Aquat. Living Resour., 1993, **6**, 185-199

Exploration of geostatistical methods for mapping and estimating acoustic biomass of pelagic fish in the Gulf of St. Lawrence: size of echo-integration unit and auxiliary environmental variables

Yvan Simard [1], Denis Marcotte [2] and Gilles Bourgault [2]

[1] *Institut Maurice Lamontagne, Ministère des Pêches et Océans, C.P. 1000, Mont-Joli, Québec, Canada G5H 3Z4.*
[2] *Département de Génie Minéral, Ecole Polytechnique, C.P. 6079, Succ. A, Montréal, Québec, Canada H3C 3A7.*

Received February 10, 1993; accepted April 22, 1993.

Simard Y., D. Marcotte, G. Bourgault. *Aquat. Living Resour.*, 1993, **6**, 185-199.

Abstract

Daytime 38 kHz mean volume backscattering strength (MVBS) of small pelagic fish in the Gulf of St. Lawrence was recorded at 60-m intervals along a regular network of 11 parallel transects in August 1991. Profiles of temperature, salinity, density and transmittance were collected on a grid of 44 punctual stations along the transects. The MVBS was contagiously distributed and autocorrelated for all transects. No significant monotone relationship with any of the environmental variables was evidenced, except that the highest concentrations of fish were located over shallow bottoms. The anisotropic variogram of MVBS indicated three different scales of spatial structures: an unresolved small-scale variability, which accounted for only 7.7% of variance, and two structure components of along-transect scales of ~500 m (~40%) and 15 km (~50%). The theoretical variograms deduced for larger echo-integration units (EIUs) than the recorded 60-m unit, provided a good fit to the corresponding experimental variograms up to EIU of 1 920 m, when the disturbing effect of outliers on the variograms was removed. Global estimates and maps produced with 240-m and 1 920-m EIU data were almost identical, showing that no significant loss of information was caused by increasing the EIU. Asymmetric confidence intervals of the global estimates are obtained using two variograms, one controlled for the effect of the outliers, the other one not controlled. In an effort to take the relation of MVBS with the bottom depth into account, four different kriging methods were compared: ordinary kriging, kriging with the bottom depth or its indicator-transform as external drift, and 3-D ordinary kriging with depth as the third dimension. Because of the weak relation of MVBS with the bottom depth, ordinary kriging had the best performance. Results are discussed in the context of applications of geostatistics to estimation and mapping of acoustic survey data.

Keywords: Hydroacoustics, fisheries acoustics, echo-integration, mapping, geostatistics, kriging, spatial statistics, pelagic fish, capelin, *Mallotus villosus*, Gulf of St. Lawrence.

Exploration de méthodes géostatisques pour cartographier et estimer la biomasse de poissons pélagiques mesurée par écho-intégration dans le golfe du Saint-Laurent : taille de l'unité d'écho-intégration et incorporation de variables environnementales auxiliaires.

Résumé

Des données diurnes d'écho-intégration à 38 kHz (MVBS) de petits poissons pélagiques ont été récoltées à intervalles de 60 m dans le golfe du Saint-Laurent le long d'un réseau de 11 transects parallèles en août 1991. Des profils de température, de salinité, de densité et de transmittance ont été réalisés sur un réseau de 44 stations ponctuelles le long des transects. Les données de MVBS montraient une distribution contagieuse et étaient autocorrélées pour tous les transects. Aucune relation monotone significative avec les variables environnementales n'a été mise en évidence, excepté que les plus fortes concentrations de poissons étaient localisées dans des sites de faible profondeur. Le variogramme anisotrope des données de MVBS montrait trois composantes de structure spatiale : une variabilité non résolue de petite échelle

représentant seulement 7,7% de la variance et deux composantes structurales d'échelles d'environ 500 m (~40%), et 15 km (~50%) dans la direction des transects. Les variogrammes théoriques déduits pour des unités d'écho-intégration (EIUs) plus longues qu'un pas de 60 m, s'ajustaient bien aux variogrammes expérimentaux calculés jusqu'à des EIU de 1 920 m, lorsque l'effet perturbant des données extrêmes sur les variogrammes était éliminé. Les valeurs estimées globales et les cartes calculées avec des EIU de 240 m et de 1 920 m étaient presque identiques, révélant ainsi que l'allongement des EIU ne causait aucune perte significative d'information. Des intervalles de confiance asymétriques sont calculés pour les estimations globales en utilisant deux variogrammes, l'un où les effets des valeurs extrêmes sont contrôlés, l'autre où ils ne le sont pas. Afin de tenir compte de la relation du MVBS avec la bathymétrie, quatre méthodes de krigeage différentes ont été comparées : le krigeage ordinaire, le krigeage avec la profondeur ou son indicatrice en dérive externe et le krigeage ordinaire 3-D avec la profondeur comme troisième dimension. A cause de la faible relation du MVBS avec la profondeur, le krigeage ordinaire a montré la meilleure performance. Les résultats sont discutés dans le contexte de l'application de la géostatistique à l'estimation et à la cartographie des données de prospection par écho-intégration.

Mots-clés : Hydroacoustique, pêche, écho-intégration, cartographie, géostatistique, krigeage, poisson pélagique, capelan, *Mallotus villosus*, golfe du St.-Laurent.

INTRODUCTION

Estimation and mapping of variables has always been an important field of research in various disciplines of earth sciences, agriculture, biology, forestry, environment, etc. (Soares, 1993 and references therein). Since most variables in these disciplines are spatially autocorrelated, measurements made in a field are generally more similar if they are taken at nearby locations. This characteristic of variables carries rich information that can be advantageously exploited for various purposes, in particular for estimation and mapping. Geostatistics (Matheron, 1971) is a theoretical framework that has been developed especially to deal with these spatially autocorrelated variables, and many methods have been successfully applied in a wide range of disciplines, including, recently, fisheries (Simard *et al.*, 1992, and references therein). In fisheries acoustics (Johannesson and Mitson, 1983), because of their continuous nature, the data recorded are strongly autocorrelated (Barbieri Bellolio, 1981; Williamson, 1982). Geostatistics appears therefore promising for fish stock estimation and mapping from acoustic data (Foote and Stefansson, 1990). The International Council for the Exploration of the Sea (ICES) has devoted some efforts to this matter through the organization of two special workshops in 1990 and 1991.

Resource *estimation* is a routine task in fisheries research. *Mapping* is however a new objective in this field, even though it is a traditional object of oceanography. Why mapping a moving resource such as fish? First, because most generally, except for fast migrations, their movement is slow compared to the speed of our research survey. For some resources, such as benthic species, it is negligible. Second, because localization of fish in space and in their environment is a powerful key to understanding the system that controls the resource and its dynamics (Margalef, 1979; Simard *et al.*, 1992 and references therein). The present rapid dispersion of the computerized Geographic Information Systems testifies the general need for this type of spatial information. The distribution of fish in space is not random but well-organized according to the physical, chemical and biological factors, that control the activity the fish are involved in, such as feeding, predator avoidance, migration, reproduction and habitat selection. Therefore thorough understanding of the fish distribution in space will help to understand the systems that maintain them, especially the specific oceanographic context, which is poorly known in most cases, and to model properly their population dynamics and their production and aggregation at fishing grounds.

The first attempts to apply geostatistics to the estimation of fish from acoustic data are due to Laloë (1985) and Gohin (1985). Despite these interesting contributions, it was not before 1987 that other trials were presented at the International Symposium on Fisheries Acoustics of Seattle (Guillard *et al.*, 1990), to estimate and map fish abundance on a lake (Guillard, 1991). Other applications to herring stocks off Norway (Petitgas, 1991) and to acoustically estimated demersal fish off Alaska (Sullivan, 1991) have been performed. ICES reports of the above mentioned workshops and other documents also presented application examples.

The objective of the present paper is to explore some geostatistic methods to estimate and map the epipelagic fish community measured by echo-integration technique in the north-western Gulf of St. Lawrence. Special attention is devoted to (1) the verification of the influence of the size of the echo-integration unit (EIU) (sampling support) on the precision of estimates and maps and (2) to the possibility of using the information from auxiliary environmental variables recorded simultaneously to the acoustic data to improve mapping and estimation.

