A multivariate statistical approach to the interpretation of chemical analyses data from the Coastal Environment of Alexandria (Egypt)
1.- Q-Mode Factor Analysis and Partitioning Model by Linear Programing Technique

Tarek A. ABOUL-KASSIM[1,2], Naim M. DOWIDAR[2] and Fatma E. EL-NADY[2]

1- Environmental Organic Geochemistry Group, Oregon State University, College of Oceanography, CORVALLIS (OR) **(U.S.A.)**
2- Oceanography Dept., Faculty of Science, University of ALEXANDRIA **(Egypt)**

During the period of 1985-1986, we have done a comprehensive research project to examine and assess the impact of sewage disposal on the chemistry of a heavily polluted South-Eastern Mediterranean basin, the Eastern Harbor (E.H) of Alexandria, Egypt. During the first phase of our research project, the main objectives were to investigate the consequences of sewage discharge on the water quality (ABOUL-KASSIM and DOWIDAR, 1990a,b), to construct the cycles of carbon (ABOUL-KASSIM, 1987; ABDEL-MOATI *et al.*, 1991), nitrogen (EL-NADY *et al.*, 1990) and phosphorus (DOWIDAR *et al.*, 1990), to fractionate the living biomass components of the harbor using adenosine triphosphate method (ABOUL-KASSIM *et al.*, 1992) and finally to study the composition of suspended matter (ABOUL-KASSIM, 1987). Our goal in the second phase of the project is to do a complete univariate and multivariate statistical analyses, just to reduce the big data set, to evaluate the pathways of pollutant transport, the regions of concentrations, and to identify some statistically significant end members; well representing the study area; and finally to conceptually computer model the environment of Alexandria.

When using the multivariate statistics approach for interpreting a chemical analyses data (ABOUL-KASSIM, 1992), there are three main goals when studying the distribution of pollution tracers in any sewage-impacted coastal environment. The first objective is to determine the number of significant systems which are contributing to the observed sample compositions. A system is defined as a source of the group of compounds studied (in our case; N, P & C species as well as some environmental parameters). One method of studying the nutrient composition of the marine environment is to inspect the concentrations of that fractions of all samples in an area in order to determine the number of systems present (Visual Pattern Recognition). A system is easily discerned if one or more samples contain a high proportion of that system and if the system's pattern is very different from the patterns of the other systems. Many times these conditions are not met because most samples are mixtures of many systems and it would be fertuitous that one sample would contain only one system. The second objective is to determine the chemical composition of the pure end member (the composition of a hypothetical sample containing 100% of one system). A major problem with the visual method is how far to extrapolate beyond the compositions of the sample richest in one system. The third objective is to determine the distribution of the systems in each sample. It is difficult to visually estimate how much of each system is in each sample except in gneral terms such as small, moderate or large amounts. So, in this paper, we will use **Q-Mode Factor Analysis** and **Partitioning by Linear Programing Technique** (LPT) just to:
1- identify specific end members in the marine environment of Alexandria harbor,
2- obtain the composition of the pure end members from factor scores,
3- construct the distribution of the systems within each sample from factor loadings.

Within this statistical approach, it is possible to assign the percentage contribution from various sources to the observed distribution of all chemical parameters in each sample.

Factor analysis has been recentally used in Environmental Chemistry (ABOUL-KASSIM, 1992), and very few studies have been done in the Mediterranean region (EL-SAYED *et al.*, 1988; FRIHY and GAMAI, 1991; GARDNER *et al.*, 1990). Q-Mode factor analysis technique is based on grouping a multivariate data set based on the data structure defined by the similarity between samples. By applying this fancy technique to our big data set, three significant principal factor loading scores were obtained, giving information about the sample variation of about 89.1%, 4.96% and 3.31%, respectively (maximum cumulative information of 97.37%).

