ICES CM 1996/ 1D:16 Ref. E

Spatial distributions of variables in marine environmental and fisheries research.

Part I: Geostatistics and autocorrelated environmental and fisheries data.

by
Boonchai Stensholt,
Institute of Marine Research, Bergen, Norway
and
Knut Sunnanå,
Norwegian Institute of Fisheries and Aquaculture Ltd, Tromsø, Norway

Abstract

The geographical information very often is lost during calculation of indices for the abundance from surveys. Isoplet diagrams showing the distribution of both fish abundance and temperature/salinity are usually produced by hand by skilled personnel. In order to find an objective approach we explored a new stochastic method for abundance calculation and isoplet diagram drawing, using the geostatistics presented in this paper. The potential of geostatistical methods is to analyse the structural pattern of spatially dependent variables, model it and use it to estimate the unknown values at the unsampled locations. Moreover it gives the variance of the estimation error at a location and the variance of the global estimated value which other methods do not give. At each location we can store the estimated values of different spatial dependent variables for the purpose of studying the relationship among them and presenting graphic plots showing the geographic distribution pattern of each variable. Furthermore the structural pattern of each variable can be used to detect the interrelationship among them. This presentation demonstrates these points with the following data : Temperature, salinity, bottom depth, and 0-group cod density in the Barents sea.

1.INTRODUCTION

When conducting surveys over geographical areas in order to study fish abundance very often the geographical information is lost during calculation of indices for the abundance. Even when measuring temperature and salinity over a geographical area, the measurements are often combined without including the geographical dimension in the result.

In some cases isoplet diagrams are presented showing the distribution of both fish abundance and temperature/salinity. Such isoplet diagrams have so far mostly been produced by hand by skilled personel.

For many years the stratified random survey approach and related methods of calculation have been the only standard procedure for calculating indices of abundance from trawl surveys. It has been clear to all working with these topics that there are strong limitations to this method. This involves both the general bias imposed by the pre-stratification of the survey area and the neglect of the variation imposed by the sampling itself.

During the later years some new approaches have been proposed, both for abundance calculation and isoplet diagrams, and among these are geostatistical procedures see(ref. 5). In this paper we will look into the potential of this approach.

In many fields of survey activity involving estimation of abundance it is an aim to describe the distribution of abundance throughout the geographical area and estimate the total quantity with a knowledge of the precision level. This also involves detecting relationships between the biological and environmental data. Among the main features of this paper is how the bottom topography influences the water movement, and the water movement influences the fish distribution. Moreover, if we find high correlation between different variables, we may use this information by treating some variables as covariates for another variable to improve the estimation of the latter.

As a case for these studies we have chosen a data series of extensive sampling, both geographically and in time, containing pelagic trawl hauls in the upper 60m watercolumn together with CTD measurements in the whole watercolumn. The data series extend from 1964 up until today, but we have chosen 1992 as a year to demonstrate the method in this paper. Work on data from other years will be continued.

We have chosen to present the work into two parts The present, first part deals mainly with the geostatistical approach and the second with model based approaches. In this paper we emphasise more on environmental data. More work will be done on different aspects of the application of spatial statistics to environmental and fishery data e.g. to estimate the stock size.

In this Part I the environmental data such as temperature, salinity and bottom depth, assumed to be highly continuous in their distribution, have been chosen as the basis. It is assumed that the measurement error can be neglected, and that the measurement therefore represents the true, and sought, distribution at the time of sampling.

Historically, when making isoplet diagrams, there has not been any desire to give the underlying digitized values and variance of the estimation error. Mostly such diagrams have been hand made by skilled and experienced personnel, but this is time consuming and may be influenced by individual subjectivity and artistic expression.

In order to study the properties of each variable and their interrelationships by means of mathematical and statistical methods, it is necessary to have digitized values. This paper presents an interpolation method (kriging) to give the required values from the observation material. Moreover, these allow us to produce diagrams fast and objectively.

The potential of kriging is to analyse the structural pattern of spatially dependent variables, model it and use it to estimate the unknown values at an unsampled location. Moreover it gives the variance of the estimation error. At each location we can store the estimated values of different spatially dependent variables for the purpose of studying the relationship among them

and for presenting graphical plots of the geographic distribution pattern of each variable. An example of this is presented in the paper, classifying watermasses according to both temperature and salinity. The area of mixing between two types of watermasses is identified and illustrated with hatching on isoplet diagrams. In our work we use ISATIS (ref. 9) and BLUEPACK 3D (ref. 2) as our geostatistical packages. We defined a 3D grid on the Barents sea and at each grid point we estimate the value of each variable from the available data by means of kriging. Thus we obtain values of different variables at the same position.

2. GEOSTATISTICS, FISHERY AND ENVIRONMENTAL DATA

We start with a brief survey of the geostatistical methods applied in the subsequent data analysis. For more detail on this and related topic can be found in (ref. 3), (ref. 8), (ref. 10), (ref. 11), (ref. 12), and (ref. 15).

2.1 Random function models

A regionalized variable in a region $D \subset R^n$ is a function z(x) of a point x, $x \in D$. For a given $x_i \in D$, $z(x_i)$ means an observable value (e.g. temperature, salinity, fish density) at x_i. We interprete "locally" $z(x_i)$ as one realization of a random variable $Z(x_i)$. Similarly we interprete "globally" the function z(x) as one realization of a random function Z(x); a random function can be seen as a set of autocorrelated random variables $\{Z(x_i)\}$, for x_i being any position within the field of study. Some applications include time dependence, and one may write Z(x,t) etc.

Typical data in the earth sciences come as a set of such observations $\{z(x_i), 1 \leq i \leq k\}$, i.e. with one realization of random function Z(x), $x \in D$. Among the regionalized variables $z(x_i)$, $1 \leq i \leq k$ a certain structure in spatial variability is usually perceptible. This is the basic observation behind geostatistical theory, and it calls for modelling a pattern of spatial correlation between the various random variables $Z(x_i)$, $1 \leq i \leq k$.

2.2 The variogram

The most crucial step in the structural analysis of a regionalized variable is a process of data analysis which leads to a variogram model. The variogram function of Z(x) defined below acts as a quantified summary of the structure of variability; it is estimated on the basis of all available information. This summary of information is then a tool in all subsequent steps of reconnaissance and estimation. The variogram function is defined by

$$2\Upsilon(x,h) = E\{[Z(x) - Z(x+h)]^2\} \qquad \text{(EQ 1)}$$

It can also be interpreted as a measure of the expectation of squared errror when Z(x+h) is used to estimate Z(x). Intuitively one may regard it as expressing how much information about Z(x) would be lost using Z(x+h) to inform about the unknown Z(x). Knowledge of the variogram can be used to minimize the loss in estimating an unobserved variable by means of neighboring observations. (This method explains only linear relationships among the data; nonlinear cases require knowledge of higher order moments.) The variogram function depends on location x, the distance |h| between location vectors x and x+h and the direction of h. In practice only one realization is available and the data will not suffice for a good estimate of a function depending on x.

However, one can treat situations where dependence on the location x can be set aside, and one may write $\Upsilon(h) = \Upsilon(x, h)$. This brings up the discussion of stationarity conditions in the next section, which again is followed by a discussion of how to estimate the variogram.

2.3 Second order stationarity[1]

From their definitions, covariance and variogram functions depend simultaneously on two points, say x_1 and x_2. Many observations of the pair of random variables $\{Z(x_1),Z(x_2)\}$ are then needed for any inference to be possible. But in a homogeneous zone of study (the relevant parameters of) these functions depend only on the difference between the two points, the vector $h=x_1 - x_2$. Thus each pair of data with location difference equal to vector h (i.e. the data points are separated by the same distance and direction as $\{x_1, x_2\}$), can be considered as a separate realization of the pair $\{Z(x_1), Z(x_2)\}$.

A random function is said to be second order stationary if

(i) The expectation $E\{Z(x)\}$ exists and does not depend on x:

$E\{Z(x)\} = m$, for all x; and

(ii) for each pair of random variable $\{Z(x_1), Z(x_2)\}$ the covariance exists and depends only on the separation vector h (modulus and direction):

$$C(h) = Cov[Z(x+h), Z(x)] = E\{Z(x+h)Z(x)\} - m^2 \quad \text{, for all x}$$

This means that the covariance of a pair is invariant under translations. If in (ii) C(h) depends only on the modulus |h| , the covariance is also invariant under all rotations; this is called isotropy.

This brief survey mainly assumes second order stationarity; it is, however, possible to adapt the development to the weaker assumption of the "intrinsic hypothesis" in case the covariance function does not exist. A random function Z(x) is called intrinsic when the variogram exists, i.e.:

(i) $E\,Z(x) = m$, for all x;

(ii) for all vectors h the increment [Z(x+h) - Z(x)] has finite variance, see (EQ1) which does not depend on x, i.e.

$$Var\{Z(x+h) - Z(x)\} = E\{[Z(x+h) - Z(x)]^2\} = 2\Upsilon(h) \quad \text{,for all x}$$

2.4 Modelling the variogram

Under the second order stationarity it possible to estimate the variogram (EQ 1) from available data. Let N(h) be the number of pairs $\{x_i,x_j\}$ of observation points with $x_j - x_i = h$. The estimator of $2\Upsilon(h)$ is an arithmetic mean:

---

1. Stronger stationarity assumptions are occationally made, like strict stationarity. see e.g. Cressie p.53.

$$(2\Upsilon)^*(h) = \frac{1}{N(h)}\sum_i [Z(x_i) - Z(x_i+h)]^2, 1 \le i \le N(h) \qquad \text{(EQ 2)}$$

A good estimate requires a large N(h); unless the data are gridded, one compromises and groups together observation pairs where the deviations of x_j-x_i from h (in modulus and direction) are within specified tolerances. As estimates by the above formula are obtained only for some discrete values of the variable vector h, techniques of smoothing and completion are then used to obtain the final variogram model. An essential requirement is to keep special theoretical properties of variogram functions. Obviously $\Upsilon(h) \ge 0$. Moreover, it is "conditionally negative definite" in the following sense:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i \Upsilon(x_i - x_j) a_j \le 0, \ whenever \sum_{i=1}^{n} a_i = 0 \qquad \text{(EQ 3)}$$

for any finite number of spatial locations $x_1,...,x_n$ and real numbers $a_1,...,a_n$. This follows from the positive semi-definiteness of covariance matrices [(ref. 3) p.86,87 and (ref. 12) p35]. In practice the modeller may choose a predefined class of functions satisfying the above requirements and nest them together to fit the estimated values as well as possible. A list of such functions can be found in [(ref. 3) p.61, and (ref. 12) p.163-171]. The expansion of (EQ 1) under second order stationarity,

$$\Upsilon(h) = Var[Z(x)] - Cov[Z(x), Z(x+h)] = C(0) - C(h) \qquad \text{(EQ 4)}$$

shows the connection between variogram and variance/covariance. If the intrinsic hypothesis holds and second order stationarity fails, one works with the variogram instead of the covariance function. [(ref. 12) p306]

In practice one estimates the variogram at different distances and directions. Analyzing the estimated variogram at different distances and directions will aid us to recover the structure of the spatial data. It can give us ideas about themes such as the variance and range of influence among the data in distance and direction, stationarity and isotropy, and non-stationarity (e.g. data with presence of a trend, or banding from periodic variation in intensity in any direction, or patchiness).

