ICES Fisheries Technology Committee
ICES CM 2004/B:07, Ref. D,G

Report of the Workshop on Survey Design and Data Analysis (WKSAD)

21–25 June 2004
Aberdeen, UK

https://doi.org/10.17895/ices.pub.9620

International Council for the Exploration of the Sea

Conseil International pour l'Exploration de la Mer

H. C. Andersens Boulevard 44–46 · DK-1553 Copenhagen V · Denmark
Telephone + 45 33 38 67 00 · Telefax +45 33 93 42 15
www.ices.dk · info@ices.dk

Contents

Executive summary

1) **TERMS OF REFERENCE.** The **Workshop on Survey Design and Analysis** [WKSAD] met in Aberdeen, Scotland, UK, from 21–25 June 2004 to: a) review methods of designing and analysing fisheries surveys; b) summarise the current methods used for survey design and analysis; c) investigate where there are similar design and analysis problems; d) identify areas of agreement and specific areas of work where progress could be made; e) prepare work plans for identified areas of development; and f) investigate methods to deal with intercalibration studies of fishing gears and survey vessels.
2) **A REVIEW OF METHODS OF DESIGNING AND ANALYSING FISH SURVEYS**. An account of previous efforts to examine survey design and analysis is given. The essential statistical elements to survey planning are described and an analytical framework for survey analysis is proposed. Some surveys perform rather well and stock assessments may be improved using alternative models which use survey data more explicitly.
3) **SUMMARY OF CURRENT METHODS.** A number of specific examples are given describing current survey practise in most of the ICES member states. Survey designs, estimation of abundance and variance, and use in assessments are covered for trawl, acoustic and other (ichthyoplankton, visual, drag or dredge) surveys.
4) **SURVEY-SPECIFIC DESIGN AND ANALYSIS ISSUES.** Tow duration in trawl surveys needs to be investigated in more detail because there are advantages to taking a short tow. Visual surveys are becoming popular and distance sampling methods should be investigated. Adaptive sampling can improve precision in cases where the target species is sedentary, but may be less efficient when more mobile patches of organisms are hard to find (small relative to the area). Biological sampling methods in acoustic surveys need improvement.
5) **GENERAL DESIGN AND ANALYSIS ISSUES.** There is an increasing array of model-based procedures. Geostatistics, for example, enable the precision of a survey to be estimated using the global estimation variance, providing the autocorrelation function can be determined. Simulations and model-based procedures which have non-linear approaches to dealing with extreme values can be informative and improve the estimation process, although the methods still need to be evaluated. Survey reports should include: a comprehensive description of estimation procedures, survey precision as the relative standard error, and measures of design efficiency.
6) **CHOICE OF SURVEY DESIGN.** In the presence of positive local autocorrelation (common in most fish surveys), a more precise estimate of the population mean will usually be obtained by systematic sampling or stratified random sampling than by simple random sampling. The optimal sampling design depends on the population and the relative importance of getting the most precise estimate of the population mean and to getting a good estimate of that precision. Fixed survey designs are common in multispecies surveys (e.g., IBTS), and can be effective for detecting trends when the spatial distribution is persistent. They are also practical in areas with significant un-trawlable seabed. However, fixed designs cannot provide unbiased estimates of the variance.
7) **INCORPORATING ADDITIONAL INFORMATION.** Information additional to that of fish density should be collected on surveys, particularly when that information is related (covariate) and can be collected more extensively. Incorporation of covariates (habitat, environment) can lead to improved precision of the abundance estimate, provided that a good relationship exists, and that the covariate is known at more sample locations than the fish density. Ideally, the covariate should be known at all locations (i.e., the whole survey area).
8) **INVOLVING STAKEHOLDERS IN FISHERIES SURVEYS**. Information from the commercial fishing industry should be considered, where appropriate, to provide guidance on survey design. A range of other options were considered and guidelines for the conduct of cooperative research surveys are given.
9) **FUTURE WORKPLANS / 2005 ToR.** The Terms of Reference for the next meeting are: a) Evaluate analyses of estimates of the abundance, associated variance, and density maps, from surveys of a simulated fish population whose abundance is known; b) Evaluate alternative analyses of seven survey datasets; c) Review the state of knowledge regarding the effect of trawl duration on fish catch rate with a view to considering a reduction in sample trawl duration; d) Evaluate analyses of covariate data which could provide improved precision of abundance estimates; e) Review methods for combining surveys of the same resource using different methods; f) Evaluate the sensitivity of methods to estimate biological parameters in terms of analytical assumptions and measurement error.
10) **INTERCALIBRATION STUDIES OF FISHING GEARS AND SURVEY VESSELS.** A number of intercalibration studies of trawl surveys and acoustic surveys were presented. If calibration factors are estimated with poor precision (as is often the case), then applying them may result in estimates whose mean-square-errors are greater than the unadjusted estimates. Suggestions and advice for intercalibration exercises are given.

1 Introduction

1.1 Terms of reference

According to C.Res. 2003/2B07 the **Workshop on Survey Design and Analysis** [WKSAD] (Co-chairs: P. Fernandes, U.K., and M. Pennington, Norway) met in Aberdeen, Scotland, UK, from 21–25 June 2004 to:

a) review methods of designing and analysing fisheries surveys;
b) summarise the current methods used for survey design and analysis;
c) investigate where there are similar design and analysis problems;
d) identify areas of agreement and specific areas of work where progress could be made;
e) prepare work plans for identified areas of development;
f) investigate methods to deal with intercalibration studies of fishing gears and survey vessels.

WKSAD will make its report available by 31 July 2004 for the attention of the Fisheries Technology, the Living Resources, and the Resource Management Committees.

1.2 Participants

Jean Adams	USA	
Doug Beare	UK, Scotland	
Nicola Bez	France	
Russell Brown	USA	
Steve Buckland	UK, Scotland	
John Cotter	UK, England	
Paul Fernandes	UK, Scotland	(Co-chair)
Rob Fryer	UK, Scotland	
Marco Kienzle	UK, Scotland	
Knut Korsbrekke	Norway	
Bart Maertens	Belgium	
Bob O'Gorman	USA	
Rainer Oeberst	Germany	
Michael Pennington	Norway	(Co-chair)
Allan Reese	UK, England	
John Simmonds	UK, Scotland	
Stephen Smith	Canada	
Dave Somerton	USA	
Bjorn Steinarsson	Iceland	
David Stokes	Ireland	
Jon Vølstad	USA	
Paul Walline	USA	
Kai Wieland	Greenland	
Juan Zwolinski	Portugal	

Participants affiliation and e-mail addresses are given in Annex I.

1.3 Structure of the report

The Terms of Reference (ToRs) are addressed within the main sections of the report: ToR (a) is addressed in Section 2; (b) in Section 3; (c) in Section 4; (d) in Section 5; (e) in Section 6; and (f) in Section 7. A comprehensive bibliography is given in Section 8. Eleven working documents were presented to the meeting: these are listed in Annex II and the documents themselves are appended.

Section 2 reviews general survey methodology. It provides a brief historical background and in particular explains the origins of WKSAD. A summary of the major considerations for undertaking fish surveys are given: these are based on an interpretation of Cochran's (1977) "steps" as applied to fish surveys. An analytical framework is then proposed which describes how survey data are collected, analysed and used in the assessment process. Then, each of these components is described in a general sense with reference to the methods applied.

Section 3 summarises the current methods used in surveys across the world in ICES member states on both sides of the Atlantic. A sub-section on survey design gives details of the rationale and descriptions of the many surveys divided into categories of trawl, acoustic and others (egg, larvae, dredge, visual). This is followed by respective sub-sections on procedures for estimation of abundance and variance, and the use in stock assessment.

Section 4 provides an account of perceived survey design and analysis problems, such as: tow duration and gear size in trawl surveys (effectively the sample size); issues associated with visual sled surveys; the merits of adaptive sampling; and the need for good documentation. A variety of analytical methods are discussed which leads to a clarification of design-based and model-based approaches. Finally, issues such as non-linear methods (to deal with the high proportion of zeros and extreme values) and biological sampling are discussed.

Section 5 considers those issues which the group felt constituted agreement; hitherto these may have been issues of contention. In particular, it considers the question of random versus systematic survey designs; the use of fixed designs; how precision and design efficiency should be reported; and an extensive section on the incorporation of additional information, including cooperative surveys with stakeholders (e.g., fishermen).

Section 6 identifies work plans for identified areas of development. A description of a proposed simulation exercise invites participants to survey a two-dimensional fish population with known properties according to some predefined rules: the results of various designs will be discussed at the next meeting. Several specific existing datasets are then identified with the objective of performing more than one type of analysis (i.e., design-based and model-based). Other datasets are identified which have covariates: these will be analysed with the objective of demonstrating how the covariate can improve the precision of the fish abundance estimate. Biological datasets are identified to perform analyses beyond that of estimating fish abundance. Finally, two reviews will be conducted: one on methods of combining different surveys of the same resource; and one on trawl tow duration – with the objective of establishing the effects of shorter tows.

Section 7 deals with intercalibration. A number of case studies are presented and advice is given on the use of data from intercalibration of trawl and acoustic surveys; further, more general advice on intercalibration is also given.

Recommendations are given in Section 8. The first of these pertain to the next meeting which is proposed to take place in 2005 with the objective of reviewing the various studies outlined in Section 6. A number of additional general recommendations concerning surveys are also given.

An extensive bibliography for the whole document is provided in Section 9 and annexes make up Section 10.

2 A review of methods of designing and analysing fish surveys

2.1 Introduction

Fish surveys were successfully conceived in the mid 20th century to provide biological sources of information which could "…improve the quality of judgement necessary in interpreting calculations based on commercial data."; "…provide better indices of pre-recruit year class strength…"; and "…yield valuable information on migration routes, or such vital biological parameters as age-at-maturity, fecundity, feeding rates and preferences." (Dickie, 1981). Smith (2002) provides a brief historical review of survey development (a précis of Smith 1994), and indicates that although attempts had been made at the end of the 19th century to estimate fish abundance, the uncertainties associated with the measurement process were too great at that time to deliver meaningful results. Technological advances in the 1940s and 1950s resulted in a gradual recognition that the fish capture process could be studied as a scientific discipline (Walsh *et al.* 2002). Subsequently a number of survey programmes were initiated, some of which still survive today, such as the Woods Hole bottom-trawl resource survey (Smith 2002) which started in 1963; and the International Young Fish Survey (Heessen *et al.* 1997) which started in 1965 in the North Sea (now the International Bottom Trawl Survey or IBTS). Improvements in the quantitative measurement of fish density using trawls (Walsh 1997) and acoustic methods (Fernandes *et al.* 2002) led to an improved ability to determine abundance from surveys.

Gradually, data collected from research vessel surveys became more important to estimate abundance in the fish stock assessment process (Clarke 1981). In the 1960s and 1970s, stock trends, as determined by virtual population analysis (VPA), were matched to commercial catch per unit effort (CPUE) data, assuming that CPUE was strictly proportional to abundance. In many cases, such as the northern cod (*Gadus morhua*) off Newfoundland, the latter assumption was badly wrong (Hutchings and Myers 1994) and led to stock size overestimation with a consequent collapse of the fishery. It wasn't until the 1980s, however, that survey data were used in a manner analogous to CPUE, as tuning indices for VPA, and later in statistical catch-at-age assessment models (Skagen and Hauge, 2002).