The factor loading matrix which represent the sample representation in the model and indicated that the importance of each of the factors or end members in each sample (i.e. the better the model the closer the communality of the sample will be to 1.0). After varimax rotation of the composition scores, three significant end members were resulted, dominating by variables as salinity (S%.), total nitrogen (TN) and total suspended matter (TSM), respectively. A modification of Q-Mode factor analysis provides an objective means for end member routine. Factor analysis has not often been used to determine the actual composition of end member sources, because the transfer of the original data variables during the analysis result in negative factor scores for some variables and negative concentrations of some variables in the end members. Because of the orthogonality constraints, one vector must have negative contributions from one of the variables. This can be solved by using a non-orthogonal rotation of end member vectors, toward the mean vector to bring the end members into the positive vector space. To solve the partitioning model, a non-orthogonal end members were used by using a rotation called **oblique rotation**. By using the LPT, a set of equations were applied so as to correct the initial end member compositions and their abundance in samples so as to better fit the observed multivariate data set. This will help specify and select the composition of the end members, so that they are very close the true end members composition. Figure 1 shows the end member after partitioning by LPT. The importance of **salinity** as end member comes from the continuous dilution of the harbor water with low salinity domestic sewage as well as the continuous mixing of the harbor water with neritic Mediterranean water. These three significant end members are well describing the data set, indicating the important part played by sewage disposed to the bay with its important portions of TN and TSM (Figure 1).

An important conclusion concerning the application of factor analysis in our chemical analyses data is that the large amount of data is reduced to a very few systems "end members". This significant end members will be usefull to construct an environmental model in the near future work. The method of factor analysis is not limited to marine water samples and could be used to study rocks, gases as well as biological samples.

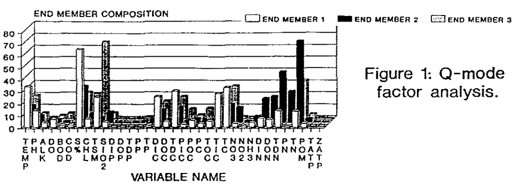

Figure 1: Q-mode factor analysis.

REFERENCES

ABDEL-MOATI A.R., ABOUL-KASSIM T.A., DOWIDAR N.M. and EL-NADY F.E., 1991.- *J. Coastal Research*, 7(2): 377 -385.
ABOUL-KASSIM T.A.,1987.- M.Sc. Thesis, Oceanography Dept., Fac. of Sci., Alexandria University, Egypt.
ABOUL-KASSIM T.A., 1992.- In preparation.
ABOUL-KASSIM T.A. and DOWIDAR N.M., 1990a.- *Rapp.Comm.int.Mer Médit.*,32 (1, L-III3), 125 p.
ABOUL-KASSIM T.A. and DOWIDAR N.M., 1990b.- *Rapp.Comm.int.MerMédit.*, 32 (1, L-III2), 125 p.
ABOUL-KASSIM T.A., DOWIDAR N.M., EL-NADY F.E. and ABDEL-MOATI A.R., 1992.- Sci. of the Total Environ. (in press).
DOWIDAR N.M., ABDEL-MOATI A.R., ABOUL-KASSIM T.A. and EL-NADY F.E., 1990.- *Rapp.Comm.int. MerMédit.*, 32.
EL-NADY, F.E., ABOUL-KASSIM T.A., DOWIDAR N.M. and ABDEL-MOATI A.R., 1990.- *Rapp.Comm.int.Mer Médit.*, 32.
EL-SAYED M.Kh., EL-WAKEEL S.K. and RIFFAAT A.E., 1988.- *Oceanol. Acta.*, 11(1), 1-11.
FRIHY O.E. and GAMAI I.H., 1991.- *Netherlands Journal of Sea Research*, 27 (2), 165-171.
GARDNER J.V., DEAN W.E. and ALONSO B.,1990.- *Mar. Geo.*, 95(3-4), 225-245.

