Geostatistics for Estimating Fish Abundance

J. Rivoirard,* J. Simmonds,† K.G. Foote,‡
P. Fernandes† and N. Bez*

**Centre de Géostatistique de Fontainebleau,*
Ecole des Mines de Paris, France
†FRS Marine Laboratory, Aberdeen, UK
‡Institute of Marine Research, Bergen, Norway

Blackwell
Science

Geostatistics for Estimating Fish Abundance

Geostatistics for Estimating Fish Abundance

J. Rivoirard,* J. Simmonds,† K.G. Foote,‡
P. Fernandes† and N. Bez*

**Centre de Géostatistique de Fontainebleau,*
Ecole des Mines de Paris, France
†FRS Marine Laboratory, Aberdeen, UK
‡Institute of Marine Research, Bergen, Norway

Blackwell
Science

Editorial Offices:
Osney Mead, Oxford OX2 0EL
25 John Street, London WC1N 2BL
23 Ainslie Place, Edinburgh EH3 6AJ
350 Main Street, Malden
MA 02148 5018, USA
54 University Street, Carlton
Victoria 3053, Australia
10, rue Casimir Delavigne
75006 Paris, France

Other Editorial Offices:

Blackwell Wissenschafts-Verlag GmbH
Kurfürstendamm 57
10707 Berlin, Germany

Blackwell Science KK
MG Kodenmacho Building
7–10 Kodenmacho Nihombashi
Chuo-ku, Tokyo 104, Japan

First published 2000

Set in 10/13 Times
by Sparks Computer Solutions Ltd, Oxford
Printed and bound in Great Britain by
MPG Books Ltd, Bodmin, Cornwall

DISTRIBUTORS
Marston Book Services Ltd
PO Box 269
Abingdon
Oxon OX14 4YN
(*Orders:* Tel: 01235 465500
Fax: 01235 465555)

USA
Blackwell Science, Inc.
Commerce Place
350 Main Street
Malden, MA 02148 5018
(*Orders:* Tel: 800 759 6102
781 388 8250
Fax: 781 388 8255)

Canada
Login Brothers Book Company
324 Saulteaux Crescent
Winnipeg, Manitoba R3J 3T2
(*Orders:* Tel: 204 837-2987
Fax: 204 837-3116)

Australia
Blackwell Science Pty Ltd
54 University Street
Carlton, Victoria 3053
(*Orders:* Tel: 03 9347 0300
Fax: 03 9347 5001)

A catalogue record for this title
is available from the British Library

ISBN 0-632-05444-1

Library of Congress
Cataloging-in-Publication Data

Geostatistics for estimating fish abundance/
by J. Rivoirard … [et al.].
p. cm
Includes bibliographical references (p.).
ISBN 0-632-05444-1
1. Fish stock assessment. 2. Geology—Statistical
methods. I. Rivoirard, Jacques.
II. ICES Statutory Meeting (73rd: 1985:
London, England)

SH329.F56 G46 2000
333.95'611—dc21 99-087566

For further information on
Blackwell Science, visit our website:
www.blackwell-science.com

Contents

Preface

The application of geostatistics to fisheries data was initially demonstrated at the 73rd Statutory Meeting of the International Council for the Exploration of the Sea (ICES), held in London in 1985. Gohin (1985) contributed a paper on developing geostatistics for estimating biomass, and Conan (1985) presented an analysis of shellfish data. In these papers, a technique that had been developed for mineral resource estimation was applied to marine biological resources. The solution to an outstanding problem in fisheries research was beginning to take shape: how to determine correctly the variance of an abundance estimate, using the pattern of spatial sampling, the observed properties of aggregations, and the extent of the stock.

ICES continued to play a central role in the development of geostatistics in fisheries, holding three workshops. Two were held in Brest, first in 1989 to consider shellfish survey data (ICES 1989), then 1 year later to consider acoustic survey data on fish (ICES 1990). A third workshop was held in Reykjavik in September 1991, resulting in an ICES Cooperative Research Report (ICES 1993). It was concluded that the geostatistical estimation variance was an appropriate variance to evaluate the spatial sampling error of abundance estimates from a single survey, and the validity of the technique was accepted. A course was subsequently held in February 1992 at the Centre de Géostatistique in Fontainebleau, France, which provided fisheries scientists with a sound formal basis to apply and develop geostatistics in fisheries research. However, the questions of how reliably these techniques could be employed and how they would perform with typical survey data remained. This was the origin of the tripartite project ‘Geostatistics for fish stock assessment’. The project was proposed for the EU FAIR programme in 1992 and accepted for funding in 1993. This book is the direct result of that project, which was carried out with the financial support from the Commission of the European Communities, AIR specific RTD programme, CT 94-1007, ‘Geostatistics for fish stock assessment’.

