This report not to be quoted without prior reference to the Council *

ICES CM 1990/D:34
Ref: B, G, H, K

REPORT OF THE STUDY GROUP ON THE APPLICABILITY OF SPATIAL STATISTICAL TECHNIQUES TO ACOUSTIC SURVEY DATA.
(4, 5, 6 APRIL 1990, IFREMER, BREST)

chairman

Gérard Y. Conan **

Department of Fisheries and Oceans
Science Branch
c/o Marine Biology Research Laboratory
Université de Moncton
Moncton N.B. E1A 3E9
CANADA

* This document is a collective report of a Study Group of the International Council for the Exploration of the Sea and does not necessarily represent the views of the Council. Therefore, it should not be quoted without consultation with the General Secretary (ICES Palaegade 2-4, DK 1261, Copenhagen K, Denmark).

** Present address:
c/o Instituto de Ciencias del Mar
Paseo Nacional s/n
08003 Barcelona
ESPANA

BACK GROUND

The study group was set up during the 1989 statutory meeting as a follow up to the 1989 Workshop on Spatial Statistical Techniques (16-19 May, Brest, France). During this meeting the participants felt that spatial statistical techniques could be particularly promising for processing acoustic survey data.

Dr K. G. Foote played an active role in promoting research on new methodologies for designing and processing non random data resulting from acoustic surveys and presented the paradigm of a very highly sophisticated technology for measuring acoustic targets which is not backed up by sufficiently elaborate statistical methods. Statistical tools need to be developed for processing the information in order to obtain accurate estimates of biomass assorted with confidence limits.

Subsequently, the present study group was set up with the following mandate:

C. Res. 1989 / 2:7. A study group on the Applicability of Spatial Statistical Techniques to Acoustic Survey Data (chairman: Dr. G.Y. Conan) will meet in Brest , France from 26-28 March 1990 to:

a) describe and discuss computational results based on processing of real or synthetic echo data prior to the meeting;

b) plan further processing exercises based on these results;

c) prepare a detailed proposal for a workshop on the Application of Spatial Statistical Techniques to Acoustic Survey Data to be held in 1991;

d) report findings and plans to the statistics committee, with reference to the Shellfish, Demersal Fish, Pelagic Fish and Fish Capture Committees, at the Council Meeting in 1990.

The meeting was postponed to April 4-6 to the request of the participants. To the initiative of the French delegate, Alain Maucorps, the meeting was hosted by Institut Français de Recherche et d'Exploitation de la Mer, Centre de Brest.

ACKNOWLEDGMENTS

A vote of thanks is extended to the Director of the Center Jean-Max De Lamare and to:Ghislaine Gourmelen, François Le Verge, Jean l'Hostis, Francis Marazanoff, Alain Maucorps, Gérard Riou, Bruno Ropars, and Veronique Trebaol, for their efficient help in facilitating access to the meeting to all participants and for a friendly local organization.

PARTICIPANTS

The following persons were nominated members of the study group or participated as correspondents:

Dr. A. APPENZELLER
Dept. of Biology
Mc Gill University
1205 Docteur Penfield Av.
Montreal, Quebec H3A 1B1
CANADA

Dr. M. ARMSTRONG
Centre de Geostatistique
35, rue St. Honoré
77305 Fontainebleau
FRANCE

Dr. Paul BOUDREAU
Dept. of Fisheries & Oceans
P.O. Box 1006
Dartmouth N.S. B2Y 4A2
CANADA

Mr.R. CHEVALIER
IFREMER
Rue de l'Ile d'Yeu
B.P. 1049
44037 Nantes Cédex
FRANCE

Dr. G.Y. CONAN
Laboratoire de Recherche en Biologie Marine
Université de Moncton
Moncton, N.B. E1A 3E9
CANADA

Dr. A. DESBARATS
Geological Survey of Canada
601 Booth street,
Ottawa, Ont, K1A OE8
CANADA

Dr. K.G. FOOTE
Havforskningsinstituttet
5024 Bergen
NORWAY

Dr. P. GAGNON
Dept of Fisheries & Oceans
Institut Maurice
Mont-Joli, Q. G5H 3Z4
CANADA

Dr. Francis GOHIN
IFREMER, Centre de Brest
B.P.70
29280 Plouzané
FRANCE

Mr. Jean GUILLARD
INRA
B.P. 511
74203 Thonon
FRANCE

Dr. F. IBANEZ
Université Paris VI
Station Zoologique
B.P. 28
06230 Villefranche/Mer
FRANCE

Dr. Zinovy KIZNER
VNIRO
17, VerkneKrasnoselskaya
Moscow 107140
U.S.S.R.
FAX: (095) 264 91 87

Dr. M. KROEGER
Institut fur Fangtechnik
Palmaille 9
2 Hamburg 50
FEDERAL REPUBLIC OF GERMANY

Mr. G. LAVOIE
Université de Montréal
C.P.6128, Succursale A
Montréal,
Québec, H2C 3J7
CANADA

Mr. D.N. Mac LENNAN
Marine Laboratory
P.O. Box 101
Victoria Road
Aberdeen AR9 8DB
UNITED KINGDOM

Dr. D. MARCOTTE
Ecole Polytechnique
Dép. de Génie Minéral
C.P. 6079, Succursale A
Montréal, Québec H3C 3A7
CANADA

Mr. J. MODIN
Institute of Marine Research
Box 4
453 00 Lysekil
SWEDEN

Mr. M.D. NICHOLSON
Fisheries Laboratory
Lowestoft NR33 OHT
Suffolk
UNITED KINGDOM

Mr. A. NICOLAJSEN
Fiskirannsoknarslovan
Debessartrod
FR - 100 Torshavn
Faroe Islands
DENMARK

Dr. A. ORLOWSKI
Sea Fisheries Institute
Aleja Zjednoczenia 1
81-345 Gdynia
POLAND

Mr. D. OWEN
Southampton University
Dept. of Oceanography
Highfield
Southampton SO9 5NH
UNITED KINGDOM

Mr. L.-E. PALMEN
Inst.of Marine Research
Box 4
453 00 Lysekil
SWEDEN

Mr. Pierre PETITGAS
Centre de Geostatistique
35, rue St Honoré
77305 Fontainebleau
FRANCE

Mr. D. RENARD
Centre de Geostatistique
35, rue St Honoré
77305 Fontainebleau
FRANCE

Dr. Yvan SIMARD
Dept. of Fisheries & Oceans
Institut Maurice Lamontagne
850, route de la mer
Mont-Joli, Québec G5H 3Z4
CANADA

Dr E.J. SIMMONDS
Marine Laboratory
P.O. Box 101
Victoria Road
Aberdeen, Scotland
UNITED KINGDOM

Dr. S. SMITH
Dept. of Fisheries & Oceans
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, N.S. B2V 4I2
CANADA

Mr. R. Mohan SRIVASTAVA
FSS International
3900 Quebec st.
Vancouver, B.C. V5V 3K8
CANADA

Dr. Gunnar STEFANSSON
Marine Research Institute
P.O. Box 1390
Skulagata 4, Reykjavik
ICELAND

Dr. Dimitri STOLYARENKO
VNIRO
17, VerkneKrasnoselskaya
Moscow 107 140
U.S.S.R.
FAX: (095) 264 91 87

Mr. E. WADE
Dept. of Fisheries & Oceans
P.O. Box 5030
Moncton, N.B. E1C 9B6
CANADA

Mr. W. WARREN
Dept. of Fisheries
& Oceans, P.O. Box 5667
St John's, Nfld A1C 5X1
CANADA

Dr. N.J. Williamson
NMFS: Alaska Fisheries Science Center
7600 Sand Way Blvd, NE
Seattle, Washington 98115-0070
U.S.A.

DATA ANALYSED

Five test data sets were provided by Dr. K.G. Foote prior to the meeting in order to allow participants and correspondents to assay methodologies (table 1). The 29 pages of data listing cannot be provided in this report but are available in the form of listings or diskettes from Dr. Foote. Additional sets of data were processed by some of the participants in order to illustrate specific practical cases.

The types of difficulties encountered for processing the sets by standard statistical methods are commonly known in acoustic survey data analysis:

1) High density of observations(up to 1712 points) along transects, but transects located far apart.

2) High level of patchiness generating strong autocorrelation along transects.

3) Very high values concentrated very locally and strongly contrasted with very low values and zero's.

4) Boundary effects set by coast lines and depth gradients.

5) Transect routes not following standard designs such as random or regular intervals. Transect routes eventually overlapping in space but not in time with considerable differences associated with time as well as space related variability.

Table 1. CHARACTERISTICS OF TEST DATA SETS.

Data set	Fish type	Region	Interval (N.M) Integra-tion	Interval (N.M) Normali-zation	No. data	Comment
1	Pelagic	Coast	5	5	664	Unbounded aggregation with concentration on geographic limits of area sampled.
2	Pelagic	Fjord	1	1	96	Bounded but extreme non autocorrelated variability
3	Pelagic	Coast	1	5	881	Mostly bounded
4	Pelagic	Coast	1	5	986	Mostly bounded
5	Benthic	Open Sea	3-5	5	1712	Two-ship survey with overlapping. Bounded aggregation

METHODS.

The following methodologies and associated software were used for processing the data sets prior to the meeting. For information on availability of software you may refer to the participant(s) identified. (In some cases an educated users, not necessarily the designers).

1) Generalized Linear Interactive Modelling (GLIM) NAG Institute. A software designed for modelling linearly response surfaces. G. STEFANSSON.

2) Spline Survey Designer Software System (SSDSS). A spline approximation algorithm derived from thin sheet theory. Incorporates information on depth as well as geographic location. D. STOLYARENKO.

3) Bluepack 3-D. A software package allowing Intrinsic Random Functions of order K. D. RENARD.

4) Gulfkrig. A software package designed specifically for fisheries data, based on ordinary kriging. G.Y. CONAN, E. Wade.

5) Box-Cox test/Power Series Analysis. A software package allowing post stratification and optimized power series transformations for calculating abundance and confidence limits. J. SIMMONDS.

6) Simple arithmetic mean and standard deviation. Standard methodology assuming random sampling.

7) Geo-EAS. A software package based on ordinary kriging. Transformation of variables optional. Y. SIMARD, A. DESBARATS.

8) Calculation of abundance based on transects used as strata. N.J. WILLIAMSON

AGENDA

At the opening of the meeting, the participants were requested to provide information on the material they had processed and on the methods of analysis they would present. The following schedule was organized thereafter:

April 4th

<u>Introductory presentations on acoustic data characteristics</u>

-Fish schools by Ken FOOTE and Gunnar STEFANSSON

-Plankton patches by Frederic IBANEZ

Statistical methods

- Geostatistics by Margaret ARMSTRONG
  - Domain
  - Stationarity
  - Variogram
  - Ordinary kriging
  - Transformations (Logarithms)
  - Intrinsic Random Functions of Order K
  - Disjunctive kriging
  - Indicative variables
  - Conditional simulations
  - Splines and kriging

April 5th

Statistical methods (Continued)

- Splines as a subset of kriging by Didier Renard

- Splines as a parallel but distinct approach to kriging for spatial data analysis and modelling by Dimitri STOLYARENKO

- Response surfaces by Gunnar STEFANSSON

- Post stratification and power function transformation of spatial data by John SIMMONDS

Data analyses

Presentation of the results obtained by the participants for each of the test data sets

Set # 1
Set # 2
Set # 3
Set # 4
Set # 5
Other specific sets of data were provided by the authors

April 6th

- Synthesis of the individual presentations, comparison of the results

- Recommendations

RESULTS.

Table 2 SUMMARY OF THE RESULTS OF THE ANALYSES OF THE TEST DATA SETS

SET	Technique	S_A m²/Nm²	CV %	Area Nm²	S_A*Area	Authors
1	Kriging	85	22	58 10^3	4.9 10^6	Conan & Wade
	Spline	77	N/A	53 10^3	4.0 10^6	Stolyarenko
	Box/Cox transform	68	9	55 10^3	3.7 10^6	Simmonds
	Arithmetic mean	75				
2	Kriging	444	2	76	33.8 10^3	Conan & Wade
	Spline	259	N/A	51	13.2 10^3	Stolyarenko
	Box/Cox transform	48	37	49	2.4 10^3	Simmonds
	Arithmetic mean	297				
3	Kriging	2534	23	45 10^2	11.5 10^6	Conan & Wade
		2089	14	90 10^2	18.8 10^6	Guillard & Gerdaux
		1911	22	83 10^3	15.9 10^6	Armstrong
	Spline				7.8 10^6	Stolyarenko
	Box/Cox Transform	1327	7	55 10^2	7.3 10^6	Simmonds
	Transects as strata	3072	30	19 10^2	5.7 10^6	Williamson
	Arithmetic mean	1793				
4	Kriging	983	52	50 10^2	4.9 10^6	Conan & Wade
		1126	55	30 10^2	3.3 10^6	Petitgas
	Spline				3.5 10^6	Stolyarenko
	Box/Cox transform	560	9	61 10^2	3.4 10^6	Simmonds
	Transects as strata	1512	31	22 10^2	3.3 10^6	Williamson
	Arithmetic mean	774				

Table 2 SUMMARY OF THE RESULTS OF THE ANALYSES OF THE TEST DATA SETS (Continued)

SET	Technique	S_A m²/Nm²	CV %	Area Nm²	S_A*Area	Authors
5	Kriging	14	18	19 10^4	266.0 10^4	Conan & Wade
	Spline				87.5 10^4	Stolyarenko
	Box/Cox transform	9	8	13 10^4	110.0 10^4	Simmonds
	Arithmetic mean	14				

GENERAL COMMENTS BY THE CHAIRMAN

Despite the apparently different methodologies used, the results for the global estimates S_A*Area are usually quite similar. However, the variance estimates widely differ. The different variance estimates are not directly comparable, and it would be an error to define as best estimator the one associated with the smallest variance. The assumptions used are far more determinant for the variance calculations than for the global abundance estimate.

The area to be used for stock estimation is not clearly defined in most methodologies. It should not be arbitrarily set, but based on the analysis of the spatial structure. Differences in area considered by the different authors were partially compensated by associated differences in global estimates of global averages within the areas.

Overemphasis of the departure of the data from the assumptions used in each model may lead to sterile statements of inadequacy of any model. Rather, assaying the robustness of the techniques would provide practical results. A biassed or imprecise estimate of known characteristics is preferable to no estimates at all.

The approaches could be roughly regrouped into two categories: those emphasizing spatial structure, and those emphasizing the shape of the sampling distribution. There is presently no answer as to which approaches are more efficient or more robust since the actual abundance and its spatial repartition cannot be known. It seems that the processing of simulated sets of data of known characteristics could provide some insight on the appropriateness of the tools presented.

Mapping can provide some very useful insights on the localization of the resource and on the appropriateness of the sampling design. Quite frequently, it becomes apparent that the main concentration is only marginal to the survey area. Shipboard data processing and mapping is possible using spatial statistics and would permit adaptative sampling schemes.

RECOMMENDATIONS.

1.- The group notes that good survey design is essential for obtaining high-quality estimates of total stock size. In particular the group recommends that acoustic surveys extend to areas of low or zero concentrations or otherwise bound the distribution.

2.- In the case of narrow fjords it is essential that the survey provide information across the fjord as well as along the length of the fjord.

3.- When a fish population is dominated by a small number of large schools and it is possible to locate all of these schools, it is recommended that the survey be designed to first locate the schools and then intensively estimate the biomass of these individual schools.

4.- It is important to design surveys in a well-ordered fashion. Where the mean or the variance of the spatial distribution is to be dependent on or affected by some external factor, data on this parameter should be recorded. Examples of such external factors are time of day, water depth, species composition and size.

5.- It is recommended that a workshop be conducted at (a place to be named) during (a time to be given) in order to do the following:

a) Present data analyses performed in advance,

b) Compare methods performed in advance,

c) Discuss these analyses and methods,

d) Prepare a digest of spatial statistical methods,

e) Decide on how the findings are to be reported formally

f) Discuss future work on the applicability of spatial statistical techniques to acoustic survey data.

6.- It is recommended that an informal group, consisting of K.G. Foote (coordinator), Z. Kizner, E.J. Simmonds and G. Stefansson, derive test data sets prior to the 1990 Statutory Meeting. The data sets will reflect the following characteristic types of fish aggregation:

Type	Aggregation	Region
Pelagic	Dense	Fjord
Pelagic	Dense	Sea
Pelagic	Dispersed	Sea
Bottom	Dense	Fjord/Bank
Bottom	Dispersed	Sea

The data sets will be extracted from repeated surveys on the same stock or will be derived by modelling, using observed characteristics of actual fish aggregations.

These data will be reported on to the Statistics Committee at the 1990 Statutory Meeting.

APPENDIX

ABSTRACTS OF INDIVIDUAL CONTRIBUTIONS FROM THE PARTICIPANTS

INTRODUCTION TO TEST DATA SETS: CHARACTERISTICS OF ACOUSTIC DATA ON FISH AGGREGATIONS.

by

K.G. FOOTE

BACKGROUND.

The use of Spatial Statistical Techniques (SSTs) is well established in many disciplines, but has scarcely been mentioned in the context of acoustic surveys of fisheries. Notable exceptions are due to the pioneering work of F. Gohin in the early 1980's and more recent work by G.Y. Conan et al. Aims in fisheries surveying are remarkably similar to applications to the mining industry, where geostatistics is a basic tool. It is hoped that this Study Group will begin to establish the applicability of SSTs to acoustic survey data.

STATISTICAL ESSENCE OF ACOUSTIC SURVEY DATA.

Briefly, the data consist of dense samples of fish density along widely spaced line transects, stationarity of the fish stock during the survey is generally assumed and is often a very good assumption. Variability in the precise form of the aggregation is, however, the rule.

EXAMPLE OF ACOUSTIC SURVEY.

Elements of acoustic surveying are illustrated by the example of Northeast Artic cod (*Gadus morhua*). This is not unique, for more than ten fish stocks are regularly surveyed by acoustics in Norway, and applications of acoustic surveying are truly worldwide.

ANALYSIS AND PROGNOSIS FOR ACOUSTIC SURVEYING.

The individual elements are separately analyzed. By analogy with a chain, the whole is no stronger than the weakest link. Significant gains may be expected in most elements of the process over the next several years. An outstanding, neglected element is that of abundance estimation over an area from line-transect measurements of fish density, including variance estimation. The relative importance of this cannot even be assessed, because of widespread neglect of spatial structure in treatments of acoustic survey data, hence this Study Group.

CRITIQUE OF PRESENT DATA SETS.

These are examples of several kinds of acoustic survey data that are routinely collected and processed in Norway. They have been compressed enormously, by integrating measurements of fish density over both depth and sailed distance over the range 1-5 nautical miles (N.M.) They do have the conspicuous advantage of being pure in species. Size or age group may also be assumed to be constant for each data set.

HIGHER-RESOLUTION ACOUSTIC SURVEY DATA.

Acoustic survey data integrated over long intervals of sailed distance are admittedly coarse. At the opposite extreme are echo data collected and stored digitally ping by ping. These are illustrated by two examples:

1) Color echograms of diverse aggregation of herring (*Clupea harengus*) and blue whiting (*Micromesistius poutassou*) displayed by the Bergen Echo Integrator (Knudsen, Proc IOA 11(3), 1989), and

2) color echograms of a dense aggregation of the 1983-year class of herring printed by the SIMRAD EK500 scientific echo sounding system (Budholt et al., ICES CM 1988/ B:10)

SOME PROLEMS ABOUT THE INTERPRETATION OF ACOUSTIC DETECTION IN PLANKTON ECOLOGY

by

Frederic IBANEZ

A great part of the in situ activity of the oceanographic laboratory of Villefranche is focused on the study of an offshore frontal zone located almost 20 miles from the coast. This zone is characterized by a sharp salinity gradient, and an upwelling of nutrients which favour an important biological productivity. For ten years, in the subsurface layer, continuous multiparametric hydrological records, associated to continuous zooplankton sampling (by a Tube Haï pumping system) have been processed on transects crossing the front. The results of this cruise showed that spatial distribution of the phytoplankton and zooplankton was related to the variations of intensity and to the displacements of the frontal zone (Ibanez & Boucher, 1987). For some species (coastal and also pelagic ones), the salinity gradient appears as a barrier never crossed, for some others the front looks like a "nursery" during the reproductive period (Boucher et al., 1987).

The use of echosounding (continuous map of echoes in the vertical plane during transects crossing the front), showed that the frontal structure does not affect only the plankton ecosystem located in the euphotic zone (Baussant, 1988). High echoes were recorded 300 or 400 m deep, corresponding to an almost continuous layer the shape of which was changing with the hydrological structure. A general oblique structure is observed, the isoclines sinking progressively from offshore to the coast.

The global estimation of plankton biomass is not the first aim of the ecologist (Ibanez, 1976). Since the scattering layers likely correspond to assemblages of several species and sizes, observations by Isaacs-Kidd net, Bioness multiple net, camera, or even submersible, were used to try to identify the targets. But, contrary to fish patches, the scattering layers do not have precise spatial horizontal limits; therefore, it is impossible to define statistic spatial blocks in order to obtain a global estimation (Ibanez, 1981). The plankton ecologist is rather interested in other properties which could be likely deduced from acoustic exploration: what produces the movements of the layers? How and why are the organisms able to follow some lines of isocline? How long should be the upwelling of nutrients in order to allow the multiplication of algae, then the appearance of the first level of heterotrophic organisms? How to separate the influence of environmental factors from biological behaviour on the spatial concentration of plankton at particular depths?

So the application of geostatistical techniques (Ibanez, 1985) is not very simple here. It seems that the enormous quantity of data prohibits the kriging computation (even after some reduction, of the vertical sampling step). A supplementary difficulty of the smoothing kriging process corresponds to the intermittency of the records: in the vertical plane, several layers are separated by large empty zones, therefore the interpolation could produce artificial limits for the patches. Photography of the screen of the acoustic device or video image now appears to be the best representation for a survey. But in this case how to relate, for instance, the distribution of plankton patches (discontinuous structure) to the hydrological (continuous) structure? Moreover, the resolution of the parameters is not so fine as acoustic sampling: their vertical variation is known only for a few stations along a transect. Estimations of the means, and hence of correlations at the scale of the sampling field are not possible.

Taking into account the intensity of the echoes and the values of hydrological parameters only at the stations where these last parameters are recorded seems much more rigorous. Rather than classical statistics (because of the absence of echo signal at particular depths), pattern recognition (syntaxic analysis: Pigeau, 1986) could be used to detect similarity between shapes (at the same depths of shifted, for instance, between blooms of chlorophyll and local high values of echoes). So the comparison of echo signal at different stations also could lead to the recognition of animal migrations. Another promising method could be the interpretation in the space of external parameters, i. e. the detection of the classes in which such acoustic intensity appears. This technique is also difficult, because it requires to separate first the different hydrological compartments.

Finally, considerations and discussions have to be made in order to find the best quantitative interpretation of the acoustic data in ecology. In my opinion, starting from the main ecological questions and not from the transposition of particular mathematical algorithms, statistics or even geostatistics seem to be of poor utility. Perhaps non-parametrics methods and semi-qualitative ones, like pattern recognition, could be the most ecologically meaningful way of interpretation.

References

Baussant, T., 1988. Contribution à la détection acoustique du plancton sur la verticale en zone frontale Ligure. Mémoire DEA Université de Paris. Research direction: F. Ibanez.

Boucher, J., F. Ibanez & L. Prieur, 1987. Daily and seasonal variations in the spatial distribution of zooplankton populations in relation to the physical structure in the Ligurian Sea Front. Jour. mar. res., 45: 133-173.

Ibanez, F., 1976. Contribution à l'analyse mathématique des événements en écologie planctonique. Bull. Inst. Océanogr. Monaco, 72: 1-96.

Ibanez, F., 1981. Immediate detection of heterogeneities in continuous multivariate, oceanographic recordings. Application to time series analysis of changes in the bay of Villefranche sur Mer. Limnol. Oceanogr., 26: 336-349.

Ibanez, F., 1984. Sur la segmentation des séries chronologiques planctoniques multivariables. Oceanologica Acta 7: 481-491.

Ibanez, F. & J. Boucher, 1987. Anisotropie des populations zooplanctoniques dans la zone frontale Ligure. Oceanologica Acta 10: 205-216.

Pigeau, F., 1986. Interprération de l'échantillonnage en continu par l'analyse syntaxique. Mémoire DEA Université de Paris. Research direction. F. Ibanez.