A univariate statistical approach to the interpretation of chemical analyses data from the coastal environment of Alexandria (Egypt)
2- Analysis of Variance (ANOVA)

Tarek A. ABOUL-KASSIM[1,2], Naim M. DOWIDAR[2] and Fatma E. EL-NADY[2]

1- Environmental Organic Geochemistry Group, Oregon State University, College of CORVALLIS **(U.S.A.)**
2- Oceanography Dept., Faculty of Science, Alexandria University, Moharam Bey, ALEXANDRIA **(Egypt)**

Alexandria is the principal summer resort of Egypt. It is one of the relatively densel populated regions of the Eastern Mediterranean (about 3.5 million inhabitants receiving about 1 million tourists in summer who come to use its beaches fc recreation. Alexandria coastal waters are highly polluted with untreated domest sewage and waste waters, discharged into the sea through several outfalls along tł coast, thus presenting a serious source of pollution in that region. In 1985-1986, research project has been done to assess the impact of sewage disposal on the wat quality of the Eastern Harbor of Alexandria. A necessary condition for any efficie water pollution study is the dominance of the inter sample concentration chang over the intra sampling site dispersion. The repeated sampling of the water netwo performed in the present study affords in principle a statistical differentiation betwee these two major causes of variation. The concentration changes can be studied terms of environmentally significant factors such as sewage influence, distance fro the source points, water depth and time of collection.

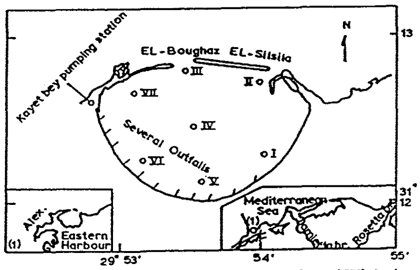

Fig. 1. Stations sampled in the Eastern Harbour (EH) during the period 1985–1986.

The study area (Eastern Harbor) is a shallow semi-closed basin (Figure 1), sheltere from the sea by an artificial break water leaving two openings through which th exchange of water between the harbor and the neritic Mediterranean water takes plac Our data set consists of 256 samples (cases) with 37 environmental parameter (variables) studied.

In this paper, we will use a powerful univariate statistical technique as ANOVA t analyze the inter/intra sample variation. The technique was performed to measur the significant difference between:
1- different stations,
2- different months, and
3- different water levels.

Based on the T-test (0.05) and F-test (0.05) values, the significant variables ONL have entered the program as dependent variables.

ANOVA between stations

The variables which recorded significant differences between stations are:
a- Transparency:
Stations III & IV (located near the entrance of the unpolluted Mediterranean wate recorded significant difference from stations VII & II as well as I, V, VI (located close t sewage disposal).
b- Salinity:
Significant difference in salinity occurred between stations (III & VII) and (I, V, VI).
c- Chlorophyll a:
Stations (V & I) recorded significant difference with (II, III, IV, VII).
d- Particulate organic carbon (POC):
Stations V varies with station IV, VII, III and station III differs from I, VI, V.
e- Nitrate:
Station I recorded significant difference with the other 6 stations.
f- Particulate organic matter (POM):
Stations I differs from III, IV and VII, while station V recorded significant differenc from stations III, IV.

ANOVA between different depths

Significant variations between surface (S) and bottom (B) water levels occurre between the following variables: O_2, S%, CHL A, DIP, DOC, POC, NO_3, DOP, PP, DO and PN were recorded. Most of the cases recorded insignificant difference betwee bottom (B) and middle (M) water levels.

ANOVA between different months

The following variables recorded significant differences between months: O_2, S% CHL, DIP, DOC, POC, NO_3, NO_2, DOP, PP, DON, DIC, PIC, NH_3, DON and PN. Month of the warm season including May, June, August and September recorde insignificant differences between themselves and significant differences with the oth months.

Conclusion

The significant differences recorded between stations located close to sewage outfal (I, V & VI) and the other stations are due to the effect of the domestic sewage dispos and its variability within the different stations. The significant difference betwee surface and bottom/middle water levels is due to the effect of mixing between the lo density sewage water with the surface water of the bay. The difference betwee summer season and the rest of the other months is coinciding with periods of hig sewage discharge rate in Alexandria city.