International Council for the
Exploration of the Sea

C.M. 1992/D:10
Statistics Committee
Ref. B + H

GEOSTATISTICAL ANALYSIS OF ACOUSTIC SURVEY DATA ON 0-GROUP HERRING IN A FJORD

by

Kenneth G. Foote
Institute of Marine Research
5024 Bergen
NORWAY

and

Jacques Rivoirard
Centre de Géostatistique
77305 Fontainebleau
FRANCE

ABSTRACT

An essentially pure aggregation of 0-group herring in Altafjord in northern Norway was acoustically surveyed on each of the first three days of December 1991. Measurements were made at 38 kHz during each survey and at 120 kHz during part of the second survey and all of the third survey. The data are expressed in terms of the area backscattering coefficient, applicable to the entire water column and averaged over intervals of 0.1 nautical mile. The spatial structure of the aggregation is characterized by the variogram. By means of geostatistics, estimates of abundance over the survey area are supplemented by estimates of variance.

RESUME: ANALYSE GEOSTATISTIQUE DE DONNEES ACOUSTIQUES DE HARENG-0 DANS UN FJORD

Une population pratiquement pure de harengs-0 dans le fjord Alta en Norvège du Nord a été reconnue par campagnes acoustiques, une pour chacun des 3 premiers jours de décembre 1991. Les mesures ont été faites à 38 kHz pour chacune des 3 campagnes, et à 120 kHz pour une partie de la deuxième et pour la totalité de la troisième. Les données sont exprimées en terme d'indice de réflexion surfacique applicable à la colonne d'eau complète, et moyenné le long d'intervalles de 0.1 mille nautique. La structure spatiale de la population est caractérisée par le variogramme. Les estimations de l'abondance dans la zone reconnue sont assorties d'estimations de variances calculées par la géostatistique.

INTRODUCTION

Acoustics is a major tool for surveying fish stocks within ICES member countries. On the basis of acoustic survey data, the abundance of a number of stocks is estimated and, ultimately, fishing quotas are established. Thus both the estimate and its goodness, in a statistical sense, are important.

Statistics is essential to the analysis of acoustic survey data and, indeed, even to the planning of surveys. A number of techniques have been developed for the estimation of abundance. These include techniques which ignore stratification or which account for it with respect to transects, blocks, or rectangles bounded by lines of latitude and longitude, in addition to contouring, according to the classification by Simmonds et al. (1991).

The same statistical techniques that are used in the estimation of abundance may also be used in some cases to estimate variance. For the most part, estimation of this second quantity is recognized to be fraught with difficulty. A particular reason for this is the evident correlation or connectedness of measurements of fish density. This reflects the simple biological fact that fish do not distribute themselves independently of one another, but, for whatever reason, aggregate.

Geostatistics is the name of a set of techniques which exploit observed correlation in geographical distributions to estimate variance associated with estimates of local concentration and of global abundance. The techniques are illustrated here for three surveys of essentially pure 0-group herring (Clupea harengus) in northern Norway in December 1991.

This work is intended to contribute to the ongoing discussion within ICES on the use of geostatistics for the analysis of acoustic survey data (Anon. 1990, 1991).

MATERIALS

An essentially pure aggregation of herring was surveyed three times during the period 1-3 December 1991 by the SIMRAD EK500 echo sounder system (Bodholt et al. 1989) with 38 kHz transducer. The first two surveys followed the same design, shown in Figs. 1 and 2. The third survey followed a different design, shown in Fig. 3. Some circumstances of the surveys, especially apropos of daylight, are shown in Table 1.

Table 1. Circumstances of the three surveys in Altafjord.

Date (1991)	Hour (GMT)	Average speed (knots)	Day/night
1 Dec	1108-1918	6.5	Civil-nautical twilight for first half, night for second half
2-3 Dec	2238-0623	6.8	Night
3 Dec	0750-1300	11.4	Night at very beginning, then nautical-civil twilight for duration

The echo sounder and Bergen Echo Integrator (Foote et al. 1991) were calibrated in Olderfjord on 2 December with standard spheres according to the ICES procedure (Foote et al. 1987). The exercise was routine and