The authors have been extensively involved in the development of geostatistics for fisheries over the last 5 years. They wish to acknowledge the contributions of a number of colleagues. Philippe Guiblin participated significantly at an early stage of the project. Marek Ostrowski contributed to data analysis and especially visualisation including many of the figures published here. Rob Fryer provided advice and help, particularly with the work on optimum survey strategies. A number of discussions have taken place at various stages with prominent fisheries scientists, which have provided a broader view of the practice. In establishing a set of guidelines for the use of geostatistics in estimating fish abundance, a number of other people should be acknowledged: Gérard Conan, Pierre Petitgas and Yvan Simard provided a broader perspective on the subject at a workshop in Fontainebleau in 1998. Earlier, Pierre Petitgas and Neal Williamson organised a workshop in 1996 in Montpellier, France, on time variability and space–time interaction in fisheries acoustic surveys. This workshop provided an excellent basis for studies of the influence of temporal variability in geostatistics.

Chapter 1

Introduction

The questions asked by fisheries managers can be deceptively simple: what is the abundance of a particular stock? How is it distributed? What is its size structure? To help answer such questions, a number of tools have been developed by resource engineers and scientists. Examples are the trawl survey and acoustic survey, performed especially on demersal and pelagic fishes, respectively (Gunderson 1993; Foote 1996); these are briefly described in Chapter 2. These and other survey tools remain objects of critical examination, always ripe for further development or extension, for their application provides valuable information about fish stocks that cannot be acquired in any other way.

Another apparently simple question that is asked by fisheries managers – in fact, invariably the next one – is: how good is the estimate of abundance? In more technical terms the question is: what is the variance of the abundance estimate? The answer to this question involves two parts. One depends on the measurement error in the determination of fish density at sampling points or stations. The other depends on the sampling error, or statistical representativity of the samples of the fish distribution over the geographical area to be surveyed. It is the random component of the measurement error and the sampling error together that are quantified by the so-called geostatistical estimation variance.

This estimation variance describes the variability of an individual survey abundance estimation for a single survey arising from, among other things, spatial sampling. It is to be distinguished from the variance of a series of survey abundance estimates over time.

The two variances, the geostatistical estimation variance for a single survey and the variance of a survey time series, may be illustrated by two contrasting cases.

(1) A very precise survey with a poor time series may be indicative of variability that is not caused by spatial sampling.
(2) Where a survey may be characterised by a geostatistical estimation variance that is numerically similar to the variance of the time series, the spatial sampling can explain much or all of the variability of the time series.

In either case, and in the general case too, knowledge of the individual survey variance may be crucial to understanding how a time-series estimate may be improved.

The importance of the variance question is illustrated by some of the extreme measures that have been taken to address it. In some cases, because fish are certainly not distributed at random, statisticians have advised fish stock biologists to change their surveying strategy. Specifically, they have recommended incorporation of large elements of randomness into the survey design; namely, in the placement of trawl stations or line transects (Jolly & Hampton 1990). The bald purpose of this tactic is to enable statisticians to estimate variance according to conventional notions, without having to consider the spatial structure of the stock. However, for fish stocks, randomness in sampling design degrades the precision of the estimate of abundance (Gohin 1985; Simmonds & Fryer 1996). This has been accepted (or in many cases ignored) because the perception has been that only random surveys can give the correct estimate of variance.

Fortunately, there is a viable alternative statistical approach to the problem of variance estimation that avoids degrading the estimate of abundance: that of geostatistics. Far from using a specific design to avoid dealing with the pattern of spatial aggregation, geostatistics exploits this through a so-called structural tool; for example, the covariance or variogram. The variance explicitly accounts for the degree of coverage and placement of the sampling stations in relation to the area to be covered and the properties of aggregation. This discussion is continued in Chapter 3, where mathematical expressions for the variance are given. Here, the larger subject is introduced.

