Measuring, Modeling Water Quality by Using Sensors and Statistical Analysis Techniques

Iman Morsi[1], Amr El Zawawi[2], Mostafa Amin[3]
[1,3]Faculty of Engineering, Arab Academy for Science and Technology, Alex, Egypt,
drimanmorsi@yahoo.com, m_a_569@hotmail.com
[2]Faculty of Engineering, Alexandria University, Alex, Egypt, E-mail: zawawi@bau.edu.lb

Abstract: Water quality is measured based on the study of several parameters for different water sources. Three different sources of water are studied in Alexandria (Egypt): Mediterranean Sea, Mariot Lake, and mineral water. The measured parameters are: Sodium (Na), Magnesium (Mg), Calcium (Ca), Nickel (Ni), in addition to pH, conductivity and temperature. The measured values are compared with the standard of water parameters given by the World Health Organization (WHO). Different devices and sensors are used in measurements which are:ICP-7500 sequential plasma Spectroscopy SHIMADZU, ASTM D1126-02, ASTM D5708-11, ASTM D4191-08, ASTM D1125-95, ASTM D4191-08, ASTM D1293-99, in addition to colorimetric and potentiometric sensors. The linear regression model based on these parameters is extracted, and the coefficients are calculated by using three statistical analysis methods namely: regstat, roubustfit, and curve fitting. The predicted values of the concentrations from the linear regression models are compared with the measured values, and the root mean square error (RMS) is calculated. The principal component analysis (PCA) method is applied for clustering the three sources of water according to the values of measured parameters. The results indicate that, Mediterranean Sea is recommended as an alternative source. Mariot Lake is contaminated to a high degree.

Keywords: Component Mariot Lake, Mediterranean Sea, Statistical Analysis, Principal Component Analysis, Water Contamination, Spectrometer, Water Sensors

1. Introduction

Lately Egypt has been facing serious threatening in shortage of water from neighboring countries, that share the same source of water namely the river Nile. Eventually, Egypt's share is decreasing gradually which led researchers to study other resources as an alternative to fulfill the demand for drinking water. Nevine et al. [1] have illustrated in their statistical analysis, at four different sites in Alexandria – Egypt, that variation in temperature, pH and dissolved phosphates are insignificant in value; however, the variation in total nitrogen and organic matter is significant.

Conway et al. [2] have introduced the future availability of water in Nile basin in Egypt which is driven by the forces operating on three distinct spatials: river basin, regional, and global. Global and regional driving forces are taken from the results of global climate model experiments. Basiouny et al. [3] have evaluated the total trihalomethane (THMs) at new Benha water treatment plant in Egypt, in terms of initial chlorine dose, total organic carbon, bromide ion, contact time, temperature, algae and pH. Discriminate analysis is widely used due to its important role as a multivariate statistical analysis with different techniques [4-11]. In Alexandria – Egypt, two different sites are studied which are considered as alternative sources. Mediterranean Sea has a surface area of 2.500.000km2, and Mariot Lake is brackish, with an area about 250 km2 in northern Egypt and is located south of Alexandria. Mineral water is taken from already packed bottles in the market and extracted from the oasis in Siwa. In the present paper seven parameters are measured by using different methods. These parameters are Magnesium, Calcium, Nickel, Sodium; in addition to, temperature, conductivity and pH.

It is found that, if these parameters increase or decrease beyond the specific level, they will inversely affect human health and cause many health hazards and diseases [4]. Two methods are introduced in this paper to analyze and evaluate the measured parameters. The first method is a statistical analysis based on predicting a linear regression mathematical model of the measured data. It is constructed between the measured parameters as independent variables, and the total concentration in each source of water. The coefficients of the linear mathematical model are calculated by using three statistical analysis methods namely: regstat, roubustfit, and curve fitting. The second method is the principal component analysis for clustering and distinguishing the three sources of water according to the measured parameters in each source.

2. Description

To detect the level of different parameters, several specimens have been taken from different locations of the three sites from January 2011 to April 2011 as shown in Figure 1.More than 100 measurements for the seven variables are used in the data analysis as a data base.

2.1 Measurements of pH

WTW Inolab terminal levels and ASTM D1293-99 are used [12]. The instrument measurements range from 0 to14 for measuring pH, and the temperature ranges from 0 to 80 oC. A pH measures the electrochemical potential between a Known liquid inside the glass electrode membrane and an unidentified liquid outside. The reference electrode is potassium chloride (KCL) electrolyte. The pH of water must be close to the neutral pH7.

