

Multivariate statistical analysis and water quality index to assessment water quality in lakes

Diaa Seif ^{a)}, Medhat Moustafa, Walid Elbarqi

Alexandria University, Faculty of Engineering, Sanitary Engineering Department, Alexandria 21544, Egypt

Abstract— This study concerns using Multivariate statistical analysis to handle the large complex datasets in lakes and use water quality index factor to estimate the water quality in lakes. This study investigated the seasonal and spatial variations of water quality parameters in Burullus Lake in Egypt as a case study. Significant seasonal changes ($p < 0.05$) were observed in temperature, pH and phosphate ($\text{PO}_4\text{-P}$), whereas significant spatial differences ($p < 0.05$) were detected in pH, salinity, dissolved oxygen (DO), ammonium-nitrogen ($\text{NH}_4\text{-N}$), nitrite-nitrogen ($\text{NO}_2\text{-N}$), nitrate-nitrogen ($\text{NO}_3\text{-N}$), $\text{PO}_4\text{-P}$, silicate (SiO_4) and Chlorophyll-*a* (Chl-*a*). The water quality index (WQI) estimated using the inputs of: salinity, DO, $\text{NH}_4\text{-N}$, $\text{NO}_3\text{-N}$, $\text{PO}_4\text{-P}$, and Chl-*a*. The (WQI) of the lake was rated as “Bad” to “Very bad”. Based on principal component analysis (PCA), the first principal component (PC1: 38.78%) represented high loadings on $\text{NH}_4\text{-N}$: 0.46, $\text{NO}_3\text{-N}$: 0.45, and $\text{PO}_4\text{-P}$: 0.45, indicating that the lake was mainly influenced by nutrients concentrations coming from agricultural lands. The results from cluster analysis and a dendrogram indicated that Burullus Lake was mainly influenced by spatial variations rather than seasonal changes.

Keywords— Dendrogram; Principal component analysis; Spatio-temporal variability; Water quality index.

I. Introduction

The quality of water resources is one of the major issues that threaten the economic and social development of Egypt. Water pollution has become a crucial environmental concern owing to industrialization and increasing dependence on agrochemicals. The Egyptian Mediterranean coast contains five lakes, namely Maryut, Idku, Burullus, Manzala, and Bardawil; also known as Northern delta lakes. The lakes are situated along the Nile delta coast, where they receive drainage water from agricultural and urban areas. The lakes exhibit coastal ecosystems for estuarine and wintering waterbirds. Moreover, they provide various recreational

activities and afford rich habitats for more than 75% of fish production in Egypt.

Burullus Lake is the second largest delta lake in Egypt, which receives a total drainage discharge of $3,904 \times 10^6 \text{ m}^3\text{-y}^{-1}$. The lake is a consistent water body for fulfilling the needs of domestic, agricultural and industrial activities. Recently, Burullus Lake has become a dumping site for huge amounts of agricultural runoff coming from sectors/lands of the Delta region. Based on flow mass balance of the lake, discharges from agricultural activities are responsible for more than 90% of the total inflows into the lake. The agricultural discharge contains washed and leached fertilizers and pesticides that increase the pollutants and nutrient influx. Additionally, the lake receives wastewater coming from urban sewage and industrial facilities. Currently, the lake is facing many challenges, such as eutrophication, high load of contamination, reduction of the area, and depletion of fish production. As a result, Burullus Lake has become unfit for drinking, agricultural and other domestic purposes. In this context, water quality assessment of Burullus Lake should be undertaken to provide a useful tool that could help the decision makers in determining the extent of pollution and ecological processes [1].

Multivariate statistical approaches, such as principal component analysis (PCA), cluster analysis (CA) and a dendrogram have been performed to identify water pollution sources and to assess water quality condition in a given aquatic system. PCA provides deep information on the most meaningful parameters (i.e. simplify the data structure) by presenting statistical correlations among constituents. PCA of the normalized variables are executed to extract significant principal components (PCs) and to reduce the data set with minimal loss of original information. CA is a group of multivariate techniques performed to describe the relationships between data records. CA splits a huge number of objects into small homogenous groups based on their correlation structure. In CA, related parameters are grouped into certain distinct classes in such a way that those within a group are more similar to each other (achieve homogeneity) than to objects in other groups; i.e. similarity among groups is minimized. A dendrogram is a tree diagram that provides a visual summary of the clustering processes. It present a picture of the groups

and their proximity with a dramatic reduction in dimensionality of the original data [2].

Mathematical modelling, in conjunction with field surveys, is a promising solution for predicting seasonal/spatial variations in water quality variables characterizing an aquatic ecosystem.

To the best of our knowledge, there is limited information about water quality of the lake systems in Egypt. Thus, the present study focused on studying the water quality of Burullus Lake by measuring several parameters, including temperature, pH, salinity, dissolved oxygen (DO), ammonia nitrogen ($\text{NH}_4\text{-N}$), nitrite-nitrogen ($\text{NO}_2\text{-N}$), nitrate-nitrogen ($\text{NO}_3\text{-N}$), phosphate ($\text{PO}_4\text{-P}$), silicate (SiO_4) and chlorophyll-*a* (Chl-a).

