NOT TO BE CITED WITHOUT PRIOR
REFERENCE TO THE AUTHOR(S)

Northwest Atlantic Fisheries Organization

Serial No. N1084

NAFO SCR Doc. 85/108

SCIENTIFIC COUNCIL MEETING - SEPTEMBER 1985

Assessment of Shellfish Stocks by Geostatistical Techniques

by

Gerard Y. Conan

Department of Fisheries and Oceans, Fisheries Research Branch
Gulf Region, Marine Biology Research Center at Universite de Moncton
Moncton, New Brunswick, Canada E1A 3E9

SPECIAL SESSION ON BIOLOGICAL SURVEYS

INTRODUCTION

Shellfish species occupy a very wide range of habitats and have very original types of life cycles from benthic sedentary ocean quahogs who may live several hundred years to pelagic schooling squids who live only one to two years and benthic decapod crustaceans either sedentary or which may crawl over several hundred miles during their life time. In a first instance, the only common feature among shellfish species seems to be diversity and it seems that custom made assessment statistical tools may be required in most cases (Conan, 1984). There is however a strong temptation to use standard tools of "proven efficiency" imported from finfish population biology.

Not withstanding the diversity of shellfish species, there is one characteristic which is shared by several species: the relative temporal scales of sampling surveys and movement of the individuals. Many shellfish species will have limited movements or migrations during a sampling survey. In finfish stocks it is frequently possible and generally assumed that the individuals redistribute themselves over the fishing grounds during the survey. The exact location of the sampling points is therefore of secondary importance, and it is implicitely assumed that the results for one sample are statistically independent from those of the next one: 1) Within a stratum all samples may share a common mathematical expectation (mean) but the random residual terms (the errors) are independent from one sample to the next. In order to enhance this independance it is generally recommended to randomize the spatial distribution of sampling points. 2) The sampling procedure has negligeable effects on the population. The removal of individuals does not affect the total number of individuals present. One may eventually sample twice in the same location. The removals made by the first sample have no effects on the second sample because fish will have redistributed over the sampling area. 3) Within a stratum the mathematical expectation in one location is the same as in another one. There are no geographic trends.

Actually these three assumptions may never be fulfilled, even for finfish. However due to swift reorganizations of the spatial distribution of the individuals it is difficult to detect any structures during a survey, other than broad geographic changes in densities, stable in time, which may be dealt with by stratification. Fisheries biologist using echosounding as a survey tool have developped sampling techniques operating fast enough to detect spatial structure of populations (schools of fish) and are familiar with the problem of non independance of samples. Samples are "autocorrelated" within a certain radius.

Many species of benthic invertebrates are sedentary, at least over the fishing season and it is no longer possible to assume that they redistribute over the fishing grounds. Fishermen will harvest microareas one by one, redistributing their effort over the fishing grounds, over the fishing season or over a series of years. If the relative geographic distributions of fishing effort and harvested populations are not taken in account, the sustainable production of the stock may be considerably overestimated. The landings appear stable until the whole stock has been (over) exploited and the fishery crashes. The geographic distribution of effort can be assessed on the basis of log book data or by direct means such as aerial surveys (Conan & Maynard, 1983; Pringle & Duggan, 1983). The assessment of spatial structure of population abundance requires a non traditional approach. The final step of modelling the exploitation of such "disaggregated" stocks has already been considered by several authors (Sluckzanowski, 1984; Mohn et al.,1984; Murawski and Fogarty, 1984). An important step consists in taking in account the behavior of the fishermen who concentrate their fishing effort on a rich area, until their costs about equal their benefits and then move to another area.

Quite clearly an assessment of overall abundance (biomass) is not sufficient, even if it is unbiased and very precise. It is not sufficient either, to model the probability distribution of the samples in the traditional way: a regular distribution by the positive binomial, a random distribution by the Poisson, an aggregated distribution by the negative binomial (Elliot, 1977 for review). One needs statistical tools specialy designed for analyzing the spatial structure of the populations.

The problems encountered with sedentary benthic marine invertebrates are somehow similar to those encountered in assessments of forestry resources (Matern, 1960) or in the mining industry (David, 1977). Ecologists in recent years have moved from the traditional approach of modelling probability distributions of independent samples to the analysis of spatial structures and of autocorrelations between samples of known geographic position. The most thorough development for this approach has been in mining. (Matheron 1968, 1969, 1971), Delfiner (1976), David (1977), Clark (1979), have described a theoretical background for a "theory of regionalized variables" for the "geostatistics" of mining resources. I have reviewed some of the techniques currently used in geostatistics and checked for possible applications in fisheries science. I applied one of these techniques "Kriging" to actual data from an exploited scallop bank in Northumberland strait (Canada) and compared the results with those of a traditional biomass survey.

