Geoscience Research

Geoscience Research
ISSN: 0976–9846 & E-ISSN: 0976–9854, Vol. 2, Issue 2, 2011, pp-49-60
Available online at http://www.bioinfo.in/contents.php?id=90

ASSESSMENT OF NILE WATER QUALITY DATA USING EXPLORATORY DATA ANALYSIS AND CLUSTERING OF VARIABLES

YOUSRY M.[1], AWADALLAH A.G.[2*] AND SALEM T.[1]
[1]Nile Research Institute, National Water Research Center, Egypt.
[2]Faculty of Engineering, Civil Dept., Fayoum University, Egypt
*Corresponding Author: Email: aawadallah@darcairo.com

Received: August 29, 2011; Accepted: September 23, 2011

Abstract- Considerable attention has been given to the monitoring of Nile River water quality and the establishment of a monitoring program carried out twice a year, in low and high flow seasons. A group of tests to quantify the physical, chemical, and biological quality are conducted for each water sample. The main objective of this paper is to assess Nile River water quality data using exploratory and cluster analysis to gain insight of spatial data grouping and relationships between variables. Having an insight of such variability is essential in detecting point source pollution.

Exploratory data analysis (EDA) is undertaken on the 8 years monitored data (2000-2007) during February and August. Comparing the mean value of each variable with the national standards of Law 48 / year 1982, it is found that water quality variables have mean values within allowable limits except COD. Grouping main Nile and its two branches results in a non-homogeneous dataset especially for salinity and biological variables. The large variability of FC, TC, TSS, OP, COD, BOD, NO_2 and NO_3, depicted through box plots, indicates that the Nile and its branches are exposed to different point sources of pollution. A significant correlation is found between several variables following their expected chemical behavior and exhibits evidence of mutual dependency between water quality variables. Cluster analysis is used to identify a hierarchy of these correlations. The statistical methods undertaken in this research gave a clear picture about the behavior of the variables and the anthropogenic activities in the study area.

Key words – Water quality, Nile River, Spatial and Temporal Variability, Cluster Analysis, Correlation

Introduction

The Nile River flows within the Egyptian territories from the downstream of Aswan High Dam northwards to the Delta Barrage for about 953 km, where it bifurcates into two branches: Damietta (244 km long) and Rosetta (236 km long). Although the Nile River is the main source of water in Egypt, it is subject to contamination from different sources such as agricultural, industrial and municipal sources. Thus, implementing a monitoring network is a must as the water quality conservation is a national target. During the last decade, considerable attention has been given to the monitoring of water quality at the national level and the establishment of a monitoring program. This program plays an important role in water resources management. Therefore, National Water Research Center (NWRC) has developed extensive monitoring networks for all the water bodies. One decade ago, the NWRC through their different institutes implemented the National Water Quality and Availability Management (NAWQAM) project. The Nile Research Institute (NRI) has implemented a monitoring network through the NAWQAM project since 1999. The main objectives of the ongoing monitoring network are to evaluate the quality of water that enters Egypt through the High Aswan Dam (HAD) and detect the seasonal variations and spatial trends of the main Nile water quality and its two branches. The program is carried out twice a year, in both low and high flow seasons; mainly during the months of February / March and August / September. A group of tests to quantify the physical, chemical, and biological quality are conducted for each water sample.

The main objective of this paper is to undertake a critical review and a thorough analysis of water quality data to gain insight of the consistency, seasonal and spatial variability of the data and the hierarchy of interrelationships between water quality variables via variable clustering. Exploratory data analysis has been used in previous studies to evaluate water quality in rivers, and to detect seasonal, spatial and anthropogenic influences [1-4 …]. It was undertaken namely by calculating descriptive statistics, plotting several univariate and multivariate plots, and

assessing data variability both seasonally and spatially. Hierarchy of variable inter-relationships was investigated using first the calculation of correlation matrices and confirmed / classified using hierarchal cluster analysis on variables and not on sampling locations.

Study area, sampling and analytical methods

The national water quality monitoring network on the Nile comprises 4 sites in Nasser Lake, 18 sites along the main river, 4 sites along Damietta branch and 3 sites along Rosetta branch. Monitoring also includes samples from the outfalls of 29 agricultural drains, and at the ends of 11 irrigation canals. However, this study only addresses the main Nile and its two branches for the available data collected in the period 2000 – 2007 during low and high flow periods (February and August). The sampling locations under consideration are listed in Figure 1. Water was sampled from the pre-specified cross section locations at three points (*A, B* and *C*). Points *A* and *C* are located 20 meters apart of the left and right banks, respectively; while point *B* is located in the middle point between *A* and *C*. Sampled water quality variables in the Nile River include Nitrate NO_3, Nitrite NO_2, Ortho Phosphorus OP, Total Phosphorus TP, Chemical Oxygen Demand COD, Biochemical Oxygen Demand BOD, Total Dissolved Solids TDS, Total Suspended Solids TSS, Total alkalinity TA, Sulfate SO_4, Chloride Cl, Sodium Na, Potassium K, Calcium Hardness Ca, Magnesium Hardness Mg, Total Hardness TH, power of Hydrogen pH, Dissolved Oxygen DO, Fecal Coliform FC, Total Coliform TC. Previously mentioned physico-chemical and biological parameters are analyzed according to the Standard Methods for the Examination of Water and Wastewater [5].

Results and Discussion

a) Comparison with Egyptian standards

Descriptive statistics are calculated for the water quality data measured on the main stem and the two branches for all sites in all years. Tables 1 and 2 provide details of quantitative summaries of each water quality variable for all stations, namely the minimum, maximum, mean and standard deviation. Moreover, comparing means and medians, we observe that the means of most water quality variables are substantially greater than medians. The distributions of the data for these variables are highly skewed to the right. Consequently, nature logarithmic (Ln) transformation helps normalize the distribution of these values. Comparing the mean value of each variable with the national standards of Law 48/year 1982, it is found that all water quality variables have mean values within the allowable limits except for COD, which recorded values above the allowable limits recommended by law 48 (10 mg/l). The high values of COD along the main Nile and its two branches indicate that the Nile receives non-biodegradable organic matter. Also total coliform recorded high values especially along Rosetta and Damietta branches. From the descriptive statistics of the raw data as well as the histograms and the confidence interval plots (not shown), it is clear that the normality of the data is not satisfied.

Fig. 1-Sampling Sites along the Nile River

b) Box plots

Box plots are an excellent tool for conveying location and variation information in data sets, particularly for detecting and illustrating location and variation changes between different groups of data [6]. Box plots provide visual summaries of the centre of the data (the median), the variation or spread (the box height), the skewness and the presence or absence of unusual values (outliers and extreme values).

Box plots for all natural logarithm (ln) transformed variables are shown in Figures 2 to 5, first for main Nile sites during February and August and then for Rosetta and Damietta sites also during February and August. The box plots (Figure 2 a) for the Ln-transformed variables of the main Nile River during February show that the variables pH, DO, TDS, TH, Mg, Ca, K, Na, HCO_3 and TA have little variability (Figure 2 b) since their inter-quartile ranges (IQR's) are narrow. NO_3, TP, BOD, TSS, CO_3 and Cl show the asymmetry reflected by the bias of the median point towards one extremity of the box of inter-quartile range. FC, TC and NO_2 show relatively large inter-quartile ranges and provide important information on the variability of data. The variables of NO_3, TP, DO, NO_2, CO_3, HCO_3, K, Mg, Cl and TH visually depart from a normal distribution not only in asymmetry, but also by the number of outliers and the extreme values. Asymmetry and the presence of outliers are causes of the failure of normality tests, as will be presented later in this paper.

