Mapping, Estimating Biomass, and Optimizing Sampling Programs for Spatially Autocorrelated Data: Case Study of the Northern Shrimp (*Pandalus borealis*)

Yvan Simard

Ministère des Pêches et des Océans, Institut Maurice-Lamontagne, C.P. 1000, Mont-Joli (Québec) G5H 3Z4, Canada

Pierre Legendre and Gilles Lavoie

Département de sciences biologiques, Université de Montréal, C.P. 6128, succ. A. Montréal (Québec) H3C 3J7, Canada

and Denis Marcotte

Département de génie minéral, École polytechnique, C.P. 6079, succ. A, Montréal (Québec) H3C 3A7, Canada

Simard, Y., P. Legendre, G. Lavoie, and 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.

The methodology for mapping and for global and cutoff estimation of autocorrelated exploitable resources is presented, based on stationary geostatistical methods. Use and performance of these methods in marine ecology are illustrated with an application to northern shrimp (*Pandalus borealis*) abundance data, collected in 1989 at 137 stations in the western Gulf of St. Lawrence. Nonstationarity of the biomass data, a proportional increase of the local variance with the local mean, and the presence of outliers all violated the stationarity assumption and strongly hindered the modeling of the spatial structure. Cross-validation tests showed that kriging estimates were better when interpolating within very local neighborhoods using a small number of points. Kriging always performed better than polygonal tessellation. A stratification scheme produced better estimations than the whole-region approach using traditional or relative variograms. The spatial organization of the shrimp biomass was composed of a trend superimposed onto mesoscale patches of 30–50 km in diameter. The area under study contained about 22 000 tonnes of northern shrimp; 70% of this biomass was concentrated in less than 30% of its surface. The spatial information is used to derive guidelines for optimizing future sampling programs.

La méthodologie pour la cartographie et l'estimation globale et par seuil d'abondance de ressources exploitables autocorrelées est présentée, suivant des méthodes géostatistiques stationnaires. L'usage et la performance de ces méthodes en écologie marine sont illustrés par une application à l'abondance de crevette nordique (*Pandalus borealis*), échantillonnée à 137 stations dans l'ouest du golfe du Saint-Laurent en 1989. La non-stationnarité de la biomasse, un accroissement de la variance locale proportionnel à la moyenne locale et la présence de valeurs extrêmes violaient tous l'hypothèse de stationnarité, et rendaient difficile la modélisation de la structure spatiale. Des tests de validation croisée ont montré que les estimés de krigeage étaient meilleurs en restreignant l'interpolation à un voisinage très local et en utilisant peu de points. Le krigeage a toujours montré une meilleure performance que la méthode de tessellation polygonale. Les estimations suivant une stratification de la région étaient meilleures que celles sans stratification utilisant un variogramme traditionnel ou relatif. L'organisation spatiale de la crevette montrait une tendance superposée à des agrégations de 30–50 km de diamètre. La région à l'étude contenait environ 22 000 tonnes de crevette nordique; 70% de cette biomasse était concentrée dans moins de 30 % de sa superficie. Les lignes directrices pour l'optimisation de programmes d'échantillonnage futurs sont déduites de l'information structurale.

Received November 16, 1990
Accepted July 30, 1991
(JA806)

Reçu le 16 novembre 1990
Accepté le 30 juillet 1991

Physical, chemical, and biological variables in natural environments often present well-defined structures in time and space, such as gradients, recurrent patches, mosaics, and other complex patterns. These structures, which can be described statistically (Legendre and Fortin 1989), are intrinsic characteristics of ecosystems. They are generated by complex spatio-temporal processes acting over a continuum of scales (Margalef 1979), like the various mechanisms controlling the distribution of chemical compounds in soils (Yost et al. 1982; Burgess and Webster 1980a, 1980b), the assemblages of plant species in forests (Marbeau 1976; Legendre and Fortin 1989), or the organization of herbivores and their predators. In aquatic environments, currents, water masses, nutrients, plankton, fish, and whales are also not distributed at random but well organized in time and space (Haury et al. 1978; Steele 1978; Margalef 1979; Mackas et al. 1985; Legendre and Troussellier 1988).

In this situation, the realization of a variable at one location is dependent on its realization at nearby locations. This spatial autocorrelation requires special considerations in estimating these variables. Commonly used classical estimation methods assume the independent selection of the samples, which allows the application of classical sampling designs dictating the

(rather rigid) distribution of the samples through space (e.g. Cochran 1977). Classical estimation methods cannot be used when the data have not been obtained following these sampling designs.

When the spatial structure is known and stable through time, the space is often discretized into homogeneous strata, where spatial structuring is assumed to be absent or negligible. Allocation of the samples may then be optimized by stratification in order to minimize the variance of the estimate and the structure-dependent bias. The precision of the mean is strongly dependent on the effectiveness of the strata definition, the allocation of samples to strata, and on the stability of the known structure (Cochran 1977). When the spatial structure is unknown or unstable through time, this method will not always reduce the variance of the estimate. Such situations are common in aquatic environments.

A third and major problem lies in the fact that the formulas of classical statistics used to compute the confidence interval of the mean of the resource are based on estimates of the variance which, in turn, assume that the error terms of the samples are stochastically independent of one another. This condition is not met by spatially autocorrelated data (Cliff and Ord 1981) so that variances and confidence intervals calculated from these classical formulas are unrealistic. Geostatistics (Matheron 1971) tells us that, besides the shape of the spatial autocorrelation and the density of samples, the variance of an estimate also depends on (1) the geometry of the volume to be estimated, (2) the spatial organization of the set of samples, through the relative location of samples, and (3) the location of the samples in the volume.

Geostatistics was developed to deal with estimation problems in spatially autocorrelated phenomena. It makes use of the additional information provided by autocorrelation when computing the estimates and their confidence intervals. It has proved useful in various disciplines such as geology, forestry, hydrology, oceanography, etc., and its application is presently extended to the estimation of exploited marine resources, where classical methods present difficulties (Conan 1985; Nicolajsen and Conan 1987; Armstrong et al. 1989; Petitgas and Poulard 1989). The general applicability of geostatistics for this latter task is, however, questionable. It assumes that the probability law describing the spatial structure is known or can be adequately modeled from a limited set of samples. This is not obvious for standard data sets in marine ecology, where trends and complex spatial structures are common, and where the number of samples is generally low compared with the high variability present. Besides, the theory of geostatistics is well described but its practice is much less documented (Isaaks and Srivastava 1989), and detailed applications for various tasks in marine ecology are lacking. The present paper explores geostatistical methods to analyze the spatial structure and estimate the biomass of the northern shrimp (*Pandalus borealis*) in the western Gulf of St. Lawrence in the fall of 1989. The reason for using geostatistics with the present data is twofold. First, our data are autocorrelated, so that a correct estimation of the confidence interval of the mean could not have been obtained from classical statistics. This confidence interval will be necessary to assess whether a significant temporal trend is present in the evolution of the stock. Second, sampling has not been carried out following one of the classical sampling plans (random, regular, stratified, and so on) and comes from three different concurrent research projects.

The exploited biomass of the northern shrimp in the Gulf of St. Lawrence is found on muddy bottoms between depths of 150 and 350 m (Fig. 1). It is spatially autocorrelated. For the present data, the value of the generalized Moran's I statistic (Moran 1950; Cliff and Ord 1981) that can be computed for the various distance classes is 0.32 for the distance class of 0–30 m, which is significantly different from zero ($p < 0.001$). The spatial organization presents aggregations in various areas that are sought and exploited by the fishermen. In the western Gulf of St. Lawrence, the northern fringe of the Laurentian Channel, between depths of 150 and 250 m (Fig. 1), tends to consistently support higher densities of shrimp. Except for this general pattern and for the transient congregation of berried females on the shallow bottoms around 150 m in winter and spring in some ares (Savard 1989), the aggregations show high, seasonal and interannual, unpredictable variations (Y. Simard, unpubl. manuscr.). Despite their primary importance for the industry and the fishermen, little is known on the spatial structure of the shrimp biomass in the Gulf or on the mechanisms that control its variability.

Changing spatial structures profoundly hinder the direct estimation of the global biomass on fishing grounds, for all sampling plans. When using a stratified random plan (Cochran 1977; Mackett 1973), the determination of effective strata is difficult because the location of the aggregations is unstable, and unpredictable without a priori knowledge of the existing structure. We therefore need a method of estimation that is tolerant to the geographic organization of the biomass, such as tessellation methods of interpolation (like Delaunay's or Dirichlet's), spline approximation, or kriging (e.g. Conan 1985; Stolyarenko 1986; Smith and Mohn 1987; Isaaks and Srivastava 1989).

An additional advantage of geostatistics is its capacity of producing not only global estimates of the total biomass in the study area but also local and cutoff estimates, with their confidence intervals, in presence of spatial autocorrelation. The cutoff estimates correspond to the total biomass incorporated in the delineated area where the biomass exceeds given threshold values. This information is desirable in the estimation of exploitable natural resources whose profitability of exploitation not only depends on the total stock but also on its distribution. The exploitable biomass is given by the cutoff estimate associated with the locations where shrimp biomass is high enough to be fished profitably. The area over which this biomass fraction is distributed and the distribution pattern are also interesting measurements because they directly influence the level of financial profit. Other potentiality of geostatistics is the use of the structure function to optimize sampling strategies (e.g. Burgess et al. 1981) and to draw inferences on the process generating the aggregations (Sokal 1986).

This paper will address the basis of kriging methodology and its application, point out its limitations, and use it to (1) compute optimal estimates of the global biomass, with their confidence intervals, (2) objectively map this biomass, describe the spatial patterns and compute cutoff estimates of the exploitable proportion above given threshold densities, and (3) analyze the structure and exploit its information to optimize sampling strategies.

Methods

Sampling

From 31 August to 10 October 1989, northern shrimp biomass was sampled at 137 stations in the western Gulf of St.

