

The significance of outlying P-PO₄ values in eutrophication studies

Michael KARYDIS* and Lydia IGNATIADIS**

*Department of Environmental Studies, University of the Aegean, MYTILINI (Greece)

**Institute of Biology, NRC "Demokritos", ATTIKI (Greece)

The existence of "extreme", "doughtful" or "anomalous" values in collected data of nutrients from the natural environment have been characterized as "outliers" or outlying values (BARNETT and LEWIS, 1987). Outliers are often seen as reducing and distorting the information provided by the data and in this case they might be removed (EVERITT, 1981). On the other hand, outliers might characterize extreme conditions of an ecosystem associated with pollution problems and thus, their presence should be taken into account in any statistical analysis concerning eutrophication assessment and coastal management.

In the present investigation outliers in nutrient (P-PO₄) data sets were identified and their spatial and temporal distribution was examined. Sampling was carried out monthly from 9 stations of Saronikos Gulf, Aegean Sea during 1981 (Fig.1, A). The net of these stations covered an area characterized by strong nutrient gradients (KARYDIS *et al.*, 1983) due to the influence of the sewage outfall. Outliers in the P-PO₄ data sets were detected by the method of "Box and Whisker Plot" described by TUKEY (1977) and OTT (1988). Data points lying outside 1.5 times the interquartile range of the upper quartile were recorded as outliers.

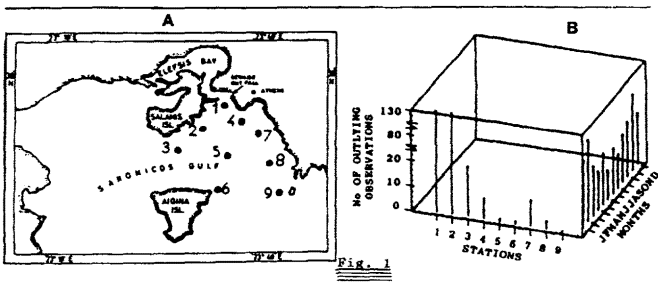


Table 1. Distribution characteristics of the raw P-PO₄ data sets.

| Data set | N | x | s | range |
|-------------------------------------|-----|------|------|------------|
| A. Before the exclusion of outliers | 942 | 0.34 | 0.67 | 0.01-13.72 |
| B. After the exclusion of outliers | 697 | 0.12 | 0.08 | 0.01- 0.33 |
| C. Outlier's set | 245 | 0.96 | 1.10 | 0.34-13.72 |

The results on the distribution characteristics of the parent data sets of P-PO₄ are given in Table 1. It is seen that the original number of the raw data (N=942) was reduced (N=697) after the exclusion of outliers (N=245). The raw data and the outliers were further classified per station in an attempt to evaluate the contribution of each station in the entire set of values before and after the exclusion of outliers. The results showed (Table 2) a well defined gradient in the distribution of the number of outliers and their mean values which seems to be related with the vicinity of the corresponding stations to the sewage outfall. For example, the highest scores of outlying values (129) and their mean (1.29 µg-at P.l⁻¹) were recorded at station S1 which was located near the sewage outfall and has been characterised as strongly eutrophic (IGNATIADIS *et al.*, in press) and they diminished gradually along with the distance of each station from it. Also, at station S1, the outlying values comprised the 75% of the total recorded observations from this station whereas this percentage diminished again along with the distance of each station from the sewage outfall. A graphical presentation of the distribution of outliers at the 9 stations is given in Fig. 1B, along with their seasonal distribution. It is seen that seasonality might also play an important role in contributing outlying values in the P-PO₄ annual data.

Table 2. Summary statistics of the raw and outlying values of P-PO₄ data sets per station. (N=total No of observ., x=means, s=stand. dev.)

| | Raw data (r) | | | Outliers (o) | | | % Outliers over Raw data | |
|----|--------------|------|------|--------------|------|------|--------------------------|------|
| | N(r) | x(r) | s(o) | N(o) | x(o) | s(o) | N(%) | x(%) |
| S1 | 173 | 1.00 | 1.29 | 129 | 1.29 | 1.38 | 75 | 77 |
| S2 | 174 | 0.35 | 0.30 | 74 | 0.60 | 0.31 | 43 | 58 |
| S3 | 174 | 0.18 | 0.20 | 19 | 0.54 | 0.41 | 11 | 33 |
| S4 | 72 | 0.17 | 0.41 | 8 | 0.81 | 1.02 | 11 | 21 |
| S5 | 71 | 0.12 | 0.09 | 1 | 0.41 | - | 1 | 29 |
| S6 | 72 | 0.10 | 0.07 | 1 | 0.40 | - | 1 | 25 |
| S7 | 68 | 0.16 | 0.19 | 10 | 0.55 | 0.21 | 15 | 29 |
| S8 | 69 | 0.12 | 0.14 | 3 | 0.68 | 0.22 | 4 | 18 |
| S9 | 69 | 0.09 | 0.04 | 0 | - | - | 0 | 27 |

The results of this work indicate the possibility of using outliers as an index of eutrophication and applying statistical tests among outlying data as a means for detecting differences concerning nutrient loadings in the marine environment.

This work is part of Project supported by a FAO/UNEP grant (GRE/68-H) of the MED POL program.

REFERENCES

- BARNETT V. and LEWIS T., 1984.- Outliers in statistical data. Second Edition, J.Wiley and Sons, Chichester, 463 p.
 EVERITT B., 1981.- Cluster analysis. Second Edition, J. Wiley and Sons, 136 p.
 IGNATIADIS L., KARYDIS M. and VOUNATSOU P. - A possible method for evaluating oligotrophy and eutrophication based on nutrient concentration scales. *Mar. Poll. Bull.* (in press).
 KARYDIS M., IGNATIADIS L. and MOSCHOPOYLOY N., 1983.- An index associated with nutrient eutrophication in the marine environment. *Estuar. Coast. Shelf Sci.*, 16: 339-344.
 OTT L., 1988.- *An introduction to statistical methods and data analysis*. Third Edition, PWS-KENT Publ. Co., Boston, 835 p.
 TUKEY J.W., 1977.- *Exploratory data analysis*. Addison-Wesley Publ. Co. 668 p.