Geostatistics is a relatively young field, whose theoretical foundations were established initially by G. Matheron (Matheron 1965, 1967). A relevant and still contemporary exposition of the subject is available in Matheron (1971). Selected recent expositions in the form of textbooks are given in David (1977, 1988), Journel & Huijbregts (1978), Isaaks & Srivastava (1989), Cressie (1991), Armstrong (1998), and Chilès & Delfiner (1999). Some notable recent reviews are found in Rossi *et al.* (1992) and Petitgas (1993a, 1996).

Admittedly, the mining application of geostatistics dominates many of these works, but the generality of the subject and diversity of applications are apparent in Cressie (1991). Not only is geostatistics applied to such subterranean resources as diamonds, gold, coal, oil, gas, and water, but also to terrestrial problems in hydrology (Bardossy 1992) and forestry, for example, and to marine problems in bathymetry (David *et al.* 1986), hydrography (Kielland & Dagbert 1992), mapping sea surface temperature (Gohin 1989), and the estimation of various marine biological resources. Included in this latter class are shellfish (Conan 1985; Gohin 1985; Nicolajsen & Conan 1987; Conan *et al.* 1988a; Armstrong *et al.* 1989), crustacea (Conan *et al.* 1989; Simard *et al.* 1992; Gonzalez-Gurriaran *et al.* 1993; Maynou *et al.* 1996) and Chironomidae (Smit *et al.* 1992), in addition to fish eggs (Bez *et al.* 1996, 1997), plankton, and the present subject, fish.

An immediate question to be asked is how a methodology devised for physically stationary resources can be applied to such conspicuously mobile resources as migrating fish, drifting fish eggs, and plankton. The answer is that spatial information on such resources can often be gathered over time periods that are rather short compared with those of large- or even intermediate-scale movements of the stock being surveyed. In addition, as in the case of ore reserve estimation for commercial exploitation, synoptic surveys of fish stocks may be performed only during a single, limited period of time. It is not generally possible to collect additional samples after the survey is completed; it may be impossible for reasons of movement in the case of fish, and infeasible for economic reasons in the case of ore reserves.

The number of applications of geostatistics to fish up until about 1990 was very modest. Some individual works – for example, Conan *et al.* (1988b), Petitgas & Poulard (1989), and Guillard *et al*. (1990) – suggested the potential of the technique. This was further acknowledged at workshops held in 1990 and 1991 (ICES 1990, 1993), which were organised by the International Council for the Exploration of the Sea (ICES).

At the same time the need for improved information from surveys to assist methods of management was becoming apparent, as evidenced by the development of the precautionary approach. This approach is being increasingly adopted by management authorities, as recommended and defined by the FAO (1995). Accordingly, management advice reflects the quality of information on the fish stock. It requires specific consideration of uncertainties in estimates of abundance [FAO, 1995: paragraph 67 (a)]. Implementation of a systematic survey can give a more precise estimate of abundance, and geostatistics provides the measure of uncertainty associated with the sampling process. Geostatistics can also be used to map the spatial distribution of the stock. Application of its structural tools, related to the spatial correlation, may enable changes in stock abundance to be detected at an early stage. An example is provided by the Canadian northern cod stock, where a trend was observed from a population with strong spatial structure in the middle of the 1980s, to one with little or no structure in 1992, coincident with the collapse of the stock (Warren 1997).

In order to explore the application of geostatistics in fish stock assessment and to make forthcoming results available to those involved in fish stock assessment, as well as to the larger research community, a proposal was submitted to the European Community Specific Programme for Research, Technological Development and Demonstration in The Field of Agriculture and Agro-Industry, Including Fisheries in autumn 1992. The proposed shared-cost project 'Geostatistics for fish stock assessment' was eventually approved, and work commenced formally in July 1994.

The specific objective of the project was to develop geostatistics to estimate fish abundance and associated variance, from:

(1) acoustic measurements of fish density along line transects; and
(2) trawl measurements of density at finite stations.

The project was conducted within the framework of five tasks:

(1) data selection;
(2) application of geostatistical techniques;
(3) preliminary publication;
(4) establishing guidelines for applying geostatistics; and
(5) publication of a comprehensive document.

