The use of geostatistics for abundance estimation by echo integration in lakes: the example of Lake Annecy

Jean Guillard, Daniel Gerdeaux, Jean-Marc Chautru

Rapp. P.-v. Réun. Cons. int. Explor. Mer, 189: 410–414. 1990

The use of geostatistics for abundance estimation by echo integration in lakes: the example of Lake Annecy

Jean Guillard, Daniel Gerdeaux, and
Jean-Marc Chautru

Guillard, Jean, Gerdeaux, Daniel, and Chautru, Jean-Marc. 1990. The use of geostatistics for abundance estimation by echo integration in lakes: the example of Lake Annecy. – Rapp. P.-v. Réun. Cons. int. Explor. Mer, 189: 410–414.

Geostatistical calculations were applied to the results of an echo-integration survey on Lake Annecy, France, in order to determine estimates of fish biomass. Initial sampling along transects, which was systematic but not exhaustive, did not allow the satisfactory adjustment of variograms. A special sampling method was necessary. The survey gave an estimate of biomass with a confidence limit of ±20 %. An optimal sampling procedure in lakes is proposed within the framework of geostatistical calculations.

Jean Guillard and Daniel Gerdeaux: Laboratoire d'Hydrobiologie Lacustre, INRA, BP 511, 74203 Thonon-les-bains, France. Jean-Marc Chautru: Centre de Géostatistique, 35 rue S' Honoré, 77305 Fontainebleau, France.

1. Introduction

Estimation of fish biomass in lakes allows us to expand our knowledge of the ichthyological resources of such environments and, above all, enables us to follow stock fluctuations over a period of time (Forbes and Nakken, 1972; Marchal, 1985). One must not only obtain an estimate, but also establish a confidence limit. This problem has been addressed frequently in recent work (Gohin, 1985; Laloe, 1985; Burczynski and Johnson, 1986). An approach using classical statistical methods often results in very large confidence limits owing to wide sampling variance (Williamson, 1982; Francis, 1984). Those methods which take into account the spatial distribution of sampling locations should result in estimates with smaller confidence limits, except in rare cases where no spatial correlation exists. Geostatistics allows us to analyse this type of data and to determine an unbiased estimator (Matheron, 1971; Journel and Huijbregts, 1978; Delhomme, 1979).

2. Materials and methods

2.1. Materials

Lake Annecy is situated in the northern region of the French Alps at an elevation of 446.5 metres. It is 13.7 km long with a surface area of 27 km^2 and consists of two basins separated by a narrow strait. Its maximum depth is 65 m. The detectable biomass consists essentially of six species: perch (*Perca fluviatilis*), roach (*Rutilus rutilus*), coregonids (*Coregonus lavaretus*), trout (*Salmo trutta*), char (*Salvelinus alpinus*), and pike (*Esox lucius*). The echo-sounding survey was conducted at night with a boat 8 m long, sailing at a constant speed of 10 km/h. The acoustic equipment used consisted of a Simrad EY-M sounder (70 kHz, TVG = 20 log R) fitted with a transducer with a full beam angle of 11°. The signal received, converted into frequency, and the trigger pulse were digitally recorded on tape. Echo integration was performed by an AGENOR echo integrator (Person *et al.*, 1982).

2.2. Sampling method

The survey was systematically run in sections over the entire surface of the lake using a classical procedure (Marchal and Laurent, 1977). Between each transect, the boat followed the shore for about six minutes. So, the distance between any two transects was never exactly the same, and we can estimate that this distance is randomly around 500 m. We consider that we made a stratification of the lake in a regular grid and place each transect randomly along the grid knots. This grid was arbitrarily chosen as a function of the time to carry out a complete survey in one night (Fig. 1).

