Not to be cited without prior reference to the authors.

ICES CM 1996/S:23

Individual based statistics for a spatially distributed population, with an application on mackerel.

by

N. Bez , J. Rivoirard and M. Walsh

Abstract

When studying statistically the spatial distribution of a fish versus environmental parameters (e.g. temperature), fish density at a point is often considered as a response to the parameter at this location (regression techniques). However, since for instance favorable temperatures can extend outside the area of presence of fish, regressions depend on the domain chosen for their computation.

Another approach can be used, which starts from the individuals of the population. As fish are all the more numerous at a location that the density is larger, the densities give the probabilities by which statistics are to be weighted. As examples:

- the mean location is given by the center of gravity of coordinates, with the associated inertia as variance;
- the distribution of temperature for individuals is given by the relative abundance of fish (sum of densities) within each temperature class. It can be summarizes by the mean (center of gravity) and the variance (inertia) of temperatures per individual.

This individual based approach is illustrated on the mackerel eggs distribution observed during the ICES triennal surveys.

Keywords: Spatial distributions, individual based statistics, preferendum.

Nicolas Bez and Jacques Rivoirard : *Centre de Géostatistique, Ecole des Mines de Paris, 35 Rue Saint-Honoré, 77305 Fontainebleau, France [tel: +1 64 69 47 00, fax: +1 64 69 47 05, e-mail: bez@cg.ensmp.fr and rivoi@cg.ensmp.fr].* ***Martin Walsh :*** *Marine Laboratory, PO Box 101, Aberdeen, AB9 8DB, UK [tel: +44 1224 87 65 44, fax: +44 1224 29 55 11, e-mail: walsh@31H.dnet.marlab.ac.uk]*

1 Introduction

In most cases, geostatistics is used to describe a regionalized variable (e.g. fish density) within a domain. Some stationarity hypotheses (of the variable, of its increments, etc...) are usually required to describe the variable statistically, enabling sophisticated estimations. When studying a spatially distributed abundance of fish a more basic approach can be used, which is not based upon the definition of a domain and a stationarity hypothesis for the variable and which is appropriate when there are few high peaks of density, and a lot of very small values.

This approach, described in the first two parts, is related to the transitive geostatistics and to the covariogram (Matheron, 1971). It involves the entire population and its interpretation gets a biological meaning as it calls upon statistics per individual. The problem of the estimation of the main tools can be found in Bez *et al.* (1996).

2 Presentation of the data

The dataset chosen for the illustration of this paper comes from the ICES triennal "mackerel and horse mackerel egg production" surveys. The sampling is designed to cover the production of eggs in space and time. It is based on a regular grid 0.5° longitude × 0.5° latitude (about 15 x 30 nautical miles (n.m.)). Some additional samples, and some difficulties at sea cause irregularities in the final sampling (Fig. 1). Each spawning season is divided into 4 or 5 periods. This leads to 23 irregular 2D datasets (from 1980 to 1995).

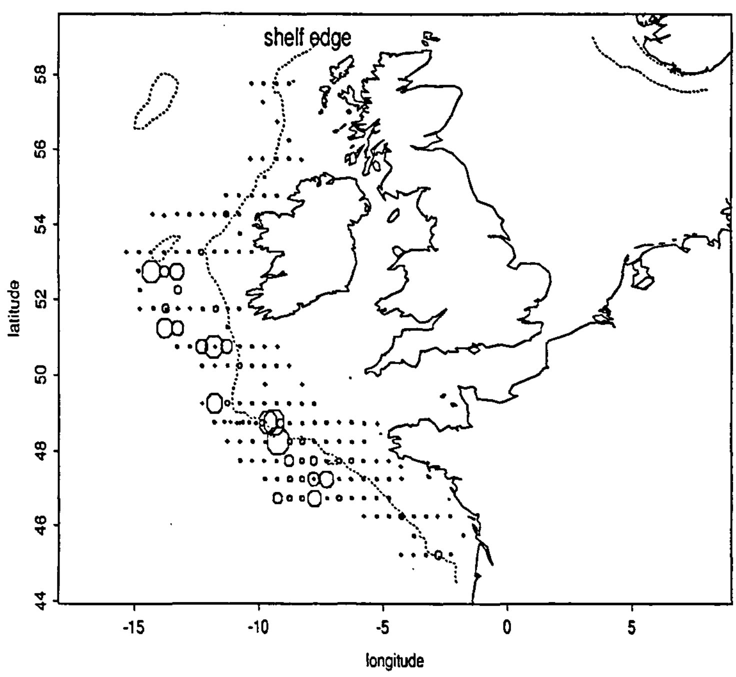

Figure 1: Proportional representation of the egg density. 1995 period 2.