It is this fifth and final task that is being addressed here.

Central to this book is a series of case studies and simulations. Presentation of these is preceded by chapters on data collection and preparation, and geostatistical methods. The chapter on data collection and preparation considers elementary survey methodology, the measurement of fish density, basic statistics, geographical referencing, and dimensionality. The chapter on geostatistical methods gives an overview of fundamentals, emphasising methods that are used in the case studies, especially structural analysis, global estimation of abundance and variance, and mapping. A total of six case studies are described in some detail. These correspond to a range of surveying situations, which are outlined immediately in advance of the case studies. Three topics are examined through simulation: robustness of variography, temporal change, and survey design. In a final chapter, a number of issues are addressed: recommendations on survey design, scope of geostatistical techniques, and guidelines for applying geostatistics. A brief guide to the geostatistical literature is presented in Appendix A. Some potentially useful software for geostatistical analysis is briefly reviewed in Appendix B.

Chapter 2
Data Collection and Preparation

2.1 Survey design

Some basic goals of surveys are to determine the spatial distribution of a population, estimate its abundance, and, if possible, quantify its precision. This is achieved by means of samples from the population within its domain, and it is the placement of these samples that is the essence of survey design.

There are many elements that need to be considered for the design of an abundance survey, and a comprehensive treatment of these goes beyond the scope of the present text. Details about statistical considerations in survey design can be found in Cochran (1977). The survey types considered here are trawl surveys and acoustic surveys. Specific design considerations for these can be found in Doubleday and Rivard (1981) and Simmonds *et al*. (1992), respectively. There are, however, some general rules that should be considered, particularly with regard to working at sea.

The first aspect to be considered is the area to be surveyed. This should extend beyond the boundaries of the fish distribution in order to ensure total coverage of the population. By their very nature, fish populations inhabit and often move within rather large areas, presenting one of the major difficulties that set fisheries surveys apart from other natural resource surveys. To minimise effects of temporal variability due to fish movement, as well as to make best use of expensive ship time, the survey should be conducted as expediently as possible.

In many cases, however, it may be known in advance that some areas are likely to contain more fish than other areas. In almost all cases, areas of high abundance are associated with high variability and this leads to a reduction in precision if the same sampling intensity is used in all areas. In such cases it is prudent to sample the high density areas more intensively than others and so the survey area is split into two or more sub-areas, known as strata, with greater levels of sampling intensity in the areas with high abundance and variability. The concept of effort stratification is elucidated in Section 6.1.

In other cases, there may be physical and/or other biological reasons to divide the survey area into strata. In such cases, differences between strata may be responsible for part of the overall variability, and by separating them the total variability is effectively reduced. Examples include hydrography or the use of depth to stratify bottom trawl surveys, exemplified in Section 4.5. Navigational constraints provide another reason for stratification. Differences in degrees of coverage imposed by navigation may be addressed simply at the analysis stage. Examples are described in Section 4.1.

There are a number of ways of locating samples within strata. A systematic design locates samples on a regular grid within the stratum. In the case of acoustic surveys, where the samples are taken continuously, the grid is formed from a number of equidistant parallel transect lines. In a ‘systematic centred’ design the grid is centred on the stratum. Some trawl surveys are also based on a systematic design, where the stratum is divided into many ‘blocks’ of equal size. Trawl samples are taken in a punctual manner, rather than continuously, such that a systematic centred trawl survey is obtained by locating each sample at the centre of the block. An element of randomisation may be added to a systematic survey by incorporating a random start point for the whole grid. Another element of randomisation may be added by locating each sample or transect of samples randomly within a block. Finally, there is a random design where the samples are placed at random throughout the stratum.

An example of a trawl survey consisting of a single stratum where each sample is located randomly within a block is given in Fig. 2.1.1 (this survey is analysed in Section 4.2). An example of a typical stratified acoustic survey design is given in Fig. 2.1.2; in this case the number of transects within strata were different (e.g. twice the number of transects in strata where high densities were expected). In the case of a trawl survey a change in effort within a stratum may be implemented by taking more samples per block or by decreasing the block size within a stratum with high abundance and variance.