Table 1- Statistics of Water Quality Variables (Main Stem Nile River in February and August)

Variable	February				August				Law 48/1982
	Mean	St Dev	Min.	Max.	Mean	St Dev	Min.	Max.	
NO_3	1.18	0.84	0.20	8.90	0.95	0.49	0.20	2.20	45 mg/l
Total - P	0.16	0.11	0.03	0.90	0.17	0.07	0.05	0.44	
COD	12.63	6.22	3.00	38.00	13.25	9.88	2.00	68.00	10 mg/l
BOD	3.17	2.30	0.80	14.00	3.15	2.12	1.00	13.00	6 mg/l
T.D.S	191.81	26.96	147.00	256.00	192.29	20.82	161.00	265.00	500 mg/l
T.S.S	11.62	7.12	2.00	42.00	10.48	4.96	2.00	30.00	
pH	8.28	0.20	7.74	8.70	7.99	0.27	7.46	8.73	7 - 8.5
DO	8.83	0.74	5.96	10.32	7.43	0.89	4.08	9.27	> 5 mg/l
FC	617.5	1070.2	10.00	9533.3	504.0	817.9	33.3	6303.3	
TC	1856.5	2613.5	78.33	20000.0	1769.2	3382.6	100.0	30866.7	
NO_2	0.04	0.07	0.00	0.20	0.09	0.22	0.00	1.61	
Ortho- P	0.06	0.03	0.01	0.18	0.08	0.04	0.00	0.18	
CO_3	4.34	5.53	0.00	20.80	1.68	4.94	0.00	24.00	
HCO_3	129.24	15.99	106.00	175.60	128.76	14.20	100.00	163.90	
T.A	132.43	18.27	14.11	175.60	130.47	13.57	112.00	163.90	20 - 150 mg/l
SO_4	19.64	4.77	12.05	32.94	20.31	5.46	12.00	40.00	200 mg/l
Cl	12.29	5.87	3.88	30.00	12.57	6.20	4.85	35.00	
Na	20.10	5.01	12.00	35.00	19.07	6.16	7.43	38.00	
K	5.84	2.33	2.06	12.96	5.25	1.68	2.21	12.30	
Ca	27.72	4.88	21.20	45.60	27.19	5.34	20.00	44.00	
Mg	11.01	2.81	7.20	21.60	11.53	3.53	4.80	27.36	
T.H	117.08	19.67	94.00	175.00	115.99	21.29	88.00	204.00	

Table 2- Statistics of Water Quality Variables (Rosetta and Damietta Branches in February and August)

Variable	February				August				Law 48/1982
	Mean	St Dev	Min.	Max.	Mean	St Dev	Min.	Max.	
NO_3	6.45	6.67	0.20	26.70	2.92	2.29	0.20	9.19	45 mg/l
Total - P	0.54	0.70	0.05	3.20	0.35	0.37	0.12	1.89	
COD	15.70	9.21	1.00	46.00	18.76	8.99	7.00	52.00	10 mg/l
BOD	6.32	5.48	1.00	22.00	5.62	2.83	1.04	13.00	6 mg/l
T.D.S	339.34	76.25	193.00	483.00	301.12	58.80	205.00	524.00	500 mg/l
T.S.S	12.07	6.67	4.00	30.00	13.49	5.66	5.00	28.00	
pH	7.73	0.23	7.36	8.47	7.67	0.21	7.28	8.10	7 - 8.5
DO	6.81	1.86	2.06	13.78	5.68	1.28	2.10	8.50	> 5 mg/l
FC	718.6	987.0	10.00	5133.3	1746.5	3540.1	22.7	19173.3	
TC	4024.6	6740.6	60.00	42466.7	20275.3	98334.9	180.0	684666.7	
NO_2	0.15	0.07	0.00	0.20	0.38	0.78	0.00	3.90	
Ortho- P	0.27	0.21	0.01	0.83	0.19	0.13	0.01	0.62	
HCO_3	183.58	33.12	126.00	262.50	171.22	22.64	121.00	228.00	
T.A	183.58	33.12	126.00	262.50	171.22	22.64	121.00	228.00	
SO_4	38.72	13.91	16.00	80.00	30.38	11.68	6.00	64.00	20 - 150 mg/l
Cl	37.99	13.57	15.67	70.18	32.27	14.72	4.60	80.00	200 mg/l
Na	40.56	19.03	21.60	99.14	33.42	14.27	4.30	92.00	
K	9.44	3.48	4.60	19.00	7.63	2.44	5.13	15.30	
Ca	41.13	9.33	26.80	60.80	40.40	10.10	24.00	83.40	
Mg	17.36	6.59	7.68	46.08	16.10	7.09	3.36	36.00	
T.H	179.93	44.52	117.00	324.80	166.16	33.85	100.00	244.00	

During August, Figure (3 a and b), the box plots of the main Nile water quality variables show that the pH, DO, TDS, HCO_3, TA, K, Mg and TH have narrow inter-quartile ranges. However, FC, TC and NO_2 have a high variability (inter-quartile ranges are relatively larger). It is worth mentioning that there are many outliers and extreme values for the TC, DO, TDS, COD, TP, NO_3, OP, TA, SO_4, Cl, K, Mg and TH variables. The NO_3, OP, DO, NO_2, CO_3, K, Ca, and SO_4 variables visually depart from a normal distribution not only in skewness, but also by the number of outliers and the extreme values.

During February, the box plots along Rosetta and Damietta Branches (Figure 4) show that most variables (e.g. pH, TDS, K, Na, Mg, Cl, SO_4, HCO_3, TA, Ca and TH) have low variability (IQR's are narrow) compared to BOD, COD, TSS, TP, OP, NO_3, FC and TC (IQR's relatively larger).

During August, Figure 5, the Rosetta and Damietta Branches variables NO_3, TP, NO_2, Cl and K show asymmetry reflected by skewness (distance between the mean and the median) and a number of outliers and extreme values (especially for NO_2). FC, TC, OP variables show large variabilities compared to other variables. pH shows the smallest variability followed by NO_2, TDS, DO, HCO_3 and TA. Generally, the box plots show that the small variabilities of some variables like DO, pH, TDS, cations and anions indicate that the sampling locations have relatively constant sources of these substances. However, the large variability of FC, TC, TSS, OP, COD, BOD, NO_2 and NO_3 indicates that the sources of these substances are highly variable and the Nile and its two branches are exposed to different point sources of pollution. The high variability of some variables is considered as an indicator of importance of continuous monitoring program and intensive sampling frequency.

c) Normality Tests

Normality tests are used to determine whether a random variable is normally distributed or not. Many data analysis methods (t-test, ANOVA, regression, as well as multivariate techniques) depend on the assumption that data are sampled from a normal (Gaussian) distribution. The results of the application of Anderson-Darling test [7], Kolmogrov-Smirnov test with Lilliefors correction [8] and Shapiro Wilk test [9] show that most of the water quality variables in the Nile do not pass the normality test in the main stem Nile section in February, even with the Ln-transformed variables. The list of important variables not passing the tests includes NO_3, pH, NO_2, TP, OP, TSS and DO. In August, the problem persists only for NO_3, NO_2, TDS and OP. As for the two branches, due to their homogeneity, most of the variables pass the tests except for NO_3, TDS and NO_2. Most of cations and anions do not pass normality tests especially in the main Nile section. Variables that pass most of normality tests (at 5% level of significance) in February and August, on main Nile and its branches, are: TDS, TSS, pH, DO, BOD, COD, FC, and TC. TDS is chosen as the variable representing the salinity along with TSS which is occasionally attributed to the organic load.

d) Seasonal variation

Since most of the variables are not passing normality tests, non-parametric tests are used to verify if there is a significant difference between measurements of February and August. Two approaches are used: independent samples Mann-Whitney U test and the paired Wilcoxon signed-rank test. The Mann-Whitney U test is the most popular of the two-independent-samples tests. Mann-Whitney test verifies that two sampled populations are equivalent in location. The observations from both groups are combined and ranked, with the average rank assigned in the case of ties. As described, no pairing is assumed between observations of February and August. On the other hand, the Wilcoxon signed-rank test considers information about both the sign of the differences and the magnitude of the differences between pairs of samples measured at the same location and the same year.