A copper ball 32 mm in diameter, attached by three

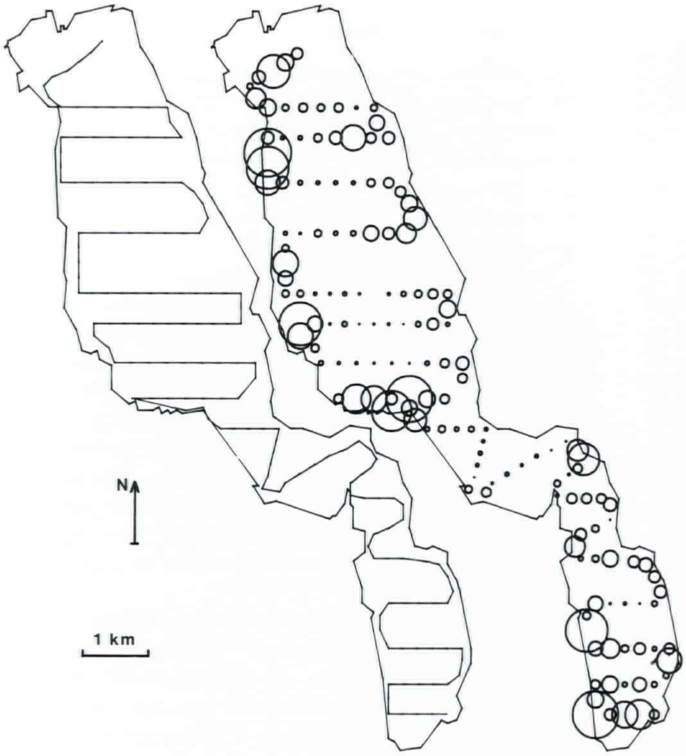

Figure 1. Echo-integration results of Lake Annecy. Left: location of hydroacoustics transects. Right: plots of echo-integration values, for two-minute sequences. Each circle is proportional to the echo-integration units for the layer 0–25 m.

wires at a distance of 7 m from the transducer was used to calibrate the measuring equipment.

The duration of echo-integration sequences was first established at two minutes for eight 5-m layers of water (from the surface to 40-m depth), and two 10-m layers (from 40 to 60 m). To facilitate data processing the integration layers were combined into two depth strata (0–25 and 25–60 m). The study objectives did not require us to develop biomass estimates. Therefore results were expressed in echo-integration units ($V^2 \times m^{-2}$). Later on, the duration of sequences was decreased to 30, then to 10 seconds, to fit the variograms and for kriging.

2.3. Overview of the theory of regionalized variables

The calculations were carried out within the probabilistic framework of the theory of regionalized variables (Matheron, 1971), where each experimental datum is considered as one realization of a random variable.

Our objective was to estimate the average value of any additive variable (e.g., biomass density) in a given field, starting from a linear combination of samples. In each case it is necessary, in practice, to adjust a model starting with the experimental data. Since the variogram is the tool currently used in this field, it is useful to define its most important properties. The variogram $\gamma(h)$ is defined by the expression:

$\gamma(h) = 0.5\ E\ \{[z(x+h)-z(x)]^2\}$ where z(x) is the observed value at point x (Fig. 2).

The sill and range only exist in the case where the field studied is sufficiently homogeneous. If not, one will be

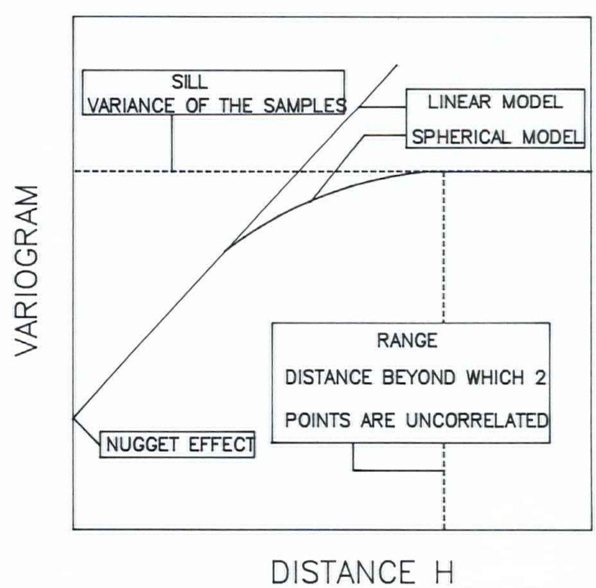

Figure 2. The variogram: definition of terms.

