INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA

CONSEIL INTERNATIONAL POUR L'EXPLORATION DE LA MER

C.M. 1991/D:40
Statistics Committee

Report of the Workshop on
the Applicability of Spatial Statistical Techniques to Acoustic Survey Data

Reykjavík, 5-9 September, 1991

General Secretary
ICES
Palægade 2-4
DK-1261 Copenhagen K
Denmark

CONTENTS

1. Introduction

1.1 Participants

Gérard Y. Conan	Canada
Marsha Daniel	Iceland
Eduardo Ferrandis	Spain
Kenneth G. Foote	Norway
Knut Korsbrekke	Norway
Rainer Oeberst	Germany
Lars-Erik Palmén	Sweden
Pierre Petitgas	France
Jacques Rivoirard	France
John Simmonds	U.K. (Scotland)
Gunnar Stefánsson	Iceland
Pat Sullivan	U.S.A.
Gordon Swartzman	U.S.A.
W.G. Warren	Canada

1.2 Terms of reference

The terms of reference are given in C. res. 1990/2:11:

A Workshop on the Applicability of Spatial Statistical Techniques to Acoustic Survey Data, with Dr. G. Stefánsson (Iceland) as Chairman and Dr. G. Y. Conan (Canada) as Vice-Chairman, will be held in Reykjavík from 5-9 September 1991 at national expense to:

a. present data analyses prepared in advance;

b. present comparisons of methods prepared in advance;

c. discuss analyses, methods, and comparisons of methods;

d. prepare plans for an ICES Cooperative Research Report.

1.3 Background

The acoustic data under analysis consist mainly of the mean area backscattering coefficient. The meaning of this is explained here.

The fundamental quantity that is measured in echo integration surveys (MacLennan 1990) is the mean volume backscattering coefficient s_v. This is the cumulative backscattering cross section in the sampled volume V_s (Stanton *et al.* 1987). For a single ping, or sounding,

$$s_v = \frac{1}{V_s} \sum_{j=1}^{n} \frac{\sigma_j}{4\pi}, \tag{1.1}$$

where σ_j, in this section only, is used to denote the backscattering cross section of the j-th scatterer of n in V_s. In the limit of a large number of scatterers or pings,

$$s_v = \rho \frac{\bar{\sigma}}{4\pi}, \tag{1.2}$$

where ρ is the number density of scatterers with respect to volume, and $\bar{\sigma}$ is the mean backscattering cross section of a scatterer.

While the dependence of s_v on echo range or depth z can be quite useful for some applications, the data quantity is generally voluminous and unwieldy for ordinary surveying work. A much more useful quantity is the area or column backscattering coefficient s_a (Clay and Medwin 1977). This is the integral of $s_v(z)$:

$$s_a = \int_{z_1}^{z_2} s_v(z) dz, \tag{1.3}$$

where z_1 and z_2 are the limits of integration. Strictly speaking, these define the inner and outer radii of a spherical shell centered at the transducer, for short pulses and in the transducer farfield. For the highly directional transducers that are almost universally employed in echo integration work, most echoes come from the central lobe of the beam patterns, hence echo range is tantamount to depth.

The quantity s_a is dimensionless, but is typically very small, say of the order of 10^{-7} to 10^{-1}. It is more conveniently expressed with respect to 1 square nautical mile, or 1 NM ^{2} (Knudsen 1990). This quantity, denoted s_A, is derived from s_a thus :

$$s_A = 4\pi 1852^2 s_a. \tag{1.4}$$

This is the basic quantity that is analyzed in echo integration surveys. Its units are square meters of backscattering cross sectional area per square nautical mile.

The importance of s_A may be emphasized by substituting Eq. (1.2) in Eq. (1.3), and the result in Eq. (1.4) :

$$s_A = \rho_A \bar{\sigma} \tag{1.5}$$

where ρ_A is the number density of fish with respect to an area of 1 NM^2. If a measurement of s_A is representative of a particular species and size or age class of scatterer for which $\bar{\sigma}$ is known, then computation of ρ_A is immediate.

In the test data, s_A is assumed to be monospecific with respect to the fish scatterer. The various manipulations are performed mainly on s_A, without division by $\bar{\sigma}$, this last step being extraneous to the aim of the study. In one case, that of Data set 6, the division has been performed and the number further converted to mass density with respect to area.

1.4 Nomenclature

The following convention is used throughout this work. Measurements made along transects are of **density**. This may be acoustic density, s_v or s_A; number density with respect to volume, ρ; number density with respect to area, ρ_A; or mass density with respect to area. The result of integrating a density field defined over an area specifies the **abundance** of the animal.

1.5 Working Papers

Working papers were available on some of the topics. These are listed in Appendix A.

1.6 Acknowledgements

The work described in this report would not have been possible without the aid of several individuals not included in the participant list. In particular, A. Aglen, I. Røttingen, K. Sunnanå, P. Reynisson, Á. Guðmundsdóttir and N. J. Williamson are thanked for their contributions of acoustic data. Z. Kizner is thanked for the contribution of simulated data. Finally, E. Wade, D. Stolyarenko and N. J. Williamson prepared analyses which were available for the meeting.

2. Basic data analyses and survey design

Each survey needs to be explored initially with data display tools before analysis begins. A key to proper spatial analysis of survey data lies in the graphical display and preliminary exploratory analysis of abundance and, where relevant, ancillary environmental data. Through the comparison of several surveys, general patterns may emerge that would suggest fruitful general approaches. Even after initial processing, an acoustic survey data set still provides much data, including the depth distribution of fish abundance and the depth of the water column, sometimes at the individual fish species level. The data are generally agglomerated further, as seen in the data sets analyzed here, to produce a total abundance estimate that applies throughout the entire water column. In some cases, depth has been provided (although presumably it is available or easy to obtain in all cases).

It is useful to present the survey tracks on a map of the area. A variety of means have been used to show the spatial distribution of abundance along the tracks including scaled histograms (Figure 2.1), scaled rectangles (Figure 2.2), and scaled circles (Figure 2.3). These figures present the data in an explicit spatial context, with abundance proportional to the size or length. Contour and gray-scale image plots were also used to display abundance data (Figure 2.4). While these permit more complete spatial coverage than the scaled transect plots, they can introduce artifacts into the data resulting from the interpolation used to cover unsurveyed areas between the survey tracks, such as apparent smears or gradients of abundance. Careful attention should be paid, in interpreting such plots, to where the survey tracks actually are. Confidence in interpolated predictions is often low away from surveyed areas, especially if the data are extrapolated out of the study area. This was not done here but it is done and it is often misleading. An example is in Figure 2.5, which shows a bilinear interpolation of the Iceher1 survey as a 3-D plot. Here, regions of low acoustic density appear as four regularly spaced areas at the bottom of the surveyed area (front of the plot). These areas are intersected by the survey transects. Between them are areas of purportedly higher density, which are higher solely because they are not on the survey tracks

and so this interpolator produces spatial biases. The same observation can be made for the high abundance areas near the top of the surveyed area (back of the plot). Color plots can be made analogously to the transect, grayscale, contour and 3-D plots.

The presence of ancillary information, such as depth, can be informative in the analysis of acoustic data. The gray scale and contour plots for abundance (shown in Figure 2.4) have been compared with depth gray-scales for the same area, which can help to suggest possible relationships between these variables. For the Bering Sea there appears to be a strong relationship between depth and abundance. This suggests further analysis using a spatial trend detection model such as GAM or GLM to correlate ancillary information with abundance. This is discussed further in section 7.

A number of classical summary statistics may be used in an exploratory sense as well. The variogram, which is used in geostatistical estimation procedures, is a good example. In a restricted sense it represents the correlation between sets of observations a distance h apart. Patterns in variograms fit to the data can indicate patterns in the data. Each of the the variograms obtained from untransformed test data sets 1-3 (Figures 2.6, 2.7 and 2.8) represents a different covariance pattern. The first indicates covariance that continues to decrease with distance, possibly indicating some large-scale pattern of variation in the observations. The second indicates a pattern more like a global nugget effect indicating no pattern or covariance on any scale. The third shows a covariance that initially decreases with distance, only to increase later, indicating both small-scale as well as large-scale interactions. The variogram can also provide diagnostic information about the existence of correlation in a population at the spatial scale of the collected data.

Fish depth profiles (showing the fish depth distribution) can give important clues to species and provide additional information to locate trends in fish distributions. As such, these data should not be summarized by a single measure, such as column scatter strength or overall fish abundance, until preliminary display is made. The variogram and other simple statistics like scatterplots of abundance versus ancillary variables can help in preliminary data examination and suggest further direction. An important aspect of such preliminary analysis is to point out the potential relative importance of autocorrelation and trend (drift) to fish spatial distribution and thereby suggest whether the analysis should ignore or include these factors. Graphical display of model predictions and residuals can also be important after analysis to indicate whether model assumptions were met and the need for possible further analysis.

2.1 Multivariate data

In addition to density integrated through the water column, survey data usually provide information on the depth distribution of abundance as well, which may be useful in providing 1) a good indication of where an abundance pattern has changed (i.e. change in both the magnitude and depth distribution of abundance) and 2) relationships of species interactions and possibly help in species identification. An example of display of such data is given in Figure 2.9.

Developing tools for multivariate spatial data analysis (as, for example, the Barent Sea data shown in Figure 2.9) remains a challenge. Few methods exist even for the display of such data, let alone tests of statistical significance or measures of trend. For example, how can the depth profile of fish abundance be related to covariates? The challenge is compounded if the data are not all collected simultaneously or in the same region, as, for example, using sea surface temperature or ocean color data collected from satellite images as covariates for fish abundance in a nonparametric regression. That such variables are important to fish distribution is attested to by the use of spatial cross-correlation between fish catch at satellite-collected sea surface temperature data to successfully predict areas of high catch. For at least two species (Shinomiya and Tameishi 1988) these 'hot spots' are on the edges of eddies of cold or warm water breaking off from major ocean currents. In this area, exploratory data analysis plays a central role in helping to choose variables of importance, to reduce the dimension of the problem to its bare essentials, to suggest analysis tools, to characterize an area, and to provide clues to possible univariate measures that can serve as surrogates for multivariate aggregates.

2.2 Model evaluation

Just as graphical display is important before analysis, so is it important after analysis. The model fit should always be graphically compared with the original data. For models involving transformations of the original data (e.g. GAM and GLM) this comparison should be made both with the transformed and the natural data. Graphical displays of the residuals through, for example, two-sided rectangular histograms (e.g. right for a positive and left for a negative residual) along the transects just like Figure 2.3, but for residuals, can indicate lack of fit of the model or the possibility of having correlated residuals. A variogram on the residuals in a trend model can also indicate the need for further analysis if significant correlation is shown. Other statistical methods, such as cross-validation, can be used to evaluate the method applied.

2.3 Survey Design

This section is provided as a brief overview of the choice within the design of a survey and the track layout. It is mostly based on Simmonds *et al.*, 1991. Only the major elements in the choice of cruise track are considered. Other elements in the survey design such as biological sampling requirements and allocation or estimating overall sampling effort are ignored. The survey design consists of a series of choices of strategy. There is no one single optimum strategy for all objectives. The choices that are appropriate are determined firstly by the objectives of the survey, secondly by any knowledge of the stock distribution, and thirdly, the analytical method to be employed for data analysis. In all cases the use of appropriate *a priori* information will improve the survey design and the subsequent estimates. However, care must always be taken to ensure that any survey design is capable of producing adequate results if the fish distribution or its behavior differs from the expected. It is unlikely that a good survey design can be completely free of assumptions, and the best results will be obtained by understanding the fish stock and its distribution.

Objectives

There are a number of possible objectives, such as; an overall abundance estimate for a population or an area, an estimate of precision for that abundance, a map of the spatial distribution, or possibly the location of major exploitable concentrations. In addition, there may be subsidiary criteria that affect the choice of strategy, such as; the absence of bias in the estimate and minimum variance, minimization of mean square error, or that the estimates are obtained with the minimum number of assumptions. It is important to be clear about both the objectives and their relative importance.

Definition of survey area and Stratification of effort

Selecting the boundaries of the survey area is important. Removal of areas that contain no fish has considerable benefit. For most stock distributions there appears to be a link between variance and mean density. Predicting in advance areas of high and low density and allocating sampling effort accordingly can give considerable gains in precision. Depth, hydrography, and a knowledge of the distribution from previous occasions are all possible stratification criteria.

Adaptive / Predetermined Strategies

Predetermined strategies require fewer assumptions about the stock distribution. More information is required to design an adaptive survey than to use predetermined designs. Adaptive strategies are particularly applicable when the stock is highly contagious in its spatial distribution but unpredictable in location. A number of adaptive methods have been used; scouting or outline surveys followed by intensive local surveys, adaptive transect lengths, and increased survey effort in areas of high density. Each of these methods requires assumptions about the distribution of the stock. If these assumptions do not hold, the estimates will be biased. Adaptive strategies may preclude calculation of survey precision without making further important assumptions.

Transect Direction

Choice of direction is controlled by a number of factors.

a. Minimization of between-transect variance. This is relevant for areas with anisotropic distributions and requires transects to be placed in the direction with the greatest rates of change.

b. Direction of migration. To minimize errors caused by systematic horizontal movement of a population the survey should be conducted with transects alternately with and against the direction of migration. If this is in conflict with criteria a) then an 'interlaced' survey design should be considered.

c. Minimization of inter-transect time. In the absence of other information the transects should be across the short axis of an area.

d. Operational considerations such as weather may necessarily override these considerations, but may compromise the results.

Systematic / Random track designs

The choice of track design is strongly influenced by the objectives of the survey and the method chosen for data analysis. However, some basic guidelines can be given. If the overriding requirement is for an estimate of total abundance, in the absence of spatial periodicity, systematic sampling generally provides the best estimate. If the spatial correlation is ignored, then random strategies should be employed to allow for calculating the variance. But if spatial information is modeled, random sampling is not required for the variance calculation and systematic sampling is believed to be more efficient.

Parallel / Zigzag transects.

For random designs independence of transects is essential. For this reason, parallel transects are useful. For adaptive designs, both the transect length and spacing will be changed by the use of zigzag transects. This requires additional assumptions that are difficult to justify and should be avoided.

For systematic designs, the choice of transect design is not so clear.

For parallel transects, a proportion of survey time will be unusable if inter-transect data is excluded from the data analysis. In most cases where the boundaries of the survey area are determined by the stock distribution, including coastlines, this must be the case.

For zigzag transects, there is increased correlation between data from the vertices. The *raison d'étre* of systematic sampling is to ensure efficient coverage of the sampled area.

For wide areas with long transects, and thus low proportions of unusable time, parallel strategies are preferred. For narrow areas, considerable survey time will be wasted if parallel transects are used. In these situations, the increase in survey effort will improve the estimate despite the loss of independence at the vertices. However, because of the high correlation at the vertices, it is important that they are not located preferentially and, where possible, they must be located outside the boundaries of the population.

3. Test data sets

3.1 Data sets 1-5 Norwegian fish stocks

These data sets are derived from acoustic surveys of Norwegian fish stocks. The presented data are believed to be monospecific within each set. The gross characteristics of the data are summarized in Table 3.1. Further details, including statistical features, are given below. Maps showing transects and acoustic density values are presented in Figs. 3.1-5.

Table 3.1. Gross characteristics of test data sets 1-5.

Data set	Fish type	Region	Interval (NM) Integration	Interval (NM) Averaging	No. data
1	Pelagic	Coast	5	5	664
2	Pelagic	Fjord	1	1	96
3	Pelagic	Coast	1	5	881
4	Pelagic	Coast	1	5	986
5	Benthic	Open Sea	3-5	5	1712

Data set 1 This describes an unbounded fish aggregation with concentrations on the survey boundary. The observations are averaged over 5-NM intervals, the transects are spaced at intervals of about 15 NM. Data on longitude are relative. The data were contributed by A. Aglen.

Data set 2 The distribution is bounded by fjord walls, but is extremely patchy. Cross-fjord samples are not available nor are more fine-grained data on one small but exceedingly dense concentration. The source of data is A. Aglen.

Data set 3 The aggregation is mostly bounded by the survey. Data are provided at 1-NM intervals. The parallel transects are locally concentrated and strongly contrasted with low values including zeroes. The source of the data is I. Røttingen.

Data set 4 This is the result of repeating the survey represented by Data set 3 after one year. The sampling interval and distance between parallel transects retain their previous survey values of 1 and 5 NM, respectively. The statistical characteristics are less extreme than in Data set 3, but concentrations exist along boundaries. I. Røttingen also contributed this data set.

Data set 5 Two ships collected the data on this survey, and the survey grids overlap in space but not in time. The data are distinguished by survey grid. Considerable differences are observed with respect to time and space. The source of data is K. Sunnanå.

3.2 Data set 6 Bering Sea walleye pollock

The data were derived from the 1988 summer survey of the eastern Bering Sea shelf. The survey region showing the 27 parallel transects is shown in Fig. 3.6. Each data record consists of position, time, bottom depth, distance sailed, and surface density. Bottom depths greater than 400 m are recorded as 400 m. The surface density expresses the fish density in terms of biomass per unit area. The units are kilograms of fish mass per square meter. The source of the data is N. J. Williamson.

3.3 Data sets 7-9 Icelandic herring

A major part of the stock of Icelandic summer-spawning herring was surveyed repeatedly in the region indicated in Fig. 3.7 under similar conditions on the night of 25-26 November 1988. The three surveys reported here had the following starting and ending times: 19:00-22:15, 22:40-02:05, and 03:15-07:00 local time. The horizontal resolution of the data is 0.1 NM. Bottom depth is given as an ancillary datum associated with each acoustic datum. P. Reynisson contributed the data which Á. Guðmundsdóttir prepared for distribution.

3.4 Data sets 10-15 Simulated data

Two fish aggregation density fields were simulated over a square 300 by 300 matrix. The first field was simulated by means of an algorithm devised by Z. Kizner and exercised on the basis of actual survey data for Myctophidae that were collected during a cruise of the Soviet vessel VOZROZHDENIE in the waters north of South Georgia, 27 September - 16 October 1988. The second field was derived from the first by a transformation. A smoothed version of the first field is shown in Figs. 3.8-9.

On the basis of each simulated data field, three variants were derived: (1) without noise, (2) with normally distributed additive noise, with standard deviation of 20, and (3) with multiplicative, lognormally distributed noise, with standard deviation of 0.1.

Survey data are simulated by superimposing a grid of ten equally spaced parallel transects on the two density fields in each of the three variants. The grid is indicated in Fig. 3.10. Data from the first simulated density field in its three variants are averaged over a series of three successive values. Each of these simulated surveys consists of 994 data. Data from the second simulated density field are averaged over series of five successive values. Each of the resulting simulated surveys consists of 596 data.

4. Classical analyses

In the classical approach to survey data analysis, the data should be collected on a largely uniform grid of either systematic or stratified random design. The grid density need not be uniform over the whole survey area but if different levels of survey effort are used then these areas must be treated separately. The survey grid is constrained so that at least one transect passes through each element of area used in the data analysis. The data are analyzed to give some geographical or spatial distribution and an overall estimate of abundance. The area is broken up into sub areas or strata. These may be large parts of the area or small 'rectangular' strata based on lat/long positions. These strata are not selected on the basis of the abundance values but rather on the spatial variability and should be determined

prior to the survey. Typically, the strata dimensions have turned out to be between two to four times the limit of sample correlation.

4.1 Method

The data from each stratum are analyzed separately to give estimates for each stratum. At the end of the analysis, the strata are combined to give a total abundance and associated variance. The data within each stratum are treated as independent and identically distributed. The strata are assumed to be independent. An arithmetic mean and variance may be calculated for each stratum. However, the amplitude distribution of data found in each stratum may not be normal, and a more efficient estimate of the stratum mean may be possible. The data are examined visually to check that the amplitude distribution is not multimodal. A Maximum Likelihood estimation procedure as described by Box and Cox (1964) is used to estimate a suitable power transform to the Gaussian distribution. This is combined with a delta function (Aitchison 1955) to remove the zero values. The Box-Cox transform is performed separately on all strata, but the results are combined to give significant results. If the results of this test give a maximum for the power transform between +0.5 and 0, a power transform of 1/2 1/3 1/4 1/6 or *ln* is selected. For each stratum, the mean and variance of the transformed data are calculated. As the distribution is Gaussian, confidence limits may be calculated in the transform domain. The inverse transform is performed and the effects of the delta function removed (see MacLennan and MacKenzie 1988). An unbiased mean and variance are determined for each stratum. The abundance of each stratum is calculated using the area of each stratum, taking into account the proportions of land and sea as appropriate. The variance of the abundance is the variance of the mean scaled by the area squared. Finally, the total abundance and its variance is calculated assuming independence of strata. To check the process, the means calculated by the arithmetic and transformed methods are compared.

4.2 Data Sets analyzed

The choice of rectangular strata sizes and the selected power transform for the data sets that have been analyzed, are shown in table 4.1.

Table 4.1. Strata sizes and power transforms used for data sets 1-5.

	Strata Sizes		
Data set	Latitude	Longitude	Power transform
1	0.5	1.0	1/6
2	1/12	1/6	*ln*
3	0.5	1.0	1/2
4	0.5	1.0	1/6
5	0.5	1.5	*ln*

The results of the data analysis are given in section 8.3. With the exception of data set 2, the survey designs and data distributions are suitable for this analysis technique. In all these cases, it was possible to select one power transform unambiguously and the differences between arithmetic and transformed means were negligible. However, it was not possible to select a unique transform for data set 2, since although the Box-Cox test indicated that the best transform was the logarithmic one, the confidence intervals included other transforms. It is also interesting to note that for data sets 1 and 4 the transform estimate exceeded the arithmetic mean, for data set 3 they were equal, and for data sets 2 and 4 the transform estimate was less than the arithmetic mean, indicating, at least from this small sample, that there is no evidence of bias in this technique.

4.3 General Applicability

This technique, when applied on a grid structure similar to those shown for data sets 1 and 3 to 5, provides some geographical information, total abundance and variance estimates along with confidence limits. It works best with a systematic sampling strategy and uniform sampling intensity. It is most applicable to large ocean areas (data sets 1,3-6 and 10-15) with little spatial correlation and non-stationarity of the density distribution. It is relatively simple to use and requires no real operator skill with the exception of choice of area size. It is not suitable for estimates of single schools (data sets 7-9) or complex areas with highly aggregated distributions (data set 2). The assumptions are that the within stratum data are uncorrelated and the strata are independent.

5. Kriging

Spatial covariation can be used in the estimation of fish density locally at a point or globally over an area. A number of approaches have been developed for using spatial covariation in this way and the geostatistical literature is a particularly rich source of such applications (Matheron 1963 1965 1971, Journel and Huijbregts 1978, Cressie 1989). These techniques are now being applied in fisheries research (Crittenden 1989, Guillard, *et al.* 1990, Sullivan 1991, Conan 1985, Conan *et al* 1988a 1988b, Conan and Wade 1989, Gohin 1985, Nicolajsen and Conan 1987).

Estimates, such as that of fish abundance at a given location, may be derived as a weighted average of the observations taken near the point of interest. The observations are weighted in the estimate according to their correlation with other observations and with the point or area to be estimated. The shape of the area of the estimate and the coverage of the survey will also affect the weights through the computed correlations. The correlation is generally given as a function of inter-point distance, and may be derived directly from another measure of interpoint variation known as the variogram (Matheron 1971, Journel and Huijbregts 1978). The variogram is often the measure of choice because of its generality, since it does not require stationarity in the mean. The variogram is defined as half the expected value of the squared difference between two random variables that are located a distance 'h' apart.

The estimation (or prediction) variance, σ_e^2, is the expected deviation of the estimator from the random variable describing the density at a point, that is

$$\sigma_e^2 = Var\left[Z_V^* - Z_V\right] = E\left[\left[Z_V^* - Z_V\right]^2\right]. \qquad (5.1)$$

Note that this may differ from the variance of the estimator,

$$Var\left[Z_V^*\right] = E\left[\left[Z_V^* - E(Z_V^*)\right]^2\right], \qquad (5.2)$$

a statistic more commonly used in classical statistical approaches, but inappropriate here except under the right conditions. In terms of the covariances, the estimation variance may be computed as

$$\sigma_e^2 = \overline{C(V,V)} - 2\overline{C(V,v)} + \overline{C(v,v)} \qquad (5.3)$$

where V represents the total area of interest and v represents the area sampled. The average (noted by bar) will depend on the weighting used in computing the estimator.

This formula contains the mean covariance between two arbitrary points independently describing the volume, $\overline{C(V,V)}$, the mean covariance between a sample observation and an arbitrary point describing the volume, $\overline{C(V,v)}$ and the mean covariation between sample points, $\overline{C(v,v)}$.

Several alternative approaches for estimating global fish abundance using these principles are presented here. The general methods will be described first, followed by results and discussion. A comparison of the results from this spatial geostatistical approach, sometimes referred to as kriging, with other approaches discussed in this report is given in Section 8.

Application 1: Point kriging with possible trend removal

W. G. Warren applied point kriging (WP6-7), taking into account the following considerations:

For the Icelandic herring data, a density surface was described by taking as coordinates the distance from the coastline and the distance parallel to the coastline from an arbitrary origin. The non-zero data exhibited noticeable positive skewness. The Box-Cox (1964) transformation was used to determine a transformation that would yield an approximate normal (Gaussian) distribution. The square-root transformation appeared suitable for all three cases.

A rectangular box with sides parallel and perpendicular to the coastline was then constructed about the patch separately for each survey. Each box was divided into rows and columns to form cells of approximately equal area. The number of data points that fell into a cell varied and for some cells this number was zero. Trend removal was accomplished by an unbalanced analysis of variance (ANOVA), with rows and columns as factors.

The ANOVA-estimated cell values were subtracted from the transformed data values at points falling within the appropriate cell. The residuals were then used to construct a conventional spherical variogram where isotropy was assumed. The global estimate is

approximated by computing the point estimates over a grid on the area, multiplying by the mean area about each point, and then summing. The variance of the estimate is computed similarly using the correlation between grid point estimates that are derived from the spatial correlations. Details relating to distance calculations, choice of variogram, and the global approximation were provided in working paper W6.

The simulated data were similarly analyzed but no trend removal mechanism was applied. The Box-Cox approach on the non-zero data suggested that a logarithmic transformation would be appropriate. The variogram was estimated in two directions but no systematic difference was found, so a single conventional spherical variogram was computed by combining the two estimates. Further details were provided in working paper W7.

Application 2: Global block kriging

The estimate of the average density and associated estimation variance over a global area of interest is obtained from all points sampled, in a one step procedure as described in Matheron (1971). The information from the variogram, γ, and from the respective distance between the points and to the area, and the shape of the area are used for calculating an estimation variance of the form:

$$\sigma_\varepsilon^2 = 2\sum_{i=1}^{n} w_i \bar{\gamma}(v_i, V) - \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \bar{\gamma}(v_i, v_j) - \bar{\gamma}(V,V) \quad (5.4)$$

where w_i are the statistical weights attributed to the n point samples v_i, and V is the global area studied. The variogram may or may not be isotropic, i.e. identical in all directions. This estimation variance is minimized by differentiating with respect to each of the weights and to a lagrangian parameter, λ, in order to optimize the estimate of the weighted average and to avoid bias under the constrain $\Sigma w = 1$.

The resulting optimized variance or kriging variance is:

$$\sigma_k^2 = \Sigma w_i \bar{\gamma}(v_i, V) + \lambda - \bar{\gamma}(V,V) \quad (5.5)$$

In the particular Gulfkrig software application designed by Conan and Wade, the numerical calculations of an average variogram over an area V can be made over any irregular shape. If the global area over which the estimate is to be made is not predefined prior to calculations, it may be defined by first calculating contours of local estimation variances by point kriging and then chosing a contour beyond which it is felt that extrapolating would be unsatisfactory.

The global kriging estimate can also be approximated by calculating a large number of point kriging estimates over a fine mesh grid, and then averaging. This procedure is useful when the number n of point observations is very large, since a matrix of $n+1$ x $n+1$ must be inverted in the direct estimate procedure.

In working paper W10, E. Ferrandis proposed a simplified computational procedure for calculating the statistical weights. This procedure reduces the dimensions of the matrix to be inverted to n by n.

Data sets 1 through 4, the Icelandic herring data sets, and the simulated data sets were all analyzed using the following methodology. No trend removal or variance transformation was applied to the data. Spherical isotropic variogram models were fit to each data set and a global estimate and its variance were derived by ordinary kriging applied within an irregular block defined by the variance contour around the area studied.

The GulfKrig software developed by G. Conan and E. Wade calculates global and local (either block or point) estimates of abundance and their variances using the method of ordinary kriging.