The results of the Mann-Whitney and the paired Wilcoxon signed-rank tests indicate that only NO_3, OP, TP, TDS, pH and DO show significant seasonal variation. Organic and bacterial variable values measured in February are not significantly different of those measured in August. Consequently, the seasonal variation is mainly due to agricultural activities. It is worth mentioning that variations in DO can occur seasonally in relation to temperature and biological activity (photosynthesis and respiration). Biological respiration, including that related to decomposition processes, reduces DO concentration. Moreover, variations in pH can be caused by the photosynthesis and respiration cycles of algae in eutrophic water [10].

e) Spatial variation and trend

To study the variation of the different water quality variables along the Nile, box plots of water quality variables are plotted against the distance of the sampling location measured from the High Aswan Dam. Box plots for the variables in February (Figure 6) show no significant trend for NO_3, significant decreasing trend for OP and significant increasing trends for BOD, COD, TDS and TSS. This increasing trend may be related to waste water discharging from different point sources of pollution, transported to the Nile water by agricultural drains along the Nile. For pH and DO, the trend is concave quadratic, showing an increase in middle locations and again a decrease towards the downstream location near Cairo. A convex quadratic trend is depicted for TC and FC marking the close relationship between pH and DO on one side and TC and FC on the other. Almost the same results can be found for August (Figure 7) except that NO_3 and COD trend lines are significant for 0.05 significance level. In general, the most significant relations are the linear trends of TDS and TSS and the concave quadratic lines of pH and DO. This behavior may suggest subdividing the Nile stem locations into three homogeneous regions formed by locations 8 to 13 (from 0 to 277 km), locations 17 to 24 (from 448 to 683 km) and locations 28 to 34 (from 815 to 938 km).

As for the temporal trend, 7 years of data is not enough to detect such trend. However, analyzing the trends in August and February data separately, it shows that the variability is random and not consistent between February and August, except for BOD and TSS where a positive upward trend is consistently present in August and February. However, as previously stated the record length is not enough to perform such temporal trend analysis.

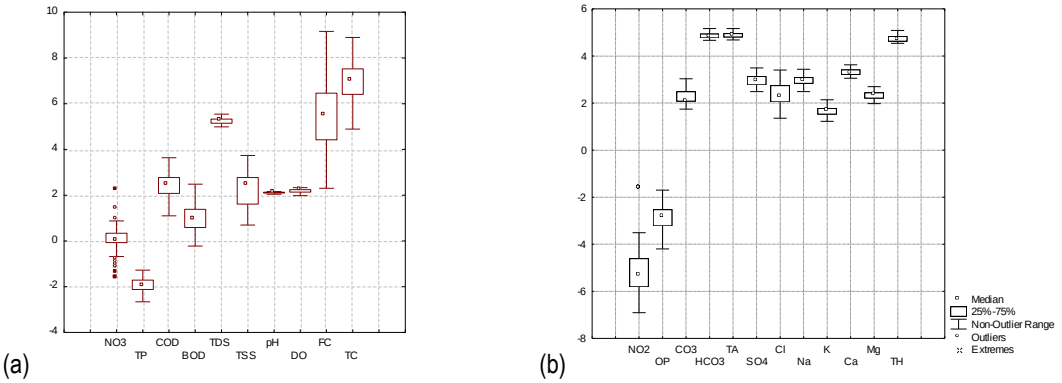

Figure 2: Box Plots for Ln-Transformed Variables (Main Nile Sites during February)

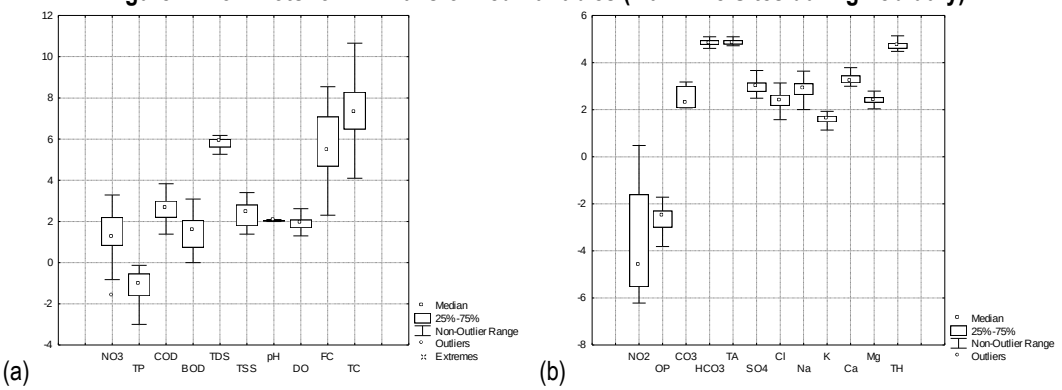

Figure 3: Box Plots for Ln-Transformed Variables (Main Nile Sites during August)

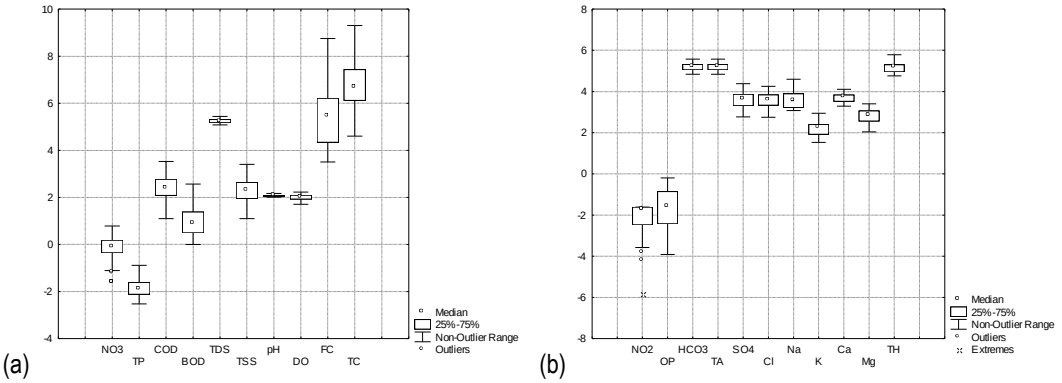

Fig. 4-Box Plots for Ln-Transformed Variables (Rosetta & Damietta Branches during February)