Here we will concentrate on how the variogram structure is used for the purpose of estimating unknown values at certain points from the neighboring known data. The task is to assign the right amount of influence of each neighboring datum. It is intuitively clear that this involves a delicate balance between two natural arguments: the neighbor data closer to the unobserved point should count more to the extent they are likely to bring more reliable information, but neighbors close to each other should have reduced weight to the extent that they bring the same information. Kriging is a technique which makes use of the spatial correlation information carried by the covariance/variogram (EQ 4) and the geometrical structure of the location of the known data to obtain these weigths. It constructs an estimator which is unbiased and has the minimal variance in the class of linear estimators. These concepts are introduced and discussed below.

2.5 Introduction to kriging

Kriging is a local estimation technique which provides the best linear unbiased estimator

(BLUE) for any unknown value $z(x_0)$. The information used are n available data $\{z(x_i) | 1 \leq i \leq n\}$ in a neighborhood of the location x_0 where the value $z(x_0)$ is to be estimated. (ref. 10)

The need for kriging arises e.g. when values are required on a set of regular grid points, as an intermediate step in the production of maps with isolines for the attribute observed; x_0 is then such a grid point. Another situation is the study of interrelations among different spatially dependent attributes; one then needs values for these attributes at the same location.

Here we will explain how we derive the kriging equations which can be solved to obtain the optimal solution for the weight λ_i used in the estimation equation below. For simplification we will present only the case of point kriging which suffices for contour mapping.

In (EQ 5) is explained what it means that the estimator is linear; then comes the determination of its unknown weights λ_i. First the requirement of unbiasedness puts restrictions on them in (EQ 6) - (EQ 8); in the case of "ordinary kriging" this amounts to restricting the λ_i's by the equations (EQ 9), i.e. the estimator is as shown in (EQ 10). Then the meaning of being the best (least squares error) is explained in (EQ 11) - (EQ 12); this requirement leads to the equations (EQ 13). Together (EQ 10) and (EQ 13) suffice to determine the weights and get the estimator.

One should keep in mind that this whole process must be repeated for each kriging of an unknown value $z(x_0)$ since the equations depend both on x_0 and on the selected set of neighbors $\{x_i, i=1, 2, ..., n\}$.

Each $z(x_i)$ is considered as an observed value of a random variable $Z(x_i), 0 \leq i \leq n$, and the estimator for the unknown value $z(x_0)$ is of the form

$$Z^*(x_0) = \lambda_0 + \sum_i \lambda_i Z(x_i), 1 \leq i \leq n, \qquad \text{(EQ 5)}$$

the coefficients λ_i being determined by kriging theory as described below.

The error of the kriging estimator is

$$Y = Z(x_0) - Z^*(x_0) = Z(x_0) - \lambda_0 - \sum_i \lambda_i Z(x_i) \qquad \text{(EQ 6)}$$

To get unbiased estimates, one must have EY=0, i.e.

$$0 = -\lambda_0 + E[Z(x_0)] - \sum_i \lambda_i E[Z(x_i)] \qquad \text{(EQ 7)}$$

The λ_i will be determined below for $1 \leq i \leq n$, but clearly some knowledge of the expectations is required. Second order stationarity implies that these are all equal, so

$$0 = -\lambda_0 + \left(1 - \sum_i \lambda_i\right) \cdot E[Z(x_0)] \qquad \text{(EQ 8)}$$

When this common expectation $E[Z(x_0)]$ must be considered unknown, one has to choose

$$\lambda_0 = 0, \sum_i \lambda_i = 1, \quad 1 \leq i \leq n \qquad \text{(EQ 9)}$$

The kriging method which determines the λ_i's constrained by (EQ 9), $1 \le i \le n$ is called "ordinary kriging". ("Simple kriging" assumes $E\{Z(x_0)\}$ known; there is no constraint, and λ_0 is finally determined by (EQ 8). On the other hand, "universal kriging" treats data with a trend by introducing more constraints on the λ_i's to filter out the trend.) Then

$$Z^*(x_0) = \sum_i \lambda_i \cdot Z(x_i) \; with \sum_i \lambda_i = 1, \; 1 \le i \le n \qquad \text{(EQ 10)}$$

The coefficients $\lambda_i, 1 \le i \le n$, are determined to minimize the variance of Y in (EQ 6). We rewrite Y using (EQ 9) and the stationarity which implies $E[Z(x_i)] = E[Z(x_0)], \; 0 \le i \le n$:

$$Y = \sum_i \alpha_i [Z(x_i) - E[Z(x_i)]], (0 \le i \le n), \alpha_0 = 1, \alpha_i = -\lambda_i, \text{ for } i > 0 \qquad \text{(EQ 11)}$$

$$Var(Y) = \sum_i \sum_j \alpha_i \cdot Cov[Z(x_i), Z(x_j)] \cdot \alpha_j, \; 0 \le i,j \le n \qquad \text{(EQ 12)}$$

subject to the constraints (EQ 9); in this sense we obtain the best estimator within the class of linear estimators described in (EQ 10). The first order equations

$$\frac{\partial}{\partial \alpha_i}[Var(Y) - 2\mu(1 - \sum \lambda_i)] = 0, \; 1 \le i \le n \; .$$

with Lagrange multiplier 2μ, may be written

$$\sum_j Cov[Z(x_i), Z(x_j)] \cdot \lambda_j + \mu = Cov[Z(x_0), Z(x_i)] \qquad \text{(EQ 13)}$$

The "normal system" (EQ 13) and (EQ 9) of ordinary kriging may be expressed in matrix form: Let $C_{ij} = Cov[Z(x_i), Z(x_j)]$, then

$$\begin{bmatrix} C_{11} & \cdots & C_{1n} & 1 \\ & \cdots & & 1 \\ C_{n1} & \cdots & C_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} C_{01} \\ \\ C_{0n} \\ 1 \end{bmatrix}; \; for \; short \; K \cdot \Lambda = C_0 \qquad \text{(EQ 14)}$$

The positive definiteness of the $n \times n$ covariance matrix C_{ij}, $1 \le i,j \le n$ implies the invertibility of the $(n+1) \times (n+1)$ coefficient matrix in (EQ 14)[1]. Thus (EQ 14) has a unique solution.

If second order stationarity fails but the intrinsic hypothesis holds, one can rewrite (EQ 13) equivalently in terms of the variogram function instead of covariances. [(ref. 12) p306]

Certain properties of the best linear unbiased estimator determined by (EQ 9) and (EQ 13) are now immediately visible:

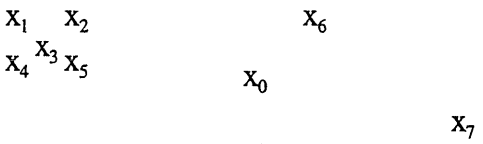
Clustered samples. The kriging method assigns weights to the sample points of the kriging

---

1. Generally let C be a symmetric nxn matrix and M an nxm matrix of rank m<n and consider the (n+m)x(n+m) block matrix $K = \begin{bmatrix} C & M \\ M' & 0 \end{bmatrix}$. Then if C is positive or negative definite, $det(K) \neq 0$, which implies K is invertible. See (ref. 1) and (ref. 4)

equation by accounting for the relationship between the unknown $Z(x_0)$ and the sample variables $Z(x_i)$, i = 1,2,...n, through C_0; and also for the degree of redundancy between sample points themselves, through K. This puts the right amount of weight that will correct the effect of clustered samples, which some other methods such as inverse distance can not do. [see also (ref. 11) p17.]

For an intuitive understanding of kriging regard 7 sample points x_1, x_2, ..., x_7 (containing a cluster) and x_0 where the kriging is done.



A sample point x_i should have a high weight λ_i if it is close to x_0 (when the variables at x_i and x_0 are highly correlated). On the other hand sample points x_i and x_j that are close together tell very much the same story and should have correspondingly small weights λ_i and λ_j. Each of the points in the 5-point cluster above should have a weight just slightly more than $\frac{1}{5}$ of the weight a single point representing the cluster area would get.

Interpolation property. Kriging is an interpolation algorithm in the sense that if $x_0=x_1$ (say), then the estimate is the "exact" value, i.e. $\sum_i \lambda_i \cdot Z(x_i) = Z(x_1)$

This is seen simply by checking that $\mu = 0$, $\lambda_1 = 1$, $\lambda_2 = \ldots = \lambda_n = 0$ solve the normal system. Thus the kriged surface passes exactly through all the experimental points. However, the kriging method may be extended to treat data with measurement errors; the real field can be obtained by filtering out the noise, thus it will not necessarily pass through these points.

Standard regression as special case. If the $Z(x_i)$ with $1 \le i \le n$ are uncorrelated, we have a case of linear auto-regression under constraint. Then (EQ 13) becomes

$$Var[Z(x_0)] \cdot \lambda_i + \mu = Cov[Z(x_0), Z(x_i)] \qquad \text{(EQ 15)}$$

Letting ρ mean correlation, the solutions of (EQ 15) and (EQ 9) are

$$\lambda_i = \rho[Z(x_0), Z(x_i)] - \mu / Var[Z(x_0)], \quad where$$

$$-\mu / Var[Z(x_0)] = \left[1 - \sum_j \rho[Z(x_0), Z(x_j)]\right] / n \qquad \text{(EQ 16)}$$

If also the n observed variables are uncorrelated to $Z(x_0)$, all weights become equal to 1/n, i.e. the estimator is the arithmetic mean. In the general case ordinary least squares regression amounts to ignoring the structural properties conveyed by covariances or variogram, and will therefore by comparison of (EQ 16) with the solution of (EQ 13) and (EQ 9) give an estimator different from the optimal one.