It is now recognised that there is a continued need to invest in survey indices of abundance no matter what assessment methodology is used (Walters and Maguire 1996). Notwithstanding the aforementioned problems with commercial CPUE data, there are also concerns about the quality of the catch data which form the basis for assessments: these may be corrupted by misreporting (of area and/or quantity), and discarding or slippage (Patterson 1998). In conjunction with the poorer samples available from restricted or closed fisheries, these concerns have added to the importance of survey data as "fishery independent" to determine the abundance and distribution of fish for effective stock assessment (NRC 1998). Finally, the move towards new management measures such as closed areas (Pauly *et al.* 2002) and the ecosystem approach (Sainsbury *et al.* 2000) requires the type of information that only surveys can provide, such as the abundance and distribution of non-commercial fish species.

In recognition of their growing importance, the science of surveying fish stocks has gained increased prominence. Methods of determining fish density have been the subject of extensive research. The International Council for the Exploration (ICES) Fishing Technology Committee (FTC) has two working groups which meet annually to discuss such issues: the Working Group on Fishing Technology and Fish Behaviour (WGFTFB) and the Working Group on Fisheries Acoustic Science and Technology (WGFAST). These groups have provided authoritative documentation in their respective fields and organised formal international conferences (see Fernandes *et al.* 2002; Walsh *et al.* 2002). The use of survey data in assessment also receives much attention in the variety of assessment working groups convened by ICES. In addition, there is a specific working group on Methods of Fish Stock Assessments (WGMG) that can be called upon to investigate developments in this field.

Issues of survey design and analysis, however, have only been addressed in *ad hoc* workshops. Canada's Department of Fisheries and Oceans (DFO) convened a workshop on bottom trawl surveys in 1980 (Doubleday and Rivard, 1981). ICES has convened workshops on the analysis of trawl surveys (ICES, 1992); and several on spatial statistics (ICES, 1989; ICES, 1990b; ICES, 1993). A comprehensive review of acoustic survey design and analysis procedures arose from discussions at WGFAST (Simmonds *et al.* 1992).

In a review of FTC activities, ICES (2003e) recommended that a third Working Group on Survey Design and Analysis (WGSAD) may be warranted in the future due to the increasing emphasis on surveys. In order to gauge interest and demand, a workshop with a proceedings format should first be convened: this was approved and Council Resolution 2003/2B07 defined a set of Terms of Reference for the workshop (WKSAD).

This section aims to review methods of designing and analysing fish surveys as part of the first (a) of the terms of reference for WKSAD. Initially at least, it is not intended as a comprehensive guide of current practice, rather as a source of reference so that detailed information may be sought elsewhere.

2.2 Survey planning steps

Cochran (1977) describes eleven steps involved in planning and executing a survey which are useful to define prior to any theoretical considerations. What follows is an attempt to interpret Cochran's scheme from a fish survey perspective. These describe the major facets that need to be taken into account; those that concern survey design and analysis are then discussed in more detail.

1) **Objectives of the survey.** The assessment of fish stocks involves the estimation of the quantity of fish and the prediction of the quantity into the near future such that appropriate management measures can be implemented. In most cases, the estimation or assessment process is based on commercial catch-at-age data tuned with data from fishery independent surveys. Surveys are, therefore, conducted to determine the abundance, age structure, and geographical distribution of most of commercial fish stocks. All surveys are designed to accomplish at least two essential objectives and additional desirable ones:
    i) They must provide an estimate of average fish density over the entire spatial range of the stock. This can then be used as a relative measure of abundance (e.g., average number of fish per hour towed) or it may be extrapolated to a global measure (e.g., total abundance of fish at age) depending on the assumptions of the technique to measure fish density.
    ii) The spatial distribution of the stock must not only be properly mapped, but also contained (i.e., on average, fish should not be missed, and should not occur beyond the borders of the survey)[1].
    iii) It is desirable, and in some cases essential, that the survey should be able to detect changes in stock size between time periods; and if possible, to detect changes in the size of year class or cohorts. This requires consistency in certain methods within the survey time series (standardisation). If methods are changed, then the influence of the change on the survey time series should be estimated.
    iv) For the purposes of prediction, it is essential that the number of young fish (pre-recruits) that are about to enter the fishery are quantified. Most of these fish are too small to be caught by commercial fishing gear, so surveys using specialised nets are required to estimate their abundance.
    v) Finally, it has always been important to collect ancillary data of either a biological (e.g., maturity, sex-ratio, weight, stomach contents) or physical (e.g., temperature and salinity) nature. This objective may be particularly important when considering an ecosystem approach to fisheries management.

1 An individual survey may not singularly contain the entire spatial distribution of a stock, particularly where the area is too vast for one ship alone to cover. Containment is, therefore, often achieved by combining information from multiple surveys, conducted through international cooperation with other vessels: this requires some level of standardisation and intercalibration.

2) The **population to be sampled**. This denotes the aggregate (target population) from which the sampled population is chosen. In the case of fish, the population is usually already defined: organisations such as ICES and the North Atlantic Fisheries Organisation (NAFO) have geographic divisions and sub-divisions which they use to allocate fish populations by species. As the survey data are usually tuned from commercial catch-at-age data, these two data sources need to be consistent and the geographic boundaries serve this purpose. Defining populations in this way typically yields, as Cochran (1977) noted, "... a sampled population that is more restricted than the target population." Concerns about the biological validity of defining populations by arbitrary boundaries, e.g., see McQuinn (1997) for a discussion of Atlantic herring, are worth considering when planning a survey.

3) The target population may also change on a temporal scale, on either a diel (24 hr) or seasonal cycle. The season for conducting a survey should be selected carefully to ensure that all age groups of interest will be catchable and that all species of interest will be available. Conducting the survey more than once a year, i.e., dividing available ship time between two survey cruises, may be efficient if it provides a better estimate of recruiting year-classes or provides a better definition of the time series of abundance. This is a question of which provides more information about the stock in one year: two points with medium-sized standard errors, or one point with a (relatively) small-sized standard error.

4) **Data to be collected**. The data required are, by and large, entirely dependent on the objectives of the survey. In most cases, the fish need to be counted and/or weighed within a specific sampled area to determine fish density as numbers, or weight, per unit area or time sampled. Depending on the techniques employed and the nature of the fish density unit (see Section 2.4) other data may be required, e.g., calibration factors of the acoustic instrument. The position of the sample in geographic coordinates is also critical and advances in satellite positioning technology have enabled samples of fish density taken at sea to be positioned with a potential accuracy of 3-5 metres. Counts or weights of fish may be subject to an additional subsampling process (Westrheim 1967; Westrheim, 1976), particularly if the initial sample taken is very large (Hughes 1976). Subsampling may also be employed to take ancillary data such as individual fish length, weight, maturity, fecundity, stomach contents and ageing structures such as otoliths.

5) **Degree of precision required**. The precision of a survey is determined by the quantity and quality of samples: the location of samples relative to the population (design); and the instruments used to measure the samples. Methods of determining precision are discussed in Section 2.6.2. Setting targets for survey precision is a relatively recent phenomenon, and no specific requirements were found relating to fish surveys.

6) **Methods of measurement**. Gunderson (1993) provides a comprehensive overview of the sampling equipment and methodology used in conducting fish surveys. He divides survey sampling equipment into four categories: trawl (otter and beam); acoustic (vertical echo sounders); egg and larval (plankton) nets; and direct (visual) counts. The choice of equipment and method is entirely dependent on the biology and behaviour of the target fish species. These methods are described in Section 2.4.

7) **The frame**. The frame is a list or map of sampling units used for performing the sampling operation - it divides the population into units which cover the whole population and do not overlap (Cochran 1977; Jessen 1978). In the case of trawl and ichthyoplankton surveys, the frame might be the number or locations of all possible non-overlapping tows, each defined by a tow length in distance or, more commonly, time (assuming a fixed, consistent speed). The choice of sampling frame should take into account such criteria as statistical efficiency, costs, bias, and logistics. Optimisation of sampling unit size for trawl surveys - tow duration - was studied by Pennington and Volstad (1991). In the case of the acoustic survey, an Equivalent Distance Sampling Unit (EDSU) must be chosen (MacLennan and Simmonds, 1992). The EDSU defines a distance, or time, over which many acoustic measurements (typically 1 is taken every second) are averaged to give one sample.

8) **Selection of the sample**. This is the survey design. Fish surveys are designed using a variety of approaches, typically incorporating one or more of the following to locate samples in the area: randomisation; semi- or pseudo-randomisation; fixed sites; stratification; systematic selection; adaptive clustering; and adaptive stratification. There is still much debate on which of these methods are best employed: this is discussed further in Section 2.5.

9) **The pretest**. This is a small scale trial to test methods and applies more to one-off surveys or newly designed surveys. Trials are often conducted on research vessels to investigate improvements of new or modified survey sampling equipment.

10) **Organisation of the field work**. The logistics of organising surveys at sea are considerable; both on a national and international scale, and are well beyond the scope and remit of this report. ICES assists in this process internationally by convening expert groups for the major coordinated surveys in the northeast Atlantic2; reports from these groups are compiled annually. The groups discuss results from the previous survey, make plans for the

2 Baltic International Fish Survey Working Group (WGBIFS), International Bottom Trawl Survey Working Group (IBTSWG), Planning Group for Herring Surveys (PGHERS), Planning Group on Aerial and Acoustic Surveys for Mackerel (PGAAM), Planning Group on North East Atlantic Pelagic Ecosystem Surveys (PGNAPES), Planning Group on North Sea Cod and Plaice Egg Surveys in the North Sea (PGEGGS), Working Group on Beam Trawl Surveys (WGBEAM), Working Group on Mackerel and Horse Mackerel Egg Surveys (WGMEGS).

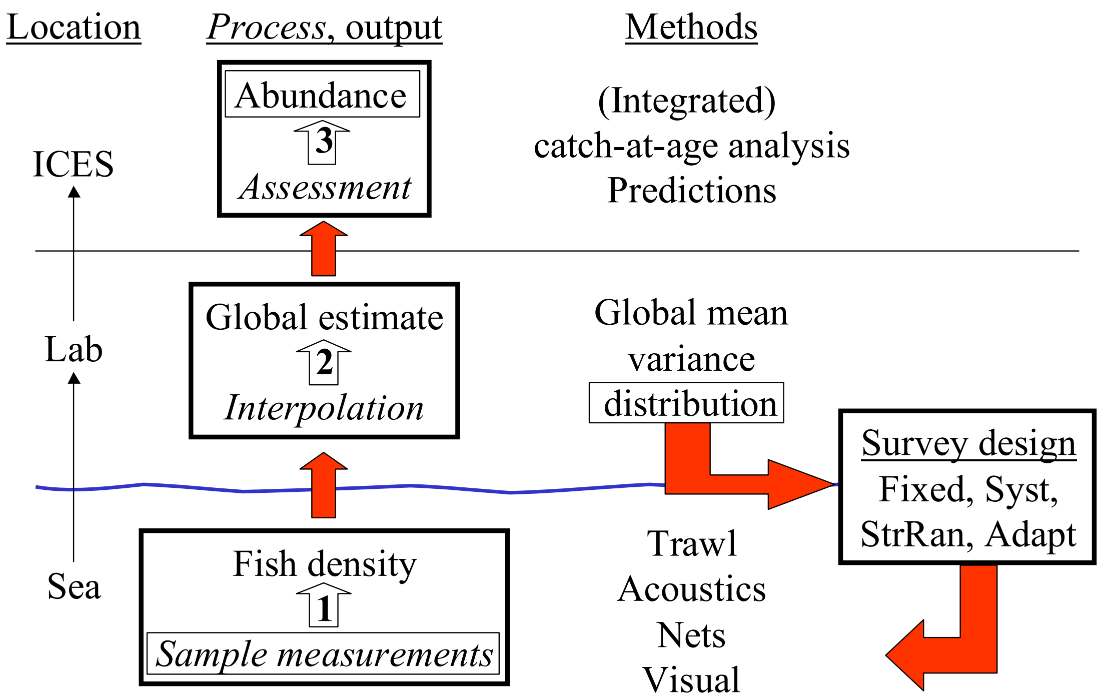


Figure 1. Schematic for an analytical framework for the processing of fish survey data from the collection of samples (1) to the ultimate objective of determining abundance (3).

following year, arrange survey overlaps and or intercalibrations, and discuss methodological problems. On a national scale the impetus is on the institute to carry out training of the individuals participating in the survey, to provide appropriate resources, and to ensure quality control of the data.

