb has the maximum strong loading value in PC1. Thus, Pb is considered as a dominant parameter. Total Phosphorus is considered as the second dominant water quality parameter as it is loaded strong in PC2 with the highest positive values. The Ammonia is considered as the third dominant water quality parameter as it is loaded strong in PC3 with highest value (0.9039). Ammonia may result from fertilizers that are present in soil and it is relatively easily oxidized to nitrite and finally to nitrate (Karavoltzos et al., 2008) and it possesses a serious threat to public health. Turbidity, the fifth dominant parameter, is loaded strong in PC5 with the highest positive values (0.8592). The FC is considered as the fifth dominant water quality parameter as it is loaded strong in PC5 with highest value (0.8379). EC, TDS and Total Hardness loaded with positive values, and they have strong effects on PC6. EC has the maximum strong loading value in PC6. Thus, EC is considered as a dominant parameter. The Iron is considered as the seventh dominant water quality parameter as it is loaded strong in PC7 with the highest positive values (0.8621). The AL is considered as dominant water quality parameter as it is loaded strong in PC8 with highest value (0.8418).

## **4.4 Cluster analysis**

### **4.4.1 Optimum number of clusters**

K-Means Algorithm is applied to determine the optimum number of clusters (k) that has the maximum percent of variation. Figure (4) illustrate the results of clusters number determination.

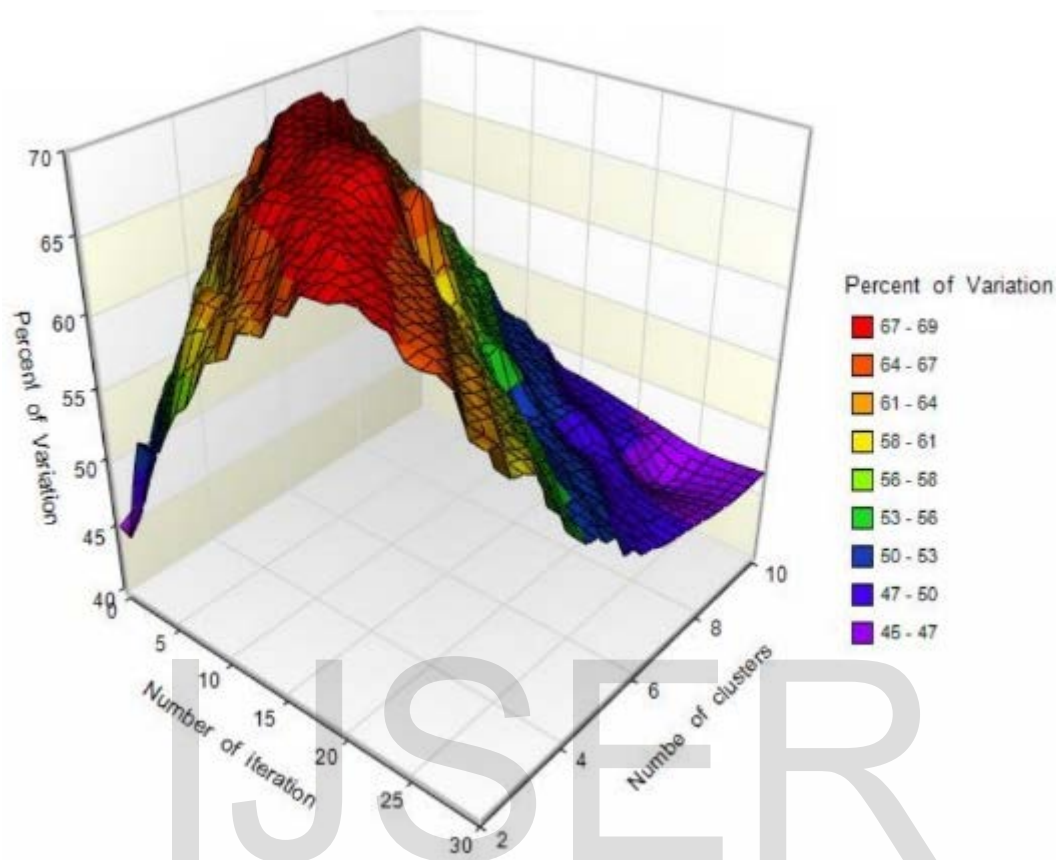


Figure (4) Results of clusters number determination

From Figure (4), it noticed that the optimum number of clusters for the Ismailia canal is five clusters which have the maximum variation percent.

**4.4.2 Clusters characteristics**

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, (Kaufman and Rousseeuw 1990). Table (4) and Figure (5) illustrate generalized characteristics mean values and the six dominant parameters mean values of the three clusters respectively.