Precision of the "BLUE". The "ordinary kriging variance" $\sigma^2_{OK} = var(Y)$ of the error of

the optimal estimator is easily found simplifying (EQ 12) by means of (EQ 13) and (EQ 9):

$$\sigma^2_{OK} = Var[Z(x_0)] - \sum \lambda_i \cdot Cov[Z(x_i), Z(x_0)] - \mu \qquad \text{(EQ 17)}$$

This gives some idea about the precision of the estimate locally, which nonstatistical methods of interpolation do not give. The kriging variance depends on how strong the linear relation is between the estimated $Z(x_0)$ and the neighboring estimators, $Z(x_i)$, $1 \le i \le n$. 'The kriging variance gives the "order of magnitude" of the estimation error, but does not show the spatial correlation of this error, nor does it indicate how to take this error into account.' This can be perceived through simulation. (ref. 2) p.6

2.6 Cross-Validation (CV)

The cross-validation (cv) is a general procedure which checks the compatibility between a set of data and their structure model. Each one of the test data points is temporarily removed from the active data set and a reestimation $\hat{z}$ is produced by kriging using its neighboring information; then the estimate is compared to the real known value. From the estimated values we can calculate the estimation error and its distribution. We hope that the mean, variance, and skewness of the errors (residues) will all be as close to 0 as possible. The kriging procedure also produces a prediction σ^2_E of the estimation variance σ^2. So we hope that the variance of the standardized error (defined below) will be close to 1.

Let Z be the removed target value,

let $\hat{Z}$ be the estimated value, and

let error(cv) = Z-$\hat{Z}$. Then

standardized error (cv) = $\frac{Z-\hat{Z}}{\sigma_E}$,

mean error = $\frac{1}{N}\sum_N (Z-\hat{Z})$,

sample variance of error = $\frac{1}{N}\sum_N (Z-\hat{Z})^2$,

mean standardized error = $\frac{1}{N}\sum_N \left(\frac{Z-\hat{Z}}{\sigma_E}\right)$,

variance of standardized error = $\frac{1}{N}\sum_N \left(\frac{Z-\hat{Z}}{\sigma_E}\right)^2$.

It is tempting to use the variance of standardized error to adjust the sill value of the variogram. But the discussion in (ref. 8) p. 514-517 points out the reason not to do so.

2.7 What can be kriged?

'Since it is a linear combination of data, kriging can be used to estimate any quantity obtained by a linear operator applied to the initial variable.

Thus, we can obtain:

- Kriging of the average on a given support (area, volume);
- Kriging of the drift, in other words the average at each point of the random function;
- Kriging of drifts along the axes of coordinates or gradient kriging,
- Kriging of a convolution or deconvolution of the variable, for example to estimate a point value when measurements are regularized on a given support.' (ref. 6)

3.ENVIRONMENTAL AND FISHERIES RESEARCH DATA

In the study of data from fish density distributions and environmental data such as temperature and salinity the following observation is most often noted: If we grid such data and calculate the experimental variance σ_V^2 within grid cells of size V, the variance often increases continuously with V. This is a logical consequence of the existence of spatial correlations: the smaller V is, the closer the data points and hence the closer their values are. When they are sufficiently far apart, the mutual influence is negligible. And thus σ_V^2 approaches a constant which is the a priori variance of the random variable (assuming second order stationarity). So the experimental variance is practically constant once the grid side passes a certain size; this size indicates the range of influence among spatial samples in the region. However, various circumstances, in particular nonstationarity, can cause the experimental variance to keep growing with the grid size.

We may investigate the autocorrelation of the data in relation to the geographical structure of the locations by means of analysing the estimated variogram or covariances. From this we can get information about the geographic variation of the data, for example the primary variance of the data, the range of influence of data in different directions, in which direction the data have the greatest range of influence and in which direction that the values of data change fastest. The variogram may indicate various patterns of spatial variation, e.g continuity, smoothness, high fluctuation, drift, etc.

The information in the variograms can be used in planning surveys and sampling. For example from the variogram plot we find out which is the direction of the longest/shortest range of influence. We can intensify the sampling in the direction that the data seem to change faster (direction of largest gradient). In the case of 2 dimensions, instead of having a square grid design we can have a rectangular grid design where the axis directions are the directions of longest/shortest range and the rectangle sides are proportional to the ranges.

The importance of second order stationarity has been explained above. However, most fishery and oceanographic data, e.g. temperature and salinity, are nonstationary, especially when the study area is large. Mainly we are interested in drawing isolines of a certain variable, where it is clear that there must be a trend/drift in the data.

There are various ways to deal with nonstationarities. Universal kriging has been mentioned. Alternatively, we can use a variogram model for data with a trend/drift, such as a "linear model", or we may define a moving neighborhood inside which it is acceptable to treat the data as quasistationary and apply the above method locally. To find out in detail we refer to (ref. 13).

3.1 Sampling

The data were collected by 5 research vessels during the 0-group survey in the Barents Sea and adjacent water (for short 'the Barents Sea' in the rest of the paper) in August and September 1992. It was assigned to cover a certain area from west to east by sailing along NS line transects and collect samples when they reach the predefined stations (see (ref. 7)). The trawl and CTD data have been sampled along the course track at approximately 30 nautical miles distance, which is also the approximate distance between the transects. We can regard the sampling stations as covering the study area in the Barents Sea homogeneously (see figure 2 and 3). In figure 2 and 3 we show, respectively, the trawl and hydrographic station positions. For convenience the plot in figure 1 also indicates some of the 70-82 degree parallels and the 15-60 degree meridians.

All these data are temporally and spatially dependent. This may be a problem with e.g. the day/night distribution of fish density. But in this study we disregard the temporal dependence.

Of the observed data some are attached to 2-dimensional coordinates (fish density calculated from 0-group trawl data and bottom depth), and others to 3-dimensional coordinates (temperature, salinity).

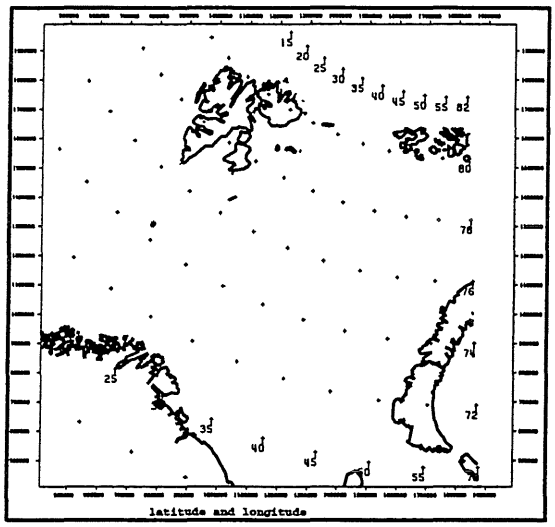


Figure 1
The plot shows positions at longitudes 15, 20, 25,....,60 degrees east and latitudes 70, 72, 74, ..., 82 degrees north in Polar Stereographic projection. In this and subsequent maps the origin of the axes is (400000m, 400000m) and the increment of the tick marks is 100,000m.

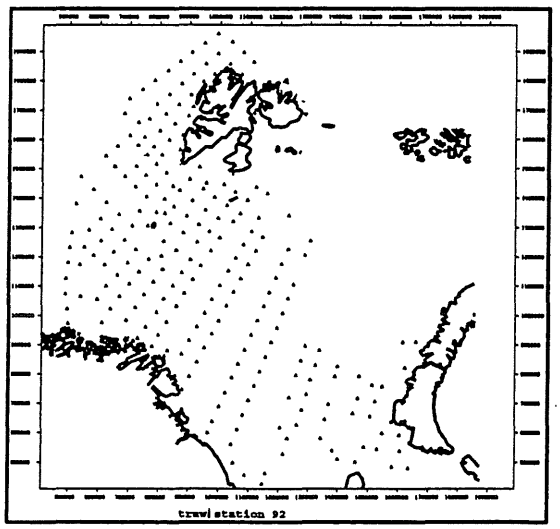


Figure 2

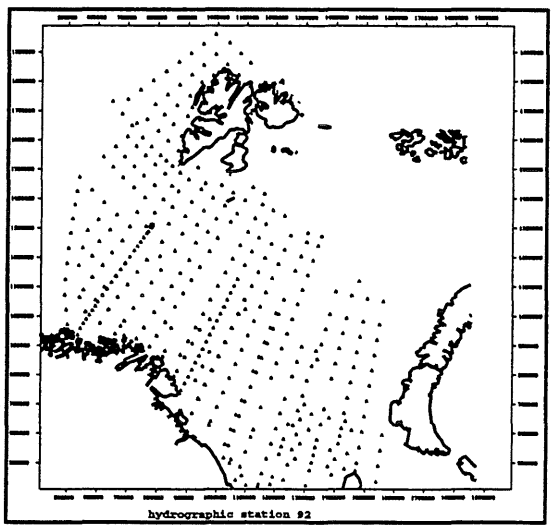


Figure 3

3.2 Gridding the data

Using Polar Stereographic projection to transform the latitude and longitude position on the global surface into the xy-coordinates on a flat surface[1] (see figure 1), the area in the Barents Sea is subdivided into a 20 km square grid, which is extended into a 3-dimensional grid with 5 meter depth intervals. The grid point data are estimated from the observed data with the use of

1. Bjørn Ådlandsvik provided the FORTRAN program.

the kriging method. At any unsampled point x_0 where an estimate is required, the kriging method is applied using the software ISATIS. Since the data cover a large area we use a moving neighborhood which is small enough to justify the assumption of second order stationarity (quasi-stationarity) (see (ref. 13)). Then all coefficients of the normal equation system may be estimated under stationarity assumption.

4. RESULTS OF THE DATA ANALYSIS

In order to prepare data for contour mapping we kriged temperature and salinity on to the 3D grid of cell size 20kmx20kmx5m, to the depth of 100 meters. Sea bottom depth, and cod fish density were kriged on to the 2D grid of cell size 20kmx20km, then attached to the 3D grid to study the interelationship between these variables. In this introductory study we will consider only these variables.