METHODS

Sampling

During August 11-15, 1991, daytime echo-integration of epipelagic fish in the upper 7-47 m of the water column was carried out along a network of 11 parallel cross-shore transects, spaced regularly every 10 nautical miles (18.5 km), in an area of about 8 000 km^2 in the northern Gulf of St. Lawrence (*fig.* 1). An additional transect was done in the longshore direction on August 24. The acoustic system used was a Biosonics Model 102 sounder, equipped with a dual-beam (10° and 25°) 38-kHz transducer and linked to an IBM-PC computer through the Biosonics-ESP interface and echo-integration software version 2.01. The pulse repetition rate was 1 s and the echo-integration time interval was 20 s. The ship was sailing at about 6 knots and the position from a Magnavox-200, Global positioning system (GPS) receiver, was recorded for each echo-integration unit, EIU, which corresponded to a distance of

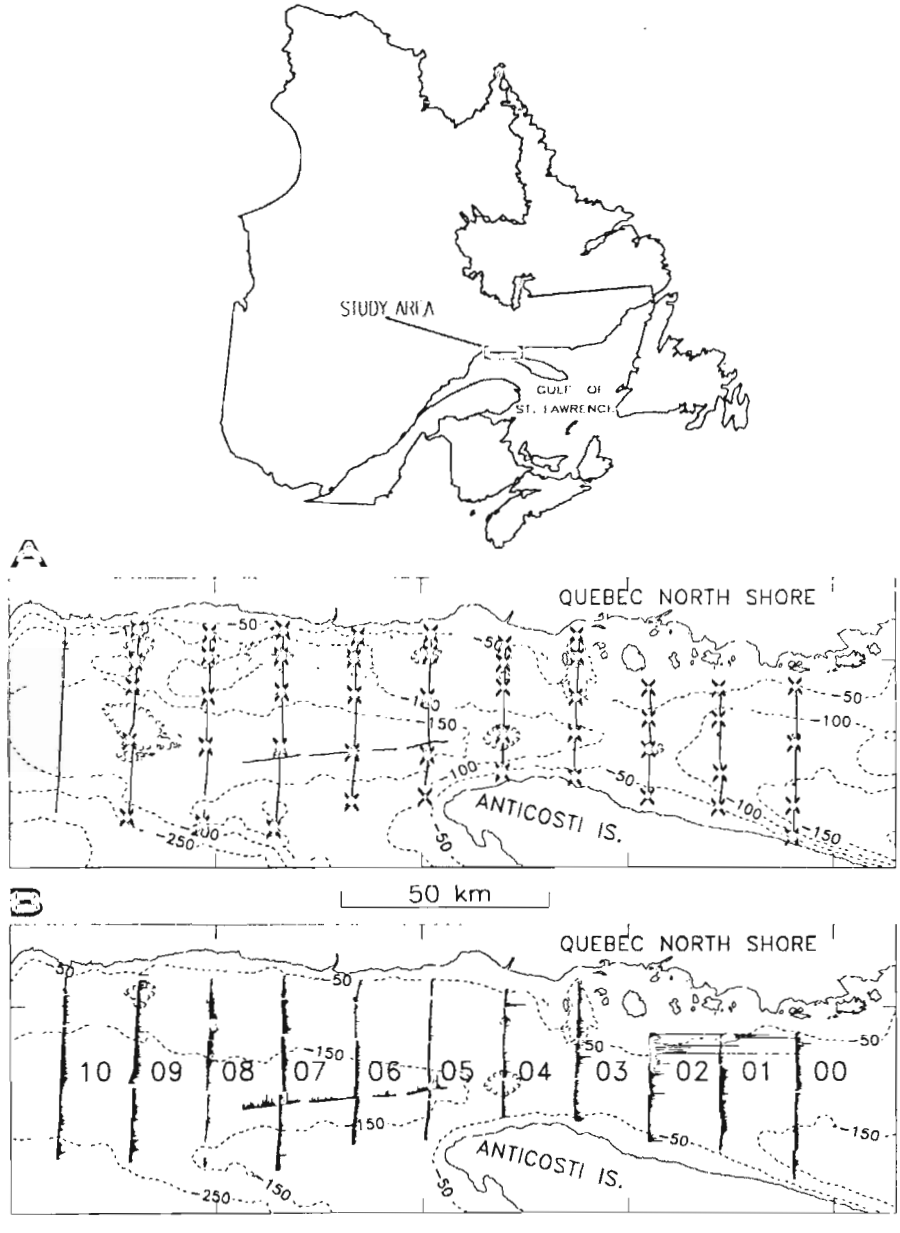

Figure 1. – Maps of (A) the study area showing the hydroacoustic transects and the applied microsystem CTD profile stations (x) and (B) of the distribution of the mean volume backscattering strength (MVBS) in the upper 7-47 m indicated by proportional lines along the hydroacoustic transects. Depth contours in metres.

60 m on average. The mean volume backscattering strength in the sampled stratum, MVBS (in m^2 per m^3, units=m^{-1}), was obtained from standard echo-integration technique (Johannesson and Mitson, 1983). MVBS is proportional to the number of fish per unit of volume, assuming a constant size and target strength. Along the transects, profiles of temperature, salinity and light transmission were recorded regularly at a grid of 44 stations (*fig.* 1) by means of a S-12 Applied Microsystem CTD and a Sea-tech transmissiometer.

Numerical analysis

The existence of a relation of MVBS with the auxiliary environmental variables was checked with correlations and cross-variograms. Two types of variogram were computed: one was the traditional variogram using the raw data, the other eliminated for the disturbing effect of outliers. To compute reliable variograms and to minimize the disturbing effects of outliers, extreme MVBS values (0.7% of the data) were truncated to the value of 24 $m^{-1} \times 10^{-9}$, which corresponded to a change in the slope of the cumulative frequency distribution (*fig.* 3). These outliers did not represent the most significant part of the data, but only exceptional values, and the truncation done concerned only 4.7% of the biomass. As the "robust kriging" method described in Cressie (1991) for handling outliers, this truncation method greatly enhanced the variogram continuity. It has, however, the advantage of not spreading the effect of outliers in computing the maps. The two types of variogram were used in the estimation of the confidence intervals of the global estimates. To test the effect of the length of EIU, the elementary 60-m EIU has been averaged over 2, 4, 8, 16 and 32 samples to obtain larger EIUs. The data set was trimmed to contain an integer number of the largest EIU used (32×60 m = 1 920 m) in order always to compare the same data. This trimmed data set was used in further analysis. The performance of various alternative kriging methods was compared with cross-validation, using the raw data and the truncated data with their respective variograms. Four methods were compared: (1) 2-D ordinary kriging; (2) 2-D ordinary kriging using the bottom depth as external drift; (3) 2-D ordinary kriging using the indicator-transform of the bottom depth as external drift; and (4) 3-D ordinary kriging using the bottom depth as the third dimension. The equations for these 4 kriging systems are given below (for a general procedure, *see* Journel and Huijbregts, 1978; Cressie 1991, Deutsch and Journel, 1992).

(1) 2-D ordinary kriging:

$$\mathrm{MVBS}_i^*(x) = \sum_{\alpha=1}^{n} \lambda_\alpha \mathrm{MVBS}_i(x_\alpha)$$

$$\sum_{\beta=1}^{n} \lambda_\beta \gamma(x_\alpha, x_\beta) + \mu = \gamma(x_\alpha, x)$$

$$\forall\ \alpha = 1, \ldots, n$$

$$\sum_{\alpha=1}^{n} \lambda_\alpha = 1$$

where x: position of the location to be estimated in the 2-D system

x_α: position of a sample in the 2-D system

λ: kriging weight

n: number of samples in the neighbourhood used for kriging

i: EIU

γ: variogram of MVBS_i

μ: Lagrange parameter

(2) 2-D ordinary kriging with the bottom depth (B_i) as external drift. As in (1) but with an additional constraint:

$$\mathrm{MVBS}_i^*(x) = \sum_{\alpha=1}^{n} \lambda_\alpha \mathrm{MVBS}_i(x_\alpha)$$

$$\sum_{\beta=1}^{n} \lambda_\beta \gamma_R(x_\alpha, x_\beta) + \mu_0 + \mu_1 B_i(x_\alpha) = \gamma_R(x_\alpha, x),$$

$$\forall\ \alpha = 1, \ldots, n$$

$$\sum_{\alpha=1}^{n} \lambda_\alpha = 1$$

$$\sum_{\alpha=1}^{n} \lambda_\alpha B_i(x_\alpha) = B_i(x)$$

Where γ_R is the variogram of the residuals of MVBS_i, once the linear effect of the bottom depth is removed. In practice, γ is used instead of γ_R (Ahmed and De Marsily, 1987), both covariances are equal in areas or along directions where the trend can be ignored (Deutsch and Journel, 1992), which is generally the case for small kriging neighbourhoods. Also, when the correlation with the auxiliary variable is low, both covariances are very similar.