Application 3

A third application by P. Petitgas and J. Rivoirard was presented in working paper W9 and is given as Appendix B. For the Icelandic herring data set, since the data are regularly located throughout the field, the abundance is estimated using a simple arithmetic mean. Furthermore, since the field is large compared with the range of the correlation, the mean covariance between two arbitrary points independently describing the area $\bar{C}(V,V)$ and the mean covariance between a sample observation and an arbitrary point describing the area $\bar{C}(V,v)$ are found to be small. Thus, the estimation variance simplifies to the mean covariation between sample points $\bar{C}(v,v)$ which may be computed from the mean variance among samples plus the mean covariance between samples.

For data set 4 (Norwegian Herring) and for the Bering Sea walleye pollock data (data set 6) a

different approach is taken. Assuming the transects are parallel and that each transect traverses the entire width of the stock the integrated transect values may be taken as being one-dimensional observations on the stock. A variogram was then estimated in one dimension and geostatistical theory was applied to the overall abundance estimation and associated variance computation. Observing that the field is small with respect to the range of covariation the estimation variance must now include the mean covariance between points and the mean covariance between points (in 1D) and the sample points. But since the problem is one dimensional the computations are straightforward.

Two-dimensional computations were performed using Bluepack (1991), whereas one-dimensional computations required no software.

6. Generalized linear models

6.1 Description of method

The basic GLM assumes that the structure of the schools is of the form of a mean plus a random error, where the mean is a function of location (and potentially other variables), but the error contains no structure. The mean is parametrized as a function (the inverse link function) of some linear terms and the distribution of the measurements is from the exponential family.

Generalized linear models are described in several texts, including McCullagh (1983), McCullagh and Nelder (1989), and Nelder and Wedderburn (1972). A clear introduction to their use, using the GLIM package (Baker and Nelder, 1978) is given in Aitkin *et al* (1989). GLMs can also be fitted within the Splus package (Becker *et al.* 1988, Anon. 1991).

6.2 Application of method

Only the three herring data sets were considered, since GLMs can only be expected to work well with this type of acoustic data, when there is a single aggregation of fish, within a limited area. In all other data sets considered (with the possible exception of walleye pollock), there tend to be aggregations with low values in between or around. Polynomials of reasonably low order cannot fit such data well. An analysis using GLMs on one of these data sets was introduced briefly in Anon. 1990.

Assuming a gamma density and a log-link seems a reasonable assumption, but high-degree polynomials are needed to fit the data well. Working paper W2 found that polynomials of up to the 6th degree were needed for some of the data, and even in this case, (pseudo-) R^2-values were only at the 0.5-level.

However, since there seems to be one large "lump" in each aggregation, only the results from fitting a simple paraboloid as a function of location for each data set are presented in this report.

The numbers obtained are given in section 8. The areas and gridpoints used were based on a grid of 0.2 NM by 0.2 NM cells, which were defined in such a fashion as to cover the survey tracks with a minimal amount of extrapolation, yet retaining a roughly convex region.

6.3 Discussion

The actual values obtained (80775, 76933 and 81250) are quite close, the range of the three being only 5% of their average. This is in stark contrast to the "confidence bounds", based on integrating an estimated one standard error in each direction from the surface, all of which are over 13% in each direction from the corresponding estimate. It must be noted that these bounds are only approximate and further, they approximate the 68% confidence interval, corresponding to one standard error in each direction. They are used only to obtain an approximate "C.V." ratio (standard error/mean). The approximate 95% confidence interval will be correspondingly wider.

It would seem, therefore, that although the log-polynomials do not fit very well, there is considerable smoothing involved in the integration and this is not appropriately reflected in the variance estimate behind the confidence bound.

Some concerns were raised during the meeting that the reverse transform (exponential), required to evaluate the surface on a grid for

integration, would introduce a bias. Although this may be the case, it is not obvious what the precise effect is, or how it should be corrected for, since the equations used for estimating the parameters in the GLM model are different from simply log-transforming the data before fitting a model. These equations are based on the differences between the actual (untransformed) observations and their means according to the model.

In lieu of the results in working paper W6, the residuals from the GLM are expected to be correlated, reducing the validity of the error estimates.

7. Generalized additive models

7.1 Introduction

Generalized additive models are used here as methods for detection of spatial trends. They can be used as a tool in abundance estimation, but more importantly as an aid to demonstrating or quantifying relationships between the spatial distribution of abundance and environmental factors. In cases where the average value of a variable changes explicitly over space, this change is assumed to be a trend. A *spatial trend* is assumed to mean a change in the average density which is a function of the spatial location. Besides detecting trends in abundance over space, these changes can be related quantitatively not only to spatial location but to environmental factors such as depth and temperature. The existence of such quantitative relationships strengthens understanding of the factors that influence the explicit spatial distribution of fish species abundance and also gives a degree of explanation of this distribution that may serve to reduce the variance in abundance estimates by providing additional information about abundance distribution through covariates that are easy to measure.

Generalized additive models relate the changes in abundance to spatial covariates, without restricting the functional form of the relationship (Kaluzny, 1987; Hastie and Tibshirani 1986, 1990). This method allows nonlinear trends and includes covariates which potentially determine the spatial patterns in the data. Bootstrap methods give information on the variability around the trends and permutation tests are used to determine the significance of trends. The use of generalized additive models for analyzing survey data is quite general in that the surface which is fitted to the data is only restricted to be a sum of smooth non-parametric functions. The form of the functions is not restricted to polynomials as in generalized linear models (GLM; McCullagh and Nelder, 1989). The functions are instead determined by a smoothing technique that reflects local spatial trends, while allowing trends over the entire space to be observed (if they exist).

A Generalized Additive Model (GAM) is a nonparametric generalization of multivariate linear regression. Both methods relate the dependent variable to possibly important covariates. However, in GAM the covariates are assumed to affect the dependent variable through additive, unspecified (not linear) functions. Scatterplot smooths (Chambers *et al.* 1983) in GAM replace least square fits in regression. In GAM, the data can come from any distribution in the exponential family (which includes the normal, Poisson and binomial distributions). Because of the flexibility of GAM in detecting and testing for trends in abundance, they are valuable in uncovering factors influencing fish distributions over several years. The theory and method for applying GAM, using the gam function in Splus (Chambers and Hastie 1991), is given in Appendix C.

7.2 Application of GAM to data sets

The primary focus in the GAM analysis of the data sets provided was on uncovering relationships between fish abundance and environmental factors. Only depth was provided as an ancillary variable (except for latitude and longitude of the sampling locations) and that only for the Icelandic herring and Bering Sea surveys. Analysis was most fruitful for the Bering Sea survey, where a significant trend for abundance with depth was found. This analysis is presented in Appendix C.

8. Comparisons across methods

8.1 Introduction

In the following sections, the results from applying different methods to each data set are compared in terms of the estimated abundance and the estimated error in that number. The density estimate and area used are also considered, since in some cases these severely affect the results.

Within tables in the comparisons subsections (8.3-8.6), A denotes the arithmetic mean, B denotes the method based on the Box-Cox transform, G1-G3 will denote different geostatistical methods, S is used to denote the spline approximation method, L denotes the method based on generalized linear models and T denotes the approach of accumulating along a transect, followed by analyzing the sample using the ratio method (as described in Anon. 1990, p. 80)

8.2 Variance estimation

When the results from the different computations are compared, several issues must be borne in mind. One of these is the definition of the quantity of interest. The term "total abundance estimate" can be - and has been - interpreted in different mathematical ways, resulting in entirely different estimates of associated variances.

The approaches which have here been called "classical", as well as the GLM and GAM methods have as their underlying purpose the estimation of a "response surface" which can be of the form of a step function, a polynomial in location or an abundance-depth relationship. The surface estimates the **expected value of the response** at each location. The associated abundance estimate is the volume under that surface.

The method of point kriging, however, fits a surface which estimates **the unobserved individual responses** at each location. The associated abundance estimate is also the volume "under the surface", albeit a different surface.

A fundamental difference in approach is thus evident. This sometimes has drastic consequences for the variance estimate.

Under the "classical" approaches, including GLM and GAMs, the existence of autocorrelation in residuals reflects a redundancy of information, which reduces the effective degrees of freedom, and increases the variance in the integral/abundance estimate. However, the kriging school of thought is the exact opposite, essentially stating that since there is autocorrelation, there is better information to interpolate between data points, resulting in a better estimate of abundance.

A simple example will suffice to illustrate the difference quite clearly. If the transects are parallel, they can be added up to reduce the problem to one dimension. Suppose that there is no trend in the data, so that the expected value is constant in the remaining dimension. The "classical" approach is to attempt to estimate this single mean. The kriging approach is to estimate the entire curve (which will not be a straight line due to the autocorrelations).

If the item of primary interest is the expected value, an increase in the autocorrelation will obviously reduce the effective degrees of freedom. In fact, as the autocorrelation goes to 1, the information in the data set is reduced to just one observation, as far as the estimation of the expected value is concerned.

If the item of primary interest is the curve itself along with the integral of that realization of the process, then an increased autocorrelation will lead to more information about the behavior between data points, thus reducing the variance. In the limiting case, as the autocorrelation goes to 1, the curve will become perfectly known, as will the abundance.

It must be noted that in cases when a grid is regular, both approaches may simply be using the arithmetic mean as an estimator, but the variance estimates may be totally different, with one giving CVs as low as a few percent, the other yielding CVs which commonly range from 20 to 50%. As is described above, this simply stems from the choice between estimating a mean surface and predicting an unobserved surface. Whether the CVs are really as low or high as indicated is not known *a priori*, but can be ascertained through other methods, such as cross-validation.

It is therefore essential to precisely define the quantity of interest: should it be the surface of expected values or the unobserved measurements between the transects ? This question can be at least partly answered by investigating the source of the autocorrelation. The acoustic measurements involve several levels of variation, which for convenience can be separated into "process error" autocorrelation (the structured variability of the resource) and "observation error" autocorrelation (the structured variation in the measurement instruments). If most of the autocorrelation stems from the observation error, then there is good reason to treat it as true error and consider its effects negative ones.

Acoustic measurements are capable of detecting sharp changes in density, so most of the autocorrelation along and across transects will be due to contiguous behavior of the resource. This implies that when autocorrelations are observed along and across transects, they include important information about the resource itself and should be utilized as best possible for the estimation of the resource.

This leads immediately to the use of the criteria and language used in kriging, specifically with respect to the term "total abundance" which is defined as the abundance that would have been measured if the area had been completely covered - not the expected value of that quantity.

In mathematical notation, the variance of primary interest is the prediction variance,

$$Var\left[Z_v^*-Z_v\right]=E\left[(Z_v^*-Z_v)^2\right] \qquad (8.1)$$

where Z_v is the double integral of the process and Z_v^* is an estimate. The variance used in the alternate ("classical") approach is

$$Var\left[I_v^*-I_v\right]=E\left[(I_v^*-I_v)^2\right] \qquad (8.2)$$

where I_v^* is an estimator of the integral of the expected value of the process (the above equations assume unbiasedness of the estimators). As explained above, these two variances may be totally different, even if the estimators are both equal to the sample mean: $I_v^*=Z_v^*=\overline{Z}^*$.

8.3 Data sets 1-5

Tables 8.1-5 give, for data sets 1-5, the estimated densities (s_A), the corresponding C.V. (defined here as 100 times the standard error of s_A over s_A), the area used and the total abundance. Analyses of these data sets were also given in Anon. (1990), but many of the values have been revised.

Table 8.1. Summary of the results of the analyses of test data set 1.

ID	Method	s_A m^2/NM^2	CV %	Area 10^3 Nm^2	s_A*Area /10^6	Analyst
A	Arithmetic mean	75				
B	Box/Cox transf.	68	9	55	3.7	Simmonds
G1	Kriging	85	43.5	54	4.6	Conan & Wade
S	Spline	77	N/A	53	4.0	Stolyarenko

It is noted that in table 8.1, methods G1 and S both give higher abundance estimates than method B, but the CV estimate in B is much lower than for G1. It must be borne in mind that these two CV-values are estimates of different quantities, as described in section 8.2.

Table 8.2. Summary of the results of the analyses of test data set 2.

ID	Method	s_A m^2/NM^2	CV %	Area Nm^2	s_A*Area /10^3	Analyst
A	Arithmetic mean	297				
B	Box/Cox transf.	48	37	49	2.4	Simmonds
G1	Kriging	443	4	47	20.9	Conan & Wade
S	Spline	259	N/A	51	13.2	Stolyarenko

In table 8.2, the CV of B is much larger than that obtained in G1 (although these two have different interpretations). The actual abundance estimates also vary widely, with the Box-Cox transform (B) giving the lowest, the spline approximation (S) intermediate and global kriging (G1) giving the largest estimate. It must be pointed out that the areas used by the different analysts are different, but this does not fully explain the differences. The group noted that this dataset is particularly difficult to analyze and few methods would be applicable to this kind of data (c.f. section 3 and Fig. 3.2), since it is to a large extent due to the different area definitions.

Table 8.3. Summary of the results of the analyses of test data set 3.

ID	Method	s_A m^2/NM^2	CV %	Area 10^2Nm^2	s_A*Area /10^6	Analyst
A	Arithmetic mean	1793				
B	Box/Cox transform	1327	7	55	7.3	Simmonds
G1	Kriging	1558	33.8	63	9.8	Conan & Wade
G2		2089	14	90	18.8	Guillard & Gerdaux
G3		1911	22	83	15.9	Armstrong
S	Spline				7.8	Stolyarenko
T	Transects as strata	3072	30	19	5.7	Williamson

In test data set 3, the CV estimates vary widely (table 8.3). The geostatistical methods (G1-G3) give abundance estimates which are up to two to three times the estimates obtained by the other methods.

Table 8.4. Summary of the results of the analyses of test data set 4.

ID	Method	s_A m^2/NM^2	CV %	Area Nm^2	s_A*Area /10^6	Analyst
A	Arithmetic mean	774				
B	Box/Cox transform	560	9	6100	3.4	Simmonds
G1	Kriging	1062	51	3000	3.2	Conan & Wade
G2		1690	12%	1975	3.3	Petitgas
S	Spline				3.5	Stolyarenko
T	Transects as strata	1512	31	2200	3.3	Williamson

In data set 4, the difference between the results from the geostatistical methods G1 and G2 is considerable. The main explanation for this probably lies in definition of the area over which the estimation was performed. The area is 1975 sq. NM for G2 and 3000 sq. NM for G1. In the G2 approach, the zero values at the extremities of the transects are interpreted as zeros exterior to the fish spatial extension. Therefore, the area over which the estimation errors are made is reduced. In the G1 application, on the other hand, a much larger surface was defined. Further, estimation errors in areas were assumed, whereas no error was assumed in the G2 application.

The CV of T is very high in data sets 3 and 4. It is believed that this stems from this method not taking into account the inter-transect spatial correlation.

Table 8.5. Summary of the results of the analyses of test data set 5.

ID	Method	s_A m^2/NM^2	CV %	Area 10^4Nm^2	s_A*Area /10^4	Analyst
A	Arithmetic mean	14				
B	Box/Cox transform	9	8	13	110.0	Simmonds
G1	Kriging	14	18	19	266.0	Conan & Wade
S	Spline				87.5	Stolyarenko

Three methods were applied to this test data set. The resulting estimates varied widely.

8.4 Data set 6 Walleye pollock

Two estimates of transect mean density were provided, as indicated in table 8.6.

Table 8.6. Abundance estimates for walleye pollock (data set 6)

ID	Method	Abundance	CV, %	Analyst
G1	See sect. 5 (appl. 2)	13.220	2.3	Petitgas & Rivoirard
G2	See below	13.019	3.5	Warren

The G2 estimate is based on the total of the mean densities over the number of elementary sampling units in each transect, Σd_i, say. However, the lengths, l_i, of the elementary sampling units vary slightly and the G1 estimate is based on $\Sigma l_i d_i$.

Although the original data set consist of 27 transects, a transmission glitch of some sort erased one transect from some of the diskettes sent to participants. This omission is unlikely to have affected the results to any noticable extent.

The G1 method is described in Section 5, above. In G2, the transect was also taken as the sampling unit but the transects were treated as a systematic sample with a random start. Variance estimation was then accomplished by assuming a polynomial trend on the transect totals and applying the formula given in Cochran (1977) as extended by Kingsley and Smith (1980). A quartic was judged to be appropriate. Details are

given in working paper W8. Essentially, it was assumed that the residuals, after removal of the fourth-order trend, would be independent. The slightly greater CV obtained in G2 relative to G1 suggests that some residual serial correlation may have remained in the residuals.

8.5 Data sets 7-9 Icelandic herring

Five different analyses for each of the 3 surveys were available for the meeting. The results are summarized in table 8.7.

Table 8.7. Summary of results of analyses of Icelandic herring data. Abundances and averages divided by 1000.

Method	ID	Statistic	1	2	3	Analyst
GLM	L	Abund.	80.8	76.9	81.3	Stefán-sson
		Mean Density	5.72	4.34	5.32	
		Area	14.1	17.7	15.3	
		CV, %	16	18	13	
3D-Spline	S	Abund.	110.0	55.4	94.3	Stolyar-enko
		Mean Density	n.a.			
		Area	n.a.			
		CV, %	n.a.			
Block kriging	G1	Abund.	127.5	117.6	113.4	Wade
		Mean Density	4.94	3.24	2.77	
		Area	25.8	36.3	40.8	
		CV, %	19.9	18.6	22.5	
Point kriging	G2	Abund.	103.1	93.3	107.1	Warren
		Mean Density	5.54	3.52	3.87	
		Area	18.7	26.5	27.7	
		CV, %	16.6	n.a.	n.a	
Mean abund.	G3					Petitgas Rivoirard
		Mean Density	5.53	3.25	3.59	
		Area				
		CV[1)]				

[1)] 12% for area 15.0, 14% for area 33.5

No estimate of area size or mean density of precision was provided with the application of S.

The methods of area estimation differed from one method to the next:

- The boundaries for the GLM model application (L) were chosen to include all locations of observations.
- The areas for the application of point kriging (G1) were determined as those locations for which the estimated density was non-zero.
- Areas in the block kriging application (G2) were defined as the outline of an variance contour line of an arbitrary level (value not specified). They correspond approximately to the outline of the sample points plus a corridor of width slightly smaller than the range of influence.
- The area used in G3 was limited to the zone that was swept. It was taken as 15 sq. NM for all sets. An extension on each side of the survey was also considered, giving an area of 33.5 sq. NM.

Consequently, the areas, as used for the GLM model, were smaller than those used in point kriging, which, in turn, were smaller than those used in block kriging.

Application G2 provided a CV estimate for survey 1 only. Little change in the mean density, area and abundance estimates is anticipated by using a finer grid for G2, but the CV estimates may be somewhat reduced.

The GLM estimates of CV, obtained in the L application, are not comparable in that they represent the pointwise integration of one-standard error confidence limits and should therefore only be considered approximations.

Other choices of distribution, link function and degree of polynomial in L gave alternative abundance estimates ranging from 78225 to 103734, from 67514 to 96185 and from 74195 to 90543 for surveys 1, 2 and 3, respectively. With high-degree polynomials, slight changes in area definitions can drastically change the results.

The estimates of mean density are, not surprisingly, inversely related to the estimates of area. The relationship is, however, not that of exact inverse proportionality so that the G1 estimates of abundance turn out to be greater than those of G2 which, in turn, are greater than those of L.

The mean density estimates in G3 (the sample means) are closest to those of G2 differing by

0.1%, 8.1% and 7.7% for surveys 1, 2 and 3, respectively.

The spline approximation (S) abundance estimates for surveys 1 and 3 are also closest to those in G2 differing by 6.7% and 12.0%, respectively. The S estimate for survey 2 is clearly unrealistically low, since the data are supposed to represent three surveys of the same aggregation. By the same token, the fixed areas assumed in G3 are also unrealistic since this would imply abundances for surveys 2 and 3 of approximately 60% that of survey 1.

Since the data sets represent three surveys of the same aggregation, it was expected that the abundance estimates obtained by any one method would be consistent over the three surveys. Table 8.8 expresses, for each analysis, the range of the three estimates of abundance as a percentage of their mean. The differences between the estimates are relatively small in relation to the estimated CVs.

Table 8.8. Comparison of between-survey results for Icelandic herring.

Application	Range (of 3)	Mean (of 3)	Range / Mean %
L	4317	79653	5.4
S	54600	86567	63.1
G1	14173	119505	11.9
G2	13875	101178	13.7
G3	n.a.		

The estimate of CV in G2 (survey 1) seems comparable to that obtained in G3, and perhaps somewhat less than that obtained in G1. This is consistent with the conjecture that, in employing ordinary kriging, somewhat greater precision would be attained by removing any seemingly well defined trend.

8.6 Data sets 10-15: Simulated data

For ease of tabulation, all total abundance values have been scaled down by 100000. For all six surveys the true abundance is 87.67 and the mean density is 97.41. The population of survey 2 is that of survey 1 rotated through 90 degrees.

Three different analyses (or partial analyses) of data sets 10-15 were available for the meeting. The results are summarized as follows:

Table 8.9. Comparisons of results from simulated data sets 10-15.

	Data set	10	11	12	13	14	15
	Simulated survey	1.1	1.2	1.3	2.1	2.2	2.3
ID / Method / Analyst	Statistic						
A	Mean density	95.8	95.9	95.8	91.9	92.6	92.3
Arith. mean	Abund.	86.20	86.30	86.20	82.70	83.10	83.00
	CV	n.a.					
	Diff. from true	1.7%	1.6%	1.7%	5.7%	5.2%	5.3%
S	Mean density	97.1	97.2	97.1	95.8	96.3	96.3
2D-spline	Abund.						
	CV						
Stolyarenko	Diff. from true	2.8%	3.0%	2.8%	1.5%	2.0%	2.0%
G1	Mean Density	94.5	94.4	94.1	94.7	95.7	92.7
Kriging	Abund.	85.06	84.96	84.96	85.23	86.13	83.43
	CV	12.2%	10.4%	10.2%	12.8%	10.8%	8.8%
Wade	Diff. from true	-3.0%	-3.1%	-3.1%	-2.8%	-1.8%	-4.8%
G2	Mean density	102.59	n.a.	n.a.	97.32	94.89	96.98
Kriging	Abund.	92.33			87.59	85.40	87.29
	CV	n.a.					
Warren	Diff. from true	5.3%			-0.1%	-2.6%	-0.4%

Only estimates of abundance were provided with the S method (no measure of precision).

The G2 abundance estimates for Surveys 2.2 and 2.3 are preliminary. They were based on a smaller critical distance than intended; i.e. the distance of data points used from the interpolated locations.

The S estimates are all slightly above the true abundance by an average of approx. 2.4%. Conversely, the G1 estimates are all slightly below the true value, by an average of 3.1%. The G2 estimates are above the true value for survey 1 and below for surveys 2.1, 2.2 and 2.3.

W. Warren also presented estimates obtained by treating the major transects as a systematic sample (Kingsley and Smith 1980) with a single random start, although clearly, a random start was not employed. Not all the data were used, as the short transects linking the ends of the long transects were omitted. The results were as given

in table 8.10.

Table 8.10. Results based on assuming a random start.

Data set	10	11	12	13	14	15
Simul. survey	1.1	1.2	1.3	2.1	2.2	2.3
Abund.	94.36	94.51	94.32	75.75	75.13	74.97
Diff. from true	7.6%	7.8%	7.6%	-14.7%	-14.3%	-14.5%
CV	6.9%	6.9%	6.9%	4.9%	4.9%	4.9%

These results are interesting in that, as noted above, the underlying population for survey 2 was that for survey 1 rotated through 90 degrees. It can be seen from Fig. 3.10 that the populations consists of a "mountain range" running through the center of the region and parallel to one pair of sides. Consequently, the transects of survey 1 cut across the "mountain range" thus giving transects totals that exhibit relatively moderate variability but with no clear trend. For survey 2, the transects run parallel to the "mountain range" so that the transect totals exhibit much greater variability but also an essentially quadratic trend. Since, for survey 2, a quadratic trend was assumed in the variance estimation, this accounts for the smaller CV estimates. The lower abundance estimates are due, in part, to the omission of the short end transects which cross the "mountain range".

8.7 Discussion

The above results are, perhaps, as notable for the consistencies as for the discrepancies, most of which can be explained, at least in part.

During discussion, the group considered the described fish stocks and a number of others. There was general agreement that some structure could be assumed in all cases considered. There was evidence of large scale changes in mean density in most cases. In addition to these "trends," additional spatial autocorrelation was **always** expected to be present.

Based on these conclusions, the group agreed that there was in many cases potentially great benefit involved in utilizing the spatial structure when estimating the abundance of the resource, and, in particular, there is potential gain when estimating the precision of that quantity.

There is no doubt that spatial analysis can give a more realistic measure of precision of a survey than classical methods and, under certain circumstances, a better measure of abundance or mean density. It is not, however, a panacea. It would be a fallacy to assert that there exists a "black box" that can be used to process spatial data and that will yield viable results under all circumstances.

Depending on the severity of the trend, it may need to be removed before applying covariance techniques, although Journel and Rossi (1989) have shown that equivalent results may, in some cases, be obtained by using appropriate data windows when applying techniques which do not assume the existence of trend.

Spatial analysis can be viewed as a sequence of steps at each of which a choice must be made of the several options that are available (e.g. transform or not, if so which transform ? Trend removal or not, if so how? Should ancillary data be used ? Do the two-dimensional data lend themselves to being collapsed into one dimension?). There are as yet no well defined rules as to which choice would be best. While general guidelines can be given, each situation must be treated on its merits, and the viability of the results depends, to some extent, on the skill and experience of the analyst.

Spatial analysis cannot be divorced from survey design. While in theory it is possible to analyze spatially any configuration, spatial analysis appears to be most effective under systematic designs.

9. Conclusions and recommendations

9.1 Applicability

The aim of the workshop was to examine the applicability of spatial statistical techniques to acoustic survey data, with particular attention to global abundance estimation, variance estimation, and mapping. This has been done with respect to so-called classical or traditional statistical techniques, generalized linear models (GLMs), generalized additive models (GAMs), and geostatistical or kriging techniques.

In the course of comparing the several methods, workshop participants managed to clarify a matter of long-standing contention,

namely that of variance estimation. This is described in detail in Section 8.2. In essence, a distinction should be made between viewing the fish stock as a "pure" random process (independent, identically distributed random - i.i.d. - variables) and viewing it as a process with structure. This affects the estimation of variance. If the distribution of the fish stock is a pure random process, the variance is that of the mean estimate. If the spatial distribution is structured, i.e. has autocorrelation, then the estimated variance is based on the difference between the process as observed, and as predicted making explicit use of autocorrelation. Insofar as fish stocks do have structure in space and this can be ascertained by sampling along transects crossing the aggregation, the second view must be the preferred one. Since the whole point of spatial statistical techniques is to exploit structure as observed, these must at least be recommended as useful techniques.

In fact, the general discussion conducted on the basis of specific data analyses supports a stronger recommendation. This is that spatial statistical techniques be regarded as integral in the analysis of acoustic survey data. In other words,

> **Recommendation 1** Spatial statistical techniques are applicable to acoustic survey data and are recommended as suitable for the following: (1) estimating global abundance of acoustically surveyable fish stocks, (2) obtaining an associated estimate of precision, and (3) mapping the spatial distribution of the stock.

The precision of the global estimate is defined here in terms of the mean-square difference between the observed distribution and that predicted on the basis of observed autocorrelation.

9.2 Association of techniques with spatial features of the stock

A number of different kinds of fish distribution are recognized. These may be characterized *a priori* by the range of the autocorrelation with respect to the extent of the distribution or *a posteriori* by the scale of variation with respect to the inter-transect distance.

The following general situation is considered first: a stably located fish distribution is confined to a known geographical region, which is surveyed according to a grid of parallel, equally spaced transects. The statistical characteristics of the fish distribution can be categorized as follows:

1. The scale of variation is large compared to inter-transect distance. Exs. Test data sets 6 (walleye pollock), 7-9 (Icelandic herring), 10-15 (simulated data),
2. The scale of variation is comparable to inter-transect distance. Ex. Test data set 4 (Norwegian pelagic stock off coast),
3. The scale of variation is small compared to inter-transect distance. Ex. Test data set 2 (Norwegian pelagic stock in fjord).

Geostatistical techniques of analysis can be applied in each of these situations. In the first and second cases, they will be able to exploit the observed structure, as characterized by the autocorrelation, and the resulting variance estimate will be lower than the classical variance estimate. In the third case, the geostatistical and classical variance estimates could be similar.

Strictly speaking, the choice of analysis should also be based on the scale of variation relative to the area sampled.