Table 1: Variogram model specification

	bottom depth	salinity	temperature	cod density
Number of valid samples	595	8463	8463	256
Angle around Z	45	0	135	135
Angle around Y	0	0	0	0
Angle around X	0	0	0	0
Number of basic structures	2	3	2	1
Structure #1	Spherical	Spherical	Spherical	nugget effect
Range	340 km	70km	330 km	
Sill	797333	0.0178	2.3	928771
Anisotropy	(1 ,1 ,1)	(1 ,1 ,1)	(1 ,1 ,2000)	
Structure #2	Spherical	Spherical	Spherical	Spherical
Range	115 km	270 km	60 km	300 km
Sill	67675	0.0658	0.3127	1440222
Anisotropy	(0 ,1 ,1)	(1 ,1 ,3500)	(5 ,1 ,1)	(1 ,1 ,1)
Structure #3		Spherical		
Range		500 km		
Sill		0.0508		
Anisotropy		(.2 ,1 ,1)		

4.1 Variogram analysis

First we must analyse the structure of the data and estimate the variogram model to be used in the kriging procedure. See table 1.

The variograms of the above data were all estimated in four horizontal directions: along 0 (x-direction) , 45 , 90 (y-direction), 135 degrees. There is a tolerance angle of ±22.5 degrees, so that the x-direction variogram, say, is influenced by variations in the sector from -22.5 to 22.5 degrees. The x-direction is roughly ENE. Also variograms for the vertical direction were fitted for salinity and temperature. These experimental variograms are based on the survey data, and therefore only reflect conditions in the surveyed areas of the Barents Sea and the structural information of data which are approximately 30 NM or more apart. The horizontal dotted line shows the primary variance of the data set, and variable values above this line suggest the existence of a drift. The topic of the following discussion is how the experimental variograms capture the known dominating features in the general pattern of the spatial structure of the studied variables. Some additional remarks indicate how the variogram models are motivated. Figure 4 shows the variogram of temperature, salinity, sea bottom depth and 0-group cod density against distance.

Bottom topography: Along the x- and the 135 degree directions the variograms rise, fall and rise again, which indicates a "sinusoidal" pattern; i.e. how these directions generally cut across banks and channels in the sea bottom. There is a drift along the y-direction.

The 45 degree variogram levels off (at about 300-350 km.), which shows stationarity in this direction.

The bottom topography strongly influences the flow direction of the water masses, generally in the 45 degree direction, as shown in figure 5 and (ref. 14) figure 1 p.6 and (ref. 18) p. 24-31.

Salinity: The experimental variograms for salinity show drift along the 45 deg.- the ' y-' and the depth directions. There is stationarity along the 'x-' and 135 deg.- direction with range about 300-350 km. We fit the mixed zonal and geometric anisotropy model to the experimental variogram. We can see from the plot of isolines of salinity that the Atlantic water mass with almost constant salinity is extremely elongated in the x-direction into the Barents Sea. This slices up the domain along x thus creating a zonation along y, which explains the additional variability in the variogram of that direction.

Temperature: The estimated variograms of the temperature along the 45 degree directions and along the depth direction have a parabolic form which is characteristic of a highly continuous. There is a drift along the x-, 45 deg.- and depth directions. We chose nested spherical variogram models over gaussian variogram model because they are more stable and give realistic estimates over the whole region and the results from crossvalidation are satisfactory. Since the variogram appears stationary along the y- direction and the 135 degree direction, we use the maximum value of variogram in the 135 degree direction as a guide-line for fitting the sill in the variographic model. The observed range of variogram in 135 degree direction is about 300-350 km. So we use two nested spherical models with 60 and 330 km range.

Cod density: The variograms of cod density show drift in all directions and a large variability. This is due to a very high density in one area. Again the variability along the 135 degree direction is lower than along other directions and the variogram reaches the sill at around 300-350 km range (where the variogram levels off). The downward slope of the variogram at larger distance can come from the single-peaked shape of the cod density distribution: When the observations are made in august/september, the O-group has already been transported by

the Atlantic current into the Barents Sea and the waters west of Spitzbergen. The distribution of the O-group is not uniform. Isolines show a concentration around the Bear Island, and this single-peak pattern is captured by the variogram.

Similarities and differences. The variograms of temperature, salinity, bottom depth, and cod density show some similarities and some differences.

Let us discuss temperature and salinity first. Similarities must be expected, as both temperature and salinity are linked to the bulk of Atlantic water masses. Both have the similar variogram pattern on y, 45, 135 degree and depth direction. However, the salinity and temperature are changed by different physical factors. The movement into the Barents Sea with a colder climate causes a gradual temperature drop, while the salinity remains constant for a longer distance. Thus the two x-direction variograms are very different.

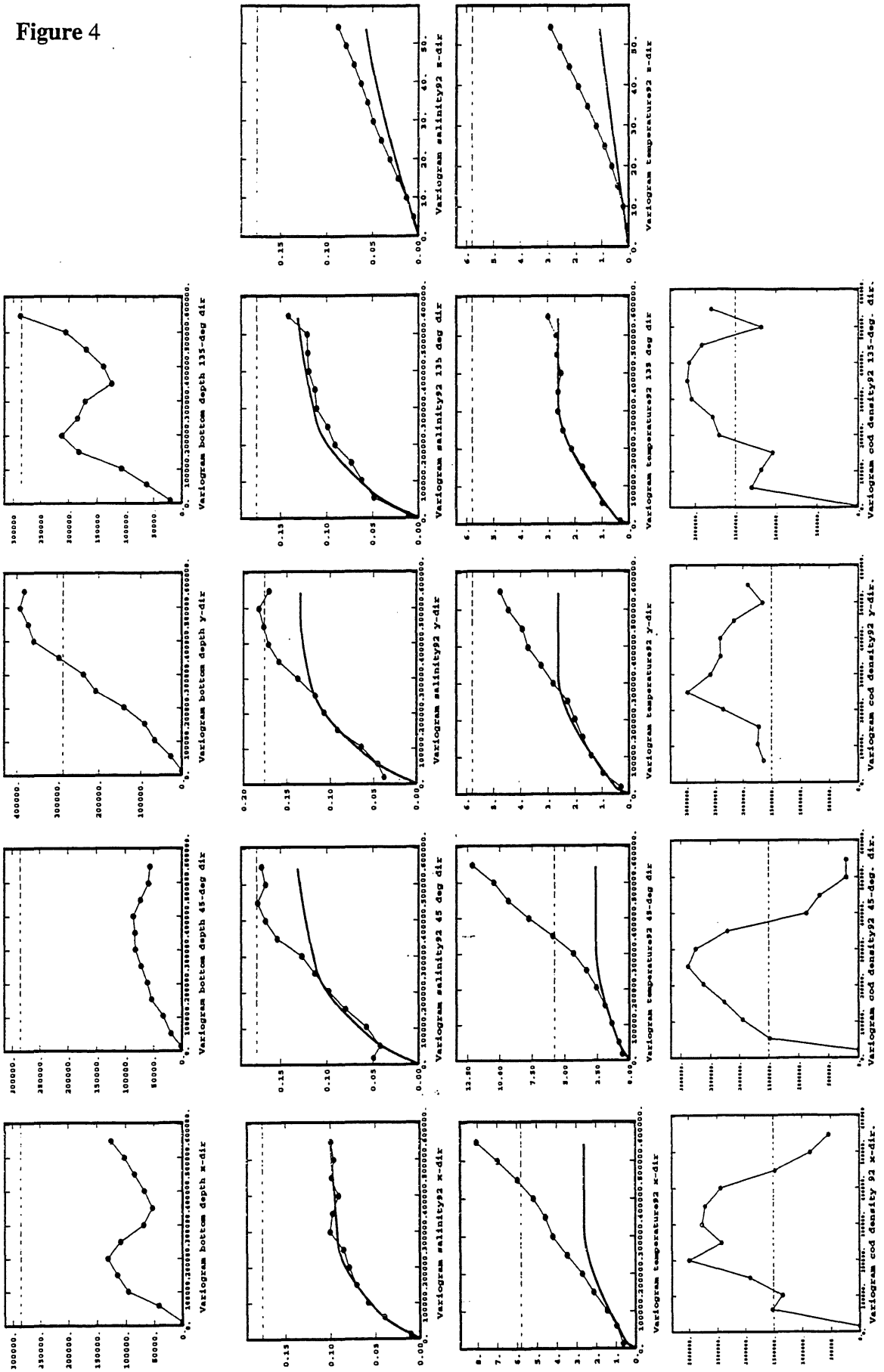
These considerations link both temperature and salinity to the bottom topography. The bottom depth is stationary in the 45 deg. direction. The Atlantic water masses generally move northeast and confront the Arctic water masses moving southwest. That explains why the variograms of temperature and salinity show drift along the 45 degree direction and stationarity in the perpendicular (135 deg.) direction.

The cod density variogram along the 135 degree direction has the lowest variability; this is the direction of stationarity for salinity and temperature. The ranges also match well. On the 45 degree direction there is a drift similar to the temperature and salinity variograms. This reflects that the distribution pattern of the 0-group cod is mainly influenced by the current.

Neigborhood specification

Because the estimation area is very large and trends exist, we use the ordinary kriging with the moving neighborhood method in the estimation. A stationarity neighborhood of radius 150 km. was chosen (with 15 m depth in 3D case), which seems a reasonable size in the present geographical conditions and the variographic analysis of each variable. In each neighborhood should be at least 4 data points, while 20 are considered optimal.

Figure 4


Variogram bottom depth x-dir
Variogram bottom depth 45-deg dir
Variogram bottom depth y-dir
Variogram bottom depth 135-deg dir
Variogram salinity92 x-dir
Variogram salinity92 45 deg dir
Variogram salinity92 y-dir
Variogram salinity92 135 deg dir
Variogram temperature92 x-dir
Variogram temperature92 45-deg dir
Variogram temperature92 y-dir
Variogram temperature92 135 deg dir
Variogram cod density 92 x-dir.
Variogram cod density92 45-deg. dir.
Variogram cod density92 y-dir.
Variogram cod density92 135-deg. dir.

4.2 Kriged isolines

The kriging was done with the software ISATIS. We estimate the temperature and salinity in 3-dimensions (3D) using a 3D moving neighborhood. The horizontal kriged isolines for temperature and salinity at depths of 50 meters and 100 meters are shown in figure 5 overlaid on the hatched area showing the bottom topography of the Barents Sea. The distribution of inflowing warm water from the Atlantic is obviously influenced by the bottom topography. The Atlantic water confronts cold water coming from the Arctic Sea (in the north and north east) and a small amount from the Kara Sea (south of Novaja Zemlja). In the south there is also a coastal current and an outflow of cold fresh water from several rivers, eg. the Pechora. [see (ref. 14)]

Figure 6 shows isolines in six vertical sections perpendicular to the x-axis. The sharp transition between Atlantic and Arctic water masses , i.e. the polar front, appears clearly. The Atlantic water usually has salinity above 35.0 and temperature above 3 $°C$, while Arctic water is fresher and colder, at least down to 100 m.