MATERIAL AND METHODS

Data

The data was provided by Dr. Jean Worms (Marine Biology Research Center). It was obtained during a survey of Sea Scallop *Placopecten magellanicus* on the banks of Pictou and Indian Rock in Southern Northumberland Strait. One hundred and nineteen short tows of Digby scallop dredge over a rectangle of 10 by 25 nautical miles were obtained along transects crossing the banks. The tows being very short are assumed to provide punctual information on density. The density is expressed in tons per square km (Fig. 1).

Standard statistical treatment

The standard approach assumes no autocorrelation between samples $Z(x)$. We shall simply compute a mean density $\bar{Z} = \frac{\Sigma Z(x)}{N}$ and a variance $S^2 = (\Sigma Z^2(x) - (\Sigma Z(x))^2/N)/(N - 1)$. From the area covered, A, the mean density $\bar{Z}$ and its standard deviation $S(\bar{Z}) = S/(\sqrt{N})$ we may provide an estimate of the standing stock (biomass) and of its 95% confidence limits. We assume that the distribution of the mean Z is normal because we have more than 30 samples (central limit theorem). The biomass B is the product of the average density $\bar{Z}$ by the constant A, its variance is therefore $S^2(B) = A^2 S^2(\bar{Z})$. The 95% confidence limits are $\pm$ 1.96 S(B). We also provide an histogram of the sampled estimates of densities.

Geostatistical treatment

The first step is to analyze the autocorrelation between data points by calculating an experimental variogram $\hat{\gamma}(h)$ where h represents the distance between points.

$$\hat{\gamma}(h) = 1/(2N) \sum_{i=1}^{N} [Z(x_i) - Z(x_i + h)]^2$$

N is the number of couples of points separated by a distance h. The variogram, when plotted against h provides a first insight on the spatial structure of the data. A most common shape is a curve starting from a non zero value in ordinate the "nugget effect" which represents the variability between replicate samples taken at the same site. The nugget effect may be due to a microstructure, measurement or location errors. The slope of the curve progressively decreases towards an asymptotic value in ordinate, the "sill" which is reached when the samples become fully independent after a value of h called the "range". Beyond the range there is no autocorrelation effect.

A very general model used for the variogram is the "spherical" model:

$$\gamma(h) = C\,[3/2\frac{|h|}{a} - 1/2\frac{|h|^3}{a^3}] \qquad |h| \leq a$$

$$\gamma(h) = C \qquad |h| > a$$

a is the range, C is the sill

In fisheries biology the range could represent the size of "patches" in an aggregated distribution modelled by the negative binomial for instance. In the most simple case, if the mean of the random variable $Z(x)$ does not vary for any point x, neither does the variance $\sigma(Z(x))$. The experimental variogram $\hat{\gamma}(h)$ is an estimate of the variogram $\gamma(h) = \frac{1}{2}\,\sigma^2[Z(x + h) - Z(x)]$. If there is no autocorrelation (no covariance) between samples (when the sill is reached for instance) we may write:

$$\gamma(h) = \tfrac{1}{2}[\sigma^2(Z(x + h)] + \sigma^2[Z(x)] = \frac{2\sigma^2[Z(x)]}{2} = \sigma^2[Z(x)]$$

The departure from these simple conditions will allow to draw some inference about the spatial structure of the data. For instance in case of a "drift", when the mean varies according to a directional trend

$$E[Z(x + h) - Z(x)]^2 =$$

$$\sigma^2[Z(x + h) - Z(x)]^2 \quad \text{variance}$$

$$+$$

$$(E[Z(x + h) - Z(x))]^2 \quad (\text{Bias})^2$$

$$\hat{\gamma}(h) = \gamma(h) + \tfrac{1}{2}[\bar{Z}(x + h) - \bar{Z}(x)]^2$$

There will be no asymptotic value in ordinate, no "sill".

Once a "drift" has been identified it may be modelled by a sum of monomials (functions of h of increasing degree) and filtered out. This is performed by "universal kriging", a technique that I have not used in the present, simple, case.

Once the spatial structure has been analyzed and modelled we shall proceed by taking the best linear unbiassed estimator (B.L.U.E.) of the quantity y_o which is a linear function of the variable $Z(x)$.

For instance:

1) $y_o = Z(x_o)$ the value taken by Z at a point unsampled $x = x_o$, i.e. the quantity of fish at location x_o

2) $y_o = \frac{1}{v}\int_{v_o} Z(x)dx$ the average value of $Z(x)$ over the subarea v_o of area v centered at $x = x_o$

i.e. the density of fish in a subarea: is it worth harvesting?

3) $y_o = \frac{1}{V} \int_V Z(x)dx$ the average value of $Z(x)$ over the whole area V

i.e. the average density of fish over the whole fishing area.

The B.L.U.E. will consist to give a weight to the different available samples so as to obtain an estimator with minimum squared error. For instance in case 3, for calculating an average density from uncorrelated "random" samples, the traditional procedure would consist in giving a weight of 1/N to each sample and summing. Now that we know that the samples may not be independent, i.e. may be autocorrelated, we shall allocate the adequate weight it really deserves to each of the samples according to its position in the spatial distribution in order to provide an adequate estimate of the average density by summation.