A second general situation is illustrated by the Icelandic summer-spawning herring. The bulk of the stock exists at the autumn survey time in one or two dense aggregations of initially unknown location. These must be found in order to estimate the abundance. When an aggregation is found, it is usually possible to sample this very densely. An application of geostatistics to the Icelandic herring found that the range of covariation was too small to obtain the benefits associated with high spatial correlation. This resulted in an appropriate increase in the variance estimate as compared to an estimate assuming independence.

A third general situation is that of migration, which requires special surveying tactics. These are described in Simmonds *et al.* (1991). This situation requires detailed examination, not undertaken at the workshop.

An underlying assumption employed here is that the biology of the fish stock being surveyed is known, at least in its gross whereabouts at the time of the survey. Given this knowledge, the following recommendation can be made:

Recommendation 2 Among spatial statistical techniques, geostatistics, i.e. analysis using the variogram, is specifically recommended for the analysis of acoustic survey data.

If the acoustic survey has been performed over a grid composed of parallel transects reaching the boundary, then the variance can be estimated according to a quite simple procedure. Each value of density is exact for the particular, small interval of sailed distance. The total density along each transect is computed by simple summation. The resulting set of numbers constitutes a one-dimensional distribution. This is necessarily less rough, or spatially more correlated, than the underlying two-dimensional fish distribution. Illustrative examples are found in the analyses of test data sets 4 (Norwegian pelagic stock off coast), 6 (walleye pollock), and 10-15 (simulated data). In these particular examples, the range of spatial correlation of the one-dimensional data is large in comparison to the extent of the distribution. Application of geostatistics here will give both a lower and more realistic estimate of variance than is obtainable by classical statistical analysis.

It is noted that for the general acceptance of geostatistical techniques, some form of education and dissemination of information is required.

When synoptic knowledge of the whereabouts of the fish stock is lacking, estimation of abundance is not generally possible. Knowledge of fish biology is a precondition for conducting a proper survey, thence analyzing resulting measurements of fish density in order to estimate abundance over a region.

9.3 Analysis procedures

The phases of an analysis of acoustic survey data are:

1. Exploratory data display and analysis, to learn about the characteristics of the data, including possible connection with other variables, namely covariates,
2. Diagnosis, or selection of the best analysis technique,
3. Analysis, or exercise of the selected technique with the particular survey data,
4. Evaluation, including judgement of the quality of the analysis in the context of the degree of coverage of the stock by the survey grid and how well the analysis assumptions are met.

In the analysis phase, generalized additive models (GAMs) may be useful for associating other variables with the fish distribution. Examples include those of bottom depth, as in test data set 6 (walleye pollock), and temperature, as considered by Shinomiya and Tameishi (1988), among others, but not considered at the workshop. These techniques are particularly valuable for facilitating interpolation of measurements of fish density between transects, hence aiding the process of mapping fish distribution, as discussed in Appendix C. Hence,

Recommendation 3 Generalized additive models should be considered for use in exploratory data analyses to aid in choosing the specific analysis technique, and in the analysis process itself, as to map the distribution.

Association of the pattern of fish distribution with other variables can have major significance for the conduct of acoustic surveys. The potential to improve both the survey design and quality of analysis result is emphasized.

REFERENCES

Aitchison J. 1955. On the distribution of a positive random variable having a discrete probability of mass at the origin J Am Stat Soc 50, 901-908

Aitkin, M., Anderson, D, Francis, B., and Hinde, J. 1989. Statistical modelling in GLIM. Clarendon Press, Oxford. 375 pp.

Anon. 1990. Report of the study group on the applicability of spatial statistical techniques to acoustic survey data. ICES C.M. 1990/D:34.

Anon. 1991. S-plus user's manual. Statistical Sciences, Inc. Seattle, Washington, USA.

Baker, R. J., and Nelder, J. A. 1978. The GLIM system, release 3. Oxford: Numerical Algorithms group.

Becker, R. A., Chambers, J. M. and Wilks, A. R. 1988. The New S language. A programming environment for data analysis and graphics. Wadsworth & Brooks/Cole Advanced Books & Software. Pacific Grove, California.

Bluepack 1991. User manual. Center of Geostatistics. School of Mines. Fontainebleau.

Box, G. E. P. and Cox, D. R. 1964. An analysis of transformations. J Roy Stat Soc B 26 211-252

Chambers, J. M. and Hastie, T. J. 1991. Statistical models in S. Wadsworth and Brooks. Pacific Grove CA. 608 pp.

Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey. 1983. Graphical Methods for Data Analysis. Wadsworth, Belmont, CA.

Clay, C. S., and Medwin, H. 1977. Acoustical oceanography: Principles and applications (Wiley, New York, 1977).

Cochran, W. G. 1977. Sampling techniques. John Wiley, New York.

Crittenden, R.N. 1989. Abundance estimation based on echo counts. Ph. D. Thesis. University of Washington, Seattle WA.

Conan, G.Y. 1985. Assessment of Shellfish Stocks by geostatistical techniques ICES C.M. 1985/K:30, 19 pp.

Conan, G.Y., Moriash, M., Wade, E. and Comeau M., 1988a. Abundance and spatial distribution surveys of snow crab stocks by geostatistics ICES C.M. 1988/K:10, 23 pp.

Conan, G.Y., Bucrkle, E., Wade, E. Chadwick, M. and Comeau, M. 1988b. Geostatistical analysis of spatial distribution in a school of herring. ICES C.M. 1988/D:25, 18 pp.

Conan, G.Y. and Wade, E. 1989. Geostatistical analysis, mapping and global estimation of harvestable resources in a fishery of northern shrimp (*Pandalus borealis*) ICES C.M. 1989/D1, 29 pp.

Cressie, N. 1989. Geostatistics. American Statistician, 44:197-202.

Efron, B. 1979. Bootstrap Methods : Another Look at the Jackknife. The Annals of Statistics, Vol. 7, p 1-26.

Efron, B. 1982. The Jackknife, the bootstrap and other resampling plans, SIAM. Philadelphia.

Friedman, J.H. 1984. A variable span smoother, Tech. Rep. No. 5, Dept. of Stat., Stanford Univ., Stanford, CA

Gohin, F. 1985. Planification des experiences et interpretation par la theorie des variables regionalisees: application a l'estimation de la biomasse d'une plage. ICES C.M. 1985/D:3. 11 pp. (mimeo)

Guillard, J., Gerdeaux, D. and Chautrin, J. M. 1990. The use of geostatistics for abundance estimation by echointegration in lakes: the example of Lake Annecy. Rapp. P.-v. Reun. Cons. Int. Explor. Mer. 189:410-414.

Hastie, T. and Tibshirani, R. 1986. Generalized additive models, Statist. Sci. vol. 1. pp. 297-318.

Hastie, T. and Tibshirani, R. 1990. Generalized additive models, Chapman and Hall. London.

Journel, A. G., and Ch. J. Huijbregts. 1978. Mining geostatistics. Academic Press.

London. 600 p.

Journel, A. G. and Rossi, M. E. 1989. When do we need a trend model in kriging. J. Int. Assoc. Math. Geol., 21: 715-739.

Kaluzny, S. 1987. Estimation of trends in spatial data. PhD dissertation. University of Washington. Seattle. 165 pp.

Kingsley, M. S. C. and Smith, G. E. J. 1980. Analysis of data arising from transect sampling. In symposium on Census and Inventory Methods for Populations and Habitats. Ed. F. L. Miller and A. Gunn. Forest, Wildlife and Range Experiment Station, University of Idaho, Moscow, ID, Contribution No. 217. pp 40-48.

Knudsen, H. P. 1990. The Bergen Echo Integrator: an introduction. J. Cons. int. Expl. Mer, 47: 167-174.

MacLennan, D. N. and MacKenzie, I. G. 1988. precision of acoustic fish stock estimates. Can. J. Fish. Aquat. Sci. 45, 605-616

MacLennan, D. N. 1990. Acoustical measurement of fish abundance. J. acoust. Soc. Am., 87: 1-15.

Matheron, G. 1963. Principles of geostatistics. Econ. Geol., 58:1246-1266.

Matheron, G. 1965. Les variables regionalisees et leur estimation. Ed. Masson, Paris.

Matheron, G. 1971. The theory of regionalized variables and their application. Les cahieres du Centre de Morphologie Mathematique, Fascicule 5, Ecole des Mines de Paris, Bibliotheque de Fontainebleau. (English version).

McCullagh, P. 1983. Quasi-likelihood functions. Ann. Statist., 11: 59-67.

McCullagh, P. and Nelder, J. A. 1989. Generalized Linear Models, 2nd edition. Chapman and Hall, London.

Nelder, J. A., and Wedderburn, R. W. M. 1972. Generalized linear models. Journal of the Royal Statistical Society, Ser. A, 135, 370-384.

Nicolajsen, A., and Conan, G.Y. 1987. Assessment by geostatistical techniques of population of Icelandic Scallop (*Chlamys islandica*) in the Barent Sea ICES. C.M. 1987/:14 22 pp.

Shinomiya, H. and Tameishi, H. 1988. Discriminant prediction of formation of saury fishing ground by satellite infrared imageries. Nippon. Suisan Gakkaishi/Bull Jap. Soc. Sci. Fish. 54(7):1093-1099.

Simmonds, E. J., Williamson, N. J., Gerlotto, F. and Aglen, A. 1991. Survey design and analysis procedures: A comprehensive review of good practice. ICES C.M. 1991/B:54

Stanton, T. K., Nash, R. D. M., Eastwood, R. L., and Nero, R. W. 1987. A field examination of acoustical scattering from marine organisms at 70 kHz. IEEE J. Ocean. Eng., OE-12: 339-348.

Sullivan, P.J. 1991. Stock abundance estimation using depth-dependent trends and spatially correlated variation. Can. J. Fish. Aquat. Sci 48:0000-0000.

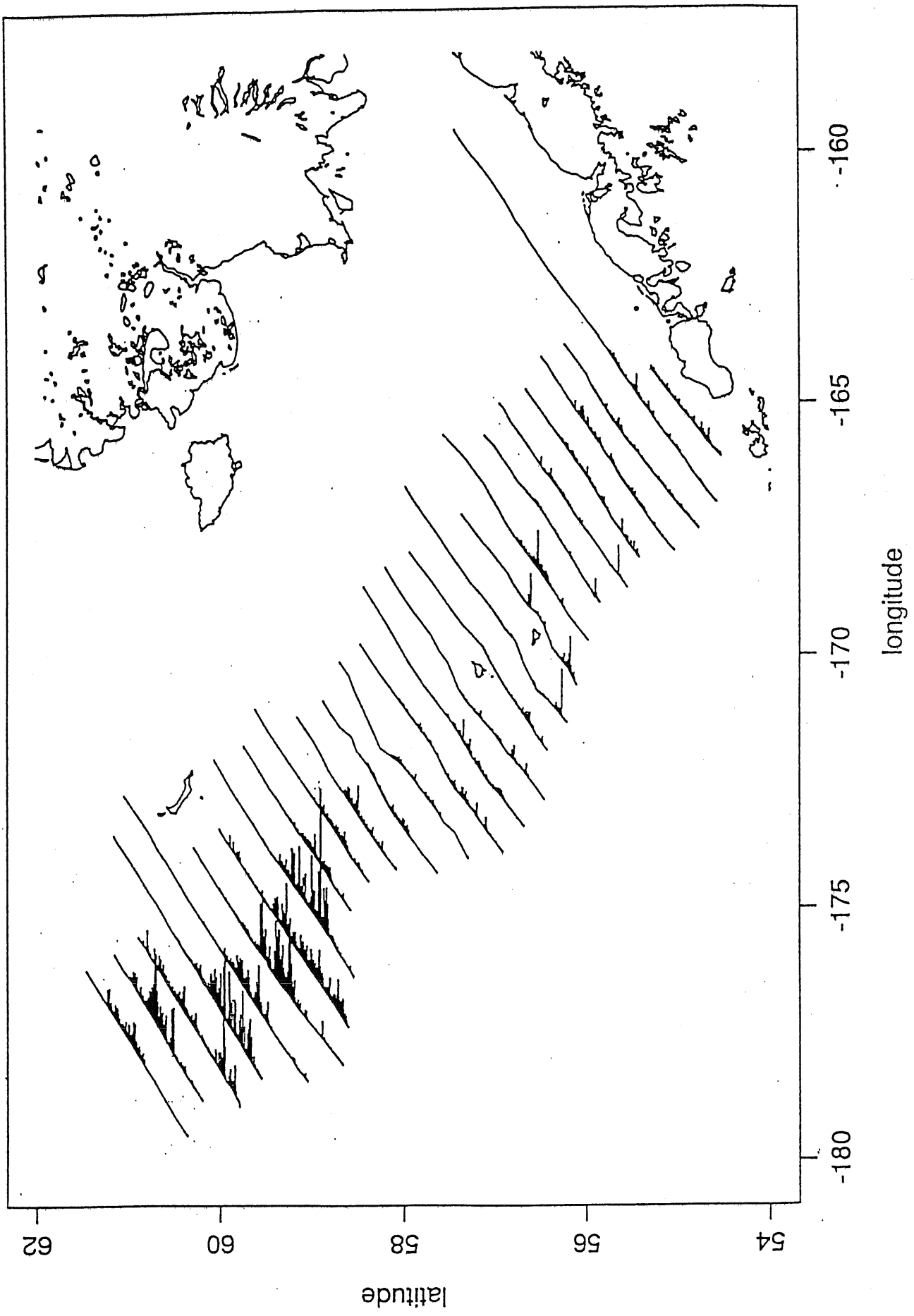


Figure 2.1 Bering Sea 1988 walleye pollock survey showing cruise tracks and a histogram of abundance along the tracks.

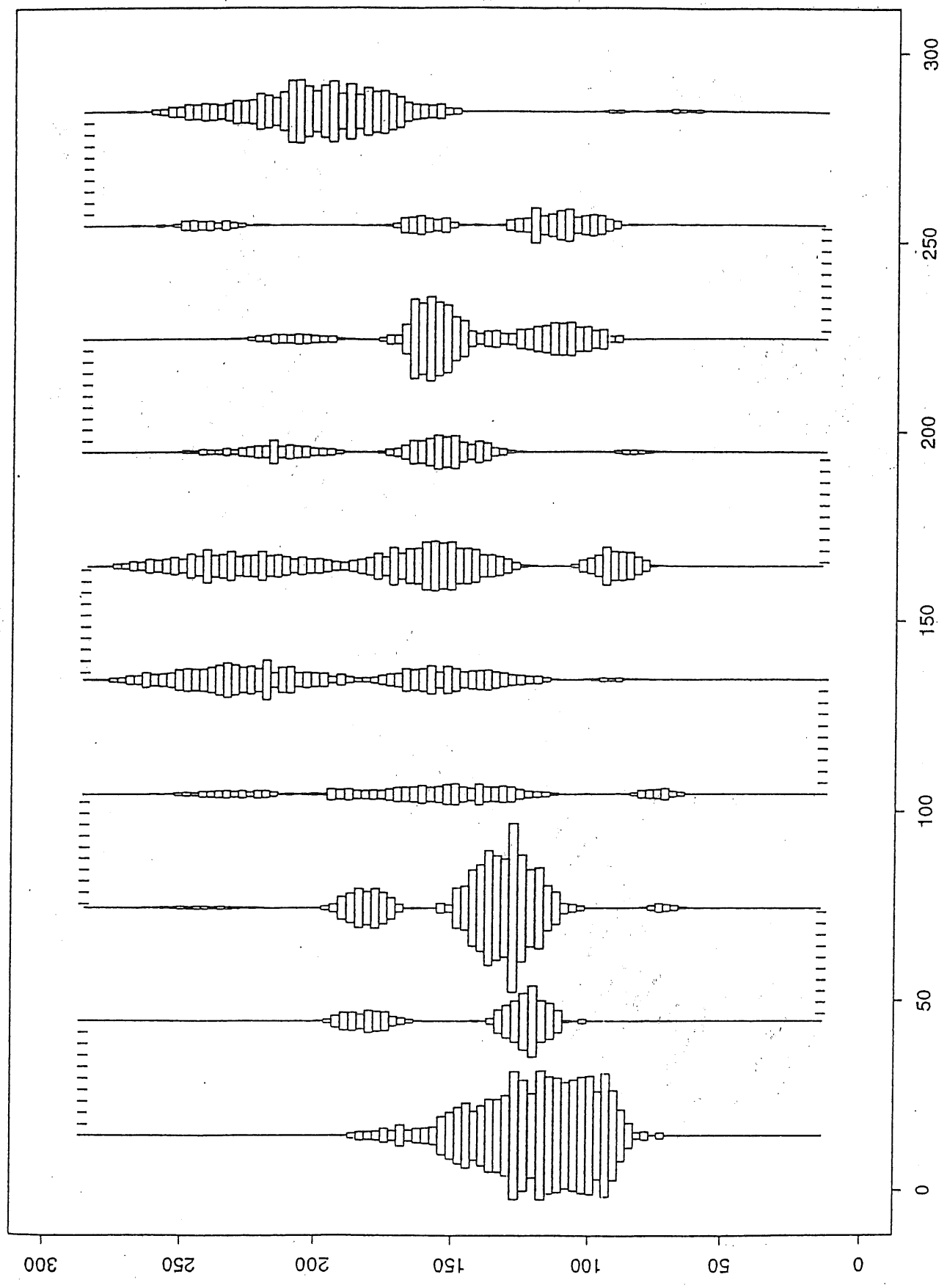


Figure 2.2 Acoustic survey tracks based on simulated data showing abundance along the track as a series of scaled rectangles.

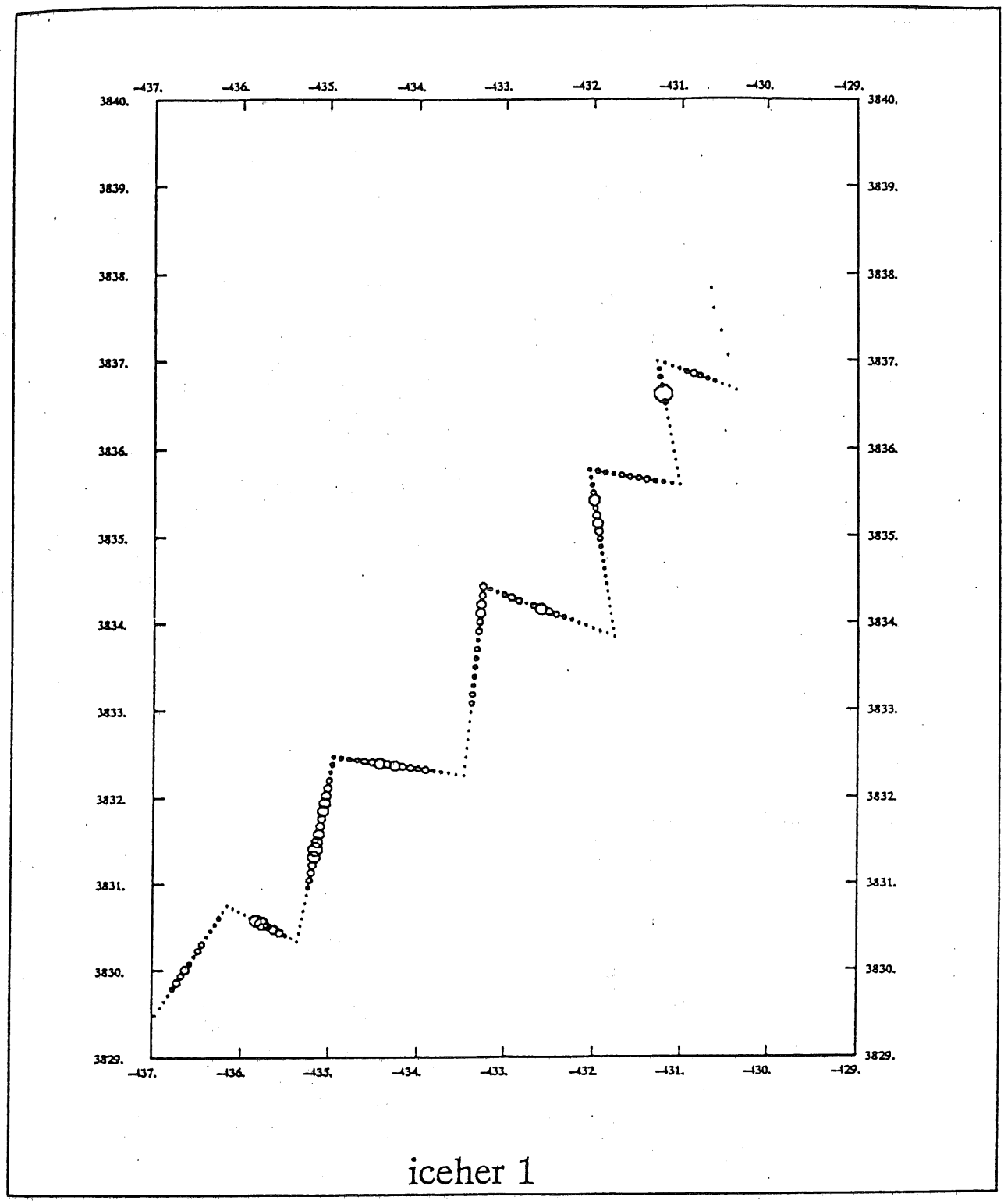


Figure 2.3 Acoustic survey track for Icelandic herring survey test data set Iceher1 showing abundance along the track as a sequence of scaled circles.

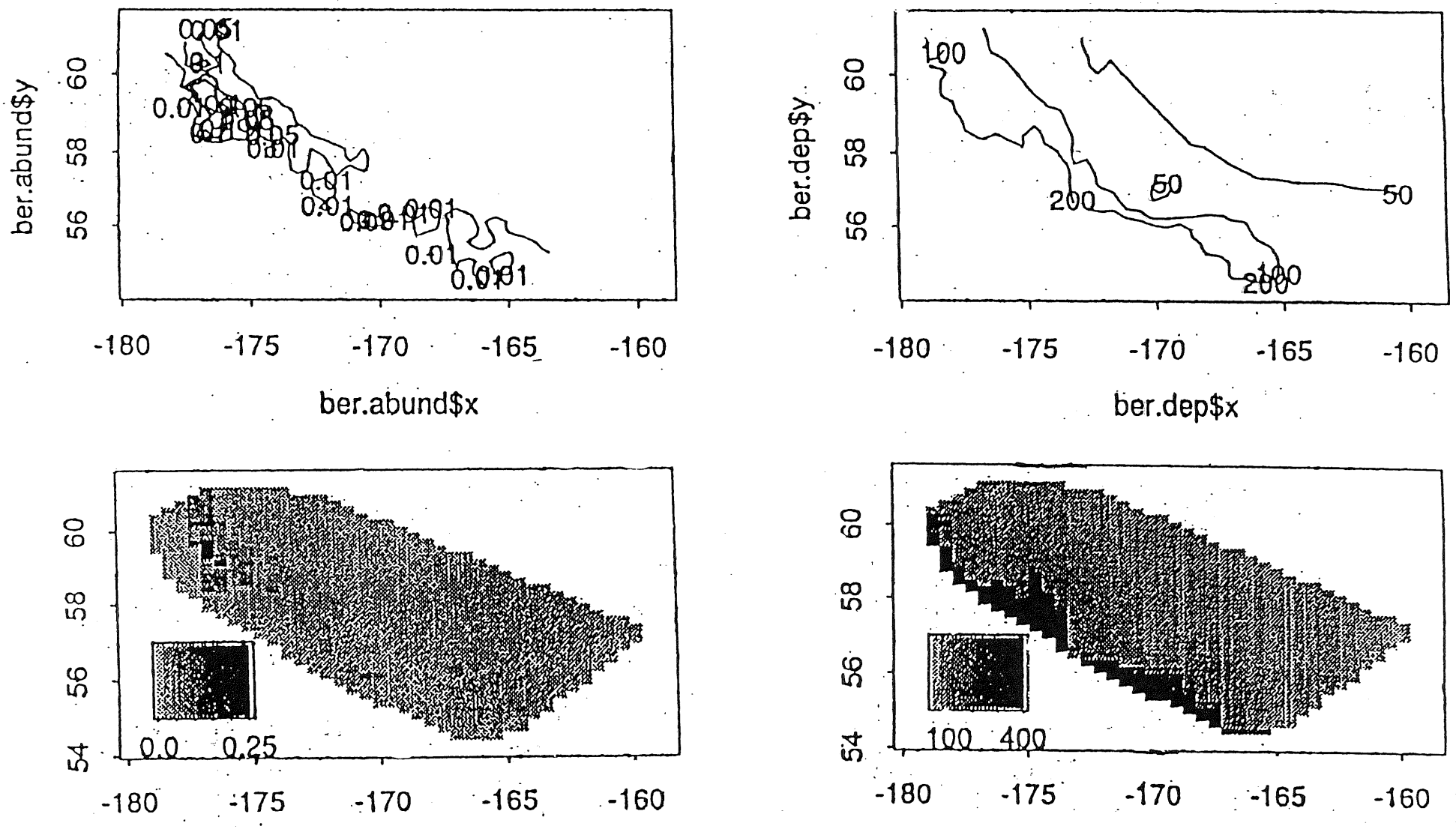


Figure 2.4 Gray scale and contour plots of acoustically determined fish density (left side) and depth (right side) for the Bering Sea acoustic survey test data.

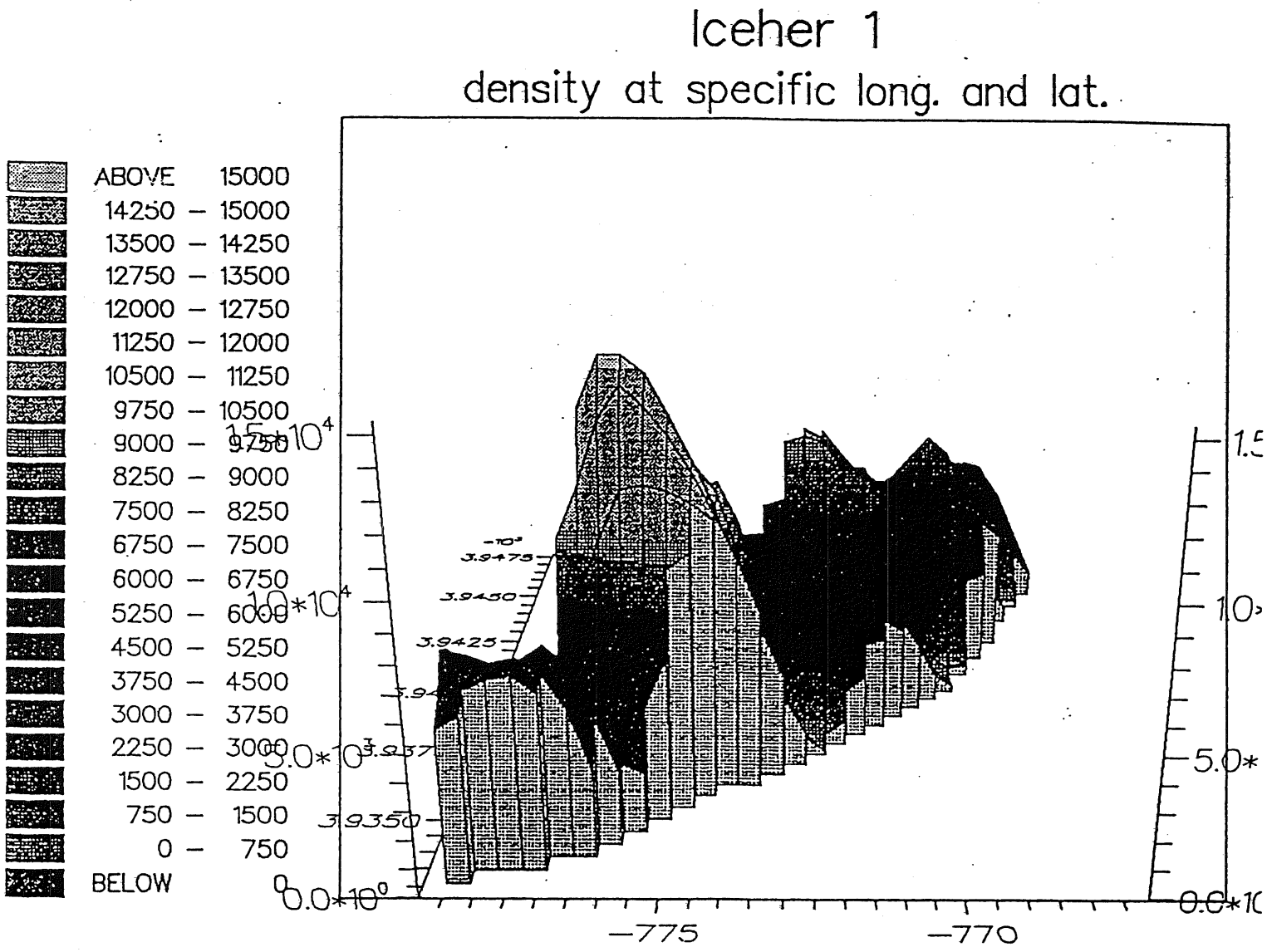


Figure 2.5 3-D plot of herring density interpolated from the Iceher1 acoustic survey data using a bilinear interpolation routine.