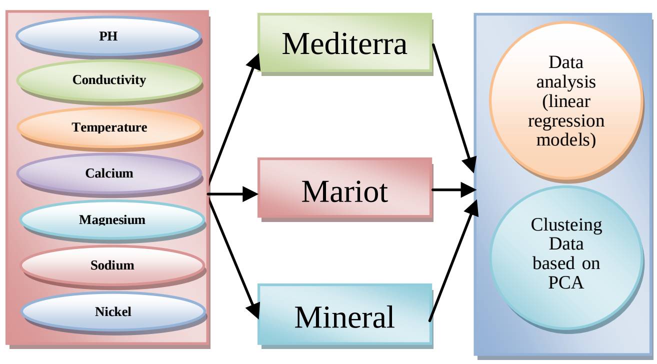

Fig 1- The procedures of the experimental work

2.2. Measurements of Conductivity

TOACM-40v and ASTM D1125-95 are used to measure the electric conductivity based on potentiometric and colorimetric sensors [12]. The instruments measure the conductivity based on two plates in the sample.

A potential is applied across the plates, and the current that passes through the solution is measured. Multiple measurements can be taken to get the whole range of values for different parameters. The measurements using colorimetric sensors are based on either the direct discovery of an analyte via changes in the colorimetric dye alone, or the colorimetric sensor encompasses luminescent and colorimetric dyes in which the luminescent dye is the photon donor, and the colorimetric dye is the photon acceptor. The reagents differ according to the parameter being measured.

2.2 Measurements of Calcium and Magnesium

ASTM D1126-02 is used in measurements based on the EDTA titration method [12]. It is possible to differentiate between the hardness due to Calcium ions, or due to Magnesium ions using the titration method. It is based on changing the color from red to blue, after adding specific reagents with certain concentrations according to the measured parameter. This method is applied in clear water, that is free from chemicals, and contains Calcium or Magnesium. The minimum level of detection is about 2 to 5mg/L as CaCO3, and the maximum level can be extended to all concentrations by sample dilution.

2.4 The Measurements of Nickel and Sodium

ASTM D5708-11 is used to measure the low concentration of Nickel [12], while ASTM D4191-08 is used to measure the low concentration of Sodium [12]. ICP-7500 sequential plasma spectrometer SHIMADZU is used to measure the high concentration of Nickel and Sodium. It depends on the intensity of light emitted from plasma, at a specific wave length, which is used to determine the concentration of the sample to be tested [13].

3. Results

The measured data from the three sources of water have been evaluated and analyzed by using two methods. The Statistical analysis method, and the principal component analysis method. All measurements have to align with the standard of water parameters given by the WHO as shown in TABLE 1. The measured data is collected and averaged over the range of the recorded temperature.

4. The Statistical Analysis Method

The measurements for the three different sources of water are conducted for different parameters. The relation between the different parameters and the recorded temperature are shown from Figures 2 to 7.

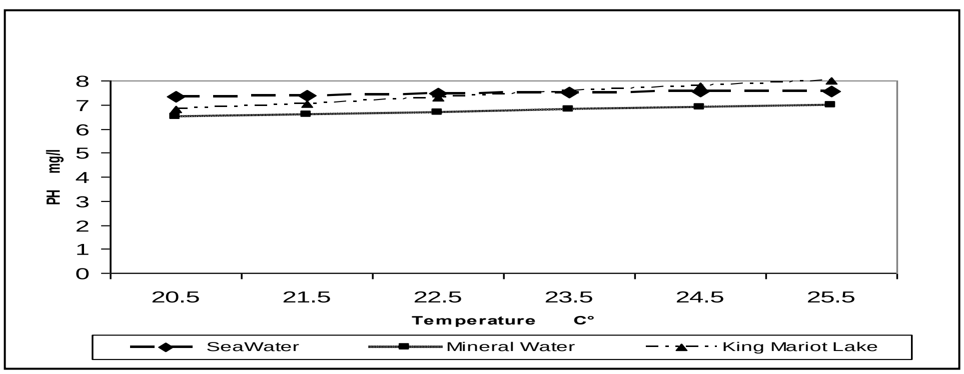

Fig 2- The comparison between average values of pH and temperature among three different sources of water