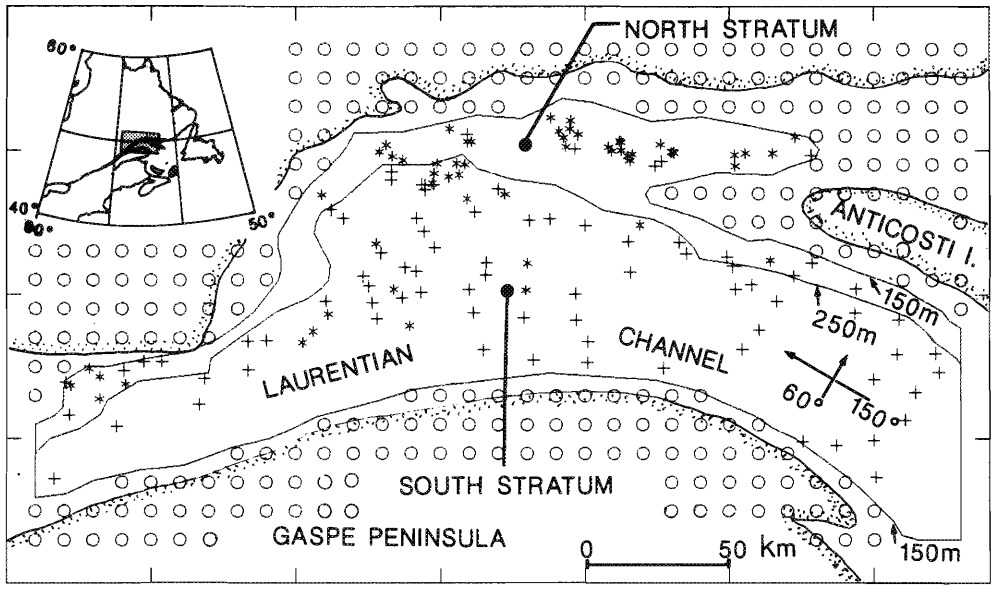

FIG. 1. Map of the western Gulf of St. Lawrence with sample locations and the contour of the studied region at depths greater than 150 m. The isobath of 250 m on the north separates the two strata, north and south. *, >1000 kg shrimp/km^2; +, <1000 kg shrimp/km^2; ○, imposed zeros at the nodes of the interpolation grid outside the studied region for guiding contours (see text).

Lawrence (Fig. 1) by the trawlers M.V. *Bradelle* (26.5 m) and F.R.V. *Alfred Needler* (50 m). The *Bradelle* visited most (120) of the stations and 16 of them, located in the richest area, were paired by the *Alfred Needler* which sampled simultaneously at a nearby station. At each station, bottom shrimp trawls were towed for 30 min at 2.5–3.0 knots. The *Bradelle* used a Western II-A trawl and the *Alfred Needler* used a URI 81/114 trawl. The stretched mesh aperture was 38 mm and the codend was lined with a 19-mm mesh. Since the *Alfred Needler* caught significantly (Wilcoxon test, $p < 0.05$) more shrimp (17% on average) than the *Bradelle*, because of her more efficient trawl, her catches per square kilometre were converted to *Bradelle* catches by means of a linear regression ($r^2 = 0.82$). All sampling was carried out during daytime because vertical migrations of the shrimp off the bottom at night (Barr 1970; Apollonio et al. 1986) could change their availability to the gear. Because of shrimp availability to the gear, avoidance of the gear, and trawl sampling efficiency, the catch does not represent 100% of the shrimp present. Since this is neglected here, the computed estimates of shrimp biomass are conservative. Although the samples covered an approximate area of 14.5 m per about 2.5 km (≈0.035 km^2), they were considered as point samples centered at the starting tow coordinates.

Geostatistical Model

The local and global estimations computed here were obtained using the usual ordinary punctual kriging technique of stationary geostatistics, briefly outlined below; for complete presentations, see Clark (1979) or Isaaks and Srivastava (1989).

Let the shrimp biomass, SHR, at n locations of the sampled region be described by the series of random variables SHR(x_1), SHR(x_2),. . .SHR(x_p),. . .SHR(x_n). Since the shrimp biomass is spatially autocorrelated, the random variables are linked together by a random function, $Z(x)$, which describes the dependence of the random variables SHR(x) on each other. The characteristics of this model can be estimated from the sample data set, if we assume some stationarity conditions.

Stationarity

In the case of ordinary kriging (i.e. the global mean is unknown), the following second-order stationarity conditions are assumed:

$$(1)\quad E[Z(x + h) - Z(X)] = 0$$

$$(2)\quad \mathrm{VAR}[Z(x + h) - Z(x)] = 2\gamma(h)$$

where h is the distance separation between two locations (which can have a directional component) and the function $\gamma(h)$ is the semi-variogram (commonly called the variogram). These two stationarity conditions are referred to as the "intrinsic hypothesis." In this situation, the *difference* in $Z(x)$ does not depend on the locations x but only on the geographic distance separation h, and the average of this difference over the studied region must be zero. These stationarity conditions ask for a certain degree of regional homogeneity of the shrimp biomass and for absence of trend.

These conditions must be carefully checked for proper estimation through stationary geostatistics. Spatial trends in data sets must be removed prior to the analysis and special consideration given to outliers in data sets and to highly skewed probability distributions (e.g. lognormal), which greatly affects the estimation and interpretation of the variogram. Normal distributions are also more prone to provide kriging estimates close to conditional expectations and kriging errors that are normally distributed. For lognormal distributions, normalization of the data by transformation to logarithmic units and the use of the lognormal kriging technique can help minimize problems (Rendu 1979; Journel 1980). Results are, however, very sensitive to slight departures from the strict lognormal distribution and to the choice of the variogram model, and therefore this method must be used with caution (Armstrong and Boufassa 1988). Another way to minimize the effect of increasing variance with the mean is to compute a relative semi-variogram,

$\gamma(h)$/mean^2, using the semi-variogram of the log-transformed data, which is then back-transformed to the raw units according to Journel and Huijbregts (1978).

An implicit assumption of geostatistics is that the structure is stable in time, i.e. the spatial organization of the biomass is fixed, at least for the duration of the sampling period. This constraint might be important for highly mobile species but it is not very restrictive for shrimp. Even though they are not sessile animals, the displacement of their biomass is rather slow compared with the survey speed. Another underlying assumption is that the capturability by the sampling gear is not structured through space. This is likely to happen when the sampling grid is visited around the clock and day and night stations are mixed together, because of diel vertical migrations of the shrimps. This was minimized here by sampling only during daytime, when shrimp are more fully available to bottom trawls than during the night.

Local estimation

Local estimation[1] of the biomass $Z(x_0)$ at a point x_0 is done by interpolation according to the following equation using the data points $Z(x_i)$ available in the surrounding neighborhood chosen:

$$(3)\quad Z^*(x_0) = \sum_{i=1}^{n} w_i \cdot Z(x_i)$$

where w_i are weights, which sum to 1 to insure that the estimate is unbiased ($E[Z^*(x_0) - Z(x_0)] = 0$). The w_i are estimated in such a way that they minimize the variance of error of the estimate ($\sigma_k^{*2}(x_0)$), called the kriging variance (VAR$[Z^*(x_0) - Z(x_0)]$ = minimum). The solution to minimize $\sigma_k^{*2}(x_0)$ is obtained by partial differentiation relative to the weights w_i, taking into account that they must sum to 1 (Isaaks and Srivastava 1989). In matrix notation:

$$(4)\quad \underset{C}{\begin{bmatrix} C^*_{11} & \dots & C^*_{1n} & 1 \\ \vdots & & \vdots & \vdots \\ C^*_{n1} & \dots & C^*_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}} \cdot \underset{W}{\begin{bmatrix} w_i \\ \vdots \\ w_n \\ \mu \end{bmatrix}} = \underset{U}{\begin{bmatrix} C^*_{10} \\ \vdots \\ C^*_{n0} \\ 1 \end{bmatrix}}$$

where the C^*_{ij} are the covariances between the random variables corresponding to the samples ($C^*_{ij} = E\{[Z(x_i) - E(Z(x_i))] \cdot [Z(x_j) - E(Z(x_j))]\}$) that describe the spatial structure of the probabilistic model. The Lagrange parameter, μ, is a mathematical artifice introduced to constrain the weights to sum to 1. The C^*_{ij} are calculated by means of a function that expresses the covariance in terms of the distance separation between samples $C^*(h) = \gamma^*(\infty) - \gamma^*(h)$ where $\gamma^*(h)$ is the estimated variogram and $\gamma^*(\infty)$ is its sill. This sill is equal to the variance of the population (σ^2) under the assumed stationarity conditions, when the sampled field is larger than the spatial structure. Since $\gamma^*(\infty)$ is a constant, the covariance terms, C^*, in equation (4) are replaced by the variograms γ^* without changing the equality. The weights are found by solving $W = C^{-1} \cdot U$, and they are substituted in equation (3) to estimate $Z^*(x_0)$. The corresponding minimized error variance of this estimate, $\sigma_k^{*2}(x_0)$, is obtained from the following matrix equation:

$$(5)\quad \sigma_k^{*2}(x_0) = \sigma^{*2} - W \cdot U$$

[1]The asterisk distinguishes estimates from true values.

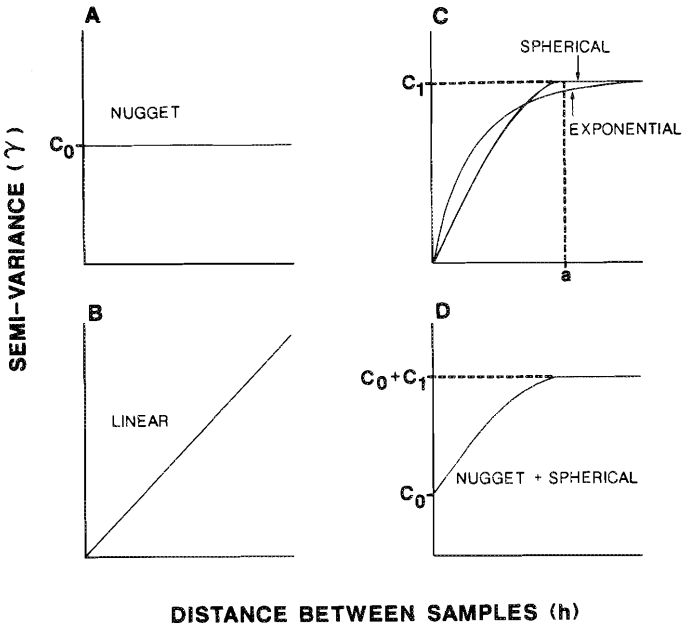

FIG. 2. Common variogram models. (A) Nugget model, $\gamma(h) = C_0$; (B) linear model, $\gamma(h) = bh$; (C) spherical model, $\gamma(h) = C_1\,[3h/2a - h^3/2a^3]$ if $h \leq a$ and $\gamma(h) = C_1$ if $h > a$, where a is the range of autocorrelation and C_1 is the plateau of semi-variance or the sill; and exponential model, $\gamma(h) = C_1[1 - \exp(-h/a)]$; (D) nested nugget + spherical model, $\gamma(h) = C_0 + C_1\,[3h/2a - h^3/2a^3]$ if $h \leq a$ and $\gamma(h) = C_0 + C_1$ if $h > a$.

where σ^{*2} is the variance of the samples. When W and U are obtained from a relative semi-variogram, the computed σ_k^{*2} must be multiplied by the kriged estimate squared, Z^{*2}, to obtain the absolute minimized error. Kriging is referred to as the best linear unbiased estimator (BLUE). It is also an exact interpolator: at a sampled location, x_p, the interpolation always passes through the data point (i.e. $w_p = 1$ and all other $w_i = 0$). This is true even if the variogram does not pass through the origin (nugget effect, see below); the interpolation surface is then discontinuous at the data point.