The study aimed at:

- 1) providing an insight in seasonal and spatial variations of water quality parameters over the period 2000 – 2013,
- 2) applying multivariate statistical techniques (correlation matrix, PCA, CA, and a dendrogram) to enhance the understanding of water quality and to aid the interpretation of large complex water quality datasets,
- 3) estimating the water quality index (WQI).

II. Materials and Methods

Study area

Burullus Lake is located in the north-western region of the Nile Delta, between Damietta and Rosetta branches of the Nile River. Geographically, it is situated between longitude $30^\circ 33' - 31^\circ 07'$ E and latitude $31^\circ 22' - 31^\circ 26'$ N. The eastern side of the lake is connected to the Mediterranean Sea through an opening, namely Boghaz Al-Burullus. The lake has dimensions of 70 km in length, 6 – 14 km in width and 0.42 – 2.07 m in depth, with an approximate surface area of 515 Km^2 [3].

Data collection

The recorded data over a period ranging from 2000 to 2013 were quality-certified by checking for errors and anomalies. The comprehensive database elucidated seven zones in the lake, as follows: zone 1 (Z1) north-east, zone 2 (Z2) south-east, zone 3 (Z3) north-middle, zone 4 (Z4) south-middle, zone 5 (Z5) north-west, zone 6 (Z6) south-west, and zone 7 (Z7). This classification was proposed in order to define the best spatial framework for an efficient and reliable water quality assessment of the Lake. Winter data included water samples harvested during January – March, spring data represented those taken during April – June, summer data included those collected during July – September, while autumn data comprised samples from October to December.

Analytical analysis

In this study, the selected water quality parameters were temperature, pH, dissolved oxygen (DO), salinity, ammonium-nitrogen ($\text{NH}_4\text{-N}$), nitrite-nitrogen ($\text{NO}_2\text{-N}$), nitrate-nitrogen

($\text{NO}_3\text{-N}$), phosphate ($\text{PO}_4\text{-P}$), silicate (SiO_4) and Chlorophyll-*a* (Chl-a). The measured parameters were compared with guidelines of the water quality standards of the United Nations Economic Commission for Europe (UNECE), Environmental Protection Agency (EPA), and World Health Organization WHO [4].

All measurements were conducted in accordance to standard methods of American Public Health Association (APHA). Temperature and pH were monitored on site (at the sampling stations) with a mercury thermometer (range 0° to 100°C) and a digital portable pH meter, respectively. Dissolved oxygen was estimated according to Strickland and Parson (1968). Water salinity was measured using Bechman salinometer (Model NO. R.S.10). Nutrient salts (N and P species) and Chl-a were measured using a Shimadzu double beam spectrophotometer UV-150-02.

Calculation of water quality index

Water quality index (WQI) is basically a mathematical tool used to calculate a single output (indicator) from multiple water quality parameters. The WQI designates the overall water state/condition at various locations over a period of time in a particular water basin.

The calculation of WQI for Burullus Lake was conducted using weighted arithmetic index method. In this method, each water quality parameter is transferred to a weighting-curve chart in order to obtain a numerical value (Q_i). The Q_i is multiplied by a weighting factor (W_i), and the resulting values are aggregated to obtain an overall WQI, as described in Eq. 1. Basically, WQI value “water quality status” can be classified as, 0 – 25 “Very bad”, 26 – 50 “Bad”, 51 – 70 “Medium”, 71 – 90 “Good”, and 91 – 100 “Excellent” [5].

$$WQI = \sum_{i=1}^n Q_i W_i \quad \text{Eq. 1}$$

Where

Q_i is the sub-index for water quality parameter; W_i is weight associated with water quality parameter; n is number of water quality parameters

Principal component analysis

PCA is designed to rotate the original variables into new and orthogonal variables (axes), called principal components (PCs). This can be achieved by transforming the original data matrix into a product of two matrices, one of which contains the information about the samples and the other about the variables. The PCs lie along the directions of maximum variance, and thus, provide information on the most meaningful parameters. The eigenvalues measure the associated variance of the PCs. The loadings are the weights of the original variables in each PC, whereas the scores are the projection of the samples on PCs [6].

PCA involves four major steps, as follows: 1) all parameters are z-scale standardized with zero mean and unit variance, that is, to ensure that all the measurements have equal weight in the analysis; 2) singular value decomposition is performed to the sample matrix A with the rows (samples)

and columns (variables), described as $A=UAVT$; 3) the loading matrix (V) and eigenvalues (the square of Λ) are obtained; and 4) the data are projected onto the PCs (loading), and then the score matrix $T=AV$ is estimated [7].