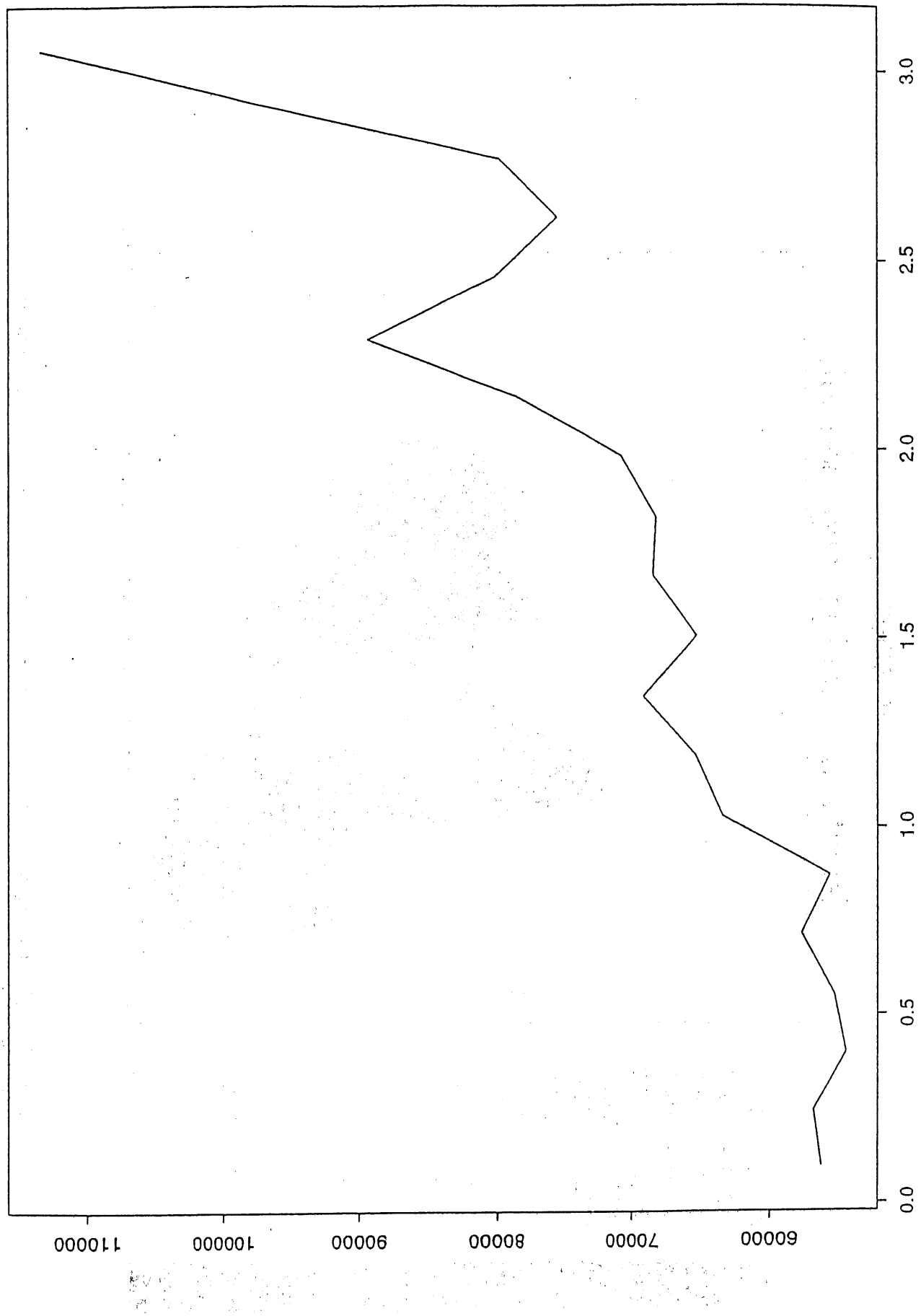


Figure 2.6 Variogram from the untransformed data in test data set 1.

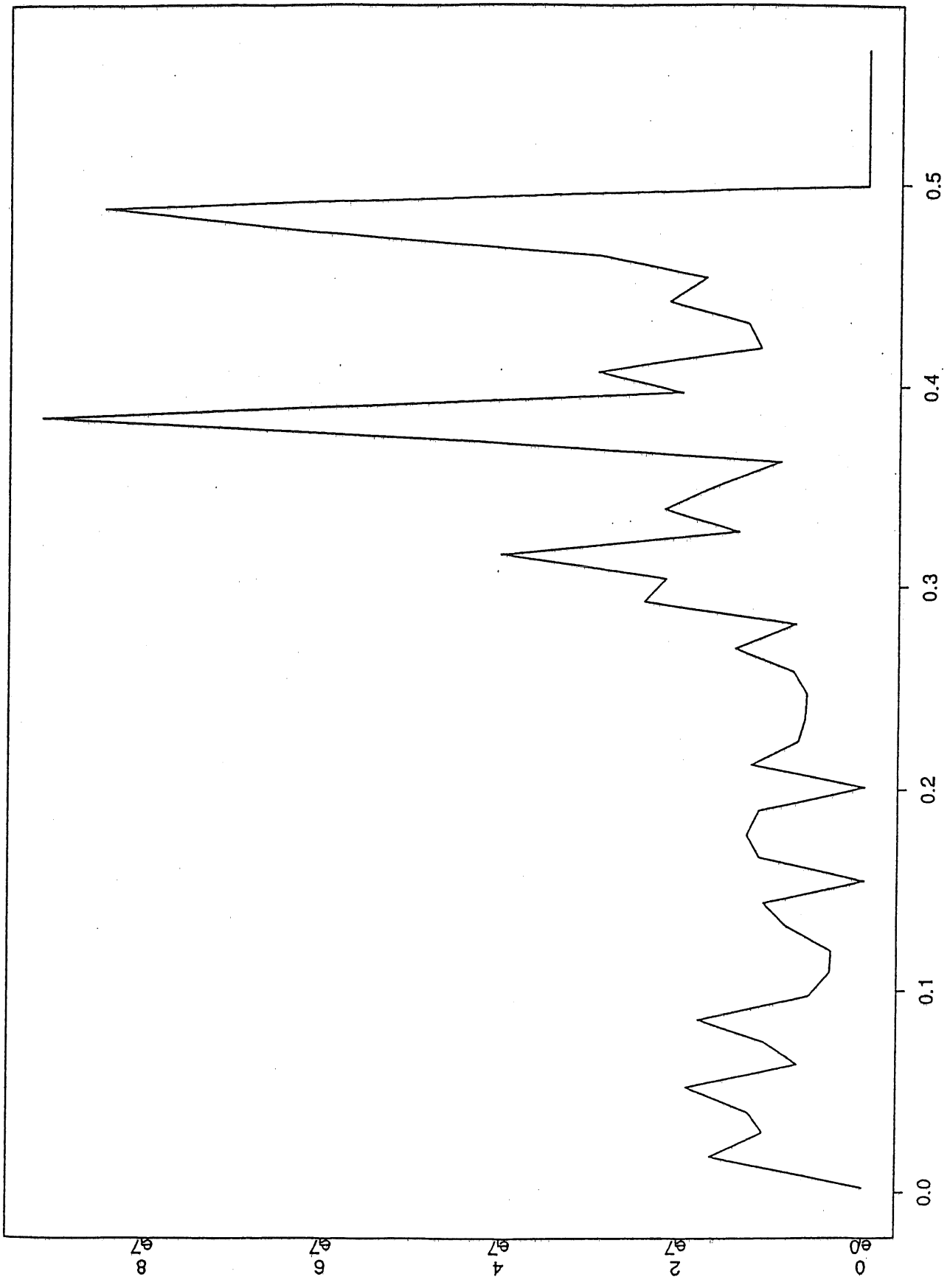


Figure 2.7 Variogram from the untransformed data in test data set 2.

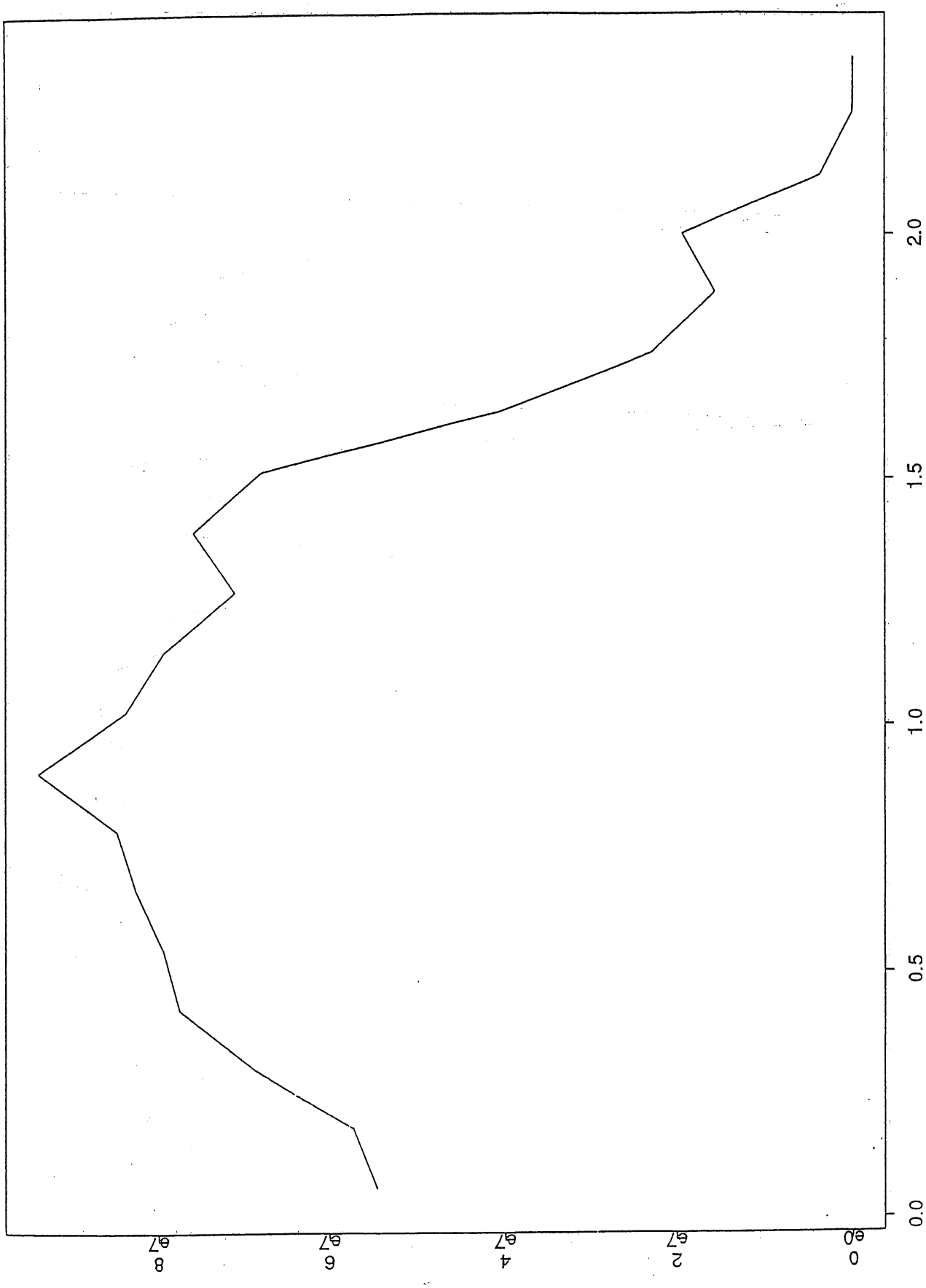


Figure 2.8 Variogram from the untransformed data in test data set 3.

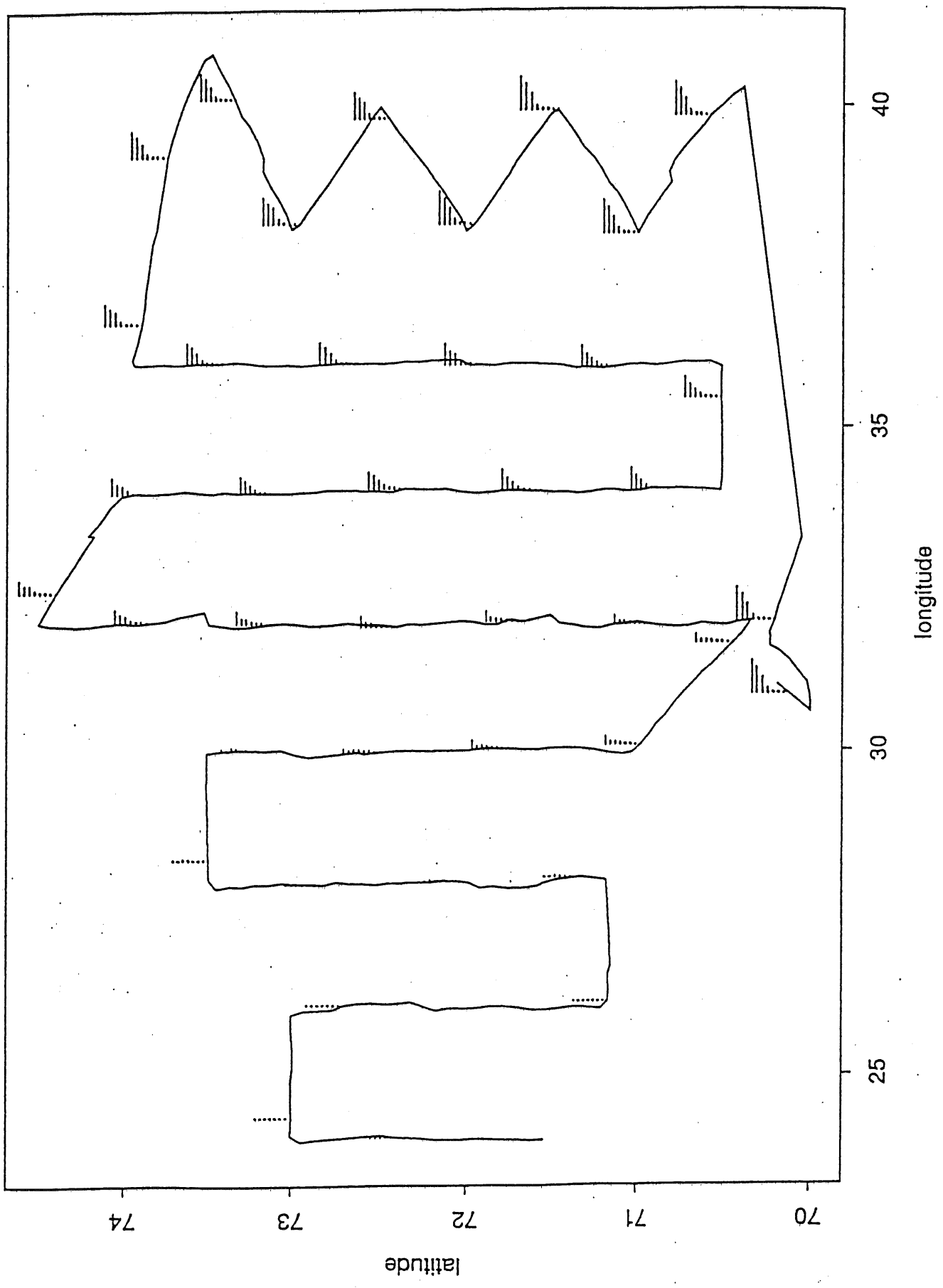


Figure 2.9 Display of the depth distribution of fish abundance along the survey track for an acoustic survey in the Barents Sea conducted by the Institute of Marine Research, Bergen. Depth distribution for abundance runs from top to bottom in 50 m depth increments. Profiles are drawn for 5 NM segments every 50 NM along the survey track.

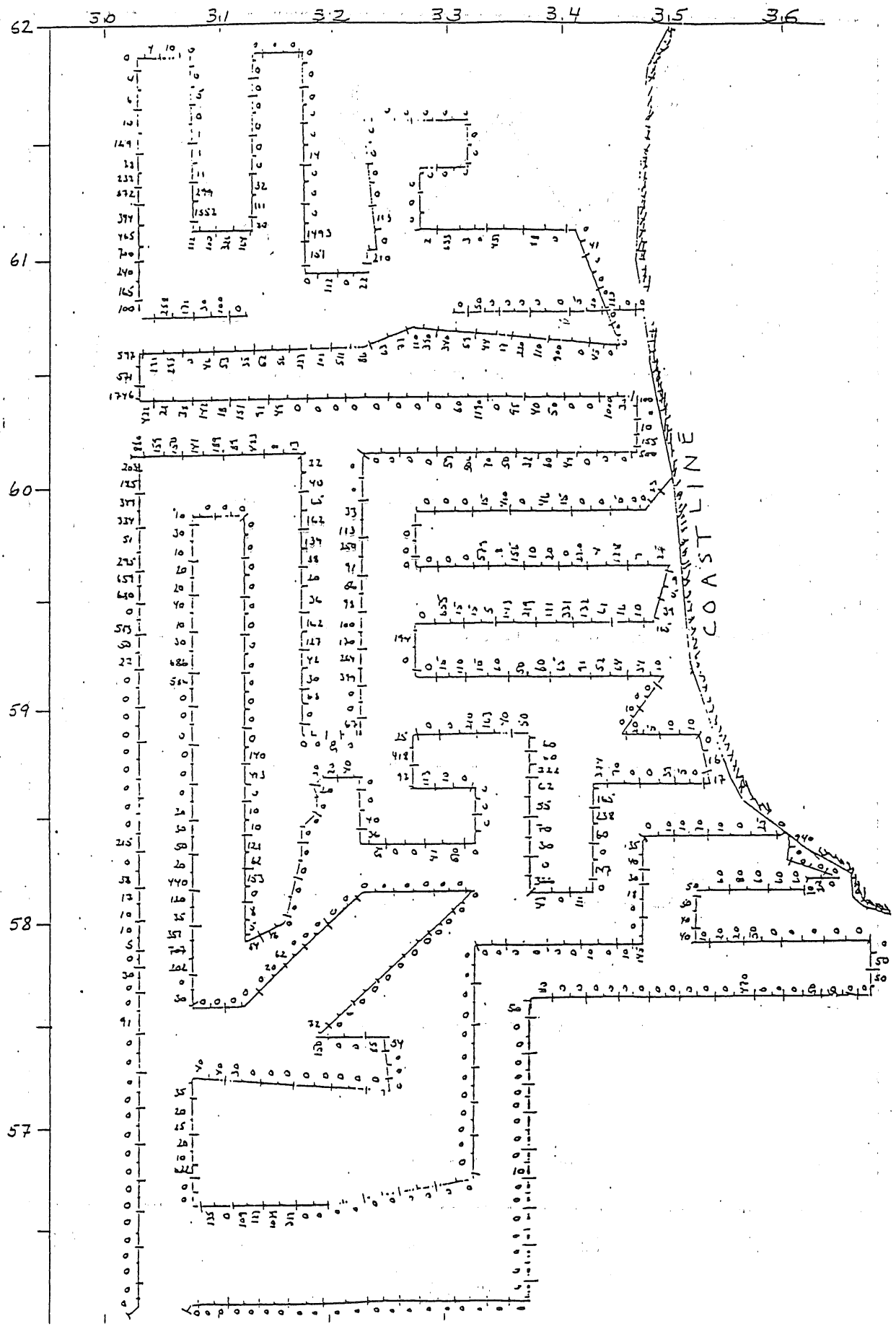


Figure 3.1 Acoustic density of fish for test data set 1.

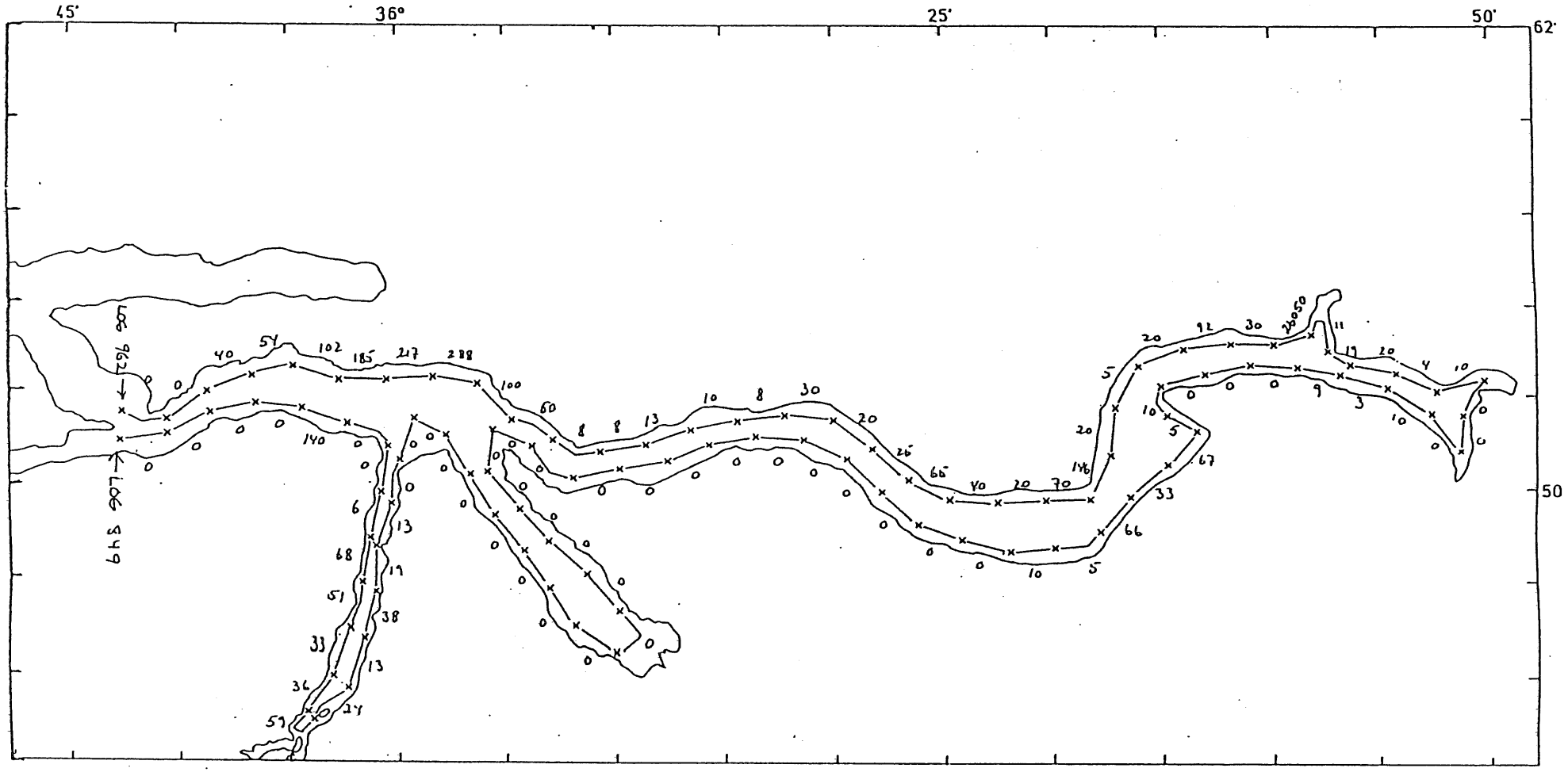


Figure 3.2 Acoustic denstiy of fish for test data set 2.

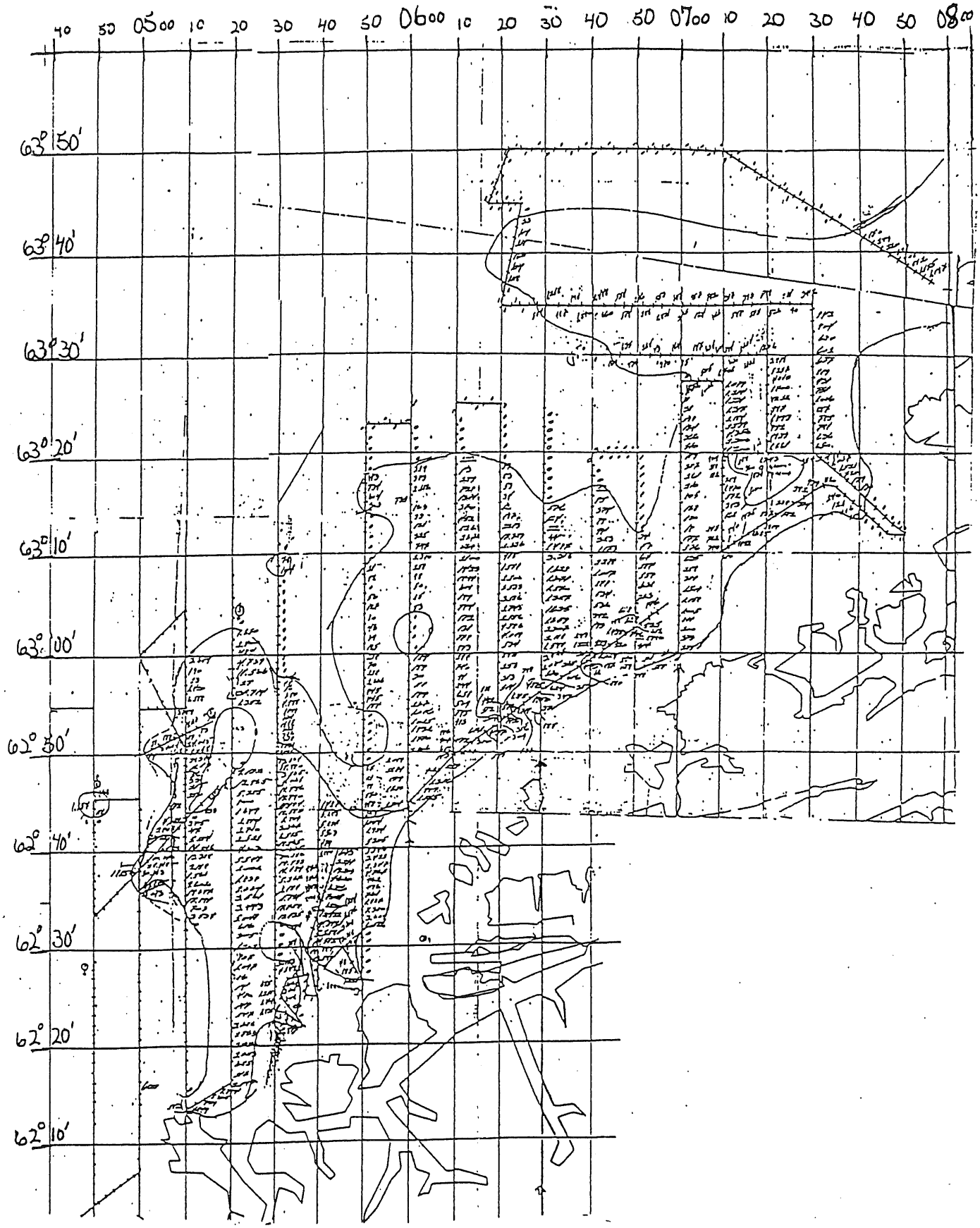


Figure 3.3 Acoustic denstiy of fish for test data set 3.