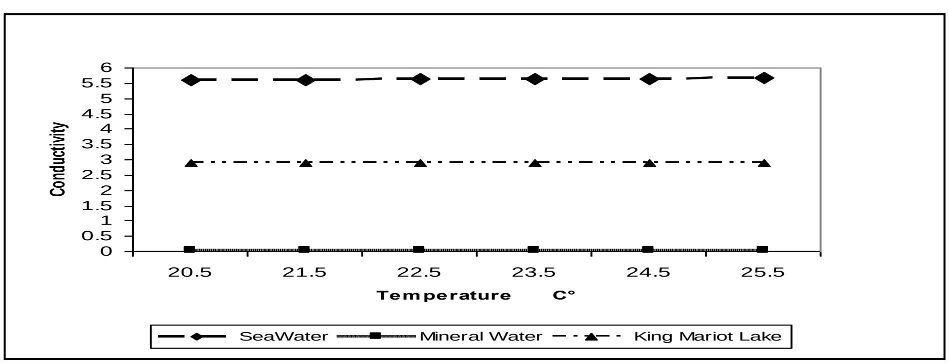

Fig 3- The comparison between average values of conductivity and temperature among three different sources of water

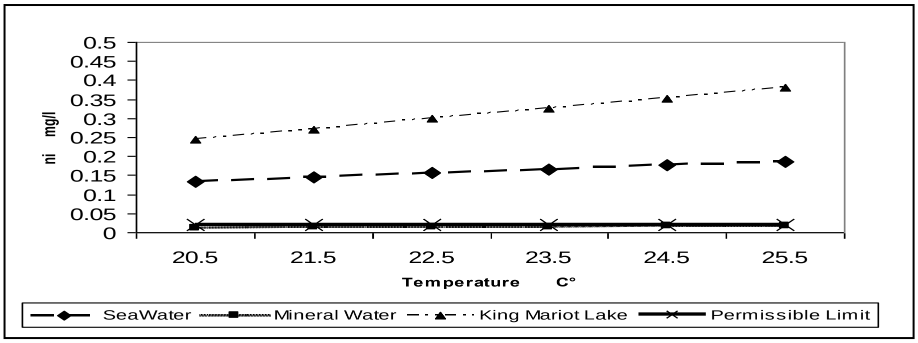

Fig 4- The comparison between average values of Ni and temperature among three different sources of water

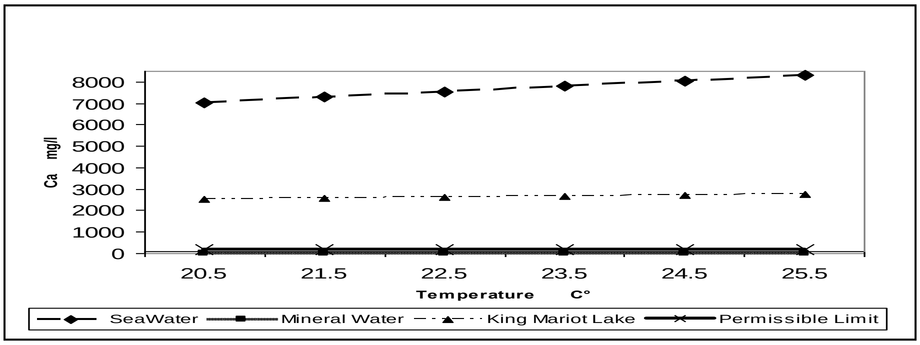

Fig 5- The comparison between average values of Ca and temperature among three different sources of water

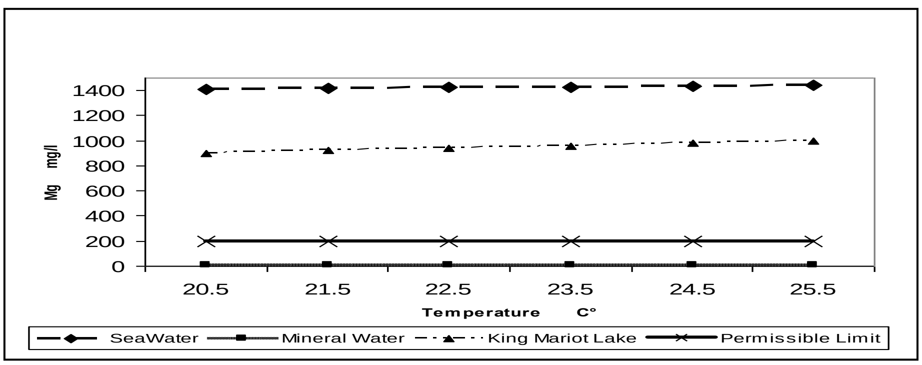

Fig 6- The comparison between average values of Mg and temperature among three different sources, of water