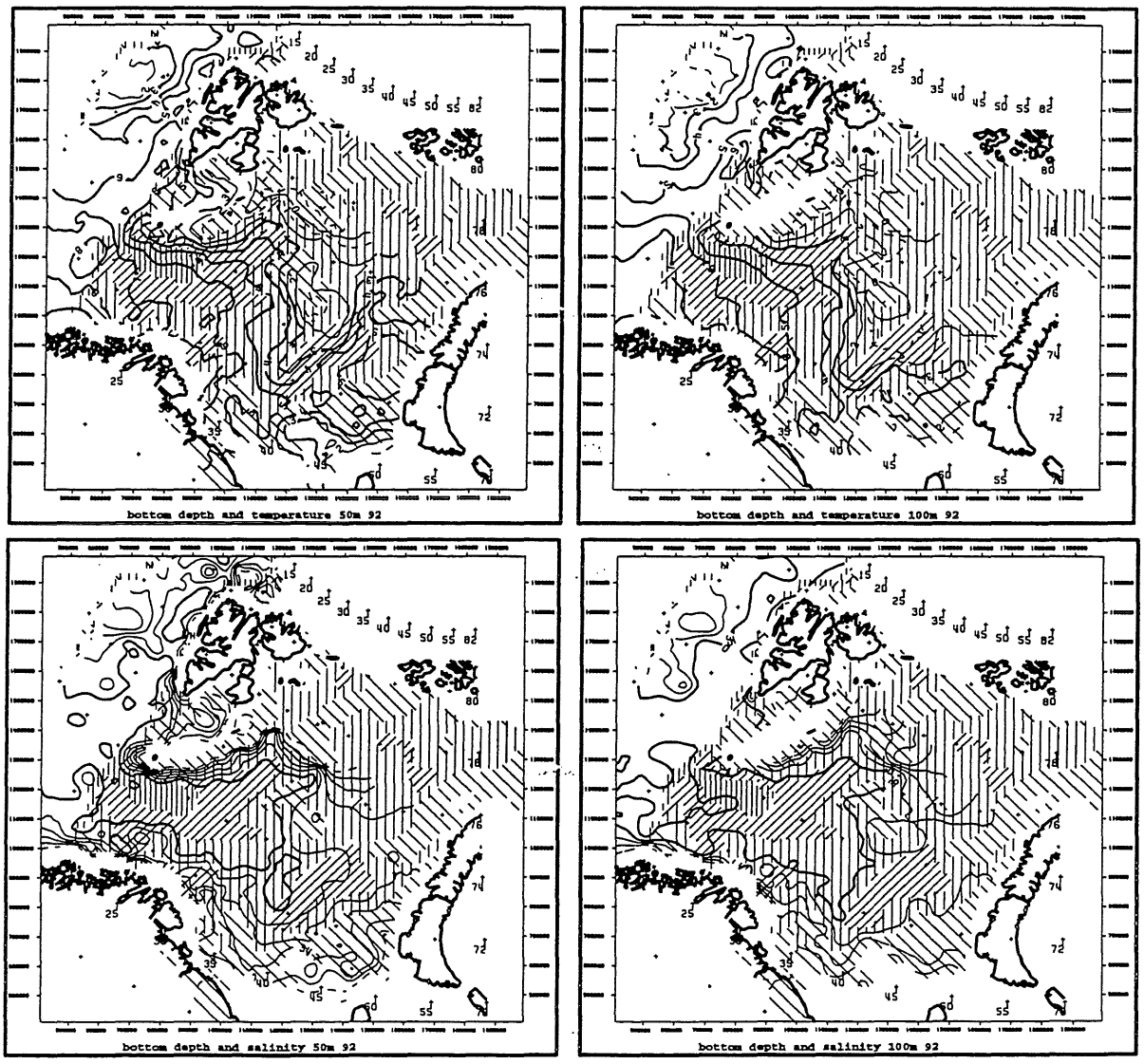
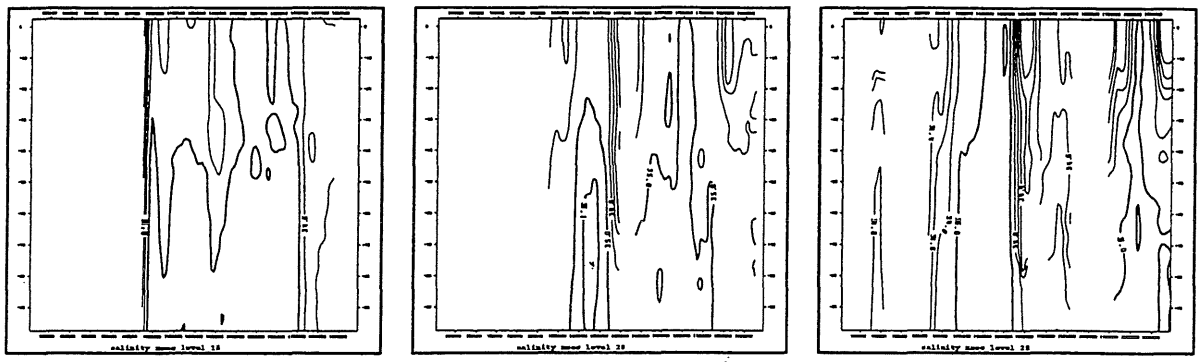


Figure 5:The hatched area of sea bottom depth |||| 400-500 m., //// 300-400 m., | | | | 200-300 m, . \ \ \ \ 100-200 m.
The isoline of temperature; _ _ _ -2,0,2, _____ -1,1, _____ 3,4,5,6,7,...

The isoline of salinity at step of 0.1; _ _ _ 34.0-34.5, _____ 34.6-34.9, _____ 35.0-35.2



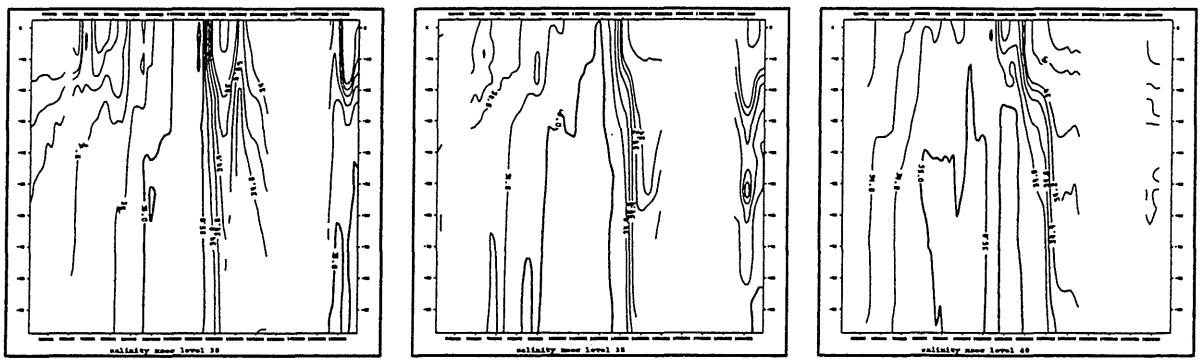


Figure 6: Isolines of salinity in vertical section perpendicular to x-axis of figure 5. The observer looks from east to west, so the north is on the right hand side.
The depth is from 0 to 100 meters, with the ticks marking steps of 10 meters.
The isoline of salinity at step of 0.2; ____ 34.0-34.8, at step of 0.1 _____ (double thickness line) 35.0-35.2

The polar front is an important boundary between Atlantic and Polar water. At the same time it is a very important area for the phytoplankton production [see (ref. 18) p.94-95]; this is the basis for the zooplankton, which again is the main food source for the 0-group fish.

The distribution of various water masses and the boundaries where Atlantic and Arctic water masses mix are shown in figure 2 of (ref. 14). Table 1 of (ref. 14), part of which is given in table 2 below, gives some characteristics of the water masses.

Table 2: Water masses in the Barents Sea

Name of water masses		Temperature, $^{\circ}C$	Salinity
North Atlantic Water	NAW	>3.0	>35.0
Arctic Water	AW	<0.0	34.3-34.8
Polar Front Water	PW	-0.5-2.0	34.8-35.0

A digitized polar front definition: Here we use gradient kriging (see section 2.7) and the characteristics of water masses in table 2 to define the polar front area. Computers can give a quick and objective information on the values of salinity, temperature, bottom depth, and the

lenght of their gradients. The "robotic" nature of the computer and the limitation of the data can give weird results; with e.g. a pure temperature gradient definition the computer may declare the boundary between Atlantic and Coastal waters as "polar front". By adding more criteria we get a rough but objective polar front area, which, if needed, may be improved by human editing. With more work on this, and perhaps with more available data, one may hope to improve the digitized definition.

In the present paper the lenght of the horizontal gradients of temperature and salinity were calculated and a high length value used as one criterion in defining the polar front area. We investigate the plot of isolines along the depth profile and the horizontal profile where the polar front occurs. We can see that the salinity is the best indication of the difference between the two water masses.

Here the polar front area is given as the intersection $GT \cap GS \cap S \cap NN$ where

GT is the area where the length of the horizontal temperature gradient is above K_T percentile (here we choose $K_T = 70$).

GS is the area where the length of horizontal salinity gradient is above K_S percentile (here we choose $K_S = 52$).

S is the area where the salinity is above 34.6.

NN is the area above the 1000000 along the y-axis. (During the survey period the polar front is well within NN)

These criteria will give an objective map of the polar front area. The isolines of temperature, salinity of 35.0 are shown in figure 7 for the depth of 50 and 100 meters, overlaid on the hatched area of the polar front.