(3) 2-D ordinary kriging with the indicator-transform of the bottom depth ($I(B_i)$) as external drift:

As in (2) but the second constraint becomes:

$$\sum_{\alpha=1}^{n} \lambda_\alpha I(B_i(x_\alpha)) = I(B_i(x))$$

According to the scattergram of MVBS against B (*fig.* 4A), which showed that the high values of MVBS were encountered when bottom depths were shallower than 70 m, the indicator-transform, $I(B_i)$, was 1 when

the bottom depth was smaller than 70 m and 0 when it was larger.

(4) 3-D ordinary kriging with the bottom depth as the third dimension:

As in (1) except that x and x_α are 3-D. The variogram of $MVBS_i$ in the vertical direction is assumed to be the same model as the 2-D but the range is 0.014 times the horizontal northing range. This ratio of 0.014 is the mean gradient of bottom depth computed from two successive samples along the transects: $| \Delta B_i | / | \Delta y |$. The introduction of this third dimension requires that the ranges of the horizontal variogram increase by a factor of $\sqrt{2}$ to take into account the effect of adding a third dimension in computing the distances and the semi-variance in the kriging system. In fact, for a horizontal distance Δy, the semi-variance should be kept to $\gamma(\Delta y)$ but, in this 3-D system, it becomes $\gamma\left\{\sqrt{\Delta y^2 + \frac{\Delta B_i^2}{0.014^2}}\right\}$, which is $\gamma\left(\Delta y \sqrt{2}\right)$ for points separated by the mean depth gradient. This distortion effect is corrected by increasing the variogram ranges by a factor of $\sqrt{2}$.

The kriging systems (1) to (4) are all unbiased and systems (2) and (3) take into account the influence of a drift (Ahmed and De Marsily, 1987) imposed by a linear relation with an auxiliary variable, which here is the bottom depth. In systems (2) to (4), the auxiliary variable or third dimension must be known at every point of the kriging grid. It can be obtained from bathymetric charts or from kriging from an available depth data set. Except in unexplored regions, this information is most generally available and its error is negligible for the present purpose of estimating the primary variable. The measured MVBS data were cross-validated (*i. e.* jackknife method) with the kriging systems, using the 8 nearest samples on the same transect. The mean errors, their variances, the mean absolute deviation and a non-parametric criteria were used to compare the 4 methods. The non-parametric criteria was the winner: loser ratio for precise estimation. When the performances of the two methods compared are equal, the value of the criteria is 1.0; the best method is indicated by a value > 1.0 and the worst by a value < 1.0.

The MVBS was mapped with ordinary block kriging (estimate is the average of the block) using two different EIUs, 240 m and 1 920 m. The block size was 2×2 km. They were discretized with a grid of 3×3 points. The moving neighbourhood contained 16 samples, constrained to include 8 samples from each of the 2 nearest transects. Because the size of this neighbourhood depended on the size of the EIU, the difference in maps may depend on the different smoothing due to the different neighbourhoods. Therefore a third map was computed for the 240-m EIU, using a larger neighbourhood containing 48 points and it was compared to the two other maps.

To compute the global MVBS in the study area, a rectangle of influence is defined for each transect. Then block kriging is performed in each rectangle using only the data of the centred transect. This ensures that the kriging errors are approximately uncorrelated. The global estimates are then computed as the weighted means of the kriging estimates for each rectangle, where the weights for the global mean MVBS are the ratios s_i/S of the surface of the rectangle (s_i) to the surface of the study area (S) and those for the estimation variance are s_i^2/S^2. Alternatively to kriging, each rectangle could be estimated by the extension of the central transect mean value to the rectangle and by computing the extension variance from geostatistical charts as in Petitgas (1991). Both approaches give similar results.

RESULTS

The data

From trawl samples taken during the survey, the main fish target in the area, in the upper water column, was juvenile capelin, *Mallotus villosus*, but other secondary species were also present (J.-D. Lambert, Maurice-Lamontagne Institute, pers. comm.). The high-resolution echogram showed that the fish were generally distributed in thin layers, of half a metre to a few metres, mostly in the first 10 m from the transducer, but going sometimes down to 40 m. This high-resolution information is lost by averaging the MBVS over bins of 40 m vertically by 60 m horizontally, corresponding to the elementary EIU. Therefore the density of fish in the only volume they actually occupied is much higher than that presented here by the MVBS averaged over the whole 40-m layer.

The distribution of MVBS was not uniform in the study : transects 01, 02 and 03 in Jacques-Cartier strait were generally richer than the average (*fig.* 2), with some high concentrations in shallow waters near the 50 m depth contour (*fig.* 1B). This eastern pattern was separated from the western part of the study area by a sharp drop in transects 04 and 05, which were the poorest (*fig.* 2). Sudden exceptionally high values were observed on transects 01, 02, 04 and 06 due to the crossing of dense surface schools (*fig.* 1B). Maximum value observed was 316 $m^{-1} \times 10^{-9}$ and the mean 5.1 $m^{-1} \times 10^{-9}$. The histograms of MVBS were asymmetric, a feature whose importance varied among transects.

Auxiliary variables

The bottom depth exhibited a very weak negative correlation ($r = 0.08$, EIU = 60 m, raw data) with the MVBS, even though evident higher abundance in shallow waters was observed on some transects, as mentioned above. The cross-variogram of MVBS

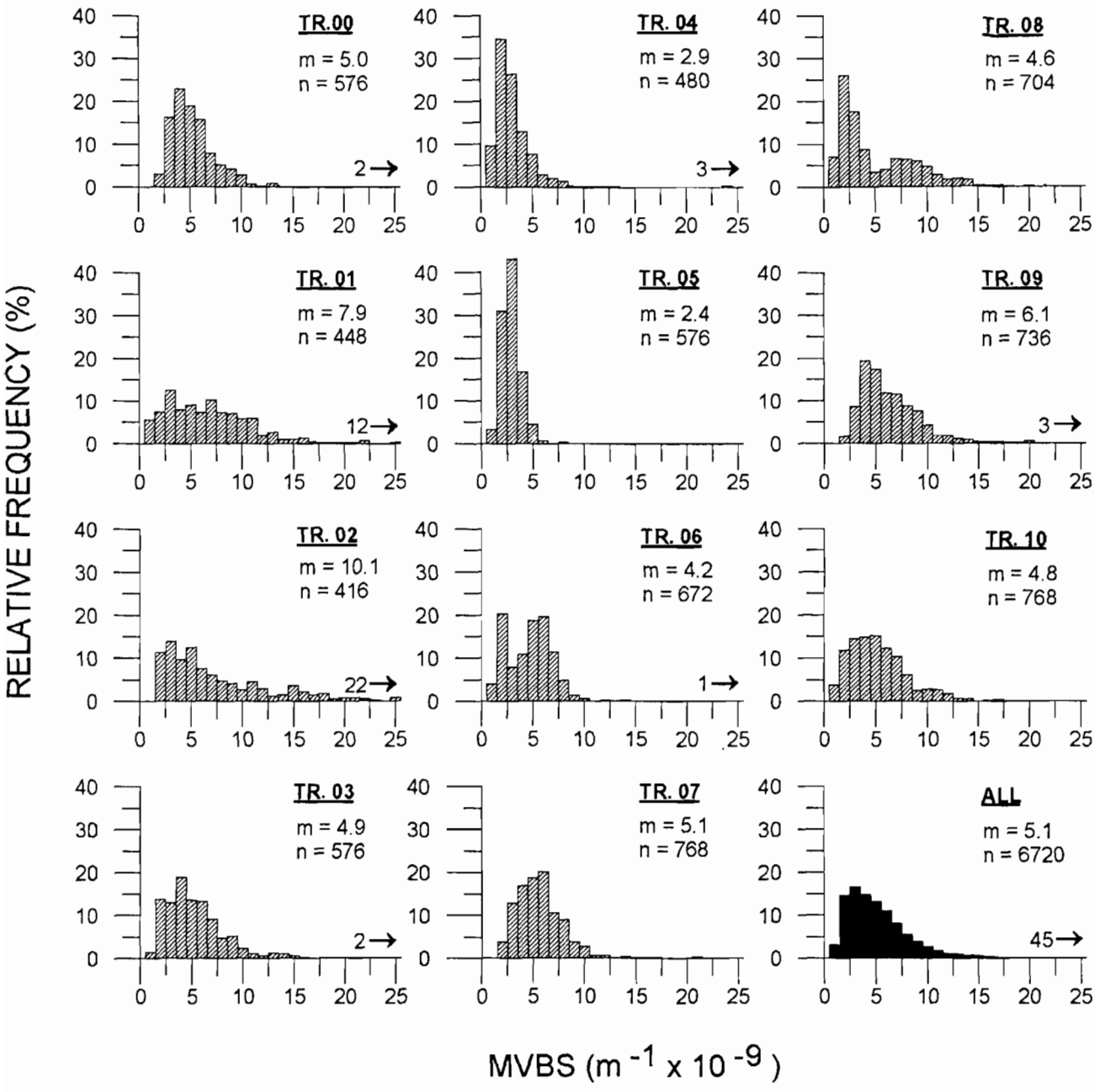

Figure 2. – Histograms of MVBS in the upper 7-47 m along the transects and for all transects together. M = mean, n = no of samples; echo-integration unit = 60 m.

with the bottom depth also oscillated around a zero value. The scattergram of MVBS with the bottom depth (*fig.* 4A) evidenced a non-linear relation, the few extreme high values being all found at depths shallower than 70 m. The kriging methods (2), (3) and (4) were attempts to take this relation into account in the computation of the estimates, in allowing higher weight to samples taken at bottom depths similar to the one of the point to be estimated on the kriging grid. Because this relation with bottom depth implies the high values of MVBS (*fig.* 4A), taking it into account could be advantageous to adequately circle the rich areas and estimate their biomass.