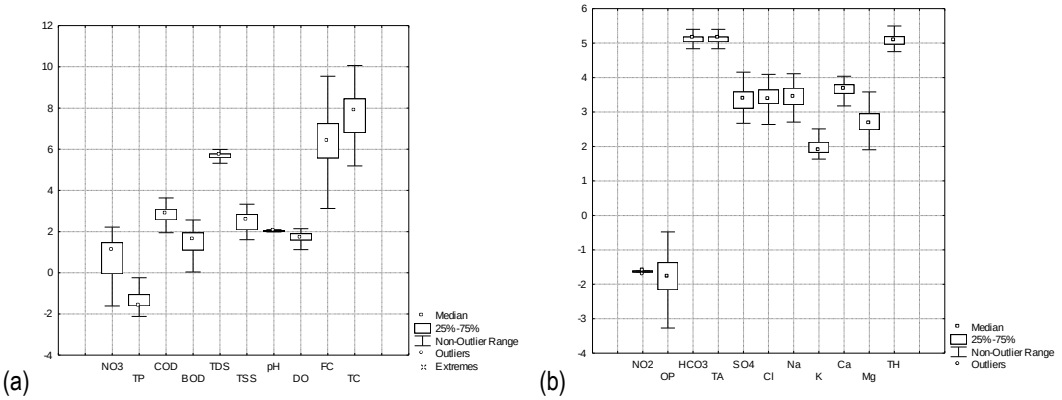

Fig. 5-Box Plots for Ln-Transformed Variables (Rosetta & Damietta Branches during August)

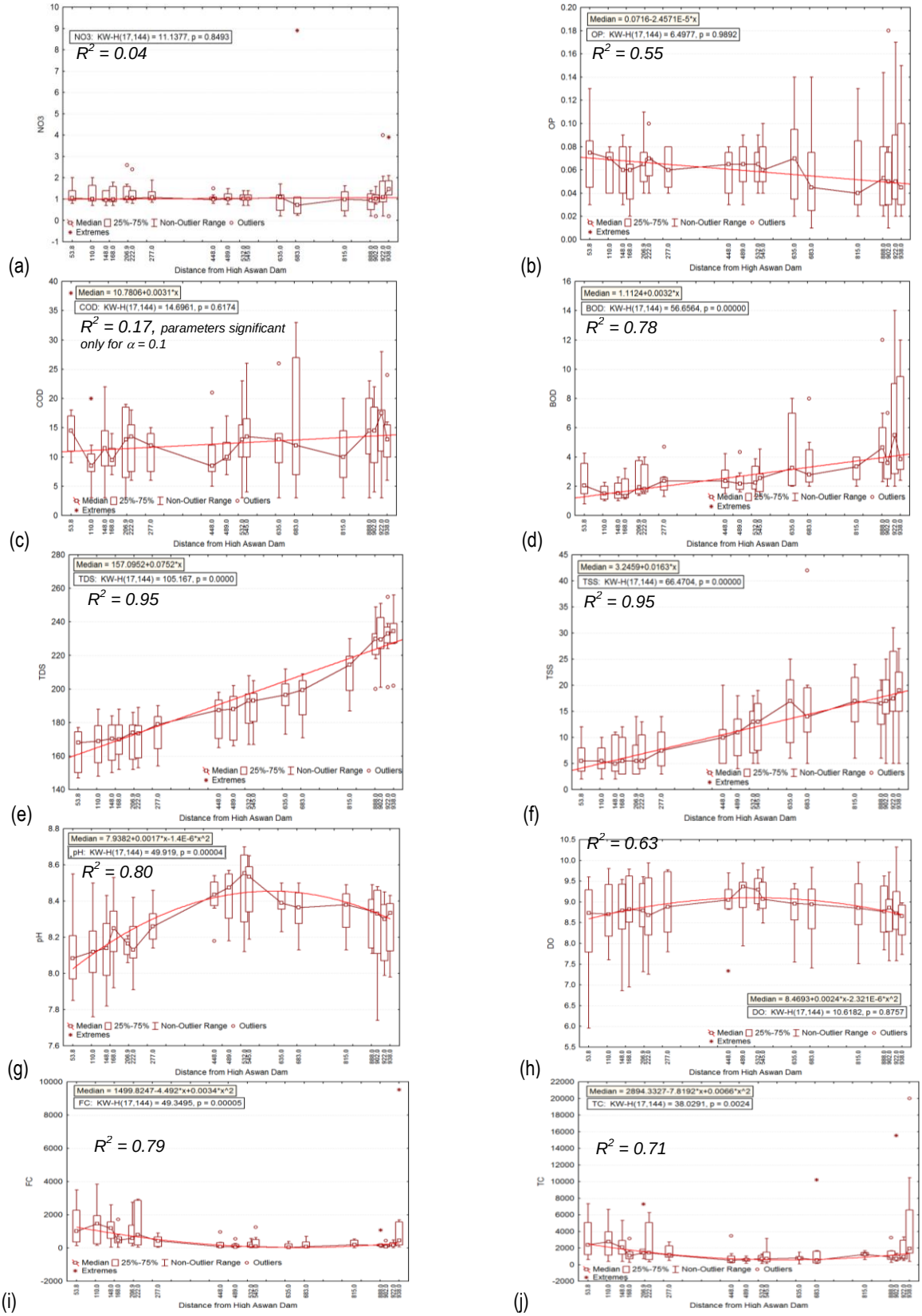

Fig. 6-Box Plots of Water Quality Variability in February with Distance from High Aswan Dam on the X-Axis
(a) NO_3, (b) OP, (c) COD, (d) BOD, (e) TDS, (f) TSS, (g) pH, (h) DO, (i) FC and (j) TC