OVERVIEW OF GEOSTATISTICS

M. Armstrong, Centre de Géostatistique, Fontainebleau, France

OBJECTIVES

The objective of this chapter is to give an overview of geostatistics, and in particular to explain the main concepts: stationarity, variogram, kriging,... The term "stationary" can lead to misunderstandings. Sometimes used in the statistical sense while at others it refers to the mobility or immobility of the fish. Here it will always be used in the first sense.

MODELLING REGIONALIZED VARIABLES

The term "Regionalized Variable" was chosen by Matheron to emphasize the two apparently contradictory aspects seen in most spatial variables:– a random aspect, which accounts for local irregularities, and a structured aspect, which reflects the large scale tendencies of the phenomenon.

A Regionalized Variable is characterized by the joint distributions of any set of variables $Z(x_1)$, $Z(x_2)$, ... $Z(x_k)$, for all k, and for all points $x_1, x_2, \ldots$. Of course, it would be impossible to do anything with this model unless we are prepared to make some assumptions about the characteristics of these distributions. In particular since only one realization is usually available we have to make some assumptions about its stationarity.

STATIONARY AND INTRINSIC HYPOTHESES

In statistics it is common to assume that the variable is stationary, i.e. its distribution is invariant under translation. In the same way, a stationary Regionalized Variable is homogeneous, and statistically self-repeating in space. This makes statistical inference possible. In its strictest sense stationarity requires all the moments to be invariant under translation but since this cannot be verified from the limited experimental data, we usually only require the first two moments (the mean and the spatial covariance) to be constant. This is called "weak" or second order stationarity. In other words we require that

(i) the expected value (or mean) of the function Z(x) is constant for all points x. That is, $E(Z(x)) = m(x) = m$ which is independent of x.

(ii) the covariance function C(h) between any two points x and x+h is independent of the point x. It depends only on the vector h. That is,

$$E[Z(x)\, Z(x+h)] - m^2 = C(h)$$

In particular, when $h = 0$, the covariance C(h) is just the variance of Z(x) which must also be the same at all points .

In practice, it often happens that these assumptions are not satisfied. Clearly when there is a marked trend the mean value cannot be assumed to be constant. We shall see how to take account of trends later. For the moment we shall only consider cases where the mean is constant. However, even when this is true, the covariance need not exist. So it is convenient to be able to weaken our stationarity hypothesis. Under the "intrinsic hypothesis" we suppose that the increments of the function are weakly stationary: that is, the mean and variance of the increments $Z(x+h) - Z(x)$ exist and are independent of the point x.

$$E[Z(x+h) - Z(x)] = 0 \qquad \text{intrinsic hypothesis}$$
$$\mathrm{Var}\,[Z(x+h) - Z(x)] = 2\gamma(h) \qquad \text{with zero mean}$$

The function $\gamma(h)$ is called the variogram. It is the basic tool for the structural interpretation of phenomena as well for estimation.

In practical situations the variogram is only used up to a certain distance. This limit could be the extent of a homogeneous zone within a deposit or the diameter of the neighbourhood used in kriging (i.e. estimation). Consequently, the phenomenon only has to be stationary up to this distance. The problem is to decide whether we can find a series sliding neighbourhoods within which the expected value and the variogram can be considered to be stationary and whether there are enough data in these zones to give meaningful estimates. This assumption of quasi-stationarity is really a compromise between the scale of homogeneity of the phenomenon and that of the sample density.

THE VARIOGRAM PROPERTIES

The variogram is defined as the variance of $Z(x+h) - Z(x)$. As it has been assumed that the mean of $Z(x+h) - Z(x)$ is zero, the variogram is just the mean square value of this difference. That is,

$$\gamma(h) = 0.5\, E\,[Z(x+h) - Z(x)]^2$$

Here x and $x+h$ refer to points in a n-dimensional space where n could be 1, 2 or 3. For example, when $n = 2$ (i.e. in the plane), x denotes the point (x_1, x_2) and h is a vector. Consequently, in a 2-dimensional space the variogram is a function of the two components h_1 and h_2. Transforming to polar coordinates, it is a function of the modulus of the vector h and its orientation. For a fixed angle, the variogram indicates how different the values become as the distance increases. When the angle is changed, the variograms disclose the directional features of the phenomenon such as its anisotropy.

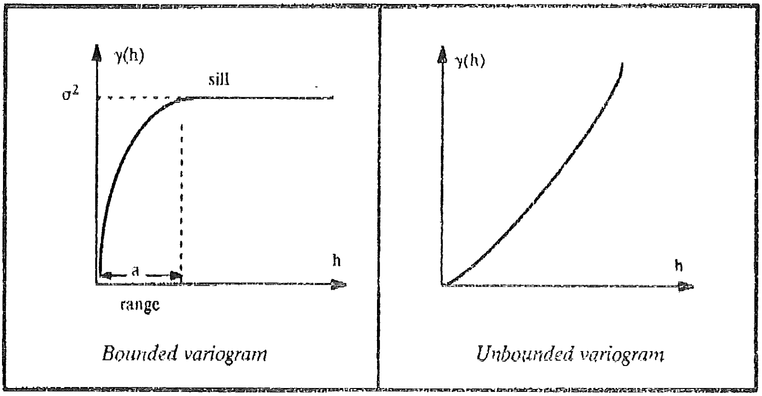
The graph of $\gamma(h)$ plotted against h presents the following features. It always starts at 0 (for $h = 0$, $Z(x+h) = Z(x)$). It generally increases with h.. rising up to a certain level called the sill and then flattening out. Alternatively it could just go on rising.

RANGE AND ZONE OF INFLUENCE

The rate of increase of the variogram with h indicates how quickly the influence of a sample drops off with distance. After the variogram has reached its limiting value (its sill) samples this far apart no longer correlated. Theory shows that the sill value of the variogram is exactly the variance of the population.

The range need not be the same in all directions. This merely reflects the anisotropy of the phenomenon. What is more, even for a given direction there can be more than one range. This occurs when there are several nested structures acting at different scales of distance.

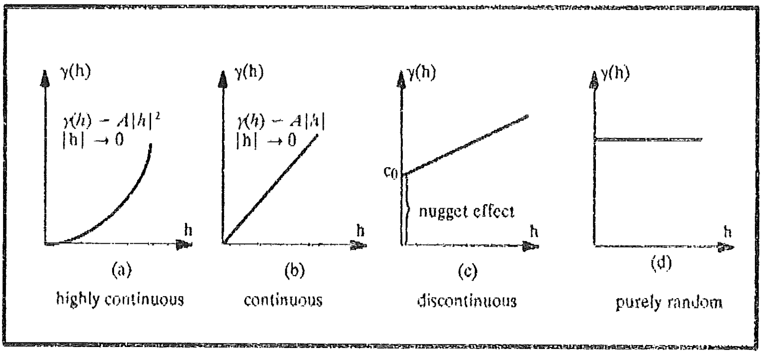
Not all variograms reach a sill. Some like the one shown on the right just keep on increasing with h. This is one fundamental difference between the variogram and the covariance. The latter exists only for stationary variables.



Bounded variogram *Unbounded variogram*

BEHAVIOUR NEAR THE ORIGIN

We have just examined the behaviour of the variogram for large distances. But it is also most instructive to study its behaviour for small values of h because this is related to the continuity and the spatial regularity of the variable. Four types of behaviour near are shown below.



(a) highly continuous (b) continuous (c) discontinuous (d) purely random

(a) A parabolic shape. This indicates that the regionalized variable (Re. V.) is highly continuous and even differentiable. A parabolic shape can also be associated with the presence of a drift.

(b) A linear. In this case the Re V is continuous but not differentiable, and thus less regular than in (a).

(c) A discontinuity at the origin. This means that the variable is not even continuous in the mean square. It is, therefore highly irregular at short distances.This jump at the origin is called a nugget effect because it was first noticed in gold deposits in South Africa where it is associated with the presence of nuggets on the ore. It is convenient to apply the term "nugget effect", to this sort of short range variability even when it is known to be due to some other factor e.g the micro-structure, measurement error or errors in location.

(4) A flat curve. Pure randomnesss or white noise. The regionalized variables $Z(x+h)$ and $Z(x)$ are uncorrelated for all values of h. no matter how close they are. This limiting case shows a total lack of structure. It is incidently the model adopted in trend surface analysis.

ANISOTROPIES

When variograms are calculated for all pairs of points in certain directions such North-South and East-West, they sometimes show different types of behaviour (i.e. anisotropy). If this does not occur the variogram depends only on the magnitude of the distance between points h and is said to be isotropic.

Two different types of anisotropy can be distinguished: geometric anisotropy and zonal anisotropy.

(a) Geometric Anisotropy (also called "elliptic" anisotropy). In this case the anisotropy can be corrected by an affine change of coordinates.

(b) Zonal (or stratified) Anisotropy. More complex types of anisotropy than geometric anisotropy exist. For example, in 3-D the vertical direction often plays a special role because there is more variation between strata than within them and so the sill of the vertical variogram is often higher than that of the horizontal ones.

PRESENCE OF A DRIFT

Theory shows that for large distances the variogram of a stationary or intrinsic regionalised variable must increase more slowly than a parabola. To be more specific.

$$\frac{\gamma\ (h)}{h^2} \rightarrow 0 \ as\ h \rightarrow \infty$$

However in practice we often find variograms which increase more rapidly than h^2 for large h. This indicates the presence of a drift.

HOW TO CALCULATE EXPERIMENTAL VARIOGRAMS

The following formula is used to calculate the experimental variogram from the data.

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

where $Z(x_i)$ are the data values; x_i are the locations of the samples and $N(h)$ is the number of pairs of points $(x_i, x_i + h)$; that is the number of pairs of points separated by a distance h..

VARIOGRAM MO LS

Not all mathematical functions can be used as variogram models. They must have the property of being positive definite. To be more precise $-\gamma(h)$ must be conditionally positive definite The following ones are.

All those listed here except the power models correspond to stationary random variables; the others are associated only with intrinsic ones. This list of variogram models is not exhaustive.

1) Nugget Effect

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C & |h| > 0 \end{cases}$$

2) Power Functions:

$$\gamma(h) = |h|^{\alpha} \quad \text{with} \quad 0 \leq \alpha < 2$$

As a particular case we have the linear model $\gamma(h) = |h|$

3) Spherical Model:

$$\gamma(h) = \begin{cases} C\left[\frac{3}{2}\frac{|h|}{a} - \frac{1}{2}\left(\frac{|h|}{a}\right)^3\right] & |h| < a \\ C & |h| > a \end{cases}$$

The spherical model is probably the most commonly used model. It has a simple polynomial expression and its shape matches well what is often observed: an almost linear growth up to a certain distance then a stabilization.

(4) Exponential Model:

$$\gamma(h) = C\,[\,1 - \exp(-|h|/a)]$$

For practical purposes, the range can be taken as 3a.

(5) Gaussian Model:

$$\gamma(h) = C\,[1 - \exp(-h^2/a^2)]$$

KRIGING

The problem is as follows: we have N data values $z(x_1), .. z(x_N)$ at our disposal and we want to estimate a linear function of the variable Z(x). For example we might want to estimate the value of the variable at a particular point or its average over a certain region. To avoid having to write out all the cases separately we shall denote the quantity to be estimated as:

$$y_0 = \frac{1}{V}\int_V Z(x)\,dx$$

where the volume V would reduce to a single point in the case of point estimation. To estimate this, we consider weighted average of the data:

$$y^*_0 = \sum_{i=1}^{N} \lambda_i\, Z(x_i)$$

(By convention the star will be used to denote the estimated value as opposed to the real but unknown value). The problem is to choose the weighting factors λ_i in the best way. This is where we make use of the statistical model. We consider the random variable:

$$y^*_0 = \sum_{i=1}^{N} \lambda_i Z(x_i)$$

We choose the weights so that the estimator is

1. unbiased: $E(Y_0^* - Y_0) = 0$
2. minimum variance: $E(Y_0^* - Y_0)^2$ is a minimum.

Some fairly simple calculations lead to a set of N + 1 linear equations:

Kriging system

$$\sum \lambda_j \gamma(x_i - x_j) + \mu = \bar{\gamma}(x_i, V) \quad i = 1, 2 \ldots N$$
$$\sum \lambda_j = 1$$

The minimum of the variance which is called the kriging variance, is given by:

Kriging variance

$$\sigma^2_K = \sum \lambda_i \bar{\gamma}(x_i, V) - \bar{\gamma}(V, V) + \mu$$

To solve the system numerically, it is convenient to write it in matrix from. We get

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & & & \gamma_{1N} & 1 \\ \gamma_{21} & \gamma_{22} & & & \gamma_{2N} & 1 \\ & & & & & \\ \gamma_{N1} & \gamma_{N2} & & & \gamma_{NN} & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \\ \lambda_N \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{\gamma}(x_1, V) \\ \bar{\gamma}(x_2, V) \\ \\ \bar{\gamma}(x_N, V) \\ 1 \end{bmatrix}$$

NONSTATIONARY GEOSTATISTICS

The intuitive way of coping with non-stationarity is to fit a polynomial trend by some method such as least squares, calculate the differences between the experimental values of the variable and the fitted ones and then fit a variogram model to the square of the differences (or residuals as they are usually called). Unfortunately the variograms of the residuals are almost always biased as there is no clear cut distinction in reality between the trend and the residual. They seriously under-estimate the underlying variogram. For more details see Chiles (1977). This approach which is called universal kriging only works when either the variogram or the drift are known a priori, which is rarely true in practice.

When the variogram is not known a priori, a more sophisticated method involving intrinsic random functions of order k (called I.R.F.-k. for short) should be used. The idea behind this method is to filter out polynomial drifts up to degree k without ever explicitly calculating or estimating their coefficients in the same way that ordinary kriging filters out the unknown constant mean.

FILTERING POLYNOMIAL DRIFTS

In ordinary kriging the role of the nonbias condition (called the universality condition) $\Sigma\ \lambda_i = 1$ is to filter out the (unknown) mean of Z(x) without explicitly calculating or estimating its value. By adding more nonbias conditions we can filter out polynomial drifts without ever estimating their coefficients. For example suppose that Z(x) is defined in a 1-D space and that we want to estimate the value of $Z(x_0)$ at point x_0. If we want to remove up to quadratic terms we have to include the following nonbias conditions:

$$\sum \lambda_i = 1$$

$$\sum \lambda_i\, x_i = x_0$$

$$\sum \lambda_i\, x_i^2 = x_0^2$$

This can be extended to 2-D. Similar types of relations must be satisfied by both the x and y coordinates of the sample points, and also their cross products.

So to filter out a linear drift the following conditions are needed:

$$\sum \lambda_i = 1$$

$$\sum \lambda_i\, x_i = x_0$$

$$\sum \lambda_i\, y_i = y_0$$

To filter out a quadratic drift, the following conditions would also have to hold:

$$\sum \lambda_i\, x_i^2 = x_0^2$$

$$\sum \lambda_i\, y_i^2 = y_0^2$$

$$\sum \lambda_i\, x_i\, y_i = x_0\, y_0$$

Because of the complexity of the mathematics involved in I.R.F.-k we shall not go into much detail here. For more information see Matheron (1973) and Delfiner (1976).

The main difference between the matricial kriging systems for ordinary kriging and kriging using I.R.F.-k is that there are several extra non-bias conditions at the end of the matrix, corresponding to the conditions described above for filtering out the drift.

Another difference between ordinary kriging and its non-stationary equivalent is in the range of "variogram" models that may be used. In the same way that the intrinsic hypothesis with its single universality condition allows us to use a much wider range of models for the variogram than we could for the stationary or the intrinsic cases, so here we have an even wider choice for the generalized covariance model as it is called. This allows us to use some higher order polynomial models such as cubics, as well as more unusual models like $h^2 \log(h)$. The latter is particularly important since kriging with an I.R.F.-1 and this covariances is equivalent to a thin-plate spline interpolation.

SOME SPECIAL TYPES OF APPLICATIONS

Geostatistics is now widely used in both the mining and petroleum industries for estimating point and block values. One of its main uses in the oil industry is in estimating the shape of geological horizons (surfaces) from seismic data. In a seismic campaign, readings are taken at very closely spaced points along lines called profiles. This very special spatial arrangement of data (very close readings along widely spaced lines) resembles the data configuration in the ship-board measurements made for acoustic measurements of fish, and also for measurements of the sea-floor, and of the gravimetric and magnetic fields. As the particular estimation and computing problems posed by this arrangement of data have already been successfully solved for seismic readings for the oil industry, there is every reason to think that the same method can be applied successfully to acoustic measurements of fish.

REFERENCES

Matheron, G. (1973). The intrinsic random functions and their applications. Adv. in applied Prob. Vol 5. pp 439 – 468.

Chilès, J.P. (1977). Geostatistique des phénomènes non-stationnaires. Doc.Ing. thesis, CGMM, Fontainebleau 152 pp.

Delfiner, P. (1976). Linear estimation of non-stationary phenomena. Proc NATO ASI Rome 1975 "Advanced geostatistics in the mining industry" ed M. Guarascio et al. Reidel Pub. Co., Dordrecht, Holland pp 49 – 68.

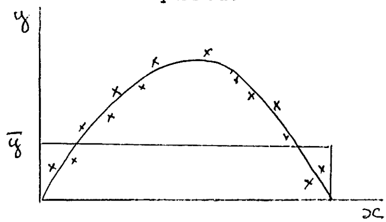
GENERALIZED LINEAR MODELS.

by

Gunnar STEFANSSON

INTRODUCTION

Models for bivariate data need to take into account the error structure and the mean response, as expected at each point. A common historical approach has been to model acoustics data as estimates of a common mean (possibly within squares). Residuals from such computations will automatically exhibit much autocorrelation along transects. This has been used in the past as an indicator of the wrong error structure, and the autocorrelation has been incorporated in the variance estimate, usually with the result of raising it considerably. A basic fallacy in this approach becomes obvious when simple univariate sampling of a quadratic response is considered, and the area under the curve is required.



The integral can easily be estimated by computing the average response and multiplying by the range in the x-values. This approach is equivalent to assuming that all measurements are really measuring a constant level. If the response is heavily quadratic, then a test of serial correlation will yield a significant result.

This, however, is an indication that the underlying structure should be taken into account when computing the integral.

In the example illustrated, it would be easy to fit a quadratic response, and then to integrate the response function.

Similarly, for spatial data it is possible to fit models to the response, z, e.g. using polynomials in **x** and **y**.

OVERVIEW

A generalized linear model (**GLM**) contains a description of the expected response at each location along with a description of the probability distribution at each point.

Thus a typical **GLM** for spatial data might be:

$$E(Z) = \alpha + \beta x + \Gamma y + \delta d$$

$$Z \approx \text{Gamma distribution}$$

Here, (x, y) represent the location of the response, Z, and δd denotes a depth effect.

The expected response at a given location is therefore described in the above as a linear function of location, plus a depth effect. Since the effect of depth is not known, it is usually entered as a factor with several levels. The resulting model is an **ANCOVA** model.

GLM IN GROUNDFISH SURVEYS.

Generalized linear models have been used for analyzing groundfish surveys. Typically the models are of the following form:

$$E(Z) = \exp\left(p(x,y) + \beta y + \Gamma w + \delta d \right)$$

Here, p(x,y) is a step function (i.e. region effect), a polynomial or even a station effect (in the case of fixed stations). This particular model uses data from several years, estimating a biomass index by including a year effect (βy). Other terms can be included as necessary, e.g. wind speed, depth strata etc.

To complete the definition, a distribution needs to be assumed. Typically it is found that the variance is proportional to the square of the mean. This suggests either a log-transform or explicit **GLM** modelling using a log-link and gamma (or negative binomial) distribution.

GLM models can be fitted using the **GLIM** statistical package.

APPLICATION TO ACOUSTIC DATA.

Multi-year models clearly do not apply in this case and in fact for schooling pelagic species the model for the structure of the mean will mainly include a function of the location.

A simplified analysis of acoustic measurements of a single school was attempted by fitting polynomials in location to Z, $\log Z$ and $\log(Z+1)$. Numerical problems required the use of orthogonal polynomials. Even in this case, a log-linear model using up to a fourth degree polynomial in x and y yields an R value of only about 0.5.

It is therefore obvious that acoustic data will be hard to model using ordinary response surfaces. Since the surfaces are quite complex, even for small schools, a very high degree polynomial may be required as a rule.

This may be an indication that smoothing techniques are to be preferred to response surface techniques, although the issue should not be considered quite settled yet.

REFERENCES.

GLIM77 User manual.
Numerical Algorithms Group.

AN ILLUSTRATION OF THE ORDINARY KRIGING PACKAGE "GULFKRIG" FOR MAPPING AND ESTIMATING ABUNDANCE OF THE RESOURCE SURVEYED BY SETS OF DATA 1 TO 5.

by

Gerard Y. CONAN and Elmer WADE

The purpose of this exercise was to demonstrate the advantages and disadvantages of using the straightforward technique of ordinary kriging and to identify possible adaptations of this technique suitable for the characteristics of the sets of data provided.

The process of ordinary kriging assumes that, in the absence of information on neighbouring values of the variable studied, the mean and variance of the estimate at a given point will remain the same, whatever the location of this point. Further, the variance will be independent from the mean. Emphasis is set on the spatial covariance effects, i.e. on the similarities in departure from the overall mean among values observed or expected within a limited vicinity. The covariance effects are assumed to be of an identical nature for all locations of the area surveyed.

Ordinary kriging allows to somehow correct preferential sampling, involuntary or not, in areas of high or low densities by attributing lower weights in the estimation of the overall mean for sample points set closely apart. It also allows to generate a fine mesh grid of estimated points suitable for drawing a high definition map.

As any statistical tool, ordinary kriging provides only approximate estimates. The robustness of the tool is defined as how well it will resist to departure of the data from the basic assumptions and still provide reliable estimates.

Traditionally in fisheries data, random, or at the least, non preferential sampling is assumed. Emphasis is set on the shape of the sampling distributions in order to define confidence limits, but spatial covariance effects are totally ignored.
A strong relationship between the variance and the mean is generally recognized.

Ordinary kriging emphasises spatial covariance effects, but assumes that there is no relationship between the variance and the mean. The data points do not necessarily need to be chosen non preferentially, and a ship course, as in acoustic surveys, is an acceptable sampling scheme.

GENERAL PROCEDURE

For each set we first mapped the course of the survey and the location of the data points along the Norwegian coast. A digitized map of the coast of Norway was provided to us by NOAA, Woods Hole USA. We then calculated the experimental variogram for the data points, and fitted where possible a spherical model. A contour map and a three dimensional representation of fish abundance were generated along a fine grid calculated by point kriging.

A contour map of the kriging variance was calculated , and the area within which a global estimate of average fish abundance could be reasonably calculated, was measured within a chosen contour of isovariance. For certain data sets, the coast line was used as a boundary more restrictive than the variance contours.

The global average density within the so defined polygon was calculated by block kriging and the associated variance was estimated.

SPECIAL CASES

The practical difficulties encountered were of 4 types

1) Lack of information
   In the case of set number 2, transects follow the coast of a narrow deep fjord, but there is no information on the variability across the fjord. Anisotropy could have been easily incorporated in the calculations (differences in covariance range along and across the fjord). In the absence of information we blankly assumed that there was no anisotropy, information taken along the fjord was extrapolated across. This is likely to have generated overestimates of biomass if the resource was concentrated along the coasts.

2) Misleadingly redundant information.
   In set 5, the route of the ships overlapped, but after a time lag. The values sampled from the two ships may not be equivalent due to changes in spatial distributions through time. No corrections could be made.

3) Overabundant data
   For Block kriging within a large polygon in order to draw a global estimate, our algorithm requires the inversion of a matrix of N*N, where N is the number of data points. Some of the sets exceeded the 8 Meg. memory capacity of our workstation. We had to partition the data into geographic subunits.

4) High patchiness of data
   In all sets the fish are concentrated into discrete patches separated by areas of abundance either null or extremely low. The structure within the patches sometimes strongly differs between the patches. The variograms calculated for the overall area was meaningless in case 5, while patch variograms consistently showed neat spatial structures.

   We therefore resolved to treat as different entities each of the patches and the overall areas of low density. We simply identified the patches on a preliminary contour map, but kriging using indicative variables (see Desbarats) would have provided a similar preliminary information. Global estimates were obtained separately for the subdomains, and then pooled after weighting their contributions by the area of their respective domain.