To check the relation of MVBS with the other auxiliary environmental variables collected at the 44 profile stations, MVBS was averaged over a radius of 0.5 km around the station, and its correlation with the values of the environmental variables in the upper

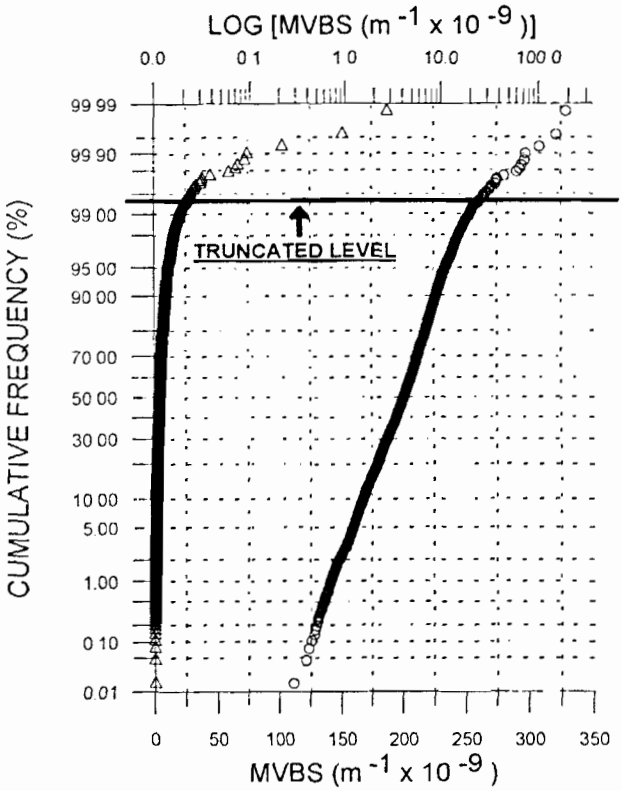

Figure 3. – Cumulative frequency distribution curves of MVBS for echo-integration unit of 60 m showing the truncated level at a change in slopes. Linear and log scales. Data symbols are drawn at every 2 points only.

5 m of the water column was computed. This 0-5 m stratum was chosen for the environmental variables because it was the layer where the temperature and salinity horizontal patterns were the most pronounced. *Figure* 4B indicates that MVBS was significantly correlated with the temperature of the upper water column. The correlation was, however, only due to two high values of MVBS at low temperatures. Similarly, the correlations with the other environmental variables (*fig.* 4C-F) would be close to zero, if the few high values were removed. Because no direct linear relation of MVBS with the environmental variables was evident, and because of the scarcity of environmental data which, among other consequences, did not allow computation of significant cross-variograms, no further attempt was made to see if they could help the estimation of MVBS at unvisited locations.

Variography

The variograms were computed in the transect direction (cross-shore) and longshore, the two main axes of anisotropy and of bathymetric features (*figs.* 5 and 7). The variograms of the truncated data for 60-m EIU (*fig.* 5) exhibited an autocorrelation range (a) in the longshore direction (~27 km) which was estimated to be about 1.8 times larger than in the cross-shore direction (~15 km) (*table* 1). This was supported by the data from the longshore transect (*fig.* 5, 60-m and 240-m easting variograms). At large distances (h) the semivariance was oscillating, possibly because of occasional intermittent mesoscale patches (*fig.* 1B), which were not systematic enough to be considered in modelling the variograms. At small distances, the 60-m northing variogram showed that the pure unresolved random structure was low (nugget, c_0=1.0), and accounted for only 7.7% of the total variance. This unresolved variability may partly reflect the averaging of the vertical spatial pattern over the 7-47 m stratum and background noise. The sharp changes in the slopes of the variogram at small h were indicative of the presence of more than one structure component. A small structure of an approximate range of 0.5 km, which represented about 40% of the total variability, appeared to be nested in the larger 15 km-range structure representing the rest 50% of the variability. The overall 60-m unit variogram was thus modelled with 2 nested spherical functions $C_m \cdot (1.5 \cdot h/a_m - 0.5 \cdot (h/a_m)^3)$, one for each structure (m), combined with a small nugget constant (C_0) (*table* 1).

Table 1. – Parameters of the northings variograms of the mean volume backscattering strength, MVBS $(m^{-1} \times 10^{-9})^2$ for the different echo-integration units EIUs plotted in *figures* 5 and 7. A: Truncated data. A: Raw data. Fitted: manually adjusted variogram; bold values indicate variograms used for estimation and underlined values indicate those used for mapping. Derived: derived from geostatistical charts. Eastings ranges (a_{E1} and a_{E2}) are 1.8 times northing ranges (sills and nuggets are equal to northing ones).

Echo-integration Unit	Nugget C_0	Spherical 1 C_1	Spherical 1 a_{N1} (km)	Spherical 2 C_2	Spherical 2 a_{N2} (km)
	A: Truncated data				
Punctual-derived	–	5.25	0.44	7.00	14.94
60 m-fitted	1.00	5.00	0.50	7.00	15.00
120 m-derived	0.50	4.50	0.56	7.00	15.06
240 m-fitted	**3.00**	–	–	**7.80**	**15.00**
240 m-derived	0.25	3.90	0.68	6.90	15.12
480 m-derived	0.125	2.90	0.92	6.80	15.42
960 m-derived	0.062	1.80	1.40	6.70	15.90
1 920 m-fitted	**0.00**	–	–	**8.00**	**15.00**
1 920 m-derived	0.0312	0.95	2.36	6.60	16.86
	B: Raw data				
60 m-fitted	20.00	13.00	0.20	19.00	4.00
240 m-fitted	**8.00**	**7.00**	**2.40**	**10.00**	**30.00**
1 920 m-fitted	**0.00**	**3.00**	**4.00**	**11.00**	**25.00**

In order to see the effect of lengthening the EIU on the variograms, experimental variograms were computed for various sampling supports on the same data, for some transects separately (*fig.* 6) and for the whole study area (*figs.* 5 and 7). *Figure* 6