Cluster analysis

Hierarchical is the most commonly method that can demonstrate distinction among different types of clustering. Cluster analysis involve the following four major steps: 1) all parameters are z-scale standardized into zero mean and unit variance; 2) Euclidean distance of the standardized data is performed to calculate the distance between objects; 3) pairs of objects in close proximity (distance) are linked; 4) as objects are combined into binary clusters, the newly formed clusters are grouped into larger clusters until a hierarchical tree is formed; and 5) the hierarchical tree is naturally divided into distinct, well-separated clusters [8]. A dendrogram is a tree diagram that illustrates the arrangement of the clusters produced by hierarchical clustering. Ward's minimum variance hierarchical linkage method is employed to construct the similarity levels [9].

III. Results and Discussion

Seasonal variations

The measured water quality parameters among the four seasons were analyzed using a one-way analysis of variance (ANOVA1). Figure(1) shows the seasonal variations of water quality parameters for Burullus Lake represented by their median values and percentiles (whiskers). According to the results, clear seasonal variations in temperature, pH and PO₄-P were detected, however statistically non-significant changes were witnessed for the other water quality parameters.

Temperature:

The temperature of Burullus Lake showed a significant seasonal cycle during the study period ($p < 0.01$). Summer was the season with the warmest water, having an average value of $27.63 \pm 0.74^\circ\text{C}$ (ranging from 26.19°C to 28.33°C), whereas winter was the season with the coldest water, having an average value of $16.43 \pm 1.37^\circ\text{C}$, ranging from 14.10°C to 18.40°C . This variation in temperature could control the aquatic plant photosynthesis and respiration, affect the metabolism of fish and withstand chemical reaction rates and DO solubility.

Hydrogen ion concentration (pH):

The pH of Burullus Lake in the present study varied seasonally as follows ($p < 0.01$): 7.31 ± 0.54 (winter), 8.40 ± 0.60 (spring), 8.50 ± 0.60 (summer), and 7.89 ± 0.51 (fall). The observations of pH during winter and summer were neutral (pH 6.6 – 7.3) and strongly alkaline (pH 8.5 – 9.0), respectively, whereas spring and fall seasons were moderately alkaline (pH 7.9 – 8.4). Additionally, based on the average pH during the year of 6.5 – 8.4, the lake can be used for irrigation purposes. According to the UNECE, the water quality can be classified as Class I (pH 9.0 – 6.5) for the maintenance of

aquatic life. Thus, the pH-standards indicated that most of data were moderately alkaline in nature, recommended for irrigation, and suitable for different kinds of aquatic life [10].

Salinity:

Seasonally, the average values of salinity showed non-significant variation ($p: 0.733$), recording minimum and maximum levels during summer ($0.24 - 8.24 \text{ g L}^{-1}$) and winter ($0.74 - 12.14 \text{ g L}^{-1}$), respectively. The average salinity of the lake was medium saline ($2.0 - 4.0 \text{ g L}^{-1}$) to highly saline ($> 4.0 \text{ g L}^{-1}$). According to Ayers and Westcot, the salinity concentrations were found to exceed the severe degree of restriction on reuse ($> 2.0 \text{ g L}^{-1}$).

Dissolved oxygen (DO):

Measurements of DO are widely used to determine the ability of a water body to support a well-balanced aquatic life. The DO varied within the range of $6.04 - 8.44$, $2.96 - 7.70$, $2.09 - 7.45$, and $5.11 - 7.94 \text{ mg L}^{-1}$ during winter, spring, summer and fall, respectively. This indicated that minimum DO of $4.91 \pm 1.98 \text{ mg L}^{-1}$ was observed during summer, while winter season provided the maximum level of DO: $7.04 \pm 0.89 \text{ mg L}^{-1}$, with a p-value of 0.053. The relatively high content of DO in Burullus Lake during winter could be attributed to the low discharge of irrigation water to the lake, known in Egypt as a low-flow period of "El-Sada El-Shetwya". Other possible explanation is due to the fact that the solubility of oxygen decreases as water temperature increases, and thus DO increased in winter. Additionally, a clear drop in DO concentrations during summer season could be due to their consumption by microorganisms for the decay of organic substances. This seasonal pattern is consistent with other studies that described peak levels of DO in winter. According to UNECE, the DO standard for the maintenance of aquatic life is Class I ($\text{DO} > 7 \text{ mg L}^{-1}$) to Class II ($\text{DO} 6 - 7 \text{ mg L}^{-1}$) during winter, and Class III ($\text{DO}: 6 - 4 \text{ mg L}^{-1}$) during summer [11].

Nutrients (nitrogen and phosphorus):

The increase in nutrients concentrations reflects severe effects of domestic and agriculture effluents on the water quality. Slight variations in the $\text{NH}_4\text{-N}$ values ($p: 0.085$) were observed between $0.07 - 2.80 \text{ mg L}^{-1}$ during summer and $1.55 - 4.70 \text{ mg L}^{-1}$ during winter. $\text{NH}_4\text{-N}$ can enter the aquatic environment through different means, such as municipal effluent discharges, excretion of wastes from animals, and runoff from agricultural lands.