RESULTS:

Provided in table 2

COMMENTS

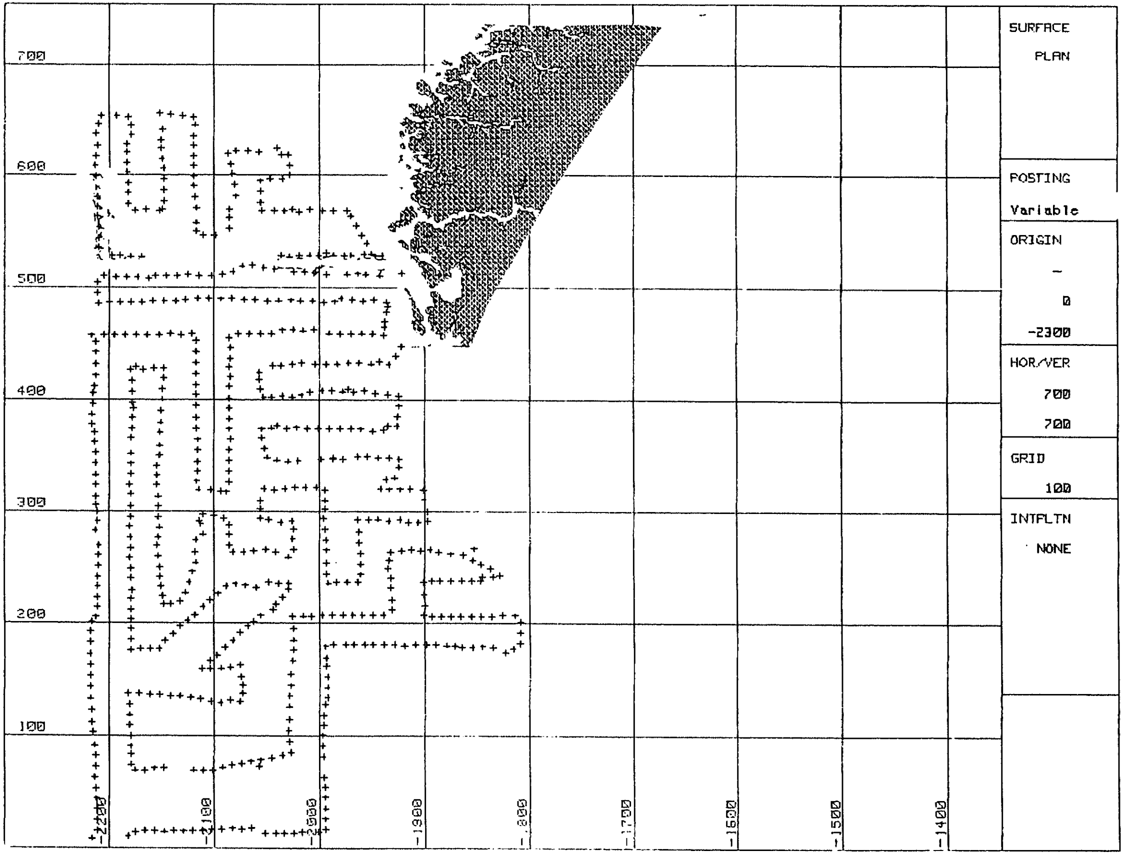
Ordinary kriging could be applied satisfactorily to all sets of data. However, lack of information in set 2 did not permit adequate estimates unless a drastic assumption of isotropy could be made.

It would have been preferable for the purpose of kriging that the data not be regularized, but provided in a ping by ping form.

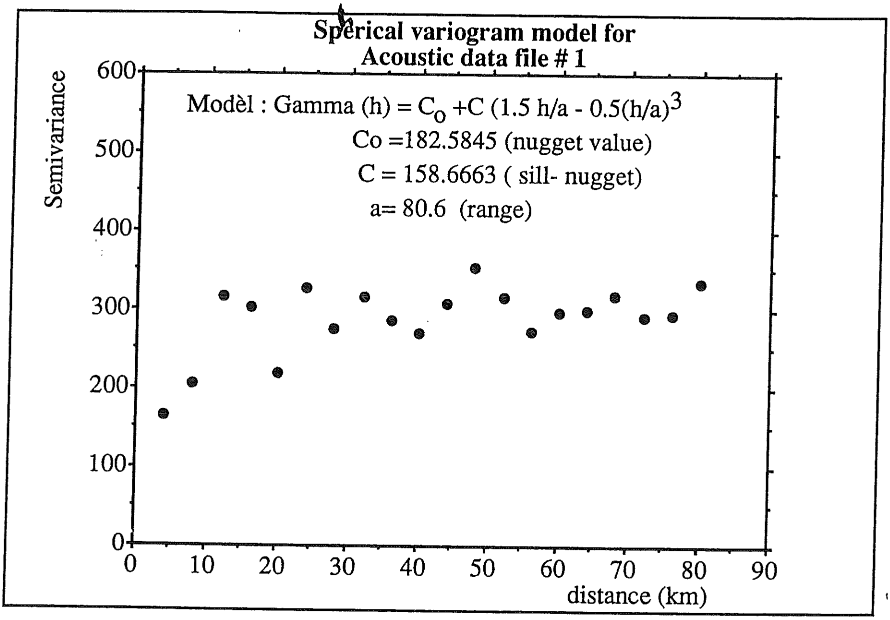
A grid coverage allowing variogram estimates in all directions would have been preferable.

REFERENCES:

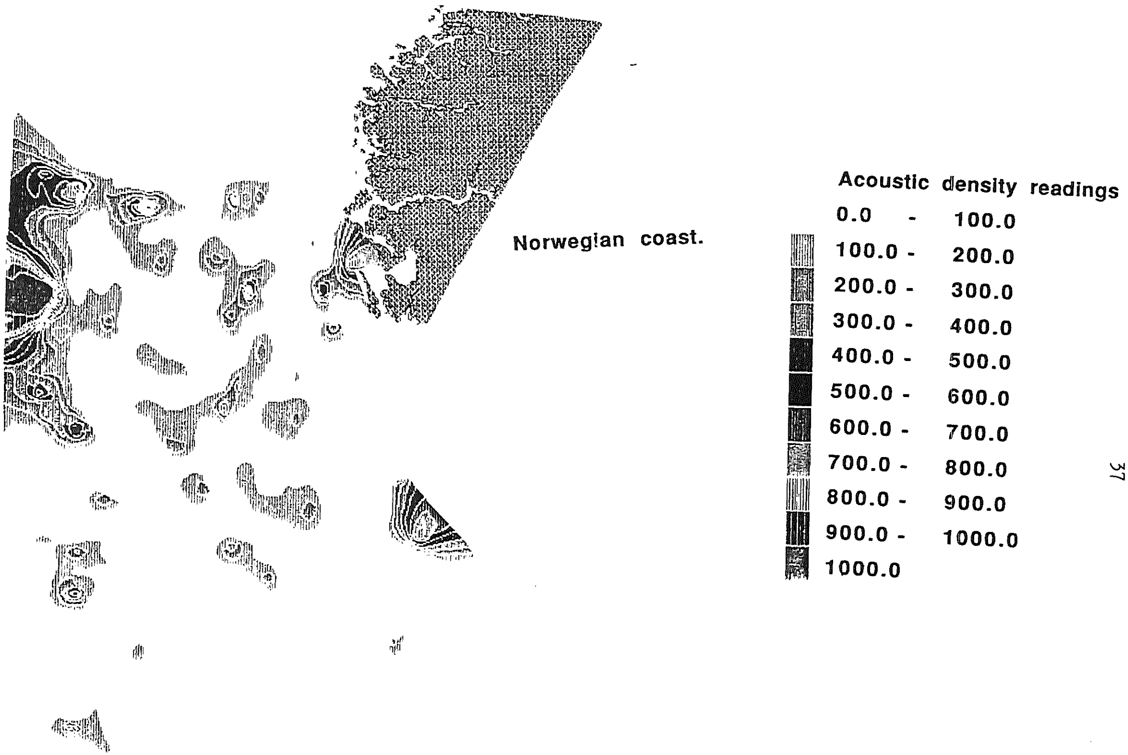
Conan, G.Y., U. Buerkle, E. Wade, M. Chadwick, and M. Comeau, 1988. Geostatistical analysis of spatial distribution in a school of herring. ICES CM.



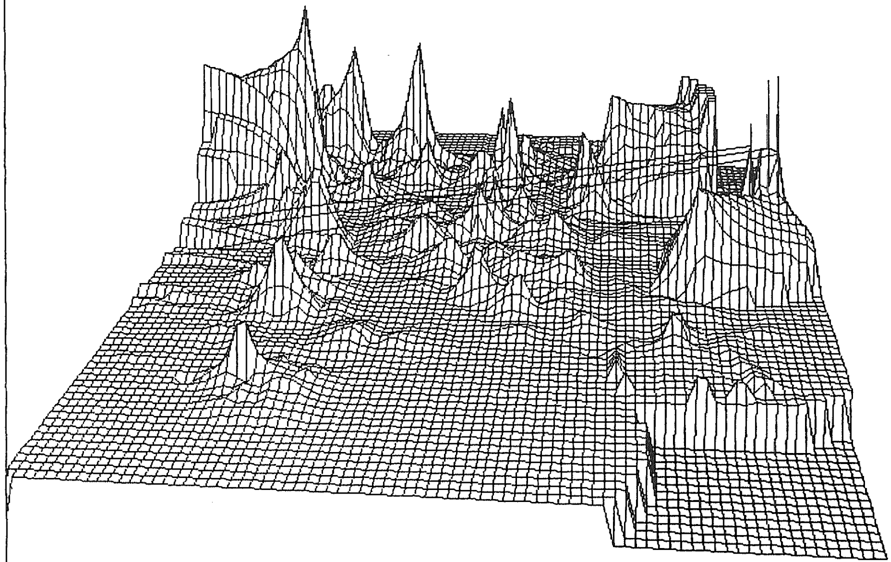
Distribution of survey sample points for acoustic density readings located near Norway. This is for test data file # 1



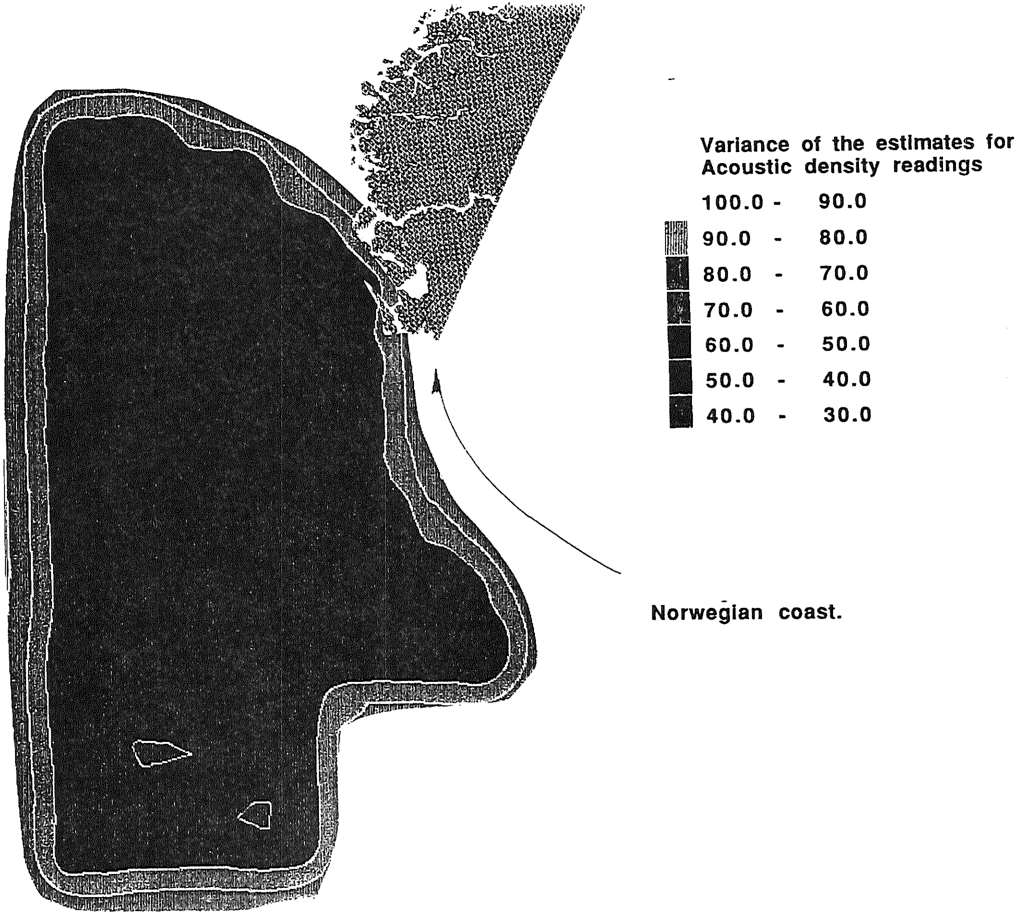
Variogram model for test data file #1.



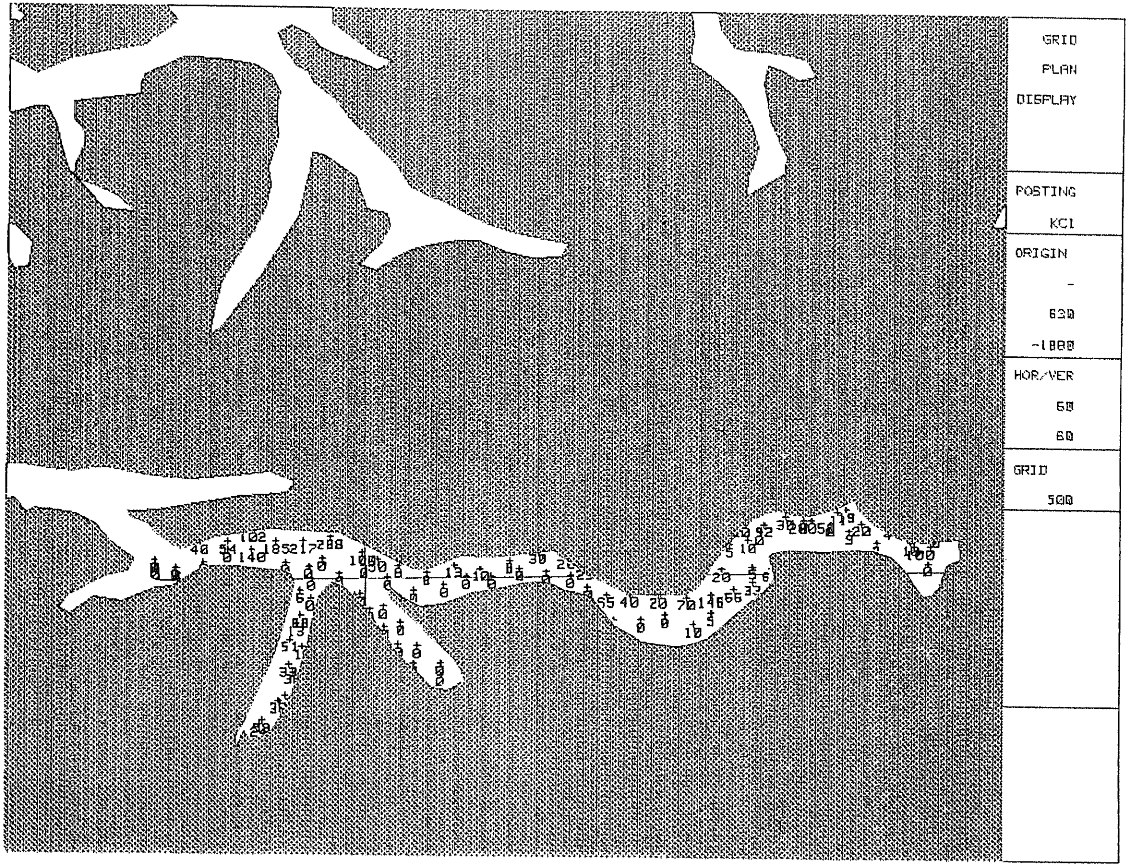
Contour diagram showing acoustic density readings as determined by geostatistical methods for test data file #1.



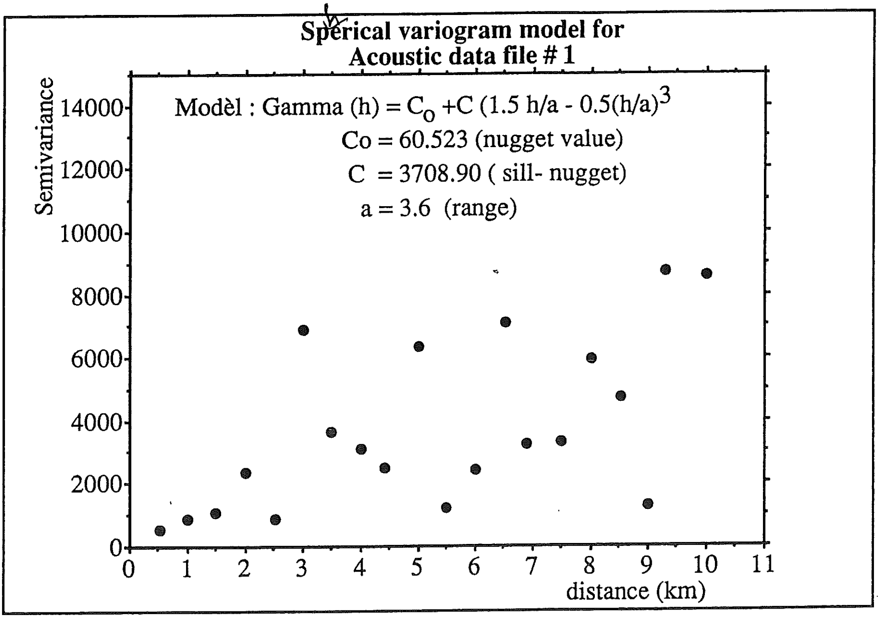
Three dimensional representation of acoustic density as determined by geostatistical analysis. for file #1

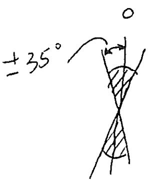


Contour diagram showing variance of estimates as determined by geostatistical methods for acoustic density readings for test data file #1.

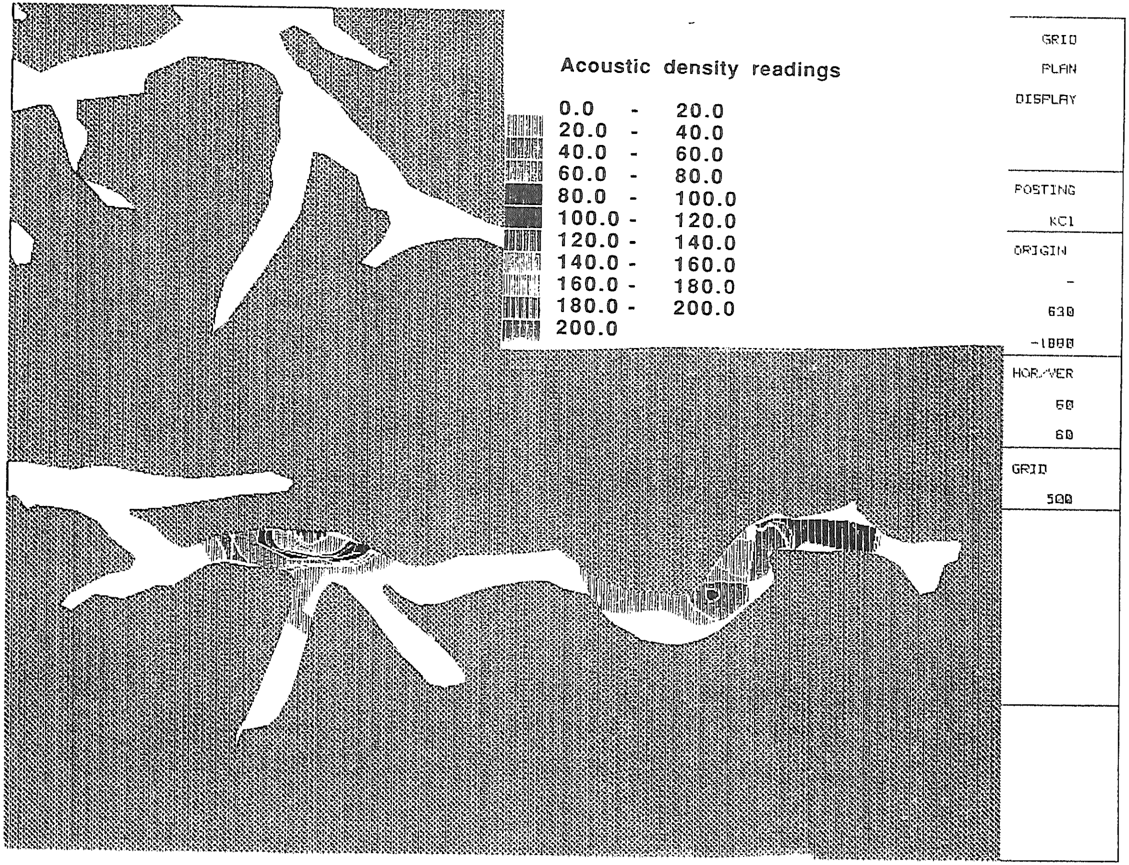


Distribution of survey sample points for acoustic density readings located near Norway. This is for test data file # 2.