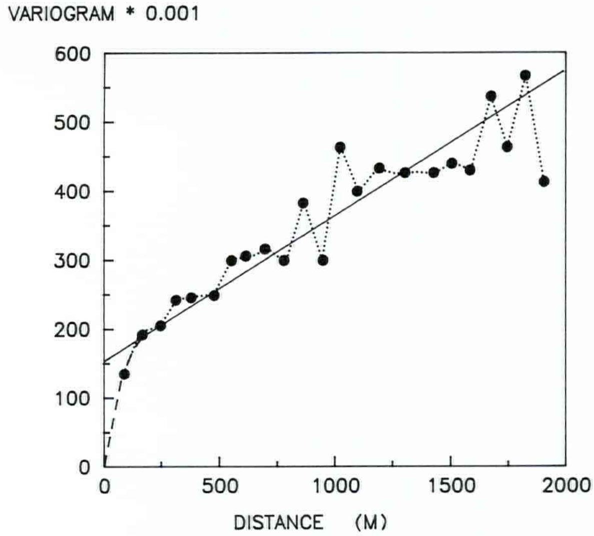

Figure 3. East−west variogram, with 30-second sequences. Possible models: $\gamma(H)$: 150 000 × (1.5 × (H/150) − 0.5 × × $(H/150)^3$ + 210. $\gamma(H)$: 150 000 + 210 × H.

able to observe a continuous growth of the variogram (the linear shape of Figure 2).

It should be noted that the variogram is direction-dependent. Different behaviour in various directions therefore demonstrates and quantifies the anisotropy of the phenomenon studied. Because of this, one must take care not to treat as a whole, any samples which are regularized using different criteria (supports), in particular on long segments in different directions. The regularized function Zv(x) with a support v is the average over the volume v of the random function Z(x) defined on a point support. Thus, to avoid serious incoherences, the echo-integration sequences along the east−west transects (sections) should be studied separately from those along the cross-transects, approximately north−south.

For more information, the reader may refer to Matheron (1971), David (1977), and Journel and Huijbregts (1978).

3. Results

Plotting of the echo-integration results shows the heterogeneity of the spatial distribution of the fish (Fig. 1). The coastal areas contain the major portion of the biomass. Note the existence of a gradient of decreasing density from the shore towards the pelagic zone. The majority of the fish are located at the thermocline, 12 to 15 m deep. Observation of these gradients demonstrates the spatial correlations between samples.

3.1. Variogram study

The calculations were carried out separately for each of the two parts of the lake, the "upper lake" and the "lower lake". Each of these zones is assumed to be homogeneous, in other words, having a unique structure. Since only the "upper lake" was sufficiently sampled to allow precise calculations, the following results are applicable to this zone only.

Starting from echo-integration sequences of two minutes' duration (63 locations), it was impossible to fit the variogram, and thus to determine a structure logically. The size of each sequence (about 330 m) perhaps places too much emphasis on the regularization of the phenomenon, masking the structures. Therefore, we had to increase the number of sampling locations and work on echo-integration sequences of 30 seconds. The number of samples for the transects of the "upper lake" was thus 202; there were, accordingly, enough pairs of data points in the principal east−west direction to fit the variogram (Fig. 3).

The arrangement of points corresponding to the first steps of the variogram leads to uncertainty regarding the ordinate at the origin: was it zero or not?; in other words, was there or was there not a nugget effect? We finally decreased the duration of sequences to 10 seconds to obtain enough points at short distances. With this additional information, we were able to select a theoretical linear model which permitted adjustment of the variogram in the principal east−west direction and which incorporated a fairly big nugget effect (Fig. 4).