3 Basic tools : statistics per individual

3.1 Density, abundance and distribution of a random individual

The fish density taken as a regionalized variable, positive or zero, will be denoted by $z(x)$. For instance x is a point in 2D. The sum of $z(x)$ over space gives the abundance $Q = \int z(x)dx$.

Individuals are all the more numerous at a location when the density at this location is larger. So if we consider an individual I taken at random, the probability density function of its location x_I is the relative density $z'(x) = z(x)/Q$, which sums to one.

Consider now another regionalized variable, e.g. an environmental variable such as sea surface temperature ($t(x)$). Then the distribution of temperature for a random individual (denoted t_I) is obtained by the distribution of temperatures weighted by the relative densities. As summing the densities within a class of temperature gives the abundance of that class, the probability that an individual belongs to a class of temperature is given by the relative abundance of that class.

3.2 Center of gravity

The average position of the population can be described by the center of gravity of locations:

$$E(x_I) = \frac{\int xz(x)dx}{\int z(x)dx} = \int xz'(x)dx$$

which is also the mean location of individuals (or the expected location of an individual taken at random).

In the same manner we can define the center of gravity for any other variable, e.g. the mean temperature of individuals :

$$E(t_I) = \frac{\int t(x)z(x)dx}{\int z(x)dx} = \int t(x)z'(x)dx$$

3.3 Inertia

Since all individuals are not located at the center, the statistical dispersion of their locations can be described by their inertia:

$$var(x_I) = \frac{\int (x - E(x_I))^2 z(x)dx}{\int z(x)dx} = \int (x - E(x_I))^2 z'(x)dx$$

which is also the variance of the locations of individuals.

This inertia, in turn, can be split into orthogonal directions in space. The first one is the direction which explains the maximum of inertia. The second one is orthogonal (in 3D it is the orthogonal direction carrying the maximum of the residual inertia, etc). This decomposition is nothing but the result of a Principal Component Analysis on the coordinates, weighted by the density $z(x)$ (i.e. the coordinates of individuals).

Inertia can also be defined for other variables, giving for instance the variance of the temperatures of the individuals.

3.4 Application: description of a serie of surveys

The center of gravity of the location of eggs for each sampling period is represented by the number of the period with a size all the larger that the abundance of the period is large (Fig. 2). We can observe that the peak of spawning (largest size number) has shifted slightly to the west from 1983 to 1995. This is not due to a shift of the sampling as was checked by computing the center of gravity of the sampling pattern (unweighted mean location of samples).

The variance (inertia) of location is displayed by 2 orthogonal segments centered on the center of gravity with directions the principal axes and with lengths twice the square root of the inertia they explain. The major part of inertia is carried by the first axis, more or less parallel to the shelf edge, for instance as in 1986 (Fig. 2).