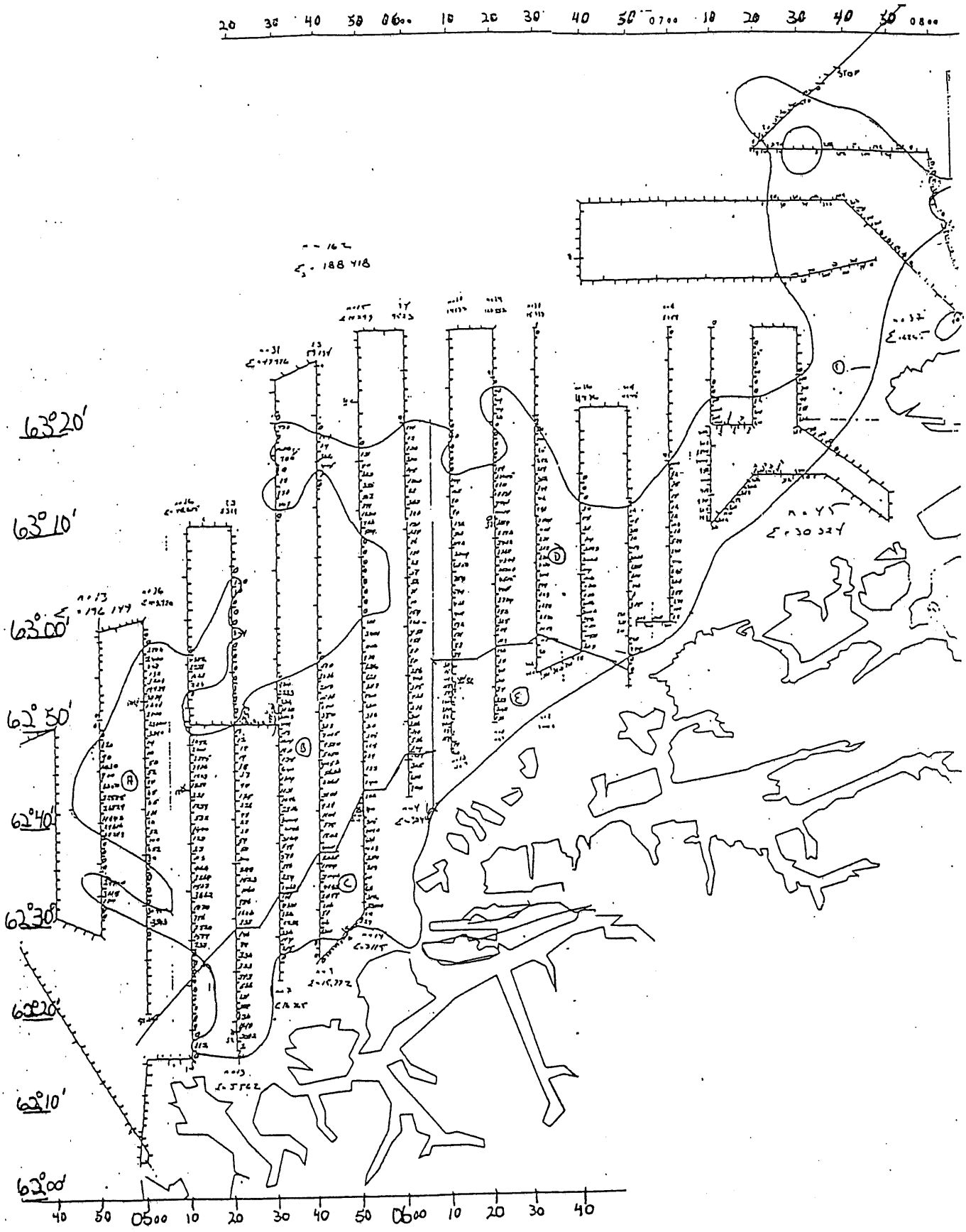


Figure 3.4 Acoustic denstiy of fish for test data set 4.

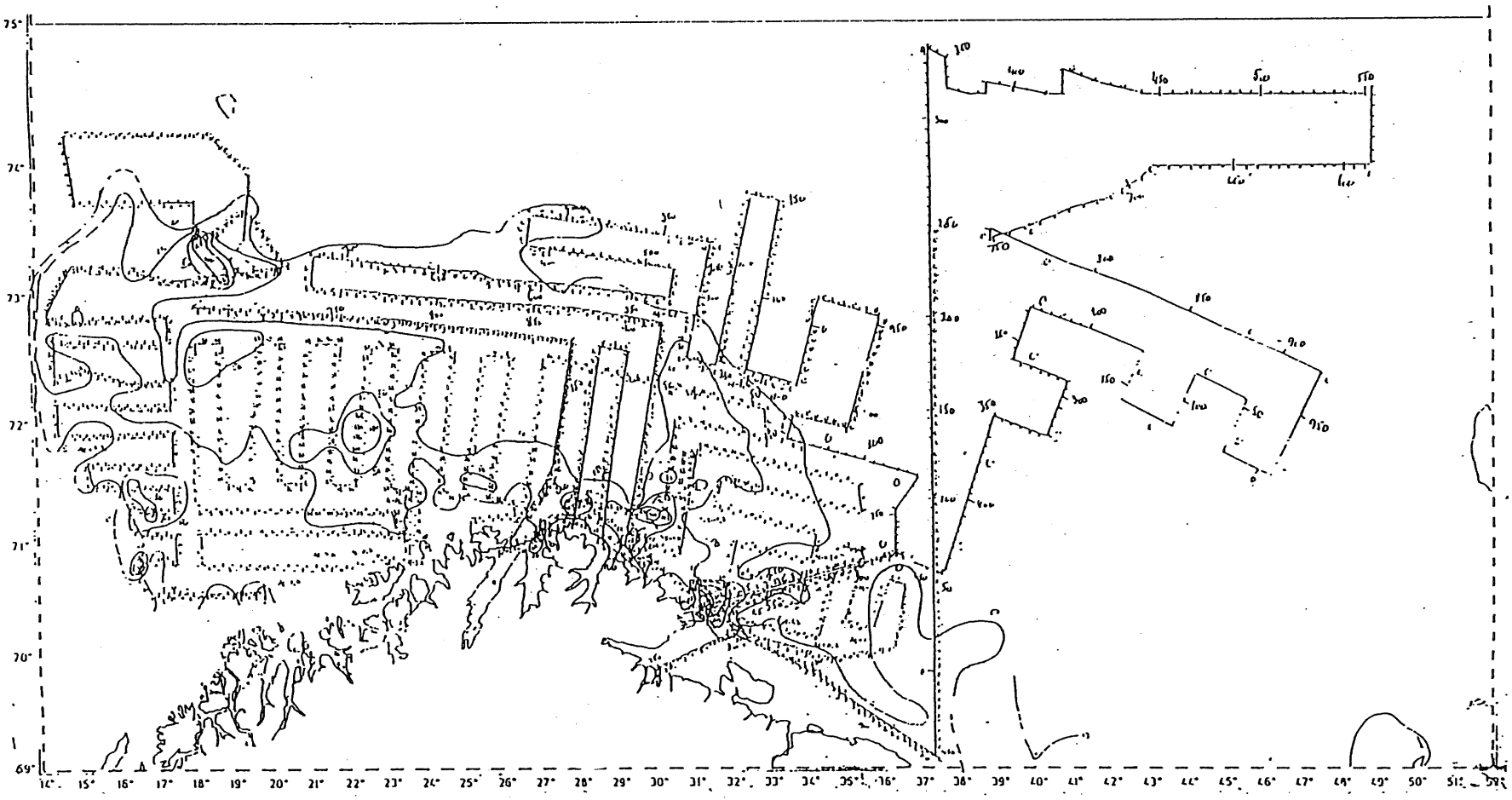


Figure 3.5 Acoustic denstiy of fish for test data set 5. The solid line indicates the survey grid for one ship; the faint or missing line, discerned from the rows of values, that for the second ship. The data are, however, to be considered together. Log numbers lack the thousands digit. Congestion in placement of numbers, especially near intersecting or coincident transects or in the vicinity of a third-ship track, shown in one or several places but not included in the data otherwise, is acknowledged. This map is presented only to give an overview of the survey area and grids for the tabulated data.

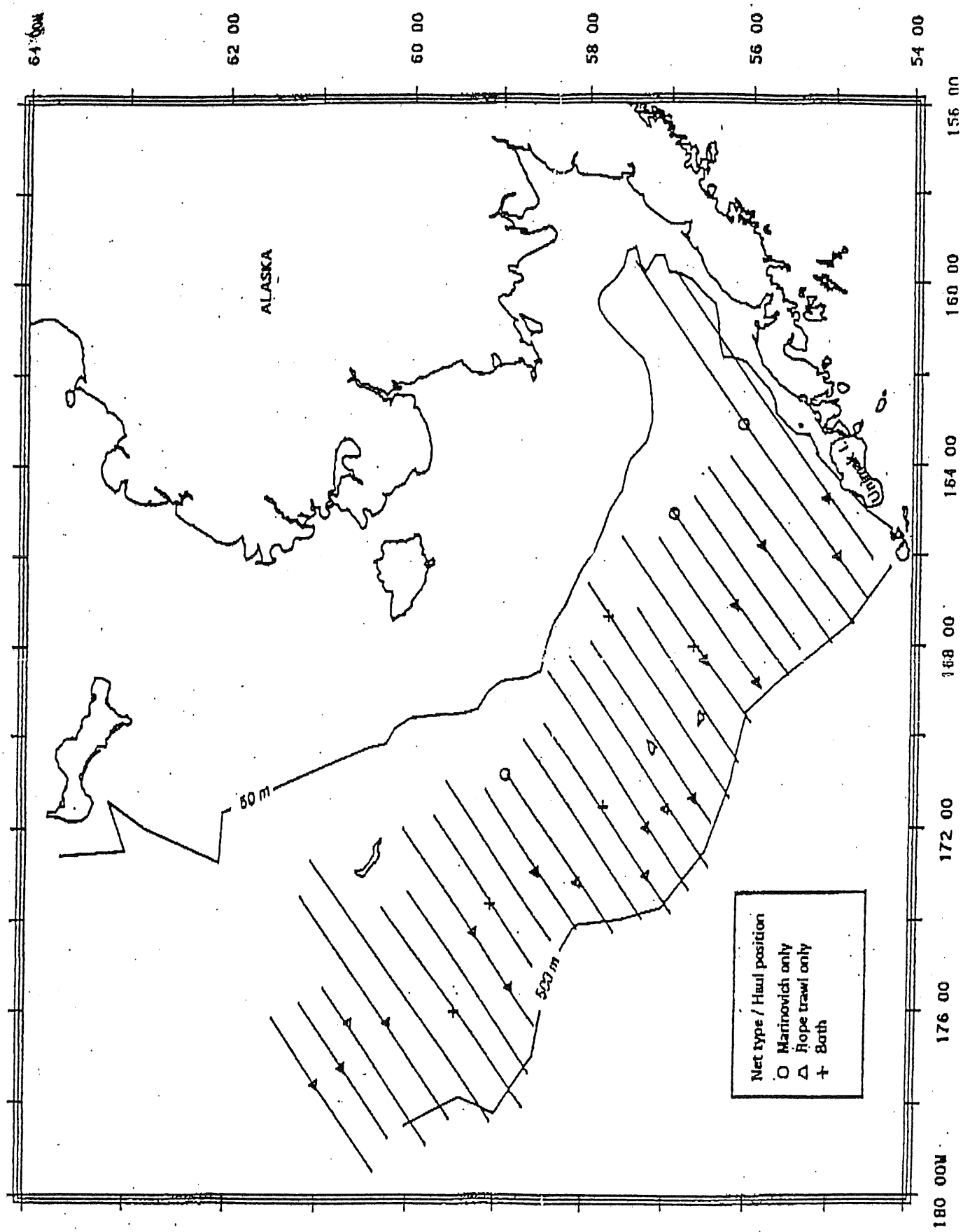


Figure 3.6 Transect lines surveyed during Summer 1988 echo integration/midwater trawl survey of adult walleye pollock on the eastern Bering Sea shelf and slope. Net type used at each haul position also indicated.

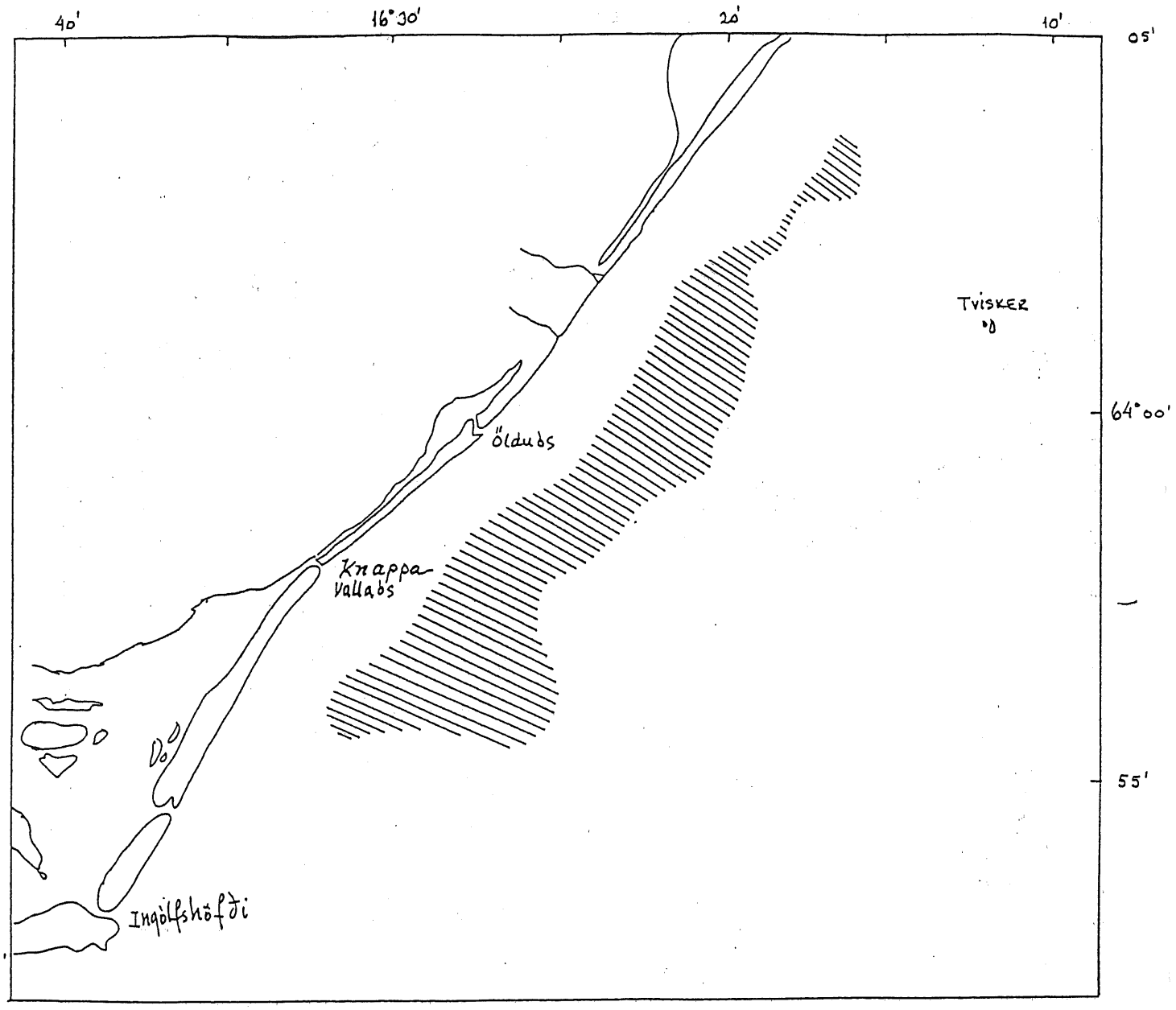


Figure 3.7 Location of Icelandic herring during three surveys in November 1988.

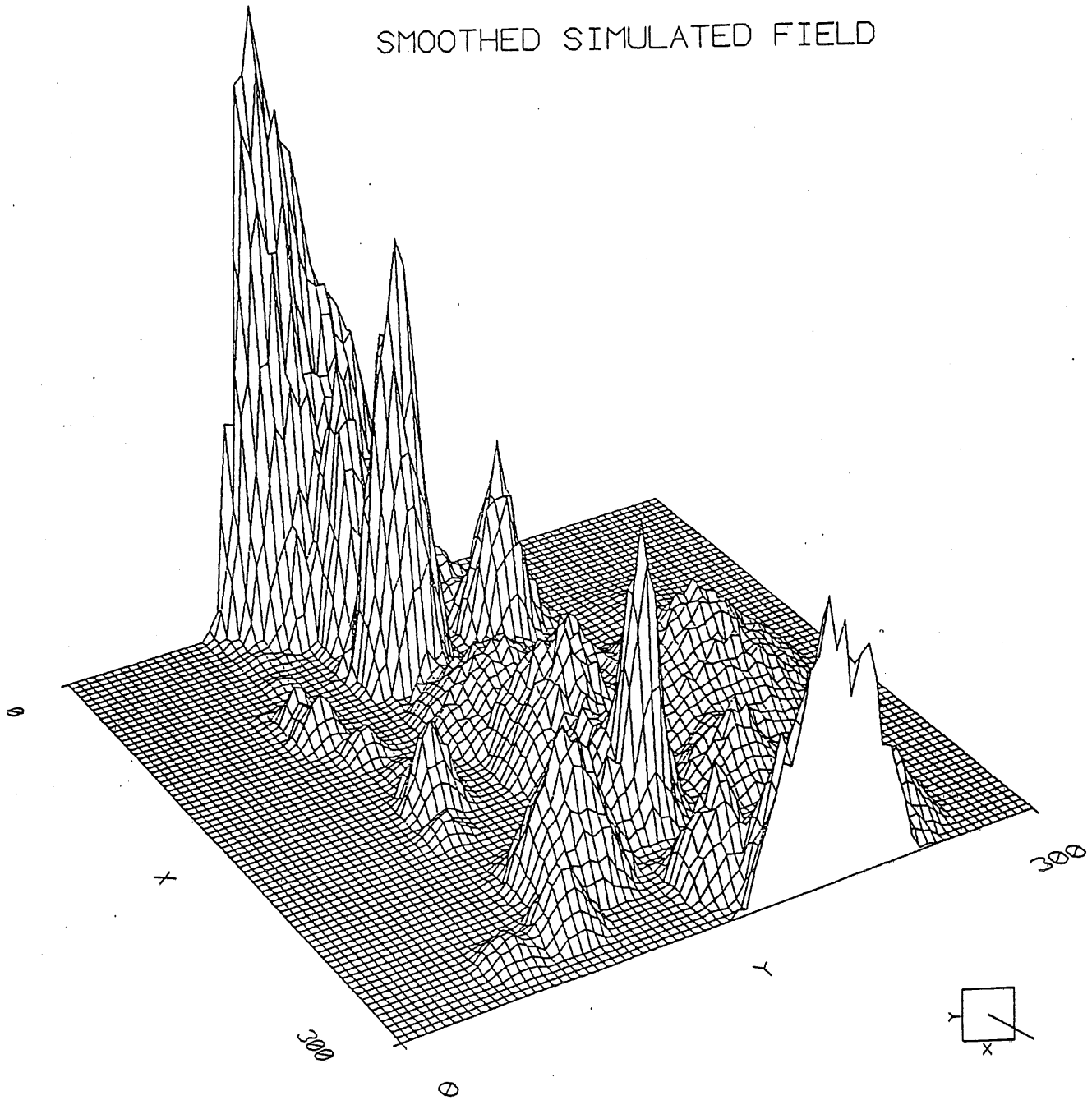


Figure 3.8 Simulated field used for data sets 10 - 15.

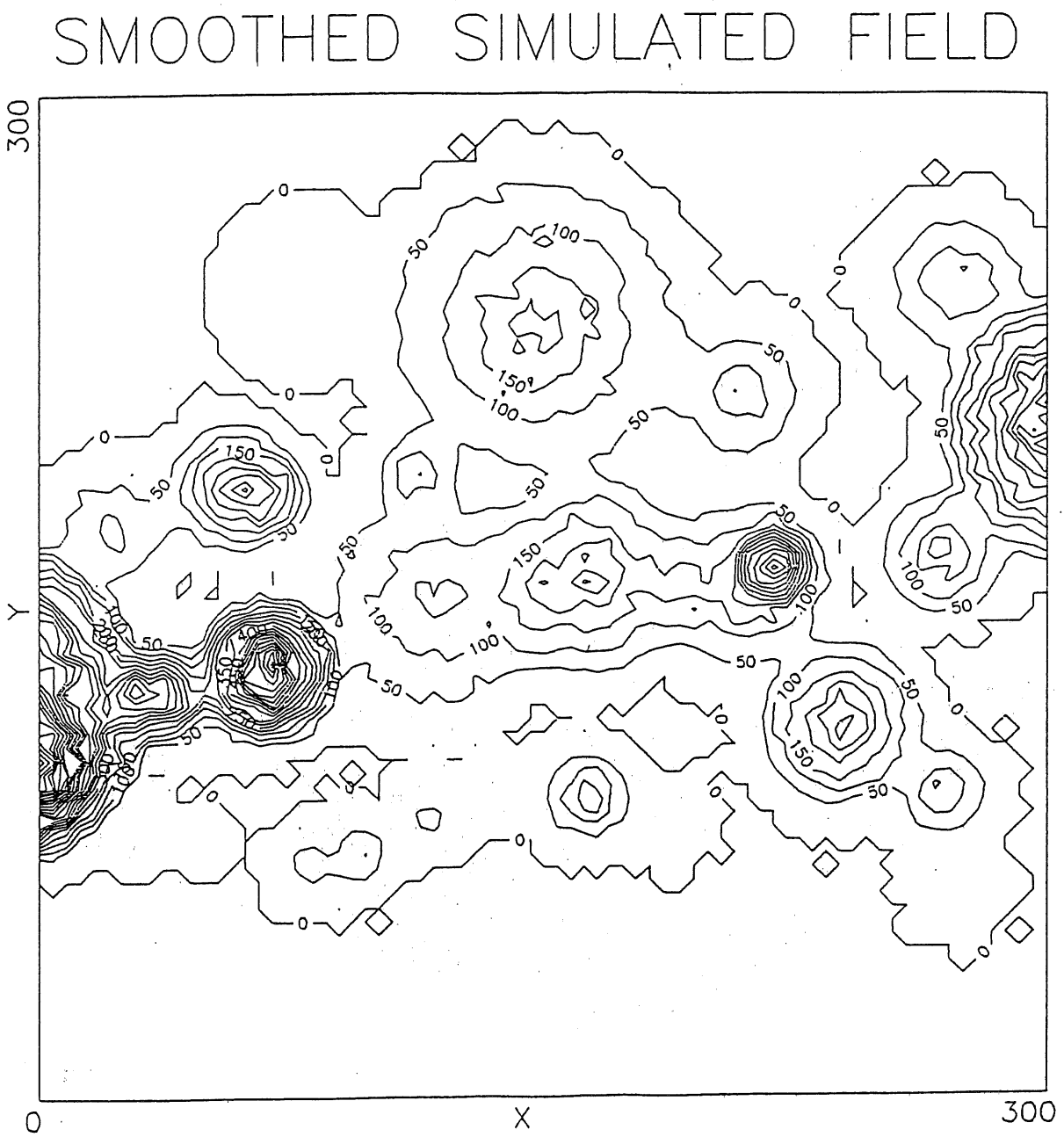


Figure 3.9 Contour map of simulated field used for data sets 10 - 15.

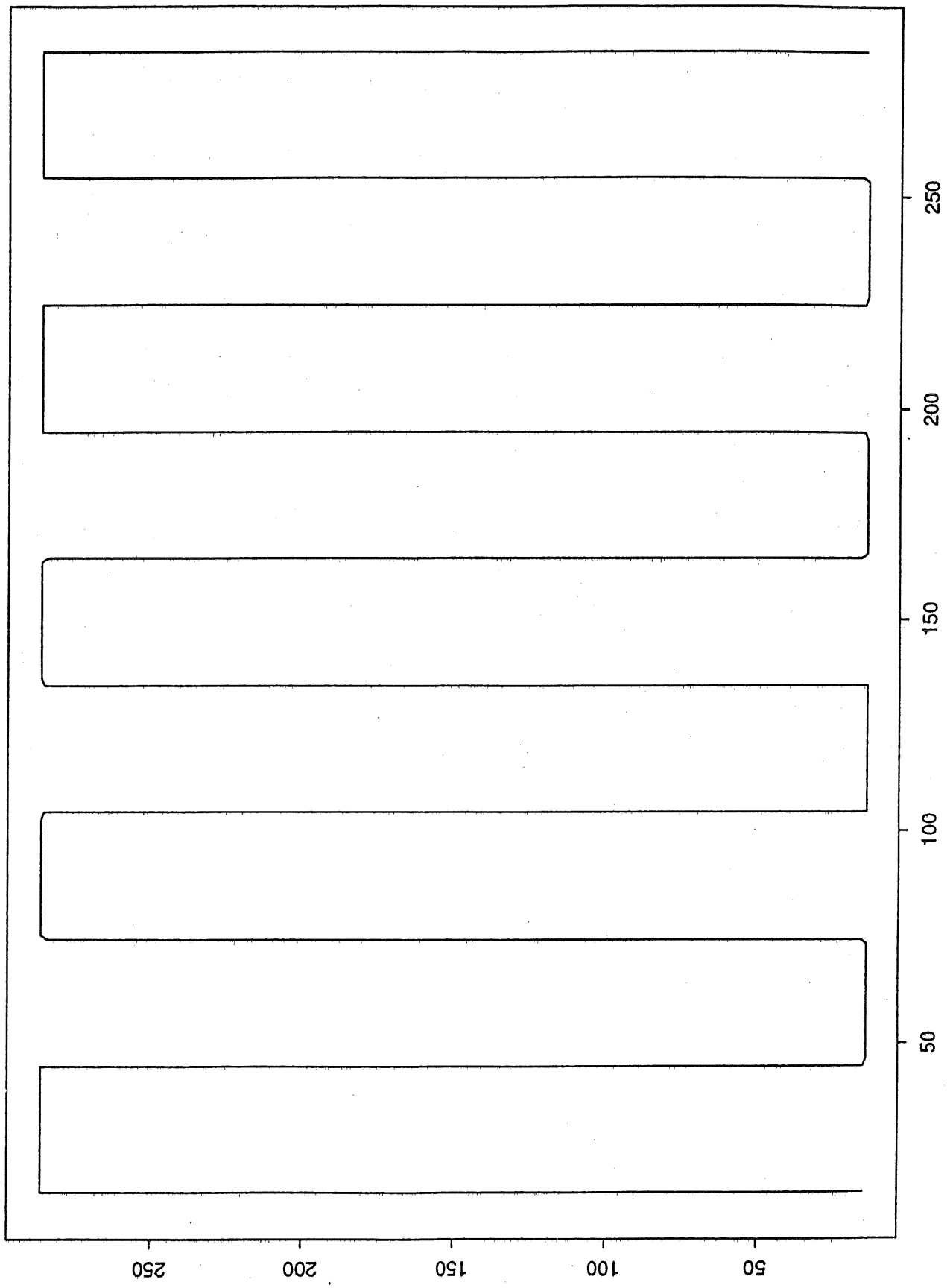


Figure 3.10 Simulated cruise tracks used as a basis for data sets 10 - 15.

Appendix A: Working papers and relevant documents available to the meeting

A.1 Working papers

W1: Kizner, Z.I. 1991. Simulating Data for Comparison of Methods of Spatial Statistics.

W2: Stefánsson, G. 1991. Analysis of Icelandic herring data using GLMs.

W3: Stolyarenko, D.A. 1991. Multidimensional Spline Approximation of Stock Density: Spline Survey Designer Software System.

W4: Swartzman, G. and Sullivan, P. 1991. Exploratory analysis of hydroacoustic fisheries survey data using statistical and graphical techniques.

W5: Wade, E. 1991. The Application of the Ordinary Kriging Package "Gulfkrig" for Mapping and Estimating Abundance of the Resource Surveyed by Acoustic Data Sets.

W6: Warren, G. W. 1991. Spatial Analysis of Acoustic Survey Data. I. Iceland Herring.

W7: Warren, G.W. 1991. Spatial Analysis of Acoustic Survey Data. II. Simulated Data Sets.

W8: Warren, G.W. 1991. Spatial Analysis of Acoustic Survey Data. III. Bering Sea Pollock.

W9: Petitgas, P. and Rivoirard, J. 1991. Global estimation: σ^2/n and the geostatistical estimation variance.

W10: Ferrandis, E. 1991. A note on the kriging weighting estimation.

A.2 Related documents, available to the meeting

Butterworth, D.S., Borchers, D.L. Miller, D.G.M. 1991. Some Comments on the Procedure for Testing Estimators of Krill Abundance which Utilise Survey Data.

Haslett, J., Bradley, R., Craig, P., Unwin, A. and Wills, G., 1991. Dynamic Graphics for Exploring Spatial Data, with Application to Locating Global and Local Anomalies. American Statistician.

Sullivan, P.J. 1991. Stock Abundance Estimation Using Depth-Dependent Trends and Spatially Correlated Variation.

Unwin, A., Will, G. and Haslett, J. 1991. Regard-Graphical Analysis of Regional Data.

Wills, G., Unwin, A. 1991. Kodiak Crabs - The View from Ireland.

Appendix B: Global estimation: σ^2/n and the geostatistical estimation variance

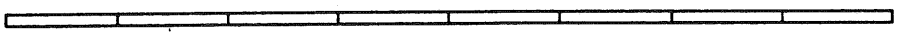
by
P. Petitgas and J. Rivoirard

Acoustics can provide a lot of data over a given domain. Here we will look at the estimation of the mean acoustic density over this domain, and in particular at the estimation variance.