Cases 1, 2, and 3 can be pooled in a single problem by taking y_o as an average value over an unspecified domain V.

$$y_o = \frac{1}{V} \int_V Z(x)dx$$

To estimate y_o we consider a weighted average of the data.

$$y^*_o = \sum_{i=1}^{N} \lambda_i Z(x_i)$$

where y^*_o is a Kriging estimate. We determine the weights λ_i so that y_o^*:

a) is unbiassed: $E(y^*_o - y_o)^2 = 0$

b) has a minimum squared error

$$E(y^*_o - y_o)^2 \text{ minimum}$$

Since $E(y^*_o - y_o) = 0$, $E(y^*_o - y_o)^2$ is also the variance of the kriging error.

In the "stationary case" where $E(Z(x_i)) = m$ for all i's and $E(y_o) = m$ condition a entails:

$$E(\sum_{i=1}^{N} \lambda_i Z(x_i) - y_o) = 0$$

or

$$\sum_{i=1}^{N} \lambda_i m - m = m(\sum_{i=1}^{N} \lambda_i - 1) = 0$$

i.e. the weights must add up to 1

The variance of the error can be expressed in terms of the model used for the variogram.

$$E(y^*_o - y_o)^2 = S^2(y^*_o - y_o) =$$

$$2\sum_{i=1}^{N} \lambda_i \quad \bar{\gamma}(x_i, V) - \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \gamma(x_i - x_j) - \bar{\gamma}(V,V)$$

$\bar{\gamma}(x_i,V)$ is the average of the variogram between x_i and the area V.

$$\bar{\gamma}(x_i,v) = \frac{1}{V} \int_V \gamma(x_i - x)dx$$

$\gamma(V,V)$ is the average of the variogram between any two points x and x' sweeping independently throughout the domain (area) V

$$\bar{\gamma}(V,V) = \frac{1}{V^2} \int_V \int_V \gamma(x - x')dx\, dx'$$

The variance $S^2(y^*_o - y_o)$ is minimized under the constraint that $\sum_i \lambda_i = 1$ by setting partial derivatives for each λ_i equal to 0. We obtain a system of N equations in λ_i, i=1...N that can be solved for each λ_i the "weight" of each point i in the system.

The minimum of the variance or kriging variance is:

$$\sigma_k^2 = \sigma^2(y^*_o - y_o) = \sum_{i=1}^{N} \lambda_i \bar{\gamma}(x_i,V) - \bar{\gamma}(V,V) + \mu$$

When the variogram is a pure nugget effect (no autocorrelation) $\lambda_i = 1/N$.

If V is a point, $\bar{\gamma}(x_i,V) = \gamma(x_i - x_o)$ and $\bar{\gamma}(V,V) = \gamma(0) = 0$

In case of a drift i.e. a systematic increase or decrease in one direction, each point x has its own mean, the assumption $E[Z(x)] = cte$ is violated and a we must introduce more variables. We may model the drift m(x) as:

$$m(x) = \sum_{u=1}^{k} a_l f^u(x)$$

I have not pursued this approach in the present case.

I have used the kriging procedure to calculate average densities inside each "block" of a high definition grid. Knowing the area of each block and its density I have calculated the overall biomass. I have also used the kriging procedure for drawing contour lines of equidensity within the fishing area and generating a three dimensional representation of the data.

Kriging computations were run on an HP9845 B using a software package provided by Geomin.

RESULTS

Standard statistics

For 108 samples the mean density over the banks was 2.98 tons per km^2 with a standard deviation among samples of 1.9824, and confidence limits for the mean of $\pm$ 0.37. The area of the banks was estimated as 177.63 km^2 and the standing biomass as 529.33 tons live weight $\pm$ 66.41. A frequency distribution of the densities in the samples is provided in figure 2.

Variogram

A representation of the experimental variogram $\hat{\gamma}(h)$ is provided in figure 3. A spherical model
$\gamma(h) = C_o + C(1.5|h|/a - 0.5(|h|/a)^3)$ with parameter values:

nugget value	$C_o = 1.0094$
sill - nugget	$C = 3.0244$
range	$a = 2.6$

fitted the data quite well. The presence of a well defined sill at 4.0338 indicates that within the banks surveyed there was no detectable drift and in first approximation we may retain the "intrinsic hypothesis" with constant mean:

$$E[Z(x + h) - Z(x)] = 0$$

$$\sigma^2[Z(x + h) - Z(x)] = 2\gamma(h)$$

There is no correlation between the samples beyond a distance of 2.6 nautical miles (the range). The variance of the population of sample points beyond the "zone of influence" of a sample is $\sigma^2 = \gamma(h) = 3.0244$.

There is a distinct "nugget effect" which reveals a small scale structure, may be a microstructure smaller than the dredge haul or a measurement error of the actual length of tow.

There is a small "nested structure" indicative of two different scales of variation, one with a range of 1.5, one with a range of 3.0.