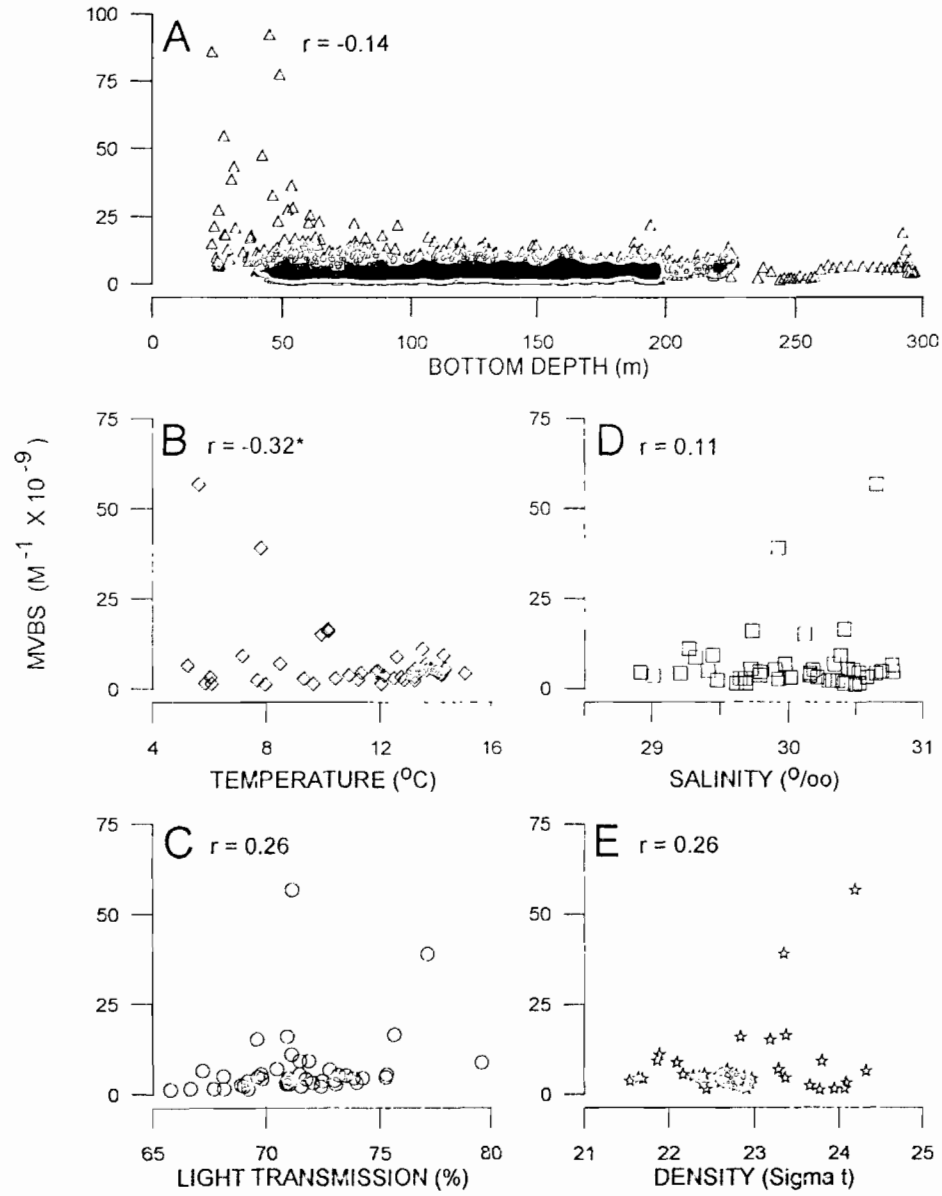

Figure 4. – Scattergram of MVBS against bottom depth along the transects, for echo-integration units of 240 m (A), and against the variables measured at the profile stations (B-E), (*: $p < 0.05$).

shows that all features of the transect spatial structure persisted when the support was increased from 20 to 480 m. This was also true for the whole study area for the variograms computed with the truncated data (*fig.* 5) for a range of supports as large as 32 times the recorded elementary EIU of 60 m. The predicted regularized (*i. e.* computed for supports larger than a point support) variograms computed for gradually enlarged supports, as derived from geostatistical charts (*table* 1), adequately fitted all the experimental variograms for the various supports for the truncated data (*fig.* 5). The variograms of the raw data (*fig.* 7) were more erratic, as expected because of the effect of the outliers, especially for the smallest EIU.

Then, to compare the MVBS global estimates and maps in the study area, we supposed that the data had been recorded on supports of either 240 or 1 920 m and variogram models were fitted to these data (*figs.* 5 and 7, dotted lines). For the truncated data, only one structure was postulated and it was modelled with a spherical function (*table* 1); the small-scale structure which was discernible at small EIUs being smoothed out at large EIUs. This latter small-scale structure was therefore imbedded in the unresolved small-scale variability, which was modelled by a larger nugget

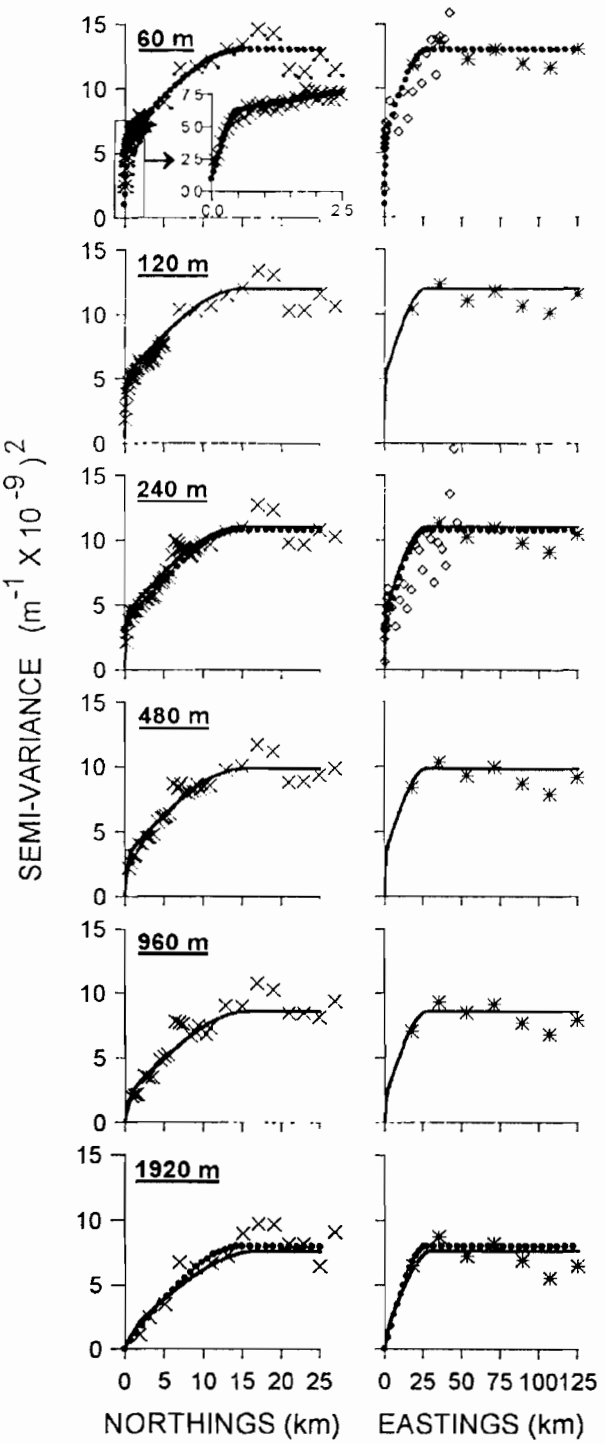

Figure 5. – Variograms of MVBS in the upper 7- 47 m computed with the truncated data for all transects for echo-integration units of 60 m to 1 920 m. Dotted lines: fitted models; continuous lines: derived models from geostatistical charts. The symbol " ◇ " on the 60-m and 240-m easting variograms represents data from the longshore transect. Models parameters are presented in *table* 1. Note the different scales for northings and eastings. The rectangle on the 60-m plot indicates a zoom on the variogram for small distances.

constant (C_0=3.0) for the 240 m-unit. At the large integration unit of 1 920 m, it was completely smoothed out, as expected by the theory. For the raw data, two structure components and large nugget constants were used to model the variograms (*table* 1, *fig.* 7).

Cross-validation

The four kriging methods were compared by cross-validation using the fitted anisotropic variograms for the 240-m EIU. The results indicated that the 2-D ordinary kriging (1) had the best performance, for both the raw data and the truncated data, for all criteria, except the mean error (*table* 2). The 2-D kriging with bottom depth indicator as external drift (3) also produced relatively good estimates. The two other kriging methods had inferior performances. Since kriging (1) was the best, it was therefore retained for estimation and mapping. Kriging (3) was not retained despite its relatively good performance, because its implementation was more complicated and its performance was not better.

Maps

The maps computed with the truncated data (*fig.* 8) showed two main aggregations, east and west of the survey area. They were longer than 75 km, and separated by a poor zone aligned with the western end of Anticosti Island. All three maps A, B and C showed this general feature. Map A, which was computed with the smallest neighbourhood (radius: 18 × 1 km), showed however more small-scale patterns than maps B and C computed with larger neighbourhoods (radius: 18 × 3 km and 18 × 8 km respectively). The differences between maps A and B and maps A and C are shown on maps D and E, respectively (*fig.* 8). Since these two maps of differences (D and E) were almost identical, we can conclude that maps aspects depended more on the size of the neighbourhoods than on the size of EIU. When the size of the EIU is changed from 240 m to 1 920 m the means would therefore be very similar if the neighbourhood size is similar.

Global estimates

The MVBS global estimates obtained using either the EIUs of 240 m or the EIU of 1 920 m were identical for the whole study area, and they were almost the same for the 11 rectangles of influence of the transects (*table* 3). Therefore the size of the EIU did not significantly influence the global estimates. As expected, the variograms computed with the raw data gave the largest variance. For this case, the use of the largest EIU reduced the variance by 25%. Given these results, the global mean of MVBS in the study

Table 2. – Results of the cross-validations of the four kriging methods: (1) 2-D ordinary kriging; (2) 2-D ordinary kriging with bottom depth as external drift; (3) 2-D ordinary kriging with bottom depth indicator (<70 m) as external drift; (4) 3-D ordinary kriging with bottom depth as third dimension. Comparisons of the estimates to the actual value of MVBS for EIU of 240 m (n = 1680). Trunc.: truncated data with their variogram. Raw: raw data with their variogram. Bold is the best performance.