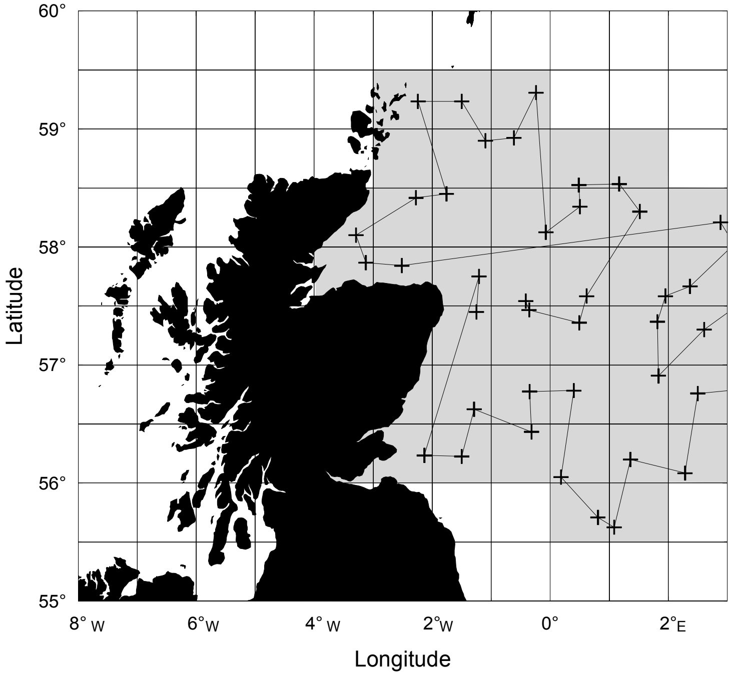

Figure 2.1.1 Cruise track (solid line) and sampling locations (crosses) for the 1989 young fish (trawl) survey. The shaded area indicates the sampling area which is based on a grid of ICES statistical rectangles (1° longitude by ½° latitude).

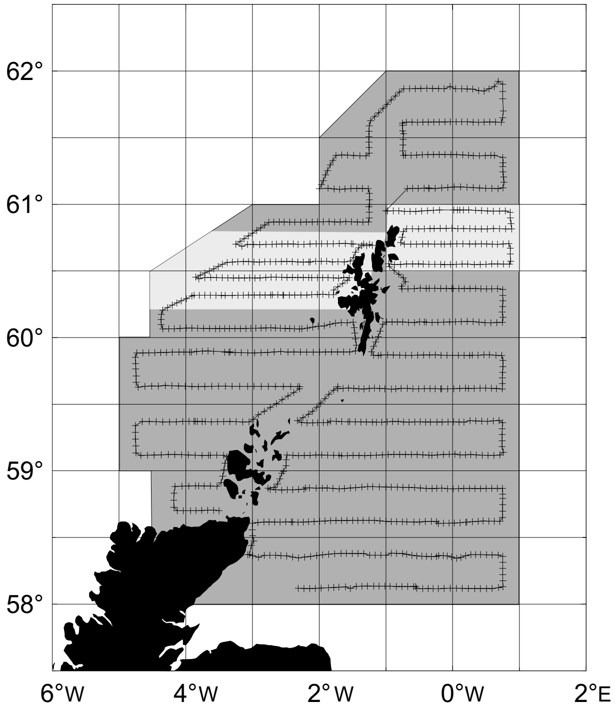

Figure 2.1.2 Cruise track (solid line) and sampling locations (crosses) for the 1991 North Sea herring (acoustic) survey. The shaded area indicates the sampling area which is based on a grid of ICES statistical rectangles (1° longitude by ½ ° latitude); lighter shading denotes strata of high intensity sampling; dark shading denotes low intensity sampling.

Although basic statistical texts may advocate the use of random sampling strategies (Zar 1984; Krebs 1989), systematic grid sampling has been advocated as the best strategy for fisheries surveys (Hilborn & Walters 1992; Simmonds & Fryer 1996). The most relevant point to bear in mind about survey design is the fact that fish, in common with most living organisms, are very rarely distributed at random (Legendre & Fortin 1989); traditional fishermen's knowledge attests to this. Moreover, patchiness in fish distributions implies that fish densities at two points close together are positively correlated (Francis 1984). This characteristic of spatial structure, also known as autocorrelation, is recognised as typical of many natural populations (Cochran 1977).