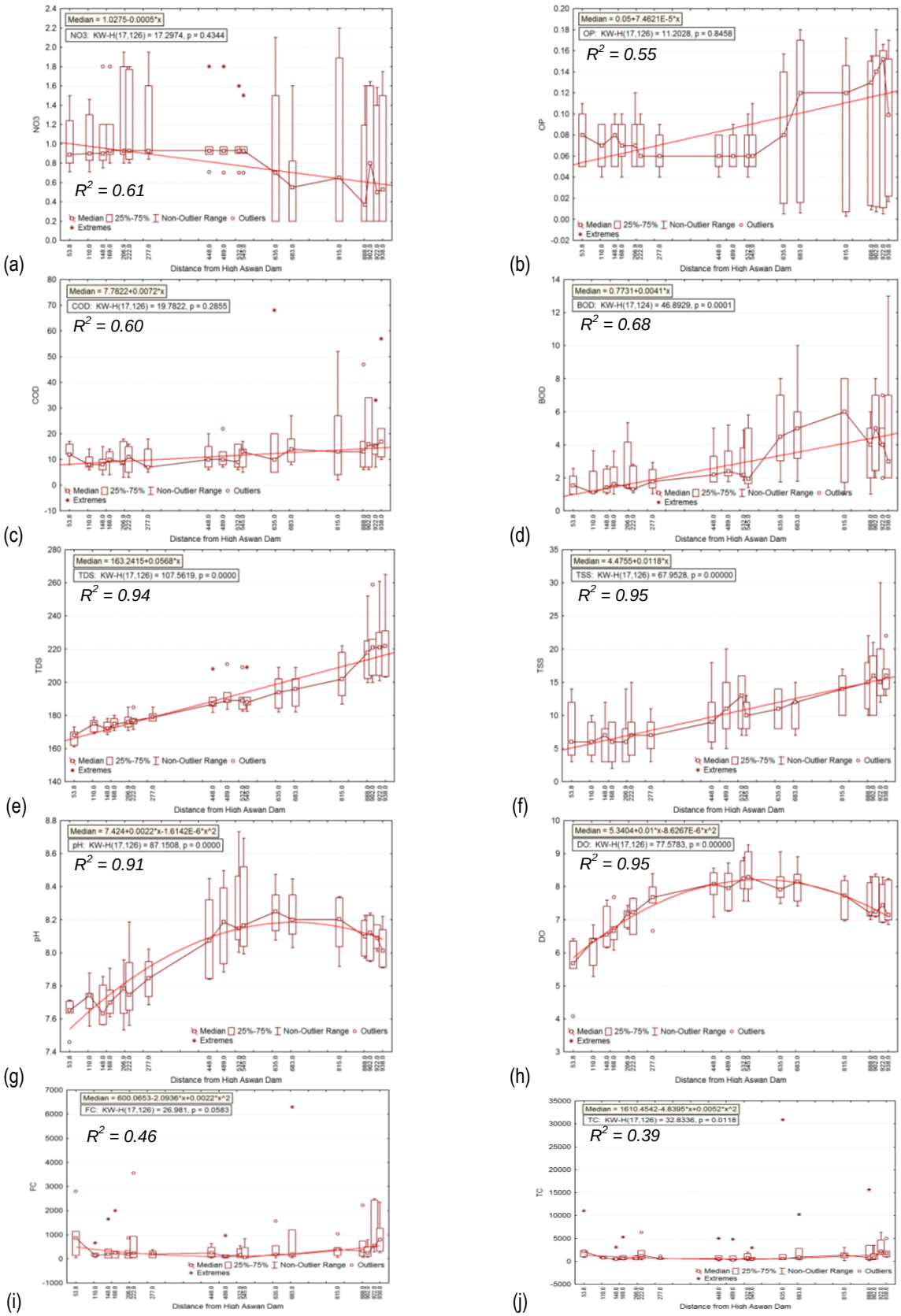

Fig. 7-Box Plots of Water Quality Variability in August with Distance from High Aswan Dam on the X-Axis
(a) NO_3, (b) OP, (c) COD, (d) BOD, (e) TDS, (f) TSS, (g) pH, (h) DO, (i) FC and (j)

f) Inter-Relationships between Variables

The correlation matrix of the variables, calculated based on raw untransformed data, is summarized in Tables 3 to 6. Most of the variables along the main stem of Nile River show statistically significant correlations to each other indicating a close association of these variables.

During February, TDS has a strong positive correlation with a number of variables like chloride, total hardness, magnesium, calcium, sodium, potassium, sulphate, bicarbonate, TSS and COD. Thus, TDS is the single variable that gives a reasonably good indication of a number of related variables. TSS is also well correlated with TP, NO_3 and BOD which indicates that the TSS provides adsorption sites for chemical and biological agents. As expected, Sodium is well correlated with chloride, sulphate and bicarbonate (the presence of salts like NaCl, Na_2SO_4 and $NaHCO_3$ are expected). Also, fecal coliform is well correlated with total coliform and OP with TP. Furthermore, COD has stronger correlation with BOD. Total hardness bears positive correlation with Ca, Mg and Cl. Hence, it is suggested that total hardness of water samples is mainly due to the presence of $MgCl_2$ and $CaCl_2$. All these correlations, following the expected chemical behavior of the variables, provide an insight on the correctness of the measurements.

During August, the pH is well correlated with DO. TSS is also strongly correlated with TP, COD and BOD which confirms the above mentioned remarks related to the adsorption sites. TDS has also a strong correlation with cations and anions. pH has a strong positive correlation with CO_3 which indicates that the increase of pH values could be related to photosynthesis and growth of aquatic plants [11]. Photosynthesis consumes carbon dioxide leading to the rise of pH value. BOD is correlated with NO_2 which indicates the presence of biodegradable organic pollutants. These pollutants may produce nitrite as intermediary product when subjected to microbial oxidation. For Rosetta and Damietta Nile branches, the NO_3 is well correlated with TP, COD, BOD, TDS, TSS, NO_2, OP, HCO_3, SO_4, Cl, Na and Ca during February. The Nitrogen cycle is linked to many quality variables. The TSS has a strong correlation with NO_3, TP, COD and BOD. It is worth mentioning that TDS is well correlated with most of the variables except pH, DO and TC. Salinity is known as a main characteristic of the Nile branches. DO and pH show a high positive correlation. Both variables, closely related, are indicators of water quality.

g) Cluster Analysis

The goal of cluster analysis of variables is to detect the hierarchy of interrelations among a set of variables of a data matrix. The procedure joins the two most similar variables to form a cluster. The amalgamating process continues in a stepwise fashion (joining variables or clusters of variables) until a single cluster is formed that contains all variables.

The algorithm used in this study first calculates dissimilarities between vectors by calculating a so-called distance. Many different types of distances (e.g. Chebichev, multi-block, power, etc.) can be calculated. In the current research, the Pearson correlation coefficient is used as the measure of similarity. Using correlation as the measure of similarity aims to determine reference location in each cluster that is representative of the cluster.

The amalgamation rules in cluster analysis are defined using linkages, i.e. rules determining the distances between clusters. Many linkages algorithms exist (e.g. nearest neighbour, furthest neighbour, Euclidean distance, etc.). Ward's method is used in the current analysis. This method is distinct from all other methods because it uses an analysis of variance approach to evaluate the distances between clusters. It attempts to minimize the Sum of Squares (SS) of any two (hypothetical) clusters that can be formed at each step. In general, this method is regarded as very efficient; however, it tends to create clusters of small size. Variables can thus be grouped into clusters and the process can be repeated to build a hierarchical cluster tree (often called a dendrogram). A selection of a cut-off distance will then determine the "optimal" number of clusters. Cluster Analysis is undertaken twice: using all variables and using only the variables that pass the normality tests. The results of the later are shown in Figures 8 and 9. Cluster analysis is performed for the main Nile stem on pH, DO, TDS, TSS, BOD, COD, FC, TC, NO_3 and OP (Figure 8). Cluster analysis along the Nile River reveals four distinct clusters for winter (February) as well as for summer (August). From Figure 8, different clusters and their members are extracted as follows:

- During February, cluster (1) includes NO_3 and OP; cluster (2) includes FC and TC; cluster (3) includes COD, BOD, TSS and TDS; and finally cluster (4) includes pH and DO.