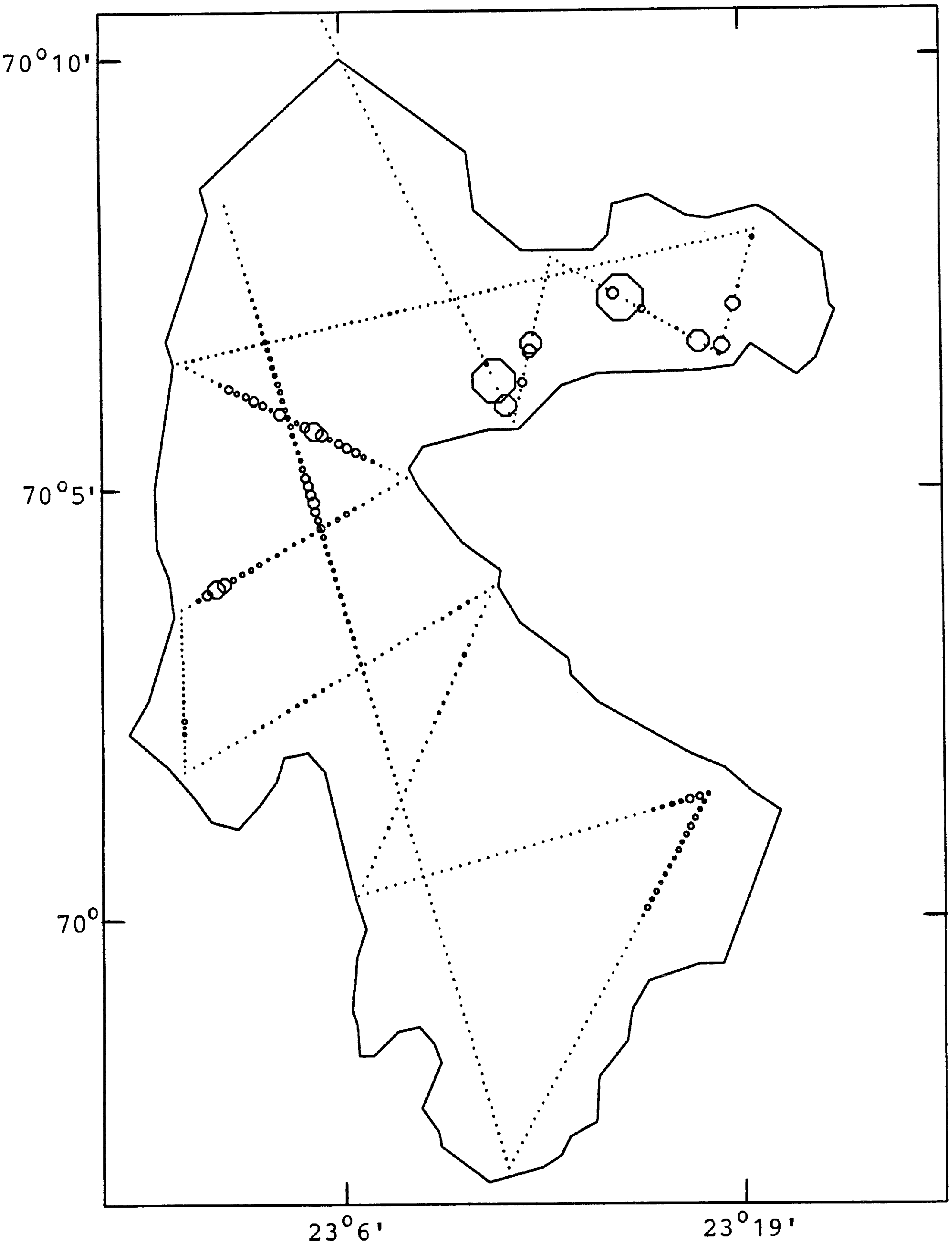

Fig. 1. Survey in Altafjord on 1 December 1991, with survey area defined by the solid-line border. Rafsbotn is excluded, for not being covered. The vessel track begins outside of the survey area and ends inside the border, in the upper left quadrant. The diameter of the dots or circles is directly proportional to the local area density or s_A-value.