As we have already pointed out, the nugget effect includes both inevitable measurement errors (precision of the apparatus, stability of adjustments, etc.), and

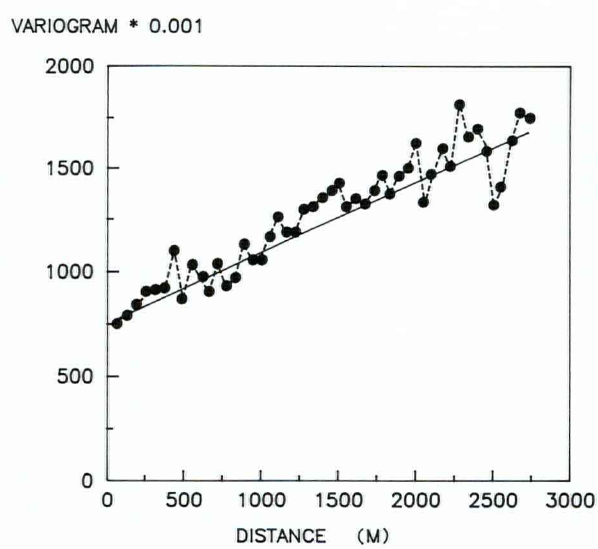

Figure 4. East−west variogram, with 10-second sequences. Possible models: $\gamma(H)$: 750 000 + 340 × H.

microstructures which are not detectable on the scale of the study. In our particular case, we can say that the behaviour of a fish depended more on its immediate neighbours than on those 30 metres away. The primary spatial structure must therefore be on the order of a metre. Because our scale of study was in tens of metres, this structure would not be detectable, and it was not surprising to notice a significant nugget effect. The nugget effect entered into the calculations as a purely random component.

It should be noted that the particular case of the nugget effect alone (the variogram still equal to its sill) corresponds to the independence of the random variables introduced at each point. The more significant the nugget effect, the closer we get to conditions where classical statistical methods apply.

Incidentally, because of the heterogeneity and small dimensions of the field studied, it is not surprising to observe a linear variogram, i.e., without a sill or range.

One can see, moreover, in Figure 1 some anisotropy of the fish biomass, with a north–south elongation, which would be useful to quantify. To do this, we have to calculate and fit the variogram in many directions. Two types of data are available: on the one hand, the samples along the east–west transects theoretically allow us to obtain a discontinuous north–south variogram, but the transects are too far apart to enable us to adjust it appropriately in the vicinity of the origin. On the other hand, the samples of the cross-transects allow us to calculate the regularized north–south variogram which is comparable to that in the east–west direction in Figure 3. However, because there are so few samples and they are concentrated toward the edge of the lake, the points on the variogram have little statistical significance and there is a risk of bias. One can nevetheless