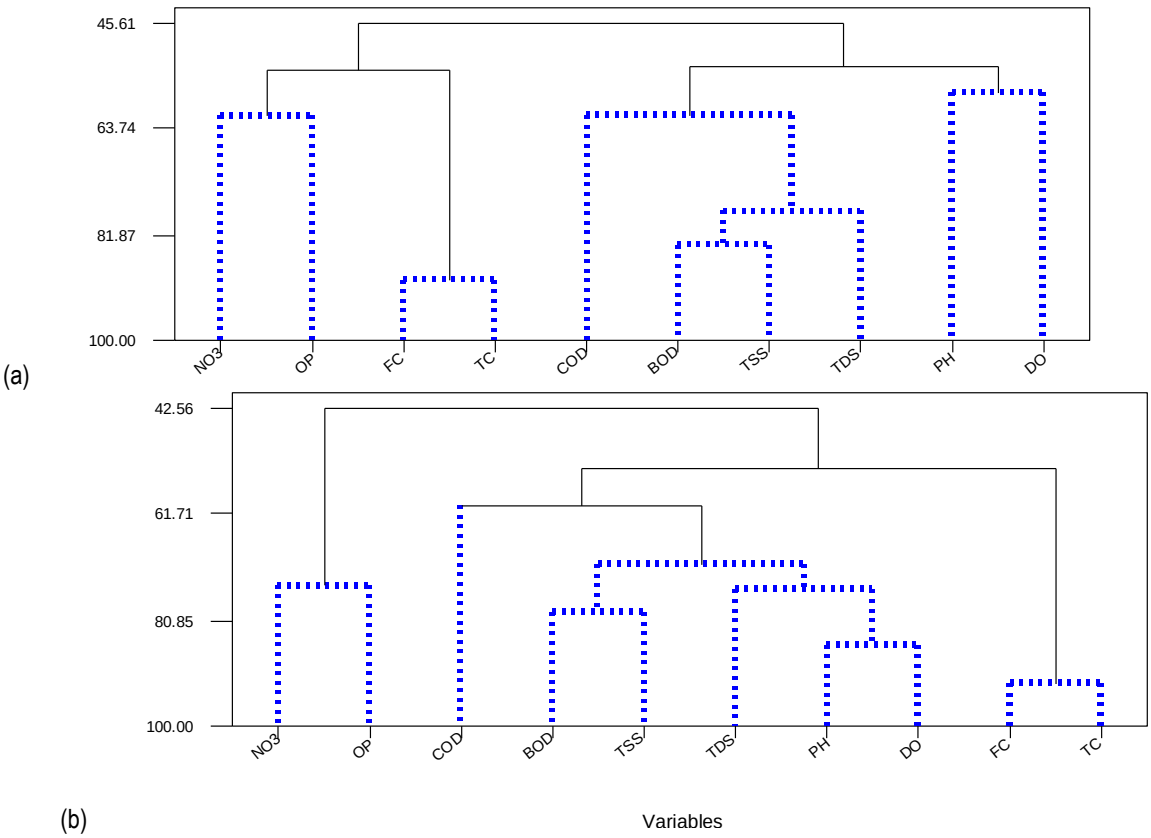

Fig. 8-Cluster analysis of variables along the Nile ((a) February & (b) August)

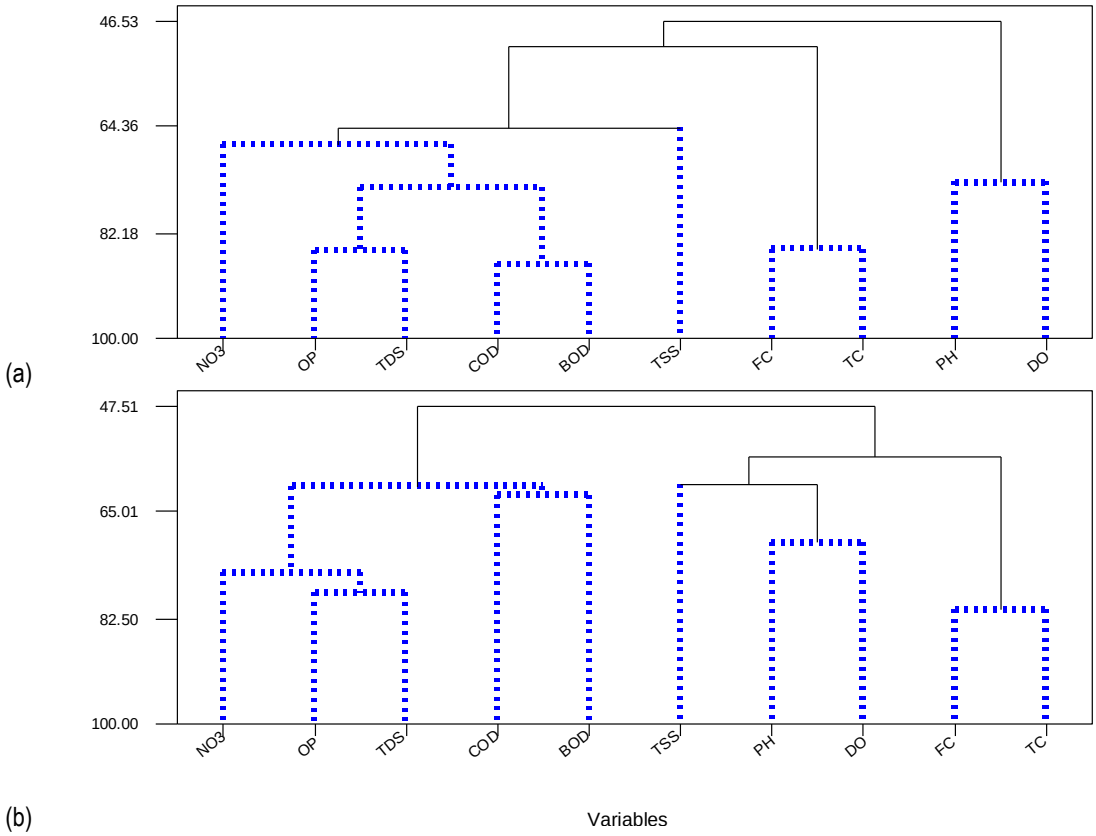

Fig. 9-Cluster analysis of variables along Rosetta and Damietta branches ((a) February & (b) August)

- During August, cluster (1) includes NO_3 and OP; cluster (2) includes COD; cluster (3) includes BOD, TSS, TDS, pH and DO; while cluster (4) includes FC and TC.

Along Rosetta and Damietta branches different clusters are extracted as follows:

- During February, cluster (1) includes NO_3, OP, TDS, COD and BOD; cluster (2) includes: TSS; cluster (3) includes FC and TC, and finally cluster (4) includes pH and DO (Figure 9).
- During August, cluster (1) includes NO_3, OP, TDS, COD and BOD; cluster (2) includes TSS; cluster (3) includes pH and DO; while cluster (4) includes FC and TC.

The above mentioned classification confirms the hierarchy of inter-relationships between water quality variables in the Nile River partly depicted through correlation analysis.

Conclusion and Recommendations

The Exploratory and Cluster Analyses reveal the following:

1. Comparing the mean value of each variable with the national standards of Law 48 / year 1982, it is found that all water quality variables have mean values within the allowable limits except COD. This indicates that the Nile receives highly non-biodegradable organic matter.
2. The large variability of FC, TC, TSS, OP, COD, BOD, NO_2 and NO_3, depicted through box plots, indicate that the Nile and its two branches are exposed to different point sources of pollution (basically of agricultural type mixed with sewage). The normality tests show that the variables, which relatively pass normality tests in February and August on the main Nile and its branches, are as follows: TDS, TSS, OP, pH, DO, BOD, COD, FC, and TC. Most cations and anions did not pass normality tests especially in the main Nile section. Although they don't pass some of the normality tests, nitrate variables will also be included in cluster analyses to assess the variability of nutrients.
3. A significant positive correlation is found between several variables following their expected chemical behavior. This finding provides an insight on the correctness of the measurements. TDS has a strong positive correlation with a number of variables like chloride, hardness, magnesium, calcium, sodium, potassium, sulphate, bicarbonate, TSS and COD. Thus, TDS is the single variable that gives a reasonably good indication of a number of related variables. The correlation between TSS and several other variables shows also that suspended solids may be considered as adsorptive agent for chemical and biological substances. Dissolved oxygen shows significantly negative correlation with all the variables except pH with which a positive correlation is found. Thus DO concentration serves as well as a useful index of water quality in the two branches.
4. Only NO_3, OP, TP, TDS, pH and DO show significant seasonal variation. Organic and bacterial variable values measured in February are not significantly different of those measured in August. Consequently, the seasonal variation is mainly due to agricultural activities.
5. The importance of TDS and TSS, depicted by correlation analysis, emphasizes the importance of the increasing trend found with the distance from High Aswan Dam.
6. Cluster analysis reveals that Nile River and its two branches are exposed to three types of pollutants, nutrients, organic and sewage pollutants.
7. It is recommended to follow this statistical approach in reporting Nile water quality status. Hence, the decision maker will be better informed.

References

[1] Aruga R., Negro G. and Ostacoli G. (1993) *Fresenius J. Anal. Chem.,* 346, 968-975.