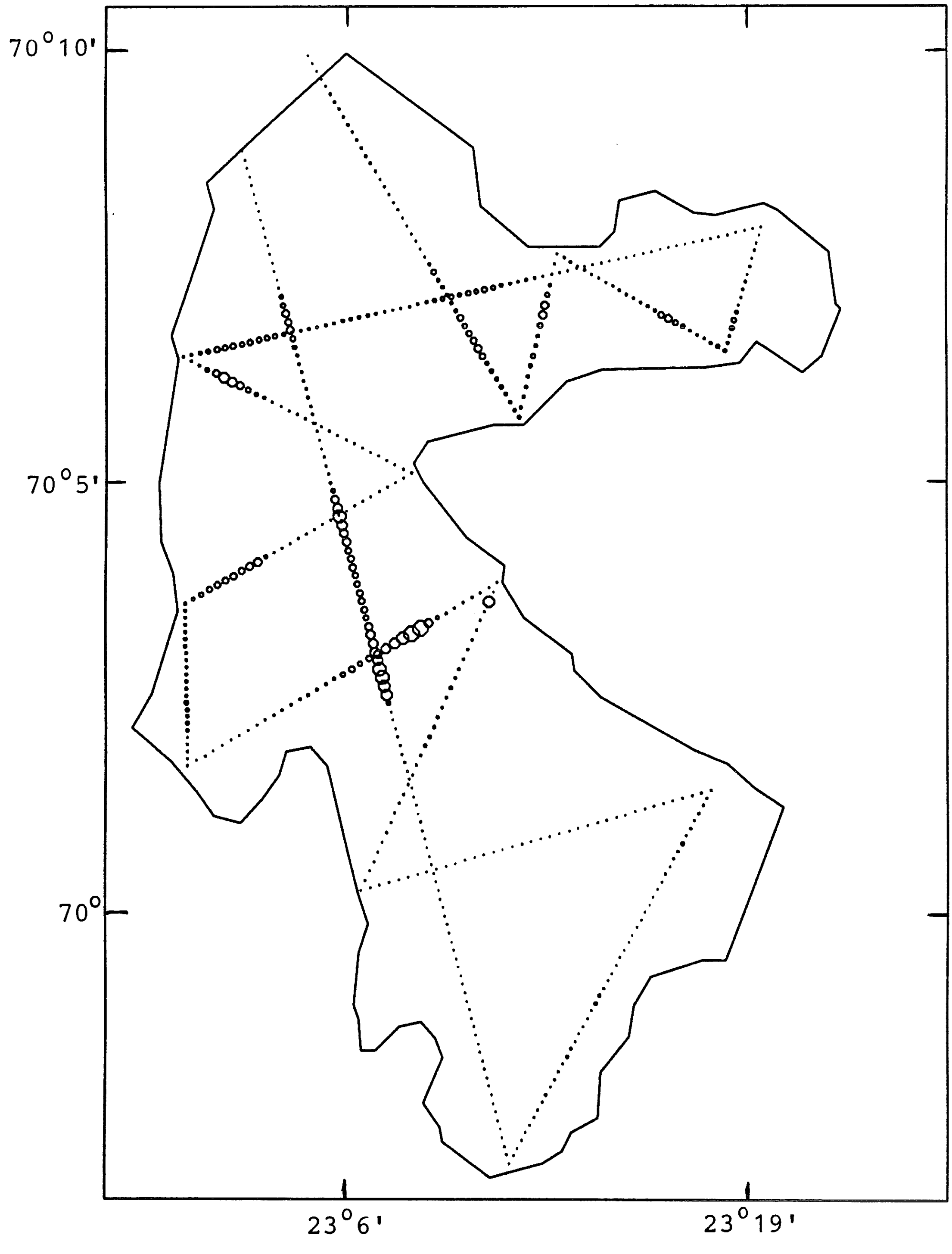

Fig. 2. Survey in Altafjord on 2-3 December 1991.