The existence and acknowledgement of spatial structure has led to recommendations to implement a random survey design within a stratum (Smith & Gavaris 1993). The values are then considered as independent, enabling statisticians to estimate variance using classical methods; the same methods applied to any other design result in an invalid estimate of variance. However, the estimate of mean abundance obtained in a random survey is not as precise as that obtained from a systematic survey design (Gohin 1985; Simmonds & Fryer 1996), and a valid variance estimate for autocorrelated populations can be obtained, regardless of survey design, using geostatistics.

Other advantages of systematic sampling include the following:

- a more precise estimate of mean density when grid points are chosen so as to cut across spatial gradients (which invariably occur in fish populations);
- the ability to map boundaries and spatial distributions more precisely;
- reduction of the risk of missing aggregation clusters or shoal groups that are of the same diameter (or larger) than the distance between grid nodes; and
- allowance for more consistent comparisons of abundance and distribution patterns within a time-series.

This particular topic is treated comprehensively using simulated datasets in Section 5.3, leading to recommendations for systematic designs in Section 6.1.

There is, however, an advantage to incorporating a small element of randomisation in a systematic design. A random starting point for the grid design, or a randomisation within blocks, ensures that every point has an equal chance of being sampled. Furthermore, by allowing the possibility of locating samples at different points in subsequent surveys, an unbiased estimate of the spatial abundance is obtained; in contrast, a fixed grid may only provide an abundance index. This unbiased estimate of abundance does, of course, depend on the accuracy of the measurement of fish density at a location; this is considered below for both acoustic and trawl surveys.

2.2 Measurement of fish density

2.2.1 Acoustic measurement of fish density

A number of acoustic devices and methods are used in the determination of fish density (Foote 1993a). Chief among these is the combined echo sounding and echo integration system. An echo sounder is a box of electronics that controls the operation of a transducer, an electromechanical device that converts an electrical signal to a mechanical vibration and *vice versa*. A directional sound wave can be transmitted in the contact medium, namely sea water, and resulting echoes from fish, among other things, received. In a scientific echo sounder, the transmit level and receiver amplification function are rigorously controlled. Such a device can, moreover, be calibrated, enabling the receiver output to be associated with the absolute echo strength of a scatterer and, ultimately, to the backscattering cross section of the same. This cross section is an inherent property of the scatterer, which also measures its echo potential.

Further processing of the so-called calibrated output signal from the echo sounder is done, for example, by echo integration (MacLennan 1990). Accordingly, the received echo signal is treated as though it were coming from an aggregation of fish that is primarily distributed in one or more horizontal layers. Range compensation is applied to remove the effects of geometrical spreading and absorption, allowing the echo signal to be expressed in terms of the volume backscattering coefficient (Urick 1983) as a function of depth. Integration of this over a given range interval yields an estimate of the area backscattering coefficient. In its most convenient form for application to fish, the coefficient is expressed in units of m^2 of cumulative backscattering cross section per square nautical mile (Foote and Knudsen 1994). This quantity is abbreviated s_A.

So far in this description, only a single transmission and echo signal have been considered. Typically, in acoustic surveying, the echo sounder and transducer are deployed from a ship, where the transmission, or pinging, is done on regular and frequent intervals along the vessel survey track. A nominal pinging rate is one pulse per second. In addition to integrating the processed echo signal over one or more range intervals, the processed signal is also averaged over a number of transmissions. The estimate of s_A improves rapidly with increasing averaging for a uniform distribution of scatterers. The proportionality of s_A and cumulative backscattering cross section can be exploited through the fundamental equation of echo integration, namely:

$$\rho_A = s_A / \bar{\sigma}$$

where ρ_A is the number of scatterers per unit area and $\bar{\sigma}$ is the characteristic mean backscattering cross section for the observed scatterers.

In the present work, visualisation and much of the analysis is performed in terms of the acoustic measure of density, s_A. However, for purposes of illustration, some of the results are converted to biological measures of fish density or abundance, the product of ρ_A and surveyed area, as indeed they would be in actual fish abundance estimations.

Because of the number of operations involved in echo integration, measurement of the acoustic density s_A is subject to a number of errors. These may be random or systematic, producing bias. An explication of errors associated with the echo integration process, hence determining the quality of the measurement of s_A, is provided in Simmonds *et al.* (1992).