Kriging estimate of biomass

The fishing area was divided in a fine mesh grid containing

1066 cells, 890 of which were defined i.e. taken in account for kriging. Each cell had a surface of 1.1147 km^2. An average density was calculated by the kriging program for each defined cell. The overall surface of the kriged area, including samples with zero densities and unsampled locations was 992.10 km^2. The arithmetic mean of the average cell densities was 0.5886 tons per km^2. The overall biomass was therefore estimated as 583.95 tons live weight.

Isocontours of density estimates

Isocontours were drawn around the kriging estimates of average densities in the cells, in two dimensions (figure 4) and in three dimensions (figure 5). The dual structure of the fishing bank appears very neatly. This is already evident in the plot of raw data (Fig. 1). The isocontour structure indicates the presence of distinct nodes in the spatial distribution within each of the two banks. These nodes could not be easily identified in the raw data (Fig. 1).

DISCUSSION

The basic principal in geostatistics is that the Best Linear Unbiased Estimator of any linear function y_o of variable $Z(x)$ sampled is not obtained by giving equal weights to all observations as is implicitely done in standard statistics. The problem of autocorrelation between samples at close distance in space is solved by attributing unequal weights to each of these.

$$y_o = \sum_{i=1}^{N} \lambda_i Z(x_i)$$

The weights λ_i are determined so that y^*_o is unbiassed:

$E(y^*_o - y_o) = 0$, usually when $\sum_{i=1}^{N} \lambda_i = 1$ and has minimum squared error: $E(y^*_o - y_o)^2$ minimum.

It is only under particular conditions (no covariance between samples) that average value of the variable studied over the geographic domain will be an arithmetic mean i.e. $\lambda_i = 1/N \quad i=1..N$. In most cases the arithmetic mean is a biased estimate. Actually the kriging theory was developed after it was discovered that in mining the standard statistical prediction of the ore content of a block (i.e. a subarea) was usually an overestimate. The same observation has also been made in forestry (Narboni, 1979). In fisheries we seldom have the possibility to compare estimates with actual values, however the same bias may very well be expected if the samples are spatially autocorrelated.

The geostatistical approach allows by to pass the constrain of "random" sampling used in standard statistics to avoid autocorrelation between samples. In geostatistics it can be demonstrated that random sampling is not the best approach. If we exploit the spatial correlation instead of avoiding them the theory shows that an unbiased result can be obtained whatever sampling strategy is used, random, stratified random or regular. In terms of precision of the estimates the variance of the averaged estimates turn out to be greater for random sampling than for stratified random sampling, but the smallest variance is for a regular sampling.

The weakest point in the kriging methodology is probably the modelling of the variogram and of the drift. The computations are based on models fitted to the data and not to the data itself. If the models chosen are inadequate the predictions may be inadequate. An empirical robustness survey of the kriging predictions as a function of deviations in the parameter values in the model and of the model itself would be required for fisheries assessments.

In the case study the conditions were particularly simple: absence of a detectable drift, good coverage of the fishing ground by the sampling survey, good fit of the spherical model to the experimental variogram. In further development the spatial analysis may require much more complex techniques:

1) a directional computation of the variogram to detect anisotropies in the orientation of covariance between samples. If the underlying structures are oriented preferentially in a geographic direction, in fisheries the "clumps" or "aggregates" of fish, the covariance between points will extend over a wider range in that direction.

2) The fitting of a "drift" function to filter out the trends. Such drifts can be expected in fisheries over gradients of depth, temperature or sediment quality for instance.

3) The assessment of fisheries in which the sampling gear is of mixed type. Geostatistics cope with a similar problem by a technique called "Cokriging".

These techniques involve matrix manipulations for solving sets of linear equations and require a good software back-up. Such software is available on microcomputers such as Hewlett Packard models 9000 series 200, 300, 500 and on HP9845[(*)].

The tools developed for the mining industry need to be adapted for fisheries assessments. In mining, emphasis seems to be set on estimates of the grade of unexploited blocks (subareas) in order to determine whether they are worth exploiting or not. In fisheries assessments of sedentary species this could be an interesting approach, however the usual goal is a standing biomass estimate for a given area or subarea and confidence limits on this estimate. Such computations can be obtained directly by kriging (estimator and standard deviation) but were not yet available in the software I used. I obtained an approximate value for biomass by taking the arithmetic mean of kriged estimates of densities within a high resolution grid of 890 cells, therefore neglecting possible covariance effects between cells. In respect of kriging theory this estimate is likely to be biased. It compares fairly well with the simple estimate obtained by standard statistics (584 vs 529 tons).

Very little, if any work has been conducted on spatial autocorrelation and structure of populations harvested by fisheries except in echo sounding surveys and it is not possible as yet to conclude whether geostatistical techniques such as kriging or contouring after spline (Dubrule, 1984) smoothing of sampled data will provide considerable improvement over more standard and simple techniques of biomass estimates. However geostatistics is an approach worth investigating since it provides a theoretical back ground for efficiently using data from non random sampling (true random samples are very seldom available in fisheries surveys, and never in log book data).