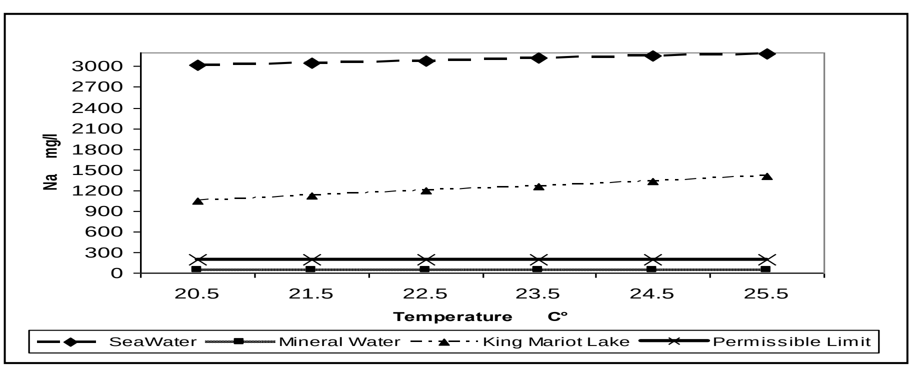

Fig 7- The comparison between average values of Na and temperature among three different sources of water

Table 1- Standard parameters of water in Egypt [4]

Parameters	Standard
PH	6.5 – 9.2
Magnesium (Mg)	150 mg/L
Calcium (Ca)	200 mg/L
Sodium (Na)	200 mg/L
Nickel (Ni)	0.02 mg/L

Table 2-Statistical analysis results of some parameters for the three sources of water

Figure Number	Sources Of water	Parameters measured	Range Of measurements	Average Values $\bar{x}$	Standard deviations (σ)
2	Sea water	pH	7.3 – 7.8	7.5	0.485
	Mineral water		6.5 – 7	6.75	0.371
	Mariot Lake		6.8 – 7.9	7.435	0.352
3	Sea water	Conductivity s/m	5 – 5.9	5.45	0.218
	Mineral water		0.002– 0.0023	0.00215	0.00239
	Mariot Lake		2.5– 2.97	2.73	0.524
4	Sea water	Calcium concentration mg/l	7050– 8050	7550	0.231
	Mineral water		100 – 111	105.5	0.0011
	Mariot Lake		2080– 2085	2082	0.0012
5	Sea water	Magnesium Concentration mg/l	1395- 1490	1424.5	0.00234
	mineral water		55 – 65	60	0.0024
	Mariot Lake		900– 1000	950	0.0156
6	Sea water	Nickel Concentration mg/l	0.12– 0.18	0.15	0.135
	mineral water		0.001– 0.002	0.015	0.001
	Mariot Lake		0.25– 0.38	0.315	0.246
7	sea water	Sodium Concentration mg/l	3000– 3080	3040	0.195
	mineral water		60 – 80	70	0.001
	Mariot Lake		1050– 1250	1150	0.127

Table 2 shows the range of measurements, the average values, and the standard deviations of all parameters. It shows the comparison of all parameters among the three sources of water. From the previous Figure s and results, it can be noticed that, the alkalinity reading is the highest in Mediterranean Sea, followed by Mariot Lake then mineral water.

In addition, the conductivity of Mediterranean Sea is the highest than in Mariot Lake, and the conductivity is close to zero for mineral water. Sodium concentration is measured at the band of 589.592 nm, and Nickel concentration is detected at the band of 231.604 nm by using ICP-7500 sequential plasma spectroscopy. Mariot Lake has high concentration of Nickel.

A linear regression model is used to create the relationship between the measured parameters as shown in equation1 [14, 15].

$$Y = B1+B2\ X1+ B3\ X2 +B4\ X3+ B5\ X4+ B6\ X5+ B7\ X6 + B8X7 \qquad (1)$$

Based on the measured parameters, Y is the predicting concentration of each water source, where as X1 denotes Na, X2 denotes Mg, X3 denotes Ca, X4 denotes Ni, X5 denotes the temperature, X6 denotes the pH, and X7 denotes the conductivity.

The parameter B is called regression coefficient and is calculated by using statistical analysis methods which are: regstat, roubustfit, and curve fitting.