11) **Summary and analysis of the data.** There are various stages involved in the analytical process that are discussed further in Section 2.6. Practices are extremely variable and even within the same survey it is rare to see the data reported in a consistent comprehensive format. One exception would be the reports of individual vessels participating in the North Sea herring survey which have a standard reporting format summarising the results from the individuals surveys (ICES 2003b).

12) **Information gained from other surveys**. Cochran (1977) states that "Any completed sample is potentially a guide to improved future sampling, in the data that it supplies about the means, standard deviations, and the nature of the variability of the principal measurements…". The information gained from one survey on the distribution of fish may lead to improved stratification in the survey design for subsequent surveys. This has certainly been the experience on the west of Scotland herring survey where initial surveys in the absence of any information were conducted without stratification and were extremely variable; as the time series progressed effort was stratified and as a result the year to year variability diminished. This experience contrasts with that of surveys with a fixed design from year to year, such as the North Sea IBTS. This is a multispecies survey and so any particular stratification scheme may not improve the estimates for all species.

2.3 An analytical framework for abundance estimation from surveys

Abundance estimation from surveys can be broken down into three, related, but quite distinct components (Figure 1). These are often carried out independently, by different groups of people, in very different places:

1) *The estimation of fish density at a point or site, carried out at sea*. This is the unique technical issue of deriving numbers or weight per unit area or time. It may be derived from a trawl (catch per unit effort), an echosounder (area of acoustic scattering per unit area sampled) or a plankton net (number of eggs or larvae per unit area sampled). The data are collected by people with specialised technical capabilities for sampling, who are often not users of the data.

2) *The interpolation of fish density to a global estimate, carried out in the laboratory.* This concerns both the way in which samples are laid down in space (survey design) and the way in which they are interpolated over the surveyed area (survey analysis). This can be as simple as deriving an arithmetic mean CPUE index, or as complex as deriving geostatistical conditional simulations to estimate the abundance and uncertainty of total biomass. This is usually carried out by individuals associated with a survey planning group or, in the case of IBTS for example, at ICES HQ in Copenhagen. Survey planning groups, of which there are many, deal more in logistics not statistics.

ICES WGFAST (Working Group on Fisheries Acoustics) has examined interpolation methods in the past (Simmonds *et al.* 1992; ICES 1990b). There are, however, no expert groups in ICES which deal with the methodological aspects of this issue: one of the objectives of WKSAD was to address this.

3) *The incorporation of global estimates into stock assessment;* in the case of European ICES member states, this is usually carried out at ICES HQ (by government scientists at expert group meetings). It concerns analysis of the time series of global estimates (at age), and other biological parameters (e.g., maturity ogive, weights at age) to estimate abundance, fishing mortality and to make predictions. Survey data are used to "tune" the catch-at-age matrix (Hilborn and Walters, 1992).

Prior to the whole process, a sampling design must be adopted. Although this is often fixed, it is important to consider how the results from previous surveys (specifically the local distribution of fish) may be used to improve the survey design. Thus, abundance estimation from surveys should contribute to an iterative positive feedback process.

2.4 Measurement of fish density

The measurement of fish density is the first process in a survey and can take different forms according to the target species. Trawls provide the most widely used means of estimating fish density (Doubleday and Rivard, 1981; ICES 1992). Fish are sampled with an otter or beam trawl, for a specified time period and the results are expressed as catch per unit effort (CPUE). The effort may be standardised to a certain time period (e.g., one hour) even if the haul duration was less. Otter trawls are used principally to provide indices for demersal roundfish species such as cod, haddock (*Melanogrammus aeglefinus*), whiting (*Merlangius merlangus*), and hake (*Merluccius merluccius*); redfish (*Sebastes spp.*); invertebrates such as the Norway lobster (*Nephrops norvegicus* hereafter referred to as *Nephrops*); and for young pelagics such as herring (*Clupea harengus*) and mackerel (*Scomber scombrus*). Beam trawls are used for flatfish, such as plaice (*Pleuronectes platessa*) and sole (*Solea solea*).

Acoustic instruments are used in fisheries science to assess the abundance, distribution, and behaviour, of fish, plankton, and other marine organisms (MacLennan and Simmonds, 1992). Within ICES, there are currently over 20 fish stocks for which acoustic estimates are carried out. Most of these are pelagic (midwater) species such as herring, sprat (*Sprattus sprattus*), mackerel, sardine (*Sardina pilchardus*), and anchovy (*Engraulis encrasicolus*). The basic tool in fisheries acoustics is the scientific echosounder. This instrument sends sounds down into the water column in an acoustic beam and receives echoes from objects in the water. Each echo is then converted to an electrical signal with intensity proportional to the echo strength and, hence, size or number of fish targets. An echosounder will transmit sound frequently, typically once every second; such that the water column is almost exhaustively sampled along the cruise track (subsequent acoustic beams actually overlap at depth). As the ship moves through the water, a two dimensional echogram is built, with distinctive patterns (echotraces) which may be characteristic of certain fish species. To confirm the identification of targets in the echotrace, trawl samples are taken. Once the species and size of the fish which have contributed to the echotrace are known, the echo intensity can be converted to fish density. The latter conversion is based on experimental evidence of the linear relationship between acoustic density and fish density which gives rise to the concept of target strength (a measure of acoustic intensity for a fish species of a particular length). The fish densities measured continuously over the course of the survey area are then interpolated to produce an estimate of total numbers. The trawl samples also enable weight, age, maturity and sex of the fish to be determined allowing estimation of numbers at age and spawning stock biomass.

There are clear advantages of using an acoustic technique for surveying for pelagic species, the principal one being the fact that most of the water column is surveyed (as opposed to a trawl, which samples only a limited portion immediately above the seabed). However, there are also uncertainties, particularly in the identification of species and in measurements of target strength, which require further research. Acoustic methods are limited to detecting fish above the seabed (typically >2 m) due to the properties of the acoustic beam; they are therefore of less use for surveying demersal fish.

Many commercially important fish and shellfish species shed eggs directly into the seawater. These eggs spend a period drifting in the plankton before hatching into fish larvae. Assuming that eggs and larvae are easier to sample without bias than the adults, and that numbers of eggs and larvae are proportional to the adult population size, then stock size can be estimated. Such techniques are currently used for the estimation of mackerel and horse mackerel (*Trachurus trachurus*). The egg surveys involve collecting samples of ichthyoplankton from predefined rectangles within the spawning area during a number of discrete periods (usually around a month long) throughout the spawning season. The mackerel eggs are then identified and extracted from the samples, staged, and counted. The numbers of stage I eggs are converted to give a density in number of eggs per metre squared per day based on the performance of the sampler and the volume of water filtered. This value is then converted to daily egg production.

2.5 Survey design

There are many elements which need to be considered for the design of an abundance survey, and a comprehensive treatment of these goes beyond the scope of this report. Details about statistical considerations in survey design can be found in Cochran (1977) and Kish (1995). Specific design considerations for trawl and acoustic surveys can be found in Doubleday and Rivard (1981) and Simmonds *et al*. (1992) respectively. There are, however, some general rules which 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 which 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. It is then prudent to sample the high density areas more intensively than the others. The survey area is, therefore, 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 and the effects on survey precision are discussed in Cochran (1977); demonstrations of how precision is increased by appropriate stratification are given in Shotton and Bazigos (1984), Jolly and Hampton (1990) and Smith and Gavaris (1993a).

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 (Azarovitz, 1981). Navigational constraints provide another reason for stratification. Differences in degrees of coverage imposed by navigation may be addressed at the analysis stage.

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 the stratified random design where the samples are placed at random throughout the stratum.

The stratified random design has proved to be the most common design for trawl surveys (Gunderson 1993). Examples of these include the groundfish trawl surveys for haddock and cod on the eastern coast of Canada (Forest and Minet, 1981; Halliday and Koeller, 1981); multispecies off the northeast USA (Azarovitz, 1981); cod off Iceland (ICES 1992); and scallops (*Placopecten magellanicus*) on Georges Bank (Mohn *et al.* 1987) and in the Bay of Fundy (Smith and Lundy 2002). Stratified random designs are also used in acoustic surveys of krill (*Euphausia superba*) off South Georgia (Brierley *et al.* 2003) and South African anchovy (*Engraulis capensis*, Jolly and Hampton 1990).

A random survey design within a stratum renders the values independent, enabling estimates of variance to be made using well known formulas. Simple random sampling formulas applied to a systematic design can result in an invalid estimate of variance. However, the estimate of mean abundance obtained in a purely random survey is not as precise as that obtained from a systematic or a stratified random survey design (Lenarz and Adams, 1980; Gohin, 1985; Simmonds and Fryer, 1996). Furthermore, a valid variance estimate for auto correlated populations can be obtained, regardless of survey design, using geostatistics (Rivoirard *et al.* 2000), providing the spatial structure can be adequately described by the variogram.

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.

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, e.g., by selecting a new random starting point, an unbiased estimate of the spatial abundance is obtained. In contrast a fixed sampling design may only provide a relative index of abundance. This unbiased estimate of abundance does, of course, depend on the accuracy of the measurement of fish density.

Systematic sampling is common in acoustic survey designs such as those for: herring in the North Sea (Bailey *et al.* 1998); Alaskan walleye pollock (*Theragra chalcogramma*) in the Bering Sea (Williamson and Traynor, 1996); krill in the St. Lawrence estuary (Simard *et al.* 2003); and Norwegian spring-spawning herring, in fjords (Foote 1993) and the Norwegian Sea (ICES 2003c). Systematic designs are also common in ichthyoplankton surveys, such as: the California Cooperative Oceanic Fisheries Investigations (CalCOFI) series in southern California (Ohman and Venrick, 2003); the herring larvae surveys in the North Sea (ICES 2004a); and the mackerel egg survey of the northeast Atlantic (ICES 2002a). Occasionally trawl surveys are also conducted using systematic sampling, such as - what is considered by the authors as “the most important commercial cod stock in the world” - northeast Arctic cod (Godø and Totland, 1994, cited in Rivoirard *et al.* 2000).