---

*The software used in the present case was obtained from Geomin Computer Services Corporation, 408 Kapilano 100, West Vancouver, B.C., V7T 1A2, Canada. (604) 922-9367.

Geostatistics also provide outputs not generally available from standard statistics such as charts of densities, charts of error of estimates (allowing subsequent improvement in sampling), unbiased direct estimates of standing stock over a whole area as well as predictions over yet unexploited subareas.

ACKNOWLEDGMENTS

Mr. Simon W. Houlding, M.Sc., P. Eng. and the staff from Geomin provided free access to their kriging software for analysis of the data and efficiently helped in adapting the software. Mr. Elmer Wade, P. Eng. from MBRC spent many hours assisting me in experimenting the software. Dr. Jean Worms from MBRC provided excellent scallop survey data for running the trial. To gain some insight in the kriging theory I greatly benefited from notes provided to me by a student who attended the basic introduction to geostatistics course given by P. Delfiner at the summer school of Fontainebleau, Ecole des Mines de Paris, France.

REFERENCES

CLARK, I., 1979. Practical geostatistics. Applied Science Publishers, Barking, Essex. England.

CONAN, G. Y., 1984. Do assumptions commonly used for modelling populations of finfish apply to shellfish species? ICES, Shellfish Committee C.M. 1984/K:49. 21 p.

CONAN, G. Y., and D. R. MAYNARD, 1983. Aerial survey of spatial distribution of effort in the lobster fishery of Southern Gulf of St. Lawrence. ICES C.M. 1983/K:13, 13p.

DAVID, M., 1977. Developments in Geomathematics 2. Geostatistical ore reserve estimation. Elsevier Scientific Publishing Company Amsterdam. 364 pp.

DELFINER, P., 1976. "Linear estimation of non-stationary spatial phenomena". In M. Guarascio et al., Eds., Advanced Geostatistics in the Mining Industry., Reidel, Dordrecht, The Netherlands: 49-68.

DUBRULE, O., 1984. Comparing splines and kriging. Computers & geosciences. 10, 2-3: 327-338.

ELLIOT, J.M., 1977. Some methods for the statistical analysis of samples of benthic invertebrates. Freshwater Biological Association U.K. Scient. Publ. 25:160p.

MATERN, B., 1960. Spatial variation Medd. Skogsforkning inst., 49, 144 pp.

MATHERON, G., 1965. Les variables régionalisées et leur estimation. Masson et Cie, Paris, 305 pp.

MATHERON, G., 1968. Osnovy Prikladnoĭ Geostatistiki, Mir Editions, Moscow, 408 pp.

MATHERON, G., 1969. Le krigeage universel. Les cahiers du centre de morphologie mathématique, n° 1, Ecole Nat. Sup. des Mines de Paris, 82 pp.

MATHERON, G., 1971. The theory of regionalized variables and its applications. Les cahiers du Centre de Morphologie Mathématique, n° 5, Ecole Nat. Sup. des Mines de Paris, 212 pp.

MOHN, R.K., G. ROBERT, and D.L. RODDICK, 1984. Status and harvesting strategies for Georges Bank scallop stock (NAFO SA 5Ze). ICES C.M. 1984/K:15.

MURAWSKI, S.A., and M.J. FOGARTY, 1984. A spatial yield model for bivalve populations accouting for density-dependent growth and mortality. ICES C.M. 1984/K:26.

NARBONI, Ph., 1979. "Application de la méthode des variables régionalisées à deux forêts du Gabon" Statistical report # 18, Centre technique Forestier Tropical, 94130 Nogent-sur-Marne, France, 51 pp.

PRINGLE, J. D., and R. E. DUGGAN, 1983. A remote sensing technique for quantifying lobster fishing effort. Can. Tech. Rep. Fish. Aquat. Sci. 1217:16p.

SLUCZANOWSKI, P.R., 1984. A management orientated model of an abalone fishery whose substocks are subject to pulse fishing Can. J. Fish. Aquat. Sci. 41:1008-1014.