[2] Bartels J.H.M., Janse T.A.H.M. and Pijpers F.W. (1985) *Anal. Chim. Acta*; 177, 35-45.

[3] Brown S.D., Skogerboe R.K. and Kowalski B.R. (1980) *Chemosphere*, 9, 265-276.

[4] Librando V. (1991) *Fresenius J. Anal. Chem.*, 339, 613-619.

[5] APHA (1999) *Standard methods for the examination of water and wastewater. American Public Health Association, 17th Ed., Washington DC., USA, 20th edition; 1325 p.*

[6] Chambers J., Cleveland W., Kleiner B. and Tukey P. (1983*) Graphical Methods for Data Analysis, Wadsworth & Brooks/Cole Statistics/Probability Series*, 395 p.

[7] Anderson T.W. and Darling D.A. (1952) *Annals of Mathematical Statistics*, 23, 193–212.

[8] Lilliefors H. (1967) *Journal of the American Statistical Association*, 62, 399-402.

[9] Shapiro S.S. and Wilk M.B. (1965) *Biometrika*, 52, 591-611.

[10] Chapman D. (Ed.) (1992) *Water Quality Assessments, Chapman and Hall, London,*

[11] Entz B.A.G. (1976) *The Nile: Biology of an Ancient River. W. Junk. The Hague*, 271-298.

Table 3- Pearson Correlation Coefficients between Different Water Quality Variables along the Main Stem Nile River during February

	NO_3	T-P	COD	BOD	T.D.S	T.S.S	PH	DO	FC	TC	NO_2	O-P	HCO_3	T.A	SO_4	Cl	Na	K	Ca	Mg	T.H
NO_3	1																				
T-P	0.21	1																			
COD	-0.08	0.04	1																		
BOD	0.05	**0.23**	**0.37**	1																	
T.D.S	0.11	**0.31**	0.13	**0.50**	1																
T.S.S	**0.29**	**0.33**	0.18	**0.60**	**0.61**	1															
PH	-0.04	-0.03	0.02	0.15	0.08	0.14	1														
DO	**-0.27**	-0.03	-0.05	0.01	0.01	0.04	0.13	1													
FC	0.05	-0.07	0.04	-0.11	-0.06	-0.09	**-0.31**	-0.01	1												
TC	**0.31**	0.06	0.03	0.07	0.16	0.19	-0.15	0.07	**0.77**	1											
NO_2	-0.14	0.12	-0.02	**0.55**	**0.56**	**0.55**	-0.04	0.10	-0.03	0.07	1										
O-P	0.06	**0.30**	0.16	**0.19**	0.11	0.04	0.13	**-0.37**	-0.03	-0.11	**0.18**	1									
HCO_3	-0.16	-0.03	**0.18**	0.12	0.07	**0.22**	**0.67**	**0.32**	-0.17	-0.03	-0.02	-0.15	1								
T.A	-0.05	0.17	0.08	**0.37**	**0.69**	**0.46**	**-0.23**	**-0.26**	0.01	0.04	**0.59**	0.11	**-0.36**	1							
SO_4	-0.12	0.10	0.17	**0.35**	**0.56**	**0.42**	-0.06	-0.08	-0.01	0.05	**0.39**	-0.06	-0.04	**0.74**	1						
Cl	0.02	0.15	**0.25**	0.17	**0.47**	**0.30**	0.00	**0.47**	**0.28**	**0.33**	-0.13	**-0.27**	**0.25**	0.08	**0.22**	1					
Na	0.17	**0.34**	0.15	**0.66**	**0.79**	**0.60**	**0.27**	0.04	-0.04	**0.20**	**0.76**	**0.19**	0.14	**0.53**	**0.37**	**0.25**	1				
K	**-0.24**	-0.10	**0.28**	**0.34**	**0.50**	**0.41**	0.13	**0.19**	0.12	0.16	**0.50**	**-0.24**	**0.31**	**0.52**	**0.55**	**0.35**	**0.50**	1			
Ca	-0.12	0.13	-0.01	**0.48**	**0.61**	**0.41**	-0.04	-0.03	**-0.19**	-0.13	**0.85**	**0.21**	-0.12	**0.65**	**0.41**	-0.12	**0.59**	**0.38**	1		
Mg	**0.21**	**0.41**	-0.09	**0.40**	**0.79**	**0.51**	**0.22**	0.11	-0.06	**0.28**	**0.53**	-0.15	0.17	**0.38**	**0.38**	**0.34**	**0.67**	**0.35**	**0.37**	1	
T.H	**-0.25**	-0.07	0.18	**0.45**	**0.36**	**0.36**	0.16	0.16	-0.15	-0.14	**0.56**	-0.09	0.12	**0.42**	**0.36**	0.04	**0.30**	**0.54**	**0.52**	**0.19**	1

Marked correlations are significant at $p < 0.05$

Table 4- Pearson Correlation Coefficients between Different Water Quality Variables along the Main Stem Nile River during August

	NO_3	T-P	COD	BOD	T.D.S	T.S.S	PH	DO	FC	TC	NO_2	O-P	HCO_3	T.A	SO_4	Cl	Na	K	Ca	Mg	T.H
NO_3	1																				
T-P	**-0.24**	1																			
COD	**-0.33**	**0.36**	1																		
BOD	**-0.25**	**0.42**	0.16	1																	
T.D.S	-0.03	**0.41**	**0.29**	**0.30**	1																
T.S.S	**-0.25**	**0.27**	**0.47**	**0.47**	**0.58**	1															
PH	-0.07	0.02	**0.27**	**0.37**	**0.52**	**0.53**	1														
DO	0.08	-0.08	-0.06	0.17	**0.42**	**0.23**	**0.72**	1													
FC	-0.01	0.13	0.12	0.02	0.10	0.13	0.01	-0.08	1												
TC	**-0.19**	0.10	**0.46**	0.13	0.05	0.16	0.13	-0.03	**0.60**	1											
NO_2	**-0.29**	**0.31**	-0.02	**0.46**	**0.30**	**0.23**	0.10	0.03	0.01	-0.01	1										
O-P	**0.25**	**0.65**	-0.09	**0.22**	**0.34**	0.03	-0.05	0.02	-0.12	**-0.22**	**0.22**	1									
HCO_3	**0.23**	**-0.28**	0.13	0.08	0.16	0.14	**0.63**	**0.36**	0.00	0.10	-0.12	**-0.20**	1								
T.A	-0.05	**0.26**	0.09	**0.34**	**0.46**	**0.33**	0.16	0.16	**0.19**	0.06	**0.27**	**0.27**	-0.30	1							
SO_4	0.04	0.16	0.14	**0.38**	**0.54**	**0.39**	**0.40**	**0.30**	**0.20**	0.10	**0.23**	**0.21**	0.04	**0.94**	1						
Cl	**0.21**	**-0.25**	-0.08	-0.14	**0.36**	0.12	**0.22**	**0.37**	**-0.18**	**-0.18**	-0.12	0.05	0.19	**-0.25**	**-0.20**	1					
Na	-0.15	**0.50**	**0.20**	**0.30**	**0.61**	**0.36**	**0.32**	**0.36**	-0.01	-0.03	**0.33**	**0.50**	-0.12	**0.33**	**0.30**	**0.25**	1				
K	**0.52**	0.08	-0.16	**0.21**	**0.63**	**0.21**	**0.37**	**0.32**	-0.02	**-0.18**	**0.18**	**0.53**	**0.31**	**0.33**	**0.46**	**0.35**	**0.31**	1			
Ca	-0.11	**0.51**	**0.26**	0.07	**0.51**	**0.19**	**0.25**	0.12	0.06	0.07	**0.21**	**0.43**	0.04	**0.28**	**0.30**	-0.16	**0.59**	**0.29**	1		
Mg	**-0.21**	**0.35**	**0.28**	**0.49**	**0.71**	**0.60**	**0.43**	**0.30**	0.16	**0.19**	**0.60**	0.13	0.09	**0.46**	**0.52**	0.02	**0.52**	**0.38**	**0.39**	1	
T.H	-0.03	-0.08	-0.01	-0.10	**0.23**	0.04	0.10	**0.36**	-0.11	-0.09	-0.01	**0.18**	-0.05	**0.30**	**0.30**	**0.31**	**0.49**	0.10	0.15	0.15	1