In practice, local estimates are computed using only the nearest samples included in a confined neighborhood. The shape and size of this neighborhood and the data points retained for the estimation affect the value of the estimate. They are chosen according to the model of spatial continuity and to the distribution of samples in the area (Isaaks and Srivastava 1989). Commonly used variogram functions are the *nugget model*, which corresponds to the absence of a spatial structure (random spatial organization) (Fig. 2), the *linear model*, the *spherical model*, and the *exponential model*, whose sill, C_1, indicates the maximum variability due to the structure and whose range of autocorrelation, a, gives the distance at which the samples are no longer autocorrelated. Combinations of these functions are also used to model nested structures, which often include a nugget model to represent the unresolved small-scale variability and sampling error (C_0). The ratio $C_0/C_0 + C_1$, where C_1 is the variability due to the structure in models with a sill, gives the proportion of this latter unresolved random variability to the total variability ($C_0 + C_1$). Optimal choices of variogram and kriging neighborhood are often developed using a jackknife cross-validation method. In this procedure, the data points are removed one by one and are estimated using the chosen variogram model and kriging parameters. The best solution is the

one which, among other properties, minimizes the mean square residuals, produces the standardized residuals ($[Z^*(x_i) - Z(x_i)] / \sigma_k^{*2}(x_i)$) the closest to the conditional expectations of a mean of 0 and a standard deviation (SD) of 1.0, and maximizes the correlation coefficient between the estimated and the observed values, whose linear regression should have a slope close to 1.0 (Samper Calvete and Neuman 1989).

Global and cutoff estimations

Global estimation of the mean biomass density (Z^*) is easily obtained by averaging the m local estimates, computed at the nodes of a regular grid covering the studied region. The total biomass of shrimp (ZD*) is then estimated by multiplying this mean density (grams per square metre) by the surface area S_D of region D:

$$(6) \quad ZD^* = S_D \cdot \frac{1}{m} \cdot \sum_{j=1}^{m} Z^*(x_j).$$

Three-dimensional numerical integration of the mapped biomass of shrimp can also be used for this computation. Cutoff estimates corresponding to the area where the biomass exceeds a given threshold are obtained similarly, using only the local estimates above the chosen biomass threshold.

The variance of these estimations is much more difficult to estimate because we have only one realization of ZD. Two possibilities using the extension variance approach are formulated below. The first procedure computes the variance of the global estimation by summing the variance of strata forming a mosaic over the whole region (Petitgas and Poulard 1989). The strata variances are assumed to be independent of each other. The size of the strata must be smaller than the range of reliability of the computed experimental variogram, which is always less than half the dimension of the studied region (Journel and Huijbregts 1978). The variance σ_D^2, of ZD is then

$$(7) \quad \sigma_D^2 = \sum_l \sigma_l^2 S_l^2$$

where

$$(8) \quad \sigma_l^2 = 2\bar{\gamma}(x_i, S_l) - \bar{\gamma}(S_l, S_l) - \bar{\gamma}(x_i, x_j)$$

where S_l are the surface area of the stratum l, $\bar{\gamma}(x_i, S_l)$ is the average semi-variance between the particular sample locations i in the stratum l and all the points of the stratum (which is an averaged point-to-stratum semi-variance), and $\bar{\gamma}(S_l, S_l)$ is the equivalent for all possible pairs of points in the stratum and $\bar{\gamma}(x_i, x_j)$ for all pairs of samples in the stratum. These average semi-variances $\bar{\gamma}$ are computed from the experimental point-to-point variogram. To estimate the point-to-stratum and the stratum average semi-variances, the strata are discretized in a grid of points and the average semi-variance is estimated using the point-to-point variogram. This variogram can be unique for the whole region when global stationarity holds, or it can be computed for each set of homogeneous strata when it does not. This computation of the confidence intervals of the global estimate is time consuming and awkward when the number of strata and samples is large.

The second procedure is the combination of the elementary sampling errors, which computes the variance of the global estimation by the summation of elementary independent errors (Journel and Huijbregts 1978). Each sample x_i is assigned a surface of influence s_i. The error made when the value of the central point sample Z_{xi} is extended to represent the values of all points in the surfaces x_i is computed according to the formula

$$(9) \quad \sigma_{Esi}^2 = 2\bar{\gamma}(x_i, s_i) - \bar{\gamma}(s_i, s_i) - \bar{\gamma}(x_i, x_i)$$

where $\bar{\gamma}(x_i, x_i) = 0$. The weighted sum of these elementary errors is the variance of the global estimate of the total biomass:

$$(10) \quad \sigma_D^2 = \sum_i s_i^2 \sigma_{Esi}^2.$$

These calculations are cumbersome when the number of samples is large, since they require the computation of the surface of influence of all samples and the estimation of each error variance. An easier approximation of σ_D^2, when the sampling density is uniform, can be obtained by computing the extension variance of a random point in the average elementary surface s, whose dimensions must respect the length/width ratio of the region D. This is given by the formula

$$(11) \quad \sigma_D^2 = \bar{\gamma}(s, s) \cdot s^2 \cdot N$$

whose solution is directly obtained from geostatistical charts (Journel and Huijbregts 1978).

The variance of cutoff estimations is obtained similarly except that the variogram used is the one computed using only the samples where the shrimp density exceeds the given cutoff (Froideveaux 1984).

In our example, we have used a grid spacing of 10 × 10 km, which roughly corresponds to the average spacing between the data points. The contour line of the stock was limited to depths greater than 150 m, the grounds occupied by adults of *Pandalus borealis* in the Gulf of St. Lawrence. Zeros were imposed at the nodes of the grid in shallower bottoms outside this isobath and on land (Fig. 1), and they were used only to force the contours of biomass levels to follow the contour of the sampled area. They were not used as information points during the interpolation process because we have found that the biomass contours obtained in this way were constrained unrealistically.

The similarity between the interpolation grids obtained by kriging was measured using the Pearson correlation coefficient. Since all these maps were obtained from the same set of 137 observations, the kriged maps are not independent of one another and thus correlation coefficients cannot be tested for significance. They remain, however, a valid measure of resemblance. If they were coming from independent data sets, corrections would still have to be introduced in the testing procedure to account for the lack of independence (spatial autocorrelation) of the data points (Clifford et al. 1989).

Interpolation and mapping by *polygonal tessellation* was also used. This method simply consists of attributing the value of the nearest sample to the estimated location, which results in a mosaic of tiles whose surfaces are the surface of influence of each sample (Isaaks and Srivastava 1989). For estimating the global mean and variance, the samples are then weighed by their surface of influence.

Results

Stationarity Conditions

The biomass of shrimp per square kilometre was not homogeneously distributed and did not show a monotonous trend that could easily be modeled. Two areas of different density were found. As expected, the northern side of the Laurentian Channel was distinctly richer than the rest of the surveyed region (Fig. 1 and 3). Because this large-scale structure violated the stationarity conditions, the region was divided into two strata, north and south, bounded by the

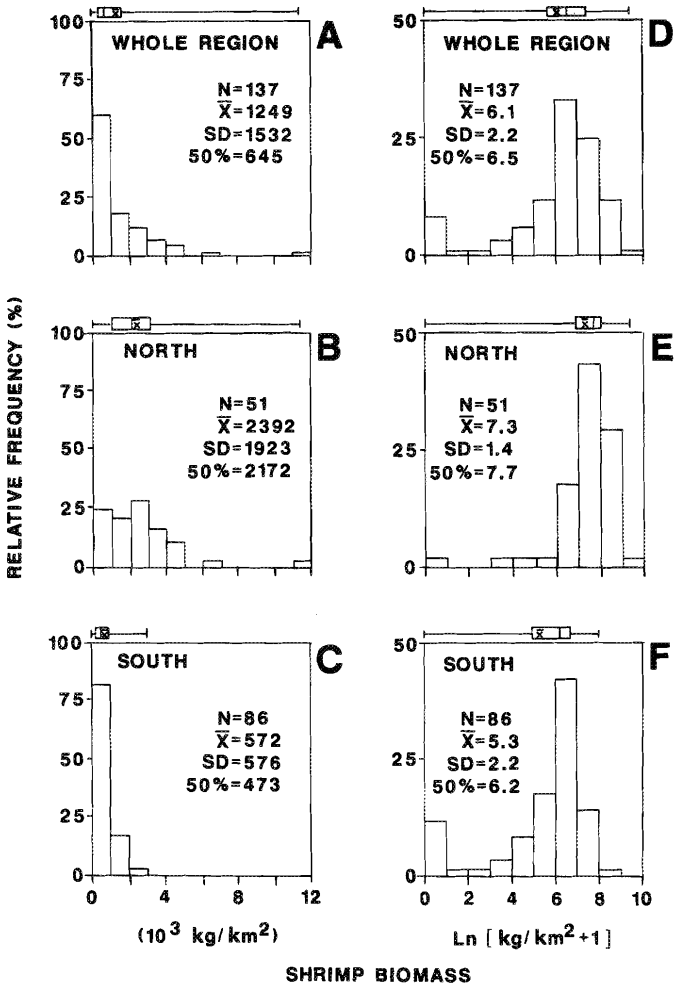

FIG. 3. Histograms with box plots and statistics of the shrimp biomass samples for the whole region and for the north and south stratum separately, for (A–C) raw and (D–F) log-transformed data.

