A simple statistical method for elucidation of the structure of the sea

by

A. CRUZADO

Instituto de Investigaciones Pesqueras, Barcelona (Spain)

Abstract*

In a recent communication to the Working Conference on Analysis of Upwelling Systems (*Investigación Pesquera* **35** (1) 261-267 (1971)) I introduced the idea of sliding statistics as a means for studying the structure of the sea when continuous physical and chemical analysis has been performed. In addition to the « sliding average » normally employed in « curve smoothing » I proposed the use of the « sliding correlation coefficient ».

The correlation coefficient is widely and indiscriminately used by many biologists in the study of relationships between physical, chemical and biological parameters. The correlation coefficient is normally defined as an index of the *association* between variables although often this interpretation has little meaning. A more useful definition would be : *the degree of closeness of a set of points to a given predictor.* In the case of the linear correlation coefficient it is the degree of linearity of the relation between two variables.

Moreover this definition fits better with the computational procedure normally used, in its determination. If $Y = a + b X + E$ is the linear predictor with Y being the response variable, X the independent variable and E the residuals (variable with zero mean and normal distribution), minimizing the variance of E gives the regression coefficients *a* and *b* and also the linear correlation coefficient $R = b\ S_X/S_Y$. This procedure, also called *least squares technique*, is of a wide use in predictive work.

In oceanography we deal with a very complex structure whose complexity appears to increase when the frequency of observation increases; thus, no simple predictor can be used ina general way. However if sufficiently small space or time intervals are studied, simple relationships (almost deterministic ones) are found permitting, at least, successful short term predictions. We can take a small interval (with a sufficiently high number of observations to assure statistical significance) and see how does the linearity of the relationship between two variables change as the sampling system moves along a given path.

In a general way the correlation coefficient can be treated as a function of two variables $R = F (t, h)$, *t* being the time of sampling and *h* the time interval. The three dimensional surface $R (t, h)$ shows the field of validity of our predictor. Strongly mixed waters show strong linear correlation coefficient in both physical, chemical and biological parameters whereas stable waters show more complicated features and sudden changes in the correlation coefficient sign indicate sharp boundaries in both water characteristics and phytoplankton distribution.

* Le texte *in extenso* de cette communication a paru *in* · *Invest. pesqu.*, **35**, 1, pp. 261-267 (1971).

Rapp. Comm. int. Mer Médit., **21**, 7, p. 353 (1973).