Marked correlations are significant at $p < 0.05$

Table 5- Pearson Correlation Coefficients between Different Water Quality Variables along Rosetta and Damietta Branches during February

	NO_3	T-P	COD	BOD	T.D.S	T.S.S	PH	DO	FC	TC	NO_2	O-P	HCO_3	T.A	SO_4	Cl	Na	K	Ca	Mg	T.H
NO_3	1																				
T-P	**0.72**	1																			
COD	**0.50**	**0.69**	1																		
BOD	**0.60**	**0.66**	**0.83**	1																	
T.D.S	**0.51**	**0.38**	**0.39**	**0.54**	1																
T.S.S	**0.38**	**0.39**	**0.41**	**0.52**	**0.28**	1															
PH	-0.08	0.07	0.10	0.17	-0.25	0.18	1														
DO	0.09	0.02	0.00	0.11	-0.25	0.18	**0.50**	1													
FC	-0.05	-0.06	-0.01	-0.02	**-0.29**	-0.01	-0.02	0.00	1												
TC	0.12	0.00	0.06	0.16	-0.17	0.14	0.04	-0.03	**0.72**	1											
NO_2	**0.48**	**0.51**	0.05	**0.39**	**0.58**	**0.62**	0.01	0.11	0.04	0.17	1										
O-P	**0.42**	**0.48**	**0.39**	**0.35**	**0.62**	**0.31**	-0.10	-0.12	-0.16	-0.04	**0.43**	1									
HCO_3	**0.35**	0.10	0.20	**0.29**	**0.78**	**0.38**	-0.20	-0.18	-0.18	-0.20	**0.67**	**0.64**	1								
T.A	**0.35**	0.10	0.20	**0.29**	**0.78**	**0.38**	-0.20	-0.18	-0.18	-0.20	**0.67**	**0.64**	1	1							
SO_4	**0.35**	**0.30**	**0.40**	**0.38**	**0.76**	0.17	**-0.33**	-0.26	-0.19	-0.21	0.27	**0.65**	**0.62**	**0.62**	1						
Cl	**0.50**	0.26	**0.40**	**0.44**	**0.82**	**0.28**	-0.24	**-0.28**	-0.25	-0.13	**0.50**	**0.60**	**0.67**	**0.67**	**0.71**	1					
Na	**0.46**	0.15	0.27	0.22	**0.60**	**0.35**	-0.10	-0.06	-0.23	-0.17	**0.46**	**0.58**	**0.59**	**0.59**	**0.51**	**0.80**	1				
K	0.25	-0.03	-0.01	0.26	**0.65**	**0.33**	-0.13	-0.06	-0.10	-0.09	**0.78**	**0.36**	**0.79**	**0.79**	**0.32**	**0.47**	**0.36**	1			
Ca	**0.41**	**0.67**	**0.46**	**0.63**	**0.60**	0.19	-0.12	-0.07	-0.01	0.08	0.27	0.20	0.16	0.16	**0.54**	**0.28**	-0.09	0.16	1		
Mg	-0.03	-0.20	-0.11	0.05	**0.54**	0.11	-0.16	-0.20	-0.17	-0.22	**0.37**	0.22	**0.64**	**0.64**	**0.34**	**0.34**	0.19	**0.65**	0.18	1	
T.H	**0.28**	0.23	0.17	**0.33**	**0.72**	0.27	-0.13	-0.21	-0.12	-0.09	**0.42**	**0.46**	**0.64**	**0.64**	**0.52**	**0.48**	0.22	**0.54**	**0.51**	**0.75**	1

Marked correlations are significant at $p < 0.05$

Table 6-Pearson Correlation Coefficients between Different Water Quality Variables along Rosetta and Damietta Branches during August

	NO_3	T-P	COD	BOD	T.D.S	T.S.S	PH	DO	FC	TC	NO_2	O-P	HCO_3	T.A	SO_4	Cl	Na	K	Ca	Mg	T.H
NO_3	1																				
T-P	**0.43**	1																			
COD	0.15	0.15	1																		
BOD	**0.33**	**0.49**	**0.37**	1																	
T.D.S	**0.41**	**0.34**	0.17	0.22	1																
T.S.S	-0.25	-0.22	0.08	0.23	**-0.44**	1															
PH	-0.03	-0.11	0.18	0.10	-0.23	0.06	1														
DO	0.04	-0.12	0.21	-0.12	**-0.36**	0.16	0.22	1													
FC	-0.07	-0.05	**0.42**	0.25	-0.03	0.18	0.24	0.12	1												
TC	-0.08	-0.05	**0.56**	0.07	0.00	0.10	0.19	-0.04	**0.63**	1											
NO_2	0.15	-0.09	-0.20	-0.11	0.03	-0.23	-0.12	-0.23	-0.07	-0.05	1										
O-P	**0.39**	**0.87**	0.28	**0.36**	**0.53**	**-0.31**	-0.07	-0.27	0.01	0.02	-0.09	1									
HCO_3	**0.50**	0.04	0.11	0.21	**0.49**	-0.15	-0.20	-0.04	0.00	0.03	-0.18	0.16	1								
T.A	**0.50**	0.04	0.11	0.21	**0.49**	-0.15	-0.20	-0.04	0.00	0.03	-0.18	0.16	1	1							
SO_4	**0.52**	**0.30**	0.13	0.15	**0.67**	**-0.37**	0.02	**-0.29**	0.02	0.06	0.17	**0.46**	**0.39**	**0.39**	1						
Cl	**0.49**	**0.73**	0.22	**0.36**	**0.73**	**-0.40**	-0.06	**-0.32**	0.05	0.00	-0.02	**0.78**	**0.33**	**0.33**	**0.70**	1					
Na	**0.50**	**0.61**	0.16	0.11	**0.60**	**-0.51**	-0.06	-0.24	-0.08	-0.08	0.13	**0.60**	**0.32**	**0.32**	**0.73**	**0.79**	1				
K	-0.01	**0.35**	0.11	0.19	**0.42**	-0.23	-0.17	-0.13	-0.08	-0.06	-0.10	**0.49**	-0.13	-0.13	0.10	**0.41**	0.06	1			
Ca	0.03	-0.19	0.01	0.07	0.27	0.01	0.10	**-0.39**	**0.38**	0.63	0.19	-0.04	0.19	0.19	**0.30**	0.08	0.01	-0.18	1		
Mg	0.03	-0.04	0.09	-0.01	**0.33**	-0.12	-0.07	0.17	0.17	0.15	-0.22	0.22	0.27	0.27	0.21	0.27	0.10	**0.53**	-0.01	1	
T.H	0.08	-0.15	0.13	0.04	**0.47**	-0.12	-0.01	-0.08	0.25	0.14	-0.03	0.17	**0.35**	**0.35**	**0.37**	**0.31**	0.13	**0.36**	**0.42**	**0.82**	1

Marked correlations are significant at $p < 0.05$