There is an argument that since one of the objectives of fish surveys is to provide interrannual trends in abundance; these are best estimated if the same stations are chosen every year (ICES 1990a). This gives rise to the fixed station design common to many trawl surveys in northern Europe, such as the International Bottom Trawl Survey for groundfish (ICES 2003a) and the beam trawl surveys for flatfish (ICES 2003d) in the North Sea; and groundfish surveys off Portugal (Cardador *et al.* 1997). Fixed station designs are also commonly used for bottom trawl surveys in the Great Lakes of North America (O'Gorman and Schneider 1986). Heessen *et al.* (1997) describe the IBTS design as "semi-random": this is presumably because of two factors: 1) the initial design was to position stations at random within an ICES rectangle; 2) the positioning could not be entirely random because stations can only be allocated to suitable fishing grounds. According to Warren (in ICES 1992), the relative merits of the fixed station approach hinge on the idea of persistence, corresponding to the condition that changes in relative abundance at the sampled stations are representative of changes in the whole population. Although the mean abundance obtained within a year with a fixed station survey will generally be biased, differences between years will be unbiased if there is persistence.

Finally, there are adaptive survey designs where the procedure for selecting the sample may depend on values of the variable of interest observed during the survey (Thompson and Seber, 1996). Adaptive surveys have been carried out for trawl surveys (Francis 1984); for visual surveys of *Nephrops* (ICES 2000); scallop surveys on Georges Bank (Robert *et al*. cited in Smith 1999); and an acoustic survey for Icelandic herring (Jakobssen, 1983).

2.6 Data Analysis

2.6.1 Abundance estimation

Despite numerous attempts to consider a variety of sophisticated estimation techniques (ICES 1990a; ICES 1992), various expert groups have concluded that for many surveys, the arithmetic mean estimate of abundance, or a weighted version, is as good as any other. In its examination of herring data from the International Young Fish Survey, ICES (1992) concluded that "The results from the locally-weighted robust estimator, the various GLM estimates and the standard index corrected for fishing power are all disappointing, since none seems to be superior to the standard index." The same conclusions were drawn from examination of the Icelandic cod data. The IBTS results are currently expressed as a CPUE index: the average number of fish (at age) caught per hour fishing (ICES 2003a). The beam trawl survey results are similarly reported (ICES 2003d). It is, however, difficult to find a definition of the standard index.

In their extensive examination of geostatistical methods, Rivoirard *et al.* (2000) also concluded that in the case of equal sampling intensity the improvement offered by geostatistical estimators (kriging) is poor compared to the traditional arithmetic mean. Many of the acoustic surveys carried out in ICES are subject to rectangular grid averaging methods to determine the abundance index (MacLennan and MacKenzie, 1988; Simmonds *et al.* 1992; ICES 2004a).

Fish survey data rarely have normal statistical distributions. Typically, they are highly positively skewed, with a few extreme values, and a large proportion of zero observations. As a result, estimates of mean abundance have large variances associated with them when standard sampling formulae are used. This has prompted many attempts to transform the data (e.g., MacLennan and MacKenzie, 1988) or to model the data using a number of parametric distributions such as the negative binomial (Taylor 1953; Lenarz and Adams 1980; Power and Moser 1999), the lognormal (McConnaughey and Conquest, 1992), and combinations thereof, such as the delta-lognormal distribution (Pennington 1983; 1996) and the delta gamma distribution (Stefansson, 1996). Problems with these methods arise when the data do not meet the assumptions required by the models (Jolly and Smith 1989; Myers and Pepin, 1990; Syrjala, 2000). For example, relatively small values can greatly affect the estimator of the mean based on the lognormal distribution. When this is the case, post-stratifying the data based on a predetermined threshold will often provide a stable and effective estimator for highly positively skewed data (Pennington 1991; Folmer and Pennington, 2000). Smith (1990) provides a criterion for selecting those models that can provide estimates of the mean which are robust to deviations from the model.

Occasionally, fish surveys yield a single catch that is many times larger than the next biggest catch. This huge catch may account for more than 50% of the total survey catch. These huge isolated catches only seem to occur for a particular species once in ten or twenty years. An example of this was given at the current workshop in which a long survey series for cod had a single catch that was 80% of the total for that year. There are at least four ways of handling these occasional extreme catches: 1) carry out estimations as usual; 2) estimate the mean using a Winsorised estimator that replaces the largest value with the next largest (Smith 1981); 3) use an estimator based on the lognormal distribution (Pennington 1996; Folmer and Pennington 2000); and 4) use smoothing techniques to deal with the spike in the survey series caused by an extreme value (see, e.g., Pennington 1985).

Various other methods have been proposed to deal with the skewed nature of trawl survey data. Kappenman (1999) used a non-parametric kernel estimator; Smith (1981) and Shotton and Bazigos (1984) described various alternative estimators based on weighted, or Winsorised means; and Chen and Jackson (1995) developed an algorithm based on concepts from a robust regression method –least median of squares. These studies indicated that where extreme values were present, the arithmetic mean was less precise than other options. There is, therefore, conflicting evidence, of which estimator to use to determine the abundance from groundfish surveys. It would be interesting to evaluate the application of extreme value statistics used by mathematical geologists (Caers *et al.* 1996) to fish survey data.

Ichthyoplankton surveys have particular difficulties in determining the abundance index because of the extrapolation from the sample to the adult stock. In the case of northeast Atlantic mackerel, for example, the data are currently analysed using the so-called "Traditional Method" (ICES 2002a), although alternatives using Generalised

Additive Modelling (Augustin *et al.* 1998) have been tested successfully. For the traditional method the process is as follows. The daily egg production data are partitioned into a small number of discrete survey 'periods', within which a reasonable proportion of the spawning area has been sampled. Any unsampled rectangle is allocated a value obtained through a simple linear interpolation from connecting sampled rectangles. If there are too few neighbouring samples, the allocated value is zero. An egg production value for each period is obtained by summing over the rectangles, and this is plotted against time at the midpoint of that period. The Total Annual Egg Production (TAEP) is then calculated by integrating under the resulting curve. The final step is to convert TAEP to biomass. This is done using female fecundity data collected from fish immediately prior to the spawning season. Fecundity is calculated as eggs per gram female. It is corrected for atresia – eggs reabsorbed during the spawning season. The method is predicated on mackerel being a determinate spawner, i.e., no de novo vitellogenesis after the start of spawning. The TAEP is then converted to grams of female fish and doubled for a sex ratio of 1:1 to provide an SSB estimate.

2.6.2 Variance estimation

A wide variety of techniques have been proposed to determine the sampling error of a survey. Many of the model-based estimation techniques mentioned in this report incorporate estimates of variance based on: negative binomial (Taylor 1953; Lenarz and Adams, 1980; Power and Moser 1999); lognormal (McConnaughey and Conquest 1992); delta-lognormal (Pennington 1983; 1996); delta-gamma (Stefansson 1996) distributions; and power transformations (MacLennan and MacKenzie 1988).

Simmonds *et al.* (1992) describe a number of other techniques to estimate the error of the estimate for acoustic surveys, including: multiple or repeat surveys (Aglen, 1989); cluster analysis (Shotton and Bazigos, 1984); the method of Jolly and Hampton (1990); and bootstrapping (Robotham and Castillo 1990). Bootstrapping has also been applied to groundfish trawl data (Smith 1997) and for evaluating pot index surveys (Kimura and Balsinger, 1985) and longline surveys for sablefish (*Anoplopoma fimbria,* Sigler and Fujioka 1988). The previously mentioned robust estimators of Smith (1981) and Chen and Jackson (1995) also include variance estimation.

Geostatistics (Matheron, 1989) also provide methods to determine the sampling error in the form of the global estimation variance, regardless of the sampling design (Petitgas, 1999). Rivoirard *et al.* (2000) provide a comprehensive review of how the method can be used to derive abundance and variance, to produce maps, and to provide an explicit description of the spatial structure. Various types of fish surveys were studied, including: acoustic surveys for North Sea herring, Norwegian spring spawning herring and Atlantic blue whiting (*Micromesistius poutassou*); trawl surveys for North Sea cod, whiting and haddock; and Barents Sea cod. The latter work, however, considered only kriging as the estimator and the global estimation variance. Another approach is to use geostatistical simulations (Chiles and Delfiner, 1999). Instead of producing a single, average case estimate, a geostatistical simulation produces several alternative joint realisations of the local values of a variable of interest (e.g., Goovaerts, 1997). Simulations, therefore, deliver a distribution of estimates which can be used for the estimation of 95% confidence intervals. An attempt to conduct a geostatistical simulation on acoustic survey data was, however, hampered by bias associated with the transformation of the data (Gimona and Fernandes, 2003).

Adaptive survey techniques also have methods to determine variance using ideas based on classical stratified survey methodology (Francis 1984; Thompson and Seber, 1996). Such adaptive approaches were adopted for the underwater visual surveys for prawns off northeast Scotland (ICES 2000) and rockfish (*Sebastes spp*.) in the Gulf of Alaska (Hanselman *et al.* 2003).

In most cases, variance estimation techniques determine the sampling error associated with the survey. The estimation of survey variance, or perhaps more critically from an assessment point of view - the year to year variability in survey estimates - includes other factors such as measurement, or instrument error. Instrument error is a collective term for all of the components that go towards deriving the estimate of fish density at a point (process 1 in Figure 1), including the vessel. Vessel 'fishing power' may vary either from year to year, or from vessel to vessel in multi-vessel coordinated surveys. This is a matter of considerable research (see Pelletier 1998). There is also the individual sampling instrument error associated with measurements of fish density from trawls (Walsh 1996) and acoustic equipment (MacLennan and Simmonds, 1992). There are very few examples where overall survey variability is taken into account: Rose *et al.* (2000) give an example based on acoustic surveys. Although challenging to develop, consistent statistical methods for propagating the error through all of the various processes in the analysis of survey data would be welcome. Bias in these measurements remains an issue for those measuring fish density, as long as the global output of the survey is an index, rather than an absolute measure. Knowledge of the variability of certain bias assumptions might, however, aid in survey analyses.

The widespread application of the precautionary approach (FAO 1995) requires uncertainties relating to the size of stocks to be taken into account in its implementation. As a result, a variety of uncertainty estimates are now included in various assessment models (Patterson *et al.* 2001), but rarely are the variance estimates of the indices of abundance from research vessel surveys included. The assessment of North Sea herring is currently weighted according to the (inverse) variance of the survey indices (Simmonds, 2003).

2.7 Use of survey data in stock assessment

For fish stocks that are monitored by scientific surveys and for which commercial catch statistics are collected, the generally accepted method for assessing these stocks is to combine the survey estimates with the catch data as outlined in Figure 2 (and see, e.g., Smith and Gavaris, 1993b). Before a particular cohort leaves the fishery, the cohort's estimated abundance, based on these virtual population analysis (VPA) type assessments, tends to vary from year to year. Not only are the annual estimates of a cohort's abundance quite variable (see, e.g., Nakken, 1998; Pennington and Strømme, 1998; Korsbrekke *et al*. 2001), there is a tendency for the catch-based estimates to decrease as more catch data becomes available, which is the so called "retrospective problem" (Sinclair *et al*. 1991; Parma 1993; Sinclair 1998; Mohn 1999).