150-m contour towards the coast and separated by the 250-m contour on the northern side of the channel (Fig. 1).

In the whole region and in both strata, shrimp were contageously distributed, and probability distributions were significantly skewed (Fig. 3). The Box–Cox procedure (Box and Cox 1964; Sokal and Rohlf 1981) to identify the best normalizing transformation gave an estimated lambda of 0.30 (95% C.I. = [0.24, 0.37]), indicating that the shrimp biomass was far from the lognormal distribution (lambda = 0), thus precluding the use of lognormal kriging (Armstrong and Boufassa 1988). Because kriging using Box–Cox transformations is not yet described and because transformations are delicate for resource estimation (David 1988), we have only used kriging on the original data. Back-transformation in the original metric often leads to substantial bias when the transformation used and/or the structure model are improper, even only slightly.

One outlier, one order of magnitude higher than the mean, was encountered in the western end of the region. To minimize the effect of nonstationarity, indicated by the increasing variance with the mean, kriging was also performed using a relative variogram computed for the whole region. A third estimation using a normal variogram for the whole region was computed for comparison.

Variograms and Spatial Structure

The omnidirectional variogram of the shrimp biomass of the north stratum was relatively erratic (Fig. 4A). Except for the initial point at an average distance of 3 km, all points were dispersed along a line of constant semi-variance of 2 100 000, indicating no spatial autocorrelation of the shrimp biomass at distances larger than 10 or 15 km. A large part of the variability in this stratum was therefore random, and the structure, if any, existed only at small scales. The variogram was tentatively modeled by a nugget effect, representing the unresolved small-scale variability, over which can be superimposed either (1) a weak monotone structure extending over the entire stratum area (Table 1, linear model, not illustrated in Fig. 4A) or (2) a fine structure representing small patches with radius of 10–15 km (Table 1, spherical model with a range of 15 km, Fig. 4A). In this latter mixed model, the variability due to the structure (C_1 = 1 600 000) was about three times higher than the unresolved small-scale variability (C_0 = 500 000). The variogram in the east–west direction (not shown) did not significantly differ from the omnidirectional one. The distribution of the stations in this stratum precluded the computation of significant variograms in other directions. Isotropy was therefore assumed and the isotropic nugget + spherical variogram (Table 1, underlined) was used for the estimations.

The omnidirectional variogram of the south stratum (Fig. 4B, circles) presented a maximum semi-variance much lower than the richer north stratum (Fig. 3). The steady increase of semi-variance with distance indicated that a trend was present. Significant anisotropy was also observed, with the directional variograms across (direction 60°) and along (direction 150°) the Laurentian Channel being different. The low semi-variance at the smallest distance classes on the omnidirectional variogram indicated that small-scale structure could be nested within the large-scale gradient. Various models reasonably fit the variograms (Table 1) and, despite the apparent anisotropy, the two-structure isotropic model consisting of a spherical model *plus* a linear model *plus* a nugget effect performed as well as the anisotropic models and was used for the kriging estimation (Table 1, underlined). The unresolved small-scale variability (C_0 = 60 000) represented a maximum of about 20% of the larger-scale variability due to the structure.

The omnidirectional variogram computed for the whole region (Fig. 4C) gave a rough average of the variograms of the two strata. Its shape was dominated by a large-scale structure with a radius of about 110 km. It was easily modeled by a spherical model with a range of 110 km superimposed over a high nugget effect (C_0 = 750 000), representing an unresolved small-scale variability corresponding to 65% of the structural variability (C_1 = 1 150 000).

The relative variogram was characterized by a plateau of semi-variance from about 20 to 60 km, followed by an increase of semi-variance at larger distances (Fig. 4D). Neglecting the large-scale variability, a spherical model was added to a nugget effect to model the small-scale semi-variance (Table 1), and this variogram was used for kriging.

Cross-validation

All the structures modeled performed poorly in estimating the true value of the samples, especially when a large search radius and a high number of points were used (Table 1). Most solutions reasonably satisfied the global unbiasness condition (mean of standardized residuals = 0) but the distributions of

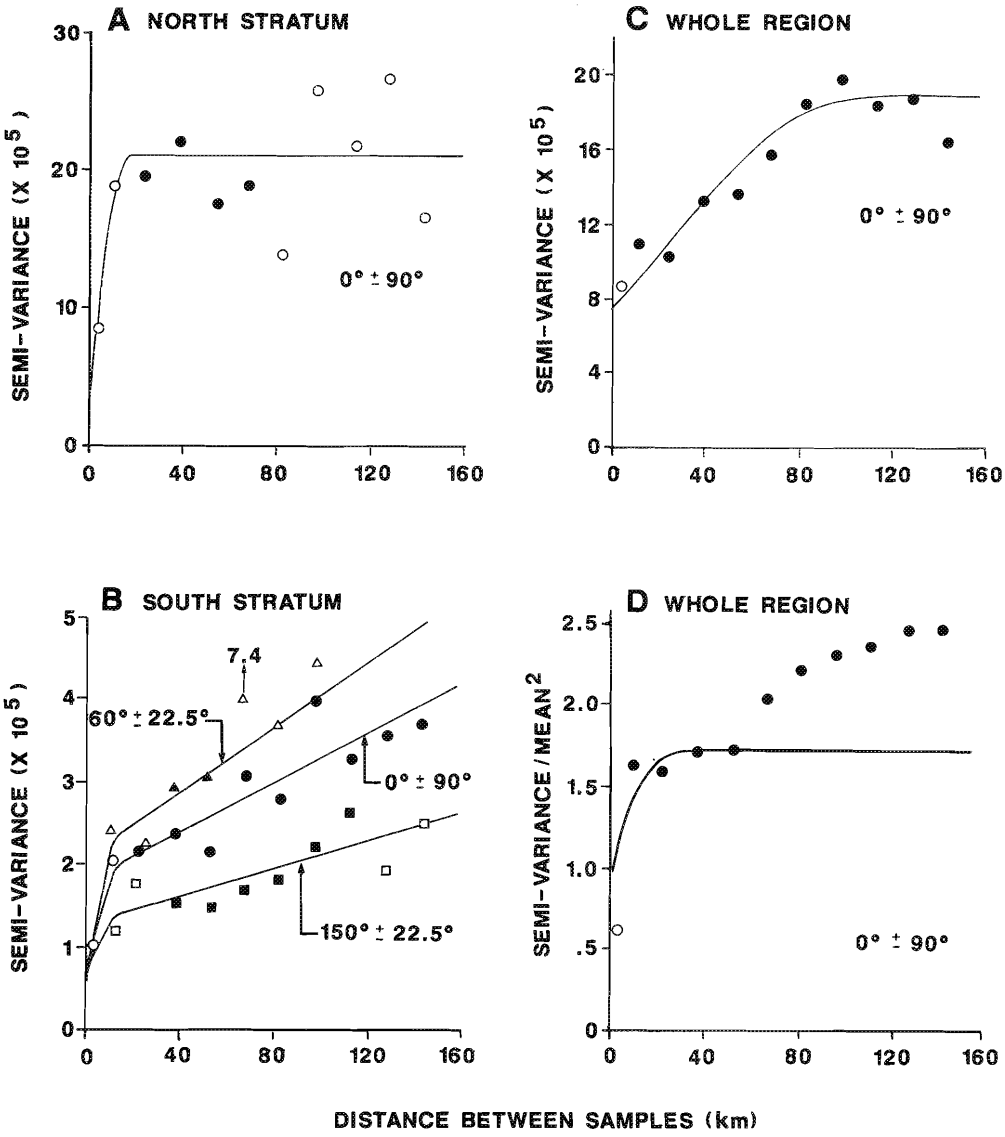

FIG. 4. Variograms computed for (A and B) the two strata separately and (C and D) the whole region. Isotropic variograms (0° ± 90°, circles) and directional variograms along and across the Laurentian Channel (60° ± 22.5°, triangles; 150° ± 22.5°, squares). Fitted models are presented in Table 1 (underlined). Except for the relative variogram (Fig. 4D), all variograms exclude the outlier value of 11 373 kg/km^2. All points include more than 100 pairs, except open symbols (20–100 pairs).

the errors of the estimates, as indicated by their SD's, varied notably from one another. This latter parameter, together with the correlation coefficients of the true values against the estimates, was used to select locally optimal solutions. The best estimations of the samples were obtained using either the two or the four nearest samples, for all the variogram models explored (Table 1, underlined). The goodness of fit was, however, weak with less than 50% of the variance being explained by the chosen spatial models.

The slopes of the regressions of the true values against the estimates were very different when computed with raw or log-transformed data. This resulted from the fact that the variance of the raw data increased with the value of the estimates. The logarithmic transformation maximized the homoscedasticity condition for the regression by giving less weight to the high values (for which the error is high) which more greatly affected the slopes. For all models, the slopes of the log-transformed data converged towards the value of 1 when the correlation coefficients were maximum; these trials were thus chosen as the optimal ones. The corresponding slopes for the raw data were smaller than 1, indicating that, on the average, the kriging estimates underestimated the low values and overestimated the high values ("conditional bias": Isaaks and Srivastava 1989). Examination of the residuals of the regressions showed that this conditional bias of the optimal models chosen was small, and mainly resulted from large errors associated with a few high values.

Interpolation by polygonal tessellation did not perform well in the cross-validation tests (Table 1). The slope parameters were always far from 1 and the SD's of the errors of the estimates were larger than the kriging alternatives.

In the north stratum, the cross-validation clearly showed that the linear model with high nugget effect was unsuitable. The estimates were uncorrelated with the true values, except when only two points were used. Lowering the nugget parameter by fitting a small-scale structure using a spherical model with a

TABLE 1. Cross-validation of the variograms models and search methods. SD of the errors of the estimates (SD_E) from the true value ($Z^*(x_i) - Z(x_i)$) obtained using the given variogram and the search method at the sample points (x_i); Pearson correlation coefficients (r_1 and r_2) and slopes (b_1 and b_2) of the linear regressions $\ln(Z + 1) = a_1 + b_1 (\ln Z^* + 1)$ and $Z = a_2 + b_2 Z^*$; mean, SD, and maximum values of standardized residuals ($(Z^*(x_i) - Z(x_i))/\sigma_k^*(x_i)$). For the relative variogram, these residuals are $[Z^*(x_i) - Z(x_i)]/[\sigma_k^*(x_i) \cdot Z^*(x_i)]$. Underlined are the chosen combinations for computing the kriging estimates of Table 3.