Comparison statistic		Kriging method			
		(1)	(2)	(3)	(4)
Mean error	Trunc.	–0.01	–0.02	–0.01	**–0.00**
	Raw	**–0.01**	–0.03	–0.02	–0.01
Variance of error	Trunc.	**2.89**	3.91	2.91	3.34
	Raw	**11.77**	15.94	11.80	12.23
Mean absolute deviation	Trunc.	**1.03**	1.14	1.03	1.12
	Raw	**1.30**	1.41	1.31	1.33
Ratio of winner: loser for precise estimation when comparing the methods by pair (1.00=equality)	Trunc.	(1) –	(1) 0.91	(1) 0.86	(1) 0.75
	Raw	–	0.99	0.91	0.77
	Trunc.	(2) **1.10**	(2) –	(2) 1.10	(2) 0.88
	Raw	**1.01**	–	0.99	0.88
	Trunc.	(3) **1.16**	(3) 0.91	(3) –	(3) 0.76
	Raw	**1.01**	1.01	–	0.80
	Trunc.	(4) **1.33**	(4) 1.14	(4) 1.32	(4) –
	Raw	**1.29**	1.14	1.26	–

area can be taken from the arithmetic mean, which is 5.1×10^{-9} m^{-1}. Its conservative confidence interval would be asymmetric, with a lower limit given by the variogram computed with the truncated data and an upper limit given by the variogram of the raw data. Using 1.96 σ; this confidence interval would be [0.27, 0.56] for the 240-m EIU, and [0.27, 0.49] for the 1 920-m EIU.

DISCUSSION

Even though very different types of data can of course often be observed, given the complexity of the 3-D distribution of fish in various marine environments, the data analyzed here are not atypical and presented all the features generally encountered in hydroacoustic MVBS data: rich and poor areas, small-scale local variability and general skewed distributions, sometimes with long tails produced by few scarce dense fish schools. We therefore can expect that the features observed here can be easily repeated in other environments.

Despite the highly skewed MVBS, the data were not transformed prior to their analysis because of the numerous problems encountered with this procedure when the goals are estimation and mapping, such as the global stationarity requirements, the high sensitivity to the model parameters, the proper back-transformation, the high weight given to low values and vice-versa, etc. (Journel, 1993; Simard *et al.*, 1992 and references therein). This is clearly presented in this citation of Journel (1993): "In presence of extreme variability it is tempting to work on a smooth transform of the data, such as the logarithm. Experimental variograms of log-data are better behaved, easier to interpret and model. Unfortunately, it is not the log-attribute which is sought after but the attribute itself. A naive inverse transform (e.g. antilog of the log-attribute estimate) in addition to yielding a biased estimator is extremely sensitive to local inhomogeneity in the data (errors are exponentiated !). What is gained in structural analysis ease and robustness by working on a smooth transform of the data is lost at the backtransform stage. This explains why the early 1980s surge in lognormal kriging applications has essentially faded away".

Spatial organization

The spatial organization of MVBS exhibited two longshore-elongated mesoscale (>50 km) aggregations that seemed to meander in the study area (*fig.* 8). These aggregations did not appear to be related to simple water mass characteristics, as measured. The highest concentrations were observed in the eastern end of the study area. This latter part of the area, where most outliers were found, is a highly dynamic environment, the Jacques-Cartier strait, where the tidal circulation interferes with the bottom topography to generate turbulence, mixing and fronts (Y. Simard, unpubl. data). These complex hydrodynamic processes as well as the frequent aggregations of predators in this area (whales) may be related to the local distribution of fish. These aspects will be explored in detail elsewhere.

The width of the two aggregations averaged about 15 km, which corresponds to the range of autocorrelation observed in the transect direction, as shown on the variograms (*fig.* 5). At smaller scales, the variograms revealed the presence of a structure of about 500 m in range, which accounted for 39% of

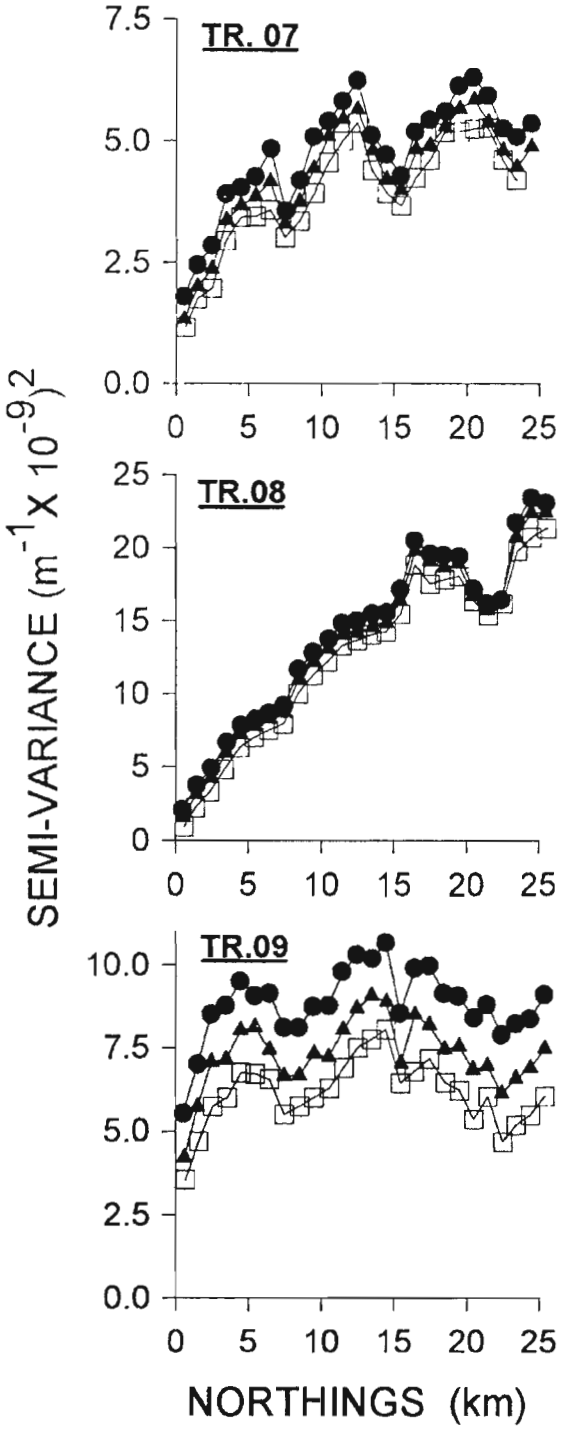

Figure 6. – Variograms of MVBS in the upper 7-47 m along transects 07, 08 and 09 for echo-integration units of 120 m (circles), 240 m (triangles) and 480 m (squares).

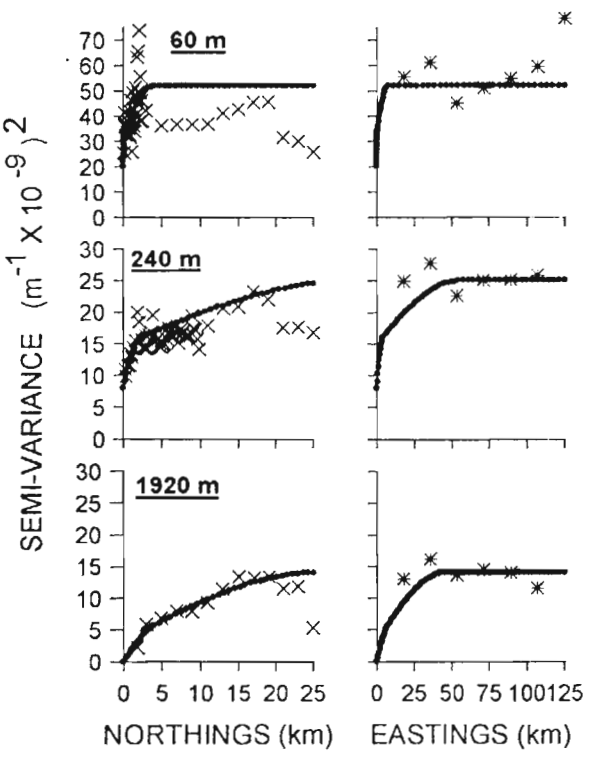

Figure 7. – Variograms of MVBS in the upper 7-47 m computed with the raw data for all transects for echo-integration units of 60 m, 240 m and 1 920 m. Dotted lines: fitted models. Models parameters are presented in *table* 1. Note the different scales for northings and eastings and for the semi-variance.

the total variance when the EIU was 60 m. The source of this small structure is likely clusters of fish schools in the epipelagic layer. The rest of the total variance that remained unresolved for the elementary EIU of 60 m was low, less than 10% of the variance that can be accounted for by the spatial structures. This elementary EIU of 60 m appears therefore adequate to resolve much of the variability of MVBS in the study area, excluding the microstructure. The use of larger EIU would rapidly mask the small structure and merge it with the unresolved variability as the variograms showed (*fig.* 5).