First we will try to explain why this variance is not always σ^2/n. Then we will give the formula using the variogram. And after that we will consider different case studies: "iceher" (herring, S-E Iceland, three surveys), "test04" (West of Norway), "k0*a" (Walleye Pollock, Bering Sea).

1. WHY THE ESTIMATION VARIANCE IS NOT ALWAYS σ^2/n

- line divided into segments



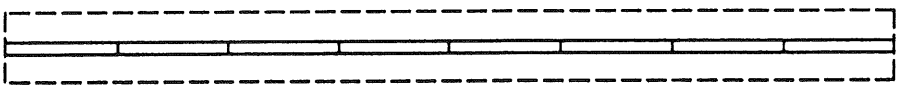
A line is divided into n segments l_i. We know the value Z_i of each of these segments. Their variance is σ^2.

Suppose we want to estimate the average value of the line L.

The estimate $Z_L^* = \frac{1}{n}\sum Z_i$ is equal to the exact value of the line $Z_L = \frac{1}{n}\sum Z_i$.

The estimation variance $Var(Z_L - Z_L^*)$ is zero, but is not σ^2/n either!

- thin block



The last example was trivial, but suppose we want to estimate the mean value over a thin block V set on the line L. We then would expect the estimate $Z_V^* = \frac{1}{n}\sum Z_i$ to be close to the real value Z_V, with an estimation variance still smaller than $\frac{\sigma^2}{n}$.

- large field

The exact mean value Z_V is now the average of many distant values.

If the data Z_i were independent, classical statistics would give $\frac{\sigma^2}{n}$ as estimation variance for Z_V.

In the case they are correlated, they count as if they were fewer but independent data. So the estimation variance will be larger than σ^2/n.

Summarising all these cases, we can see that the estimation variance is not always σ^2/n. It depends on the geometry of the field and of the data.

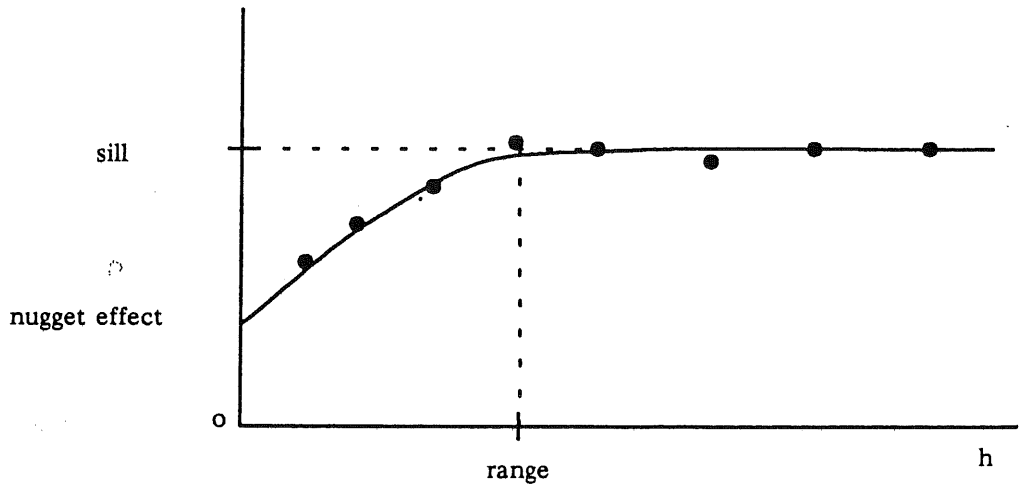
2. VARIOGRAM AND ESTIMATION VARIANCE

The variogram measures the mean variability between two points *x* and *x+h* as a function of their vectorial distance *h*:

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

The symbol *E* (expectation) denotes the average on all pairs *(x, x+h)*.

The variogram may reach a sill. In that case, there is a covariance function $C(h) = C(0) - \gamma(h)$ which represents the covariance between two values $Z(x)$ and $Z(x+h)$ distant of *h*. The covariance *C(0)* for *h=0* is the variance σ^2, and for *h* larger than the range, there is no more correlation between $Z(x)$ and $Z(x+h)$.



The variogram makes it possible to compute the variance when estimating the average value on V by the average of n samples Z_i : $Z_V^* = \frac{1}{n}\sum Z_i$. This can be written, in term of covariance:

$$\sigma_E^2 = C_{VV} + C_{ij} - 2C_{iV} \qquad (1)$$

C_{VV} is the mean covariance between two points describing V independently.
C_{iV} is the mean covariance between sample i and a point describing V.

C_{ij} is the mean covariance between samples *i* and *j*, for all n^2 possible pairs *(i,j)*: *n* pairs correspond to *i* with itself, the other n^2-n correspond to *i* different from *j*:

$$C_{ij} = \frac{C(0)}{n} + \underset{i \neq j}{C_{ij}}$$

If the field is large, compared to the range, the terms C_{VV} and C_{iV} are zero. The estimation variance σ_E^2 is reduced to the term C_{ij}, which is generally larger than $\frac{C(0)}{n} = \frac{\sigma^2}{n}$.

If the range is large compared to the field, we will see (on test04 and k0*a) that σ_E^2 can be less than σ^2/n.

3. ICEHER

The 3 surveys cover nearly the same zone (figures 1 to 4 in nautical miles). Excepting the zeroes at the North East, there are about 10 acoustic values per nm. The length of the first survey is smaller (15.5 nm, 174 values) than the two others (23 nm, 255 and 276 values). Large values are present in the first survey (max=26738), increasing its mean and variance.

Survey no	*m*	σ^2	$\frac{\sigma}{m}$	$\frac{\sigma}{m\sqrt{(n)}}$
1	5531	28 345 200	0.93	0.073
2	3253	14 120 100	1.33	0.072
3	3594	16 300 000	1.26	0.068

If the data of a given study were independent with the same law, the mean of this law would be estimated by the arithmetic average with a relative standard deviation of $\sigma/m\sqrt{(n)}$, here 7%.

In fact the acoustic data are regionalised and neighbouring data are correlated. Variograms computed at a 0.1 nm lag show structures up to 1.2 nm (figures 5–6–7). The structure is shorter for the first survey.

Knowing the variogram, we can compute the estimation variance of the mean value over the field. This supposes that the field has been delimited. Two hypotheses have been made.

– Either we limit approximatively the field to the zone which has been swept (for instance if we assume that the outside is close to zero).

– Or we extend the field on each side of the survey, admitting that the extension is not

systematically poorer than the survey.
The surfaces are respectively 15 and 33.5 nm^2.

To compute the estimation variance according to the formula (1), the field is discretised very finely. We obtain as relative standard deviation for the estimation $Z_V^* = \frac{1}{n}\sum Z_i$ from each survey:

about 12 % for the smaller field;
about 14 % for the larger one.

These two values are close, but both are larger than $\sigma/m\sqrt{(n)}$=7%. It is due to the fact that the field is large enough, and the data correlated: they count as fewer independent data.

Other approaches:

1) The data are not located regularly throughout the field. A weighted average, rather than the arithmetic one, may be used to estimated the field. Kriging corresponds to the optimal weighted average, the one which minimises the estimation variance. In our case kriging gives an increased weight to the data at the angles of broken lines. But it practically does not change the estimates and the variance (except for the estimation of the larger field from the third survey, where kriging gives about 3000 instead of 3600).

2) Tables exist, which give the estimation variance of a rectangle knowing its median line, with a spherical variogram model (Matheron 1971).
Let us take a rectangle close to the smaller field. If we replace the broken line survey by the median line (which is shorter and would contain less information), we get a relative standard deviation of about 14%.
If we unfold the broken line to be the median line of a larger but thinner rectangle, we get a deviation of 9–10%.
The reality lies between these two limits.

3) Data are correlated, which is one main reason for the estimation variance σ^2/n not to be correct. By averaging them over segments, we can build new values. Here we have regularised the data every 1 nm segment. These new values are less variable and little correlated to each other.

Survey no	n	$\frac{\sigma}{m}$	$\frac{\sigma}{m\sqrt{(n)}}$
1	17	0.45	0.11
2	24	0.73	0.15
3	24	0.76	0.15

The relative standard deviation is smaller for the first survey. This comes from the fact that the variability is shorter scaled, and has been destroyed more by the regularisation. The value of $\sigma/m\sqrt{(n)}$ is then 11% for the first survey, and 15% for the two others.

4. SURVEYS MADE OF PARALLEL REGULARLY SPACED TRANSECTS

For such survey design we suggest a simple method to calculate the variance of the estimation. We shall see that it can be calculated on the transect cumulated data using geostatistics but not using the variance $\frac{\sigma^2}{n}$.

Each echo-integrated value is the exact mean value on each ESDU segment of the acoustic fish density. The variable q(j) defined by cumulating the data Z(i,j) along each transect j represents the acoustic fish quantity along each transect j:

$$q(j) = \sum_{i=1}^{n(j)} a\ Z(i,j)$$

where i is the indice of the acoustic densities along the transects and j is the indice of the transects; and where a is the ESDU distance.

Of course, the transect should sample the limits of the fish regionalisation, i.e. should reach the bordering zeroes at both extremeties.

The cumulation transforms a 2D regionalisation into a 1D one. Obviously the 2 are related. These relations are communly used in stereology and geostatistics when the transect lengths are equal. The cumulation has 3 effects on the variogram. The sill (variance) is lowered, the nugget effect is filtered, the correlations are smoothed. Even though the transects are of different lengths the 1D data set is expected to be less rough and more regular than the 2D one.

In 1D, the estimation problem becomes the following. We want to estimate the mean acoustic quantity on a segment L when we know experimental values q(j) regularly spaced along L. The values q(j) may be regarded as punctual values because the width of the echo surveying cone is very small in comparison to the inter-transect distance. Let us call D the inter-transect distance. It is the distance between 2 successive q(j) values. We have: $L = n_q\ D$ where n_q is the number of q(j) values, i.e. the number of transects.

The estimate of the mean transect acoustic quantity is: $q_L^* = \frac{1}{n_q} \sum_j q(j)$

The estimation variance writes after equation (1) in 1D as follows:

$$\sigma_E^2 = C_{LL} + C_{jk} - 2C_{jL} \qquad (2)$$

where C(h) is the 1D covariance model of the q(j) values.

When the inter-transect distance is smaller than the range of the spatial correlations, Matheron (1965,1971) has given theoretical prove for approximating σ_E^2 of equation (2): the errors of estimation in each segment D can be considered as uncorrelated. The variance σ_E^2 then writes:

$$\sigma_E^2 \approx \frac{1}{n_q}\sigma_{elem}^2 \qquad (3)$$

σ_{elem}^2 is the variance of estimation when the segment D is estimated by the value of its central point. Equation (2) rewrites:

$$\sigma_{elem}^2 = C(0) + C_{DD} - 2C_{jD}$$

As $C(0) = \sigma^2$ (variance of q(j) values) we can write the variance of estimation σ_E^2 in the following way:

$$\sigma_E^2 = \frac{\sigma^2}{n_q} - \frac{1}{n_q}(2C_{jD} - C_{DD})$$

The mean covariance C_{DD} involves distances larger than the mean covariance C_{jD} because the point j is at the center of the segment D. So we have the inequality: $C_{DD} < C_{jD}$. Thus we expect σ_E^2 to be smaller than $\frac{\sigma^2}{n_q}$.

We did the previous calculations on 2 data sets, the one named test04 concerning herring off shore Norway and the one named K0*a concerning walleye pollock in the Berring Sea. In both cases the q(j) values are very regular and 2 values h apart stay correlated for distances h up to half of the total length L. In such situation the range of the correlations is large in comparison to the field over which the mean is estimated. The q(j) values cannot be considered as observations sampled out of an infinite field. The parameter $\frac{\sigma^2}{n}$ will over-estimate the variance of estimation.

4.1 Herring off Norway, data set TEST04 (K.Foote)

The survey design with a proportional representation of the data is given on figure 8. We shall focus only on the regular part of the survey. The Northeastern irregular part represents only 3% of the arithmetical mean of the total data set.

In order to calculate distances the longitudes and latitudes have been transformed followingly. Let y and b be the latitude and the longitude expressed in minutes and decimal fractions of minutes and let *lat* be the mean of y over the surveyed field. The transformed longitude is: $x = b \cos(lat)$. The distances are expressed in nautical miles (n.m.).

The values are cumulated along the parallel transects. We have 15 non zero q(j) data. A representation of the q(j) values is given on figure 9. We have:

$$D = 4.54 n.m. \; ; \; n_q = 15 \; ; \; L = 68.1 n.m.$$

$$q_L^* = \frac{1}{15}\sum_j q(j) = 49006. \; ; \; S_q^2 = \frac{1}{14}\sum_j (q(j) - q_L^*)^2 = 3.11 \; 10^9$$

The 1D variogram of the q(j) values is given on figure 10. No nugget effect has been modelled. The variogram model is a sum of a spherical and a linear variogram. The variabbility between 2 values h apart is lower than S_q^2 untill h is 30 n.m.. The geostatistical estimation variance σ_E^2 is caculated using formula (3). We get:

$$\frac{\sigma_E}{q_L^*} = 12.1\% \quad \text{and} \quad \frac{S_q}{q_L^* \sqrt{n_q}} = 29.3\%$$

4.2. Walleye pollock of the Berring Sea, data set K0*a (N. Williamson)

The survey is made of n_q=27 parallel transects oriented approximatively NE–SW. The coordinnates are transformed as previously (here *lat=58°).* The mean inter–transect distance is: $D = 20 \; n.m.$. Along the transects the echo–integrated data Z(i,j) are expressed in kilograms of fish per meters. The q(j) transect cumulated values derived are expressed in: $\frac{kg}{m} \; n.m.$.

The survey design with a proportional representation of the data is given on figure 11 and a representation of the q(j) values is given on figure 12. We have :

$$q_L^* = 1.983 \; ; \quad S_q^2 = \frac{1}{n_q - 1}\sum_j (q(j) - q_L^*)^2 = 2.257 \; ; \; L = 540 n.m.$$

The 9 Northwestern transects , i.e. the 9 Northwestern q(j) values which show the greater agregation of fish represent 63 % of q_L^*. Correcting the unities the estimated total quantity of pollock is: $Q = q_L^* \; D \; n_q = 3.7 \; 10^6 \; tons$

The 1D variogram of the q(j) values is given on figure 13. No nugget effect has been modelled. The variogram model is spherical. The q(j) values are very well correlated. The range of the spherical variogram is *340 n.m.* which is the equivalent of 17 inter–transect distances. The geostatistical estimation variance σ_E^2 is caculated using formula (3). We get:

$$\frac{\sigma_E}{q_L^*} = 2.3\% \quad \text{and} \quad \frac{S_q}{q_L^* \sqrt{n_q}} = 14.6\%$$

REFERENCES :

Matheron G., 1965, *Les variables regionalisees et leur estimation*, Ed. Masson, Paris

Matheron G., 1971, *The theory of regionalized variables and their application*, Les cahiers du Centre de Morphologie Mathematique, Fascicule 5, Ecole des Mines de Paris, Bibliotheque de Fontainebleau. (english version).

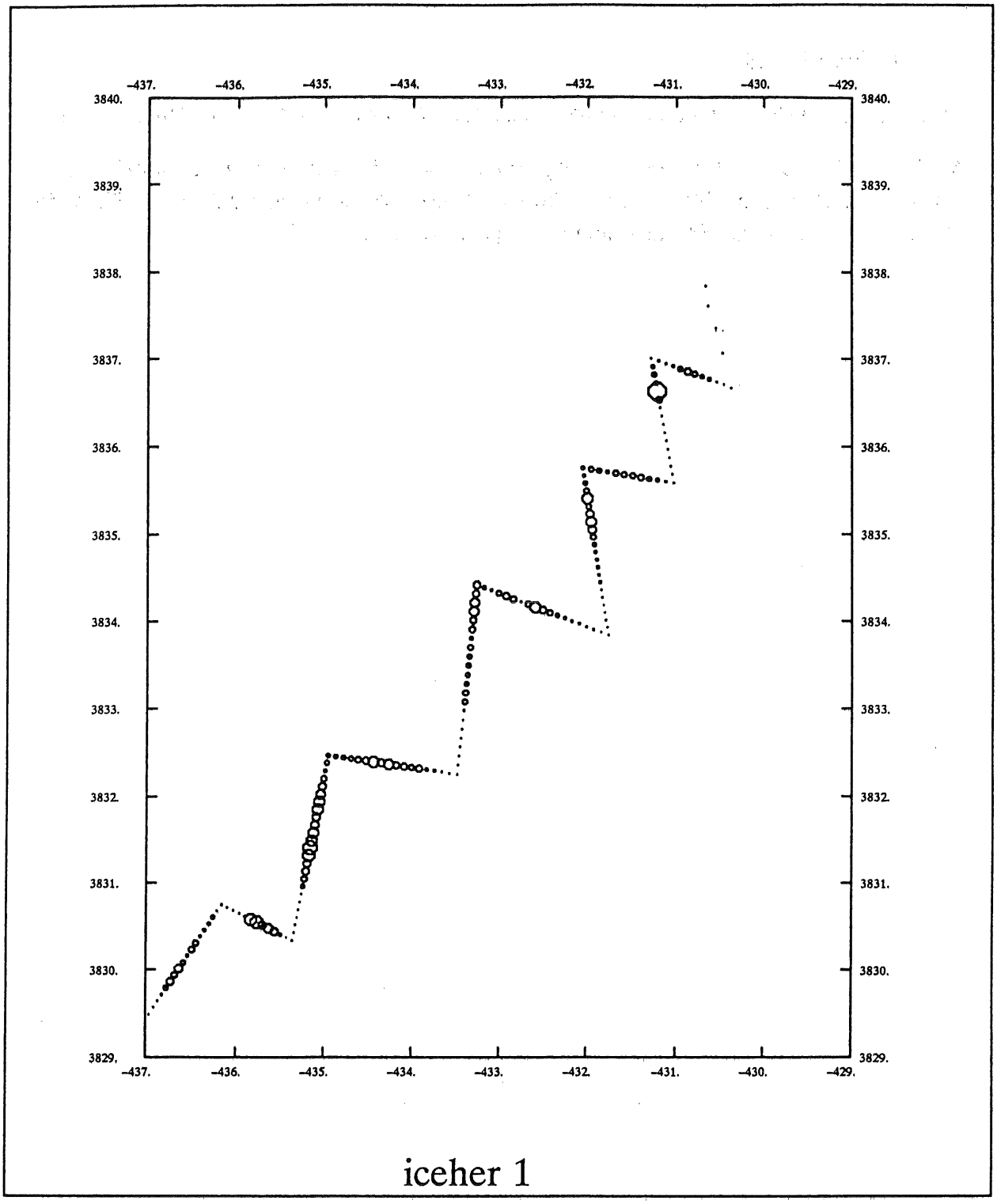

-437.
-436.
-435.
-434.
-433.
-432.
-431.
-430.
-429.
3840.
3839.
3838.
3837.
3836.
3835.
3834.
3833.
3832.
3831.
3830.
3829.
iceher 1

Figure 2.

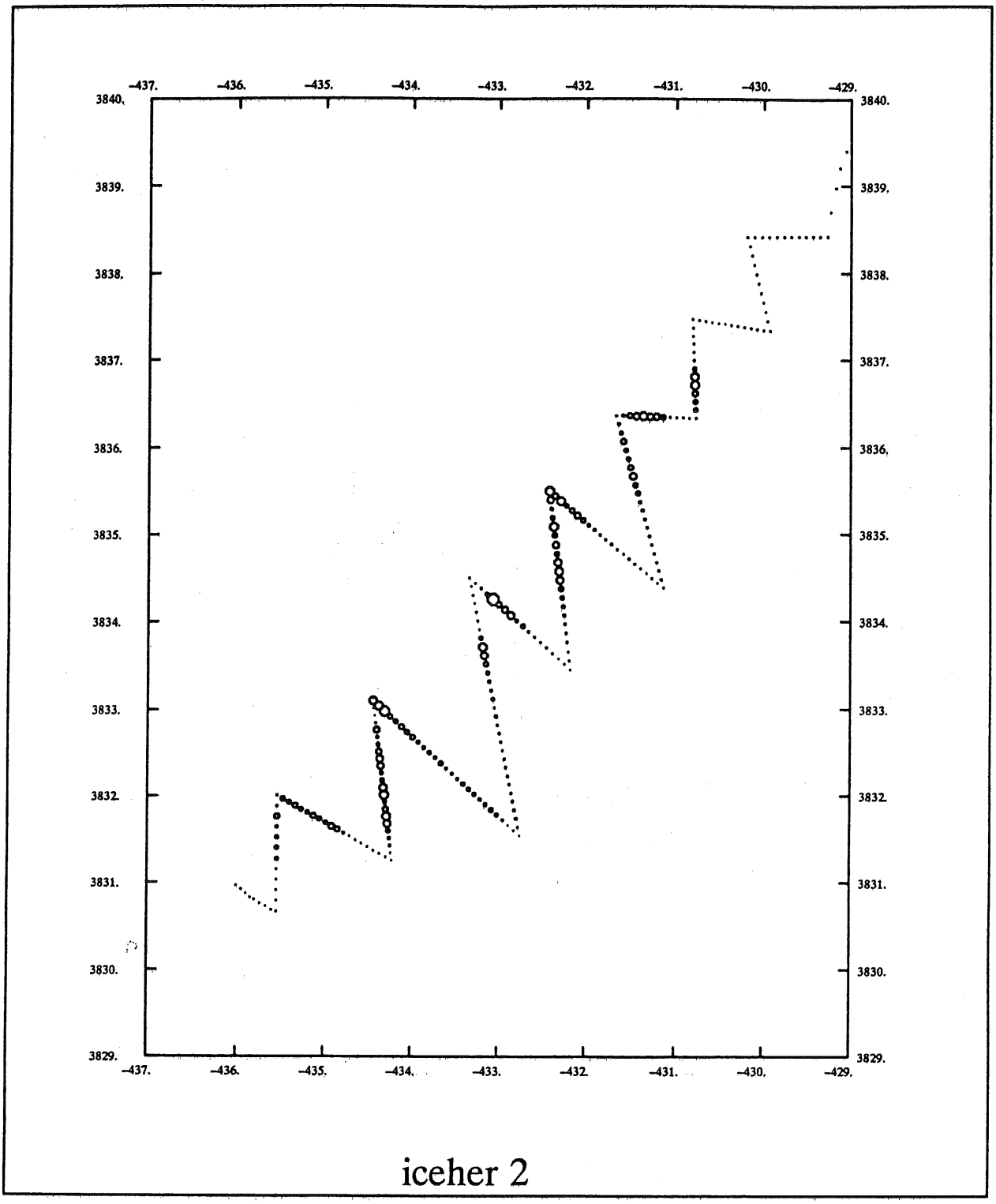

-437.
-436.
-435.
-434.
-433.
-432.
-431.
-430.
-429.
3840.
3839.
3838.
3837.
3836.
3835.
3834.
3833.
3832.
3831.
3830.
3829.
iceher 2

Figure 3.

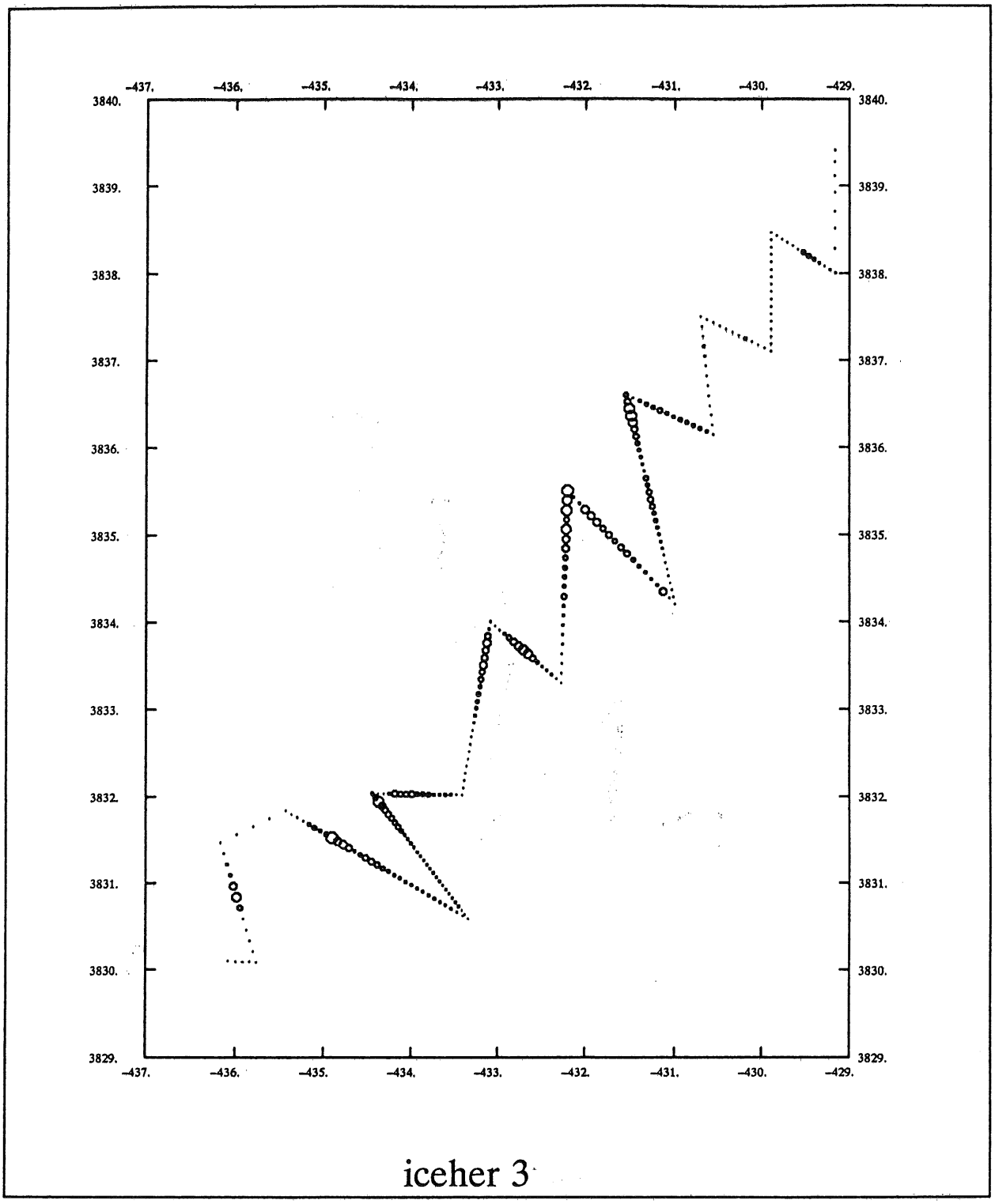
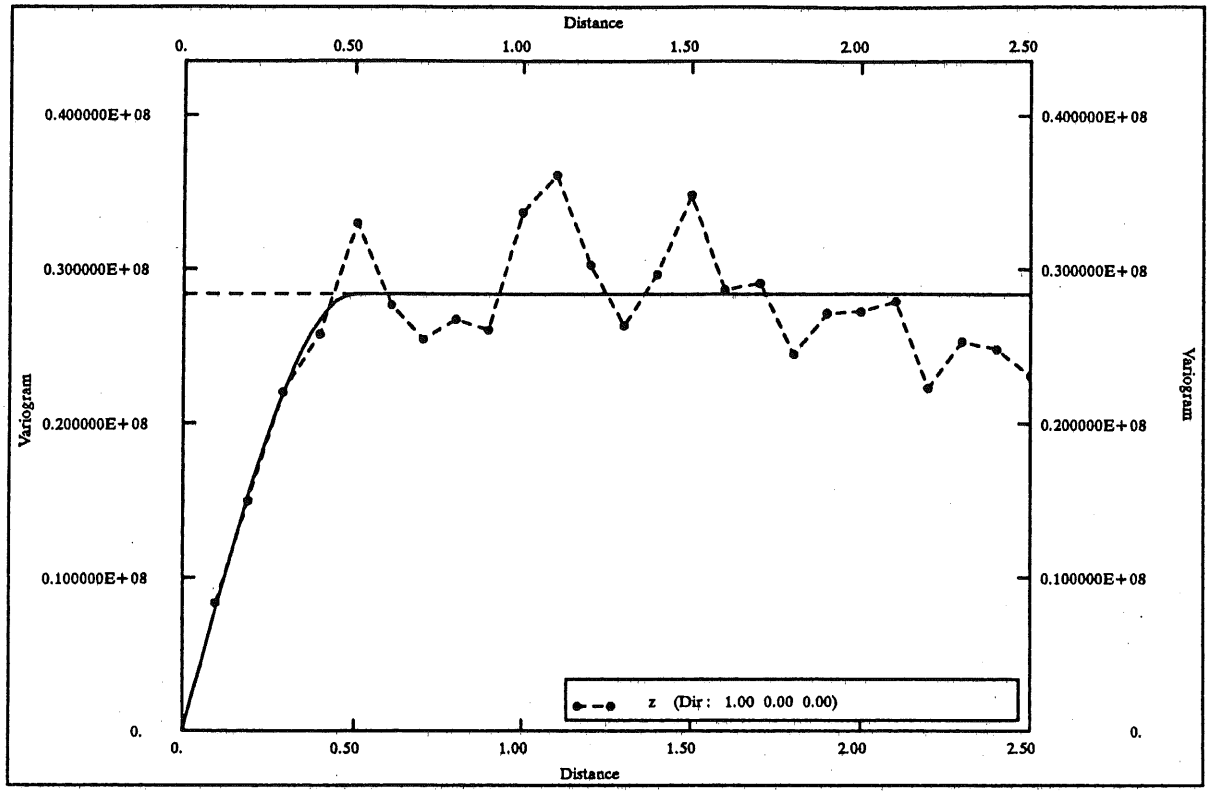

-437.
-436.
-435.
-434.
-433.
-432.
-431.
-430.
-429.
3840.
3839.
3838.
3837.
3836.
3835.
3834.
3833.
3832.
3831.
3830.
3829.
iceher 3

Figure 4.