One reason that VPA estimates of current stock size are subject to large revisions is that the relation between the commercial catch during recent years and the actual population structure is usually unknown (Figure 2). Many factors may cause this relationship to vary from year to year. One obvious factor is a change in the spatial distribution of fishing effort over time (Salthaug and Aanes 2003). If the commercial catch data are correct, then for a cohort no longer in the fishery the estimate of its historical abundance (i.e., the converged estimate) may be fairly accurate.

Abundance indices based on scientific surveys often track converged VPA estimates fairly closely, while the non-converged estimates and the survey-based indices tend to diverge (Pennington and Godø 1995; Pennington and Strømme 1998; Korsbrekke *et al*. 2001). Since recent VPA estimates will be revised, while the survey estimates will stay the same, this implies that the information contained in the survey data is not being effectively used to assess the stock.

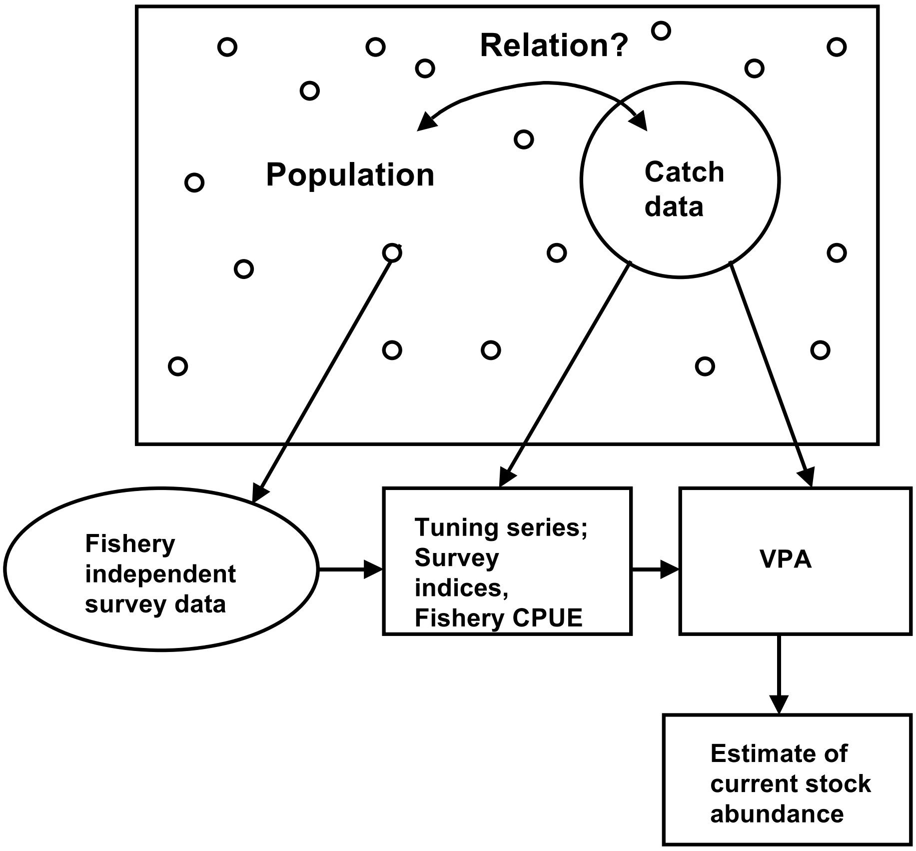


Figure 2. Diagram of the data flow for the standard VPA-type assessment of a stock for which both fishery independent survey data and commercial catch data are available.

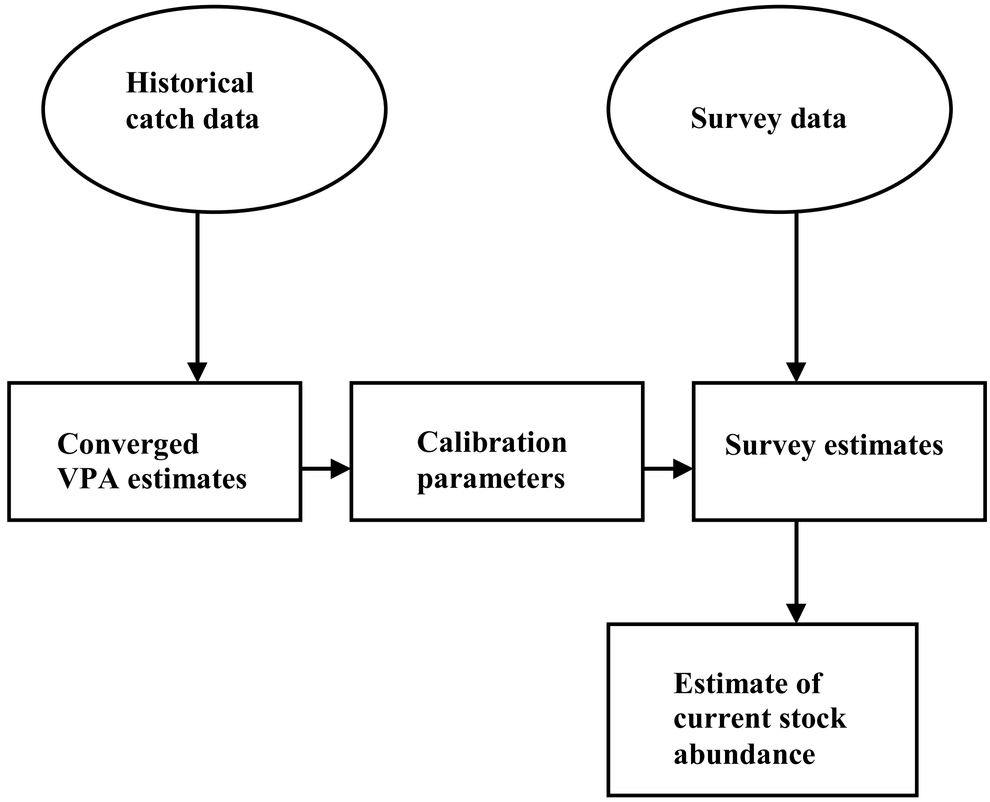


Figure 3. Diagram of the assessment procedure when historical catch data are used to calibrate the survey data.

An alternative to a VPA-type assessment of the current condition of a stock would be to base the assessment only on known, at least in theory, relations. A survey ideally covers the entire stock while converged VPA-type estimates, based on accurate commercial catch data, should provide fairly accurate historical estimates of a cohort's size. Therefore for some stocks it may be sensible to reverse the roles currently played by surveys and commercial catch data. That is, instead of using survey data to tune the current catch data, use historical catch data to calibrate the survey indices (Figure 3).As an example of this alternative assessment technique, converged VPA abundance estimates for northeast Arctic cod during an initial time period (1981–1995) were used to calibrate abundance indices generated for this stock by the winter surveys in the Barents Sea. The survey-based estimates were compared to subsequent converged estimates of cohort size and to the annual assessments (Figures 4, and 5).

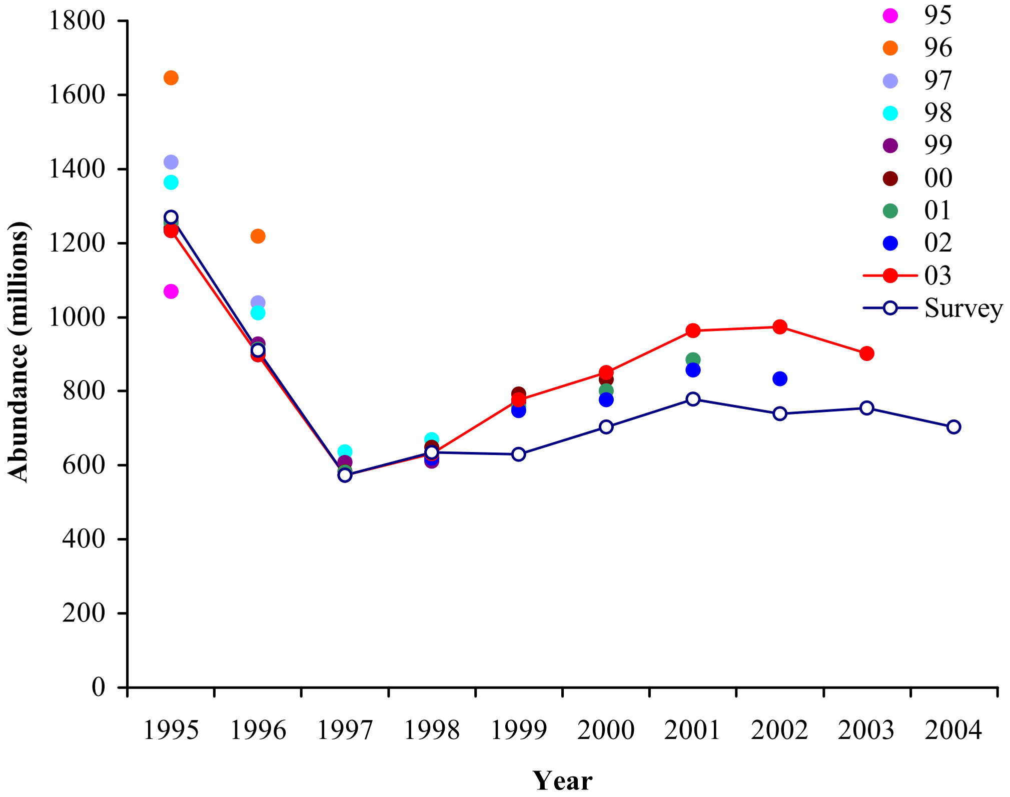


Figure 4. Calibrated survey estimates (connected open circles), ICES, 2003 estimates (connected solid circles) and the 1995–2002 ICES annual assessments (unconnected solid circles) of the total number of Northeast Arctic cod ages 4 through 6

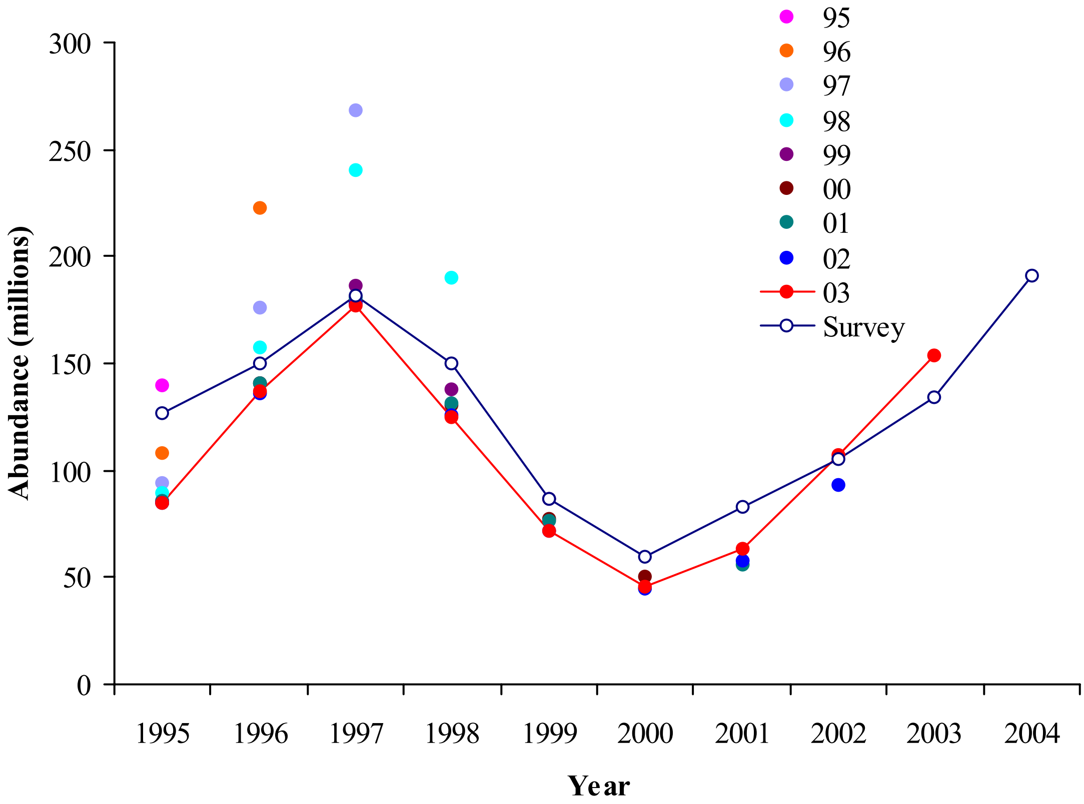


Figure 5. Calibrated survey estimates (connected open circles), ICES, 2003 estimates (connected solid circles) and the 1995–2002 ICES annual assessments (unconnected solid circles) of the total number of Northeast Arctic cod ages 7 and older.