	Variogram	Search radius (km)	Max. no. of points	SD_E	r_1	b_1	r_2	b_2	Standardized residuals: Mean	SD	Max.
North stratum[a]	Isotropic, nugget + linear	200	20	1484	0.26	2.26	0.00	0.02	0.02	1.14	−3.7
	$\gamma(h) = 1\,500\,000 + 5500\,(h)$	100	20	1486	0.27	1.95	0.02	0.09	0.01	1.14	−3.7
		50	20	1482	0.29	1.64	0.08	0.25	0.01	1.13	−3.7
		50	8	1489	0.29	1.46	0.10	0.28	0.04	1.12	−3.4
		50	4	1515	0.21	1.00	0.13	0.27	0.14	1.09	−3.2
		50	2	1350	0.67	1.55	0.45	0.62	0.06	0.89	−2.8
		25	8	1533	0.15	0.33	0.11	0.24	0.05	1.13	−3.3
		25	4	1537	0.17	0.36	0.16	0.29	0.11	1.10	−3.2
		25	2	1356	0.55	0.99	0.46	0.61	0.05	0.89	−2.8
	Isotropic, nugget + spherical	200	20	1334	0.53	2.97	0.39	0.96	0.08	1.06	−3.7
	$\gamma(h) = 500\,000 + 1\,600\,000\ \mathrm{Sph}_{15}(h)$	100	20	1315	0.65	3.00	0.42	0.94	0.02	1.05	−3.7
		50	20	1301	0.66	2.60	0.44	0.90	0.03	1.04	−3.7
		50	8	1314	0.65	2.51	0.43	0.83	0.05	1.03	−3.6
		50	4	1398	0.49	1.88	0.34	0.61	0.08	1.07	−3.5
		50	2	1313	0.69	1.51	0.49	0.66	0.04	0.96	−3.2
		25	8	1357	0.35	0.74	0.39	0.68	0.04	1.05	−3.6
		25	4	1434	0.32	0.66	0.33	0.52	0.06	1.08	−3.5
		25	2	1319	0.58	1.00	0.49	0.65	0.03	0.96	−3.2
	Polygonal method (estimate = nearest sample)	200	1	1553	0.57	0.46	0.35	0.33	—	—	—
South stratum	Isotropic, nugget + linear	200	20	531	0.73	1.33	0.41	0.77	0.01	1.23	−6.1
	$\gamma(h) = 130\,000 + 2200\,(h)$	100	20	531	0.73	1.33	0.41	0.77	0.01	1.23	−6.1
		50	20	544	0.71	1.29	0.37	0.69	0.01	1.25	−6.2
		50	8	536	0.73	1.23	0.41	0.70	0.02	1.22	−6.2
		50	4	507	0.70	0.96	0.50	0.76	0.03	1.12	−6.1
		50	2	516	0.70	0.88	0.52	0.66	−0.00	1.06	−5.5
	Isotropic, nugget + linear + spherical	200	20	496	0.76	1.67	0.51	0.90	0.01	1.09	−5.4
	$\gamma(h) = 60\,000 + 1500\,(h) +$	100	20	496	0.76	1.67	0.51	0.89	0.01	1.10	−5.4
	$120\,000\ \mathrm{Sph}_{15}(h)$	50	20	518	0.72	1.32	0.46	0.77	0.01	1.13	−5.6
		50	8	517	0.73	1.23	0.47	0.74	0.02	1.11	−5.6
		50	4	498	0.70	0.97	0.52	0.77	0.02	1.05	−5.5
		50	2	512	0.70	0.88	0.53	0.67	−0.02	1.03	−5.0
	Anisotropic, nugget + linear	200	20	539	0.71	1.82	0.38	0.74	0.04	1.26	−6.0
	$\gamma(h) = 130\,000 + 800\,(h_{150°}) +$	100	20	539	0.71	1.82	0.38	0.74	0.04	1.26	−6.0
	$3500\,(h_{60°})$	50	20	550	0.69	1.29	0.35	0.65	0.04	1.28	−6.2
		50	8	539	0.72	1.21	0.40	0.68	0.05	1.23	−6.2
		50	4	514	0.70	0.98	0.49	0.73	0.05	1.14	−6.2
		50	2	514	0.71	0.89	0.52	0.67	−0.01	1.05	−5.6
	Anisotropic, nugget + linear +	200	20	504	0.60	1.24	0.49	0.85	0.03	1.11	−5.1
	spherical[b] $\gamma(h) = 60\,000 + 850\,(h_{150°}) +$	100	20	505	0.60	1.24	0.49	0.85	0.03	1.11	−5.1
	$70\,000\ \mathrm{Sph}_{15}(h_{150°}) + 2000\,(h_{60°}) +$	50	20	526	0.69	1.27	0.44	0.72	0.03	1.13	−5.3
	$150\,000\ \mathrm{Sph}_{15}(h_{60°})$	50	8	524	0.72	1.16	0.46	0.70	0.04	1.12	−5.3
		50	4	508	0.70	0.97	0.51	0.73	0.04	1.06	−5.3
		50	2	514	0.71	0.88	0.52	0.66	−0.01	1.02	−4.9
	Polygonal method (estimate = nearest sample)	200	1	605	0.61	0.60	0.47	0.45	—	—	—
Whole region	Isotropic, nugget + spherical[a]	200	20	957	0.76	1.43	0.66	1.05	−0.00	0.97	−4.7
	$\gamma(h) = 750\,000 +$	100	20	958	0.76	1.42	0.65	1.05	−0.00	0.97	−4.7
	$1\,150\,000\ \mathrm{Sph}_{110}(h)$	50	20	958	0.76	1.39	0.65	1.05	−0.00	0.97	−4.5
		50	8	982	0.75	1.27	0.63	0.95	0.01	0.99	−4.5
		50	4	974	0.74	1.04	0.65	0.86	0.05	0.94	−4.4
		50	2	914	0.74	0.93	0.71	0.81	0.05	0.82	−3.8
	Relative variogram:	200	20	1272	0.75	1.85	0.56	1.00	0.07	0.74	−4.8
	isotropic, nugget + spherical	100	20	1272	0.75	1.85	0.56	1.00	0.07	0.74	−4.8
	$\gamma(h)/\mathrm{mean}^2 = 0.934 +$	50	20	1282	0.74	1.42	0.54	0.97	0.03	0.79	−4.8
	$0.782\ \mathrm{Sph}_{27.47}(h)$	50	8	1321	0.73	1.18	0.52	0.82	−0.02	0.92	−5.3
		50	4	1117	0.72	1.01	0.53	0.73	−0.07	1.21	−11.3
		50	2	1361	0.73	0.91	0.56	0.64	−0.31	2.45	−25.8
	Polygonal method (estimate = nearest sample)	200	1	1646	0.67	0.65	0.44	0.43	—	—	—

[a]Excluding the outlier value of 11 373 kg/km^2.
[b]Corresponds to a zonal anisotropy for the spherical model.

range of autocorrelation of 15 km resulted in much better estimations. The SD's of the errors of estimates were reduced by 9% on average. Compared with the polygonal method, the kriging model chosen improved by 15% the SD of the error of the estimates.

The results of the cross-validation of the four kriging solutions explored for the south stratum were less variable than for the north stratum. The best solution chosen had a SD of the errors of the estimates 18% lower than the corresponding SD for the polygonal method.

For the whole region, the relative variogram model did not perform as well as the traditional variogram, the SD's of the residuals being 33% larger on average. Both models were, however, much better than the polygonal method, which had SD's of residuals 47 and 77% greater, respectively.

Estimation and Mapping

Point kriging at the nodes of the 10 × 10 km gid was performed for each of the four optimal combinations of variogram and kriging parameters retained by the cross-validation (Table 1). For the two-strata estimation, the estimates were computed in each stratum by ordinary point kriging using, as information points, not only the samples of the given stratum but all the samples of both strata. This was done to prevent the establishment of an artificial discontinuity at the boundary between the two strata. All the maps of the kriged estimates (e.g. Fig. 5) were very similar, the Pearson statistics computed between them being greater than 0.90; the map obtained by the polygonal estimation method differed from the kriged maps, however, as shown by the lower values of the correlation coefficient (Table 2). The differences between the maps mainly resulted from the different smoothing effect caused by kriging with either two or four points.

By contrast, the maps of the kriging SD's were very different (Fig. 6). Because the computation of the kriging SD's involves the variance of the samples per stratum or for the whole region (equation (5)), the two-strata scheme (Fig. 6A) was clearly discernible compared with the whole-region approach (Fig. 6B). The two-strata scheme produced a lower average error but a larger range of variation compared with the whole-region model. For the relative variogram case, the weighing of the relative kriging SD's by the estimates resulted in a much larger range of errors and in a pattern mirroring the map of the estimates (Fig. 5).