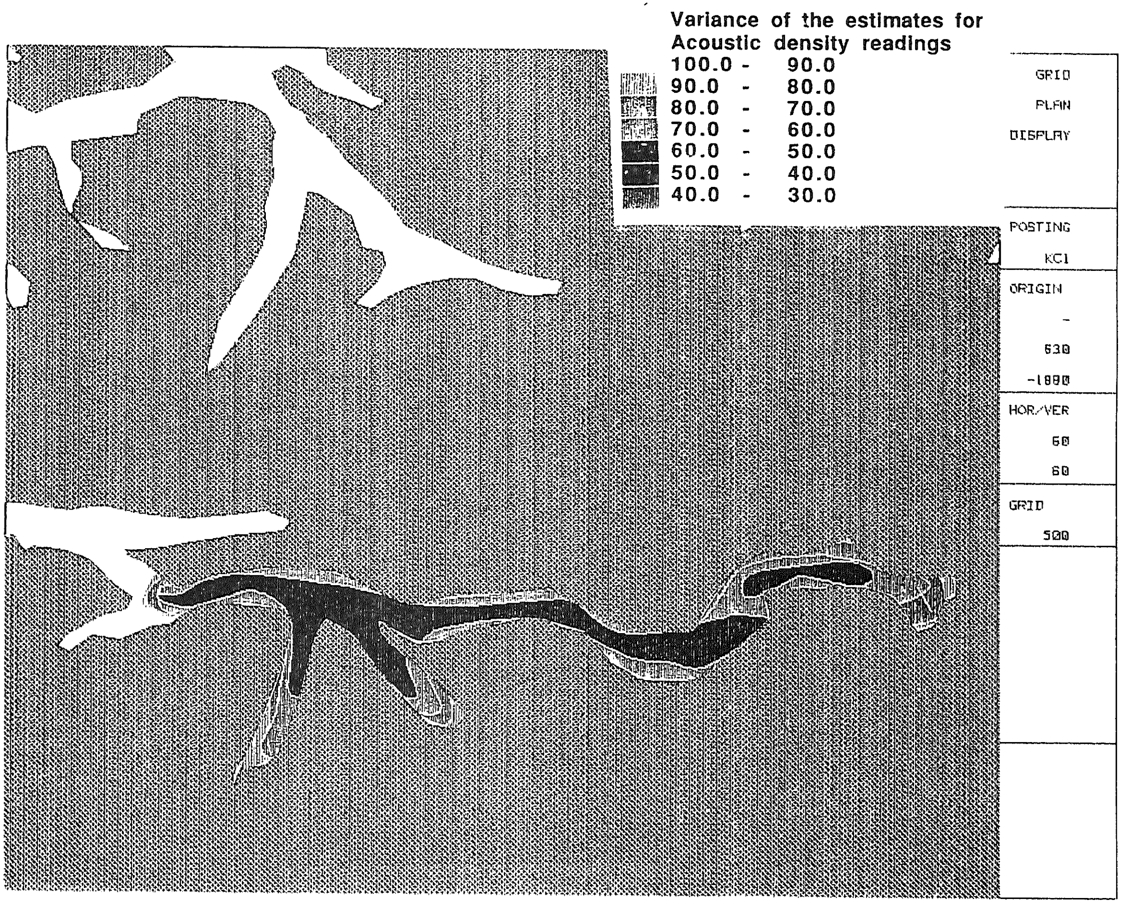




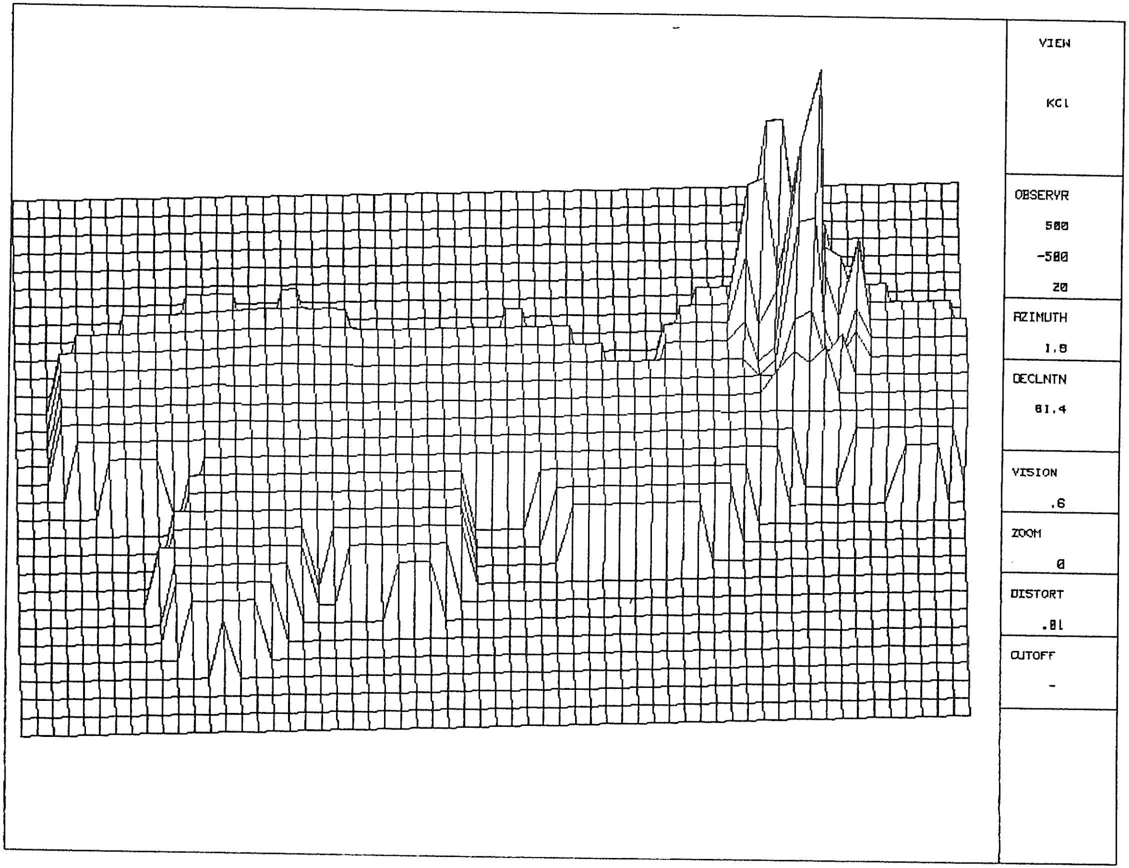
Variogram model for test data file #2.



Contour diagram showing acoustic density readings as determined by geostatistical methods for test data file #2.

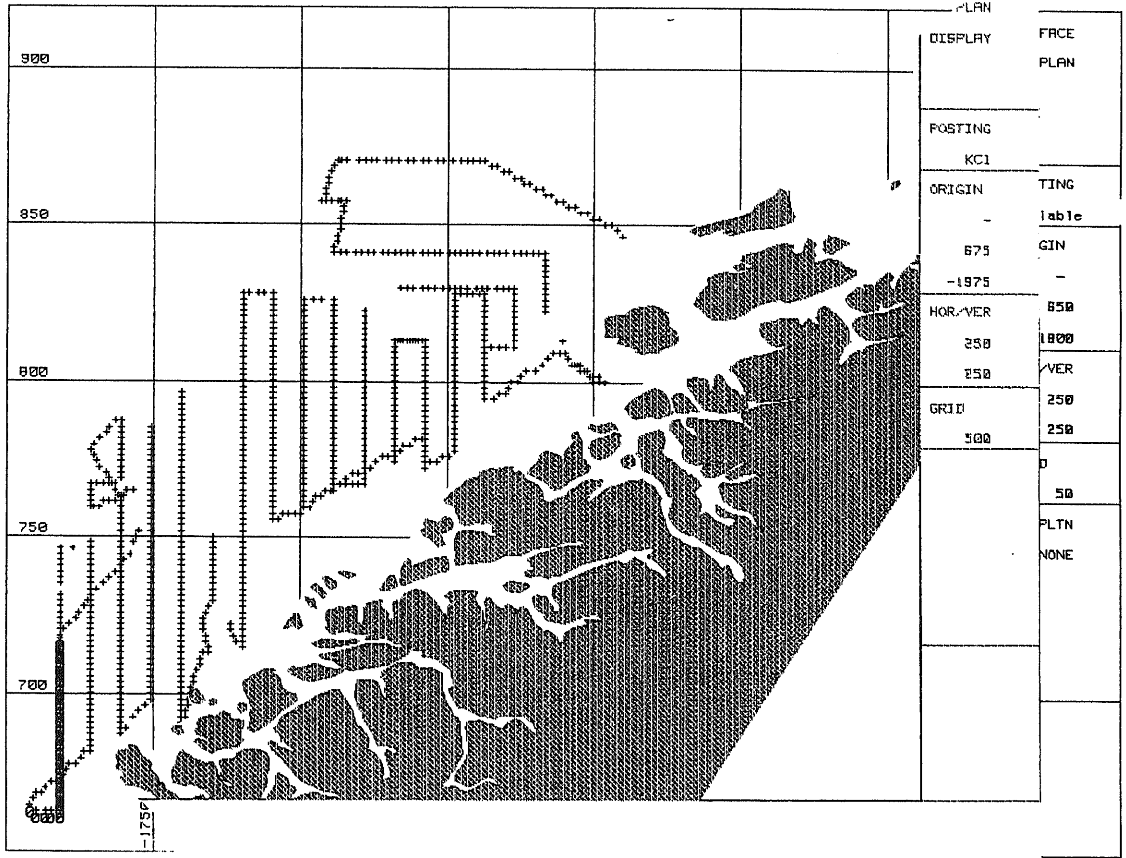


Contour diagram showing variance of estimates as determined by geostatistical methods for acoustic density readings for test data file #2

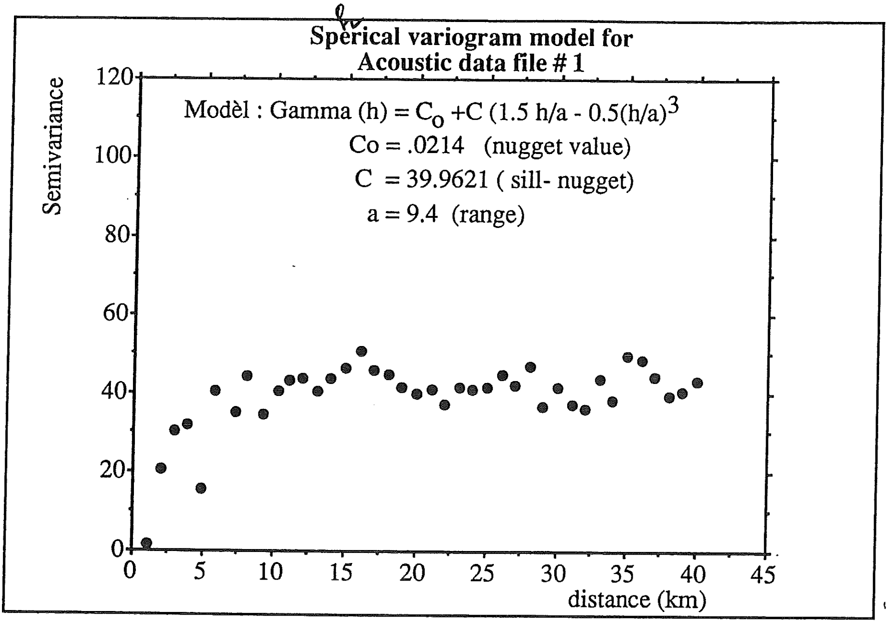


DATE: 30 Mar 1990

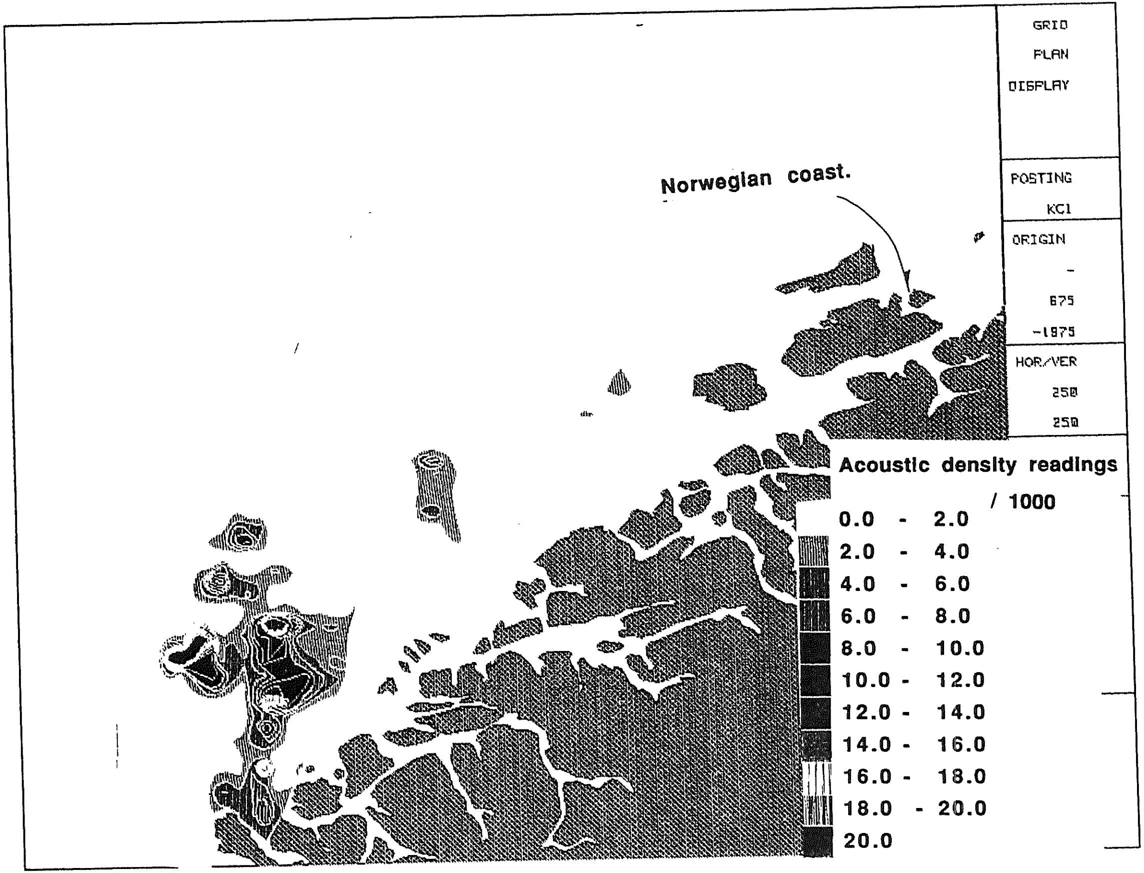
Three dimensional representation of acoustic density as determined by geostatistical analysis for test data file #2.



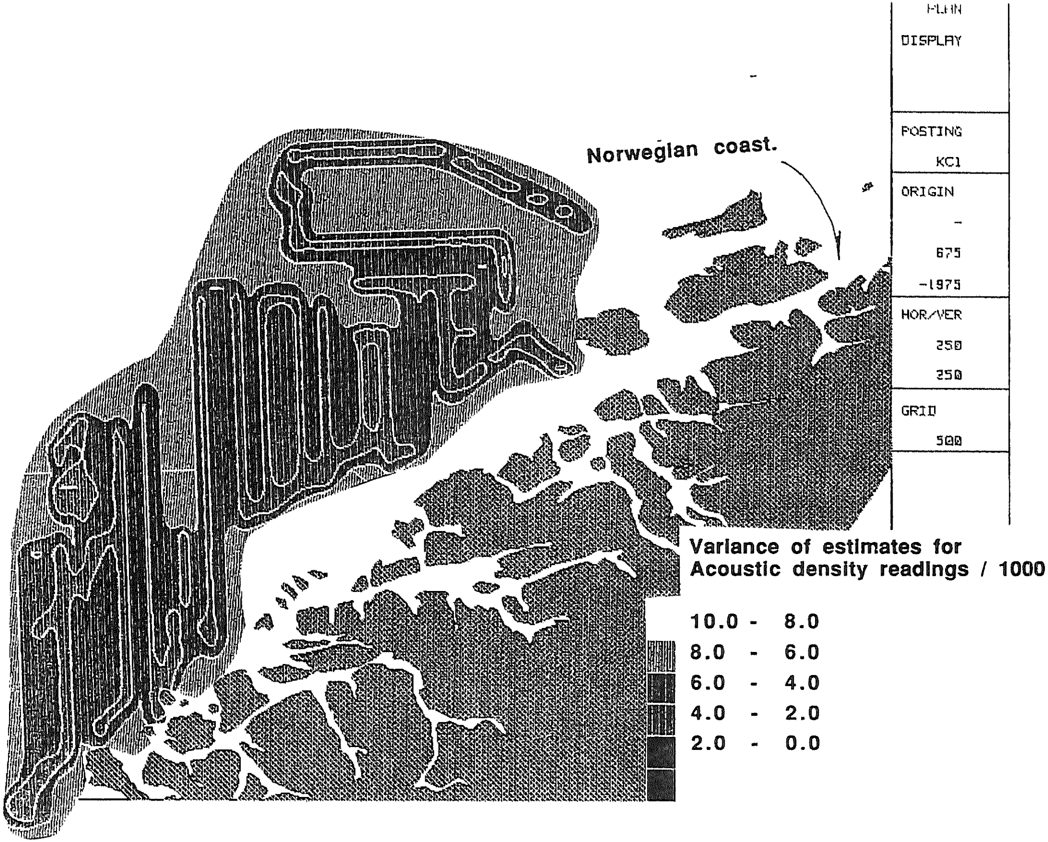
Distribution of survey sample points for acoustic density readings located near Norway. This is for test data file # 3.



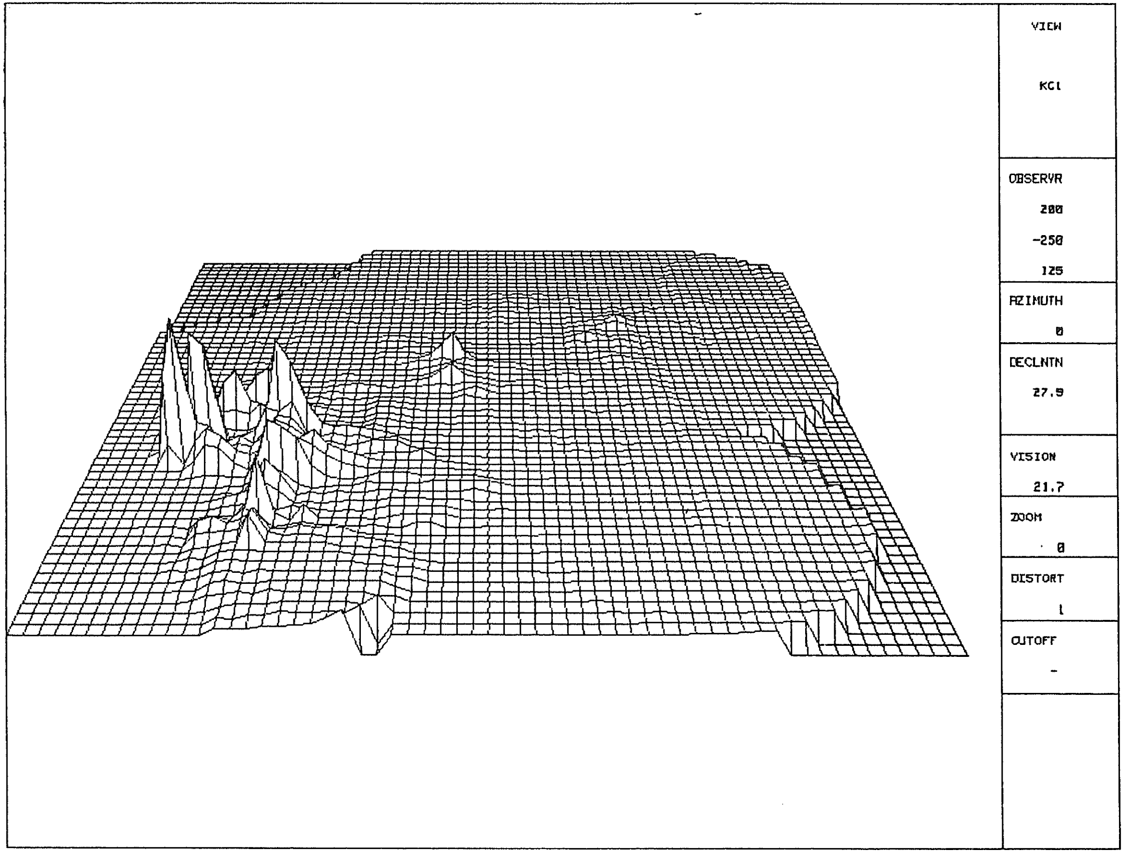
Variogram model for test data file #3.



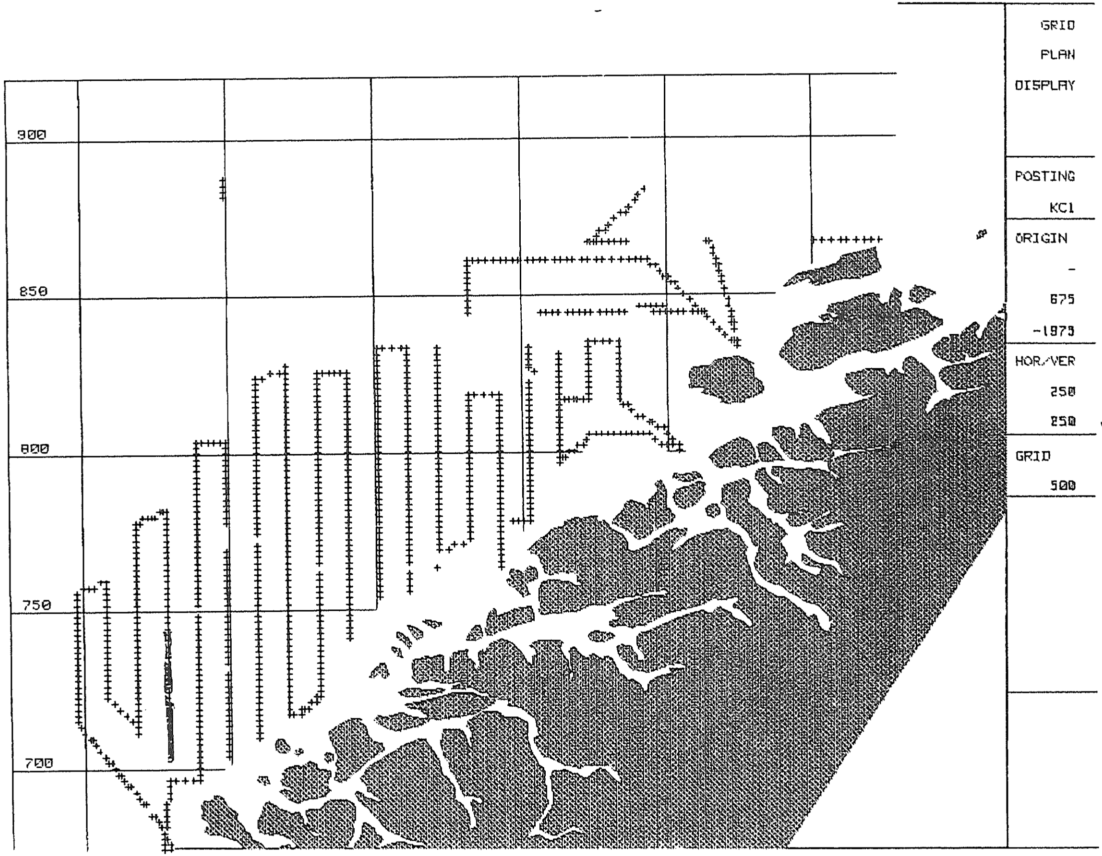
Contour diagram showing acoustic density readings as determined by geostatistical methods for test data file #3.



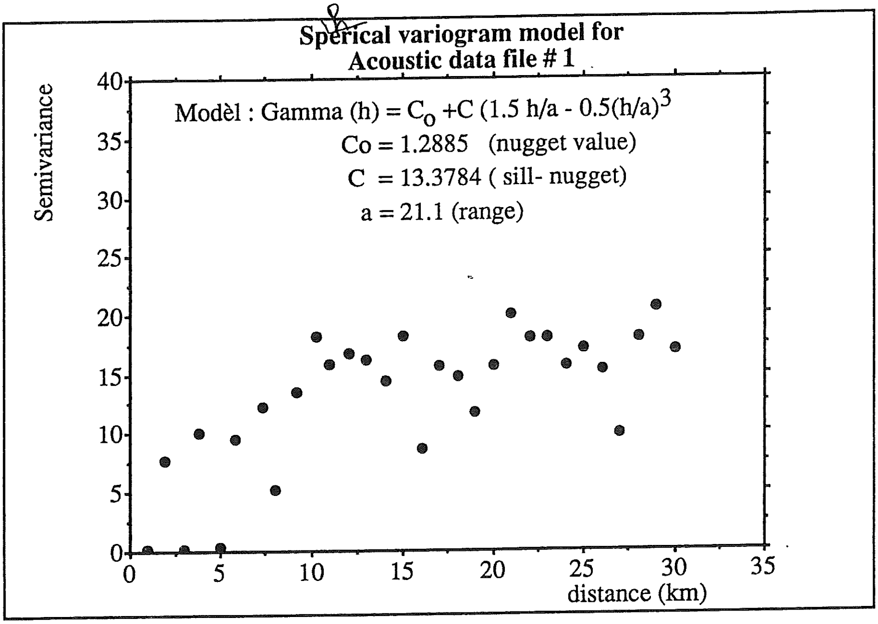
Contour diagram showing variance of estimates as determined by geostatistical methods for acoustic density readings for test data file #3.



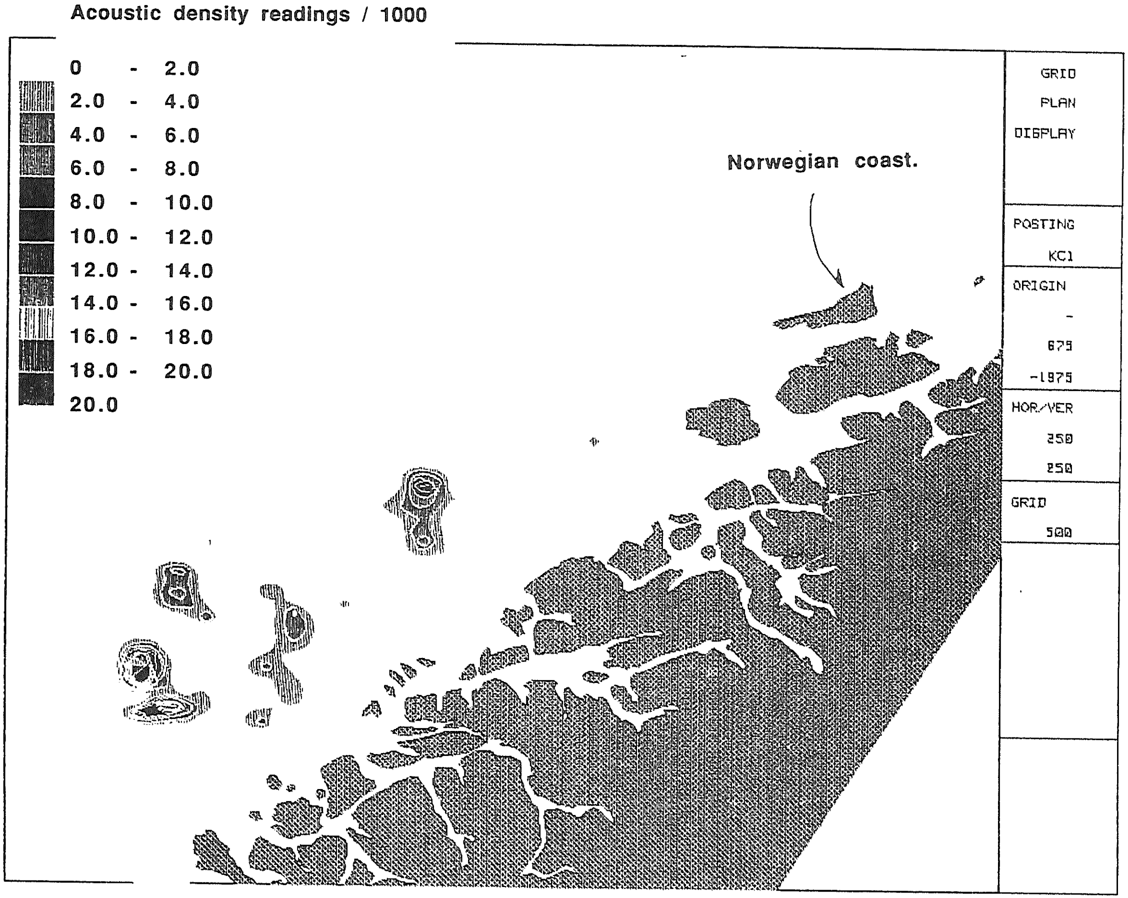
Three dimensional representation of acoustic density as determined by geostatistical analysis for test data file #3.