Table (4) Clusters centers mean values

Dominant Parameter	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
--------------------	-----------	-----------	-----------	-----------	-----------

PH	7.200	7.777	7.949	8.123	8.423
Turbidity	6.432	7.808	8.656	8.758	11.428
EC	380.589	441.297	459.258	466.025	470.258
Total Hardness	110.897	133.411	148.121	155.125	174.297
TSS	109.312	126.036	129.698	137.411	149.996
TDS	268.696	284.730	296.025	295.356	300.898
Total Alkalinity	144.589	153.378	155.888	157.698	163.985
Carbonate	6.258	7.888	8.998	9.969	10.006
Bicarbonate	169.888	174.894	176.312	177.999	180.014
Chlorides	26.411	28.387	30.962	33.257	39.411
Ammonia(NH3)	0.199	0.211	0.236	0.248	0.302
Nitrates(NO3)	0.207	0.248	0.299	0.328	0.395
Nitrites(NO2)	0.018	0.059	0.064	0.069	0.098
Phosphorus	0.139	0.144	0.158	0.163	0.174
TP	0.265	0.299	0.319	0.348	0.408
Silica	5.841	5.999	6.464	6.885	7.336
Potassium	8.789	8.833	8.914	9.011	9.114
Sodium	22.005	27.126	27.364	27.475	29.532
Iron	0.333	0.614	0.687	0.736	0.784
Aluminum	2.133	2.902	3.687	3.787	3.955
Manganese	0.189	0.204	0.218	0.229	0.238
Lead	0.020	0.023	0.024	0.026	0.027
Zinc	0.031	0.034	0.036	0.037	0.039
Cadmium	0.0009	0.0009	0.0009	0.0009	0.0009
TOC	2.842	3.002	3.524	3.613	3.788
BOD	4.034	4.034	4.034	4.034	4.034
COD	9.014	9.106	9.176	9.184	9.253
DO	7.444	7.320	7.293	7.211	7.196
FC	359.000	385.000	388.000	390.000	434.000
TC	18025.000	18369.000	18459.000	18533.000	18656.000