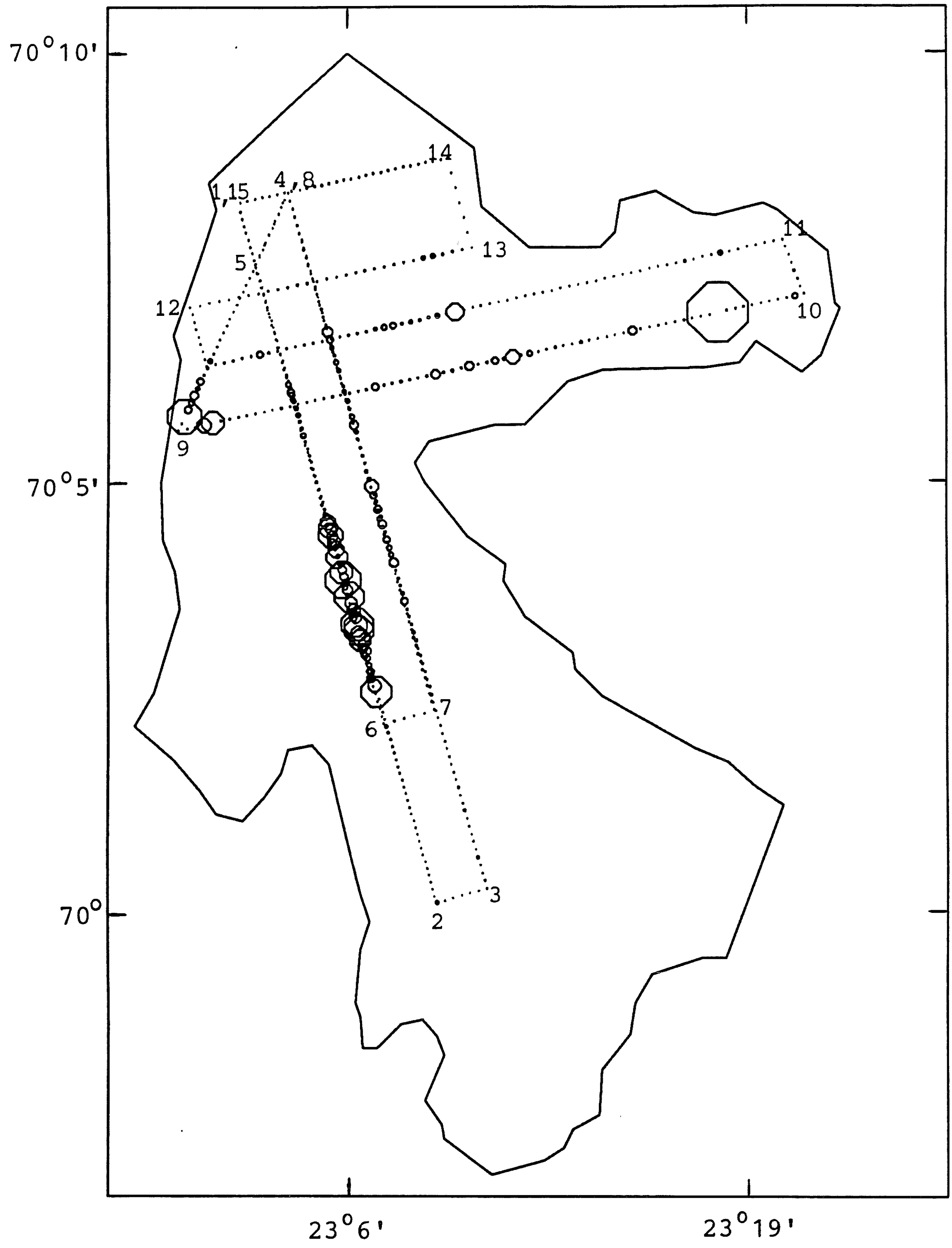

Fig. 3. Survey in Altafjord on 3 December 1991.

differences in performance with respect to the previous calibrations were negligible.

Measurements of the mean volume backscattering strength were logged on the Bergen Echo Integrator. These were interpreted by Kaare A. Hansen and Egil Ona. Results for the area backscattering coefficient s_A (Knudsen 1990) were stored in the attached database with the following resolution: 10-m-thick layers in depth and 0.1-nautical-mile (NM) intervals of sailed distance.

The measurements at 38 kHz were supplemented over some portions of the surveys by measurements at 120 kHz. These were very similar to the data at 38 kHz whenever the fish were similarly concentrated at shallow depths, i.e., whenever the signal-to-noise ration was high. The data at 120 kHz, while interesting for their bearing on the problem of the frequency dependence of target strength, add no new spatial information to the data at 38 kHz, hence are not considered further here.

METHODS

The two quantities to be estimated are the total abundance and the variance of the estimate. The key statistical quantities are the mean value of s_A and its variance estimate σ_E^2 for each survey.

Since the coverage of the fjord is more or less uniform, a weighting procedure such as that of kriging, has not been used. The mean of s_A is thus computed as the arithmetic mean of the individual measurements of s_A. If these are described by the set of numbers $\{z_i, i=1,2,\ldots,n\}$, then

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i \qquad (1)$$

is the estimate of the mean.

The global variance estimate is given by the formula (Matheron 1971):

$$\sigma_E^2 = 2\bar{\gamma}_{tv} - \bar{\gamma}_{tt} - \bar{\gamma}_{vv} \quad , \qquad (2)$$

where $\bar{\gamma}$ denotes the average of the variogram γ over regions indicated by the subscripts. The subscript t denotes the transect itself while v denotes the volume or total domain of data collection.

In terms of the measured quantity z, the variogram at lag distance h is

$$\gamma(h) = \frac{1}{2}E\{[z(x+h) - z(x)]^2\} \quad , \qquad (3)$$

where x denotes the vector position of the measurement and h, the vector of magnitude h. The distance between the positions of the data z(x+h) and z(x) is h. In practice, h is a continuous variable, but computations of γ are performed for discrete lags. Here, the continuous variable is rounded to the nearest 0.1 NM.

The definition of γ applies to each of the three terms, γ_{tv}, γ_{tt}, and γ_{vv}, but where the lag h is defined with respect to the indicated pair of domains. In particular, the transect region t is defined by the end points of each 0.1-NM interval of sailed distance. The survey region v is defined by the points of a square grid superimposed on the bounded areas shown in Figs. 1-3, with a scale increment of 0.1 NM. Since the total surveyed area is approximately 50.73 NM^2, the superimposed square grid is composed of 5073 points.

In order to derive a biological measure of total abundance, the mean value $\overline{z}$ is divided by the mean backscattering cross section $\overline{\sigma}$ of the measured 0-group herring. This is assumed to be given by the equation that is usually applied in estimating the size of the stock of Norwegian spring-spawning herring, namely

$$TS = 20 \log \ell - 71.9 \quad , \tag{4}$$

where ℓ is the mean fish length. By definition,

$$TS = 10 \log \frac{\overline{\sigma}}{4\pi} \quad . \tag{5}$$

The mean area density of fish is thus

$$\rho_A = \overline{z}/\overline{\sigma} \quad . \tag{6}$$

The total abundance is just $A\rho_A$, where A is the total area of the surveyed region.