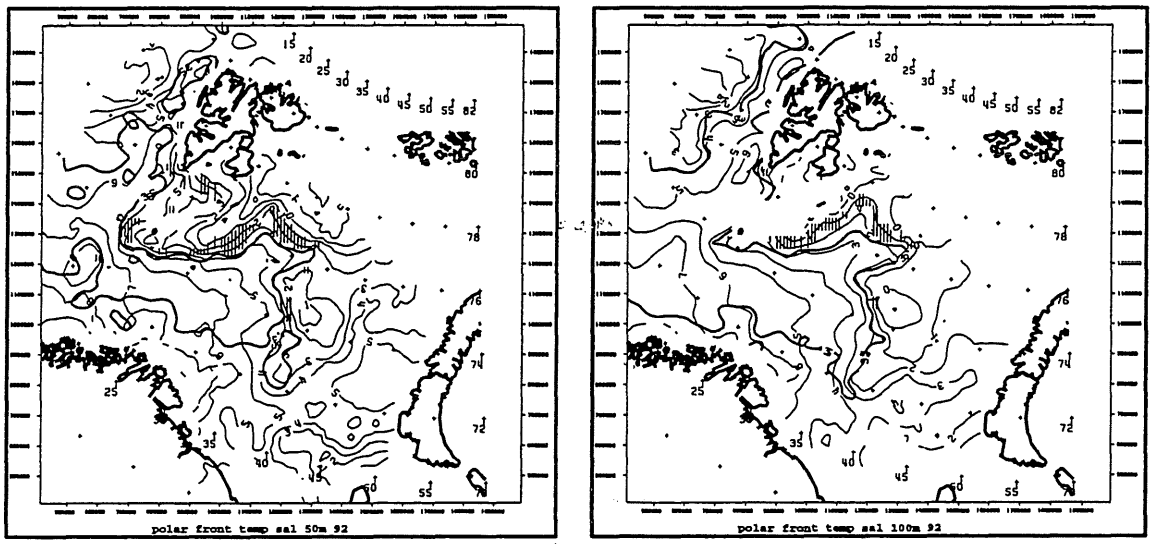


Figure 7: The isolines of temperature at step of 1 degree C _ _ _ (single thickness line) -2 - 12 degree C and of salinity at 35.0 _____ (double thickness line)
and the hatched area of polar front at 50 meters depth and 100 meters depth (year 1992).

Figure 8 shows isolines of the bottom depth, salinity isoline at 35.0, and hatched area of polar front at 50 and 100 meters depth. The salinity isoline obviously conforms to the bottom topography, which shows that the movement of water is influenced by the bottom topography. As a consequence the polar front area, which is defined by the properties of salinity and temperature alone, match with the area of high gradient of bottom depth.

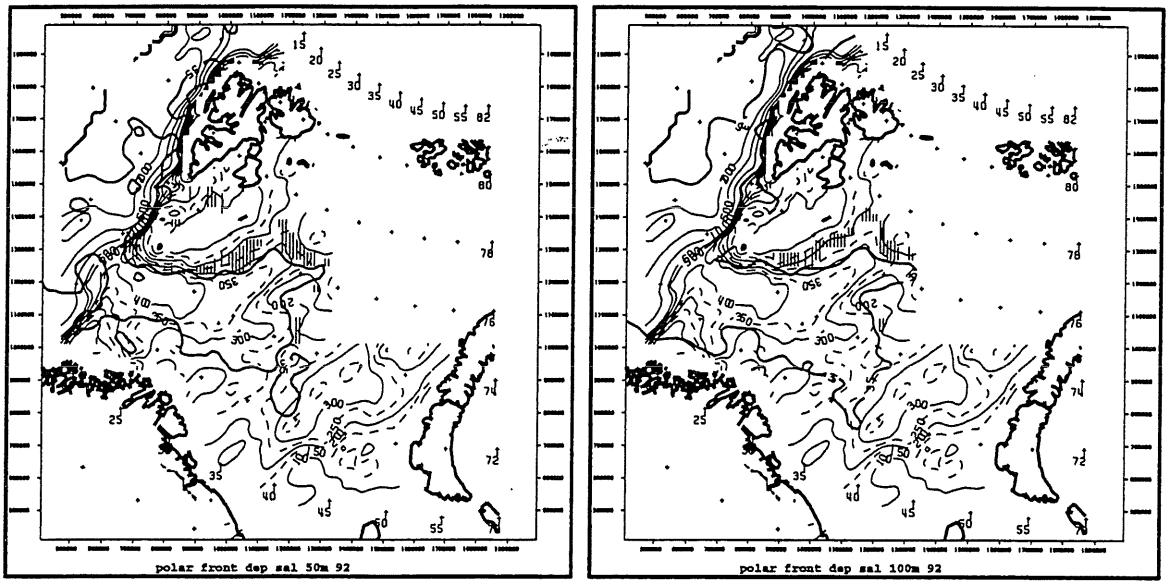


Figure 8: The isolines of bottom depth at step of 100 m.; _____ 100-600, at step of 50 m. _ _ _ 150-550 and of salinity at 35.0 _____ (double thickness line) and the hatched area of polar front at 50 m. and 100 m. depth.

4.3 Precision of the kriging method

One of the good sides of kriging method is that it also gives the estimation variance. See (EQ 17) in the case of ordinary kriging. In figures 9 and 10 we draw isolines of the estimation standard deviation σ_E. We can see that within the sample area σ_E never exceed 1 $^\circ C$ for the estimation of temperature at 50 m. and 100 m. depth. Also σ_E never exceed 0.2 for the estimation of salinity at 50 m. and 100 m. depth.

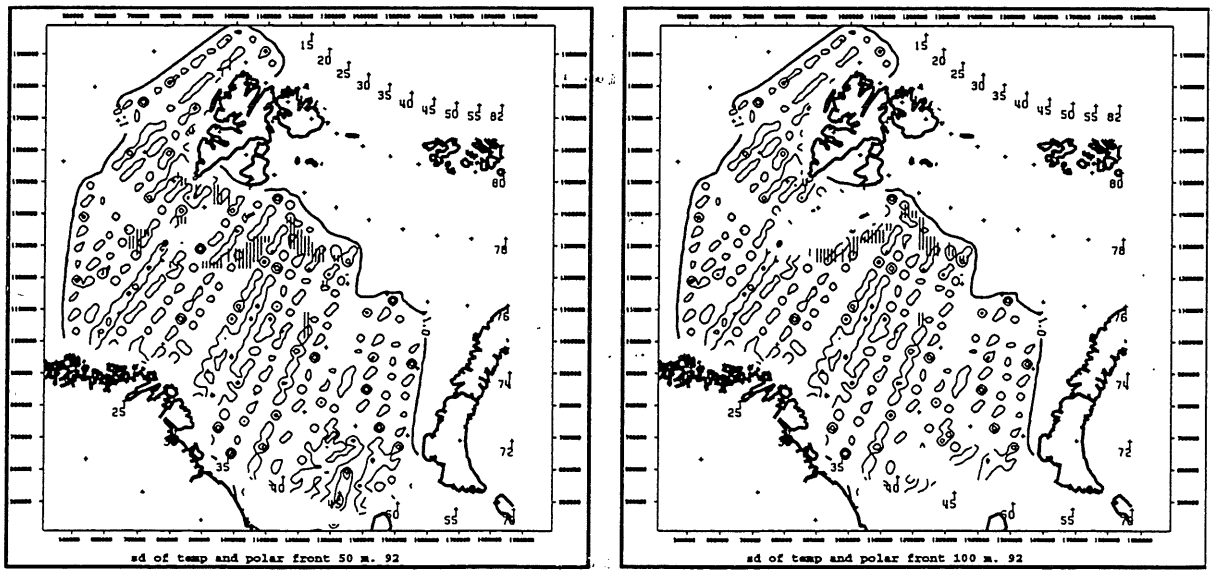


Figure 9: The isoline of estimation standard deviation for estimates of the temperature at 50 and 100 meters depth by moving neighborhood kriging method. The double thickness lines show sd. = .5 and 1. The single thickness show sd. = .75. The hatched area shows the polar front.

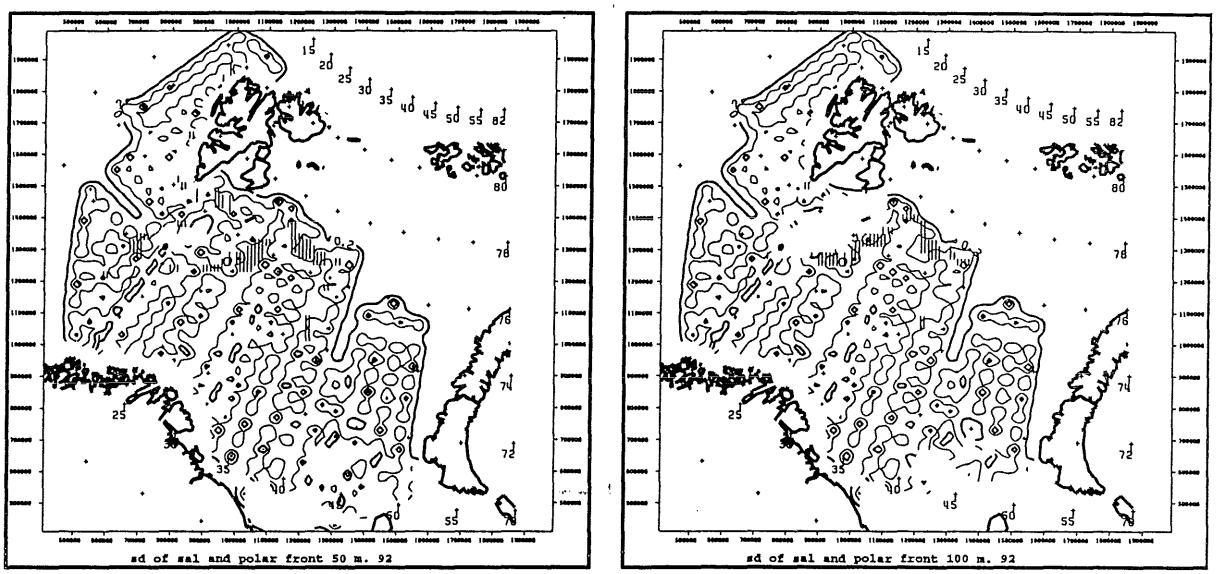


Figure 10: The isoline of estimation standard deviation for estimates of the salinity at 50 and 100 meters depth by moving neighborhood kriging method. The double thickness lines show sd. = .1 and .2. The single thickness show sd. = .15. The hatched area shows the polar front.

As described in Section 2.6 we present in tables and figure some results of the cross-validation in this section. The calculations were done with ISATIS. The histogram of errors (not included) shows a symmetric (non-normal) distribution with a high peak around 0. The number column in e.g. table 3 shows that there were 8463 sample points, of which 8341 give robust data (a datum is robust when its standardized error lies between -2.5 and 2.5.), and that there were 74855 grid points.

The plot of + and - in figure 11 shows result from cross-validation respectively under- and over estimated the removed value of cod-density.