Except for a few square kilometres in its eastern end, all parts of the survey region contained some shrimp according to the kriging performed (Fig. 5). The shrimp stock contour (Fig. 5A) was therefore not determined by the samples but by the external zeros imposed a priori to force the contours to follow the boundaries of the study region. The deepest parts of the Laurentian Channel (Fig. 1) were generally poor, especially easterly (Fig. 5B–5C). The shrimp were concentrated in a long and wide northern patch (Fig. 5B), which exhibited three rich areas (Fig. 5D), and in an isolated 40-km patch in the western end of the survey region (Fig. 5B–5D). The exploitable areas, with densities higher than 1000 kg/km^2 (Fig. 5C), covered 27–30% of the surface of the whole region and they contained slightly more than 70% of the total biomass (Fig. 5C and 5D; Table 3). Areas richer than 1500 kg/km^2 represented about 60% of the total biomass, concentrated in 20–22% of the total surface. A global biomass of about 22 kt was present in the region during the sampling period (Table 3), according to the kriging estimations. The polygonal method gave a global estimate 14%

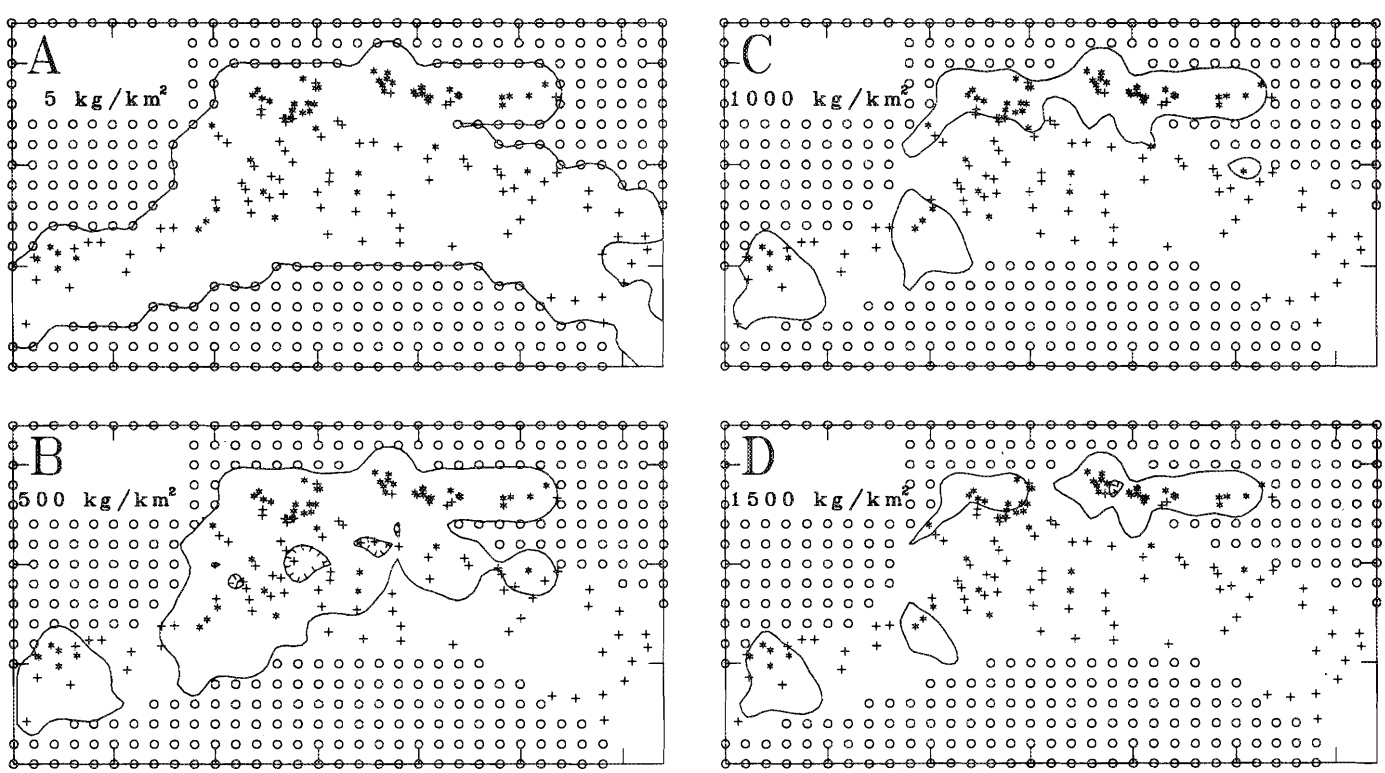

FIG. 5. Maps of the shrimp biomass contours of 5, 500, 1000, and 1500 kg/km^2 obtained from ordinary point kriging using the two-strata model presented in Table 1. Hatched contours are holes. Symbols as in Fig. 1.

TABLE 2. Comparisons of the maps of the shrimp biomass computed using the combinations of kriging parameters chosen in Table 1. Resemblance from the Pearson correlation coefficients computed between the kriged estimates of the maps. These coefficients cannot be tested (see text).

Method	Two strata	Whole region: Traditional variogram	Whole region: Relative variogram
Polygonal tessellation	0.86	0.83	0.78
Whole-region relative variogram	0.95	0.90	
Whole-region traditional variogram	0.92		

lower. The estimated relative standard errors (SE = σ_D/total biomass, σ_D computed from equation (11)) of the kriging global estimates were one third to one half lower than the SE's of the polygonal estimate, and the two-strata model produced the lowest error. The SE's of the cutoff estimates were relatively low.

Discussion

Biomass Structure

Inferences on the spatial structure from the variograms must be done with caution because stationarity conditions were not satisfied. Since the variance tended to increase with the mean and the variograms differed in the range and shapes of the autocorrelation function depending on the area considered, the direction of the variogram, and the type of data transformation, one should concentrate on the local level and on each stratum separately.

At the smallest scale of the observation window, the unresolved variability was not negligible and it increased with the mean biomass of the strata, meaning that either the sampling error or the amplitude of small spatial patterns of biomass, or both, increased with shrimp density. Sampling errors are likely to increase with shrimp density because of the proportional errors associated with the effective area sampled, the avoidance of the trawl by shrimp and the sampling efficiency. Horizontal microstructures and the variability of the vertical distribution of shrimp in the vicinity of the bottom layer sampled by the trawl may also be density dependent. The relative importance of these factors that all contribute to the high nugget variability is, however, unknown. This question deserves more research efforts.

At small distances of about 15 km, the slopes of the variograms of each stratum changed markedly. This may indicate that the shrimp tended to aggregate in mesoscale patches, having a diameter which was twice this distance, but the sampling grid used was too coarse to confirm it through the structure function. The kriged maps showed, however, four rich areas of 700–2000 km^2. Increasing the local density of samples is therefore recommended to more precisely define the intercept of the spatial structure model and to verify the existence of this mesoscale pattern.

The chief macrostructures evidenced on the kriged maps was the rarity of the shrimp along the southern shelf of the Laurentian Channel and a tendency of being richer in the northwest. The reason for this is unclear, since the sediment type (Loring and Nota 1973) and bottom temperatures (2–6°C)

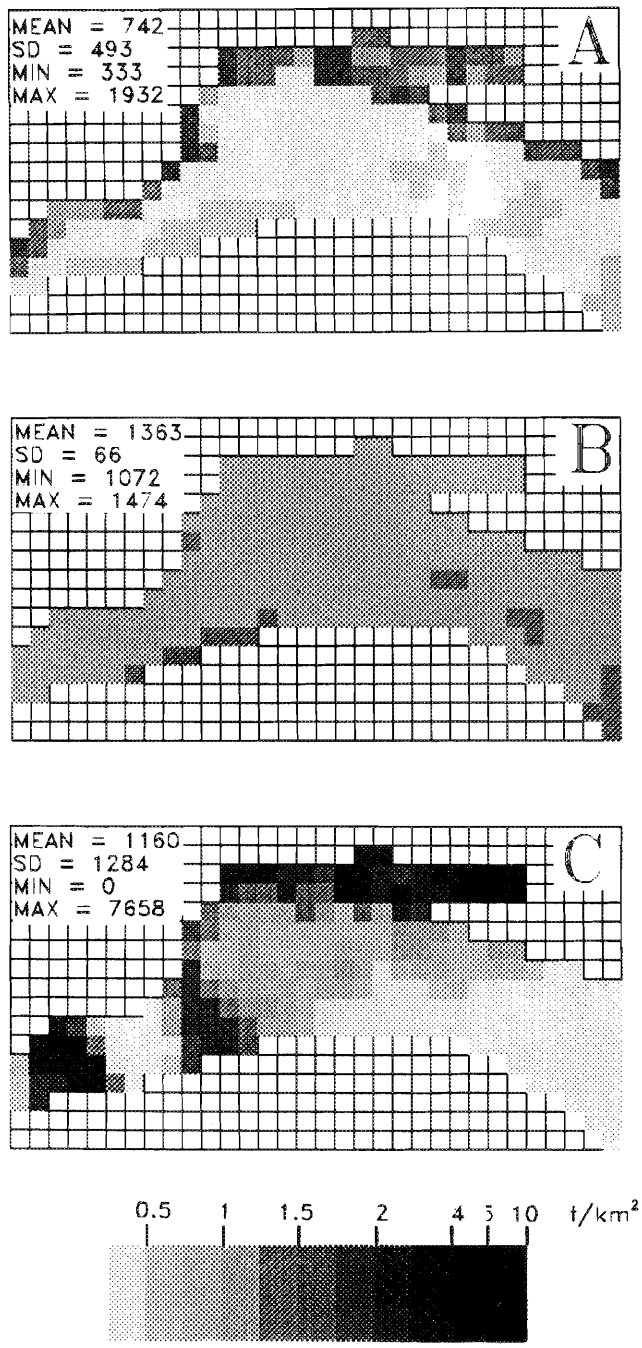

FIG. 6. Maps of the kriging SD's corresponding to the three kriging schemes presented in Table 1: (A) two-strata variograms; (B) whole region traditional variogram; (C) relative variogram. Point kriging σ_k^* ($\sigma_k^* \cdot Z^*$, in Fig. 6C) estimated at the nodes of the 10 × 10 km grid cells.

are favorable throughout the sampled region. This structure is the source of the trend we observed. It is an important regional feature of the shrimp spatial organization, which has a high degree of persistence, from unpublished shrimp fishery catch and research data. Many mechanisms can be invoked to generate and maintain this structure. We suggest the following circulation hypothesis. The location of the region at an elbow of the Laurentian Channel and the presence of a basin in the northeast surely affect the deepwater circulation and might favor the "retention" of the shrimp and/or their food in the rich areas

TABLE 3. Estimated mass of shrimp ($\times 10^3$ kg) in the study region for four cutoff densities according to discrete summation of the kriging punctual estimates (kg/km^2) at the nodes of the interpolation grid (equation (6)) or to numerical integration with the Simpson 3/8 rule, for the chosen kriging methods (Table 1), polygonal tessellation, and the arithmetic mean. The relative errors (SE = σ_D/total biomass) are computed from equation (11) for the kriging methods and from classical statistics for the polygonal tessellation.