RESULTS

The basic data are described in several ways. The spatial distribution of measured s_A-values is indicated by the radii of the circles or dots in Figs. 1-3. Histograms of the variable $z=s_A$ are presented with simple statistics in Table 2.

Two-dimensional variograms, or variograms computed on the basis of the data in two dimensions without regard to possible intervening land masses, are presented in Figs. 4-6. These are normalized to the respective sample variance. The basic fitted model is that of a nugget effect and spherical function:

$$\gamma(h) = A_1 N(h) + A_2 S(h) \quad , \tag{7}$$

where $N(h)=1$ for all h except on the transect for $h=0$, when $N(0)=0$, and

Table 2. Histograms of data from the three surveys, showing both probability density functions (pdf) and cumulative distribution functions (cdf) in percentage, together with simple statistics: number of samples n, mean $\overline{z}$, sample standard deviation s, and coefficient of variation cv. The units of the variable $z=s_A$ are square meters of backscattering cross section per square nautical mile.

Class bounds		---- pdf ----			---- cdf ----		
z_1 < z <	z_2	1	2	3	1	2	3
1	2	-	-	0	-	-	0
2	4	-	0	1	-	0	1
4	8	1	0	0	1	0	2
8	16	4	1	1	5	1	3
16	32	7	2	2	12	3	5
32	64	7	3	9	19	7	13
64	128	12	5	13	31	12	26
128	256	14	8	10	45	20	36
256	512	14	11	14	59	31	50
512	1024	11	18	17	69	49	67
1024	2048	6	19	13	76	68	80
2048	4096	12	15	6	88	83	85
4096	8192	7	11	7	95	94	92
8192	16384	4	5	5	98	100	97
16384	32768	1	0	2	100	100	99
32768	65536	0	-	1	100	100	100
65536	131072	-	-	0	100	100	100

Statistics			
n	533	525	587
z	1856	2217	2482
s	4576	2929	6168
cv	2.47	1.32	2.48

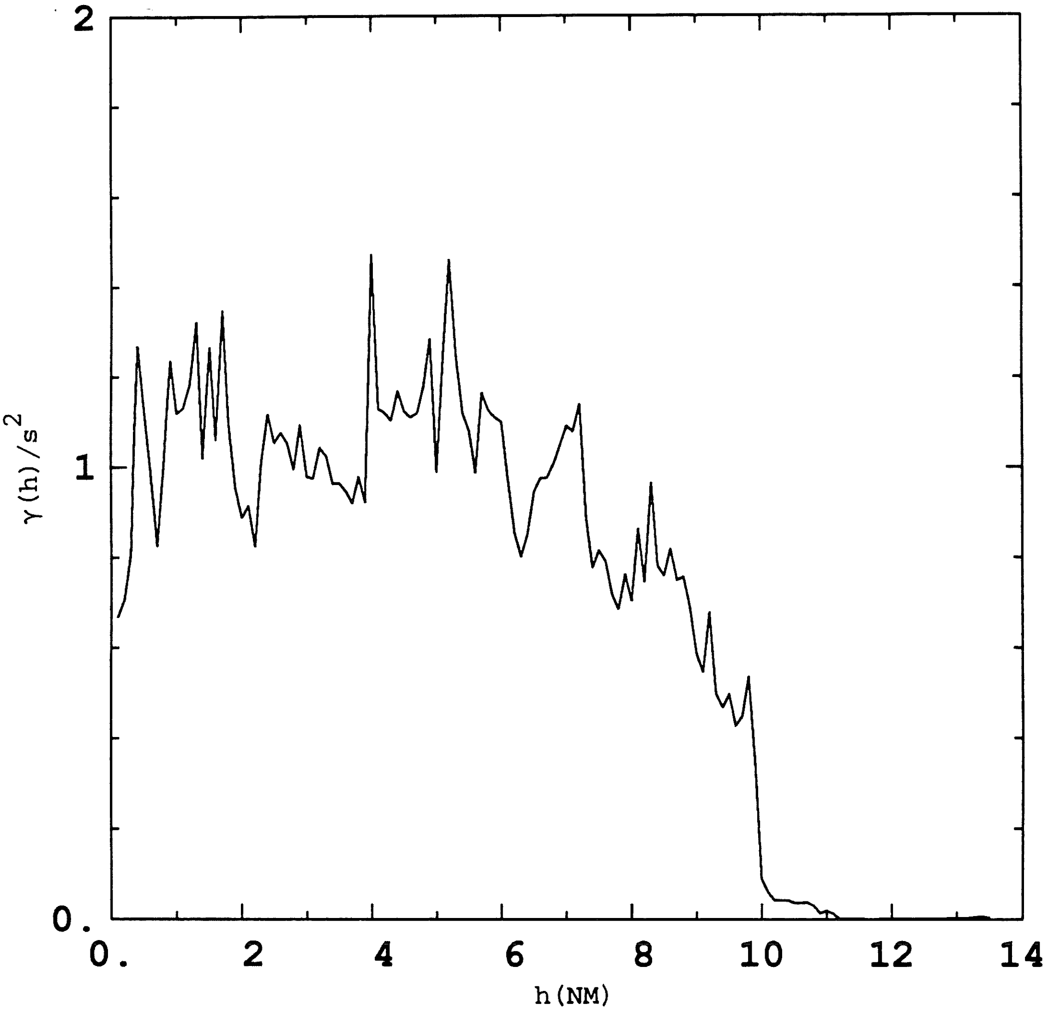

Fig. 4. Variogram of data from survey 1, normalized to the sample variance.