Table 3: Temperature

	number	minimum	maximum	mean	s.d.
temperature Z	8463	-1.726	11.68	5.19	2.407
kriged temperature at sample point $\hat{Z}$	8463	-1.70	11.68	5.19	2.37
error (cv)	8463	-2.079	2.16	0.00025	0.2
std. error (cv)	8463	-7.11	6.83	0.0056	0.69
robust error (cv)	8341			0.00019	0.0206
robust std. error (cv)	8341			0.00022	0.233
kriged temperature on to 3D grid	74855	-1.71	11.46	4.82	2.56
σ_E	74855	0.22	1.76	0.94	0.27

Table 4: Salinity

	number	minimum	maximum	mean	s.d.
salinity Z	8453	30.69	35.153	34.731	0.419
kriged salinity at sample point $\hat{Z}$	8453	31.1	35.15	34.73	0.408
error (cv)	8453	-2.065	1.18	0.00023	0.071
std. error (cv)	8453	-14.4	21.28	0.001	0.91
robust error (cv)	8234			-0.00113	0.033
robust std. error (cv)	8234			-0.01504	0.456
kriged salinity on to 3D grid	74819	30.73	35.17	34.65	0.450
σ_E	74819	0.03	0.38	0.187	0.067

Table 5: 0-group cod density

	number	minimum	maximum	mean	s.d.
cod density Z	256	0	11244.1	375.618	1226.01
kriged cod density at sample point $\hat{Z}$	256	-26.1	3539.21	376.315	655.705
error (cv)	256	-2775.16	9263.8	0.69771	1091.03
std. error (cv)	256	-7.89	2.367	0.00001	0.93
robust error (cv)	251			118.25	633.34
robust std. error (cv)	251			0.1	0.537
kriged cod dnsity on to 2D grid	3240	-36.83	4749.56	252.366	578.285
σ_E	3240	1101.58	1616.71	1228.84	112.545

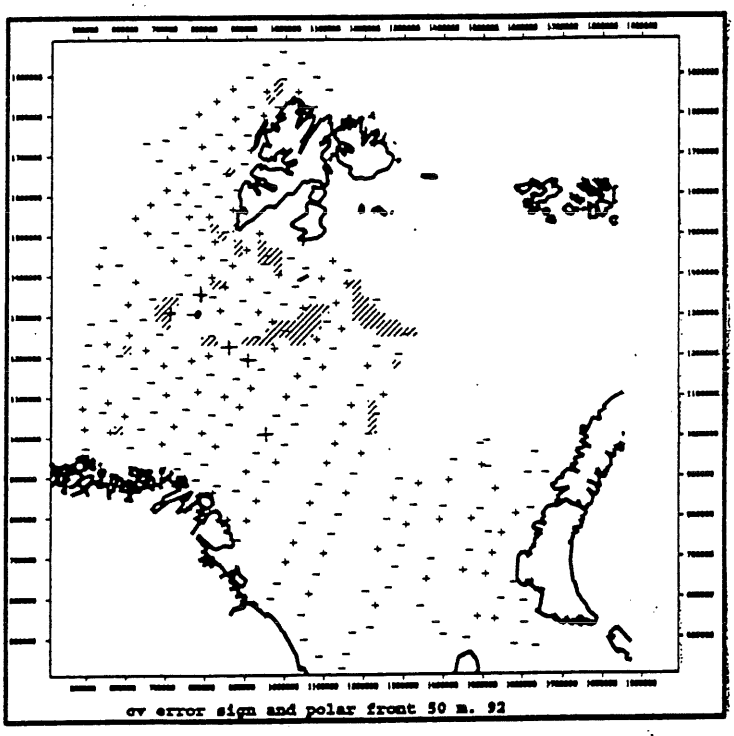


Figure 11:
The result from cross-validation of 0-group cod density.
+ marks station where observed value is larger than estimate,
- marks station where observed value is smaller than estimate, large symbols show outliers.
It seems that overestimation is more common than underestimation along the boundery where the density is low.
All 5 outliers are cases of underestimation, and occur where the density is high.

4.4 Environmental habitat of 0-group cod.

Since the 0-group fish drift along with the warm current, a study of the temperature and salinity profile in the Barents Sea helps us to identify where we can expect to find high 0-group concentrations.

Isolines of the 0-group cod density superimposed on the hatched area of the polar front and the dashed isoline showing salinity at 35.0, at 50 meters depth level are shown in figure 12. It is seen that the concentration is high in the neighborhood where the warm Atlantic water (carrying the 0-group from the spawning grounds) confronts the cold arctic water. It seems that the location of coast line and islands and the bottom topography of the Barents Sea are important factors in guiding the water movements, which is intern inflence the distribution of 0-group cod.

This is well known among fishery scientists, but our purpose is to demonstrate the usefulness of geostatistical methods in fishery science. It is a verification of its reliability that it captures important well known results. Geostatisics also give objective numerical values for calculating the interrelations between the variables. The real ecosystem, however, is too complicated to be explained by a few variables or data from a single year.

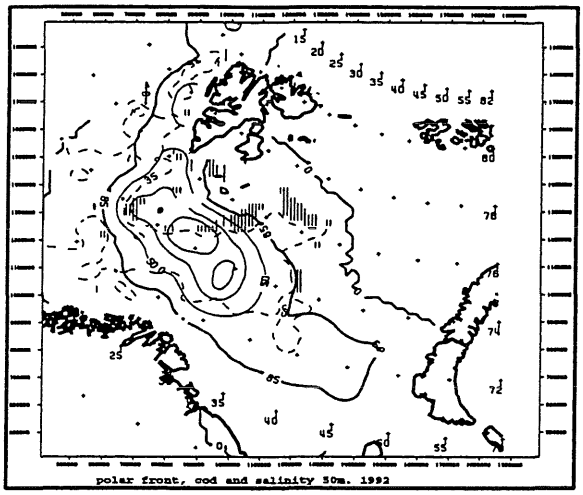


Figure 12: The spatial density distribution of the 1992 0-group cod with hatched area showing the polar front and dashed isoline showing salinity of 35.0 at 50 meters depth.
Double thickness isoline shows 85 fishes / n. mile trawled, single thickness shows 500, 1500, 2500 fishes / n. mile trawled.

The block estimate can be done for each variable. We can calculate the percentage of 0-group cod experiencing a certain level of temperature and salinity. In this calculation we assume the 0-group cod is moving from surface down to 100 meters depth, the total volume of water we use in this calculation is 120488 cubic kilometers (total volume of water in the Barents Sea is about 322000 cubic kilometers (ref. 18) p. 24).

Let

P = percentage of 0-group cod

s = salinity categorized as in table 6 into 9 levels

t = temperature categorized as in table 6 into 12 levels

c_{st} = number of 0-group cod in salinity level s and temperature level t

C = total number of 0-group cod

$$P = \sum_{s=1}^{9} \sum_{t=1}^{12} \frac{c_{st}}{C} \times 100$$

This subsection was suggested by Odd Nakken who also provided the categories for temperature and salinity. From table 6 one can see that 93 % 0-group cod live in the water of temperature between 3-<9 °C, 80% live in salinity between 34.8-<35.2. 50 % of the 0-group cod live in salinity of 35.0-<35.2 and temperature 3-<9 °C (Atlantic water, see table 2).

Table 6 gives estimated values in percent of the total amont of cod found in each temperature salinity stratum.

Table 6: Percentage of 0-group cod distributed in different categories of temperature and salinity.

salinity	32-<34	34-<34.5	34.5-<34.6	34.6-<34.7	34.7-<34.8	34.8-<34.9	34.9-<35.0	35.0-<35.1	35.1-<35.2	
tempera-ture	percent of cod									Total
-2-<0	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.03
0-<1	0.02	0.14	0.07	0.12	0.05	0.05	0.05	0.00	0.00	0.50
1-<2	0.09	0.41	0.12	0.13	0.16	0.29	0.39	0.00	0.00	1.60

Table 6: Percentage of 0-group cod distributed in different categories of temperature and salinity.

salinity	32-<34	34-<34.5	34.5-<34.6	34.6-<34.7	34.7-<34.8	34.8-<34.9	34.9-<35.0	35.0-<35.1	35.1-<35.2	
tempera-ture	percent of cod									Total
2-<3	0.23	0.76	0.30	0.31	0.39	0.44	1.19	0.28	0.00	3.88
3-<4	0.28	1.59	0.30	0.35	0.80	1.36	2.39	3.24	0.00	10.31
4-<5	0.61	1.69	0.44	0.66	0.90	1.45	2.79	11.52	0.00	20.06
5-<6	0.36	1.12	0.61	0.74	1.08	1.51	2.63	16.73	1.07	25.84
6-<7	0.02	0.53	0.57	0.76	1.13	2.03	2.78	7.97	0.69	16.48
7-<8	0.00	0.14	0.20	0.30	0.71	1.73	2.54	6.51	0.47	12.59
8-<9	0.01	0.20	0.12	0.15	0.20	0.96	4.02	2.26	0.00	7.93
9-<10	0.00	0.02	0.03	0.06	0.16	0.19	0.31	0.00	0.00	0.77
10-<11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
total	1.63	6.61	2.75	3.57	5.57	10.03	19.10	48.51	2.23	100

Table 6 is supplemented by table 7 which gives the percentages of water (by volume) in the same temperature/salinity intervals. One may see, e.g., that 73% of the water has temperature 3-9 $°C$, 58% has salinity 34.8 - 35.2, while 45.6% has both of these properties.

Table 7: Percentage of water falling into different categories of temperature and salinity.

salinity	30-<32	32-<34	34-<34.5	34.5-<34.6	34.6-<34.7	34.7-<34.8	34.8-<34.9	34.9-<35.0	35.0-<35.1	35.1-<35.2	
tempera-ture	percent of water										total
-2-<0	0.00	0.29	2.12	0.74	0.39	0.04	0.31	0.01	0.00	0.00	3.90
0-<1	0.00	0.21	0.62	0.42	0.45	0.60	0.88	0.63	0.00	0.00	3.80
1-<2	0.03	0.80	0.80	0.27	0.60	0.88	1.45	2.24	0.06	0.00	7.13
2-<3	0.15	0.76	0.61	0.25	0.71	0.91	1.60	2.86	0.90	0.00	8.75
3-<4	0.02	0.66	0.88	0.34	0.73	1.80	2.87	2.42	2.02	0.00	11.74
4-<5	0.00	0.58	1.43	0.89	1.21	1.83	3.37	2.55	4.26	0.00	16.12
5-<6	0.00	0.61	1.96	1.26	1.75	2.10	1.94	1.83	5.15	0.28	16.87
6-<7	0.00	0.40	1.52	0.65	0.89	1.08	1.71	2.40	3.83	0.64	13.12
7-<8	0.00	0.12	1.11	0.48	0.43	0.46	0.68	1.85	3.90	0.32	9.36
8-<9	0.00	0.60	1.01	0.19	0.27	0.38	0.74	1.72	1.14	0.00	6.06
9-<10	0.00	0.02	0.25	0.17	0.31	0.40	0.64	0.46	0.08	0.00	2.33
10-<11	0.00	0.09	0.33	0.07	0.09	0.05	0.13	0.03	0.00	0.00	0.80
11-<12	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
Total	0.21	5.16	12.66	5.74	7.81	10.53	16.32	18.99	21.34	1.24	100.00

Since the density of the 0-group cod has very large variance we used BLUEPACK and SAS to calculate and plot the conditional expectation curve for pairs of variables, temperature and cod, salinity and cod, sea bottom depth and cod. See figure 13. This will give us some idea how, on average, the density of 0-group cod distributes in relation to the value of those other variables.