Size of echo-integration unit

Apart from this smoothing of small structures, it appears that the use large echo-integration units, as large as 1 nautical mile, does not hinder data treatment, when the variograms are not disrupted by outliers, like the variograms computed with the truncated data. Large EIUs preserve the main characteristics of the structural information, gradually smooth the unresolved variability, produce similar maps and give identical global estimations. In the case of the raw data, large EIUs offer the advantage of smoothing the outliers, which reduces their disturbing effect on variograms (*fig.* 7). Use of large EIUs is, however, very advantageous for data management and storage, as well as for reducing the computation time. Its only disadvantage is the loss of the information on the small structures. The preservation of this may, however, be important for some studies where very

Table 3. – Mean volume backscattering strength, MVBS ($m^{-1} \times 10^{-9}$) global estimations obtained from ordinary kriging with two different EIUs, 240 m and 1 920 m. Estimates computed separately for the rectangle of influence for each transect and globally for all the study area (sum of transect estimates weighted by their rectangle of influence). Arith. Mean = arithmetic mean of the data. Trunc.: truncated data with their variogram. Raw: raw data with their variogram. Raw/Trunc.: raw data with the variogram of the truncated data.

Transect	Data	Arith. Mean	Estimated			
			Mean		Variance	
			240 m	1 920 m	240 m	1 920 m
TR. 00	Trunc.	5.0	5.1	5.1	0.21	0.20
	Raw/Trunc.	5.0	5.1	5.2	0.21	0.20
	Raw	5.0	5.0	5.0	0.95	0.71
TR. 01	Trunc.	6.8	6.9	6.9	0.23	0.23
	Raw/Trunc.	7.9	8.3	8.3	0.23	0.23
	Raw	7.9	8.6	8.5	1.08	0.82
TR. 02	Trunc.	7.7	7.6	7.6	0.25	0.25
	Raw/Trunc.	10.1	10.4	10.2	0.25	0.25
	Raw	10.1	10.9	10.2	1.18	0.88
TR. 03	Trunc.	4.9	4.8	4.8	0.22	0.21
	Raw/Trunc.	4.9	4.9	4.8	0.22	0.21
	Raw	4.9	4.9	4.8	0.98	0.73
TR. 04	Trunc.	2.8	2.8	2.8	0.23	0.23
	Raw/Trunc.	2.9	2.9	2.9	0.22	0.23
	Raw	2.9	2.9	2.9	1.02	0.77
TR. 05	Trunc.	2.4	2.4	2.4	0.19	0.19
	Raw/Trunc.	2.4	2.4	2.4	0.19	0.19
	Raw	2.4	2.4	2.4	0.89	0.67
TR. 06	Trunc.	4.2	4.2	4.2	0.19	0.18
	Raw/Trunc.	4.2	4.2	4.2	0.19	0.18
	Raw	4.2	4.2	4.2	0.86	0.64
TR. 07	Trunc.	5.1	5.1	5.1	0.18	0.18
	Raw/Trunc.	5.1	5.1	5.1	0.18	0.18
	Raw	5.1	5.1	5.1	0.77	0.57
TR. 08	Trunc.	4.6	4.7	4.7	0.19	0.19
	Raw/Trunc.	4.6	4.7	4.7	0.19	0.19
	Raw	4.6	4.6	4.6	0.83	0.60
TR. 09	Trunc.	6.1	6.2	6.2	0.18	0.18
	Raw/Trunc.	6.1	6.2	6.2	0.18	0.18
	Raw	6.1	6.3	6.2	0.79	0.59
TR. 10	Trunc.	4.8	4.9	4.9	0.18	0.18
	Raw/Trunc.	4.8	4.9	4.9	0.18	0.18
	Raw	4.8	4.8	4.8	0.80	0.59
Total	Trunc.	4.8	4.9	4.9	0.02	0.02
	Raw/Trunc.	5.1	5.2	5.2	0.02	0.02
	Raw	5.1	5.2	5.1	0.08	0.06

local variability plays an important role. The present results have shown that variograms of truncated data change with support size as is predicted by the theory, which was not the case of the variograms of raw data. This was not necessarily expected given the conditions required for the application of the theory, especially the stationarity conditions. The proper variogram has, however, to be modelled on the experimental variogram obtained with the smallest support, in order to incorporate the small-scale structure, which is hidden in the large supports. This behaviour of the variograms of the truncated data relative to the support size may be helpful for some applications involving data recorded with different supports.

Estimation and maps

The global estimates obtained were independent of the length of the EIU, when the variogram is not affected by the outliers. This is an indication of the coherent adjustment of the variograms. The easting range determines the distance from transects where the estimation will take advantage of the presence of the spatial autocorrelation in the data compared to the random case. Unfortunately, in the longshore direction, the data were scarce and the appropriate range of the variogram, to which the estimation of the variance is sensitive, was poorly estimated. The worst hypothesis is, however, the isotropic model. Here the first point

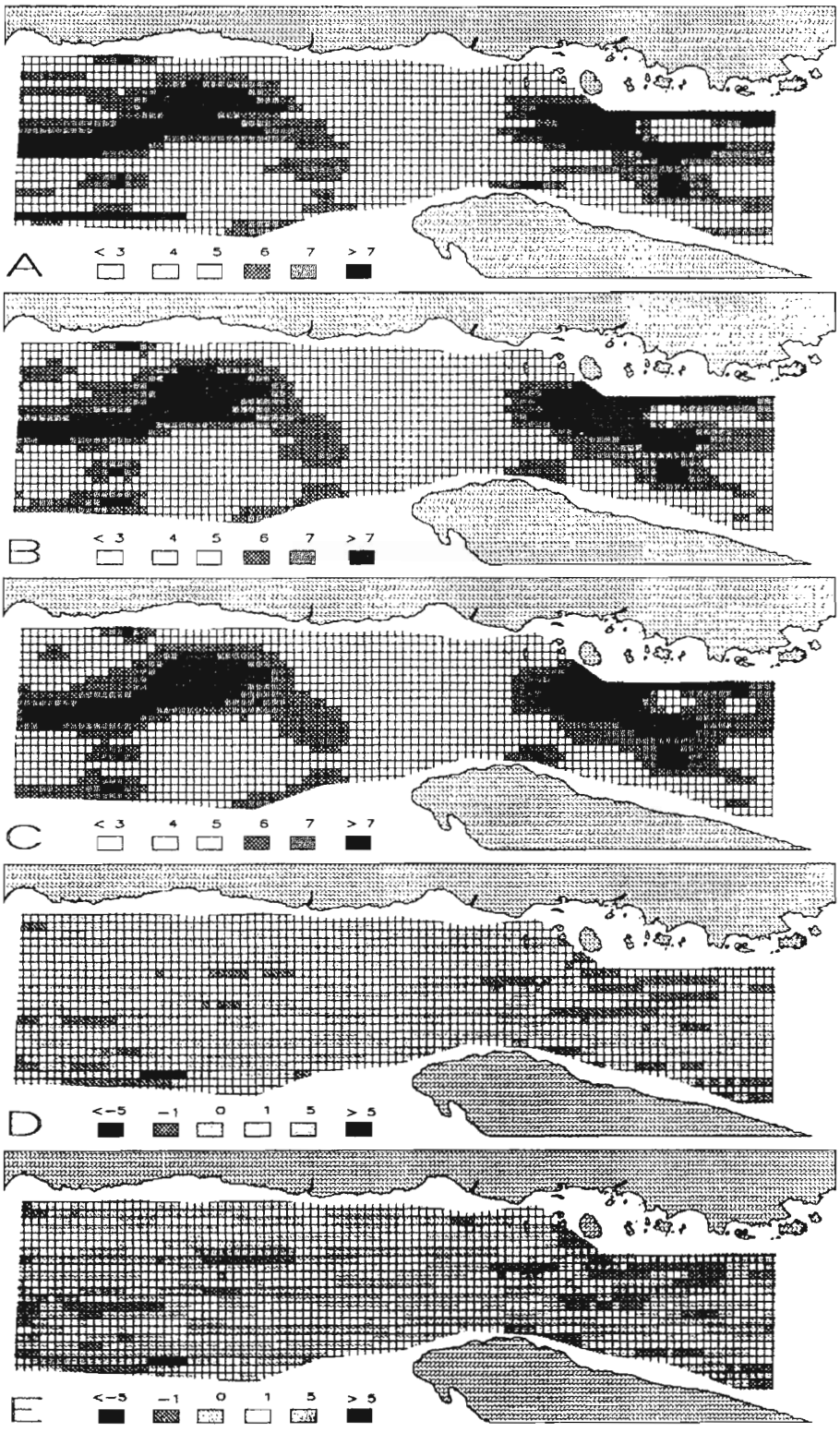

Figure 8. – Maps of MVBS ($m^{-1} \times 10^{-9}$) in the upper 7-47 m estimated from ordinary kriging for echo-integration units of 240 m (A and B) and 1 920 m (C). Neighbours used were: 16 samples in A and C and 48 samples in B. Maps of the differences between the estimates: D = A-B; E = A-C.