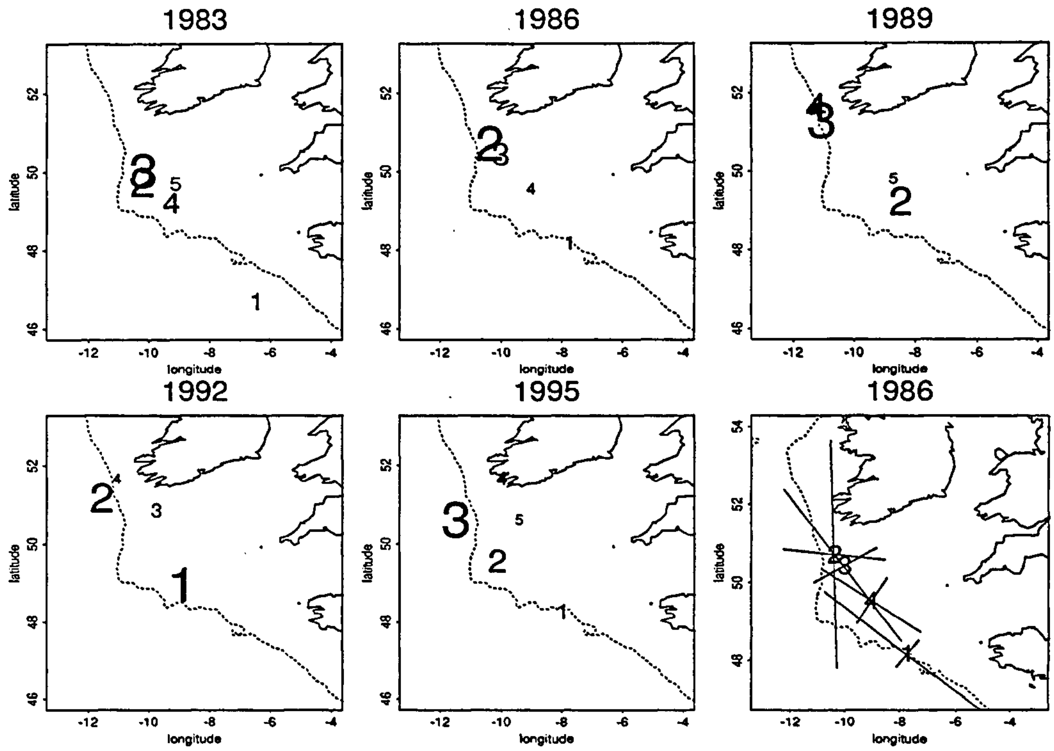

Figure 2: Centre of gravity of the location of eggs per period and per year with a size all the larger that the abundance of the period is large. Inertia of the location of eggs in 1986 split into principal axes.

3.5 Center of gravity for density or mean density of individuals

A direct mean of densities would refer to a domain, or would be zero if we consider the whole infinite space. This problem disappears by taking the sum of densities, that is the abundance. Then a way to measure the dispersion of densities could be $\int z(x)^2 dx$ (which happens to be the covariogram at distance zero).

It may be convenient to divide this by Q. As $z(x)/Q$ is the probability distribution of a random individual,

$$E(z_I) = \frac{\int z(x)^2 dx}{\int z(x) dx} = \int z(x) z'(x) dx$$

is the center of gravity of density, also the mean density of individuals.

From this we can define an equivalent surface, that is, the surface which would give the same abundance with a density constant and equal to the above mean density. It is easy to see that the inverse of this surface is nothing but $\int z'(x)^2 dx$.

Besides the mean, it is also possible to compute the inertia or variance of the density for individuals :

$$var(z_I) = \frac{\int (z(x) - E(z_I))^2 z(x)dx}{\int z(x)dx} = \int (z(x) - E(z_I))^2 z'(x)dx$$

and to deduce a coefficient of variation for the density of individuals.

Note that the mean and variance of density for individuals are different from the mean and variance of densities which could be computed directly within a domain. Generally the distribution of $z(x)$ (the usual histogram) is skew, with a large coefficient of variation. Considering the distribution of density per individual, the mean density is larger, with a lower coefficient of variation, since individuals are stacked within the high densities.

3.6 Application: statistics per mackerel egg

For each such regular 2D dataset basic, statistics are presented versus the date given by the mean julian day of each sampling period (Fig. 3). The abundance of eggs is clearly bell-shaped with a maximum production generally at the end of May (julian day=150) or one month earlier in 1992. Coefficients of variation of density per individual are lower than 1 with no clear tendancy. Except in 1983, the mean density per individual fluctuates around 450 eggs/m^2 up to the end of may and decreases afterwards. In the mean time the equivalent surface regularly increase from april to may and drops afterwards. This leads to interpret the raise of egg production essentially by a geographic extension of the spawning area (geometrical effect).