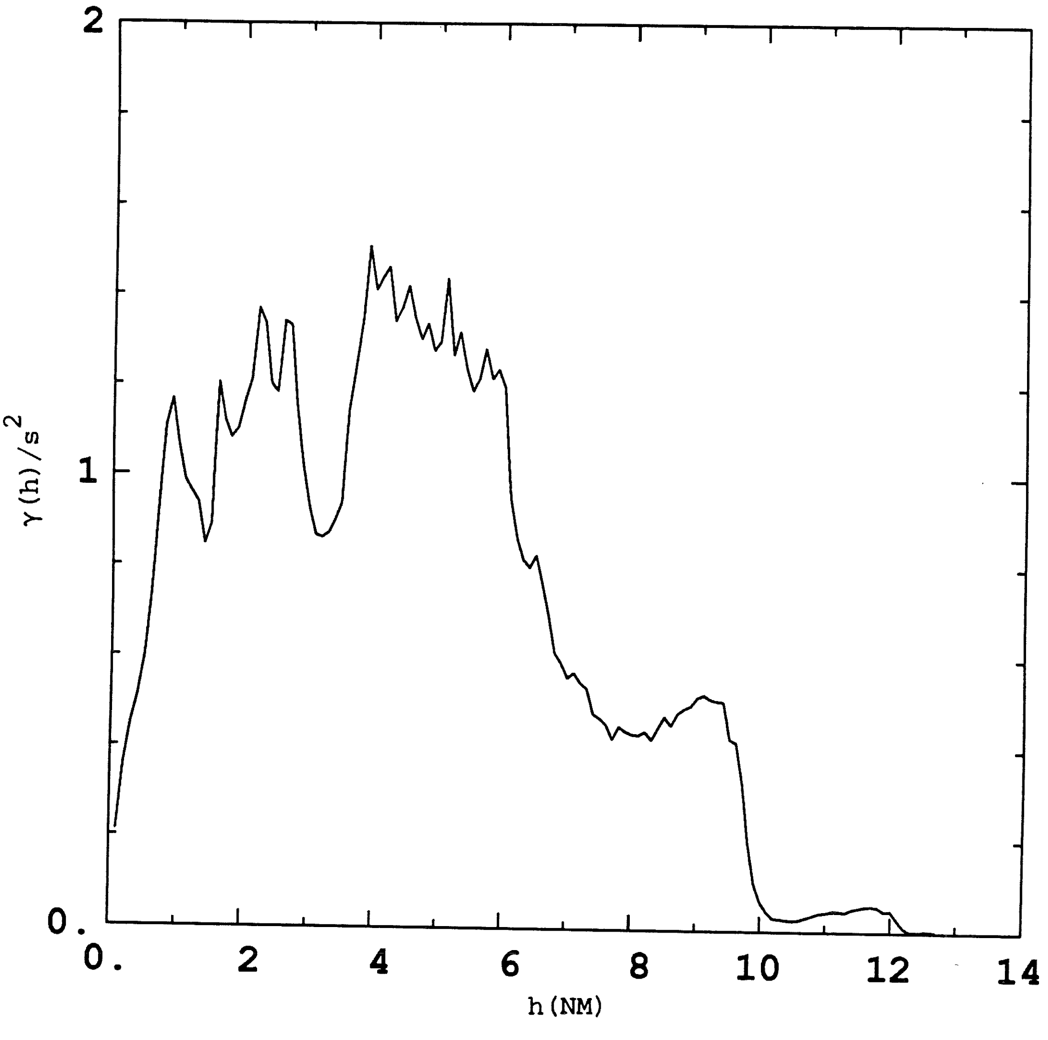

Fig. 5. Variogram of data from survey 2, normalized to the sample variance.