2.8 Conclusions

The role of dedicated surveys is increasing in importance. It is, however, very difficult and expensive to sample large areas, never mind the fact that those areas are actually huge volumes, and those volumes are under an inhospitable ocean. Survey data are, therefore, likely to be imprecise relative to landings (commercial catch) data. They are, however, likely to be more accurate, particularly in areas of illicit activity. Whatever biases survey data are likely to have, they are unlikely to be consistently unidirectional as fishery data can be (misreporting underestimates landings which leads to underestimates in stock size). More importantly, there is a significant research effort working to eliminate the biases of estimating fish density and as knowledge is gained about the bias and it should be reduced. There may also be technological progress to increase sample size through the use of emerging technologies such as multibeam sonar. Fishery science should move from biased but relatively precise landings data to imprecise but relatively unbiased survey data. Through other research efforts, survey precision can be improved and the biases eliminated; no such capacity exists to control landings data without significant coercion.

Although it would be convenient to have a consistent set of agreed methods to determine the abundance and variance of survey data, the wide variety of survey types and conditions dictate that a variety of methods are suitable. One of long term objectives of WKSAD is to decide which methods are best suited to particular objectives and to ensure that the methods are then employed to provide appropriate components of the uncertainty of survey data. Ultimately, the mistakes made in the past with commercial data should not be repeated. Survey data analysis should perhaps be subject to the kind of rigorous statistical process control (SPC) that is common in many other fields. SPC is a method of monitoring, controlling and improving a process through statistical analysis. Its four basic steps include measuring the process, eliminating variances in the process to make it consistent, monitoring the process, and improving the process to its best target value. This last step is certainly a laudable objective.

3 Summary of current methods

3.1 Survey designs

3.1.1 Trawl surveys

United States Northwest Atlantic Bottom Trawl Surveys

The U.S. National Oceanic and Atmospheric Administration (NOAA) Fisheries conducts four annual bottom trawls to index the abundance of demersal, pelagic and invertebrate marine resources in the U.S. and Canadian waters of the northwest Atlantic Ocean in depths ranging from 10–365 m (5–200 fathoms). Two of these surveys, the multispecies Spring and Autumn bottom trawl surveys are among the longest time series of their kind. The Autumn multispecies bottom trawl survey has been conducted annually since 1963, while the Spring multispecies bottom trawl survey has been conducted annually since 1968 (Azarovitz 1981).

These surveys have five principal objectives: 1) to determine the distribution, relative abundance, and biodiversity of fish and invertebrate species found along the continental shelf, 2) to collect biological samples for age determinations, growth studies, fecundity, maturity and feeding ecology, 3) collect hydrographical and meteorological data, 4) collect samples of ichthyoplankton and zooplankton for relative abundance and distribution studies, and 5) collect data and samples for cooperative researchers and programs.

Two primary research vessels, the FRV Albatross IV and FRV Delaware II, have been used to conduct the survey during the time series. These vessels have statistically different catchabilities for many species, and these catchability differences have been calibrated through a series of paired towing experiments that have resulted in excess of 1,000 pair tows. These surveys have primarily utilised a Yankee 36 roller sweep trawl with the following exceptions. From 1973 to 1981, the spring survey utilised a Yankee 41 roller sweep bottom trawl that featured a higher headrope height. From 1963–1984, each survey utilised wood/steel BMV oval trawl doors, and switched to 450 kg Portuguese polyvalent trawl doors beginning in 1985 (Despres-Patanjo *et al.* 1988). Both the net change in the spring survey and the trawl door change were calibrated through paired tow experiments to determine statistical differences in catchability.

Each survey covers a survey area from Cape Hatteras to the Scotian Shelf surveying an area of approximately 268,000 km^2 (78,000 $n.mi.^2$). Each survey employs a stratified random survey design with stratification based on region and depth. Approximately 320–350 stations are visited during each survey, with CTD casts and 30 minute bottom trawl tows conducted at each station.

In addition to the spring and autumn multispecies bottom trawl survey, NOAA initiated a winter multispecies bottom trawl survey beginning in 1992 to target flatfish and other demersal species with lower catchabilities in the multispecies spring and autumn bottom trawl survey. This survey covers an area from just north of Cape Hatteras to the southern flank of Georges Bank. The survey utilises the same stratified random design as the spring and autumn surveys and occupies 110–160 stations annually. The survey utilises a variation of the Yankee 36 bottom trawl with a flat sweep designed to achieve greater bottom contact and resulting in higher catchability of species that are "tight" to the bottom such as flatfish and monkfish (*Lophius spp.*).

NOAA also conducts a targeted bottom trawl survey for northern shrimp (*Pandalus borealis*) in the Gulf of Maine in conjunction with the states of Massachusetts, New Hampshire, and Maine. This survey utilizes a stratified random survey design with stratification that is distinct from the spring, autumn, and winter surveys. The survey utilises a fine mesh shrimp trawl with 1 3/8" stretch mesh twine and 350 kg Portuguese polyvalent doors. The survey generally occupies approximately 60 stations annually.

Survey indices, maturity, age, and growth data are used as inputs for approximately 52 stock assessments including age-based, production modelling and index-based assessments. In addition to stock assessment uses, survey data are also utilised to support a number of ecological, stock identification, and other studies designed to achieve a greater understanding of the marine ecosystems adjacent to the northeast coast of the United States.

Icelandic Groundfish Survey

Since 1985 the Icelandic groundfish survey (IGFS) has been carried out annually in March, covering the continental shelf waters around Iceland with 540–600 "semi randomly" distributed fixed stations. The survey design was based on historical information about spatial distribution of cod. Each year 4-5 identical commercial trawlers have been hired to cover the stations using standardised 105-foot bottom trawls. The horizontal net opening is estimated to be about 17 m and vertical opening about 2.5 m. The standard towing distance is four nautical miles. A conventional Cochran type method is used for calculating survey indices and variances. Various other methods have been investigated, such as geostatistical and GLM models, but they gave results very similar to the simpler methods.

Baltic International Fish Surveys

Internationally coordinated trawl surveys are carried out in the Baltic Sea in spring and autumn to estimate indices of abundance at age of cod and flatfishes. The surveys are coordinated by the ICES working group on Baltic International Trawl Surveys (BITS), and cover the total distribution area of the target species. The participating countries (6 to 7) cover overlapping areas using standard gears. The different steps of planning the surveys and aggregating the data are described in the BITS manual (ICES, 2002c) and in the reports of the Working Group (ICES, 2004d). The results are stored in the BITS database which is hosted by ICES. The indices of the age groups, expressed as catch per hour, are used by the Baltic Fisheries Assessment working group (WGBFAS) as the only fishery independent estimates.

The surveys are designed as stratified random surveys with stratification based on ICES subdivisions and 20 m depth layers. The allocation of hauls is based on a weighted combination of each stratum's area and mean stock size (5-year running mean of age groups 1+ in spring). The weight for the areas is 0.6 and the weight for the running means is 0.4. This combination was chosen to reflect the different development of the cod stock within different areas of the Baltic Sea, given that the coefficients of variations are nearly the same in the different strata. The indices are estimated as stratified means. For improving the quality of the estimates, incorporation of hydrographical conditions and further analyses are necessary.

The Norwegian winter bottom trawl survey in the Barents Sea

The survey methodology is described in Dalen *et al*. (1982) and in Hylen *et al*. (1986). The sampling trawl is a Campelen 1800 shrimp trawl with 80 mm mesh size in the front. There have been a number of changes to the survey since its inception (Table 1).

Table 1. Changes in the Norwegian winter bottom trawl survey in the Barents Sea.

Year	Change from	To
1984	Representative age sample, 100 per station	Stratified age sample, 5 per 5-cm length group
1986	1 research vessel, 2 commercial trawlers	2 research vessels, 1 commercial trawler
1987	60 min. tow duration	30 min. tow duration
1989	Bobbins gear	Rock-hopper gear (time series corrected)
1990	Stratified random bottom trawl stations	Fixed station grid, 20 n.mi. distance
1993	Fixed survey area, 1 strata system, 35 strata	Extended, variable survey area, 2 strata systems,
	Fixed station grid, 20 n.mi. distance	53 + 10 strata, Fixed station grid, 3 densities
	No constraint technique on bottom trawl doors	Constraint technique on bottom trawl doors
	5 age samples per 5-cm group at 2 stations per stratum	2 age samples per 5-cm group at 4 stations per stratum
	Weighting of age-length keys (ALK) by total catch	Weighting of ALK by swept area estimate by length group
1994	35-40 mm mesh size in cod end	22 mm mesh size in cod end
1995	Assuming constant effective fishing width of the trawl	Fish size dependent effective fishing width (time series corrected)
	2 research vessels, 1 commercial trawler	3 research vessels
1996	2 strata systems and 63 strata	1 strata system and 23 strata
	2 age samples per 5-cm group at 4 stations per stratum	1 age sample per 5-cm group, all stations with > 10 specimens

Length based indices are calculated as arithmetic means, abundance-at-age indices are calculated using age length keys (ALKs, age samples weighted with swept area estimates in stratum and length group) and biological parameters (length-at-age, weight-at-age) as weighted arithmetic means (same weights as for ALKs). The coefficient of variation of length group indices was calculated for some years in the mid-1990's, but has not been presented in the survey report in later years. The information given here is a very brief summary of some of the information in Jakobsen *et al.* (1997).

Canadian trawl surveys
The history of trawl surveys on the east coast of Canada are given in Doubleday and Rivard (1981). The longest time series date from 1970 on the Scotian Shelf. All of the surveys use a stratified random design with strata primarily delineated by depth ranges and management boundaries.

West Greenland Bottom Trawl Survey for northern shrimp
The survey has been conducted by the Greenland Institute of Natural Resources since 1988. The principal objective was to estimate the commercially fishable biomass of northern shrimp (*Pandalus borealis*) and the estimation of the numbers of age classes not yet recruited to the fishery has become an important secondary objective in the recent years (Wieland 2003).