Table 3- Regstat results

	Mediterranean Sea	**Mariot Lake**	**Mineral Water**
Beta Coefficients	B1 = 4.2640	B1 = 3.8672	B1= 0.3159
	B2 = 1.2958	B2 = 0.8619	B2= 0.1139
	B3 = 0.201	B3 = 0.251	B3= 0.0190
	B4 = 0.1782	B4 = 0.7266	B4= 0.0368
	B5 = -1.6509	B5=-1.28081	B5= -2.1036
	B6 = 0.0214	B6 = 0.0185	B6= 0.0132
	B7 =- 0.131	B7 =- 0.115	B7 = -0.105
	B8 = - 0.157	B8 = - 0.123	B8= - 0.114

Table 4- Roubustfit results

	Mediterranean Sea	**Mariot Lake**	**Mineral Water**
Beta Coefficients	B1 = 4.2388	B1 = 3.8405	B1 = 0.3150
	B2 = 1.2724	B2 = 0.7962	B2 = 0.1135
	B3 = 0.1151	B3 = 0.2155	B3 = 0.0190
	B4 = 0.0781	B4 = 0.1260	B4 = 0.0354
	B5 = -1.6349	B5 = -1.179	B5 = -2.106
	B6 = 0.0242	B6 = 0.0177	B6 = 0.0112
	B7 = -0.128	B7 =- 0.112	B7 = -0.99
	B8 = - 0.155	B8 = - 0.129	B8 = - 0.116

Table 5- Curve fitting results

	Mediterranean Sea	**Mariot Lake**	**Mineral Water**
Beta Coefficients	B1 = 3.785	B1 = 2.9861	B1 = 0.2919
	B2 = 1.2423	B2 = 0.742	B2 = 0.992
	B3 = 0.1093	B3 = 0.1956	B3 = 0.0112
	B4 = 0.0645	B4 = 0.114	B4 = 0.0254
	B5 = -1.435	B5 = -0.9879	B5 = -1.915
	B6 = 0.020	B6 = 0.0172	B6 = 0.0107
	B7 = -0.173	B7 = -0.118	B7 =- 0.110
	B8 = - 0.153	B8 = -0.124	B8 = - 0.112

Tables 3, presents the calculation of regression coefficients B using Regstat and Roubustfit, which performs a multilinear regression for different variables that consider “Xi” variables and predict “Y” as the output response of the total concentration. Table V presents the calculation of regression coefficients using curve fitting. Figures 8, 9, 10 show the output results of the predicting model of linear regression analysis based on the three statistical analysis methods. The total concentration in each water source is the y-axis with temperature ranges on the x-axis. Figure 8 shows the comparison among the three statistical analysis methods used to calculate the concentration in sea water. It is noticed that the concentrations are close to each other with respect to the value of RMS error = 0.212. Figure 9 shows the comparison of the three statistical analysis methods in Mariot Lake. It is noticed that the model is achieved with regstat and roubustfit, but the curve fitting has given a different value of concentration with RMS error = 0.834.

Figure 10 shows the comparison among the three statistical functions used in mineral water to achieve the linear model. It shows that the three statistical functions give the same concentration with RMS error = 0.0025.

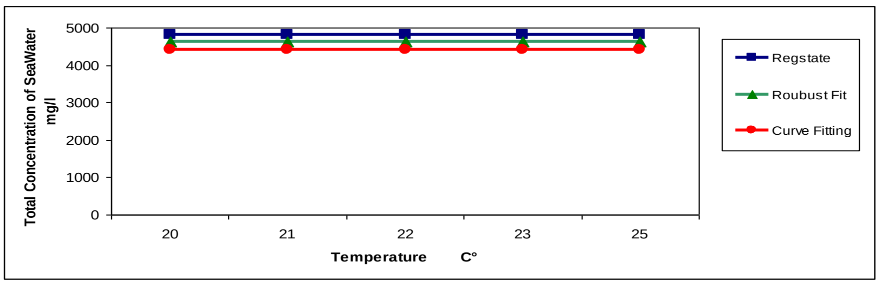

Fig 8-The predicting of the total concentration of sea water parameters with three statistical analysis methods