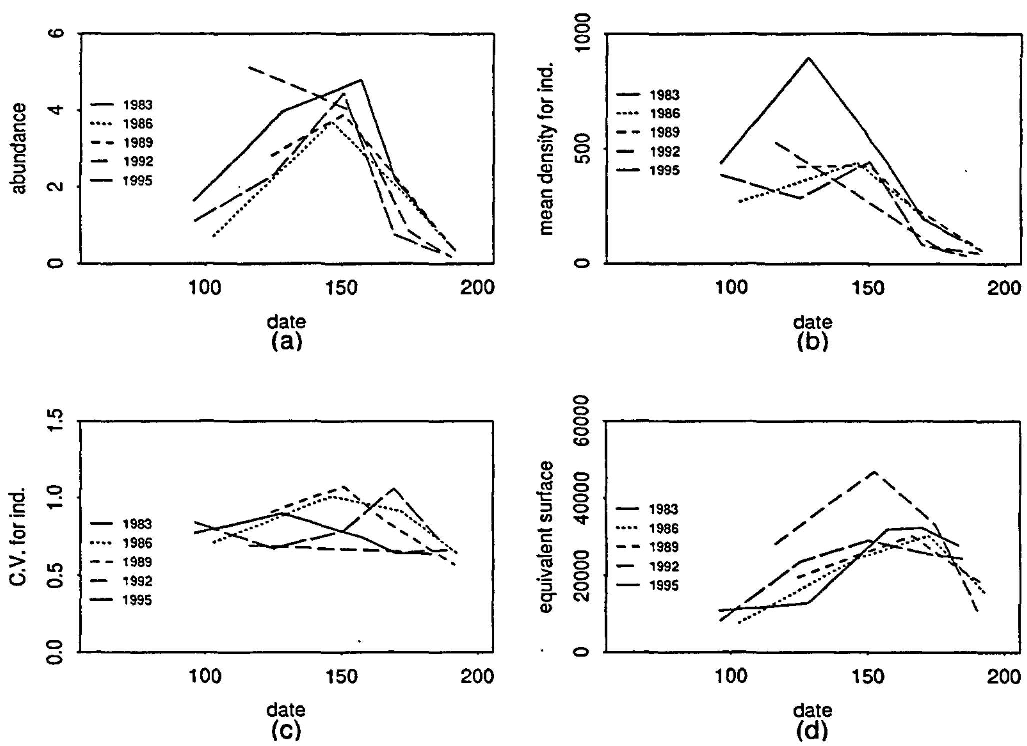

Figure 3: Analysis of a series of surveys. Statistics per period versus the date: (a) abundance (in 10^{13} eggs), (b) mean density per individual (eggs/m^2), (c) coefficient of variation of density per individual and (d) equivalent surface ($n.m.^2$).

In 1989, a finer analysis has been made looking at the early life stages of mackerel (eggs and larvae). The center of gravity and inertia of sea surface temperature per individual has been computed for the stage I and V eggs and plotted versus the date (Fig. 4). Over the spawning season the mean temperature which increases from 12 to 16°C is the same for the two stages. Nevertheless their inertia is significantly different at the beginning of the season. This indicates that the survival eggs are more concentrated around a given temperature than the spwan eggs.

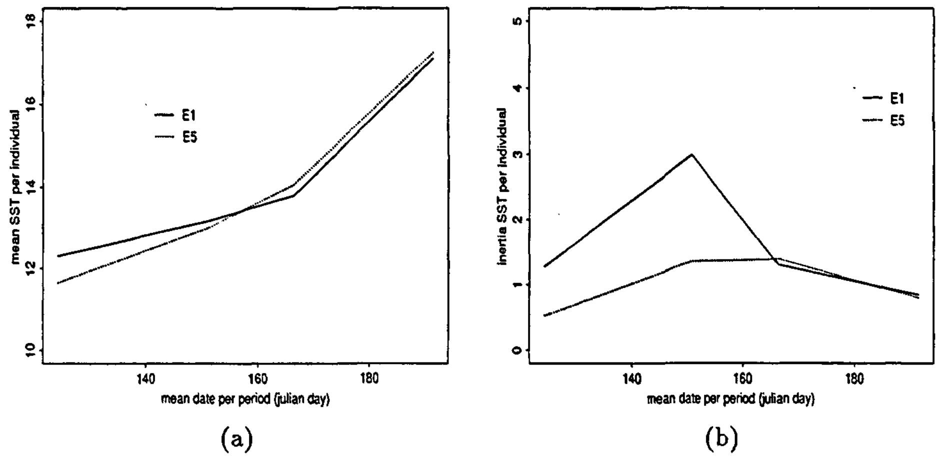

Figure 4: Sea surface temperature per individual versus the date: (a) center of gravity (b) inertia. Comparison of the stage I eggs and the stage V eggs of mackerel (1989)

4 Covariogram

4.1 Definition

A spatial description of $z(x)$ is given by its (transitive) covariogram:

$$g(h) = \int z(x)z(x+h)dx$$

The covariogram is symmetrical in h ($g(-h) = g(h)$) and maximum at the origin ($g(h) \leq g(0)$) and sums to the square of the abundance $\int g(h)dh = Q^2$. In each direction it is zero beyond a certain distance called the range giving the extension of the population in this direction. Its behavior at short distances in terms of continuity and derivability is related to the degree of spatial regularity of the regionalized variable.