Distribution of survey sample points for acoustic density readings located near Norway. This is for test data file # 4.

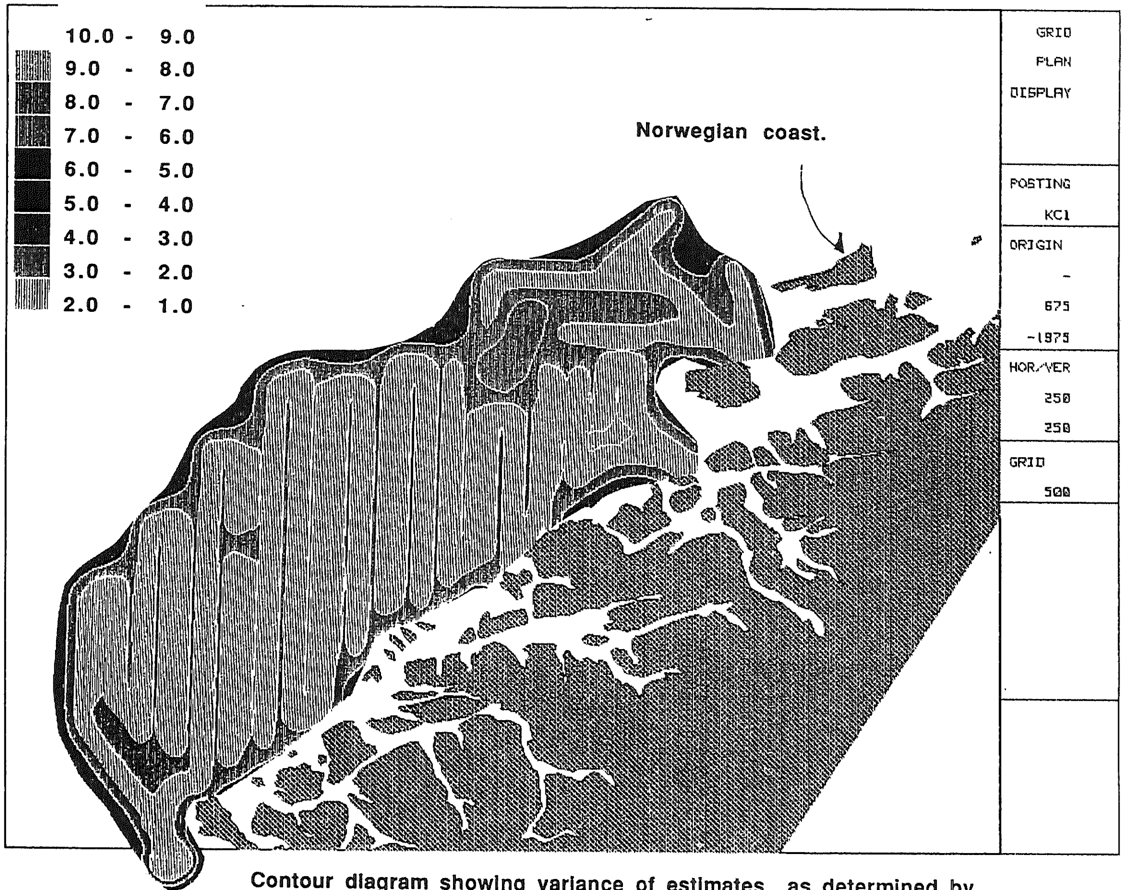


Variogram model for test data file #4.



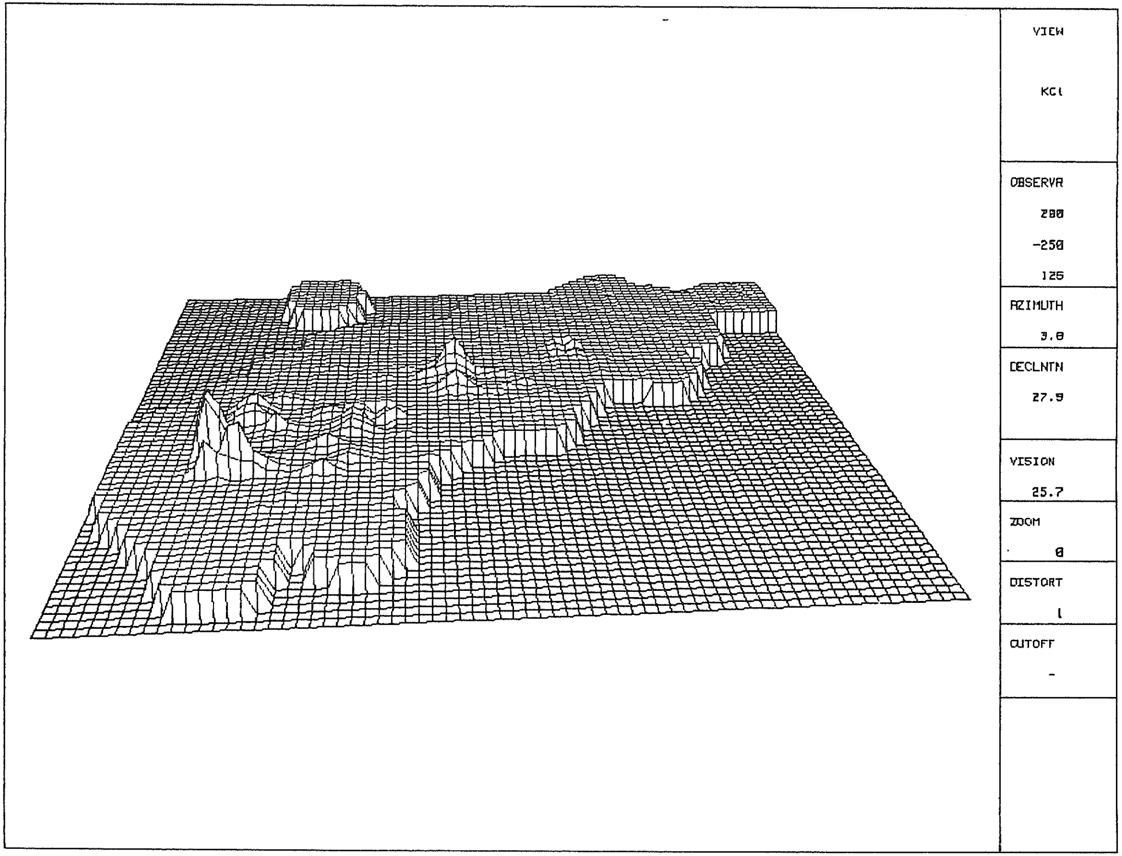
Contour diagram showing acoustic density readings as determined by geostatistical methods for test data file #4.

Variance of estimates for
Acoustic density readings / 1000

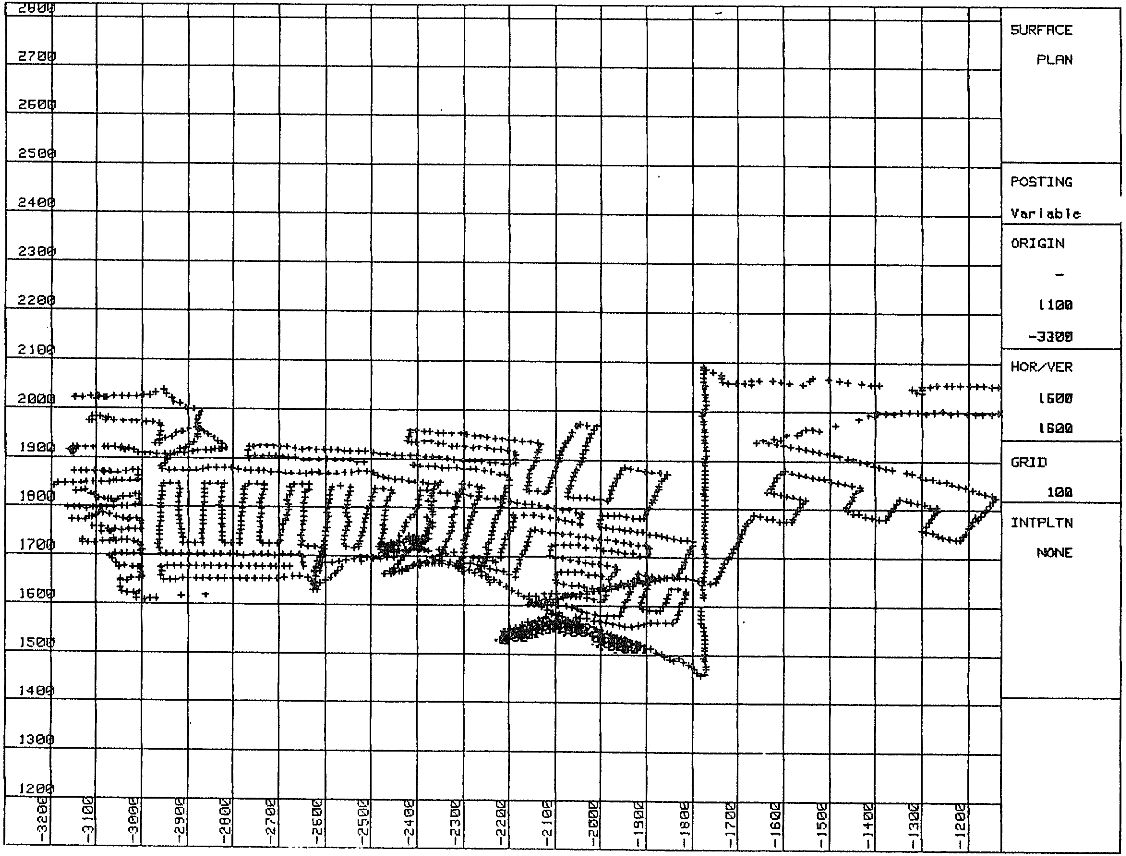


Contour diagram showing variance of estimates as determined by geostatistical methods for acoustic density readings for test data file #4

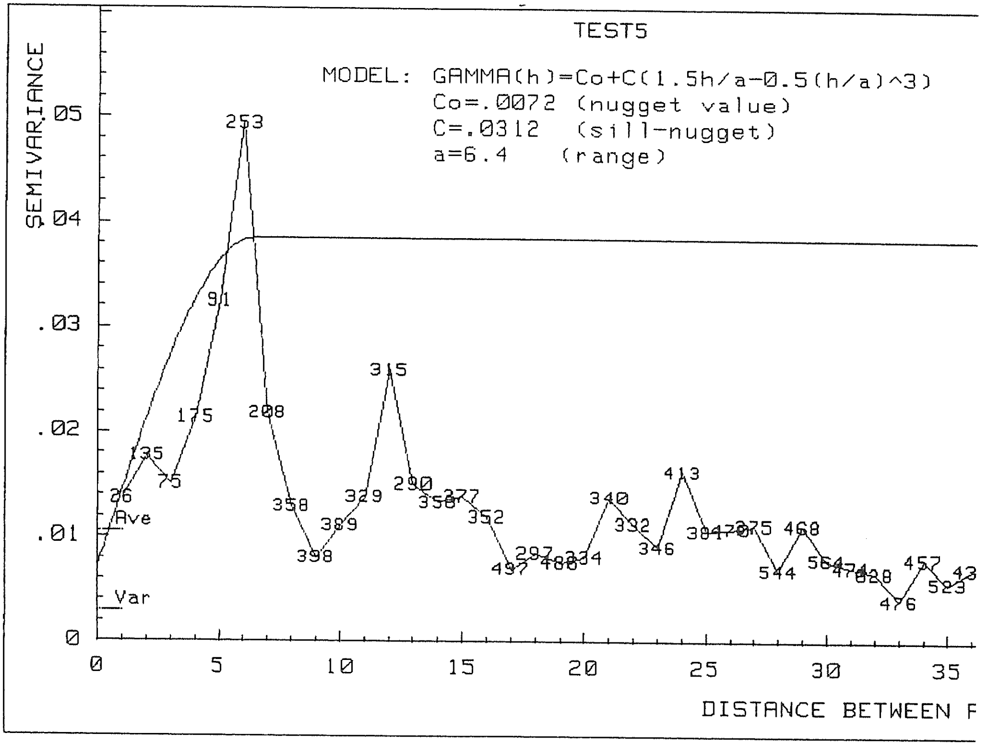
D 34 del 2



Three dimensional representation of acoustic density as determined by geostatistical analysis for test data file #4.

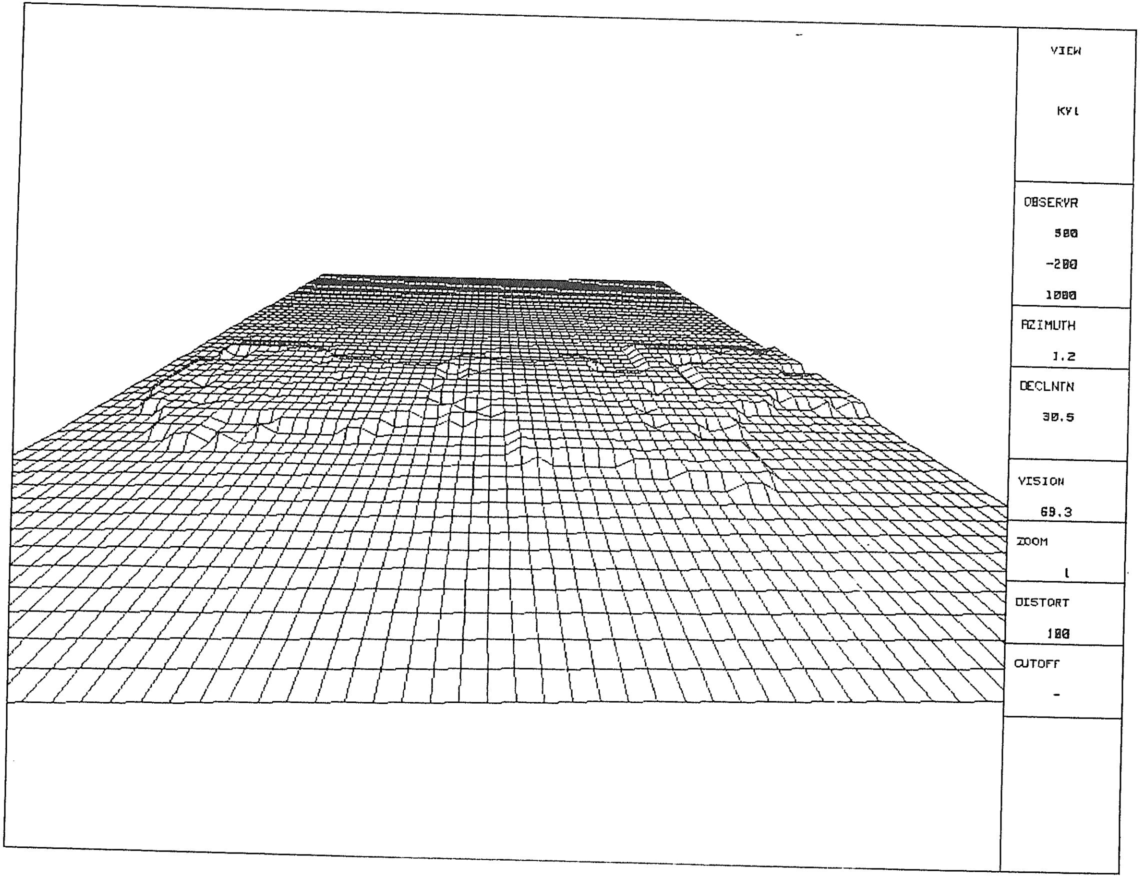


Distribution of survey sample points for acoustic density readings located near Norway. This is for test data file # 5.



Variogram model for test data file #5.

Contour diagram showing acoustic density readings as determined by geostatistical methods for test data file #5.



Three dimensional representation of acoustic density as determined by geostatistical analysis for test data file #5.

APPLICATION OF GEOSTATISTICS TO FISHERIES ACOUSTICS: EXAMPLE OF TEST3.

by

Jean GUILLARD and Daniel GERDEAUX

For this set we have considered all the data localized on transects, included zero value data, but the data localized on inter-transects are excluded. This consideration is based on the fact that the regularization on the N-S direction is not the same that the one on the W-E direction; and you can't mixed data from different supports (Guillard and al., 1987).

So all the data on the North of the map are eliminated. We have defined a polygon to limite the area (9 103 n.m2) (fig 1).

The new set is composed of 591 data point, and the arithmetic mean is: 2085.

The variography is performed on all the data; the mean variogram (fig. 2) is well modeled by a spherical model and the phenomenon is supposed to be isotropic.

nugget effect: 1.0 107, range: 12, sill: 4.0 107.

A global estimation was attempted on all the data included in the defined polygon. But the number of data points is too high for the program used (BLUEPACK) and we had to reduce this number. We regularize the phenomenon in one direction using the mean of four data in the N-S direction. The unit sample is now the mean of four data. So the variogram is the same one, but regularized (MATHERON, 1970). The new set of data is composed of 150 data points, arithmetic mean is: 2032.

The regularized variogram is: nugget effect: 2.5 106, range: 12, sill: 1.0 107.

The estimator of the mean using block kriging is 2089,and s: 298.

TEST 3

arith.mean without inter tr.	arith.mean regularized	Block Kriging
2085	2032	2089 s: 298 (14%)

REFERENCES:

GUILLARD J., GERDEAUX D., CHAUTRU J.M., 1987. The use of geostatistics for abundance estimation by echointegration in lakes: the example of lake Annecy. Int. Symp. Fish. Acoustics. June 22-26, Seattle, 17 pp.

MATHERON G., 1971. The theory of regionalized variables and its applications. Les cahiers du CMM, fasc. 5. ENSMP, 211 p.

APPLICATION OF GEOSTATISTICS TO ACOUSTIC DATA ACOUSTIC DENSITIES OF A PELAGIC FISH OFF THE COASTS OF NORWAY (62-64 lat, data set 4)

by

Pierre Petitgas

DATA.

They are collected along N-S transects regularly paced. The histogram is very skewed with a long tail. The 13 highest values (out of 653) represent 55% of the arithmetic mean and 20% of the variance. The set comprises 60% of zeros which define mainly the limits of the spatial extension of the fish.

VARIOGRAPHY.

It is performed on the raw data including the zeros. A structure is identified: range= 7 nautical miles (N.M.), nugget= 51% of the variance. No clear autocorrelation exists between transects, for the shorts distances. An isotropic variogram is retained. The bordering zeros tend to lower the variogram where as the high values tend to higher it.

GLOBAL ESTIMATION.

It is performed on the area defined by the transects. Each transect is attributed a rectangle of influence (b_i). The mean is estimated by the weighted average of the mean values of the transects (weight: b_i Σ b_i). The acoustic data are spatially continuous along the transects so that the transects are fully known (like galleries in mining). The variance of estimation is calculated using the variogram and the approximation principle developed by Matheron (1970) for sampling with regular paced galleries in a geological prospection. The variance is not a kriging variance nor is the mean a kriging estimate.

The variance is a variance of extension in space: it is the error done when estimating the mean value of the rectangle b_i by the mean value of its central transect.

Results:

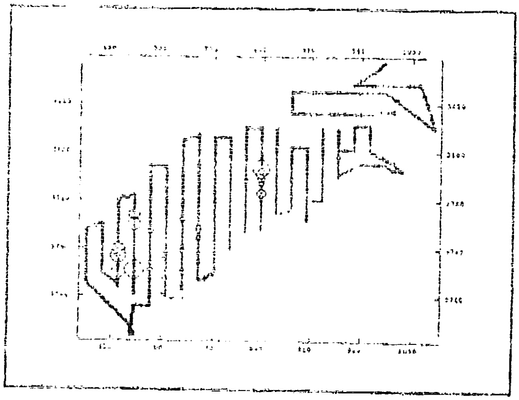
m = 1125.7 area = 3.10^3 (N.m^2)
variance of estimation = 18544.81
var. est. / m = 0.12

The nugget effect represents 74% of the variance of estimation. A biological interpretation of the nugget could lead to diminish it. Diminishing the nugget seems the only way to ameliorate precision on the global estimate.

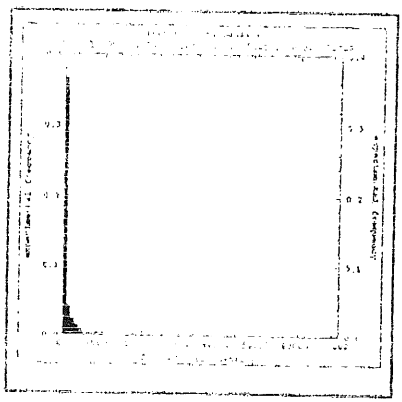
GEOMETRICAL PROPERTIES AND A DISJUNCTIVE KRIGING MODEL.

The mean is very dependent on a few very high values. The geometrical properties in space have been investigated. A model is fitted that takes into account the very quick transitions in space from one order of magnitude to another. Each cutoff on the histogram defines in space a geometrical set. It is shown that in the geometrical set defined by the values over 500, the probability that a value may trespass a higher cutoff cannot be well predicted (pure randomness with the set A_{500}). Surfaces of geometrical sets are estimated by disjunctive kriging. It is emphasized that the geometry and the localisation of the sets representing a high pourcentage of the stock (A_{500} represents 60%) may be linked to the determinism of the variations of the total quantity. The disjunctive kriging global estimate is presented as a tool when there has been preferential sampling of some set.

FIGURES



Proportional representation of the raw data. The zeros are blackened.