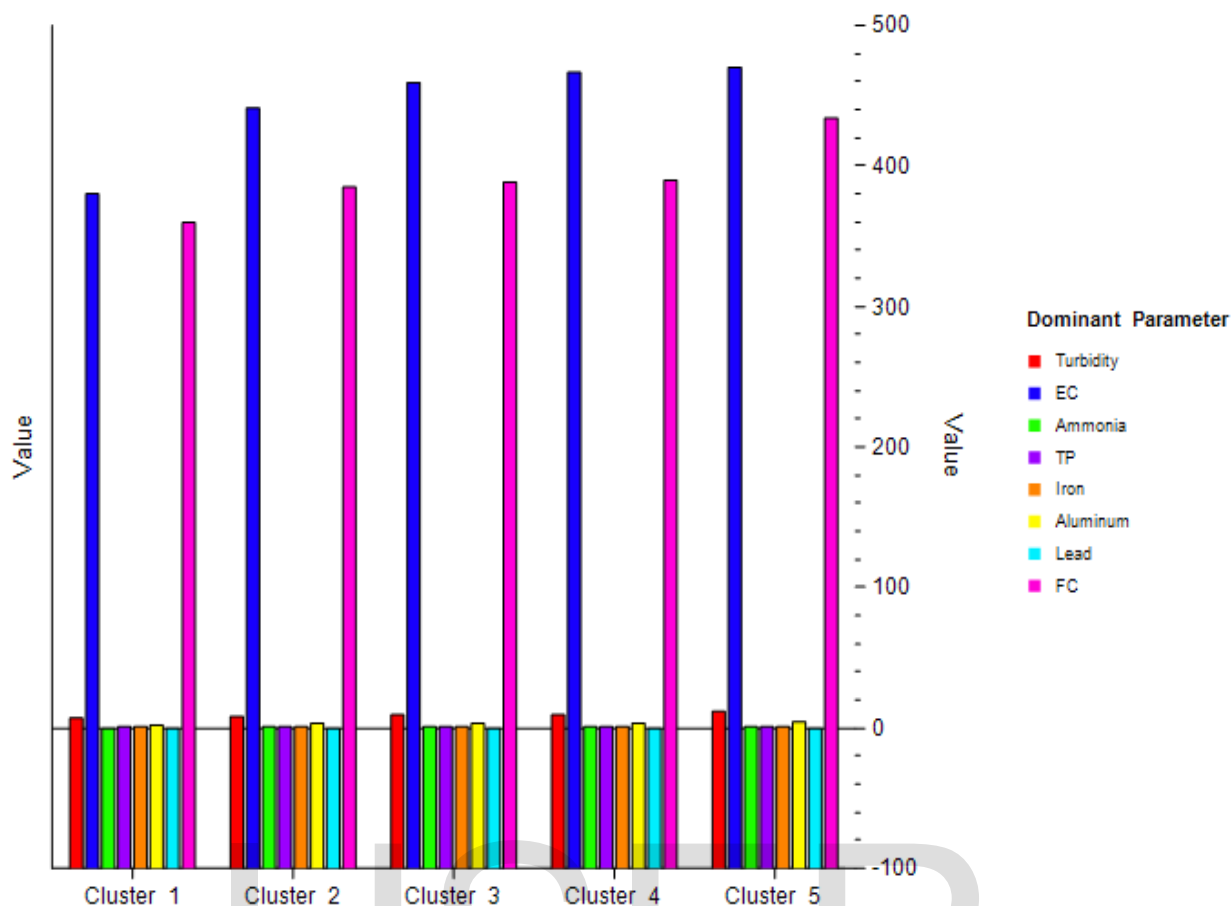


Figure (5) Clusters dominant parameters mean values

It is obvious from table (4) and figure (5) for 8 dominant parameters results, as the cluster number was changed from 1 to 5 the water quality deterioration was increased.

#### 4.4.3 Monitoring stations allocation

According to the K-means algorithm generalized clusters characteristics results, the allocation for Ismailia canal monitoring stations clusters were developed. The output of the cluster characteristics analysis is dispensed in dendrogram, Figure (6).

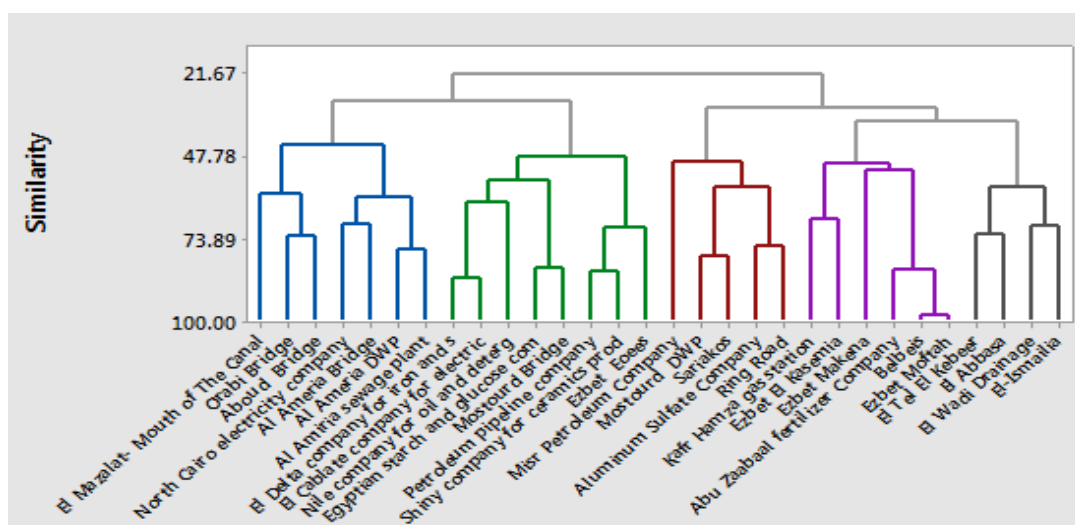


Figure (6) Ismailia canal cluster allocation

From Figure (5) for monitoring station allocation, it was concluded that:-

The first cluster, mainly located in the upstream of Ismailia canal with less polluted stations, included six monitoring stations from El-Mazalat (Mouth of Ismailia Canal) to Al Amiria sewage plant. The changes in water quality in this cluster were mainly due to the agricultural drainage water mixed with partially treated or untreated domestic wastewater and industrial wastewater and wastewater. The second cluster includes eight monitoring stations, extends from El Delta Company for iron and steel monitoring station till Ezbet Eoes monitoring station with a moderate pollution, is mainly affected by the cumulative pollution from the previous cluster in addition to the wastewater from various industrial and municipal activates in this cluster zone. The third cluster located in the middle of the canal, included five monitoring stations from Misr Petroleum Company monitoring stations. The common feature of these sites was relatively high dominant parameters concentrations compared with the previous cluster. The fourth cluster, includes six monitoring stations, begins from Kafr Hamza gas station till Ezbet Moftah monitoring station, suffers from a noted dominant water quality parameters increasing. The fifth cluster located in the downstream of Ismailia canal, includes only four monitoring stations (El Wadi Drainage, El Abbasa, El Tel El Kebeer and Ismailia town), had the highest pollution level. These stations are distinguished from other stations concerning the level of pollution and have the farthest distance from other stations.