The average concentration of $\text{NO}_2\text{-N}$ during the year was noted within the range $0.22 - 2.46 \text{ mg L}^{-1}$, with an average level of $1.03 \pm 0.71 \text{ mg L}^{-1}$ ($p: 0.755$). $\text{NO}_2\text{-N}$ is not a stable nitrogen species, and it can appear in the water mainly as a result of biochemical oxidation of ammonia (nitrification) or the reduction of nitrate (denitrification). However, $\text{NO}_2\text{-N}$ concentration in surface runoff water is generally irrelevant [4].

Nitrate ($\text{NO}_3\text{-N}$) is the most stable form of inorganic nitrogen in oxygenated water, which can be leached from the surrounding agricultural land. The decrease of $\text{NO}_3\text{-N}$ concentration ($p: 0.160$) during summer ($0.26 - 1.51 \text{ mg L}^{-1}$)

compared to winter ($0.76 - 2.32 \text{ mg L}^{-1}$), spring ($0.64 - 2.20 \text{ mg L}^{-1}$), and fall ($0.70 - 1.64 \text{ mg L}^{-1}$) could be attributed to several factors including the assimilation by plants and denitrification. According to Ayers and Westcot, 1985 the degree of restriction for $\text{NO}_3\text{-N}$ in Burullus Lake on reuse was "None". Concentration of $\text{NO}_3\text{-N}$ higher than 0.3 mg L^{-1} indicates a potential for eutrophication of streams [5].

Average $\text{PO}_4\text{-P}$ for the lake during winter ($2.21 \pm 1.89 \text{ mg L}^{-1}$) was significantly higher ($p < 0.01$) than during spring ($1.83 \pm 1.72 \text{ mg L}^{-1}$), summer ($1.53 \pm 1.55 \text{ mg L}^{-1}$) and fall ($1.89 \pm 1.67 \text{ mg L}^{-1}$). The lower $\text{PO}_4\text{-P}$ concentration in summer could be attributed to high consumption by phytoplankton crop during blooming.

Silicate (SiO_4):

The minimum seasonal distribution of SiO_4 occurred during summer ($3.18 \pm 1.19 \text{ mg L}^{-1}$), whereas the maximum value of $3.43 \pm 1.19 \text{ mg L}^{-1}$ was observed during winter ($p: 0.985$).

Chlorophyll-*a* (Chl-*a*):

Chl-*a* concentration is considered as a good indicator of the phytoplankton biomass. The temporal distribution of Chl-*a* in the lake ($p: 0.777$) attained a minimum value of $0.79 \pm 0.38 \text{ mg L}^{-1}$ during winter, tended to increase during spring ($0.92 \pm 0.41 \text{ mg L}^{-1}$) reaching a maximum level of $0.99 \pm 0.41 \text{ mg L}^{-1}$ in summer, and then dropped during fall to $0.83 \pm 0.38 \text{ mg L}^{-1}$. According to UNECE, the average Chl-*a* : $0.88 \pm 0.39 \text{ mg L}^{-1}$ during the year for the maintenance of aquatic life is Class V (i.e. Chl-*a* $> 0.11 \text{ mg L}^{-1}$) [4].

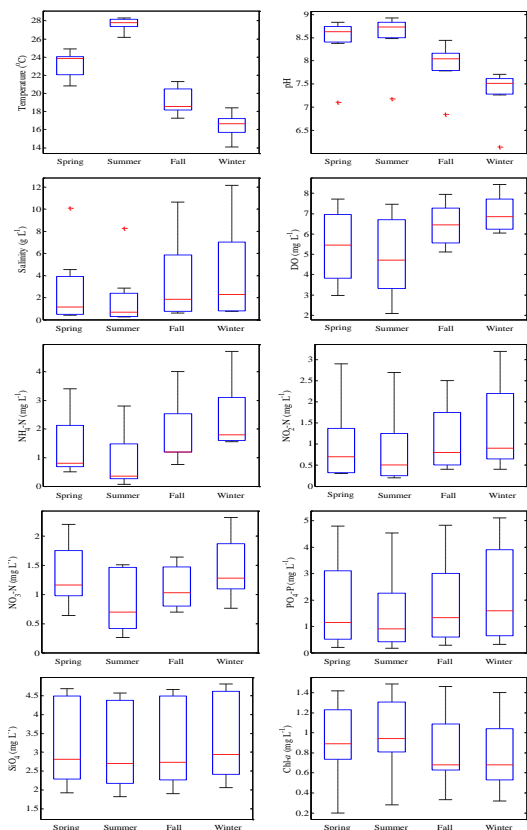


FIGURE 1. Seasonal variations of water quality parameters for Burullus Lake using box plots. In each box plot, the central line indicates the median, the box represents the interval 25 – 75% percentiles, and the segment gives the range.

Spatial variations

Figure (2) presents spatial variations of water quality parameters for Burullus Lake using box-whiskers plots. A one-way analysis of variance (ANOVA1) indicated that there were significant differences in the measured variables except in temperature and pH among the seven zones in Burullus Lake.