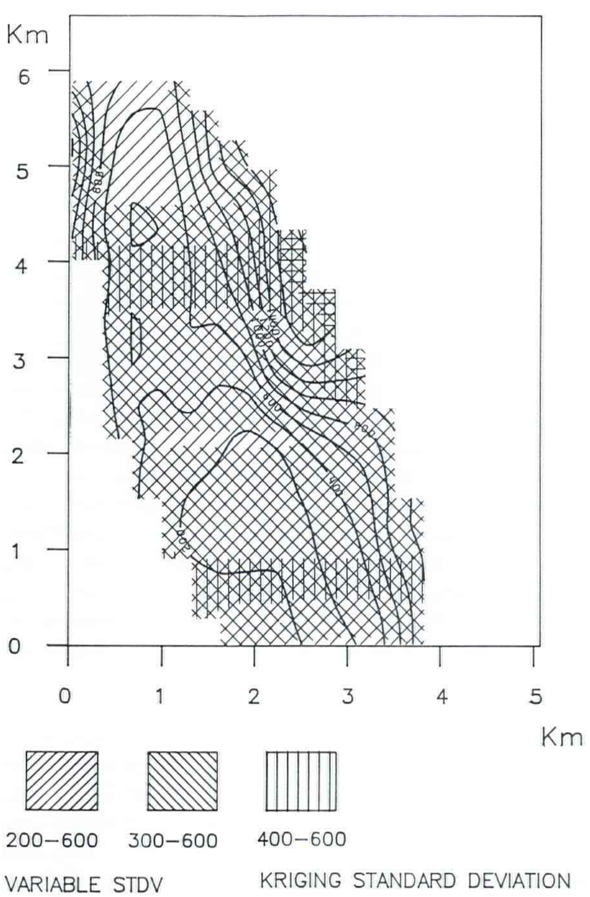

Figure 6. Contour map for kriging biomass densities (estimation in echo-integration units) in the "upper lake". Superposition (hachures) of standard deviation for local estimation.

perform the calculation and adjust the variogram based on the nugget effect and a linear model (Fig. 5).

By comparing the observed slopes in Figures 3 and 5, after standardization by the variance, we can estimate the order of magnitude of the coefficient of anisotropy to be about 2.5. Since this value is compatible with natural observations, we will hypothesize that it is correct and use it in subsequent calculations.

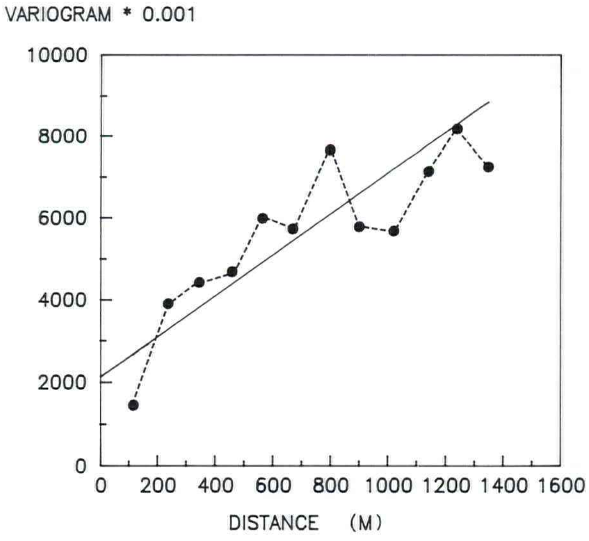

Figure 5. North–south variogram, with 30-second sequences. Possible models: $\gamma(H)$: 2 100 000 + 5 000 × H.

3.2. Estimation by kriging

The variogram models chosen (Fig. 3) were introduced into the kriging calculations (computer program BLUEPACK 3D, Centre de Géostatistique, Fontainebleau). We observed an estimator of the average biomass (in units of echo integration) for the whole of each zone sampled and for each link in the network.

With each estimation, there is an associated estimation variance which allows us to define a confidence limit for each result obtained (Fig. 6); for example, the average of the density in the "upper lake" is

$659 \pm 2 \times 62$. The estimation of total stock is made by multiplying the average value for each zone by the surface area (similarly for the confidence limit).

4. Discussion

An estimate of the biomass present in the lake of ± 18 % has been obtained. If we could obtain an even narrower confidence interval, it would make our comparison tests all the more powerful.

Calculation of the data and adjustment of the variogram must be performed with enough points to avoid contradicting the theoretical model. In our particular case, we have had to establish certain hypotheses, since the configuration of our sampling plan did not take everything into consideration. Our first hypothesis was the structural homogeneity of the chosen units. These zones were selected based on our knowledge of the environment (bathymetry, fish behaviour, etc.), and experience of local professional and sports fishermen. The sampling routine, following different directions, does not allow us to take into account simultaneously all of the sequences. To mitigate this problem, the sequences must be sufficiently short that the regularization effect is negligible. The samples can therefore be considered as points on the scale of the lake in all directions and can be studied together as a whole. Another solution would consist of sampling in only one direction, east–west for example, but bringing the transects as close together as possible and keeping the sequences relatively short, yet not necessarily treating them as point values.