The survey was designed as a stratified random survey (Cochran 1977) and a high opening shrimp trawl (Skjervøy 3000/20 mesh) with a bobbin ground gear is used. The design of the survey has been subject to various changes, and major modifications include several extensions of the survey area and a reduction of the mesh size of the cod-end liner from 44 to 20 mm in 1993 (Carlsson *et al.* 2000).

Major strata correspond to geographical areas that are sub-stratified by depth (150-200, 200-300, 300-400, and 400-600 m) where reliable depth information exists (Kanneworff and Wieland 2003). From 1988 through 1997 trawl stations were allocated to strata proportionally to stratum area, but since 1998 the allocation has been weighted towards strata with historically high densities of northern shrimp and where high variances were observed, in order to get a more precise biomass estimate. An exponential smoothing technique for the weighting was applied to give higher influence of the more recent observations to the weight factors. The minimum number of stations per stratum is set to two.

An adaptive two-stage survey approach (Francis 1984) was used from 1994 through 1997. This approach was abandoned because it involved much extra steaming time and because the second-stage stations had lower mean catches than the first-stage stations yielding downward-biased results (Francis 1991).

In 1999 a new method of choosing stations for the survey was introduced using a minimum distance between stations (a buffer zone), but still keeping the randomness in placing stations (Kingsley *et al.* 2004). The method resulted in a much more even distribution of the stations and a substantial increase of the within-stratum nearest–neighbour distance.

From 1988 through 1998 stations were selected at random by replacing sampling sites for each year. To study the stability of the stock distribution and assess the performance of a fixed-station design relative to that of re-sampling about 50% of the stations from the surveys in 1998-2003, were randomly chosen to be resampled as fixed stations in the following year. The remaining stations were reselected, using the buffer zone method.

Due to observed densities of northern shrimp in the region north of 69°30'N being consistently low, and to severe difficulties in finding suitable bottom for trawling, a fixed-station sampling design was employed in this area in 1998.

In 2003, having obtained better information on depth structure, the same procedure for stratification and selection of stations as in other offshore areas was introduced.

A mixture of different tow lengths has been used in the survey, and, since shorter tows appear to be as precise as longer tows (Kingsley *et al.* 2002), the proportion of short tows (15 min) was gradually increased throughout the years. This allowed an increase in the total number of surveyed stations, which has likely contributed to an increase in the precision of the survey estimate of shrimp biomass (Figure 6).

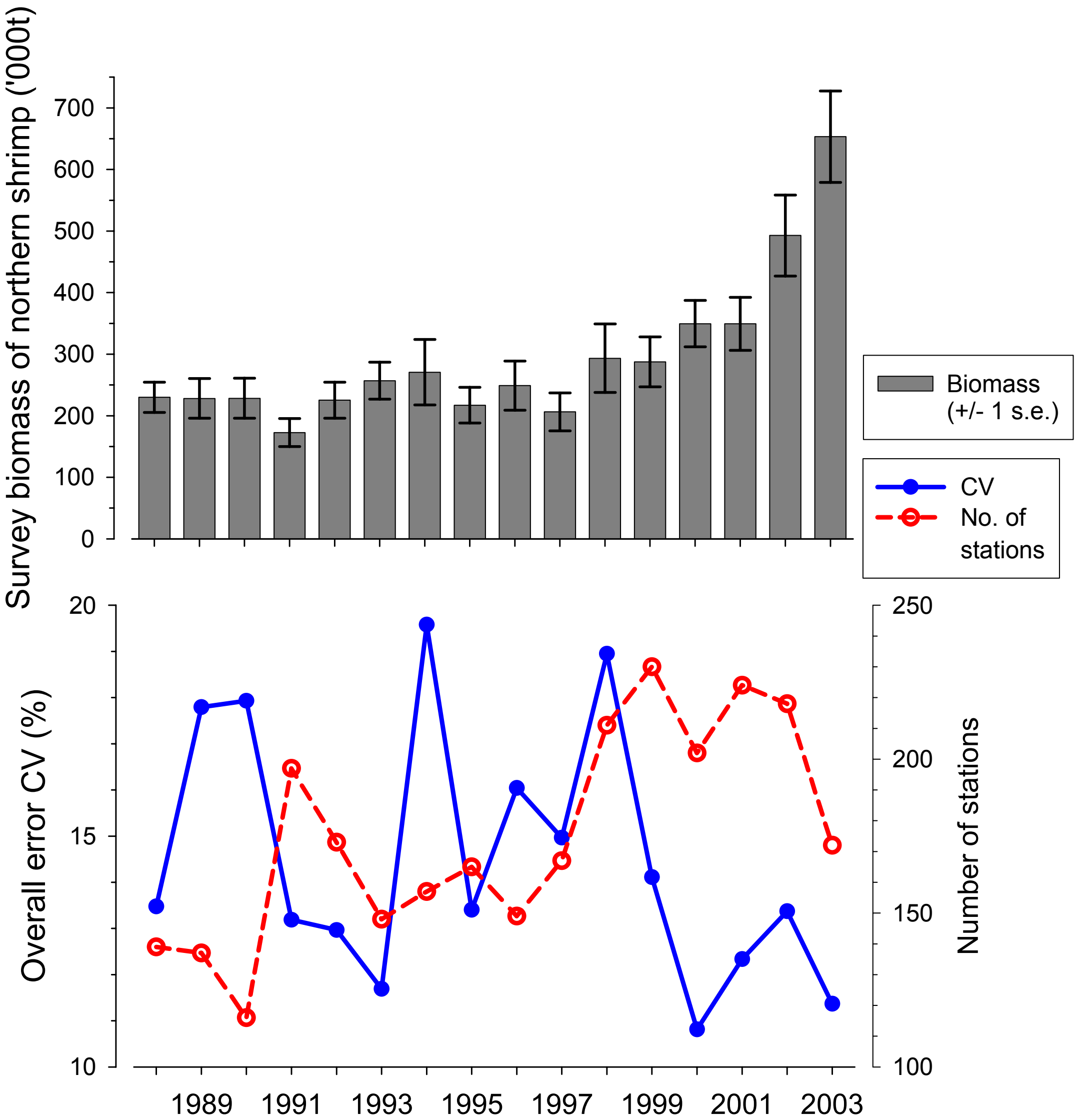


Figure 6. Improvements achieved in the West Greenland Bottom Trawl Survey for northern shrimp.

IBTS Surveys in the north east Atlantic
In 1990 an ICES Working Group (IBTS) was created to coordinate the demersal surveys in the North Sea. An evaluation of the IBTS surveys was conducted by ICES (1998) which included an examination of the use of IBTS indices, the spatial distributions of fish and their seasonal variability; the use of IBTS data in ecosystem studies; the use of correction factors (vessel effects); and considerations of survey effort. A procedural manual was also produced to document important aspects of standardisation for the use by scientists involved in these surveys (ICES, 1999).

Since 1997, representatives from countries carrying out surveys in the whole northeast Atlantic have joined the Working Group. Due to the considerable difficulties in merging the protocols used in the North Sea with those used in the western and southern divisions and it was decided in 1999 that two protocols manuals were required (ICES 2002b; ICES 2002d). For that reason much of the discussion that follows in this section will centre around the western area surveys which have proved one of the most problematic aspects of survey coordination in recent years: the reader is referred to cited literature for a comprehensive account of the IBTS surveys per se.

The general protocol for IBTS trawl surveys is to tow for 30 minutes at 4 knots (as measured over the ground) ensuring good ground contact and a minimum headline height of 3.5 m. However, tow duration even within the IBTS GOV surveys has varied and, given the variation in gear used throughout the area gear geometry will also vary considerably from survey to survey. A number of different survey designs are also employed, including random depth-stratified (Figure 7.), semi-random selection of one to two known clear tow positions per ICES rectangle, and a fixed survey grid design (IPROSTS 2001; ICES 2004b).
IBTS surveys are designed to provide a standardized relative index of abundance in the form of average numbers-at-age to tune the VPA for various assessed stocks (see Table 2 for example). The variance is not generally used by the assessment working groups to weight the survey indices, but often used by the institute concerned to estimate the relative precision from year to year of a particular survey. Variance is calculated in relation to survey design (ICES 1992; IPROSTS 2001; Armstrong *et al*. 2003) which, due to the variety of designs used, particularly, in the western part of the IBTS area, confounds the value of comparing precision across surveys.

More problematic in the implementation of the standard IBTS protocols, throughout the current range of IBTS coordinated surveys, has been the desire for a single sampling trawl. Currently, the prescribed trawl is a GOV 36/47 and allowance is made for its use in harder fishing grounds than the original North Sea area with the option of two heavier groundgear configurations (ICES 1999). The GOV trawl has been evaluated for use in the south western area around Spain and Portugal (SESITS 1999) and found unsuitable for the target species in these areas. In attempting to limit the number of sampling gears within the IBTS area while ensuring the range of ground types and target species (Figure 8, Table 3) are sampled *adequately* the Study Group on Survey Trawl Gear (SGSTG) was formed in 2002. SGSTG identified a number of short-term solutions to the issue of gear standardization within this area and defined the ideal characteristics of one or more "new" gears for the combination of grounds and target species (ICES 2003; ICES 2004b). Findings from the SGSTG have been reviewed by IBTSWG and the WGFTFB (Working Group on Fish Technology and Fish Behaviour), and the proposal is to develop a *new* survey sampling trawl(s) for integration into the Western Division IBTS surveys within the next few years. It is in relation to this expansion of an existing sampling protocol into

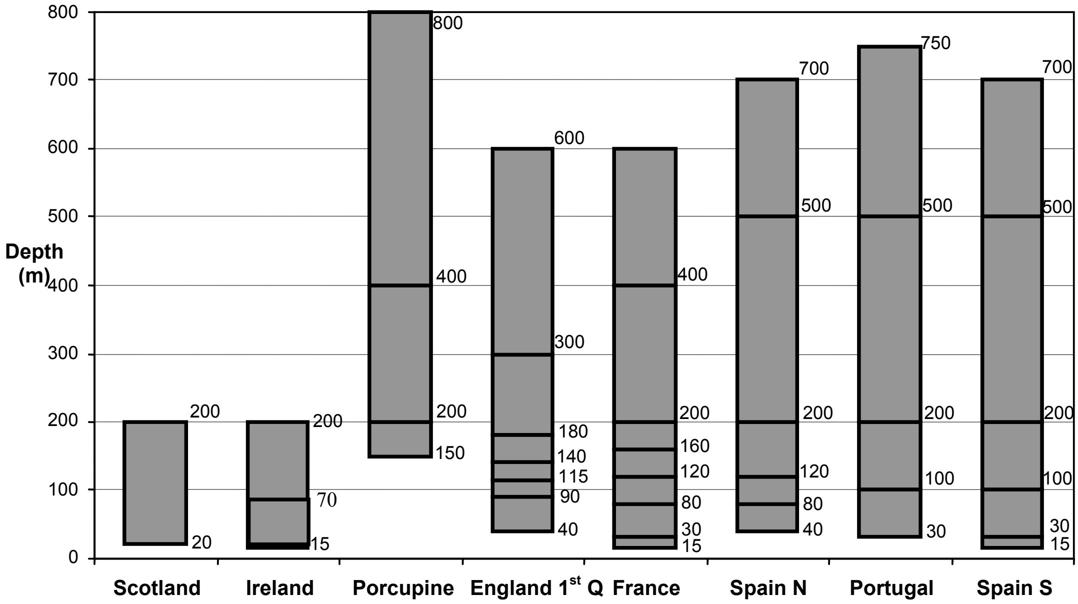


Figure 7. Depth stratification used in IBTS North Eastern Atlantic Area Surveys. Scotland, and previously Ireland, use a fixed number of stations per ICES rectangle. Stratification used in the North Sea is standard area and is species specific: see IBTS manual in ICES (1999).

a larger and contrasting survey environment that the TORs of WKSAD, particularly in relation to intercalibration and survey design, are especially pertinent. In addition, significant funding for these surveys is drawn down under the current Data Collection Regulation (EU Council Regulation 1543/2000) for which data collection and reporting of precision has become a statutory obligation.