Temperature:

The maximum temperature was attained at zone-7 reaching $23.01 \pm 4.23^\circ\text{C}$, while the minimum temperature was about $19.91 \pm 5.03^\circ\text{C}$ at zone-5. This indicated that the difference in temperature among the zones was statistically non-significant ($p: 0.984$). This non-significant variation in the lake temperature could be linked to the small region of latitude between $31^\circ 22'$ and $31^\circ 26'$ N Hydrogen ion concentration (pH):

The highest pH-value was observed at zone-2 (8.48 ± 0.55), while the lowest pH-value of 6.82 ± 0.47 was detected at zone-4 ($p < 0.01$). Based on WHO (1996), zone-2 is considered strongly alkaline (pH 8.5 – 9.0), whereas zone-4 represents a neutral condition (pH 6.6 – 7.3). A previous study by Okbah and Hussein (2006) reported that the pH of Burullus Lake was influenced by the density of phytoplankton count and the water quality discharged into the lake [12].

Salinity:

The lowest salinity ($0.50 \pm 0.22 \text{ g L}^{-1}$) was recorded at zone-7, while zone-1 demonstrated the highest salinity of $10.27 \pm 1.61 \text{ g L}^{-1}$, with a p-value < 0.01 . Zone-7 was located far from adjoining open of Mediterranean Sea, and thus its salinity was not affected by the rate of exchange with the sea water. Thus, the significant variation in water salinity could be explained by the mixing of zone-1 with salty seawater; i.e. Burullus Lake is connected with the Mediterranean Sea through El-Boughaz opening. According to FAO/WHO (1984), salinity conditions in zone-1 and zone-7 can be classified as "Very saline" and "Non-saline", respectively [4].

Dissolved oxygen (DO):

The lowest DO value was observed at zone-7 ($4.05 \pm 1.84 \text{ mg L}^{-1}$), while zone-1 showed a significant higher DO level ($p < 0.01$) of $7.88 \pm 0.42 \text{ mg L}^{-1}$. The horizontal distribution of DO revealed that its concentration increased from the west part to the east part. According to UNECE (1994), the DO standard for the maintenance of aquatic life is "Class I" in zone-1 and "Class III" in zone-7.

Nutrients (nitrogen and phosphorus):

The regional variation of $\text{NH}_4\text{-N}$ was fluctuating ($p < 0.01$) between $0.73 \pm 0.64 \text{ mg L}^{-1}$ at zone-3 to $3.73 \pm 0.81 \text{ mg L}^{-1}$ at zone-6. The significant spatial variation of $\text{NH}_4\text{-N}$ matches the chronology of discharging large amounts of drainage water

coming from agricultural sectors through drains such as Nasser drain and El-Garbiya drain. The lowest concentration of NH₄-N was observed at zone-3 (a zone far away from drains). However, the highest concentration of NH₄-N at zone-6 resulted from the discharges of drain-8 (426 × 106 m³ month⁻¹) and drain-9 (780 × 106 m³ month⁻¹), carrying agriculture contaminants to this zone.

It was observed that zone-6, which receives agricultural wastes from drain-8 and drain-9, was enriched with higher nutrient concentrations. This could be explained by the peak concentrations for both PO₄-P: 4.81 ± 0.23 mg L⁻¹ and NH₄-N: 3.73 ± 0.81 mg L⁻¹. This finding was consistent with a previous study by Tabinda and Ayub (2010), which indicated that the increase in the percentage of agricultural and urban land cover classes contributes to an increment of nutrient concentrations in lake ecosystems [13].

Silicate (SiO₄)

A spatial variation of SiO₄ reported minimum and maximum values of 1.92 ± 0.10 mg L⁻¹ (zone-5) and 4.55 ± 0.10 mg L⁻¹ (zone-2), respectively (p < 0.01). Zone-5 was not directly connected to drains, and thus recorded low SiO₄ levels. On the contrary, the high concentration of SiO₄ at zone-2 could be related to the amounts of drainage water discharged through the drains; i.e. drain-7 (478 × 106 m³ month⁻¹), Nasser drain, El-Gharbya drain and El-Burullus drain .

Chlorophyll-*a* (Chl-*a*)

The concentration of Chl-*a* significantly fluctuated (p < 0.01) between a minimum value of 0.28 ± 0.06 mg L⁻¹ (zone-1, directly connected to the Mediterranean sea) and a maximum value of 1.44 ± 0.04 mg L⁻¹ (zone-7, in the west region). The highest Chl-*a* concentration was recorded at zone 7, since this zone is connected to Rosetta Branch containing high phytoplankton biomass (floating and submerged aquatic plants). Moreover, high Chl-*a* at zone-7 was possibly associated with the use of phosphorous-based fertilizers within the watersheds (drainage basins) of drain 11 and El-Burullus drain. This Chl-*a* reflected signs of eutrophication at zone-7.