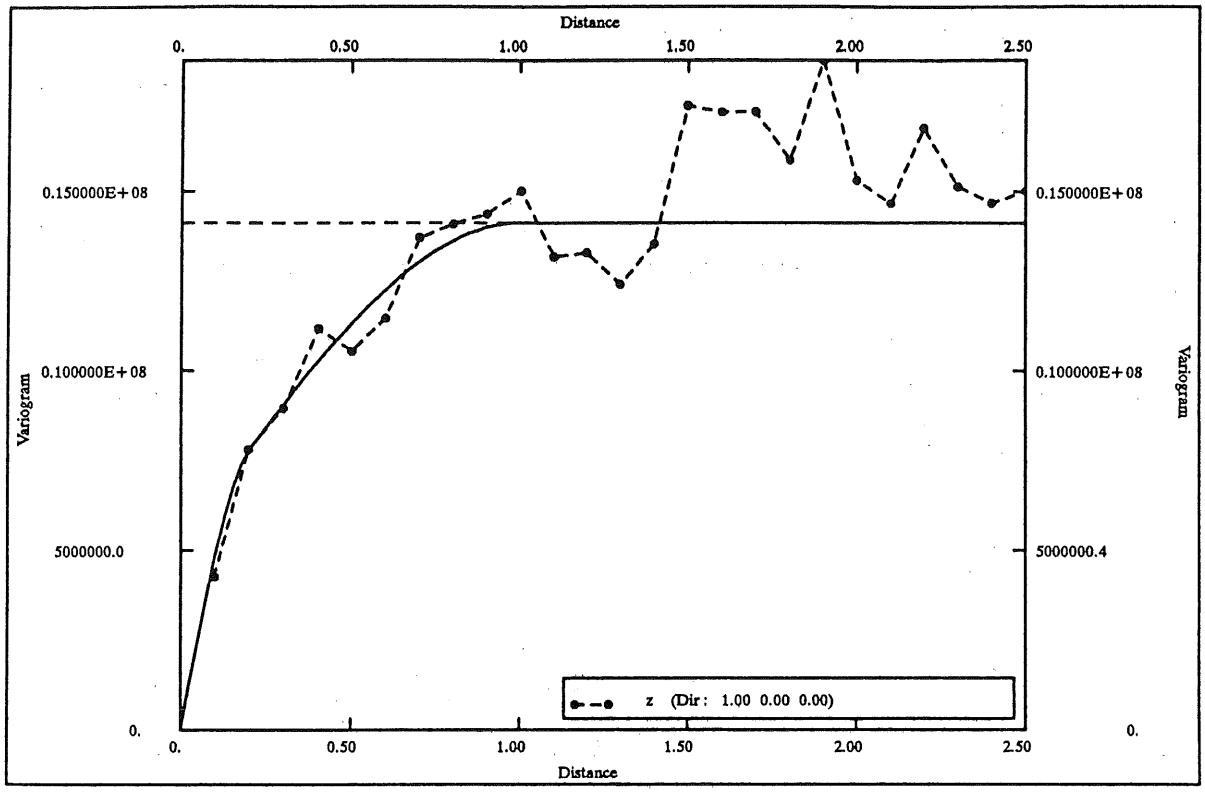


iceher 1

NUGGET: 0.00 NB. STRUCTURE(S):1 ANISOTROPY

STRUCTURE	SILL	RANGE	LAMBDA	X	Y	Z
spheric.	0.2835E+08	0.5000		1.000	1.000	1.000

Figure 5.

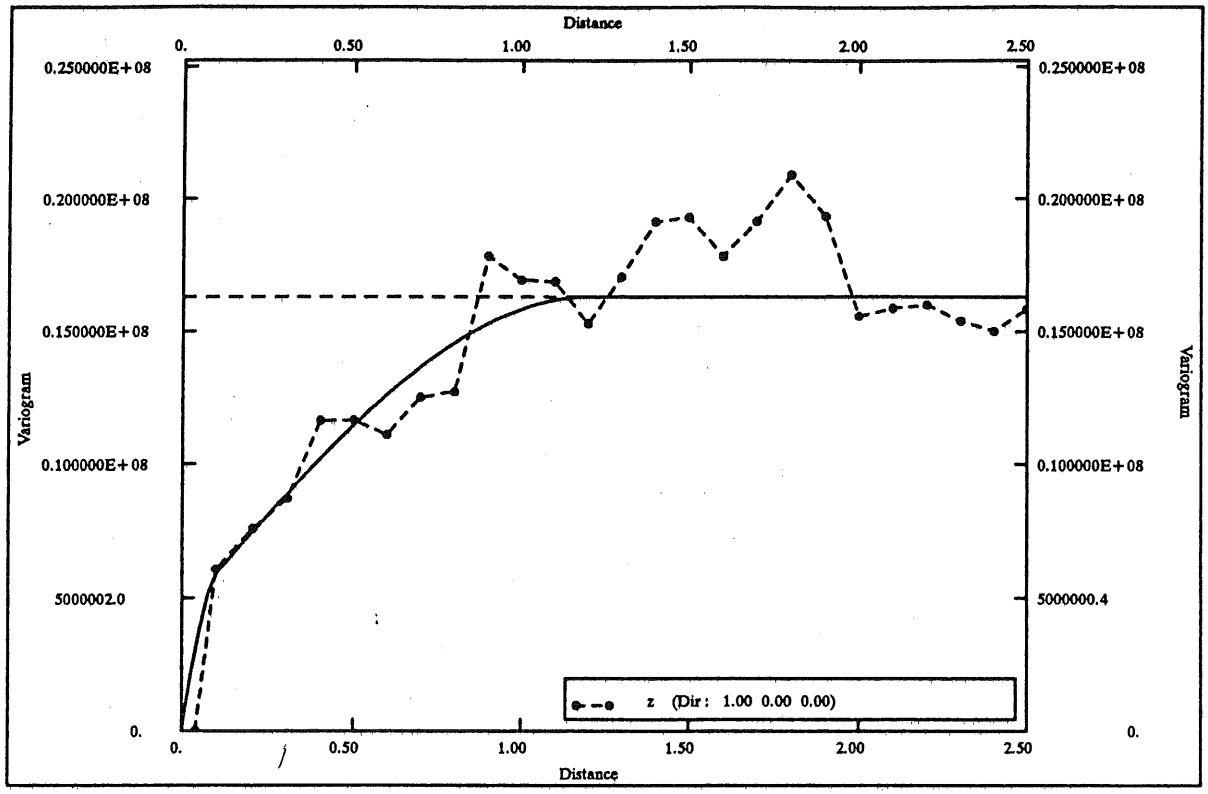


iceher 2

NUGGET: 0.00 NB. STRUCTURE(S):2 ANISOTROPY

STRUCTURE	SILL	RANGE	LAMBDA	X	Y	Z
spheric.	0.9000E+07	1.000	1.000	1.000	1.000	
spheric.	0.5120E+07	0.2000	1.000	1.000	1.000	

Figure 6.



iceher 3

NUGGET: 0.00 NB. STRUCTURE(S):2			ANISOTROPY		
STRUCTURE	SILL	RANGE	LAMBDA	X	Y	Z
spheric.	0.1180E+08	1.200	1.000	1.000	1.000
spheric.	0.4500E+07	0.1000	1.000	1.000	1.000

Figure 7.

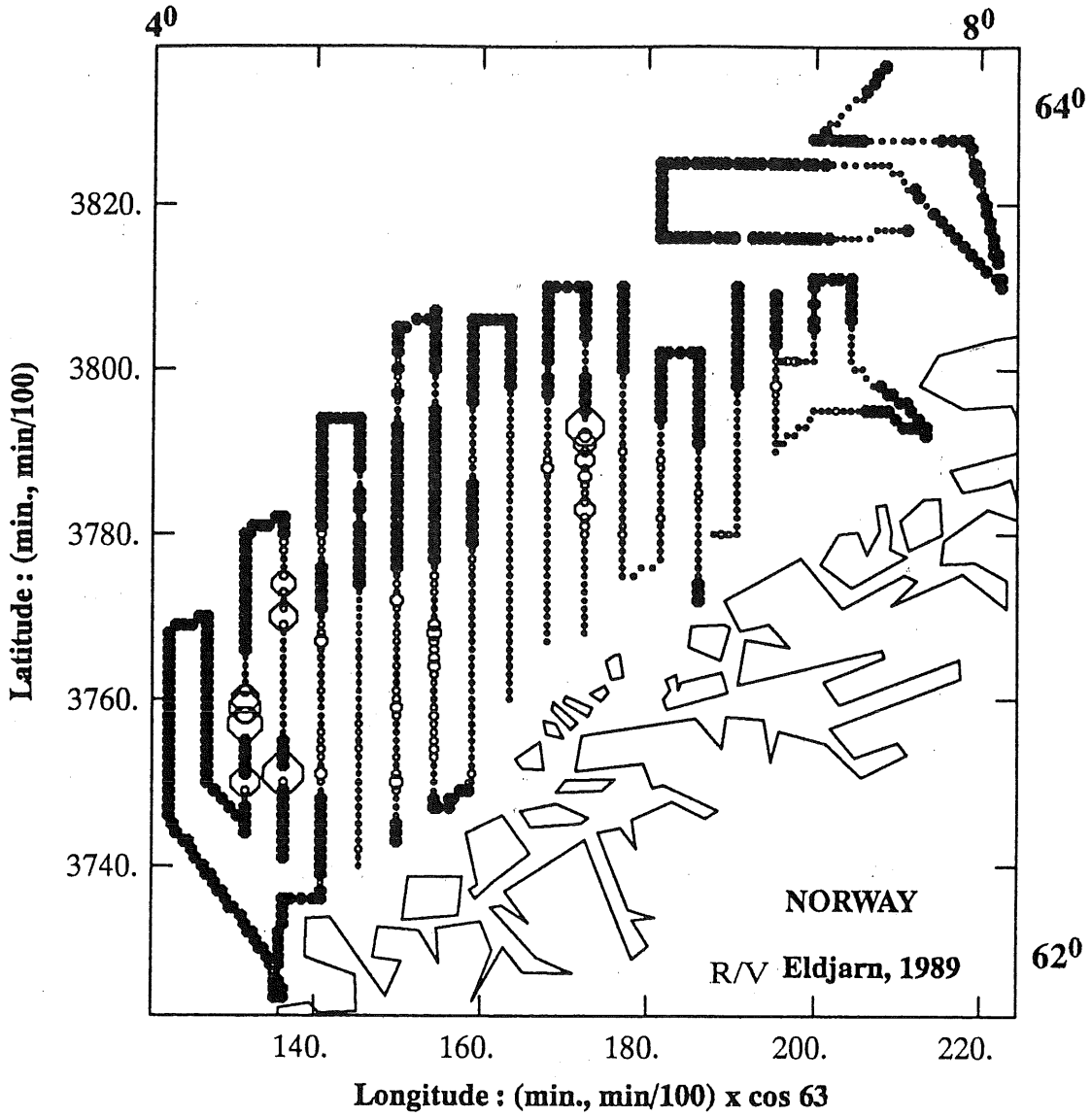


Figure 8 : Data set TEST04 :
The survey with a proportional representation of the values
The zero values are represented by black disks.

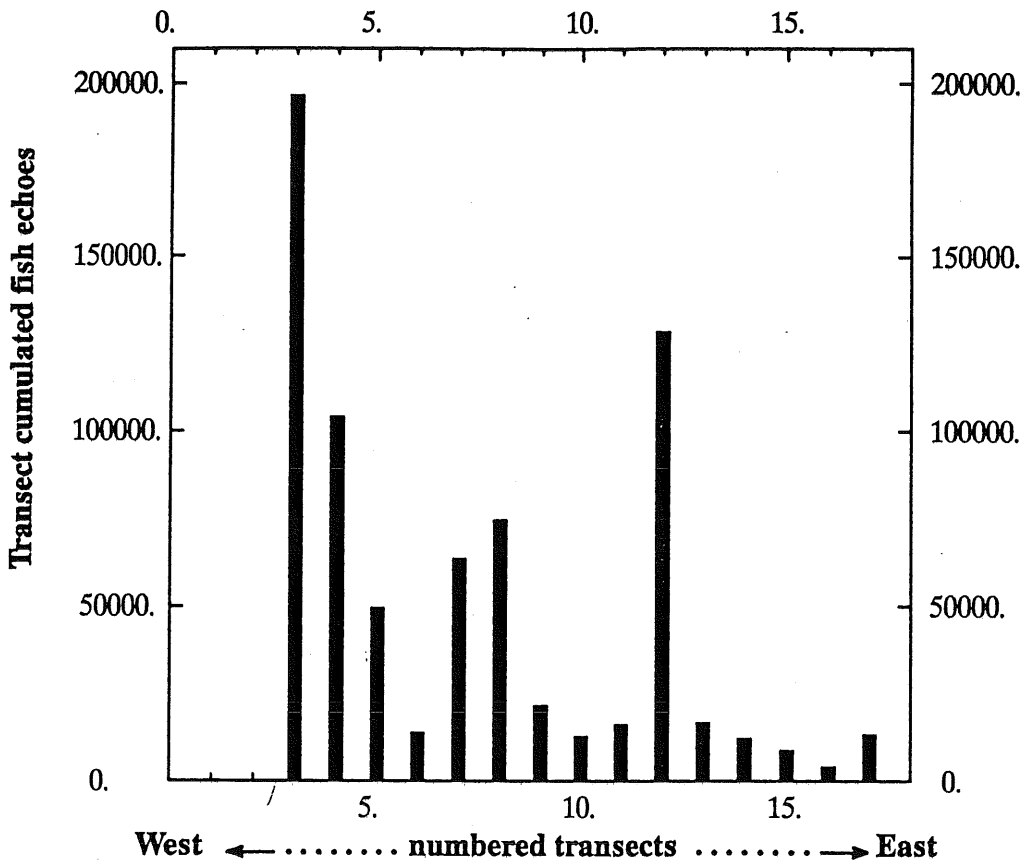


Figure 9 : Data set TEST04 :
Representation of the cumulated echos along the transects

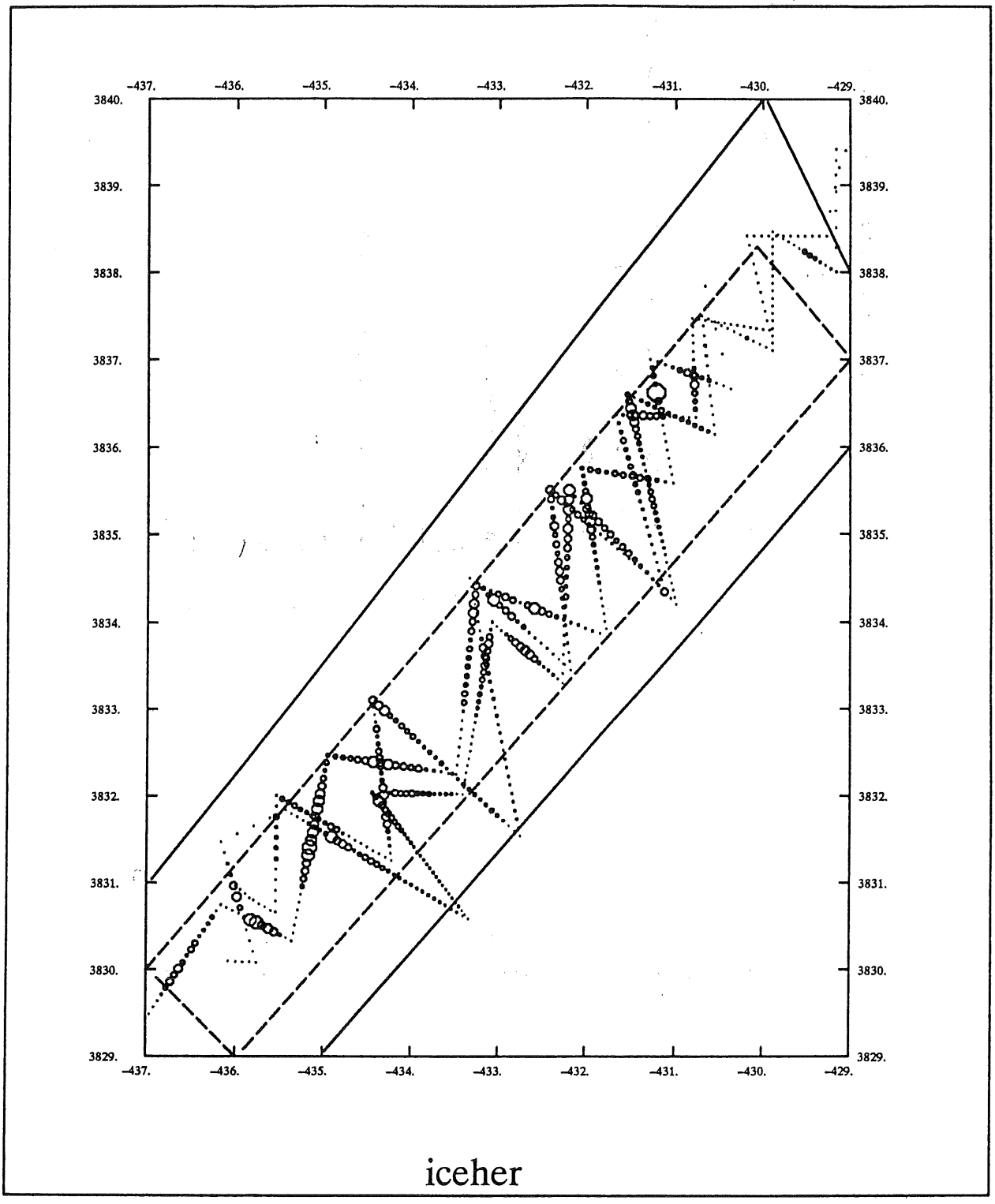


Figure 1: The three surveys with a proportional representation of the values and the two domains which have been estimated.

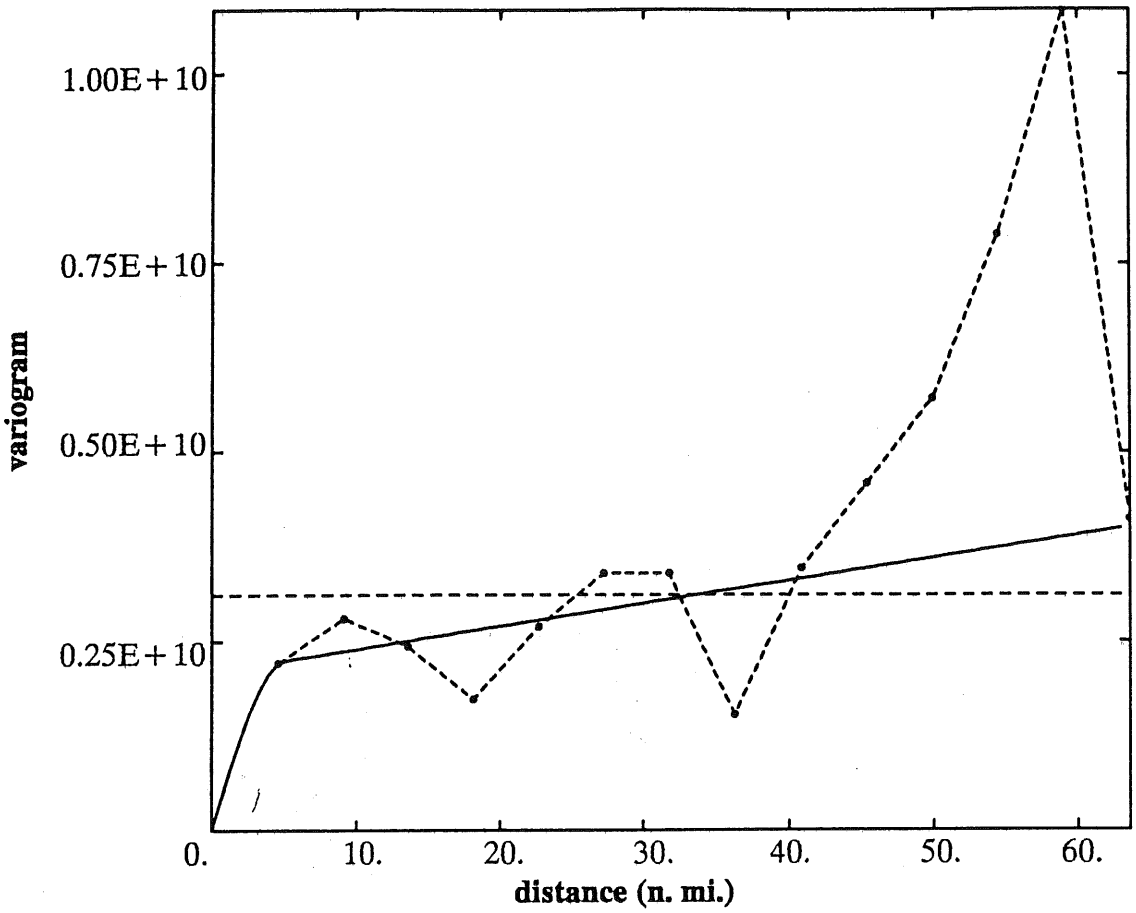


Figure 10 : Data set TEST04 :
Experimental 1D variogram of the cumulated echos and its model :
$\gamma_{1D}(h)$ = *spherical* (*sill* = 0.21 10^{10}, *range* = 5*n.m.*) + *linear model* (*slop* = 0.3 10^{8})

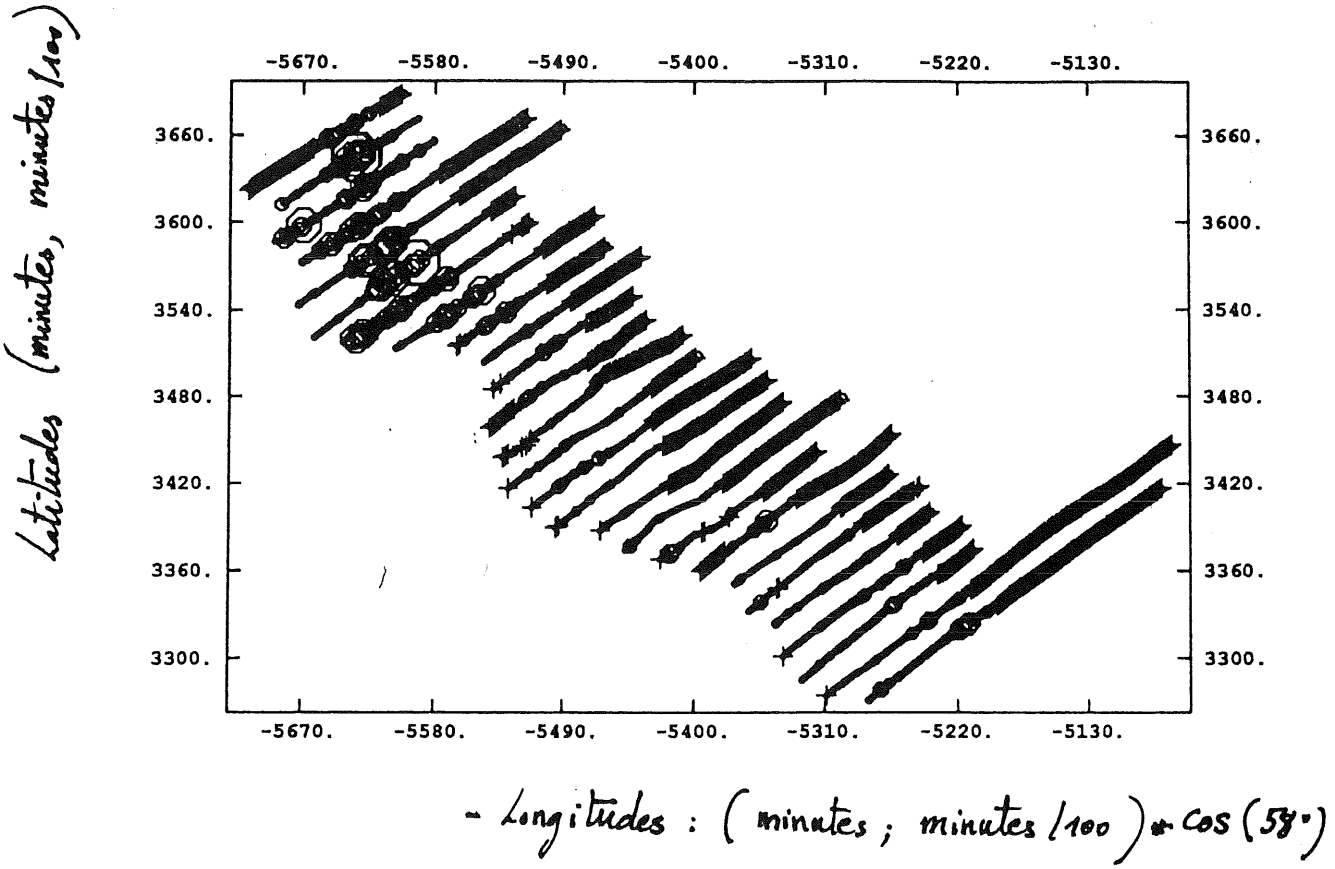


Figure 11 : Data set K0*a :
The survey with a proportional representation of the values
The zero values are represented by crosses.