2.2.2 Trawl measurement of fish density

It might be imagined that estimation of fish density by trawling is a simple matter, as the ratio of the number of fish caught to the volume of water filtered by the net defines a numerical density. This is more than a gross simplification for many reasons. For example, fish generally react, both individually and collectively, to the presence of a trawl and associated noise fields, which biases the number of fish caught. The opening area of the net and its angle of attack both change with deployment, as with lowering and raising the gear, but also with the changing degree of codend-filling, thus biasing the estimate of the volume of water filtered.

In the case of pelagic fish, catching by trawl is an operation requiring judgement about fish behaviour, and there is seldom belief, much less hope, that the numerical density of pelagic fish can be measured in this way. In the case of bottom fish, however, there is evidence that the density can be estimated by bottom trawl, at least for purposes of establishing an index of abundance. It is this form of trawling that is now considered.

Walsh *et al.* (1993) have enumerated 76 factors contributing to uncertainty in bottom-trawl catches. For a particular fish stock in a known region, experimental research fishermen can obtain, under certain conditions, catches sufficient for estimating local density, if to within some scaling factor or function. In order to estimate absolute or relative abundance, bottom-trawl data are assumed to be collected and reduced according to a standard methodology that observes the kinds of cautions described by Gunderson (1993). It is the underlying assumption in this work, as it would be in the original abundance calculations themselves, that the reduced catch data are representative of the local density.

Catch data from a single station are usually expressed by a set of numbers for each individual fish species. Such a data set typically consists of the total number of caught fish, total catch weight, and relative or absolute numbers of caught fish by length or age.

2.3 Preparation of data for analysis

Basic statistics, complemented with simple visualisation, are useful as a raw description of variables with their order of magnitude, but also as part of a control procedure. Applying geostatistics usually requires a number of steps, and errors are likely to spoil the whole sequence. Moreover, errors are often more easily detected while making simple operations than in more elaborate ones.

2.3.1 Basic statistics

Each sample from a survey data set is located by its co-ordinates and includes the values of one or several other variables (acoustic density, catch number, etc.). The description of missing values (if any) must be explicit (e.g. left blank, or given a negative value if the variable is normally positive).

For a single variable, basic statistics include for instance: the number of values, their histogram, the range of values, the mean, the variance, the standard deviation, the coefficient of variation (CV) and the skew (see Table 2.3.1). Fish density usually has a positively skewed histogram with many small values and only a few large ones (the skew parameter is positive and the histogram tail is to the right). These large or extreme values may have a considerable influence on the mean and on the variance. It is important to scrutinise the largest values, to determine whether these are legitimate extreme values or outliers.

For a pair of variables sampled together, it may be useful to compute the coefficient of correlation (Table 2.3.1). This lies between –1 and +1 but is sensitive to extreme values. When the correlation is positive, the variables tend to be large, or small, together. When negative, one variable tends to be large when the other is small. Note, however, that the correlation measures only the linear dependence between variables. A correlation coefficient of zero (or close to zero) does not mean that the variables are independent, but only that there is no linear dependency. Therefore, the correlation is best suited to variables that are linearly related within their domain. This can be seen on a scatterplot.

2.3.2 Visualisation and verification

The scatterplot or scatter diagram, a plot of the values of one variable versus the values of the other variable, is a very useful visual tool. It immediately reveals the existence of outliers for any of the variables. It may also reveal different statistical populations and inconsistencies (e.g. between catch number and catch weight for trawl data). In addition,

Table 2.3.1 Summary of some basic definitions.

Consider N samples, with values $z_1, z_2, \ldots, z_N$, expressed in units denoted u. We have:

$\bar{z} = \frac{1}{N}\sum_i z_i$	Mean, expressed in units u
$s^2 = \frac{1}{N}\sum_i [z_i - \bar{z}]^2$	Variance, ≥ 0, expressed in units u^2
$s = \sqrt{s^2}$	Standard deviation, ≥ 0, expressed in units u
$\mathrm{CV}_{\mathrm{sam}} = s / \bar{z}$	Coefficient of variation (CV), without dimension
$\dfrac{\frac{1}{N}\sum_i [z_i - \bar{z}]^3}{s^3}$	Skew, without dimension

Now consider two variables z and z' taken from the same samples, with means $\bar{z}$ and $\bar{z}'$, and standard deviations s and s'. We have:

$\frac{1}{N}\sum_i (z_i - \bar{z})(z'_i - \bar{z}')$	Covariance
$\dfrac{\frac{1}{N}\sum_i (z_i - \bar{z})(z'_i - \bar{z}')}{s\,s'}$	Coefficient of correlation, bounded by –1 and +1, without dimension

the plot can show the type of relationship between the two variables, e.g. linear when the cloud of points is elongated along a line.