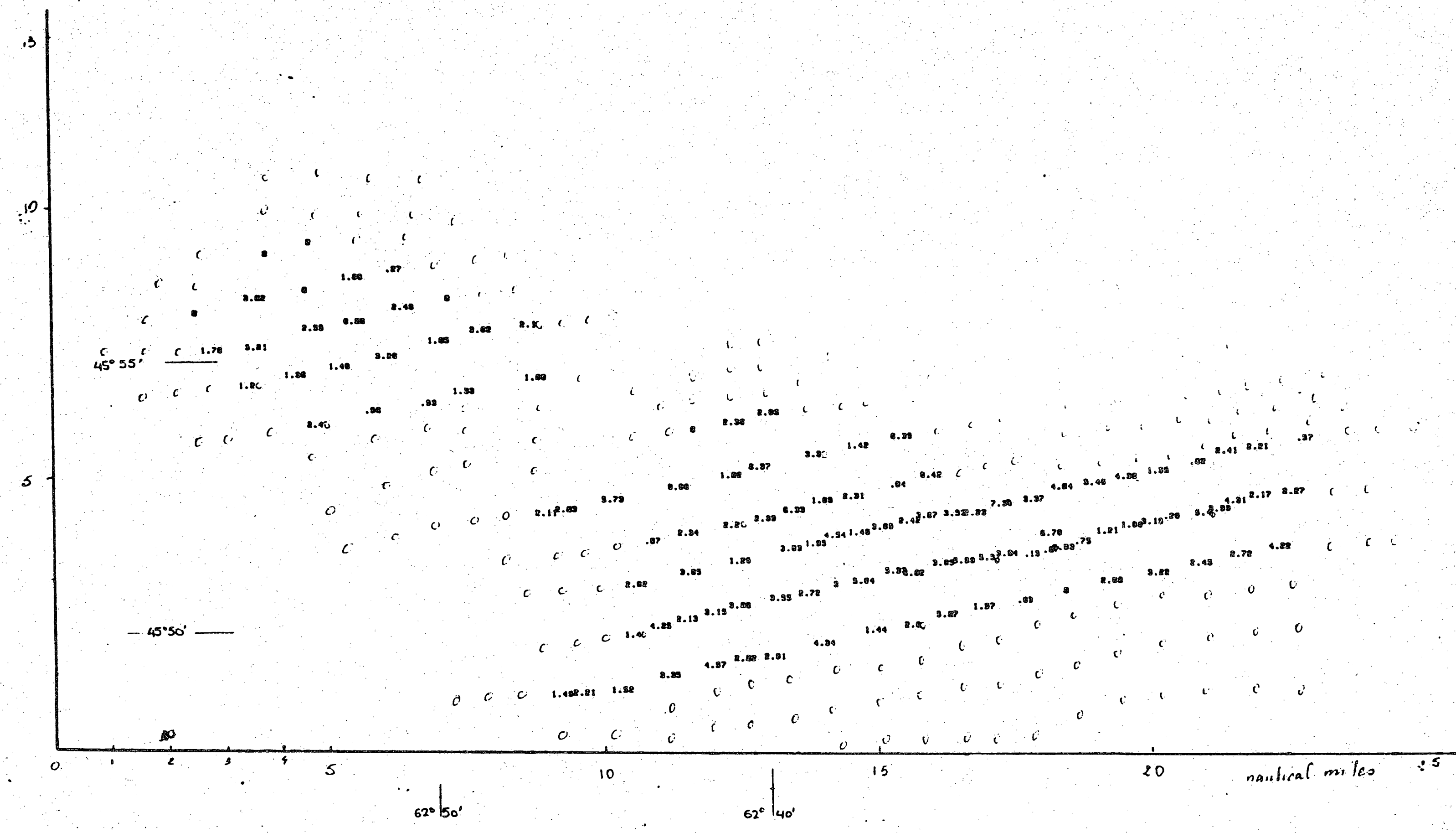

Fig. 1 - Geographic location of samples on Indian Rock (North) and Pictou (South) banks in Northumberland strait. The numbers represent densities in tons of live weight per square km.