Correlation matrix

Pearson's correlation coefficient between each pair of water quality variables was calculated. As listed in Table 1, statistically significant (p < 0.05) results were obtained for: Temperature and pH (r: 0.650), Temperature and DO (r: -0.543), Temperature and NH₄-N (r: -0.420), pH and NH₄-N (r: -0.542), pH and NO₃-N (r: -0.389), pH and PO₄-P (r: -0.421), Salinity and DO (r: 0.625), Salinity and Chl-*a* (r: -0.745), DO and Chl-*a* (r: -0.837), NH₄-N and NO₂-N (r: 0.446), NH₄-N and NO₃-N (r: 0.848), NH₄-N and PO₄-P (r: 0.909), NO₂-N and NO₃-N (r: 0.638), NO₂-N and PO₄-P (r: 0.439), NO₂-N and SiO₄ (r: 0.679), and NO₃-N and PO₄-P (r: 0.808).

TABLE 1. Correlation matrix of water quality parameters for Burullus Lake

	Temp	pH	Salinity	DO	NH ₄ -N	NO ₂ -N	NO ₃ -N	PO ₄ -P	Chl- <i>a</i>
Temp	1.000								
pH	0.650	1.000							
Salinity	-0.196	0.127	1.000						
DO	-0.543	-0.203	0.625	1.000					
NH ₄ -N	-0.420	-0.542	-0.126	-0.105	1.000				
NO ₂ -N	-0.178	-0.287	0.085	-0.064	0.446	1.000			
NO ₃ -N	-0.280	-0.389	-0.031	-0.170	0.848	0.638	1.000		
PO ₄ -P	-0.136	-0.421	-0.272	-0.267	0.909	0.439	0.808	1.000	
Chl- <i>a</i>	0.303	-0.090	-0.745	-0.837	0.273	0.143	0.251	0.316	1.000
SiO ₄	-0.101	-0.318	0.115	0.185	0.211	0.679	0.319	0.287	-0.045

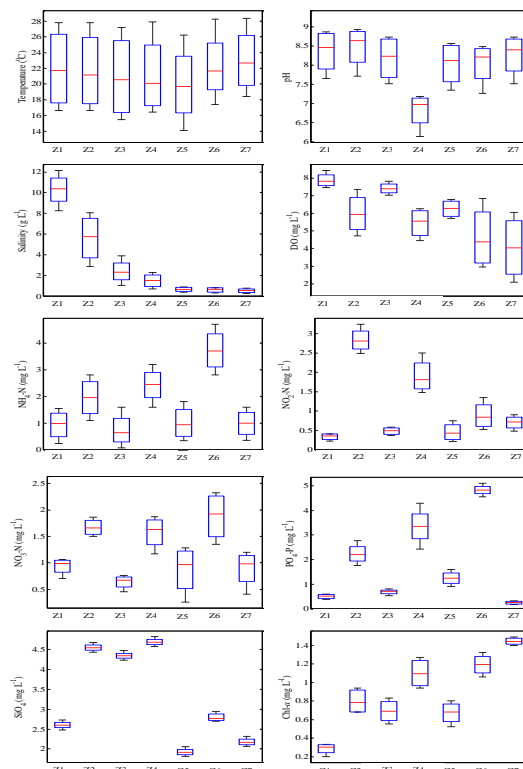


FIGURE 2. Spatial variations of water quality parameters for Burullus Lake using box plots. In each box plot, the central line indicates the median, the box represents the interval 25 – 75% percentiles, and the segment gives the range.

Water quality index (WQI)

The results indicate that no seasonal difference in WQI values can be obviously witnessed. As might be expected, the WQI decreased during the summer season. The relatively lower WQI in summer can be attributed to the decline in DO

levels to $4.91 \pm 1.98 \text{ mg L}^{-1}$; i.e. lower DO levels stipulate an unfavorable impact on WQI.

The WQIs for zones 1, 2, 3, 5 and 7 were found between 30 and 40, which refer to “Bad” water quality type. Zone-1 suffered from organic and nutrients loads coming from El-Burullus drain. This drain carries average waste of $70 \times 106 \text{ m}^3 \text{ month}^{-1}$ of untreated and industrial wastewater. Additionally, the highest salinity was reported in zone-1 due to the mixing of Burullus Lake and Mediterranean Sea. Zone-2 is located near three drains, namely El-Gharbya, Nasser and drain 7. These drains carry about $970 \times 106 \text{ m}^3 \text{ month}^{-1}$ of sewage and agricultural drainage water. Zone-3 is mainly affected by the pollution coming from drain 7 ($480 \times 106 \text{ m}^3 \text{ month}^{-1}$). Additionally, zone-3 is connected to Khalig canal and Nasser drain. Zone-5 and zone-7 are located near Brimbil canal and drain 11, discharging approximately $935 \times 106 \text{ m}^3 \text{ month}^{-1}$ of agricultural discharges.

However, WQI of zones 4 and 6 were ranged between 20 and 30, indicating a “Very bad” water quality condition. This might be due to the discharge of untreated wastewater coming from many small villages and communities distributed along drain side areas, reaching up to $1210 \times 106 \text{ m}^3 \text{ month}^{-1}$. The lower WQI in these zones could potentially increase the risk for health-related problems.