Histogram of the raw data

5

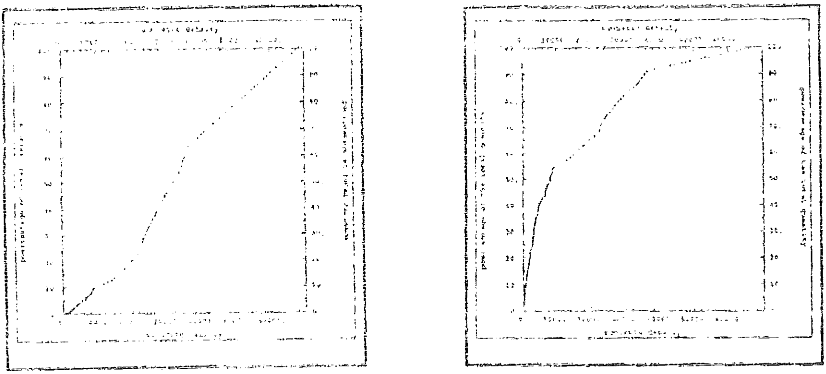
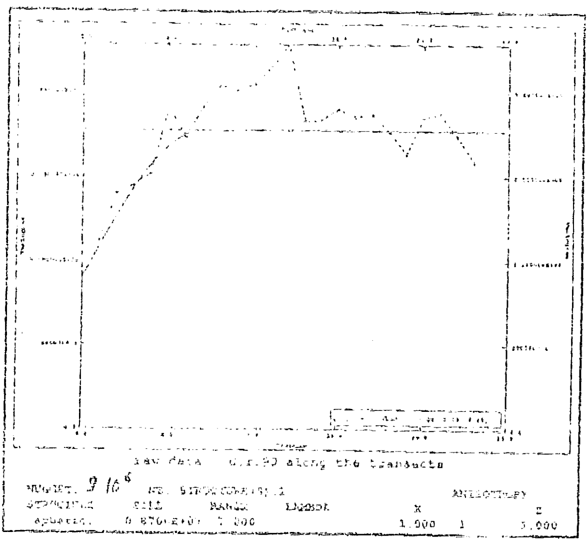
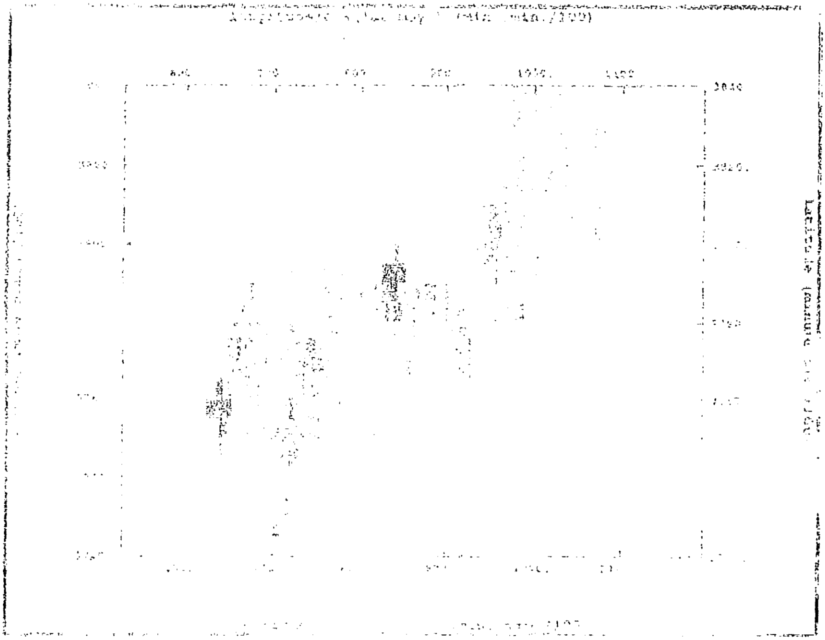


FIG.? ... of the histogram to the ... and ...





[illegible]
[illegible]
[illegible]

ANALYSIS OF SETS 1 TO 5 USING SPLINE APPROXIMATION OF STOCK DENSITY

by

Dimitri STOLYARENKO

Fig. 1 - 8 presents the results on the test data sets No.1-5 which were processed by the SSDSS - Spline Survey Designer Software System (Stolyarenko, 1987) with IBM-compatible personal computer. The major feature of the method is incorporation of depth information because fish is associated with trophic and environmental conditions which are more similar along depth contour than along perpendicular. Position of every measurement point is coupled with depth. Bathymetric information is used to describe space anisotropy. Therefore bathymetric maps have been digitized for areas of the test data sets No.1-4 and then computer maps of bottom relief was reconstructed to provide the opportunity for computation depth at all points. The test data set No.5 has been supplied with depth information. Therefore bottom relief was reconstructed only with these data. Because maximum number of measurements for SSDSS (MS DOS version) is 400, the great data sets were parted on 2 (the set No.1), 3 (the sets No.3 and 4) and 4 (the set No.5) subareas with ca. 20% overlapping.

Data set No.1. The map of stock density (Fig. 1) is coupled with the map of bottom relief (Fig. 2) which have been used to reconstruct the stock density. One of the borders is 70m depth contour.

The high concentrations near the western slope of the Norway Deep (black zone on Fig. 2) extends along depth contours. So great measurements of two parallel tracks are usually related more closely along depth contour and are to be merge in common concentration. On contrary the two great measurements are to be separated as two patches. Conventional biomass estimated equals 4.04 million units (square meters of fish backscattering cross section per square n.m. of area).

Data set No.2. Fig. 3 shows stock density for the fjord. The map of bottom relief is very rough. Therefore the weight of depth on compare with weight of distance is very low. Biomass estimated equals 13.2 thou. units.

Data set No.3 and 4. Fig. 4 and 5 shows the maps of stock density for two sequential years which are coupled with the bathymetric map (Fig. 6). Concentrations are related with banks and slopes of troughs. Estimates of biomass are 7.84 and 3.51 mil. units respectively.

Data set No.5. Fig. 7 presents the map of stock density which is reconstructed on the base of data of two vessels. Biomass estimated equals 0.90 mil. units. The part of the area with very close located points of measurements is presented with large scale (Fig. 8). The last map shows where it was necessary to carry out additional tracks (or redistribute research efforts). This example illustrates the importance of adaptive sampling during survey. The Spline Survey Designer Software System is an appropriate tool for survey design in real time on the board of research vessel.

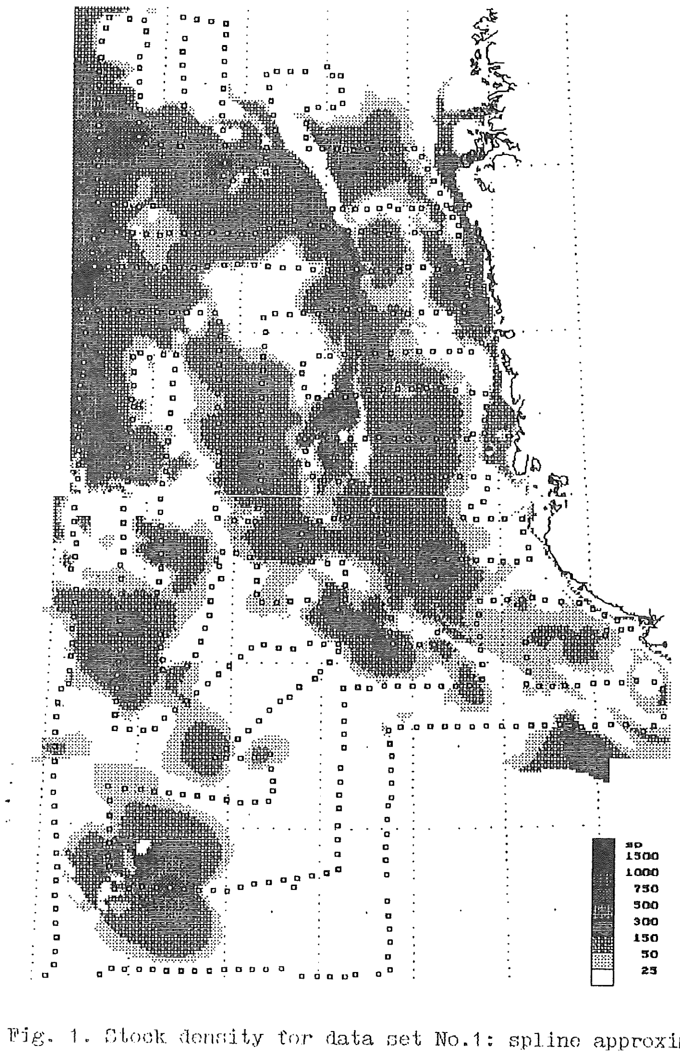


Fig. 1. Stock density for data set No.1: spline approximation.

Fig. 2. Bathymetric map for area of data set No.1.

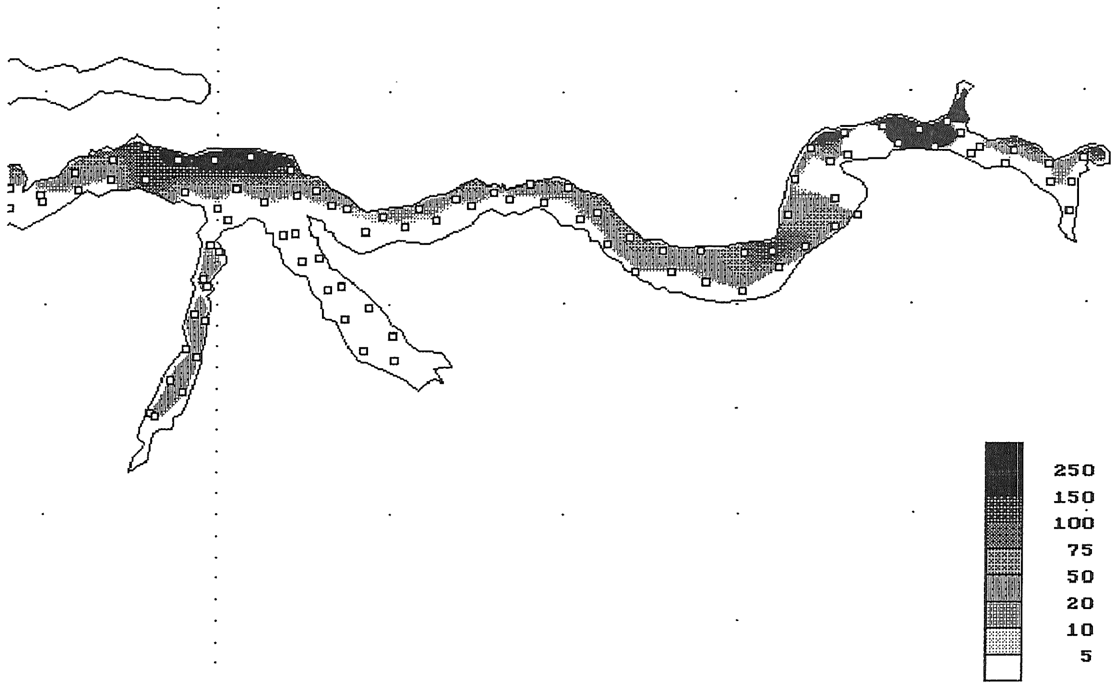


Fig. 3. Stock density for data set No.2: spline approximation.

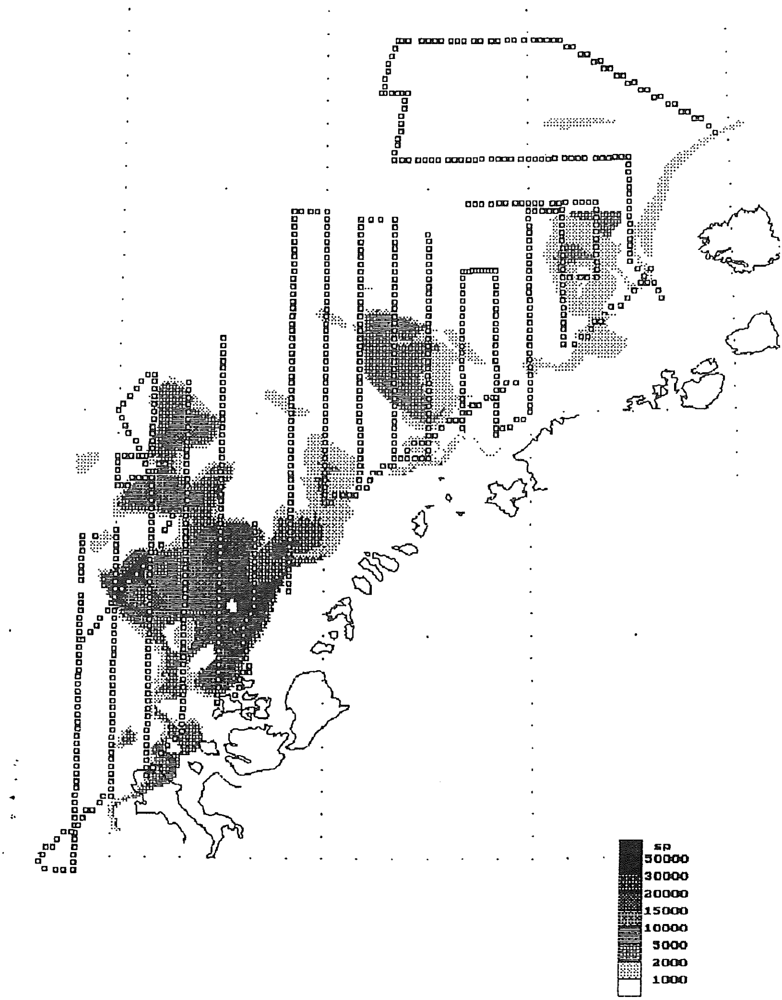


Fig. 4. Stock density for data set No.3: spline approximation.

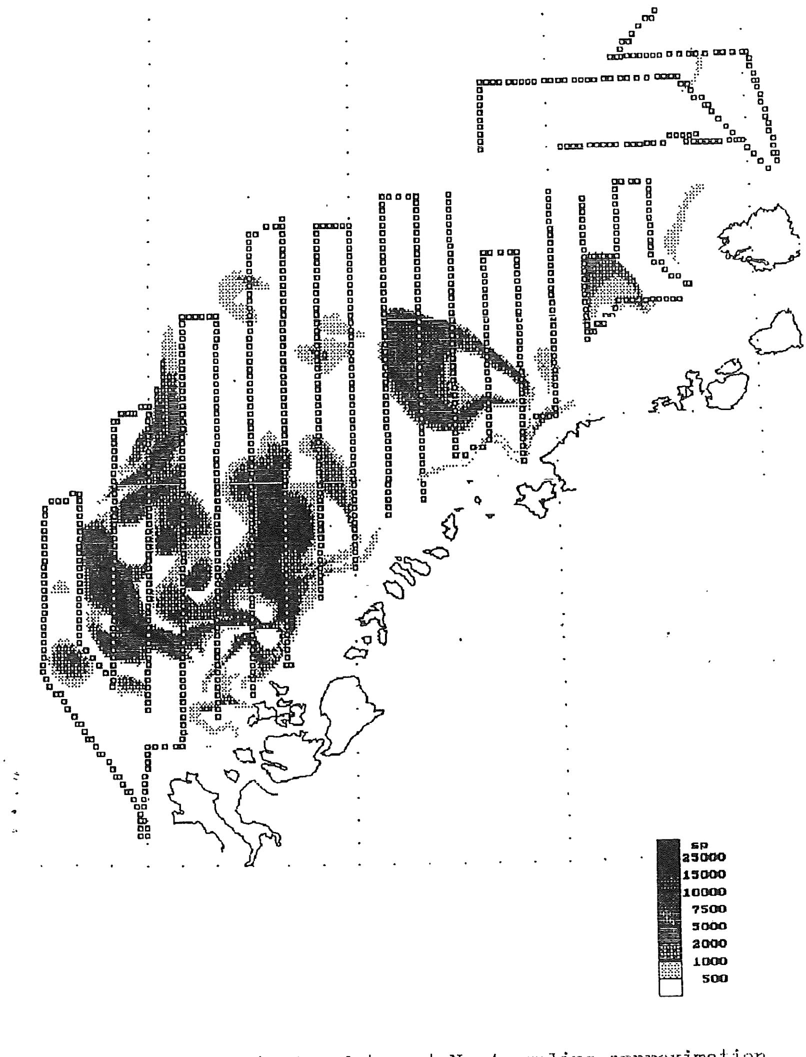


Fig. 5. Stock density for data set No.4: spline approximation.

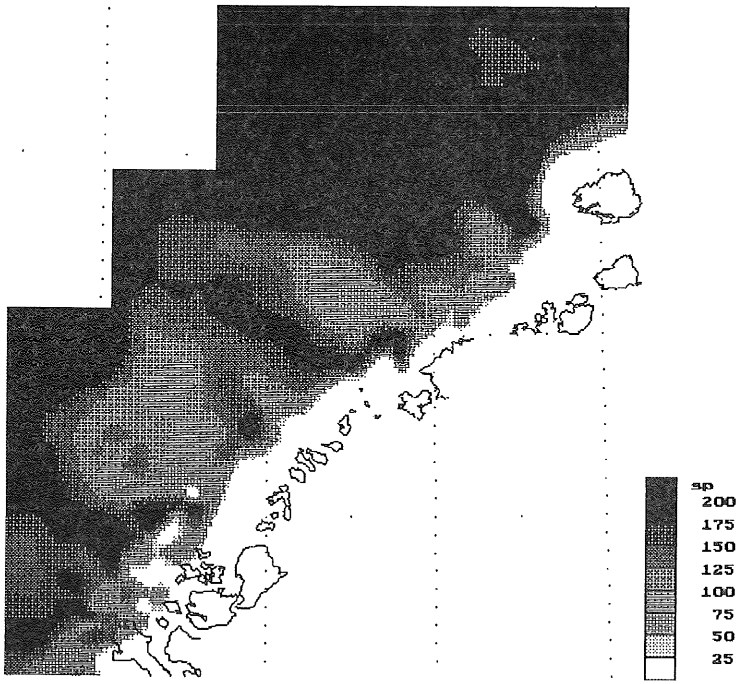


Fig. 6. Bathymetric map for area of data set No.3 and 4.

Fig. 7. Stock density for data set No.5 (the whole area): spline approximation.

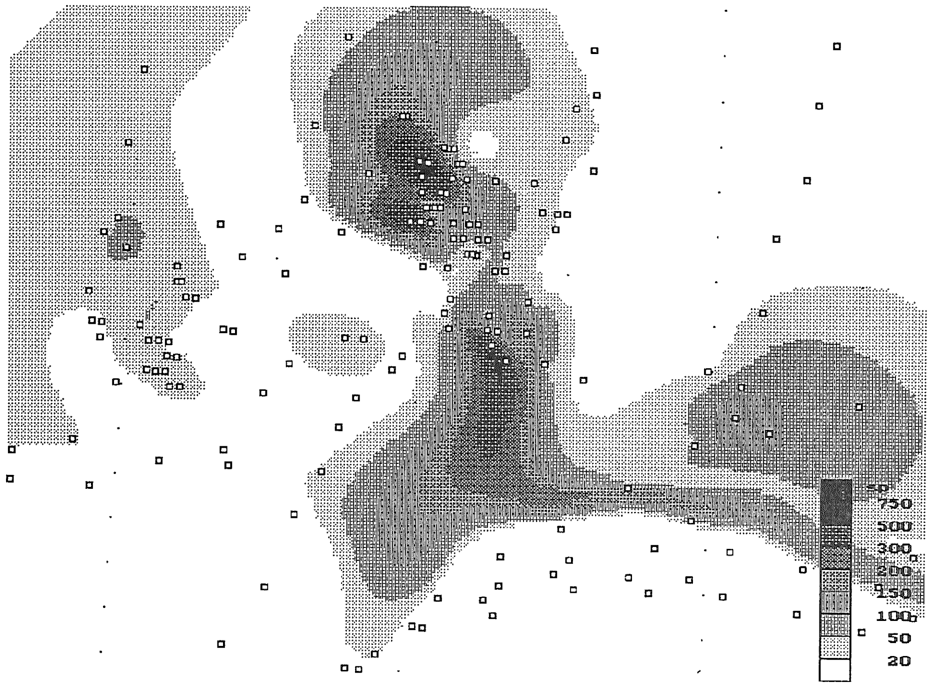


Fig. 8. Stock density for data set No.5 (the part of the area studied): spline approximation.

BOX/COX TEST AND POWER SERIES ANALYSIS OF SETS 1 TO 5

by

John SIMMONDS

Data from acoustic surveys is collected along transects with an approximate regular grid. The grid may not be uniformly spaced over the full area. The data is analysed to give some geographical or spatial distribution and an overall estimate of mean density, and total stock within the survey area. The distribution of the stock is regarded as non-stationary in a statistical sense. There will be some parts of the area that contain predictably more fish than others, giving sub-areas or regions of different mean density. In addition the nature of fish distributions suggests that the variance will be dependant on the density. A possible relationship would be that the variance is proportional to square of the mean density. The purpose of this analysis is to determine the distribution, use the most efficient estimator for the mean and to allow calculation of the confidence limits.

The data is divided into predetermined 'rectangular' strata based on a lat/long grid. The strata size are determined on the basis of the expected rates of change in mean density, and variance, and the sampling density, such that a minimum of 1 transect per strata is guaranteed and the sampling is uniform within any one strata. Typically the strata dimensions might be selected as two to four times the 'range' determined from a variogram. The choice of rectangular strata is not implicit in the analysis procedure, and depth related or any other predetermined strata may be used. Where a strata intersects the coastline the area of the strata is reduce accordingly. In order to calculate the total population the estimated mean density per strata is raised by the area of the strata, which is assumed to be flat, ie a trapezium.

The data from each strata is analysed and the results combined for the complete survey. The data is tested for a suitable power transform using a Log Maximum Likelihood test due to Box and Cox 1964. The technique may be combined with a delta function to remove all zeros, (Aitchison 1955, Pennington 1983), or it may be applied with an offset moving zero values to a positive value. Probably more correctly the zeros should be classed in two ways, first as true zeros, and secondly as zeros within a positive random function and dealt with accordingly. The analysis presented at this study group used the delta function method for all zeros.

If the maximum of this test lies between +0.5 and 0 a power transform of 1/2, 1/3, 1/4, 1/6 or Log may chosen. The individual data points are then transformed to the power domain and the mean and variance calculated. These two values are then transformed back into the arithmetic domain and corrected accordingly. The full method including the inverse transforms is described by Maclennan and MacKenzie 1988.

The underlying assumptions are that the complete area is covered by strata, the between strata variance may be ignored, the within strata statistics are stationary and that the transform predicted by the Box/Cox test is the appropriate transform. This technique ignores any spatial structure within each strata and assumes that each data point is independent. Under these circumstances the estimate of variance will be correct for the mean calculated in the transform domain. Because the distribution of the data has been defined the variance may be used to compute the confidence limits. It would, however, be inappropriate to assign the confidence limits to the arithmetic mean of the original data. The main advantages of this method are that it is compatible with existing analytical techniques, independent of the operator and may be implemented in a simple analytical package. However it is limited to data sets with skew not significantly greater than the log normal distribution.

Data Analysis of 5 Norwegian data sets.

Responses Surfaces

Box Cox Test/ Power Series Analysis

Data Set 1

This data set was analysed on a 0.5 by 1.0 latitude longitude rectangle. The Box Cox test for this data set defined the 1/6 power transform as the best power transform, with confidence limits that excluded other transforms. This transform was used to calculate the mean and the variance. These were transformed back to the arithmetic domain. The mean density for each strata was raise by its area. The results for this data set are shown in table ?.

Data Set 2

This data set exhibited a number of features. The survey consisted of two tracks which indicated significant differences between north and south sides of the fjord and a large shoal which contributed 40% or more of the stock. The data set was analysed on a 1/12 by 1/6 latitude longitude rectangle. The Box Cox test for this data set defined the log transform as the best power transform, however the confidence limits included other transforms. The log transform was used to calculate the mean and the variance. These were transformed back to the arithmetic domain. The mean density for each strata was raise by its area taking into account the ratio of sea and land in each strata. The results for this data set are shown in table ?. There must be serious reservations about the applicability of this analysis for this situation. The uncertainty of the correct transform, and the fact that the assessment using the arithmetic mean gives a stock estimate 4.6 times the size. This problem is caused primarily by the non stationarity of the data. Analysis of this data in a real situation requires very careful scrutiny of the full detail of all acoustic data and separate assessment of the single aggregation. <u>The Power Transform method of assessment is not suitable for this spatial distribution.</u>

Data Set 3/4

These data sets are for two surveys in the same area on different occasions and have been treated similarly. Analysed on a 0.5 by 1.0 latitude longitude rectangle. The Box Cox test for these data sets defined the 1/2 power transform as the best for set 3 and the 1/6 power transform for set 4. In both cases confidence limits excluded other transforms. These transforms was used to calculate the mean and the variance. These were transformed back to the arithmetic domain. The mean density for each strata was raise by its area. The results for this data set are shown in table ?.

Data Set 5

This data set was analysed on a 0.5 by 1.5 latitude longitude rectangle. The Box Cox test for this data set defined the log power transform as the best power transform, with confidence limits that excluded other transforms. This transform was used to calculate the mean and the variance. These were transformed back to the arithmetic domain. The mean density for each strata was raise by its area. The results for this data set are shown in table ?.

General

It is interesting to note that in two cases, sets 1 and 4 the transformed and corrected estimates exceeded by a small amount the arithmetic estimate. In the case of data set 3 they were equal, and in sets 2 and 5 the arithmetic estimate was higher than the transformed estimate. This Confirms in a very small way that provided the correct transform is chosen bias is not introduced by this procedure.