PROJECT :SCALLOP DISTRIBUTION - PICTOU NB

DATE :JULY 23 1985

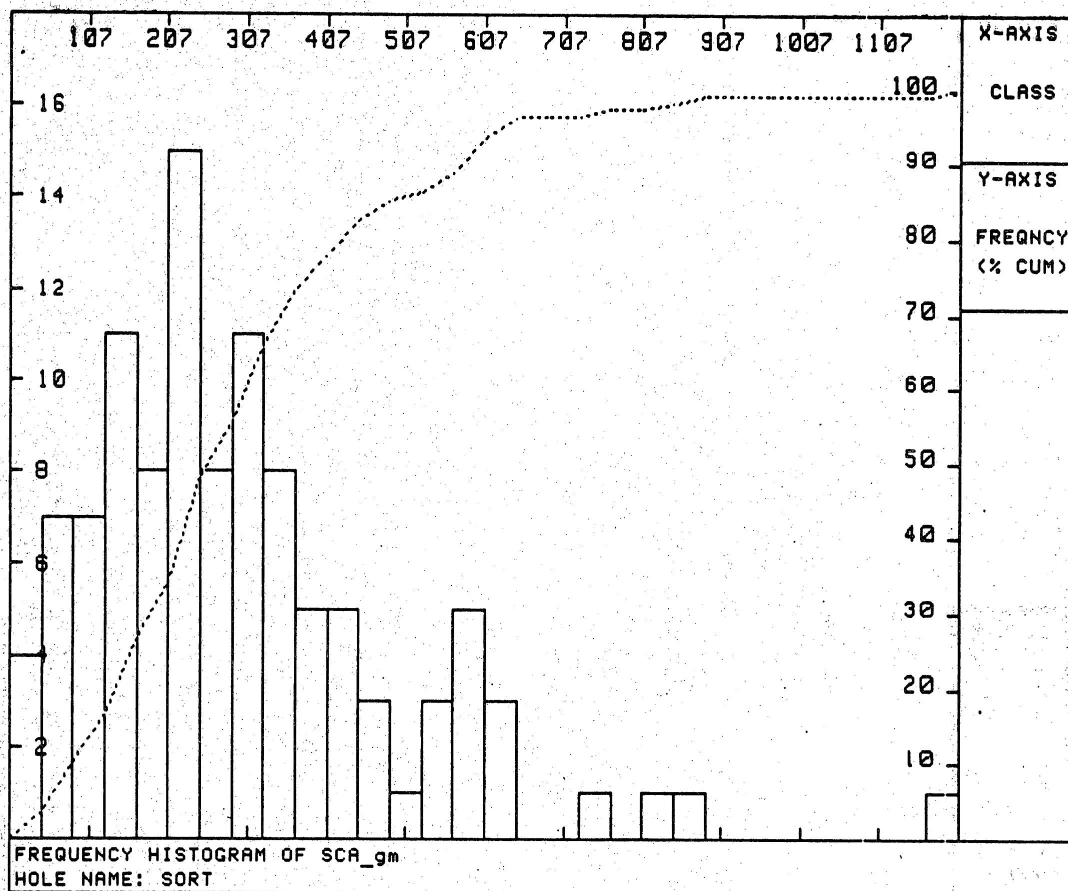

Fig. 2 - Frequency distribution of non zero densities of scallop live weight encountered in the samples in 10^{-2} tons per square km.

PROJECT :SCALLOP DISTRIBUTION - PICTOU NB

DATE :JULY 23 1985

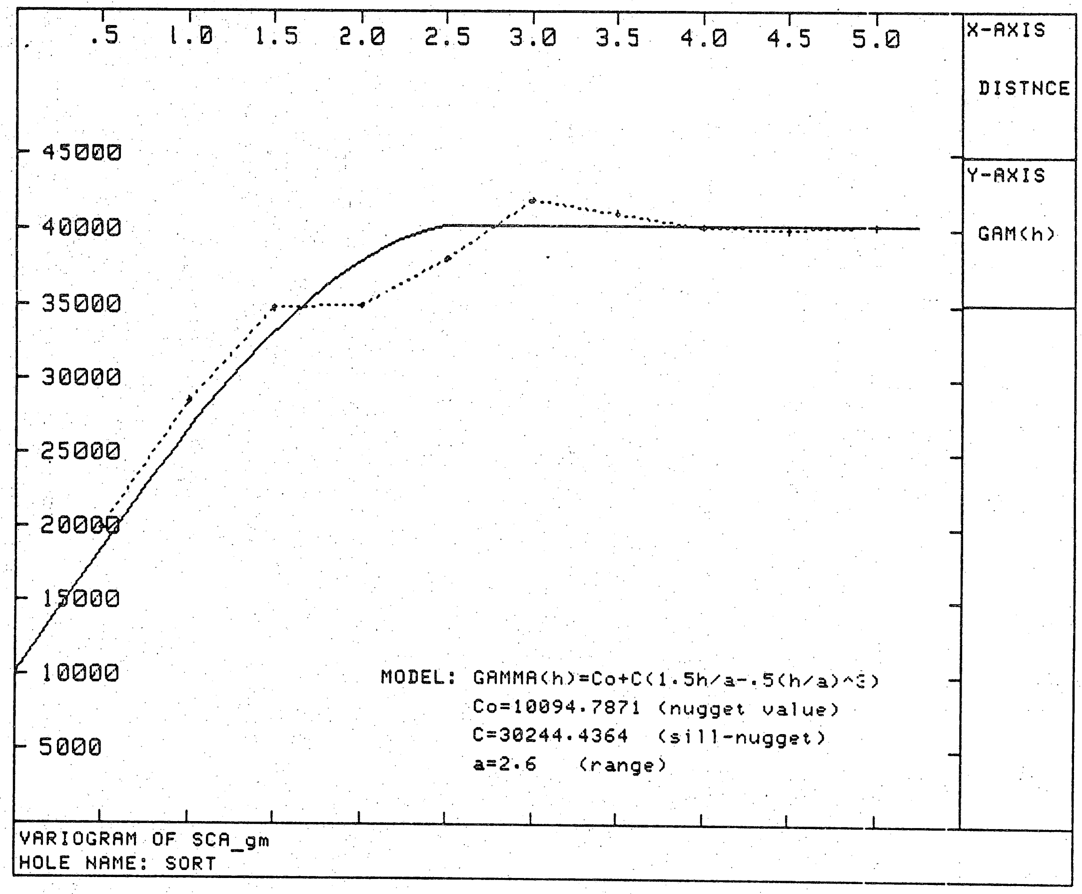

Fig. 3 - Experimental variogram $\hat{\gamma} = 1/(2N) \sum_{1}^{N} (Z(x) - Z(x + h))^2$

The variogram is well represented by a spherical model. There is a well marked sill (plateau) indicating that no auto-correlation is found among samples distant by more than 3 miles (the range). The "nugget" value indicates that there is a variance among replicates due either to a microstructure smaller than a dredge tow or to a measurement error such as length of tow. Units in ordinate, parameters C and Co should be multiplied by 10^{-4} for reading.