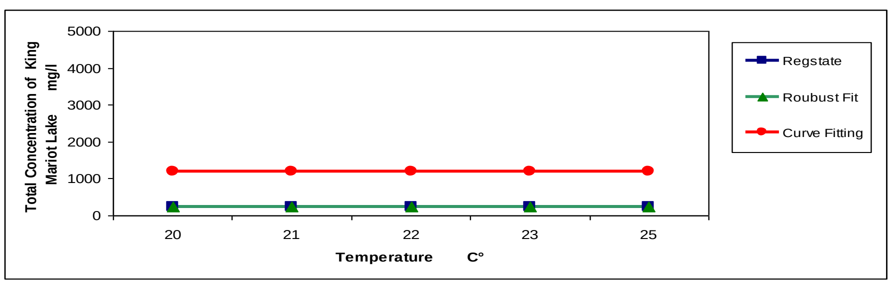

Fig 9- The predicting of the total concentration of Mariot lake parameters with three statistical analysis methods

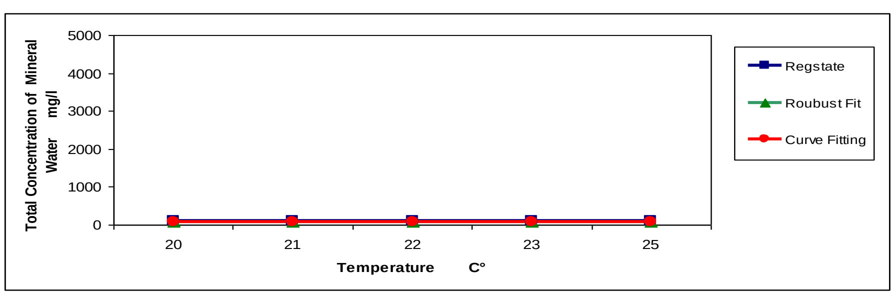

Fig 10- The predicting of the total concentration of mineral water parameters with three statistical analysis methods methods

B. The Principal Component Analysis Method

In this paper PCA is used to design and transform the original variables into new uncorrelated variables, which are linear combinations of the original variables. The new axes lie along the directions of maximum variance [5,14]. In the present case, the original data are expressed for six different parameters namely: Ca, Mg, Na, Ni, pH, and conductivity. The original data of the different parameters are measured with sensors and different devices.

The data is visualized on the two dimensional planes, as illustrated in figures 11, 12 and 13. Different colors indicate different parameters' concentration in each source of water. In Figure s 11, 12 and 13; the concentrations of Ca, Mg, Na, Ni, pH, and conductivity are indicated by numbers from 1 to 6 respectively. The concentration of Ca (No.1) is located on the upper left in sea water, while it is located on the upper right in Mariot Lake and mineral water. The concentration of Mg (No.2) is located in the upper right in sea water and mineral water, while it is located in the upper left in Mariot Lake. The concentration of Na (No.3) is located in the upper right of sea water and upper left in Mariot Lake, and mineral water.

The concentration of Ni (No.4) is located in the lower right in sea water, upper right in Mariot Lake, and lower left in the Mineral water. The pH (No.5) is located in the lower right in sea water and Mariot Lake, while it is located in the lower left in mineral water. The conductivity (No.6) is located to the lower left on the three sources of water. Clustering is based on the different locations in the patterns and different scales which are relative to the different ranges of concentration in each source.

Large scale is found in sea water as shown in figure 11 followed by figure 12 in Mariot Lake, then Figure 13 in mineral water. This is due to high concentration of Ca, Mg, Na in sea water .For this reason, these parameters are located above the axes, followed by Mariot Lake, then mineral water. Figure 12 shows that in Mariot Lake the high concentrations of Ca, Mg , Na, are above the axes ,while Ni indicates high value of concentration relative to the other parametrs.Figure 13 shows that in mineral water the high concentrations of Ca, Mg, Na are above the axes ,which indicate high value of concentrations relative to the other parameters. Distinguishing among different sources of water is due to different locations of the water parameters around the axes of PCA method.