	Discrete summation				Simpson integration	
Method and cutoff	Mean × (kg/km^2)	area = (km^2)	mass (t)	SE (%)	Area (km^2)	Mass (t)
Cutoff: >0 kg/km^2						
Traditional variogram	933.00	22 800	21 272	8.7	23 241	21 684
Relative variogram	912.71	24 100	21 996	9.7	24 596	22 449
Two-strata variograms	911.28	24 100	21 962	6.5	24 650	22 463
Arithmetic mean						
Whole region	1249.28	24 300	30 358			
Two-strata	2391.73	6 500	15 546			
	571.77	17 800	10 178			
	1058.69	24 300	25 724			
Polygonal tessellation	818.68	23 214	19 005	13.1		
Cutoff: ≥500 kg/km^2						
Traditional variogram	1621.99	11 800	19 139	7.1	12 120	19 658
Relative variogram	1498.21	13 400	20 076		13 902	20 828
Two-strata variograms	1538.57	13 000	20 001		13 492	20 759
Cutoff: ≥1000 kg/km^2						
Traditional variogram	2454.76	6 100	14 974	5.9	6 292	15 444
Relative variogram	2198.44	7 100	15 609		7 463	16 404
Two-strata variograms	2279.43	6 900	15 728		7 247	16 518
Cutoff: ≥1500 kg/km^2						
Traditional variogram	2913.60	4 500	13 111	5.8	4 626	13 478
Relative variogram	2563.41	5 100	13 073		5 447	13 961
Two-strata variograms	2779.76	4 700	13 065		4 980	13 844

where they are found. The residual circulation in the region is upstream on the northern side of the channel and downstream on the southern side (Trites and Walton 1975). The cumulative effect of this cyclonic circulation on the shrimp, during their excursions in the water column at night (Simard et al. 1990), would result in a higher probability of (1) retention of the shrimp (and their food) upstream, north, and (2) exportation downstream, outside of the region on the southern side.

Kriging Performance

The variograms indicate the presence of a structure, but very little of this information was useful for the estimation, as indicated by the jackknife cross-validation tests. The best reestimations were always associated with a small number of information points, which contradicts the idea that taking the spatial structure into account is advantageous when computing the local estimates. This resulted from the two above-mentioned factors: the high unresolved variability and the lack of stationarity.

When stationarity does not hold and when the variance increases with the value of the variable, efforts must be made to minimize the deviation from stationarity, at least locally. This can be done in many ways such as removing the trend and kriging on the residuals, transforming the data to logarithmic units, computing relative variograms, stratifying the region in more homogeneous strata, or using more elaborate geostatistical models. All these tools have, however, their own limitations and conditions of application. The relative variogram appraoch used here gave no advantage for modeling the spatial structure, its results being worse than the traditional variogram for the whole region. Likewise, the logarithmic transformation was probably inappropriate for the type of nonstationarity encountered. The partitioning of the region into two more homogeneous strata provided a better fit of the spatial model. This approach gave a SD of the errors of the estimates of the spatial model over the whole region of 718 kg/km^2, which was 21% lower than the whole-region approach with the traditional variogram (SD_E = 914 kg/km^2). The importance of having the proper structure model in each stratum (or the whole region) to minimize the SD of the errors of the estimates was also clearly depicted in the cross-validation trials. The drawback of stratifying is that it requires more data handling and computation than with the whole-region approach, and the number of strata is limited by the number of points needed to compute a significant variogram (>50 points). This extra work was worthwhile because our ultimate goal is to identify *significant* changes in the temporal evolution of the stock, which is in part dependent on the narrowness of the confidence intervals of the estimates.

The three kriging models produced very similar global and cutoff estimates and maps of the shrimp biomass, but their estimated errors were quite different. As expected, the stratification of the region in two strata of different density and variability produced the smallest SE of the global estimation. The gain was a SE reduction by 25% compared with the whole-region traditional approach. That the errors of the cutoff estimates above increasing threshold densities of shrimp were not

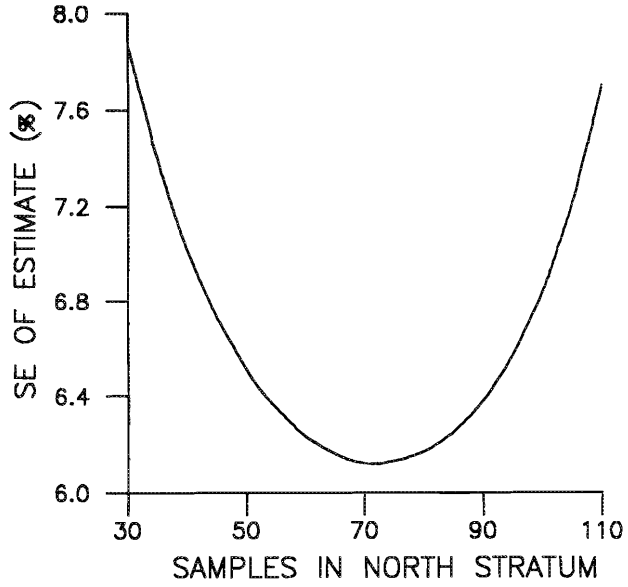

FIG. 7. Standard error of the global estimate (SE = σ_D/total biomass) as a function of the number of samples in the north stratum for the two-strata model for a fixed total number of 137 samples.

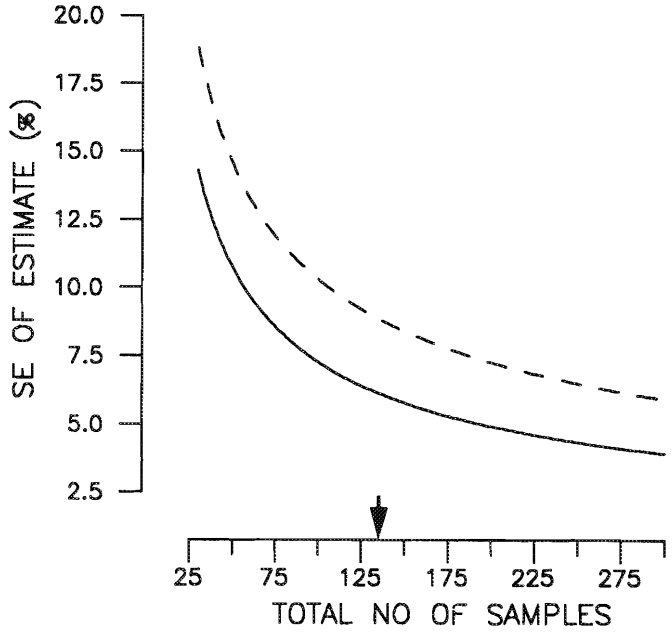

FIG. 8. Standard error of the global estimate (SE = σ_D/total biomass) as a function of the total number of samples for the two-strata geostatistical (solid line) and classical (broken line; σ^2 = variance of samples) models, for a fixed optimal allocation of 53% of samples in the north stratum from Fig. 7. The arrow indicates the present total number of samples.

larger than the error of the global estimate, but slightly smaller, indicates that the samples were adequately distributed relative to the density of shrimp.

These geostatistical errors of the global and cutoff estimates must be interpreted with caution. They are representing the error of the estimate according to the chosen structure model. If this model adequately describes the spatial structure, then the computed errors of the estimates are appropriate. The cross-validation of our optimal kriging models showed that they reproduced the true sample values with a correlation of about 0.70. This is the minimum correlation we may therefore expect from the kriging estimates because they were computed with closer information points (always smaller than the distance between samples) than the cross-validation estimates, where the information points were more distant (on average equal to distance between samples). An increase of the sample density would increase this correlation of the spatial model and reduce the errors of the global and cutoff estimates computed from equation (11) by reducing the surface of influence of the samples. At these small scales, the presence of a structural variability and its consideration in computing the errors of the estimates in equation (11) would result in a higher error reduction than it would in the random case, which is assumed in classical statistics. This advantage of taking into account the spatial structure was also clearly depicted here; kriging always produced much better estimations (with smaller errors) than the alternative polygonal method, which did not involve a structure function.

Sampling Optimization

Stationary geostatistics tells us that the variance of the estimates of a spatially structured variable will be minimized by sampling over a regular grid (Hughes and Lettenmaier 1981; Munoz-Prado et al. 1989). This can be seen from equations (9) and (10) where the size and geometry of the area of influence of the samples are involved in the computation of the variance of the global estimate. The precision of this estimate will therefore be partly determined by the spacing between samples. When geometric anisotropy is present, the grid should be stretched in the direction of the smallest variability by the ratio of the slopes of the directional variograms computed for the directions of minimum and maximum variances. This distribution of samples also ensures obtaining the optimal map of the variable over the whole studied region.

These theoretical guidelines are true only if the structure function is adequate and perfectly known. This is not the case in practice, since the variograms are unknown and they must be computed from the samples. Unless the grid is very tight, the definition of the structure function at small scales will not be precise enough to separate the structural from the random variability. It seems therefore worthwhile to direct a proportion of the samples to increase the local density of the grid (Warrick and Myers 1987; Fortin et al. 1989). This would enhance our knowledge of the small-scale variability and better define the variogram at small distance classes, in order to obtain a more precise estimate of the global variance. When the proportion of the structural variability at the scale corresponding to the separation between samples is small compared with the unresolved variability, the use of a regular grid instead of random sampling will not significantly improve the precision of the global estimate. It will, however, offer the advantage that the map will have a uniform level of uncertainty. Drawbacks of a regular grid are the risks of aliasing, when cyclic variations are interfering with the sampling frequency, and of having an improper sampling step relative to the variability of the samples.

Results showed that stratification minimized the error of the global estimate. Could we reduce again this error by changing the allocation of the 137 samples between the two strata? Using equation (11), the SE of the global estimate was computed for different allocations of the samples between the two strata (Fig. 7). The minimum SE is obtained when 72 samples (53%

of total) are placed in the north stratum and 65 in the south stratum. This changes the allocation of 21 samples in favor of the north stratum compared with the present case. The SE of the global estimate associated with this reallocation is 6.1%, which is a slight reduction compared with the present value of 6.5%.