on the longshore variogram, when only cross-shore transect data were considered, was at a distance of 18 km and it was lower than the sill, indicating a larger range in this direction than in northings. The inclusion of data from a supplementary longshore transect confirmed that this postulate was reasonable. This longshore anisotropy is common for many variables in the marine environment, such as the bathymetry, the currents, the water mass characteristics, the plankton composition and abundance, etc., where cross-shore gradients are much stronger than longshore ones. The degree of anisotropy (given here by the ratio of easting: northing ranges) was estimated at about 1.8. The number of points on the variogram was, however, too low to precisely define the degree of anisotropy, and sampling along a few more, longer longshore transects, distributed over the whole study area, is therefore recommended to better define the autocorrelation model and, consequently, to help to refine the estimation.

The maps also did not notably differ when computed for different supports using similar neighbourhoods. The main causes of differences in mapping with different supports remain the smoothing effects of the moving neighbourhoods and the border effects at the boundaries of the study area. To prevent sharp steps in mapping when changing transects, the neighbourhood must of course always include samples from at least the 2 adjacent transects. Mapping with small neighbourhoods using small EIUs minimized smoothing and thus generated more small spatial patterns, but the distribution density of samples in the study area was not sufficient to support them. Such fine mapping would however be interesting to evidence small-scale patterns and it should be done using the information from more densely-distributed samples of auxiliary variables, such as the bottom depth.

Auxiliary variables

Despite the effort to take into account an expected relation of MVBS with water mass characteristics, no significant correlation could be evidenced with any of the environmental variables profiled at the punctual stations. Either there was no such correlation, or the data were inappropriate to show it. Instant correlation may not exist because of temporal effects, such as the time-lag needed in recently upwelled or vertically-mixed waters to grow food for herbivores and primary carnivores. The variables measured may not be the appropriate ones, by their nature, their location or their density. This latter possibility is likely, with too few environmental data available compared to the continuous hydroacoustic data. Local or lag-correlation may exist but the environmental data were too scarce to show it. It is also likely that fish may respond to the multivariate flavour of the habitat, to a special combination of the environmental variables and not simply to the single variables individually. Also, the relation is probably often non-linear and may involve threshold-type responses. In any case, in order to elucidate correlation with the environmental variables, and make use of it in estimation, much more effort should be put into their collection, quantitatively, by using towed vehicles for continuous monitoring along the transects, ideally in a "tow-yo" mode, and qualitatively, by using a variety of probes for measuring more variables potentially involved in determining the distribution of fish. This is unfortunately far beyond the reach of most acoustic surveys. Another use of the auxiliary variables is to characterize the MVBS outliers in order to define the part of the survey area that requires more sampling effort because of the high variability. For that purpose, the association of high MVBS values with some characteristics of the habitat, even at a few points only as observed here (*fig.* 4), may help to define the area where sample density should be increased, and which should be isolated from the rest of the study area for data analysis. This requires almost real-time analysis of the MVBS and environmental data though, in order to exploit this information during the survey. Fortunately, this is becoming almost standard on research ships with the present micro-computer and sampling technologies.

Bottom depth is a different auxiliary variable because of its continuous nature, like the acoustic data, and because it is easy to obtain at unsampled sites from bathymetric information. Correlation of MVBS for pelagic fish with bottom depth was unexpected, and this was generally true here, except in the densest areas, where a break in the relation occurred. The effort to take this relation into account in the estimation failed; the estimates did not improve, as the cross-validation showed. Geostatistical tools to incorporate possible relationships with bottom depth, even complex, however are numerous. Modelling the relation as a trend and removing it before kriging may also sometimes be advantageous (e.g. Sullivan, 1991). Here, there was no evidence of a trend that could be modelled and removed. To improve the estimation and mapping, it is therefore worthwhile to investigate such relationships with inexpensive auxiliary continuous information such as bathymetry, bottom substrate, and distance to the shore or to particular depth contours.

Acknowledgments

We thank all people who participated in the data collection at sea. Special thanks are addressed to P. Jalbert and to J. Benoit for their contribution to data preparation. We are grateful to F. Gerlotto for providing helpful unpublished literature on acoustic data analysis. Discussions with members of the GEOSPACE working group on spatial organization of aquatic organisms have been stimulating for this study. Comments from anonymous referees were beneficial for improving the manuscript.

REFERENCES

Ahmed S., G. De Marsily, 1987. Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. *Water Resour. Res.*, **23,** 1717-1737.

Barbieri Bellolio M. A., 1981. Variabilité des données acoustiques utilisées dans l'évaluation des biomasses halieutiques par écho-intégration. *Thèse dr.,* Univ. Bretagne Occidentale, Brest, No. 133, 197 p.

Cressie N. A. C., 1991. Statistics for spatial data. John Wiley & Sons, Inc., 900 p.

Deutsch C. V., A. G. Journel, 1992. GSLIB: Geostatistical software library and user's guide. Oxford University Press, New York, 340 p.

Foote K. G., G. Stefansson, 1990. Definition of the problem of estimating fish abundance over an area from acoustic line-transect measurements of density. ICES C.M. 1990/D:25, 17 p.

Gohin F., 1985. Geostatistics applied to fish distribution as derived from acoustic surveys. ICES, FAST WG, Tromso, 22-24 mai 1985, 6 p.

Guillard J., 1991. Etude des stocks pisciaires lacustres par échointégration : problèmes méthodologiques. *Thèse dr.,* Univ. Claude Bernard-Lyon-I, No. 9291, 153 p.

Guillard J., D. Gerdeaux, J. M. Chautru, 1990. The use of geostatistics for abundance estimation by echointegration in lakes : the example of lake Annecy. *Rapp. P.-v. Réun. Cons. int. Explor. Mer*, **189,** 410-414.

Johannesson K. A., R. B. Mitson, 1983. Fisheries acoustics. A practical manual for aquatic biomass estimation. FAO Fish. Tech. Pap., 240, 249 p.

Journel A. G., 1993. Geostatistics: roadblocks and challenges. *In*: Geostatistics Troia '92, A. Soares Ed., Kluwer Academic Publishers, Dordrecht, Vol. 1, 213–224.

Journel A. G., Ch. J. Huijbregts, 1978. Mining geostatistics. Academic Press, N.Y., 600 p.

Laloë F., 1985. Contribution à l'étude de la variance d'estimateurs de biomasse de poissons obtenus par échointégration. *Océanogr. trop.*, **20,** 161-169.

Margalef R., 1979. The organization of space. *Oikos*, **33,** 152-159.

Matheron G., 1971. The theory of regionalized variables and its applications. Les cahiers du CMM, 5, Ecole Nat. Sup. Mines Paris, Fontainebleau, 211p.

Petitgas P., 1991. Contributions géostatistiques à la biologie des pêches maritimes. *Thèse dr.,* Ecole Nat. Sup. Mines de Paris, 157 p. + 63 figs.

Simard Y., P. Legendre, G. Lavoie, D. Marcotte, 1992. Mapping, estimating biomass and optimizing sampling programs for spatially autocorrelated data: case study of the northern shrimp *Pandalus borealis*. *Can. J. Fish. Aquat. Sci.*, **49,** 32-45.

Soares A., 1993. Geostatistics Troia '92, Vol. 1 and 2. A. Soares Ed., Kluwer Academic Publishers, Dordrecht.

Sullivan P. J., 1991. Stock abundance estimation using depth- dependent trends and spatially correlated variation. *Can. J. Fish. Aquat. Sci.*, **48,** 1691-1703.

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.