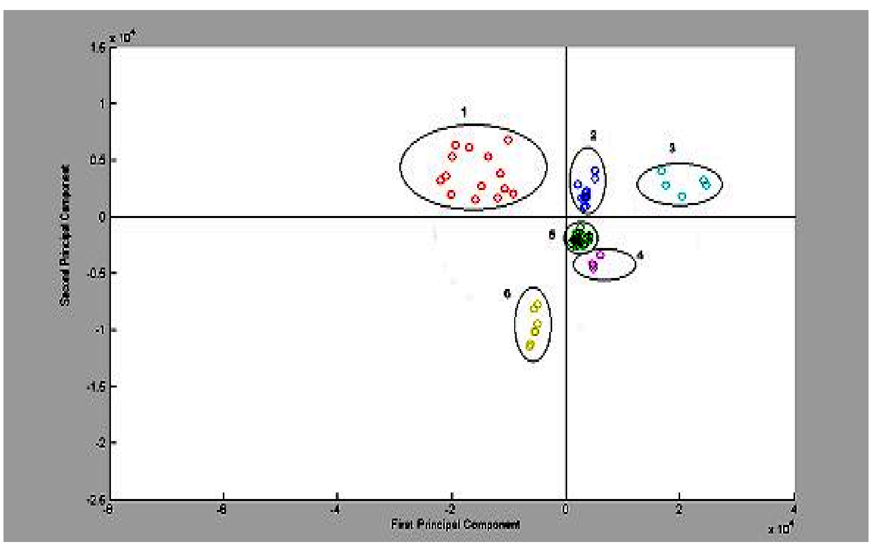

Fig 11-The clustering using PCA for different parameters in sea water
1-Ca 2- Mg 3- Na 4- Ni 5- pH 6- Conductivity

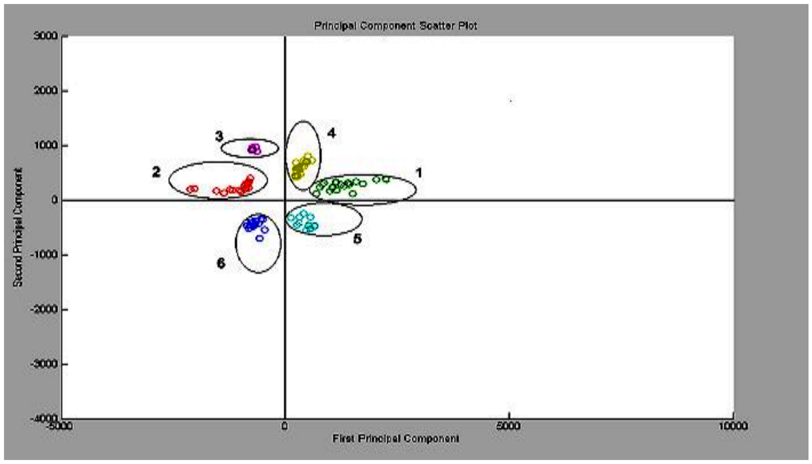

Fig 12- The clustering using PCA for different parameter in Mariot Lake 1-Ca 2- Mg 3- Na 4- Ni 5- pH 6- Conductivity

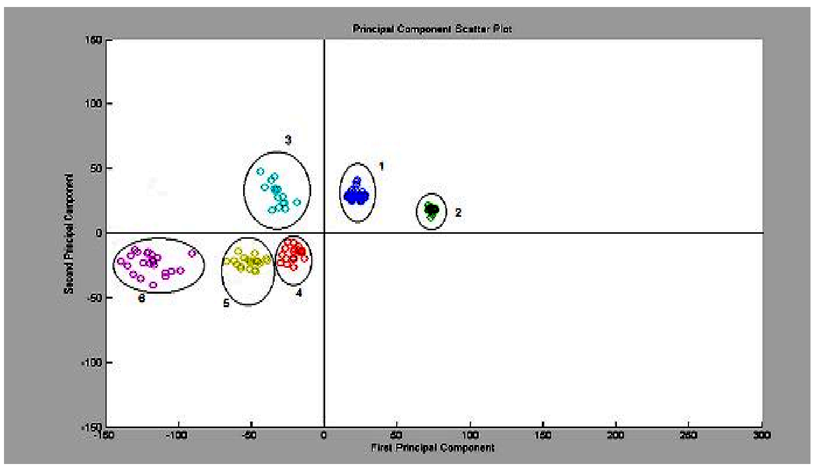

Fig 13- The clustering using PCA for different parameters for Mineral water 1-Ca 2- Mg 3- Na 4- Ni 5- pH 6- Conductivity

Shortage of water is attributed to the inflation of population. Other sources of water have been studied to detect the suitable source, which can be used beside the original source namely Nile river after treatment. Specimens of water have been taken from Mediterranean Sea, Mariot Lake, and mineral water, considering the maximum allowable level of minerals in drinking water defined by the WHO. The measurements are done by using different devices and sensors.

The linear regression mathematical model is predicted, and used to calculate the total concentration of each specimen. Three statistical methods are used: regstat, roubustfit, and curve fitting to calculate the regression coefficient in the predicted linear model. It is concluded that the prediction of the concentration values by using the linear regression model , regstat and roubustfit has given the value of concentration close to the measured value in Sea water and Mariot Lake , according to the value of RMS error. It is worth mentioning, that they have given similar values of concentrations in mineral water. Curve fitting has given the predicted output of concentration similar as the measured values in sea water and mineral water. However, it has given different results in Mariot Lake.