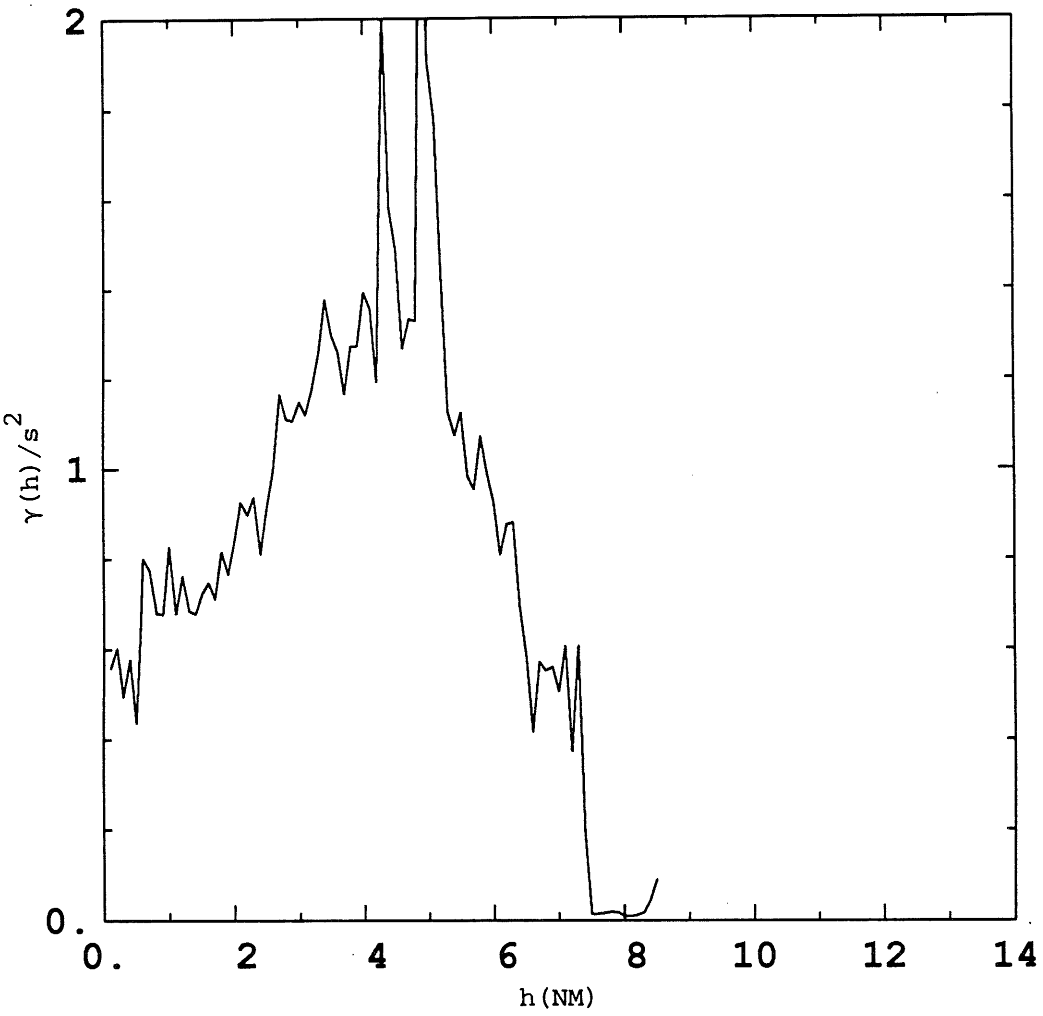

Fig. 6. Variogram of data from survey 3, normalized to the sample variance.

$$S(h) = \begin{cases} 1.5a/h - 0.5(a/h)^3 & a \leq h \\ 1 & a > h \end{cases} \quad . \qquad (8)$$

The range a of the spherical function is determined in the present examples by multiplying the range where the function crosses the sill $\sigma_E^2/s^2=1$ by 1.5, where s^2 denotes the sample variance. The amplitudes A_1 and A_2 and range a of the models are presented in Table 3.

Table 3. Variogram model parameters: A_1 denotes the nugget amplitude, A_2 the spherical amplitude, and a the range of the spherical function.

Survey	A_1	A_2	a(NM)
1	0.5	0.5	0.51
2	0	1.0	1.10
3	0.5	0.5	4.11

Further statistics are presented in Table 4. These include both the sample variance, s^2, the variance of the mean, s^2/n, and the variance of estimation, σ_E^2, defined in equation (2). For the sake of comparison, some normalized measures are also presented.

Table 4. Variance estimates of the data. The variance of the mean is s^2/n, and the estimation variance, σ_E^2.

Survey	n	$\bar{z}$	$s/n^{\frac{1}{2}}$	σ_E	$s/(\bar{z}n^{\frac{1}{2}})$	$\sigma_E/\bar{z}$
1	533	1856	198	283	0.107	0.153
2	525	2217	128	265	0.058	0.120
3	587	2482	255	1294	0.103	0.521

Global estimates of abundance are presented in Table 5. These assume the observed mean fish length of 9.52 cm, which is based on a sample size of 213, with sample standard deviation of 0.93 cm. According to equations (4) and (5), TS=-52.3 dB and $\bar{\sigma}=0.74$ cm^2. The survey area is 50.73 NM^2. The quality of the global estimates is measured through the quantity $\sigma_E/\bar{z}$, expressed as a percentage.