References

Box G E P and Cox D R 1964 An analysis of transformations J. R. Sat. Soc. B 26: 211–252

Aitchison J. 1955 On the distribution of positive random variable having a discrete probability mass at the origin. J Am. Stat. Assoc. 50: 901–908

Pennington M. 1983 Efficient estimators of abundance for fish and plankton surveys. Biometrics 39: 281–286

MacLennan D N MacKenzie I G 1988 Precision of Acoustic Fish Stock Estimates. Can. J. Fish. Aquat. Sci Vol 45: 605–616

Results
(Box Cox Transform)

Data Set	Mean SA	CV	Area	SA*Area
1	68.2	9.5%	54.7E3Nm^2	3.7E6
2	48.1	37%	48.9Nm^2	2.4E3
3	1327	7.4%	5.52E3Nm^2	7.3E6
4	560	9.0%	6.14E3Nm^2	3.4E6
5	8.9	8.1%	126E3Nm^2	1.1E6

ANALYSES OF TEST DATA SETS 3 & 4 USING TRANSECTS AS STRATA

by

Neal J. Williamson

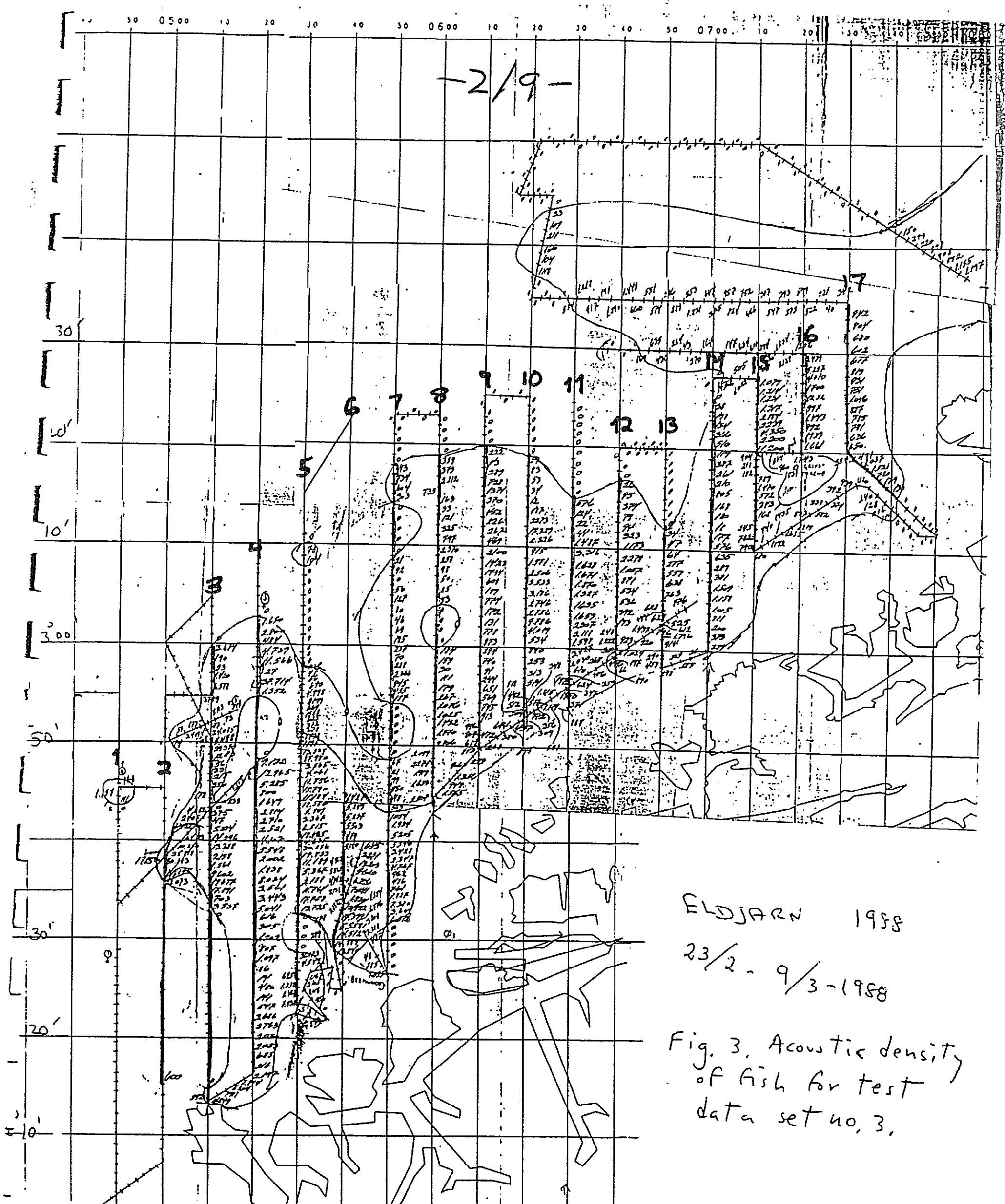
Only parallel, equally-spaced transects were considered for analysis. Inter-transect cross pieces were not not included in the analysis. Zero-valued data at the beginning and end of transects were also excluded. In data set 3, I exercised some poetic license. I chose to ignore the small aggregations at the top ends of transects 1 and 5. I also chose to include data at the bottom ends of transects 6 and 17 even though these data do not strictly adhere to the condition of equal spacing between parallel lines. [See attached figures] The area with non-zero fish density was calculated by multiplying the average transect length by the mean distance between transects (approximately 4.5 nmi) by the number of transects. Mean SA is an average of transect SA's weighted by transect length. Variance SA is estimated from the variation between transect SA's (Williamson 1982). This calculation is an application of the ratio method (Cochran 1977) where the variates are transect SA and transect length. This approach was proposed by Dr. G. Jolly during the 1987 Acoustics Symposium in Seattle (Jolly and Hampton 1987). One important difference is that Jolly stresses the necessity of randomly placed transects. I do not believe this condition is necessary (or even desireable) in many acoustic survey situations.

References

Cochran, W.G. 1977. Sampling techniques. 428 p.

Jolly, G.M. and I. Hampton. 1987. Some problems in the statistical design and analysis of acoustic surveys to assess fish biomass. Paper presented at the 1987 Fisheries Acoustics Symposium in Seattle Wa. USA.

Williamson, N.J. 1982. Cluster sampling estimation of the variance of abundance estimates derived from quantitative echo sounder surveys. Can. J. Fish. Aquat. Sci. 39(1): 229-231.



ELDJARN 1958

23/2 - 9/3-1958

Fig. 3. Acoustic density of fish for test data set no. 3.

Data Set 3

SA (m**2/nmi**2)

TR 2	TR 3	TR 4	TR 5	TR 6	TR 7	TR 8	TR 9
600	6578	216	47	13840	6512	359	612
0	1420	685	56	14889	3609	373	804
0	53	2023	690	14512	7380	2812	996
0	190	2070	4998	5558	1827	735	913
0	3674	2763	8199	14399	761	163	745
0		3686	669	12922	486	33	569
0	5458	547	1587	7182	962	121	651
0	29370	141	625	7131	4767	325	244
0	12113	480	372	4142	2347	797	91
0	56	14	678	2925	3428	2370	740
0	33	86	1738	4969	5540	258	294
0	215	1097	1195	5140	5205	98	873
0	256	807	13534	479	1934	50	778
0	172	1202	87980	5503	1054	25	131
0	233	305	3565	5237	563	33	172
0	375	616	4061	5249	418	0	774
0	47	5041	11756	1427	550	0	199
0	5224	3443	15790		99	0	608
0	14296	3561	11184		21	0	1741
38073	13388	8034	18580		67	184	1423
49145	2158	1838	5809		0	159	2100
50183	1561	2002	3367		0	30	469
6000	9602	5548	6515		0	111	267
	19677	4107	18465		0	179	226
	17891	2521	4635		0	267	452
	703	2710	30556		0	1076	370
	3537	2814	19433		187	1025	1371
	0	1649	11189		455	1432	728
	0	800	5367		945	1770	289
	0	5285	2181		2666	2806	43
	0	12965	11764		288	2099	222
	0	9132	17758		70	2271	
	0	0	13735		237	1919	
	0	0	0		185	1633	
	0	0	0		69	2807	
	0	0	0		46		
	0	0	0		80		
	0	0	36173		127		
	0	1352	4549		56		
	0	37714	695		0		
	0	27	1838		92		
	0	11566	8464		21		
	0	4739	7938		0		
	0	484			0		
	0	2800			0		
	292	1650			0		
					0		
					303		
					604		
					734		
					413		

TR 10	TR 11	TR 12	TR 13	TR 14	TR 15	TR 16	TR 17
59	371	66	34	299	1079	1661	130
83	347	197	47	373	1214	1939	179
53	359	890	64	200	1231	992	720
34	2030	1222	277	811	1317	1443	1521
12	3427	248	558	1005	2884	997	1057
487	1598	193	638	1158	2239	1252	491
2273	2111	392	763	1509	4350	1700	354
17329	2307	536	628	381	2200	4070	650
2336	1659	534	444	289	1200	5257	636
915	1635	871	612	635	904	3439	781
1971	1327	1057	796	576	211		785
2506	1870	2278	914	177	182		887
3533	1671	1173	290	18	789		1046
3186	1628	323	458	180	1470		751
2746	3316	94		168	572		931
2756	1487	78		805	343		819
9786	41	374		210	168		677
4019	22	95		261	245		602
534	124	26		387	782		680
890	576			119	790		804
253				510			842
313				366			
514				154			
482				198			
512				38			
588							
309							

Data Set 3

	TR 2	TR 3	TR 4	TR 5	TR 6	TR 7	TR 8	TR 9	TR 10	TR 11
Di	144001	148572	148520	397735	125504	55108	28320	19895	58479	27906
Di**2	2.1E+10	2.2E+10	2.2E+10	1.6E+11	1.6E+10	3.0E+09	8.0E+08	4.0E+08	3.4E+09	7.8E+08
ni	23	45	46	43	17	51	35	31	27	20
ni**2	529	2025	2116	1849	289	2601	1225	961	729	400
Di*ni	3312023	6685740	6831920	17102605	2133568	2810508	991200	616745	1578933	558120
length	115	225	230	215	85	255	175	155	135	100

TR 12	TR 13	TR 14	TR 15	TR 16	TR 17	Sums
10647	6523	10827	24170	22750	15343	1244300
1.1E+08	42549529	1.2E+08	5.8E+08	5.2E+08	2.4E+08	2.5E+11
19	14	25	20	10	21	447
361	196	625	400	100	441	14847
202293	91322	270675	483400	227500	322203	44218755
95	70	125	100	50	105	140
						avg

mean SU 2744 (m**2/nmi**2)

var SU 628450

cv SU 0.29

area 10058 (nmi**2)

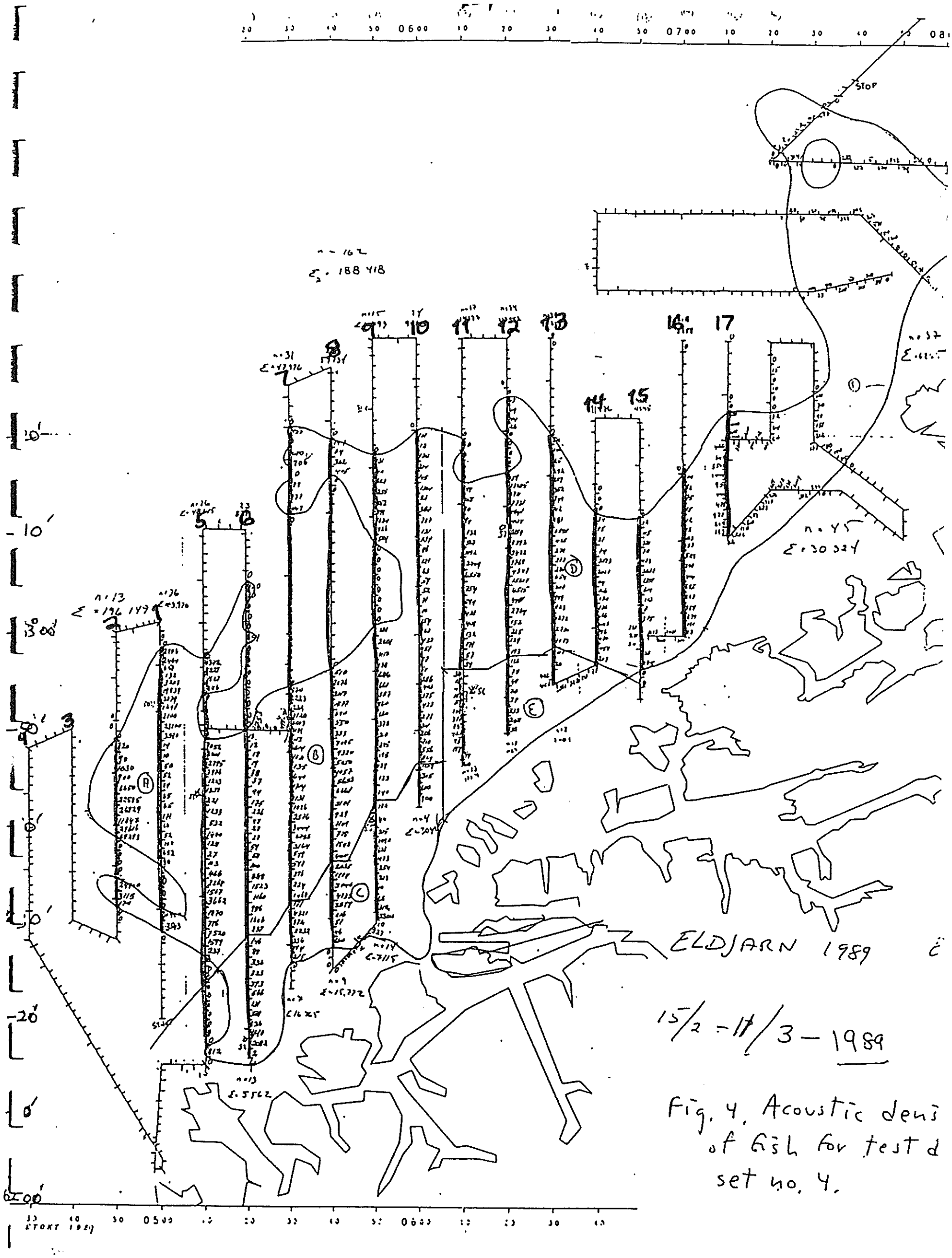


Fig. 4. Acoustic dens
of fish for test d
set no. 4.

DATA SET 4 SA (m**2/nmi**2)

TR 1	TR 2	TR 5	TR 6	TR 7	TR 8	TR 9
3543	120	812	120	215	144	3300
36400	90	0	10	194	64	312
0	1030	0	0	336	362	68
0	900	0	0	2832	405	10
0	6650	0	0	8716	0	213
0	22585	0	859	4321	0	254
80	26929	0	0	151	0	473
652	11847	0	0	1468	0	621
100	29616	0	0	224	0	1090
52	68393	0	0	396	0	215
66	0	0	0	597	0	40
171	0	237	0	599	0	182
65	0	1599	0	3164	0	140
65	0	7520	0	6786	0	133
84	0	796	13	3444	0	139
52	24700	1870	12	2576	0	195
50	3115	3662	18	1086	0	385
10	174	1567	19	1131	0	310
14		2268	18	1714	0	378
3590		466	37	640	0	660
23100		103	94	135	0	960
2100		27	175	1631	0	357
1241		128	226	1614	0	661
2879		1480	47	113	510	646
19939		532	28	1189	2876	978
3207		1233	28	2007	2017	1117
132		221	59	8120	1577	2001
858		1337	58	234	690	181
2484		1203	100	2223	550	0
2986		7976	888	580	333	0
		2795	1523	0	7085	0
		2001	1160	0	9330	0
		1052	986	0	5050	0
		0	1606	0	4057	0
		0	237	0	5653	0
		0	146	0	6661	0
		0	88	0	3101	584
		0	336	0	929	725
		0	323	0	1109	1834
		346	393	0	778	399
		1627	566	0	1842	552
		2227	131	0	4011	255
		4342	388	0	2035	667
			636	0	1984	411
			410	0	3046	131
			2082	0	4835	
			2	0	2899	
			61	0	216	
				1147	51	
				977	46	
				88	160	
				0		
				706		
				1680		
				0		
				455		

TR 10	TR 11	TR 12	TR 13	TR 14	TR 15	TR 16	TR 17
111	60	47	54	207	12	83	100
12	79	49	101	458	105	191	1000
170	197	44	25	436	338	279	800
214	95	26	292	346	70	359	193
145	421	0	257	3043	423	267	219
1364	85	0	352	666	2837	695	978
82	378	0	349	176	1351	404	478
387	169	0	42	170	614	388	440
787	59	0	981	166	103	529	952
121	187	159	3595	114	7	133	1444
134	49	33685	425	2467	375	163	5505
86	15	8711	278	3513	238	12	649
121	30	17721	373	74	271	12	420
63	59	4841	235	21	611	15	
59	57	14139	654	85	928	75	
52	171	2154	807	34	485	52	
71	132	8392	999			100	
10	409	3422	823				
154	428	1769	1272				
163	594	4347	2736				
433	754	15607	1073				
151	6550	4515	117				
87	2709	4481	20				
88	1192	2744	426				
286	507	152	426				
403	132	225					
375	94	129					
423	90	193					
276	155	166					
311	144	78					
206	0	49					
718	0	70					
556	0	39					
964	60	333					
1159		258					
305		136					
680							
900							

Data Set 4

	TR 1	TR 2	TR 5	TR 6	TR 7	TR 8	TR 9	TR 10	TR 11	TR 12
Di	103920	196149	49427	13883	63489	74406	21577	12627	16061	128681
Di**2	1.1E+10	3.8E+10	2.4E+09	1.9E+08	4.0E+09	5.5E+09	4.7E+08	1.6E+08	2.6E+08	1.7E+10
ni	30	18	43	48	56	51	45	38	34	36
ni**2	900	324	1849	2304	3136	2601	2025	1444	1156	1296
Di*ni	3117600	3530682	2125361	666384	3555384	3794706	970965	479826	546074	4632516
length	150	90	215	240	280	255	225	190	170	180

	TR 13	TR 14	TR 15	TR 16	TR 17	Sums
Di	16712	11976	8768	3757	13178	734611
Di**2	2.8E+08	1.4E+08	76877824	14115049	1.7E+08	8.0E+10
ni	25	16	16	17	13	486
ni**2	625	256	256	289	169	18630
Di*ni	417800	191616	140288	63869	171314	24404385
length	125	80	80	85	65	162 avg

mean SA	1512	(m**2/nmi**2)
var SA	219526	
cv SA	0.31	
area	10935	(nmi**2)
~~sampling index~~	~~0.22~~	

ANALYSE DE LA SERIE DE DONNEES NUMERO 4 ELDJARN A L'AIDE DU LOGICIEL GEO-EAS
(résultats provisionnels présentés au président sous forme manuscrite)

par

Alexandre J. DESBARATS

L'histogramme du log naturel de la densité acoustique est symétrique, avec une variance assez haute de 2.99. Aucun signe de populations distinctes n'est apparent. Le "probability plot" en ligne droite indique une distribution à peu près Gaussienne.

J'ai fait des histogrammes sur des indicatrices pour les seuils suivants : (ind00 : 0.0); (ind25 : 152.0); (ind50 : 588.0); (ind75 : 1992). Donc, plus de la moitié des données sont nuls.

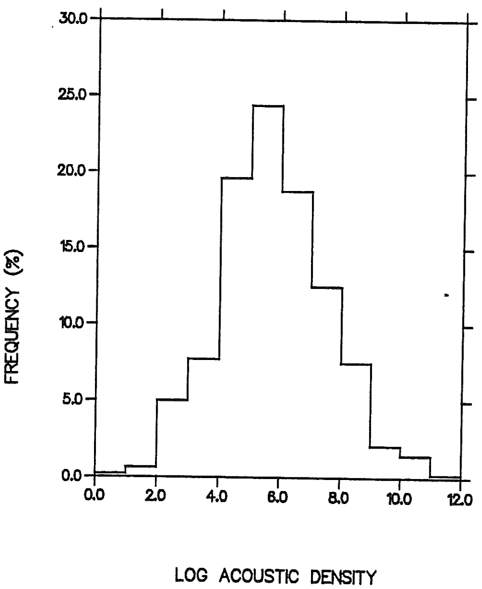
Il y a un effet proportionel très net entre la moyenne et l'écart type des valeurs pour des segments de traverses de 10 et 20 mesures. La transformation logarithmique est donc indiquée pour réduire l'hétéroscadicité des valeurs.

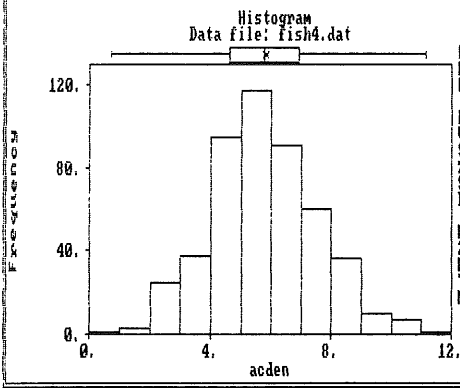
Les variogrammes sont tous dans la direction nord-sud, le long des traverses de navire. Ceci à cause de limitations de mémoire du logiciel Geo-EAS que j'ai utilisé. Aussi, la corrélation spatiale est-ouest est faible à la distance moyenne entre traverses. Les distances sur les variogrammes sont en degrés de latitude nord. Le variogramme du log de la densité acoustique est très beau. Il descend au longues portées correspondantes à la largeur moyenne du banc de poissons. Il y a donc ici un phénomène de non-stationarité à l'échelle étudiée.

Les variogrammes d'indicatrices sont beaux mais n'ont rien de particulier sauf le dernier, pour les hautes valeurs. Celui-ci montre une périodicité qui reflète la distance entre les quelques "lobes" de très hautes valeurs que l'on voit sur la carte de contours.

La carte de contours (d'ailleurs pas très belle) a été difficile à produire étant donné la disposition des points en lignes et la très grande variabilité spatiale des valeurs.

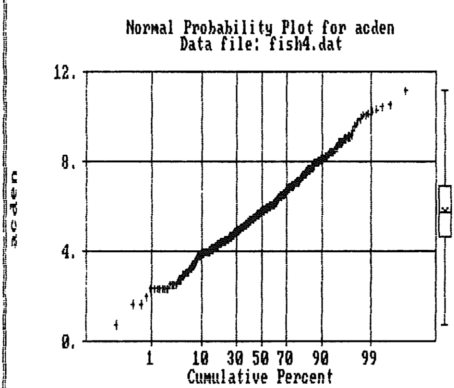
Quelques conclusions ... Le logiciel Geo-EAS n'est pas bien adapté pour les ensembles de données de plus de 500 valeurs. Les variogrammes démontrent une corrélation spatiale *indéniable* qui **réfute** les approches statistiques classiques. Cette corrélation spatiale justifierait l'utilisation du krigeage pour l'interpolation et l'estimation de stocks. L'approche des indicatrices est utile pour mettre en évidence la corrélation spatiale à divers seuils de valeurs. Les cartes de contours ne sont pas très convenables pour la représentation de la densité acoustique vu la répartition naturelle des poissons. Les cartes de pixels codées en couleurs (raster images) seraient plus appropriées. On aurait avantage à échantilloner mieux dans la direction est-ouest en faisant un patron de traverses en grillage plutôt qu'en lignes. Les méthodes de calculs de la variance d'estimation globale présentées dans David ou Journel et Huijbregts (par composition de variances d'extension élémentaires.) seraient facilement applicable ici parce que les traverses ne sont presque pas corrélées entre elles.