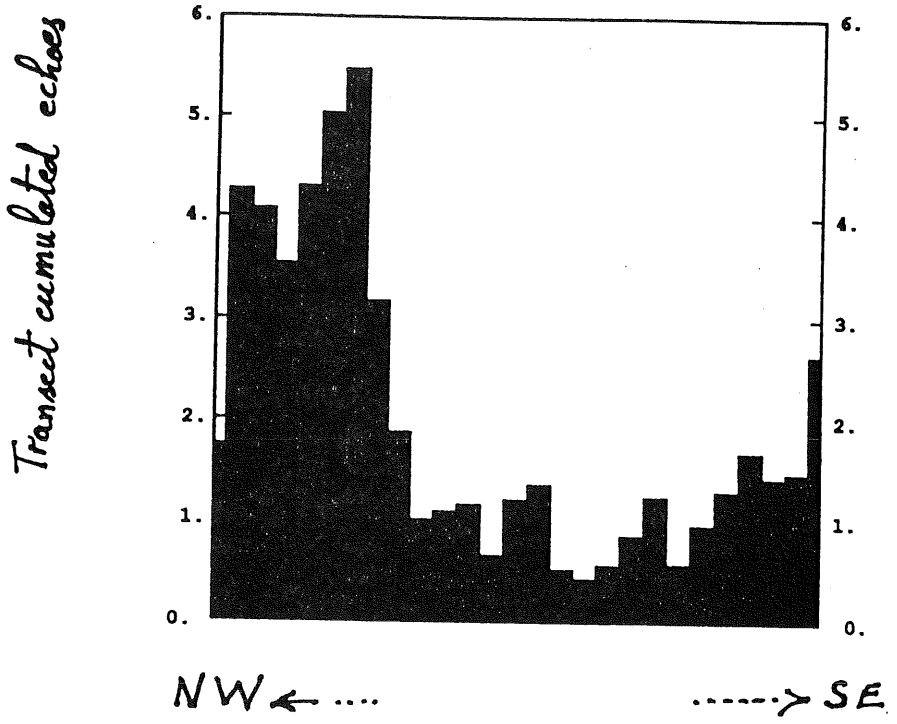


Figure 12 : Data set K0*a :
Representation of the cumulated echos along the transects

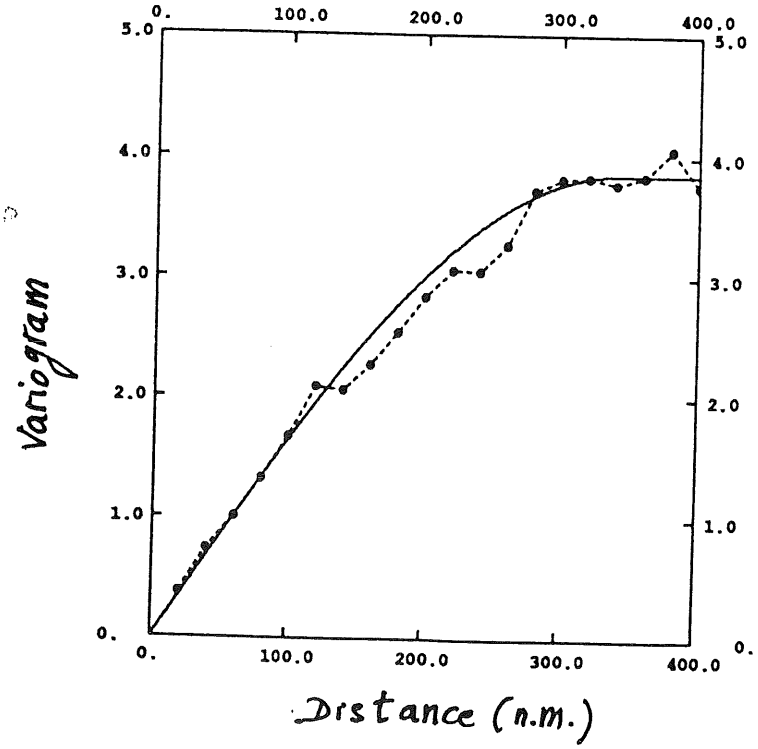


Figure 13 : Data set K0*a :
Experimental 1D variogram of the cumulated echos and its model :
$\gamma_{1D}(h) = spherical\ (sill = 3.85, range = 340 n.m.)$

Appendix C: Generalized Additive Models

by

G. Swartzman

In a generalized additive model the expected value of a random variable Y is expressed as a sum of smooth functions of the covariates. Thus

$$E(Y \mid x_1, \ldots, x_p) = \sum_{j=1}^{p} S_j(x_j), \qquad \text{(C.1)}$$

where $S_j(x_j)$ represent smooth functions of the covariates. In a generalized additive model a known function of the expected value, called the link function, is modeled as a sum of smooth functions of the covariates. This generalization of the model is easy to make for random variables in the exponential family.

If, for example, the Poisson distribution is chosen as an underlying distribution, a central assumption in the generalized additive model for spatial data is that the observations are distributed according to a nonhomogeneous Poisson distribution. The parameter of the Poisson distribution is $\Lambda = \int_{A_x} \lambda(\mathbf{u})\, d\mathbf{u}$ where $\lambda(\mathbf{x})$ is the intensity of the underlying Poisson process and A_x is the area of the observations. The expected value of the Poisson distribution is $\lambda(\mathbf{x})$ and the natural link function is the logarithm. Thus, the Poisson generalized additive model relates the expected counts to the covariates as:

$$\log\left[E(Y \mid x_1, \ldots, x_p)\right] = \log\left[\Lambda(\mathbf{x})\right] = \sum_{j=1}^{p} S_j(x_j) \qquad \text{(C.2)}$$

Or, if the additive predictor is $\eta = \sum_{j=1}^{p} S_j(x_j)$ then:

$$\Lambda = \mu = e^{\eta}$$

Since the functional form of the smooth functions, $S_j(x_j)$, $j = 1, ..., p$, is not specified, the usual estimation techniques such as maximum likelihood estimation cannot be used for generalized additive models. Instead, an algorithm that empirically maximizes the expected log-likelihood is used. The derivative of the expected log-likelihood is set to zero and the resulting equation is expanded in a Taylor series about an initial estimate of the additive predictor, η^0. The equation can then be rearranged to give a new estimate for η based on the initial estimate η^0. This update equation is used iteratively with the conditional expectation from the expected log-likelihood estimated by a scatterplot smoother. The resulting algorithm is similar to the adjusted dependent variable regression method of McCullagh and Nelder, 1989 for computing maximum likelihood estimates when the predictor, η, is a linear function of the covariates. The adjusted dependent variable for the Poisson generalized additive model at the m-th iteration is

$$z^m = \eta^m + \frac{(y - e^{\eta^m})}{e^{\eta^m}} \qquad \text{(C.3)}$$

The scatterplot smooth of z^m on x (when there is only a single covariate x) provides an updated estimate of the additive predictor, η^{m+1}.

The measure of fit for the generalized additive models is the deviance, which is twice the log of the likelihood ratio between the saturated model and the current model. For the Poisson model this is calculated as

$$Dev(\mathbf{y}, \mu) = 2\sum_{i=1}^{n}\left[y_i \log\left[\frac{y_i}{\mu_i}\right] - (y_i - \mu_i)\right] \qquad \text{(C.4)}$$

The updating iterations are continued until the deviance fails to change.

C.1 Backfitting algorithm

The above discussion of the generalized additive model was for only one covariate, x. For the spatial models that will be considered, there will be at least two covariates, e.g. longitude and latitude. To fit multiple covariates, the backfitting algorithm is used. The algorithm computes the smooth function for each of covariates by holding the other covariate functions fixed. To do this for the j-th covariate, x_j, the partial residual

$$r_j = z - S_0 - \sum_{k \neq j} S_k(x_k) \qquad \text{(C.5)}$$

where z is the adjusted dependent variable described in Eq. (C.3), is formed. An updated value of S_j is computed by smoothing r_j on x_j. The process is then repeated for each covariate.

The initial estimates for the algorithm are zero for the smooth functions S_j and the log of the overall mean count for η. The algorithm is iterated until the deviance no longer decreases or for a maximum set number of iterations.

C.2 Smoothers

The core of the generalized additive models (GAM) used is a running line smoother which is used to find the smooth functions S_i of equation

(C.2). A running line smoother fits a line by least squares to the data points in a symmetric nearest neighborhood containing n_i points around each x_i. The advantage of a running line smoother over a running mean smoother is that it reduces bias near the endpoints without sacrificing much in calculation speed (Kaluzny, 1987, Friedman, 1984).

The span of the smoother (the fraction of the data set used in estimating a line at each point) is determined using cross-validation, i.e. the smooth value for the point x_i is computed by omitting the i-th observation and the span is chosen so that the residual sum of squares is minimized. In the program used in this study the best span was found by trying the spans 0.3, 0.4, 0.5, 0.6, 0.7 and 1.0 and choosing the one which gave the smallest residual sum of squares. A span of 1.0 uses all the data to fit the least squares line and is equivalent to a simple linear regression line.

C.3 Estimation of variability

In moving from the parametric generalized linear models fit by maximum likelihood to the nonparametric generalized additive models, the likelihood theory for estimating variances is lost. However, the bootstrap methodology of Efron (1979, 1982) can be applied to the additive models to obtain estimates of variability.

A bootstrap sample of size n is drawn from the observations (x_{1j}, x_{2j}, y_j) with replacement. The Poisson generalized additive model is fit to this sample and the resulting smooth functions, S_1^* and S_2^* are saved. This is repeated N times. The spans for the running line smoothers used in the bootstrap fitting are fixed at the values chosen by cross-validation on the original data. If the span is allowed to vary for each bootstrap sample, essentially a new model would be fit when the interest lies in the variability of the model fit to the original data. The upper and lower $\alpha/2$ empirical quantiles of the f_i^* at each x_{ij} gives an approximate $(1-\alpha)\times 100\%$ prediction interval for S_i at that value of x_i.

C.4 Test of trend significance

The bootstrap prediction intervals are one method to assess the significance of the smooth functions. The intervals indicate a range of possible values the function could have. If a horizontal line can be drawn within the prediction interval then there is an indication that the smooth function is not significant. A more formal approach is to do a permutation test. The null hypothesis that is considered by the test is:

$$H_0: S_i(x_{ij}) = m \ \ (a\ constant) \ \text{for}\ all\ j,$$

i.e. the smooth function for covariate x_i does not depend on x_i. Under this null hypothesis, any permutation of $(x_{i1}, x_{i2}, \cdots, x_{in})$ should result in approximately the same overall fit. If the null hypothesis is false then permuting the values of x_i should not result in as good a fit as that obtained from the original data. Here the term "good fit" is taken to mean a small deviance. To provide a familiar measure of model fit a pseudo r^2 is computed as 1.0 minus the ratio of the deviance in the best fitting model to the deviance for the overall mean (the null or zero model). While not identical to the classical r^2 this measure is bounded between 0 and 1 and is used as a surrogate for it. Since all possible permutations cannot be examined, only a sample of size N of the possible permutations is used. The deviance from the generalized additive fit to each of the N permutations of the covariate vector $\mathbf{x}_i$ is recorded along with the other unpermuted covariates. To avoid changing the model being fit the same fixed span smoother is used for all the fits, with the span being chosen by cross-validation on the original data. If the deviance from the original data is the m-th smallest among the $N + 1$ deviances the null hypothesis is rejected at the $m / (N + 1)$ level.

C.5 Application of GAM to data sets

The primary focus in the GAM analysis of the data sets provided was on uncovering relationships between fish abundance and environmental factors. Only depth was provided as an ancillary variable (except for latitude and longitude of the sampling locations) and that only for the Icelandic herring and Bering Sea surveys. The simulated data set was therefore not addressed. Scatterplots of fish abundance against depth for each of the Icelandic surveys suggested very little relationship between abundance and depth over the surveyed area (Fig. C.1). There is a drop in abundance below 80m, however, this depth range comprised only a tiny fraction of the overall survey. GAM with latitude and longitude as covariates might provide a marginally improved fit to the data relative to GLM (see section 6). However, this fit would not help to explain the spatial distribution, and other methods appear to provide better estimates than the GLM

estimates.

Contour and image plots of depth and abundance for the Bering Sea survey (Fig. 2.4) suggest that the spatial distribution of pollock is strongly related to depth. These figures were based on spatial interpolations of the abundance and depth data provided for the survey (see section 2 for a discussion of the dangers of such interpolations). GAM was run on these data with depth, latitude and longitude as covariates. Due to the large number of data points the data were binned into a 40x40 grid. The average of all data points in each of the 1600 grid bins were taken as the value for that bin. Fig. C.2 shows the GAM smooth on depth along with the depth residuals (conditioned on the fits for latitude and longitude) and one standard error range (dashed lines). The span used for the smoother was 1/3 (i.e. 1/3 of the data distributed around each point used for estimating the value at that point). This figure indicates that almost all the high abundance points are between 100 and 130 m, just off the shelf break. 3-D plots of the raw abundance data and the GAM fitted mean are shown in Figs. C.3 and C.4 respectively. These demonstrate the quality of GAM of flattening and spreading out peaks. Also, GAM has no protection against giving negative estimates at some points, although this is only a minor affect over the entire survey region. The GAM mean values could be converted to an overall abundance estimate using the same method as used for GLM. Although variance estimates are provided by the S+ version of GAM, these are approximate and the theory is not clear. Bootstrap resampling can be used to provide pointwise variance estimates. The spatial distribution of the residuals should be examined (this was not done here) as discussed in the data analysis section 2 of this report. If the residuals appear to be spatially autocorrelated, further analysis with a variogram of the residuals suffers from bias of the residuals through the deviance minimization process inherent in the GAM algorithm.

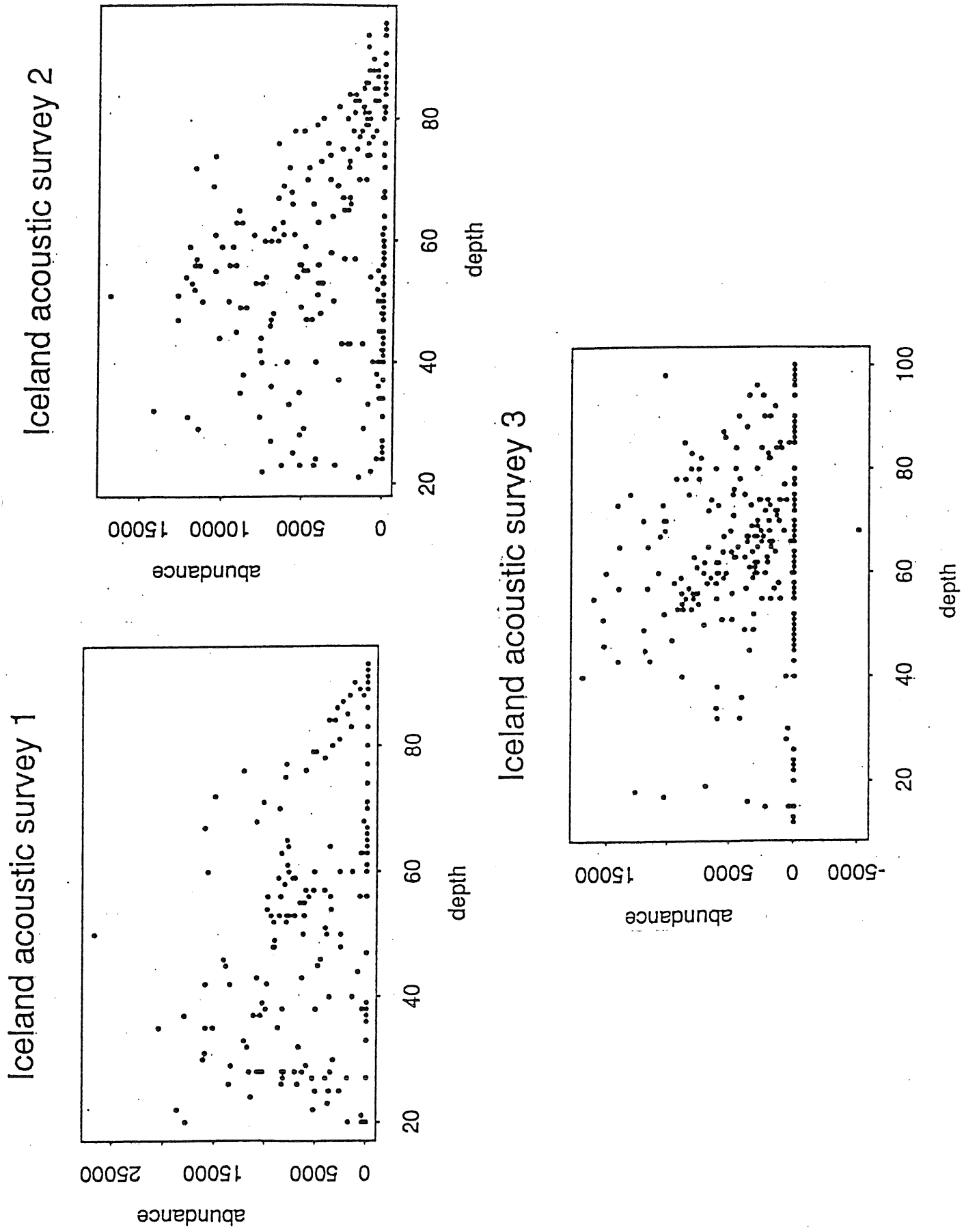


Figure C.1. Scatterplots of acoustic measurements *versus* depth for three Icelandic acoustic surveys of herring.

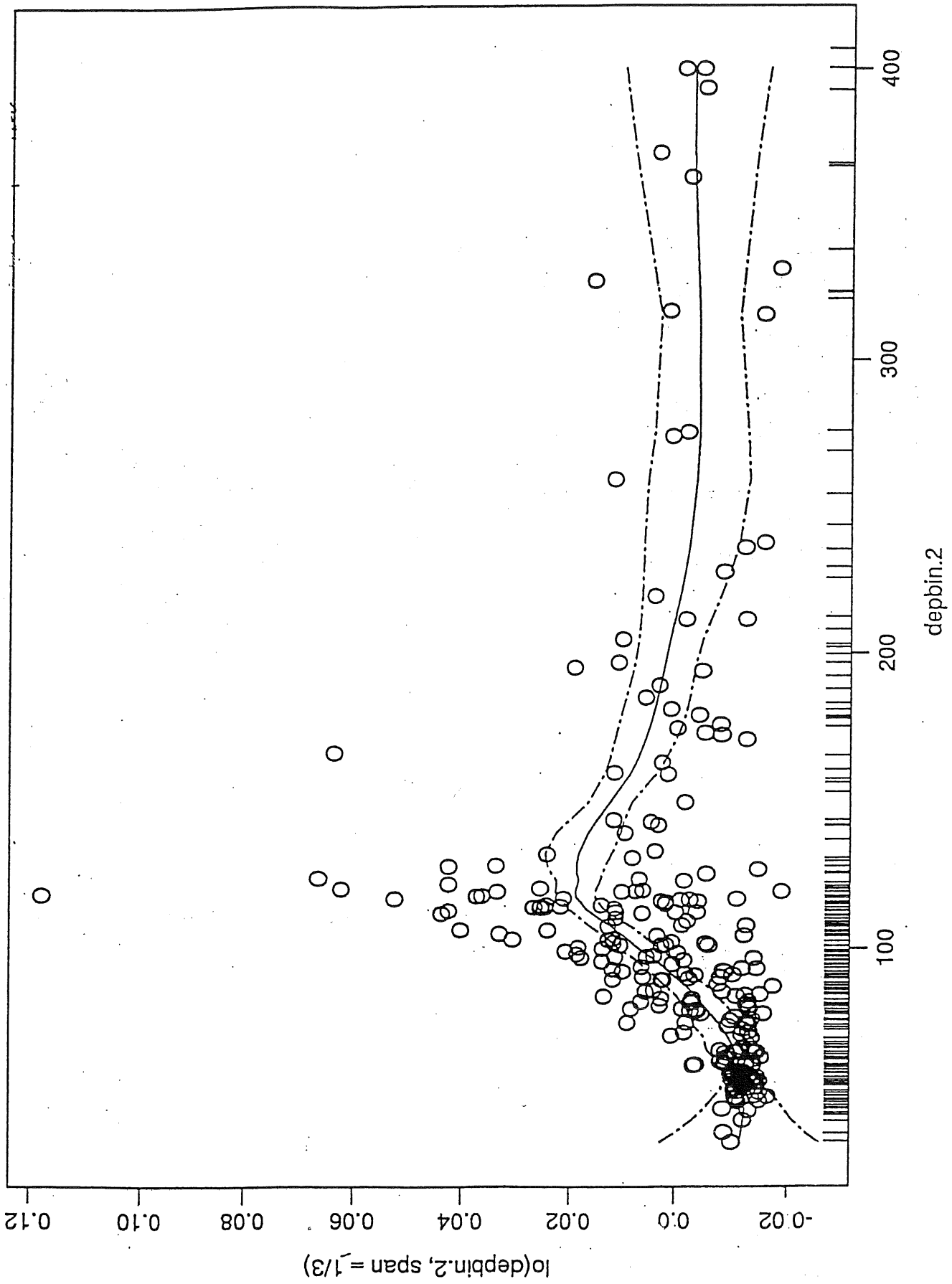


Figure C.2. GAM smooth and residuals *versus* depth from acoustic data for Bering Sea walleye pollock.

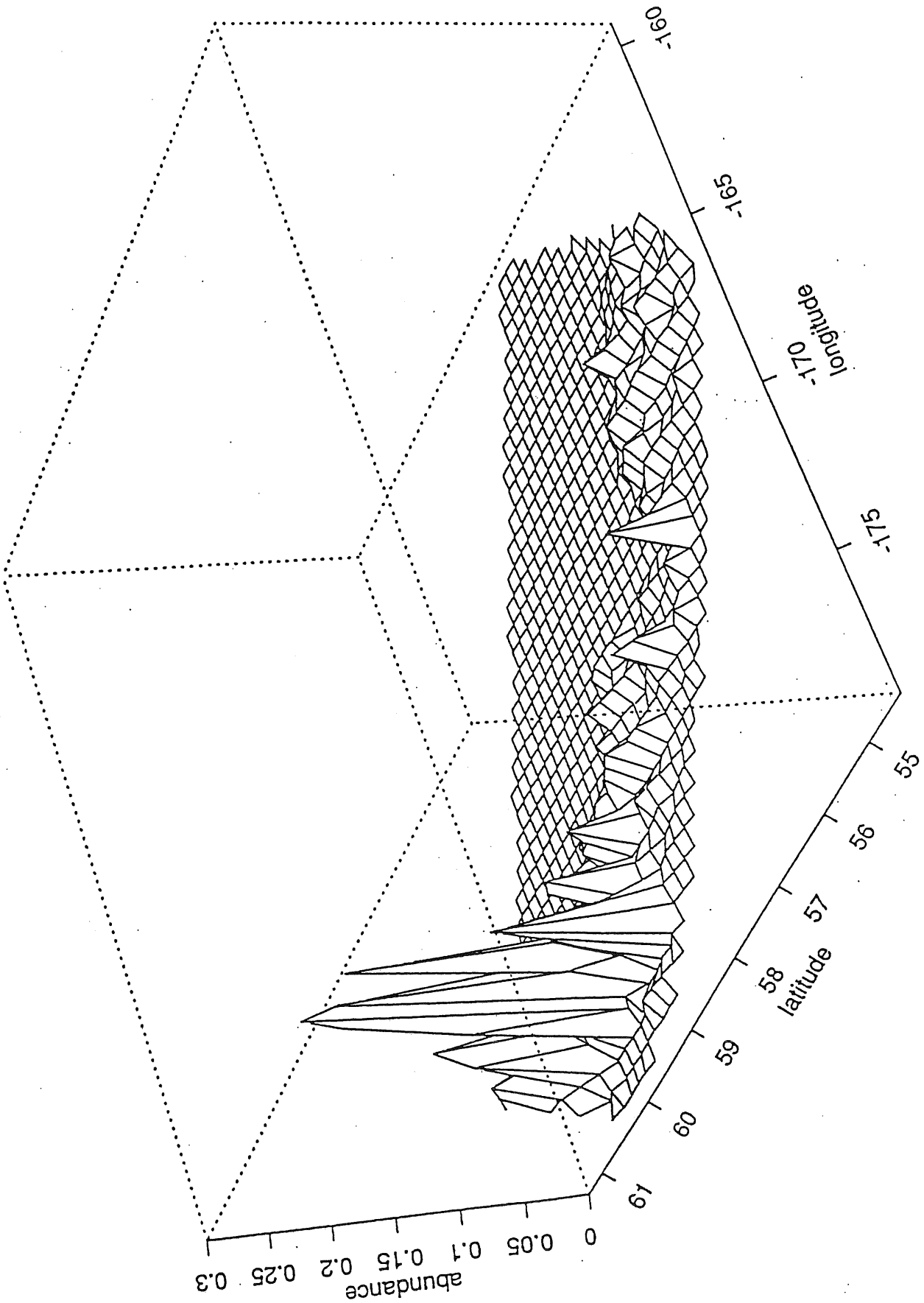


Figure C.3. Three-dimensional representation of acoustic abundance measurements.

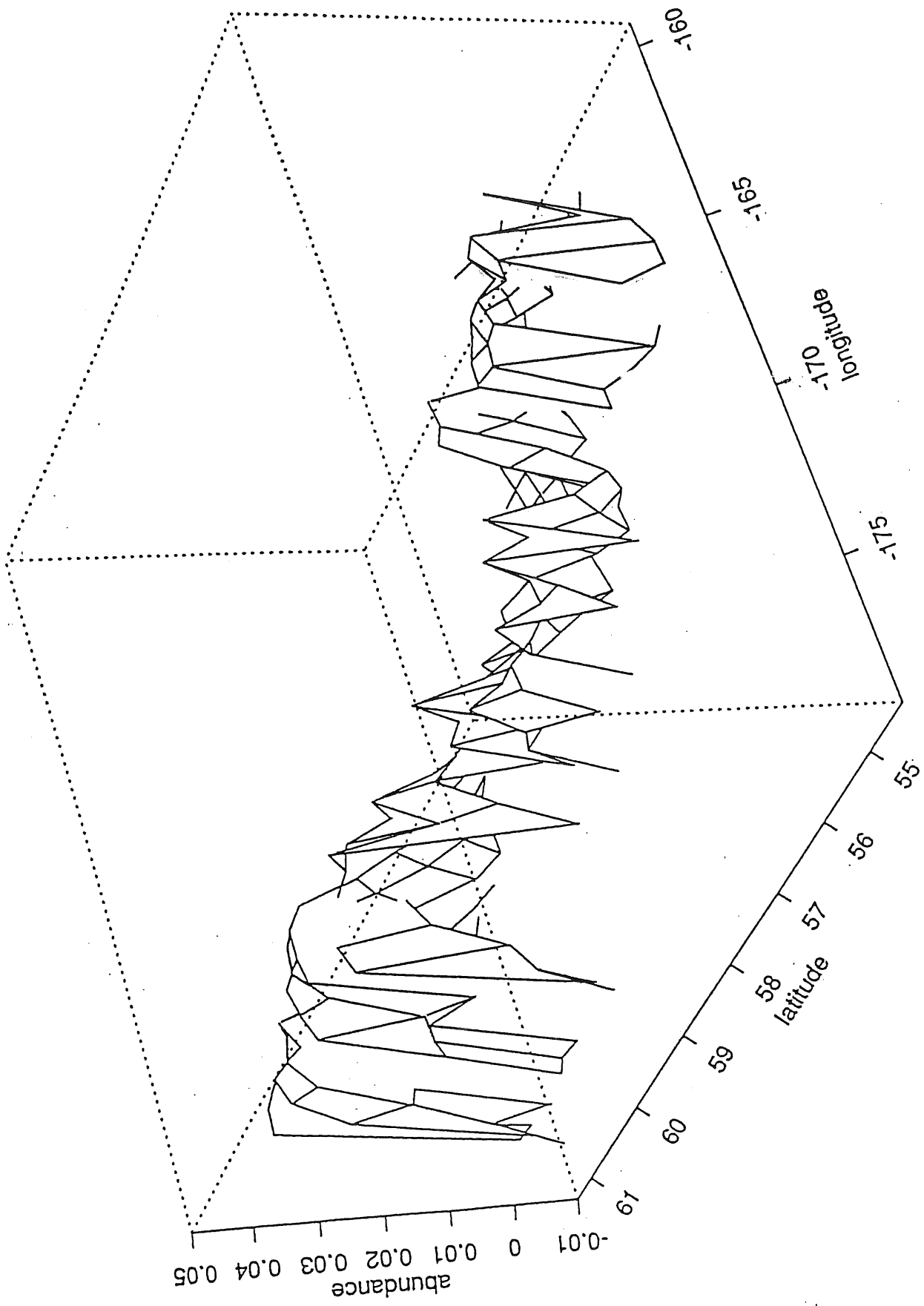


Figure C.4. Three-dimensional representation of fitted acoustic abundance using GAM.

Appendix D: Participants and addresses.

Dr G. Y. Conan
Science
Gulf Fisheries Centre
343 Archibald St. P.O. Box 5030
Moncton, NB E1C 9B6
Canada

Ms M. Daniel
(E-mail: marsha@hafro.is)
Dr G. Stefánsson
(E-mail: gunnar@hafro.is)
Marine Research Institute
P. O. Box 1390
Skúlagata 4
Reykjavík
Iceland

Prof. E. Ferrandis
Dpto De Matematicas
y Estadistica
Facultad de Ciencias
Universidad de Alicante, Apdo 99
03071 Alicante
Spain

Dr K. G. Foote (until summer, 1992)
Centre de Geostatistique
35, rue St Honoré
77305 Fontainebleau
France

Mr K. Korsbekke
(E-mail: knutk@imr.no)
Dr K. Foote (after summer, 1992)
Institute of Marine Research
P.O. Box 1870, Nordnes
5024 Bergen
Norway

Mr R. Oeberst
Institut für Hochseefischerei
und Fischverarbeitung
An der Jägerbäk 2
O-2510 RostockS
Federal Republic of Germany

Mr L.-E. Palmén
Institute of Marine Research
P.O. Box 4
453 00 Lysekil
Sweden

M P. Petitgas
Mr. J. Rivoirard
Centre de Geostatistique
Ecole des Mines de Paris
35, rue St Honoré
77305 Fontainebleau
France

Dr E. J. Simmonds
Marine Laboratory
P.O. Box 101
Victoria Road
Aberdeen AB9 8DB
United Kingdom

Dr. P. J. Sullivan
(E-mail: pat@hyacinth.iphc.washington.edu)
IPHC
P.O. Box 95009
Seattle
WA 98145-2009
USA

Prof G. Swartzman
(E-mail: gordie@apl.washington.edu)
Applied Physics Laboratory HN-10
University of Washington
Seattle, WA 98195
USA

Dr W. Warren
Dept. of Fisheries & Oceans
NW Atlantic Fisheries Centre
P.O. Box 5667
St John's, Nfld A1C 5X1
Canada