The measurements show that Mediterranean Sea has high concentration of Na, Mg, Ca, while Mariot Lake has high concentration of Ni. This is due to the presence of discharge of debris from factories that negatively influence human health. The linear model also shows that the variation in temperature is insignificant.

The principal component analysis used in clustering parameters in the three sources of water is relative to the concentrations and ranges of difference of each parameter. According to the above measurements and results, Mediterranean Sea is recommended as an alternative source after treatment by using desalination plant. The hazards, in Mariot Lake are very high due to the existence of not only Ni, but also other heavy metals and chemical debris. The result indicates that the water quality of Mariot Lake is contaminated to a high degree, while quality of sea water is better and less contaminated. It is recommended that, Alexandria governorate raises factories' awareness against littering pollutants and chemicals in Mariot Lake and Mediterranean Sea. It is necessary to follow the standards of dispensing emissions.

REFERENCES

[1] N. B. Ghanem, S. Sabry, Z. M. El-Sherif, and G. A. Abu El-Ela, Isolation nd Enumeration of Marine Actinomycetes from Seawater and Sediments in Alexandria, J. Gen. Appl. Microbiol., 46 (2000) 105-111.

[2] D. Conway, M. Krol, J. Alcamo and M. Hulme, Future Availability of Water in Egypt: The Interaction of Global, Regional and Basin Scale Driving Forces in the Nile Basin Water Resources Development, Water Resources Development. 12 (1996) 277-296.

[3] M. Basiouny, T. El Mitwalli, M. Rabee, Formulation and Modeling of Trihalomethane in New Benha Water Treatment Plant, Egypt, 12th International water technology conference, Proc. IWIC12 2008, (2008) 121-139.

[4] L. Fewtrell., J. Bartram, "Water quality: Guidelines, Standards and Health", IWA Publishing, ISBN: 924154533X (WHO), 2001.

[5] R. A. Yotter, L. A. Lee, D. M. Wilson, Sensor Technologies for Monitoring Metabolic Activity in Single Cells – Part1: Optical Methods, Journal of IEEE Sensors , 4 (2004) 395 – 411.

[6] X. Fang, V. K.S. Hsiao, V. P. Chodavarapu, A. H. Titus, A. N. Cartwright, Colorimetric Porous Photonic Band Gap Sensors with Integrated CMOS Color Detectors, Journal of IEEE Sensors, 6 (2006) 661 - 667.

[7] H. Sakai, S. Iiyama, K. Toko, Evaluation of Water Quality and Pollution Using Multi Channel Sensors, Journal of Sensors and Actuators, B66 (2000) 251-255.

[8] J.K. Atkinson, A.W.J. Cranny, W.V. Glasspool, J.A. Michell, An investigation of the Performance Characteristics and Operational Life Times of Multi-Element Thick Film Sensor Arrays Used in the Determination of Water Quality Parameter, Journal of Sensors and Actuators, B54 (1999) 215 - 231.

[9] S. Shresha, F. Kazama, Assessment of Surface Quality Using Multivariate Statistical Techniques: A Case study of the Fuju River Basin, Japan". Journal of Environmental Modelling & Software, 22 (2007) 464 – 475.

[10] L. Lei, W. Qi-Zun, S. Ping, Methods of Discriminate Analysis Classification of Hydrochemical Types of Salt Lakes, Journal of salt Research, 12 (2004) 46-50.

[11] G. A. Van, M. T., Grunwald, S. Bishop, I., Diluzio, M. Srinivasan, A Global Sensitivity Analysis Tool for the Parameters of Multi Variable Catchment Models", Journal of Hydrology, Proc. 324 (2006) 10-23.

[12] Annual Book of ASTM Standards, ASTM International, 2009.

[13] Lopez Moreno, S. Palanco, J. Laserna, Remote Laser – Induced Plasma Spectrometry for Elemental Analysis Samples of Environmental Interest, Journal of Analytical Atomic Spectrometry, 19 (2004) 1479 – 1484.

[14] W. James, F. William, G. Kenneth, Instrumentation for Engineering Measurements, Second ed., John Wiley & Sons. Inc., New York, 1993.

[15] J.B. Gupta, Electronic and Electrical Measurements and Instrumentation, Twelfths ed., S.K. Kataria & Sons , Delhi , 2005.