Table 5. Estimates of area density ρ_A (number of fish per NM^2), abundance $A\rho_A$ (total number of fish), and associated measures of confidence $\sigma_E/\bar{z}$. The assumed survey area is 50.73 NM^2.

Survey	ρ_A	$A\rho_A$	$\sigma_E/\bar{z}$(%)
1	$25.2\ 10^6$	$1.28\ 10^9$	15.0
2	$30.2\ 10^6$	$1.53\ 10^9$	11.1
3	$33.7\ 10^6$	$1.71\ 10^9$	49.8

DISCUSSION

The variability of the data appears extreme, particularly for the first and third surveys. This is evident from the spatial distributions of data indicated in Figs. 1 and 3 and from the histograms with simple statistics in Table 2.

The variability is reflected in the variograms shown in Figs. 4 and 6. Both have the character of high nuggets and rather short-range spherical models, in the language of geostatistics. The parameters of the fitted models are given in Table 3.

In contrast to the high variability of the first and third surveys, the data collected in the second survey are both less variable and more structured. This is clearly demonstrated by the variogram in Fig. 5, which lacks a nugget effect.

Reference to Table 1 may be enlightening, at least to the non-biologist. The conditions of data collection differ substantially from survey to survey. About half of the first survey and nearly all of the third survey were performed under daylight conditions, while the second survey was performed entirely at night. Apparently, as is usual, the fish were dispersed at night, and clumped with daylight.

In general, darkness is preferred for acoustic surveys. This is indeed a finding of this study, with particular applicability to hibernating 0-group herring. However, this old conclusion is accompanied by estimates of variance that explicitly account for the observed structure of the fish distribution. These are different from estimates of variance of the mean, both in number and in kind.

The variance of the mean is a simple characteristic of the set of measurements $\{z_i, i=1,2,...,n\}$ without regard to physical structure. That is, geophysical or relative positions are ignored.

There is, of course, structure in the data. This is described in geostatistical techniques by the variogram and characterized for computational purposes by a fitted model. This is not necessarily true, as it makes assumptions about the data, especially concerning their stationarity.

Much more can be, and has been, done with the present data. Three short investigations are mentioned. (1) Division of the first survey into daylight and night-time parts allows separate analyses. The two variograms show the expected pattern of high nugget effect and no nugget effect, respectively. (2) Elimination of the redundant part of the third survey is illustrative. If transects 1-4 are removed, then the estimation variance is much larger than if transects 4-7 are removed. The quantity $\sigma_E/\overline{z}$ is 69 and 47% in the respective cases. This also illustrates day/night differences. In the first case, daylight prevails and the data distribution is due to a more clumped fish distribution. In the second case, the smoothing effect of the earliest collected, night-time-like data is experienced. (3) The variogram could also be computed by averaging the individual transect variograms. This has also been done, without significant difference from the present results reported in Figs. 4-6 and Table 3, except owing to the loss of large-lag couples in equation (2).

A study that has not been performed would involve stratifying Altafjord into two geographical regions, containing respectively the highest values and the bulk of lower, more regular values. Recomputation of variograms and using these in estimating the global variance would be a worthwhile exercise. It is not, however, expected to change the present findings to any significant degree.

ACKNOWLEDGEMENTS

Kaare A. Hansen and Egil Ona are thanked for their interpretation of the original acoustic data. One of the authors (KF) wishes to express his appreciation to Ecole Nationale Supérieure des Mines de Paris: Centre de Géostatistique, Fontainebleau, for its invitation to visit, summer 1991 - summer 1992, and to Norges Fiskeriforskningsråd for a stipend during the same period.

REFERENCES

Anon. 1990. Report of the Study Group on the Applicability of Spatial Statistical Techniques to Acoustic Survey Data. ICES C.M. 1990/D:34. 103 pp.

Anon. 1991. Report of the Workshop on the Applicability of Spatial Statistical Techniques to Acoustic Survey Data. ICES C.M. 1991/D:40. 71 pp.

Bodholt, H., Nes, H., and Solli, H. 1989. A new echo-sounder system Proc. IOA, 11(3): 123-130.

Foote, K. G., Knudsen, H. P., Vestnes, G., MacLennan, D. N., and Simmonds, E. J. 1987. Calibration of acoustic instruments for fish density estimation: a practical guide. Coop. Res. Rep. Cons. int. Explor. Mer, 144: 69 pp.

Foote, K. G., Knudsen, H. P., Korneliussen, R. J., Nordbø, P. E., and Røang, K. 1991. Postprocessing system for echo sounder data. J. acoust. Soc. Am., 90: 37-47.

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

Matheron, G. 1971. The theory of regionalized variables and its applications. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fasc. 5, Ecole Nat. Sup. des Mines de Paris. 211 pp.

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. 113 pp + 30 figs.