It is possible to represent a third variable at each point of a scatterplot, when the three variables are sampled simultaneously. This very useful technique is known as a postplot. The value of the third variable can be displayed (whether this is numerical or not, e.g. type of gear), or coded with symbols (with a different symbol for each value or for each class of values). A proportional representation is particularly suited for a numerical variable that is positive and has a skewed distribution: this consists for instance of a circle, whose diameter or, better, area is proportional to the value. This identifies large values rather well.

Postplots, particularly with proportional representation, are often presented in geographical co-ordinates, e.g. longitude and latitude (see also Section 2.3.3). A single scatterplot of the sample co-ordinates is useful to control the location and the navigation route, e.g. for an acoustic survey. A line joining the locations will then reveal the vessel track.

It may be useful to identify the samples that have exactly the same location, as this will cause problems for kriging (mapping), as well as those that are close to each other, as these may also create problems.

Other consistency checks are useful; for example, the sum of catches at length being equal to total catch, or the sum of age proportions being equal to one. Such checking may seem banal, but experience indicates its diligent performance.

2.3.3 Geographical referencing

Geostatistics, like other spatial statistical methods, deals with data that are located in space; a reference system must therefore be chosen. This will be used primarily to express the distance between two locations, as one aim of geostatistics is to measure the spatial continuity of a variable, or equivalently its variability, as a function of distance. Such a distance must first be considered as a vector distance (with different components), even if in many applications only the scalar distance (a single number) is considered. The spatial continuity between two locations in three dimensions, for instance, may depend on their horizontal distance expressed in nautical miles (n.mi.), but also on their vertical distance expressed in hundreds of metres, even if this is negligible when computing the three-dimensional scalar distance. An expansion of the vertical distances may be imagined; however, using a scalar distance in such a case may not be appropriate when the horizontal and vertical variabilities are different by nature and are not comparable.

While there are 60 n.mi. in a degree of latitude, the number of nautical miles in a degree of longitude is 60 times the cosine of the latitude. So in two dimensions, working in degrees is incorrect, and absolute units such as nautical miles is preferable. Given the spatial extension of surveys, it is generally sufficient to work on a plane projection. At low or middle latitudes, a simple transformation by the cosine of the mean latitude of the survey may be sufficient for distance conversions if the north–south extent is relatively small. A more elaborate projection should be used at high latitudes (for instance a gnomonic projection (Snyder 1987), which is a projection, from the earth's centre, onto the plane tangent to the earth at a focal point, e.g. pole or centre of survey).

As suggested above, the spatial continuity may not depend simply on the scalar distance between two points. In particular, the continuity may be greater along a given direction, a condition known as anisotropy. However, it may be better to follow the curvature given by another parameter, e.g. a bathymetric contour. This is more easily taken into account by using a reference system, based on so-called natural co-ordinates.

2.3.4 Dimensionality

As a general theoretical rule, the more intensely the data are collected, the better the description of the variable. However, this also renders the analysis more complex and requires stronger hypotheses to be made. So the choice of the dimensionality (e.g. working in two or three dimensions) depends on the data, but also on the purpose of the analysis. Clearly a three-dimensional description requires data in three dimensions. On the other hand, a spatial abundance can be viewed as the sum of the three-dimensional densities (density per volume unit), as well as the sum of two-dimensional densities (density per surface unit). It is generally easier to work on two-dimensional densities when the purpose is estimating the abundance and its variance. In some circumstances (e.g. acoustic surveys) two-dimensional densities are measured nearly continuously along parallel transects. Then it is possible and simpler to work in one dimension, with the transect cumulants whose sum leads to the abundance (Petitgas 1993a).