## 5. Conclusions

From the previous mentioned discussion, we can conclude that:-

- 1- Studying of water quality parameters along Ismailia Canal revealed that Ismailia Canal is exposed to different sources of pollution, such as agricultural drainage water, domestic wastewater and industrial wastewater from industrial activities.
- 2- Principal components analysis is a useful technique to correlate and reduce a great amount of water quality variable to a few meaningful components; however it can conclude only eight principal components that represent 87.34% of total variance of the canal water quality.
- 3- K-means clustering technique might be helpful to understand the characteristics of each reach of the canal, the effect of natural processes, pollution types, and seasonal changes on the water quality of the canal water. However a five distinctive cluster centroid is driven with the aid of canal dominant water quality parameter to facilitate the spatial classification of canal water quality.
- 4- Apply the study methodology periodically on the Ismailia canal water quality sampling data to identify major principal component analysis (PCA) and extract dominant parameters can reduce the number of samples needed. This will assist in sampling priorities determination and water quality improvement from economic perspective.

## 5. References

- Abd El-Hady, H.H., Hussian, A.M., 2012. Regional and seasonal variation of phytoplankton assemblages and its biochemical analysis in Ismailia Canal, River Nile, Egypt. *J. Appl. Sci. Res.* 8 (7), 3433–3447.
- Abdel-Satar, A. M. 2005. Quality of River Nile sediments from Idfo to Cairo. *Egyptian J. of Aquat. Research*, 31(2): 182 – 199.

- Abdel-Satar, A. M. 2005. Water quality assessment of River Nile from Idfo to Cairo. *Egyptian J. of Aquat. Research.* 31(2): 200 – 223.
- 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 (4): 307-318.
- Cattell, R., D.: The scree test for the number of factors. *Multivariate Behav. Res.* 1, 1966, 245-276.
- Davis, J.C., 2002. *Statistics and Data Analysis in Geology* (Third Edition). John Wiley and Sons, Inc., New York, NY.
- Egyptian Governmental Law No. 48, 1982. The Implementer Regulations for law 48/1982 regarding the protection of the River Nile and water ways from pollution. *Map. Periodical Bull.*, 3-4: 12-35.
- El-Sayed, S. A. 2008. *Microbiological studies on Ismailia Canal, River Nile, Egypt.* M. Sci. Thesis, Fac.of Sci. Al-Azhar Univ.
- Fischer, Ii.B.,E.J. List. R.C.Koii,J, Imbehger,And N. H.Brooks. 1979. *Mixing in inland and coastal waters.*
- Geriesh, M.H., Balke, K., El-Bayes, A., 2008. Problems of drinking water treatment along Ismailia Canal Province, Egypt. *J. Zhejiang Univ. Sci. B.* 9 (3), 232–242.
- Ibrahim, H.S., Ibrahim, M.A., Samhan, F.A., 2009. Distribution and bacterial bioavailability of selected metals in sediments of Ismailia Canal, Egypt. *J. Hazard. Mater.* 168, 1012–1016.
- Jolliffe, I., T. *Principal component analysis* (2nd edn.). Springer-Verlag, New York, 2002.
- Kaufman L., Rousseeuw P. J., *Finding Groups in Data—An Introduction to Cluster Analysis*, John Wiley & Sons Inc., New York 1990.
- Karavoltos, S., Sakellar, A., Mihopoulos, N., Dassenakis, M., Scoullas, M.J. (2008). Evaluation of the quality of drinking water in regions of Greece. *Desalination*, 224: 317-329.
- Shrestha, S., Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan *Environmental Modelling & Software*, 22(4): 464-475.
- Tebbutt, T. (1998). *Principles of Water Quality Control*. 5th Ed., Hallam University.
- Terceiro, P., Lobo-Ferreira, J. P., & Leitão, T. E. (2008). Análise da qualidade da água e questões de governança Albufeirado Alqueva. *Comunicaçãoapresentada no 9º Congresso da Água—Água: Desafios de hoje, exigências de amanhã.* Cascais, Portugal.[http://www.aprh.pt/congressoagua2008/PDF/Lobo-Ferreira Alqueva.pdf](http://www.aprh.pt/congressoagua2008/PDF/Lobo-Ferreira_Alqueva.pdf). Accessed 20 January 2009 (in Portuguese).
- Wulder M (2007) *A practical guide to the use of selected multivariate statistics.* Retrieved February 5, 2007 from [http://www.pfc.cfs.nrcan.gc.ca/profiles/wulder/mvstats/index\\_e.html](http://www.pfc.cfs.nrcan.gc.ca/profiles/wulder/mvstats/index_e.html)