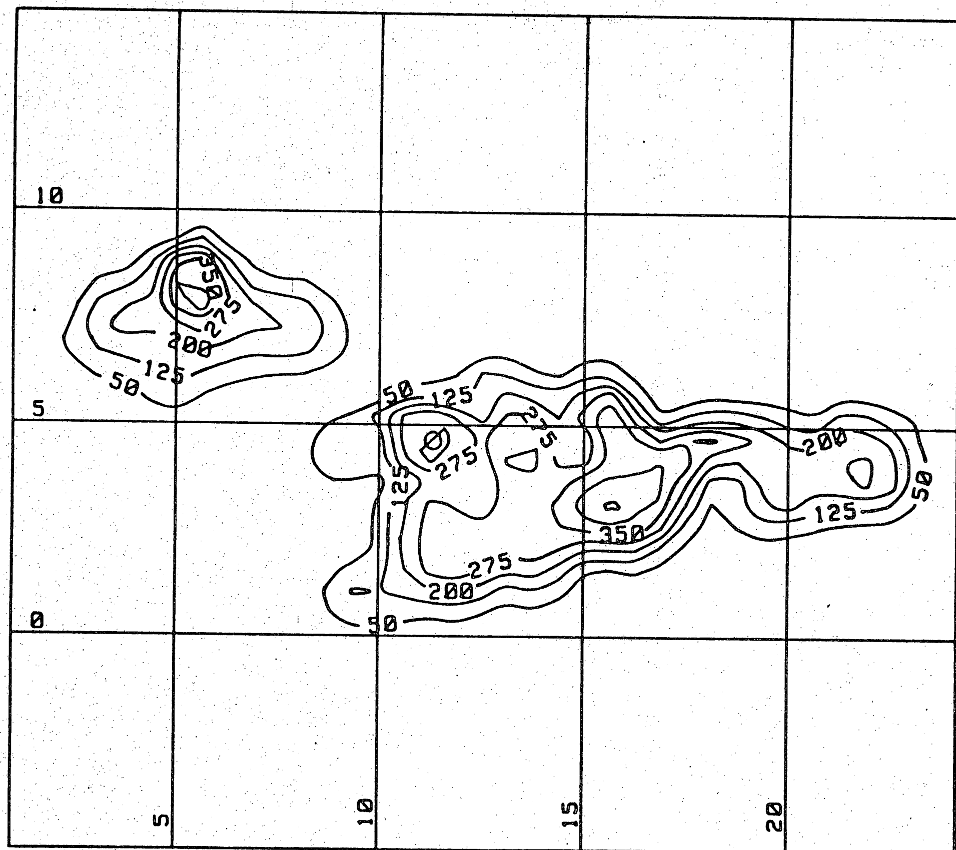

Fig. 4 - Two dimensional representation of isodensity contours for scallop biomass on Indian Rock and Pictou Banks. The numbers on the contours represent density units of 10^{-2} tons of live weight per square km.

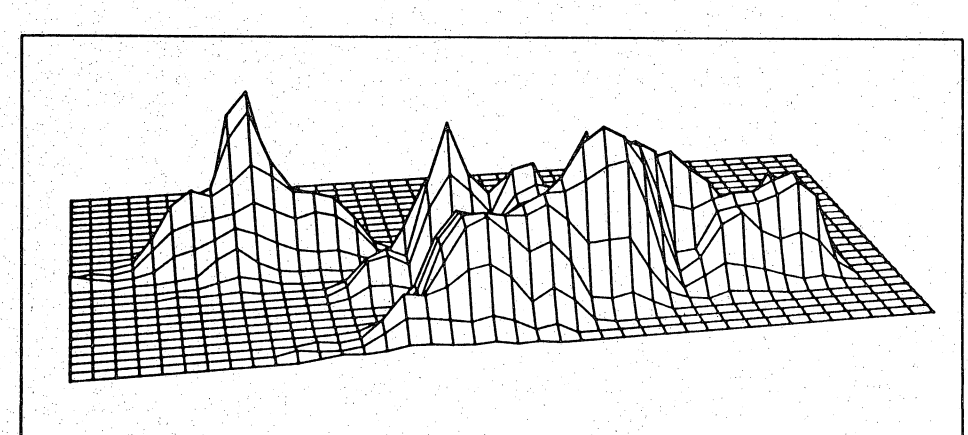

Fig. 5 - Three dimensional representation of isodensity contours for scallop biomass on Indian Rock and Pictou Banks. The numbers on the contours represent density units of 10^{-2} live weight per square km.