In general, the problems encountered during this study could be eliminated, or at least greatly reduced, if the following sampling procedure were followed. Samples should at first be taken along closely spaced transects in order to obtain an accurate variogram. These transects could begin at the shore of the lake and end in the middle; there should be on the order of 10 to have enough points in the direction perpendicular to the transects, and the distance between the two courses which are farthest apart should be greater than the assumed range of the variogram. One should try within reason to duplicate this sampling procedure in another zone in order to test the homogeneity hypothesis. This preliminary work should be carried out some time before the main echo-integration survey. It will then be possible to define the range of the phenomenon and to calculate the estimation variance for many sampling plans. Depending on the desired precision, one can select the plan best suited to one's needs and constraints, and finally perform the survey by selecting an adequate sampling procedure for each zone.

For stock estimation, there will always be an error on the surface of the lake which is not precisely known; but this type of error is common to all estimation methods, and can perhaps be calculated with the help of approximation formulas proposed by Matheron (1971).

A procedure of this type will result in a satisfactory sampling plan which enables us to obtain a reliable estimation of the biomass. If narrower confidence limits can be obtained, it will be easier to show stock fluctuations.

References

Burczynski, J. J., and Johnson, R. L. 1986. Application of dual-beam acoustic survey to limnetic populations of juvenile sockeye salmon (*Oncorhynchus nerka*). Can. J. Fish. aquat. Sci., 43: 1776–1788.

David, M. 1977. Geostatistical ore reserve estimation. Elsevier, Amsterdam. 364 pp.

Delhomme, J. P. 1979. Spatial variability and uncertainty in groundwater flow parameters: a geostatistical approach. Water Resources Res., XIV(2): 269–280.

Forbes, S. T., and Nakken, O. (Eds.). 1969. Manual of methods for fisheries research survey and appraisal, Part 2. The use of acoustic instruments for fish detection and abundance estimation. FAO Man. Fish. Sci., 5: 1–138.

Francis, R. I. C. C. 1984. Variability in hydroacoustic biomass estimates. Can. J. Fish. aquat. Sci., 41: 825–826.

Gohin, F. 1985. Geostatistics applied to fish distribution as derived from acoustic surveys. ICES Working Group on Fisheries Science and Technology, Tromsø, 22–24 May 1985. 6 pp.

Journel, A. G., and Huijbregts, Ch. J. 1978. Mining geostatistics. Academic Press, London and New York. 600 pp.

Laloe, F. 1985. Contribution à l'étude de la variance d'estimateur de biomasse de poissons obtenus par échointégration. Océanogr. trop., 20(2): 161–169.

Marchal, E. 1985. La détection acoustique dans l'étude des peuplements piscaires. *In* Gestion piscicole des lacs et retenues artificielles, pp. 107–124. Ed. by D. Gerdeaux and R. Billard. INRA, Paris.

Marchal, E., and Laurent, P. J. 1977. Première estimation de la population piscicole du lac Leman par échointégration. Cah. ORSTOM, Sér. Hydrobiol., XI(1): 3–16.

Matheron, G. 1971. The theory of regionalized variables and its applications. Cah. CMM, 5. ENSMP, 211 pp.

Person, R., Marchal, E., Terre, T., and Berthe, J. 1982. Système d'échointégration numérique pour l'évaluation des stocks "AGENOR". ICES/FAO Symposium on Fisheries Acoustics, Bergen, Norway, 21–24 June 1982.

Williamson, N. J. 1982. Cluster sampling estimation of the variance of abundance estimates derived from quantitative echo sounder surveys. Can. J. Fish. aquat. Sci., 39: 229–231.