The significant negative correlation between temperature and DO ($r: -0.543, p < 0.05$) indicated that the solubility of oxygen decreased as water temperature increased. Additionally, significant positive correlations were observed between nitrogen and phosphorus species. The explanation for this phenomenon might be based on the excessive agricultural drainage water; i.e. the most important precursor of fertilizers (nutrients) in the lake (Lai et al. 2014). The negative correlation between Salinity and Chl-*a* ($r: -0.745, p < 0.05$) indicated that Chl-*a* decreased with increasing salinity. Similar findings were reported by Hakim et al. (2014). On the contrary, the poor correlation between Chl-*a* and nutrients concentrations could be explained by the rapid utilization of N and P species by phytoplankton [14].

Principal component analysis

In order to evaluate the most significant water quality parameters in Burullus Lake, the data were further analyzed using PCA. The first three principal components (i.e. PC1, PC2, and PC3) affected the quality of the lake, with a total variability of 79.94%. The first principal component (PC1: 38.78% of the variance) represented high loadings on $\text{NH}_4\text{-N}$: 0.46, $\text{NO}_3\text{-N}$: 0.45, and $\text{PO}_4\text{-P}$: 0.45. These positive loadings indicated that PC1 increased with nutrients concentrations coming from agricultural land. It is important to remark that agricultural drainage water was the key variable with the highest influence on water quality characteristics of Burullus Lake. The second principal component (PC2: 28.07% of the variance) was associated with high positive loadings on salinity: 0.45 and DO: 0.55, and with a negative loading on Chl-*a*: -0.50. This indicated that PC2 increased with salinity and DO, while it showed an opposite trend with Chl-*a*. The salinity of the lake imposed a negative effect on chlorophyll growth, and thus Chl-*a* decreased at higher salinity (Hakim et

al. 2014). The third principal component (PC3: 13.09% of the variance) represented high loadings on temperature: 0.45, $\text{NO}_2\text{-N}$: 0.50, and SiO_4 : 0.54. This can be regarded as an indicator of the increase in $\text{NO}_2\text{-N}$ and SiO_4 during summer having a high temperature [15].

Cluster analysis

The samples were divided into two main groups, as follows (Figure 3):

Cluster 1 (containing Z2, Z4 and Z6) is clearly recognizable in the right part of PC1. Since $\text{NH}_4\text{-N}$, $\text{NO}_3\text{-N}$, and $\text{PO}_4\text{-P}$ had stronger positive loadings in PC1, thus these zones were mainly influenced by nutrients concentrations coming from agricultural lands. This could be explained by the catchment run-off processes. Zone-2 is located near a large population community, discharging domestic sewage to El-Gharbya drain and El-Burullus drain. Zone-4 is near drain 7, carrying huge amounts of agricultural drainage water ($478 \times 106 \text{ m}^3 \text{ month}^{-1}$). Zone 6 receives discharges coming from drain 8 ($426 \times 106 \text{ m}^3 \text{ month}^{-1}$) and drain 9 ($780 \times 106 \text{ m}^3 \text{ month}^{-1}$).

Cluster 2 (containing Z1, Z3, Z5 and Z7) was approximately distributed on the negative direction of PC2, with relatively lower nutrients. These zones are located far from the drainage water, and thus they were characterized by lower pollution. Z1-Wi, Z1-Fa, Z1-Sp, and Z1-Su were close together in the up-left corner (Figure 3). This indicated that Zone-1 was affected by high salinity due to the influence of the exchanges with the Mediterranean Sea. However, Z7 was distributed on the negative direction of PC2, indicating that this zone suffered from increased chlorophyll-a concentrations, and lower DO levels. This might be due to the decrease of flow rate coming from Rosetta Branch, leading to poor water circulation. Moreover, lack of hydrodynamic water circulation in the lake leads to the appearance of dead zones, such as zone 7.

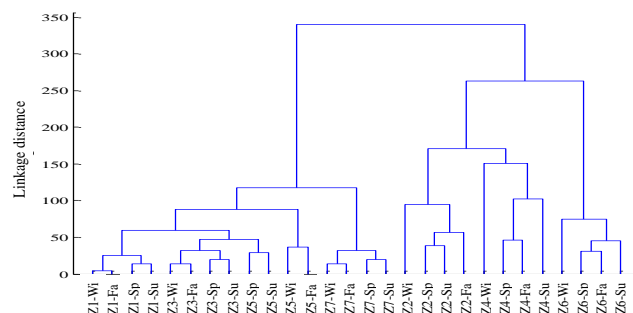


FIGURE 3. Dendrogram showing hierarchical clustering of monitoring seasons/zones.