The conditional expectation method classifies the conditioning variable (in this case the value of temperature, salinity, bottom depth) into classes and in each class one finds the mean value of the conditioned variable (in this case the density of 0-group cod).

This also demonstrates that the relationship between them are not linear. This explains why the correlation coefficients which concern only the linear relationship are not significantly high. See table 8.

Most high density cod, e.g. more than 250 fishes/1 nm trawl, live in water of temperature between 4-8.5 $°C$ and bottom depth deeper than 280 m and salinity above 34.9.

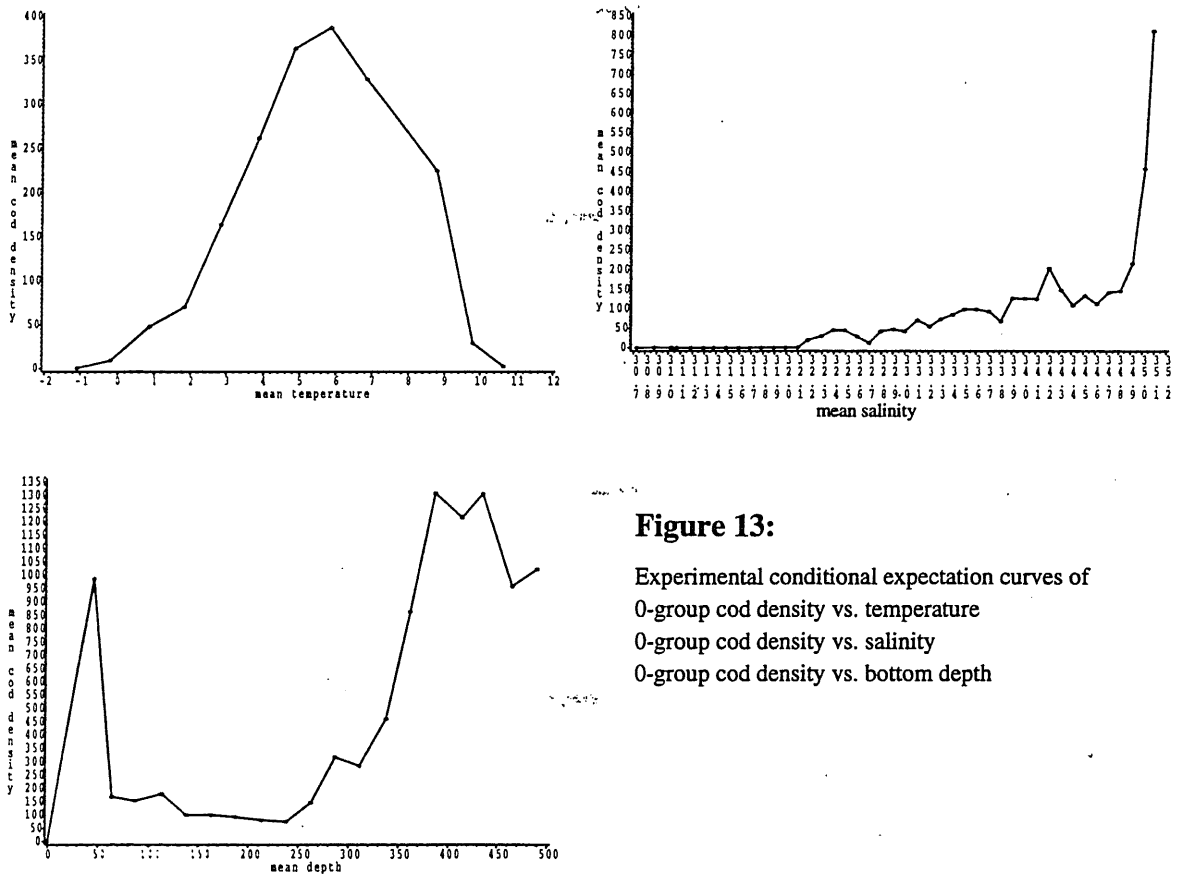


Figure 13:

Experimental conditional expectation curves of
0-group cod density vs. temperature
0-group cod density vs. salinity
0-group cod density vs. bottom depth

Table 8: Correlation coefficients

	temperature	salinity	cod density	bottom depth
temperature	1	.15	.15	.30
salinity		1	.17	.41
cod density			1	.33
bottom depth				1

With kriging we can obtain a total estimate of the 1992 0-group stock in the Barents Sea and adjacent waters of 6.6945019×10^9 individuals without correction for catchability. [See also (ref. 16) p. 193 and p. 195; acoustic estimates of the stock were 56×10^9 around Svalbard, and 51×10^9 in the Barents Sea. 'Asuming similar catching efficiencies and using the same formula as Sundby et al. (1989) for the computation of total numbers of 0-group cod in 1992, estimates of 19×10^9 and 8×10^9 ']. Such values present other indices for the 0-group stock in studying the development of the stock over time. This will be discussed in more detail in the next report.

5. DISCUSSION

Fishery research is concerned with spatial observations of several "attributes" such as fish density distribution, food source distribution and environmental variables and particularly with their interrelations. Therefore they must be treated simultaneously, which calls for an extended model where the regionalized variable is vector valued and a for a study of the cross correlations between the vector components.

Kriging may therefore be extended to cokriging. This also allows us to take into account cheaper, hence more plentiful data which have lower quality. Thus trawl data may be supplemented by acoustic data; a drawback of the latter is that the acoustic observations cannot distinguish between different species. This will be presented in the second paper.

The distribution pattern of the 0-group cod is influenced by environmental variables in a complicated way: the bottom topography, earth rotation, etc all influence the movement of different water masses which in turn influence pattern of 0-group cod density distribution. The study of variograms illustrated some of these influences. One result is the creation of an area called the polar front. This polar front area has influenced the ecosystem in the Barents sea to a large extent. The movement of polar front area also determines the amount of resource habitat for several species.

In this paper we found no significant direct linear correlations between the variables at the same location. As illustrated in figure 13 the relationships are nonlinear. What we found is an explainable relationship in the distribution pattern of bottom topography, temperature, salinity and 0-group cod density. This shows that all these variables are not related to each other in a way expressible by correlation coefficient values, but are interconnected in ways that are captured by variograms and isoline maps. This exhibits a strong side of geostatistics: it can be used to analyze the structural patterns of spatially dependent variables and take this information into account in the estimation procedure.

Many of the results above are well known among fishery scientists, but our purpose is to demonstrate the usefulness of geostatistical methods in fishery science. A verification of its reliability is that it captures important well known results. Moreover, geostatistics offer several advantages; It gives numerical estimates and also numerical values to the precision of the estimates. This allows us to make objective analysis and conclusions from the data.

Acknowledgements: Thanks to Asgeir Aglen, Odd Nakken and Bjørn Ådlandsvik for valuable discussions and to an excellent computer system administrator Bjørn-Erik Gjerde.

REFERENCES

1. Black, J and Morimoto, Y, May 1968, A note on quadratic forms positive definite under linear constraints, Economica.
2. BLUEPACK 3D the Centre de Geostatistique of the Ecole des Mines de Paris.
3. Cressie, Noel A.C. 1991 Statistics for Spatial Data, John Wiley & Sons.
4. Farebrother, R.W., 1978, Necessary and sufficient conditions for a quadratic form to be positive whenever a set of homogeneous linear constraints is satisfied, Linear Algebra and its Applications, 16.
5. Foote, K. and Stefansson, G.,1993, Definition of the problem of estimating fish abundance over an area from acoustic line-transect measurements of density, ICES J. mar. Sci., 50: pp. 368-381.
6. Geovariances, July 1996, Which Kriging ?, The Natural Resources Managers.
7. ICES Preliminary Report of the International 0-group Fish Survey in the Barents Sea and Adjcent Waters in August-September 1992, Demersal Fish Committee, C.M. 1992/G:82 Ref.H.
8. Isaaks E.H. & Srivastava R.M., 1989, An Introduction to Applied Geostatistics, Oxford University Press.
9. ISATIS TRANSVALOR and GEOVARIANCES Avon France (1993,1994)
10. Journel A.G. 1983, Geostatistics. Pp 424-431 in Kotz,S. and Johnson N.L. (eds.) Encyclopedia of Statistical Sciences vol.3 John Wiley & Sons, Inc.
11. Journel, A.G. 1989, Fundamentals of Geostatistics in Five Lessons Amer. Geoph. Union.
12. Journel, A.G. & Huijbregts Ch.J 1978, Mining Geostatistics Academic Press London, 600 p.
13. Journel, A. G. and Rossi, M.E. 1989, When Do We Need a Trend Model in Kriging? Mathematical Geology, vol. 21, No. 7, pp 715-739.
14. Loeng, H. 1991, Features of the physical oceanographic conditions of the Barents Sea. Pp. 5-18 in Sakshaug, E. et al. (eds.) Proceedings of the Pro Mare Symposium on Polar Marine Ecology, Trondheim, 12-16 May 1990. Polar Reseach 10(1).
15. Matheron, G., 1971, The Theory of Regionalized Variables and Its Applications: Fasc. 5, Paris School of Mines, 212 p.
16. Nakken, O. et al. Acoustic estimates of 0-group fish abundance in the Barents Sea and ajacent waters in 1992 and 1993, Proceedings of the sixth IMR-PINRO Symposium Bergen, 14-17 June 1994, p. 187-197.
17. Ripley B.D. 1988, Spatial Data Analysis. Pp 570-573 in Kotz,S. and Johnson N.L. (eds.) Encyclopedia of Statistical Sciences vol.8 John Wiley & Sons, Inc.
18. Sakshaug, E. et al., 1992, Økosystem Barentshavet, Pro Mare, 304 p.