4.2 Distance between individuals

More interestingly, the relative covariogram $g(h)/Q^2$ has a meaning in term of individuals within a population. If we take independently two individuals at random in the global population, the distance between them is a (vectorial) random variable, and its probability density function is the relative covariogram (Matheron, 1971, exercise 16 p.47).

4.3 Application: covariogram of mackerel egg density

In the case of the mackerel eggs distribution, the drop from the distance 0 which can be interpreted as a nugget effect is very large and more or less the same whatever the direction (Fig. 5), indicating a highly irregular variable isotropic at short distances. The range is different according to the direction which means that the distribution is anisotropic at long distances.

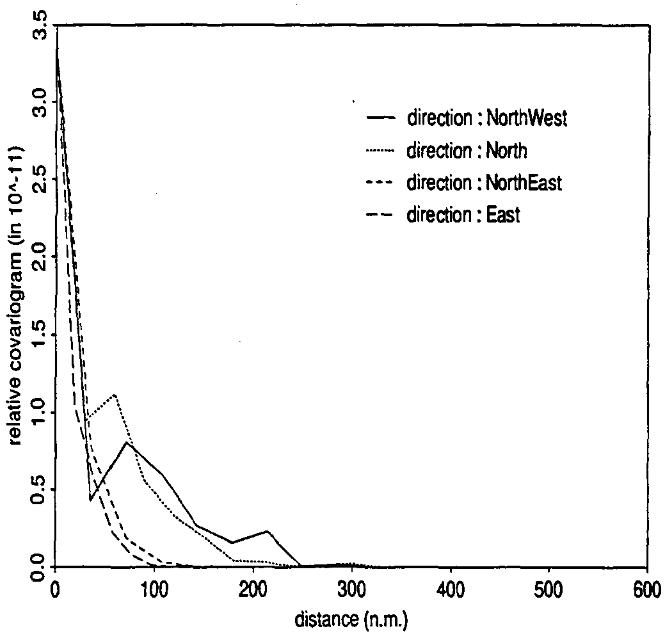

Figure 5: Experimental relative covariogram. Four directions. Mackerel eggs

5 Conclusion

In this paper we have emphasized the advantages of using sums (e.g. abundances) rather than means (averages) when analysing regionalized fish density. This does not require the definition of a domain, is not sensitive to the zeroes and the smallest values and does not necessitate any stationarity hypothesis. It leads to the probability distribution, mean (center of gravity) or variance (inertia) of any characteristic (location, density, environmental parameter) for an individual taken at random. Considering two random individuals leads to the description of spatial structure through the covariogram.

Acknowledgements: these developments have been made within the european program AIR 93 1105 SEFOS (Shelf Edge Fisheries and Oceanographic Study). Its aim is to describe and to link the spatial and temporal variability of some commercial european species with hydrographical parameters. Authors thank Lowestoft Institute (MAFF) for the data they provided.

References

[1] Anon., 1993. Report of the mackerel/horse mackerel egg production workshop. Aberdeen, 8-12 March 1993, ICES CM1993/H:4, 142 p.

[2] Bez N., J. Rivoirard, J.C. Poulard, 1995. Approche transitive et densités de poissons. Compte-rendu des journées de Géostatistique, 15-16 juin 1995 Fontainebleau, France, *Cahiers de Géostatistique*, **5**, 161-177 pp.

[3] Bez N., J. Rivoirard, Ph. Guiblin, and M. Walsh, 1996. Covariogram and related tools for structural analysis of fish survey data. Fifth International Geostatistics Congress. (*in press*).

[4] Matheron G., 1971. The theory of regionalized variables and its applications. *Les cahiers du Centre de Morphologie Mathématique de Fontainebleau*, n°**5**, ENSMP, Fontainebleau, France, 211 p.

[5] Matheron G, 1989. Estimating and choosing - An essay on probability in practice. Ed. Springler-Verlag, 141 p.