IV. Conclusion

This study succeeded to develop a large-scale picture of water quality in Burullus Lake. It can be concluded that:

Seasonally, significant ($p < 0.05$) variations were observed in temperature, pH and $PO_4\text{-P}$, however significant ($p < 0.05$) spatial differences were determined in pH, salinity, DO, $NH_4\text{-N}$, $NO_2\text{-N}$, $NO_3\text{-N}$, $PO_4\text{-P}$, Chl-*a*, and SiO_4 . According to PCA results, PC1: 38.78% represented high loadings on $NH_4\text{-N}$: 0.46, $NO_3\text{-N}$: 0.45, and $PO_4\text{-P}$: 0.45, whereas PC2: 28.07% was associated with high positive loadings on salinity: 0.45 and DO: 0.55, and with a negative loading on Chl-*a*: -0.50. The results from cluster analysis and dendrogram revealed that zones 2, 4 and 6 (located in the southern part of the lake) were mainly influenced by nutrients concentrations coming from agricultural drainage water, while zone 7 (located in the western part of the lake) suffered from increased Chl-concentrations, and lower DO levels. Zone-1 was affected by high salinity due to the influence of the exchanges with the Mediterranean Sea, while zones 3 and 5 (situated in the northern part of the lake) were subject to lower pollution.

In conclusion, the lake water was not adequate for drinking or irrigation purposes, however it can be suitable for some aquatic habitats and plants. Thus, it is recommended to: 1) control the discharge of drainage water and sewage discharged into the lake, 2) reduce the usage of fertilizers in agricultural sectors, and/or 3) improve the quality of agricultural drainage water coming to the lake by providing an appropriate treatment method. Further investigations should be conducted to determine the heavy metals contents in the lake; and that will be the focus of our future work

Acknowledgment

The authors would like to thank the staff of Environmental Affairs Agency (EEAA) and National Institute of Oceanography for their important information and support.

References

- [1] Okbah, M., & Hussein, N. (2006) Impact of Environmental Conditions on the Phytoplankton Structure in Mediterranean Sea Lagoon, Lake Burullus, Egypt. *Water, Air, and Soil Pollution*, 172(1), 129–150.
- [2] Solidoro, C., Pastres, R., Cossarini, G., & Ciavatta, S. (2004) Seasonal and spatial variability of water quality parameters in the lagoon of Venice. *Journal of Marine Systems*, 51, 7–18.
- [3] El-Zeiny, A., & El-Kafrawy, S. (2016) Assessment of water pollution induced by human activities in Burullus Lake using Landsat 8 operational land imager and GIS. *The Egyptian Journal of Remote Sensing and Space Science*, In Press, <http://dx.doi.org/10.1016/j.ejrs.2016.10.002>
- [4] WHO (1998) Health Guideline for the use of Wastewater in Agriculture, *Report of WHO Science Group, Technical Report Series 778, 10 pp, Geneva, Switzerland.: World Health Organization.*
- [5] BREABĂN, I., GHETEU, D., & PAIU, M. (2012) Determination of Water Quality Index of Jijia and Miletin Ponds. *Bulletin of the University of Agricultural Sciences & Veterinary*, 69(2), p. 160.
- [6] Ruggieri, N., Castellano, M., Capello, M., Maggi, S., & Povero, P. (2011) Seasonal and spatial variability of water quality parameters in the Port of Genoa, Italy, from 2000 to 2007. *Marine Pollution Bulletin*, 62, 340–349.
- [7] Wu, M., Wang, Y., Dong, J., Sun, C., Wang, Y., Sun, F., & Cheng, H. (2011) Investigation of spatial and temporal trends in water quality in Daya

Bay, South China Sea. *International Journal of Environmental Research and Public Health*, 8(6), 2352–2365.

[8] Noiva, K., Fernández, J., & Wescoat Jr., J. (2016) Cluster analysis of urban water supply and demand: Toward large-scale comparative sustainability planning. *Sustainable Cities and Society*, 27, 484–496.

[9] Nasr, M., & Zahran, H. (2016) Performance evaluation of agricultural drainage water using modeling and statistical approaches. *The Egyptian Journal of Aquatic Research*, 42(2), 141–148.

[10] APHA (1998) Standard Methods for the Examination of Water and Wastewater, 20th ed.. Washington, DC, USA: *American Public Health Association/American Water Works Association/Water Environment Federation.*

[11] Materna, E. (2001) Issue Paper 4, Temperature Interaction, Prepared as Part of EPA Region 10 Temperature Water Quality Criteria Guidance Development Project, *United States: Environmental Protection Agency (EPA).*

[12] House, M., & Ellis, J. (1987) The Development of Water Quality Indices for Operational Management. *Water Science and Technology*, 19(9), 145–154.

[13] Sarbu, C., & Pop, H. (2005) Principal component analysis versus fuzzy principal component analysis A case study: the quality of danube water (1985–1996). *Talanta*, 65, 1215–1220.

[14] Wu, M., Wang, Y., Dong, J., Sun, C., Wang, Y., Sun, F., & Cheng, H. (2011) Investigation of spatial and temporal trends in water quality in Daya Bay, South China Sea. *International Journal of Environmental Research and Public Health*, 8(6), 2352–2365.

[15] Ayers, R., & Westcot, D. (1994) Water Quality for Agriculture, Irrigation and Drainage, *Rome: Paper No. 29. Rev. 1, M-56. ISBN 92-5-102263-1, Rome: Food, Agriculture Organization of the United Nations (FAO).*

About Author :

Diaa Hamdy Seif, PhD Student.
• Master of Science.
*Department of sanitary Engineering,
Faculty of Engineering, Alexandria
University, Egypt.*