The effect of increasing or reducing the total number of samples of this optimal allocation is depicted in Fig. 8. Under the classical hypothesis of random distribution (pure nugget variogram), the SE of the global estimate decreases as a function of $n^{-0.5}$. It can be shown that SE decreases as $n^{-0.75}$ for a linear variogram and, provided n is sufficiently large, as $n^{-1.0}$ for a Gaussian variogram. For mixed models, this decreasing function will depend on the relative importance of each variogram component. Here, because of the high proportion of the unresolved nugget variability, the SE of the global estimate decreases with a rate close to the classical case. It was, however, always significantly lower than the SE computed under the hypothesis of randomness. Doubling the sampling effort would result in a reduction, by one third, of the SE of the estimate, which would become 4.1%. Decreasing the number of samples by one half would give a SE of 9.0%, which is a relative increase of about 50%.

Sampling in areas of high uncertainty could be reinforced using the error maps computed. Figures 6A and 6B reflect the sampling density in the stratum or region considered and point out areas where the sampling could be increased to get a better local estimate considering the local variability, independent of the value of this estimate. As expected from equation (5), the stratified scheme resulted in lower errors in the south stratum in response to its lower variance. In general, both maps point out the same areas of high errors in this stratum. The error depicted in Fig. 6C takes into account not only the sample density and local variability but also the values of the local estimates. It thus indicates the areas where it is worth increasing the sample density for estimation of the global biomass because they are sufficiently rich and they have a high variance or, conversely, decreasing the sampling density because of a uniform low abundance. For example, more samples should be collected in the south, at the junction of the gulf and the estuary, to delineate more accurately the extent of the local rich patch (Fig. 5D and 6C). Similarly, the sampling effort in the southeast could be relaxed because of its low variance.

To conclude on the geostatistical estimation of the biomass of the northern shrimp in the Gulf of St. Lawrence, the nonstationarity and the high unresolved small-scale variability are two questions that need further research in order to more precisely define the spatial model. The use of more exhaustive data sets and data from repeated surveys would surely bring more information on the shape of the probabilistic model driving the spatial organization of the shrimp. Despite this limitation the present results show that geostatistics can easily provide various precise estimates for spatially autocorrelated biological resources. Besides, it provides extra information on their spatial organization, which can hardly be obtained otherwise, and which is most helpful for understanding the ecology of the studied species and, hence, for improving their management.

Acknowledgments

We are grateful to Louise Savard for the planning and realization of sampling and to Dominique Gascon, Sylvain Hurtubise, Louis Gosselin, and Benoit Mercille for the collection of samples. We also thank anonymous referees for helpful comments.

References

APOLLONIO, S., D. K. STEVENSON, AND E. E. DUNTON. 1986.Effects of temperature on the biology of the northern shrimp, *Pandalus borealis* in the Gulf of Maine. U.S. Dep. Commer. NOAA Tech. Rep. NMFS No. 42: 22 p.

ARMSTRONG, M., AND A. BOUFASSA. 1988. Comparing the robustness of ordinary kriging and lognormal kriging: outlier resistance. Math. Geol. 20: 447–457.

ARMSTRONG, M., D. RENARD, AND P. BERTHOU. 1989. Applying geostatistics to the estimation of a population of bivalves. ICES C.M.1989/K37: 22 p.

BARR, L. 1970. Diel vertical migration of *Pandalus borealis* in Kachemak bay, Alaska. J. Fish. Res. Board Can. 27: 669–676.

BOX, G. E. P., AND D. R. COX. 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B 26: 211–243.

BURGESS, T. M., AND R. WEBSTER. 1980a. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. J. Soil Sci. 31: 315–331.

1980b. Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging. J. Soil Sci. 31: 333–341.

BURGESS, T. M., R. WEBSTER, AND A. B. MCBRATNEY. 1981. Optimal interpolation and isorithmic mapping of soil properties. IV. Sampling strategy. J. Soil Sci. 32: 643–659.

CLARK, I. 1979. Practical geostatistics. Elsevier, New York, NY. 129 p.

CLIFF, A. D., AND J. K. ORD. 1981. Spatial processes: models and applications. Pion Ltd., London, U.K. 266 p.

CLIFFORD, P., S. RICHARDSON, AND D. HEMON. 1989. Assessing the significance of the correlation between two spatial processes. Biometrics 45: 123–134.

COCHRAN, W. 1977. Sampling techniques. John Wiley and Sons, New York, NY. 413 p.

CONAN, G. 1985. Assessment of shellfish stocks by geostatistical techniques. NAFO SCR Doc. 85/108: 19 p.

DAVID, M. 1988. Handbook of applied advanced geostatistical ore reserve estimation. Elsevier, Amsterdam, The Netherlands. 216 p.

FORTIN, M.-J., P. DRAPEAU, AND P. LEGENDRE. 1989. Spatial autocorrelation and sampling design in plant ecology. Vegetatio 83: 209–222.

FROIDEVEAUX, R. 1984. Precision of estimation of recoverable reserves: the notion of conditional estimation variance, p. 141–164. *In* G. Verly et al. [ed.] Geostatistics for natural resources characterization. NATO ASI Series C, Vol. 122. Reidel, Dordrecth, The Netherlands.

HAURY, L. R., J. A. MCGOWAN, AND P. H. WIEBE. 1978. Patterns and processes in the time–space scales of plankton distributions, p. 277–327. *In* J. H. Steele [ed.] Spatial pattern in plankton communities. Plenum Press, New York, NY.

HUGHES, J. P., AND D. P. LETTENMAIER. 1981. Data requirements for kriging: estimation and network design. Water Resour. Res. 17: 1641–1650.

ISAAKS, E. H., AND R. M. SRIVASTAVA. 1989. Applied geostatistics. Oxford University Press, New York, NY. 561 p.

JOURNEL, A. G. 1980. The lognormal approach to predicting local distributions of selective mining unit grades. Math. Geol. 12: 285–303.

JOURNEL, A. G., AND CH. J. HUIJBREGTS. 1978. Mining geostatistics. Academic Press, New York, NY. 600 p.

LEGENDRE, P., AND M.-J. FORTIN. 1989. Spatial pattern and ecological analysis. Vegetatio 80: 107–138.

LEGENDRE, P., AND M. TROUSSELLIER. 1988. Aquatic heterotrophic bacteria: Modeling in the presence of autocorrelation. Limnol. Oceanogr. 33: 1055–1067.

LORING, D. H., AND D. J. G. NOTA. 1973. Morphology and sediments of the Gulf of St. Lawerence. Bull. Fish. Res. Board Can. 182: 1–147.

MACKAS, D. L., K. L. DENMAN, AND M. K. ABBOTT. 1985. Plankton patchiness: biology in the physical vernacular. Bull. Mar. Sci. 37: 652–674.

MACKETT, D. J. 1973. Manual of methods for fisheries resource survey and apparaisal. F.A.O. Fish. Tech. Rep. No. 124: 29 p.

MARBEAU, J. P. 1976. Géostatistique forestière: état actuel et développements nouveaux pour l'aménagement en forêt tropicale. Ph.D. thesis, École Nat. Sup. Mines Paris, Fontainebleau. 211 p.

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, facs. 5, École Nat. Sup. Mines Paris, Fontainebleau. 211 p.

MORAN, P. A. P. 1950. Notes on continuous stochastic phenomena. Biometrika 37: 17–23.

MUNOZ-PARDO, J., J. F. BOULIER, AND M. VAUCLIN. 1989. Étude de l'échantillonnage d'un phénomène bidimensionnel par simulation d'une fonction aléatoire, p. 837–850. *In* M. Armstrong [ed.] Geostatistics. Vol. 2. Kluwer Academic Publishers, The Netherlands.

NICOLAJSEN, A., AND G. CONAN. 1987. Assessment by geostatistical techniques of populations of Iceland scallop (*Chlamis islandica*) in the Barent sea. I.C.E.S. CM 187/K:14: 18 p.

PETITGAS, P., AND J.-C. POULARD. 1989. Applying stationary geostatistics to fisheries: a study on hake in the bay of Biscay. ICES Demersal Fish. Comm. C.M./G.62: 21 p.

RENDU, J.-M. 1979. Normal and lognormal estimation. Math. Geol. 11: 407–422.

SAMPER CALVETE, F. J., AND S. P. NEUMAN. 1989. Geostatistical analysis of groundwater quality data from the Madrid basin using adjoint state maximum likelihood cross-validation, p. 725–736. *In* M. Armstrong [ed.] Geostatistics, Vol. 2. Kluwer Academic Publishers, The Netherlands.

SAVARD, L. 1989. Evaluation des stocks de crevette (*Pandalus borealis*) du Golfe du Saint-Laurent. Comité Sci. Consul. Pêches Can. Atl. (CSCPCA) Doc. Rech. 89/07: 70 p.

SIMARD, Y., P. BRUNEL, AND J. LACELLE. 1990. Distribution and growth of pre-recruit cohorts of the northern shrimp (*Pandalus borealis*) in the western Gulf of St. Lawrence as related to hydrographic conditions. Can. J. Fish. Aquat. Sci. 47: 1526–1533.

SMITH, S. J., AND R. K. MOHN. 1987. Considerations on the representation and analysis of a spatially aggregated resource: Georges Bank scallops. I.C.E.S. C.M. 1987/K: 26: 19 p.

SOKAL, R. R. 1986. Spatial data analysis and historical processes, p. 29–43. *In* E. Diday et al. [ed.] Data analysis and informatics. IV. Proceedings of the Fourth International Symposium on Data Analysis and Informatics, Versailles France, 1985. North-Holland, Amsterdam, The Netherlands.

SOKAL, R. R., AND F. J. ROHLF. 1981. Biometry the principles and practice of statistics in biological research. W. H. Freeman and Co., San Francisco, CA. 859 p.

STEELE, J. H. 1978. Spatial pattern in plankton communities. Plenum Press, New York, NY. 470 p.

STOLYARENKO, D.A. 1986. Data analysis of trawl shrimp survey with spline approximation of stock density. ICES C.M. 1986/K:25: 15 p.

TRITES, R. W., AND A. WALTON. 1975. A Canadian coastal sea — the Gulf of St. Lawrence. Bedford Inst. Oceanogr. Rep. Ser. BI-R-75-15: 1–29.

WARRICK, A. W., AND D. E. MYERS. 1987. Optimization of sampling locations for variogram calculations. Water Resour. Res. 23: 496–500.

YOST, R. S., G. UEHARA, AND R. L. FOX. 1982. Geostatistical analysis of soil chemical properties of large land areas. II. Kriging. Soil Sci. Soc. Am. J. 46: 1033–1037.