30.0
25.0
20.0
15.0
10.0
5.0
0.0
FREQUENCY (%)
0.0
2.0
4.0
6.0
8.0
10.0
12.0
LOG ACOUSTIC DENSITY



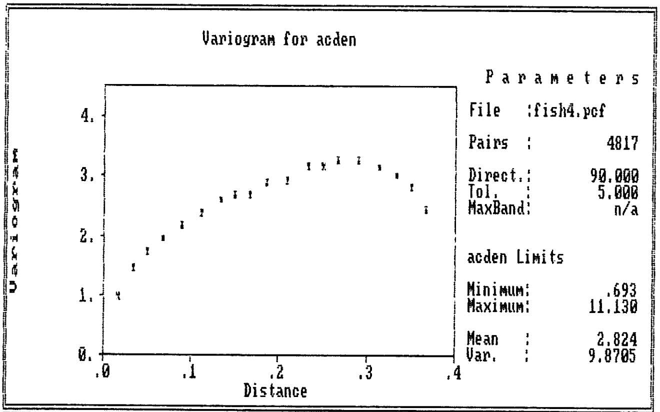
Statistics

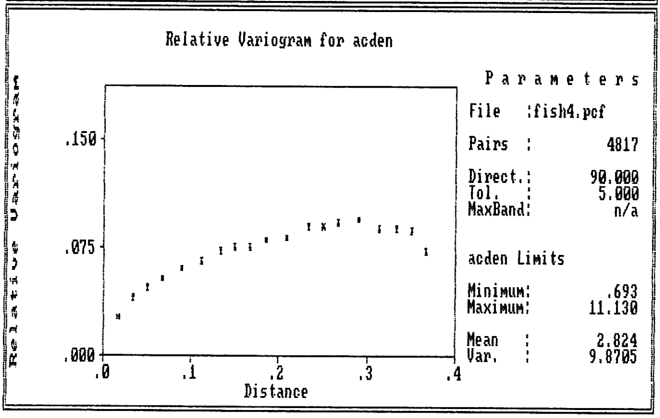
N Total :	986
N Miss :	506
N Used :	480
Mean :	5.801
Variance:	2.990
Std. Dev:	1.729
% C.V. :	29.808
Skewness:	.182
Kurtosis:	3.008
Minimum :	.693
25th % :	4.605
Median :	5.741
75th % :	6.907
Maximum :	11.130

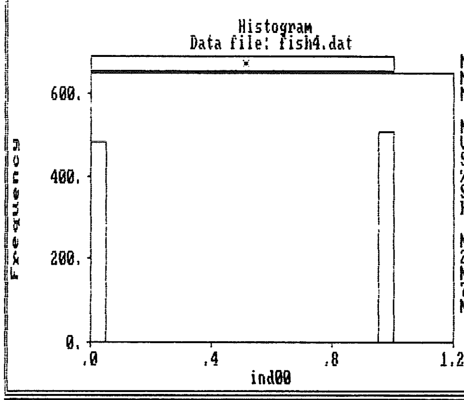


Statistics

N Total :	986
N Miss :	506
N Used :	480
Mean :	5.801
Variance:	2.990
Std. Dev:	1.729
% C.V. :	29.808
Skewness:	.182
Kurtosis:	3.008
Minimum :	.693
25th % :	4.605
Median :	5.741
75th % :	6.907
Maximum :	11.130

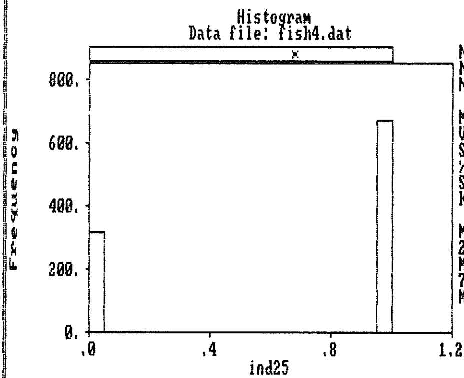

Variogram for acden
Variogram
4.
3.
2.
1.
0.
.0
.1
.2
.3
.4
Distance
Parameters
File :fish4.pcf
Pairs : 4817
Direct.: 90.000
Tol. : 5.000
MaxBand: n/a
acden Limits
Minimum: .693
Maximum: 11.130
Mean : 2.824
Var. : 9.8705


Relative Variogram for acden
Relative Variogram
.150
.075
.000
.0
.1
.2
.3
.4
Distance
Parameters
File :fish4.pcf
Pairs : 4817
Direct.: 90.000
Tol. : 5.000
MaxBand: n/a
acden Limits
Minimum: .693
Maximum: 11.130
Mean : 2.824
Var. : 9.8705



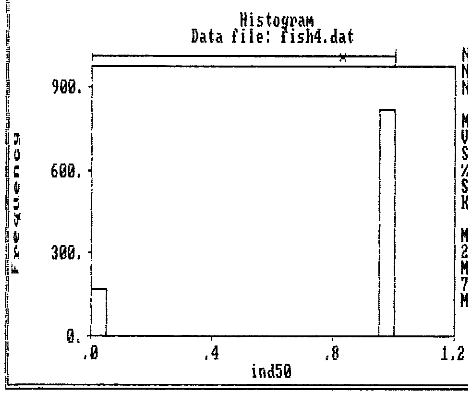
S t a t i s t i c s

N Total :	986
N Miss :	0
N Used :	986
Mean :	.513
Variance:	.250
Std. Dev:	.500
% C.V. :	97.446
Skewness:	-.053
Kurtosis:	1.003
Minimum :	.000
25th % :	.000
Median :	1.000
75th % :	1.000
Maximum :	1.000



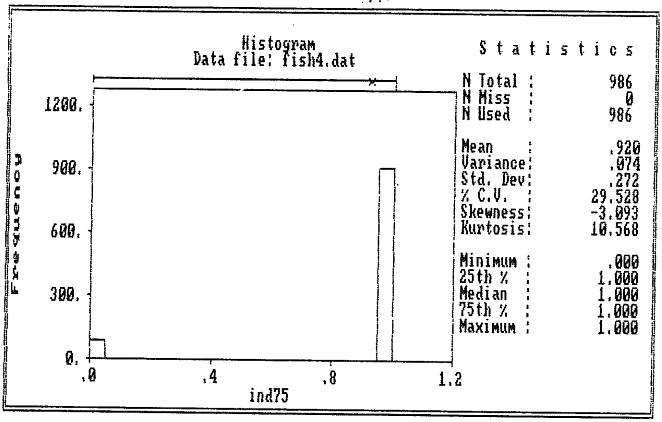
S t a t i s t i c s

N Total :	986
N Miss :	0
N Used :	986
Mean :	.678
Variance:	.218
Std. Dev:	.467
% C.V. :	68.871
Skewness:	-.764
Kurtosis:	1.584
Minimum :	.000
25th % :	.000
Median :	1.000
75th % :	1.000
Maximum :	1.000

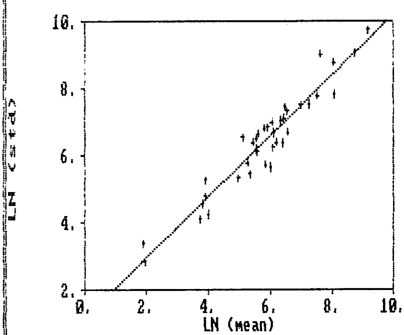


S t a t i s t i c s

N Total :	986
N Miss :	0
N Used :	986
Mean :	.828
Variance:	.143
Std. Dev:	.378
% C.V. :	45.667
Skewness:	-1.734
Kurtosis:	4.008
Minimum :	.000
25th % :	1.000
Median :	1.000
75th % :	1.000
Maximum :	1.000


Histogram
Data file: fish4.dat
Frequency
1200.
900.
600.
300.
0.
.0
.4
.8
1.2
ind75
Statistics
N Total : 986
N Miss : 0
N Used : 986
Mean : .920
Variance: .074
Std. Dev: .272
% C.V. : 29.528
Skewness: -3.093
Kurtosis: 10.568
Minimum : .000
25th % : 1.000
Median : 1.000
75th % : 1.000
Maximum : 1.000

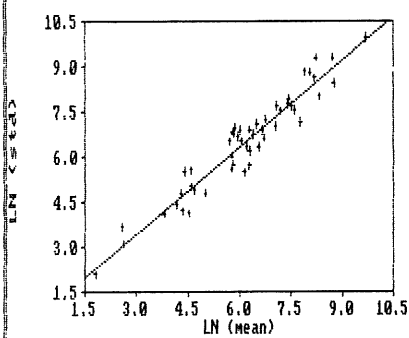
Scatter Plot
from data file blk20



Regression Results:

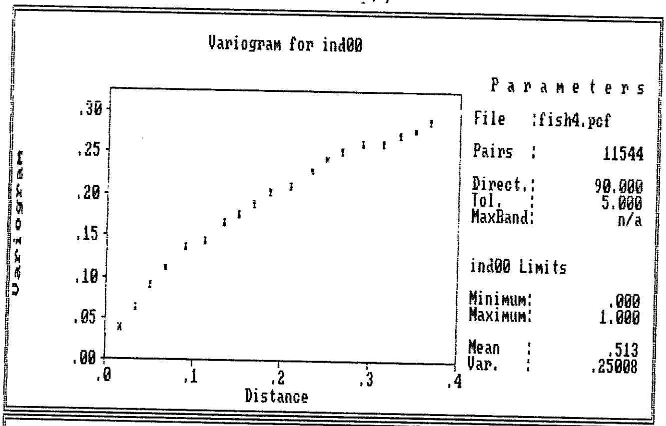
# Pairs	:	43
Slope	:	.902
Intercept	:	1.154
Correl. coeff.	:	.958

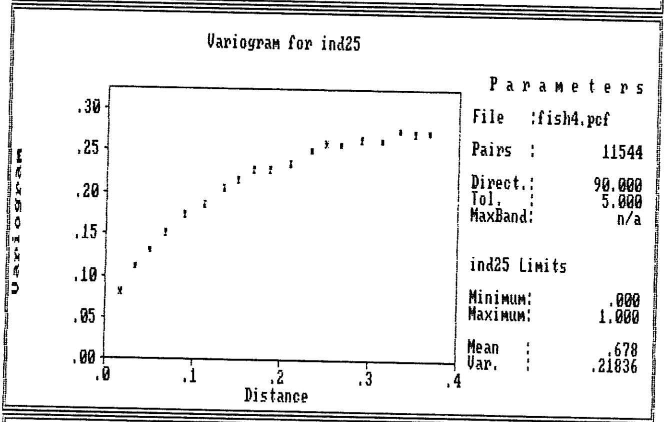
Scatter Plot
from data file blk10

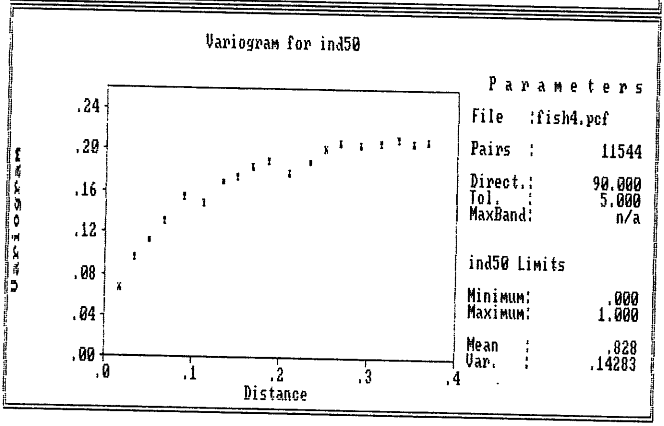


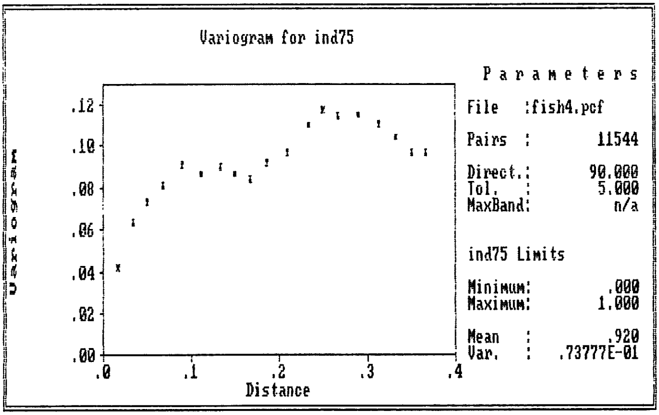
Regression Results:

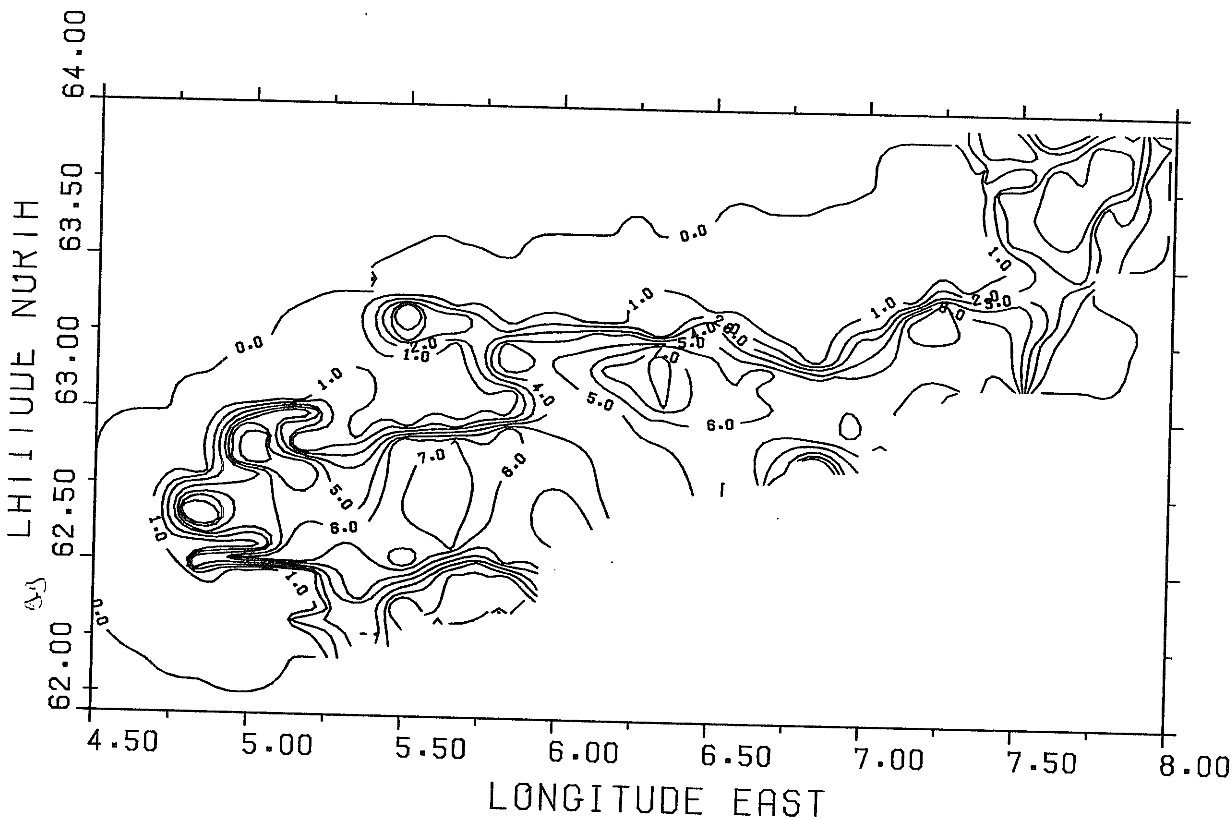
# Pairs	:	71
Slope	:	.966
Intercept	:	.523
Correl. coeff.	:	.958


Variogram for ind00
Variogram
.30
.25
.20
.15
.10
.05
.00
.0
.1
.2
.3
.4
Distance
Parameters
File :fish4.pcf
Pairs : 11544
Direct.: 90.000
Tol. : 5.000
MaxBand: n/a
ind00 Limits
Minimum: .000
Maximum: 1.000
Mean : .513
Var. : .25008


Variogram for ind25
Variogram
.30
.25
.20
.15
.10
.05
.00
.0
.1
.2
.3
.4
Distance
Parameters
File :fish4.pcf
Pairs : 11544
Direct.: 90.000
Tol. : 5.000
MaxBand: n/a
ind25 Limits
Minimum: .000
Maximum: 1.000
Mean : .678
Var. : .21836


Variogram for ind50
Variogram
.24
.20
.16
.12
.08
.04
.00
.0
.1
.2
.3
.4
Distance
Parameters
File :fish4.pcf
Pairs : 11544
Direct.: 90.000
Tol. : 5.000
MaxBand: n/a
ind50 Limits
Minimum: .000
Maximum: 1.000
Mean : .828
Var. : .14283

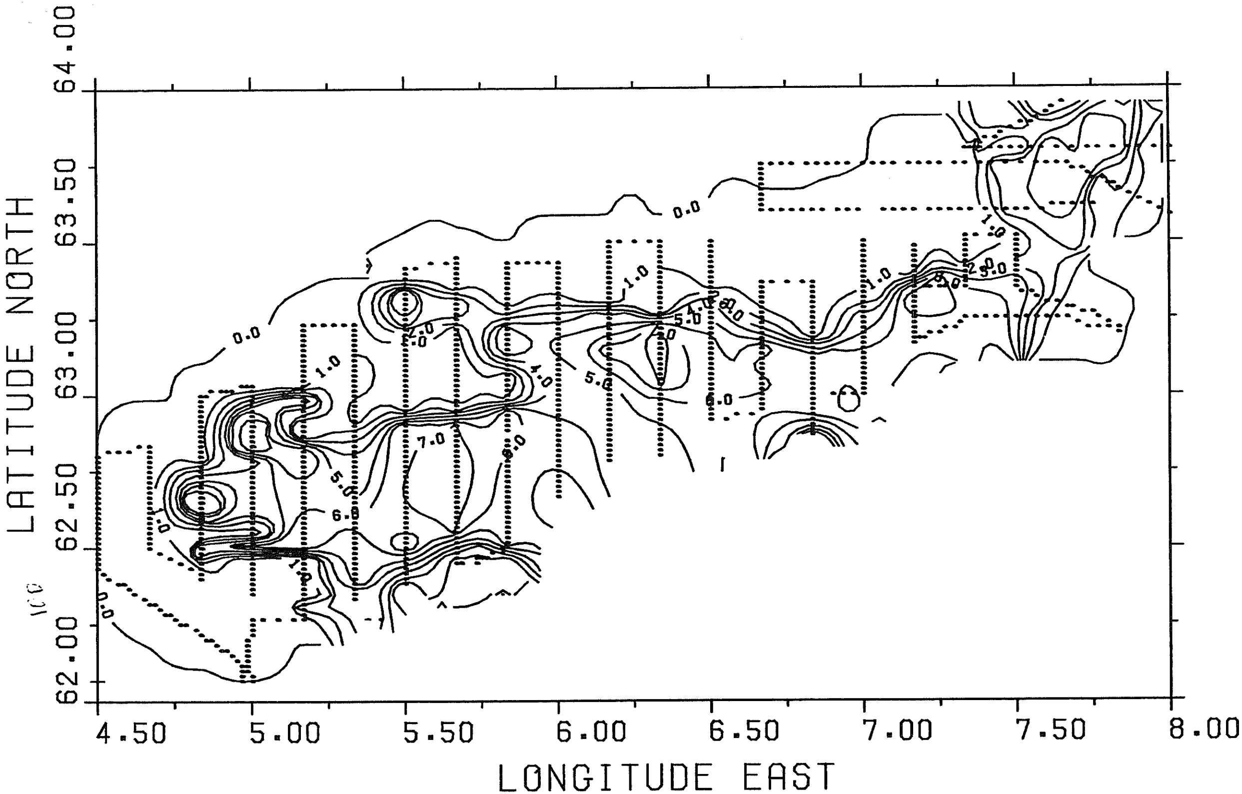

Variogram for ind75
Parameters
File :fish4.pcf
Pairs : 11544
Direct.: 90.000
Tol. : 5.000
MaxBand: n/a
ind75 Limits
Minimum: .000
Maximum: 1.000
Mean : .920
Var. : .73777E-01
Variogram
.12
.10
.08
.06
.04
.02
.00
.0
.1
.2
.3
.4
Distance



ELDJARN - 1989

LOG- ACOUSTIC DENSITY

DATA SET NO 4



ELDJARN - 1989

LOG- ACOUSTIC DENSITY
DATA SET NO 4

MEDIAN POLISH KRIGING

by

W.G. WARREN

Not surprisingly, the data of all five test data sets appear to exhibit non-stationarity. Cressie (1986) observes that "The spatial prediction method known as kriging exploits second-order spatial correlation structure to obtain minimum variance predictions of certain average values of the random function. But to do so, it must be assumed that either the mean function (the drift) is known up to a constant or the second-order structure (the variogram) is known exactly. Knowledge of the drift allows the (stationary) variogram to be estimated and leads to ordinary kriging. Knowledge of the variogram allows the drift to be estimated and leads to universal kriging. More usually, neither is known". In the paper he "shows how median polish of gridded spatial data providers a resistant and relatively bias-free way of kriging in the presence of drift, yet yields results as good as the mathematically optimal (but operationally difficult) univeral kriging". It was decided, therefore, to explore the potential for using median-polish kriging with the test data.

To develop an understanding of the procedure, it was first applied to the 29-point data set contained in the report of the 1989 Workshop on Spatial Statistical Techniques. Although these points are not perfectly aligned in space they are sufficiently so for the Cressie and Read (1986) solution to "the problem of irregularly located data by assigning each datum to the nearest node of an overlaid grid" to be applied without ambiguity. With these data the median polish apparently accounted for all structure, leaving nothing but white noise in the residuals. [One point was accidentally misplaced but the effect of this on the conclusion is believed to be inconsequential].

Because of time constraints the method was then applied to data set 3 only. This set was chosen over sets 1 and 2 because the locations being recorded in degrees and minutes, rather than degrees, minutes and seconds, simplified the placing of points on a grid. Data set 4 might have equally well been chosen. Data set 5 seemed too large to handle in the time frame available.

Of the 881 points more than 600 were located on parallel transects in the north-south direction at a constant interval of 11 miniutes of longitude. It was therefore decided, for the purpose of illustration, to ignore the points between these transects (appox. 25% of the data). Points with zero reading were also omitted. This seems clearly justified for "external" zeros. "Internal" zeros tended to occur in clusters and may also represent areas devoid of fish. It was planned to do analyses with the internal zeros included as well as omitted but, again, insufficient time was available. The remaining points were then placed on an 85 x 19 rectangular grid with the rows being 1 minute of latitute apart and the columns 11 minutes of longitude.

The non zero data exhibited positive skewness. Cressie (1989) observes that "When $\{Z_t\}$ is a Gaussian process, the best predictor is a linear predictor". He then assumes "that an appropriate transformation has been made that converts the problem into Gaussian data (with possible additive outliers, here modeled as heavy-tailed contamination in the stationary error distribution)". The median polish was, therefore, applied to the square roots of the observations. (With more time the appropriate Box-Cox transformation could have been identified - interestingly, it was subsequently learned that E.J. Simmonds found the square root to be the appropriate transformation for set 3).

The median polish apparently accounted for any structure in the east-west direction. In that direction the variogram had the appearence of white noise; the transects could, thus, be treated as independent. (One should, however, not overlook the possibility of structure at less than the 11 minutes of latitude between transects. On the other hand, in the north-south direction, while the median polish clearly accounted for a certain amout of drift, some auto-correlation structure remained. There was a well-defined variogram which rose from a relatively small nugget effect to a sill at a range of about 25 minutes of latitute.

On the basis of this, one is tempted to conclude that there is potential for median-polish kriging of acoustic survey data.

Reference.

Cressie, N. 1989. Kriging nonstationary data. Jour. Amer. Statist. Assoc. 81:625-634.

SYNTHESIZING ACOUSTIC SURVEY DATA.

by

K.G. FOOTE AND Z. KIZNER

LARGE-SCALE DATA SIMULATION.

Z. Kizner described a procedure for simulating the large-scale features of a fish aggregation density field. This might be modelled as a superposition of a number of patches (aggregations) of different sizes; these patches can overlap and create large aggregations. There is a set of histograms of density values, and every histogram corresponds to a certain patch size.

Initially each patch is represented by a circularly symmetric domain over which the density generated according to a given histogram, has a quasi gaussian smooth space distribution with superimposed noise. So, the function, which is visualized by a surface over the domain, demonstrates smaller-scale features, viz. irregularities into the patch.

The function corrugated and convoluted by the introduced noise may be further deformed, as by a diffusion process performed on the field. Statistical and geometric properties or patterns of observed fish aggregations may also be modelled. The derived distribution of fish density may be surveyed by extracting values of the simulated data along arbitrary tracks crossing the domain. Typical grid sizes are 50 x 50 or 100 x 100, but 200 x 200 is entirely feasible.

SMALL-SCALE DATA SIMULATION.

Synthesis of echogram data on the smallest ping-based scale is described in ICES CM 1989/ B:6. The model is composed of a number of stochastic processes, which may also contribute to the realism of the simulation.