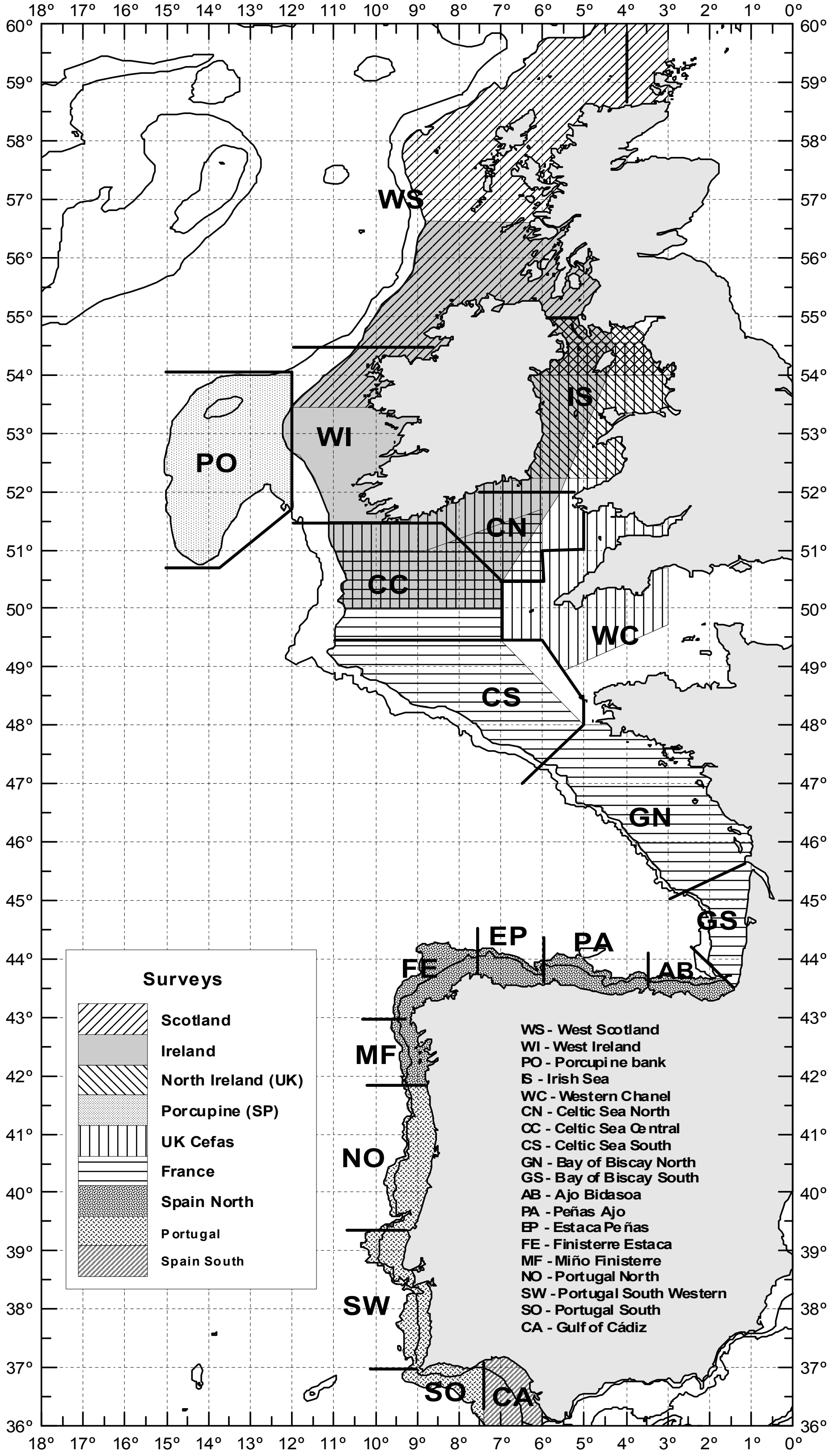


Figure 8. Coverage of the bottom trawl surveys included in the Western and Southern areas and general geographic stratification used.

Table 2. Use of the IBTS Eastern Atlantic surveys in ICES assessment Working Groups (WG – see www.ices.dk for abbreviation definitions). Stock refers to ICES fishing area statistical divisions; “Assess” indicates the assessment model used: XSA = Extended Survivors Analysis; TSA=Time Series Analysis; “Recruit” indicates whether the survey receives weight in the XSA for the pre-recruit and recruiting age classes. “Tuning” columns refer to the years and ages from the survey data that were used to tune the assessment model.

Survey	Common name	WG	Stock	Assess	Recruit	Tuning 2001		Tuning 2002	
						years	ages	years	ages
UK-ScoGFS	Haddock	WGNDS	VIa	XSA,TSA	No	85-00	1-7	85-01	1-7
	Cod	WGNDS	VIa	XSA,TSA	No	85-00	1-6	85-01	1-6
	Whiting	WGNDS	VIa	TSA	No	85-00	1-7	85-01	1-7
UK-WCGFS	Cod	WGSSDS	VIIe-k	XSA	Yes	92-00	1-2	92-01	1-2
	Whiting	WGSSDS	VIIe-k	XSA	No	93-00	2-6	92-01	2-4
	Haddock	WGSSDS	VIIb-k	XSA	Yes	93-00	1-1	1-3(2) 93-97	 1-1(2)
	Hake	WGSSDS 1	North	XSA	No	88-00	1-2	88-01	1-2
	Megrim (*Lepidorhombus whiffiagonis*)	WGSSDS 1	VIIb-k VIIIa-b	XSA	Yes			93-01	2-3
UK-NIGFSoct	Whiting	WGNSDS	VIIa	XSA		92-00	0-4	92-01	0-5
	Cod	WGNSDS	VIIa	XSA	Yes	92-00	0-2	92-01	0-2
	Haddock	WGNSDS	VIIa	XSA		95-00	0-3	95-01	0-3
UK-NIGFSmar	Whiting	WGNSDS	VIIa	XSA		92-00	1-5	92-01	1-5
	Cod	WGNSDS	VIIa	XSA	Yes	92-00	1-4	92-01	1-4
	Haddock	WGNSDS	VIIa	XSA		95-00	1-4	95-01	1-4
UK-NI_MIK	Cod	WGNSDS	VIIa	XSA	Yes			94-01	0-0
	Haddock	WGNSDS	VIIa	XSA		95-00	0-0	94-01	0-0
IR-WCGFS	Whiting	WGSSDS	VIIe-k	XSA	No	93-00	1-1		
	Haddock	WGSSDS	VIIb-k	XSA	Yes	93-00	0-1	93-01	1-1
	Plaice	WGSSDS	VIIb-c	XSA		93-00	1-4	93-01	1-4
	Plaice	WGSSDS	VIIh-k	XSA	No	93-00	2-5	93-01	2-5
	Sole	WGSSDS	VIIb-c	XSA		96-00	2-3	95-01	0-8
	Sole	WGSSDS	VIIh-k	XSA	No	93-00	2-4	93-01	2-6
IR-ISCSGFS	Haddock	WGNSDS	VIIa	XSA				97-01	0-3
FR-EVHOE	Whiting	WGSSDS	VIIe-k	XSA	Yes	97-00	0-4	97-01	0-4
	Cod	WGSSDS	VIIe-k	XSA	No			97-01	1-3
	Hake	WGSSDS 1	North	XSA	Yes	97-00	0-5	97-01	0-5
	Monkfish (*Lophius piscatorius*)	WGSSDS 1	VIIb-k VIIIa-b	XSA	Yes	97-00	0-7	97-01	0-7
	Anglerfish (*Lophius budegassa*)	WGSSDS 1	VIIb-k VIIIa-b	XSA	Yes	97-00	2-13	97-01	2-13
	Megrim	WGSSDS 1	VIIb-k VIIIa-b	XSA	Yes	97-00	2-9	97-01	1-9
FR-Ressgascs	Hake	WGSSDS 1	North	XSA	Yes	87-00	0-5	87-00	0-5
	Sole	WGSSDS	VIIIa-b	XSA	No	87-00	1-6	87-01	1-6
SP- GFS	Hake	WGSSDS 1	VIIIc- IXa	XSA	Yes	83- 00	0- 5	83- 01	0- 5
	Megrim	WGSSDS	VIIIc- IXa	XSA	Yes	90- 00	1- 6	90- 01	1- 6
	Four Spot Megrim (*Lepidorhombus boscii*)	WGSSDS	VIIIc- IXa	XSA	Yes	88- 00	1- 6	88- 01	1- 6
	Horse mackerel	WGMHSA		XSA		85- 00	0- 11	85- 01	0- 11
P- GFS- jul	Hake	WGSSDS	VIIIc- IXa	XSA	Yes	89- 00	1- 5	89- 01	1- 5
	Horse mackerel	WGMHSA		XSA		89- 00	0- 11	85- 01	0- 11
P- GFS- oct	Hake	WGSSDS	VIIIc- IXa	XSA	Yes	85- 00	0- 5	85- 01	0- 5
	Horse mackerel	WGMHSA		XSA		85- 00	0- 11	85- 01	0- 11

(1) WGSSDS (Working Group on the Assessment of Southern Shelf Demersal Stocks) was replaced by WGHMM (Working Group on the Assessment of Southern Stocks of Hake, Monk and Megrim) for these species since 2002.

Table 3. Main and secondary target species by each survey in the western area (see Figure 8 for area codes and Table 2 for survey name abbreviations). IBTS surveys covering each area as well as a first estimation of percentage ground type within each area. Bold figures in ground type indicate the desire of the surveys to cover that ground type, though not necessarily that is being covered by the survey presently. Ground type codes: 1: Sandy, muddy: trawlable with wire synthetic coat. 2: Gravel, bed rocky: trawlable with wire with double coat. 3: Moderate rocky: trawlable with rubber discs or bobbins. 4: Hard rocky: hostile trawling grounds trawlable with rockhopper gear.

Area	**Main target species**	**Other species of interest**	**IBTS Surveys** (Quarters)	**% ground type**			
				1	2	3	4
WS	Cod, haddock, whiting, mackerel	Anglers, megrim, plaice, saithe, pollack, *Nephrops*, elasmobranchs, pelagic species	SCOGFS (1&4) NIRGFS (1&4), IRGFS (4)	20	50	24	6
WI	Cod, haddock, whiting, plaice, sole	Anglers, megrims, hake, *Nephrops*, saithe, pelagic species	IRGFS (4)	80	20		
PO	Hake, megrims, anglers, *Nephrops*	Witch, Deep water species, elasmobranchs	SPGFP (4)	30	45	15	10
IS	Cod, haddock, whiting, plaice, sole	*Nephrops*, elasmobranchs, pelagic species	SCOGFS (1&4) NIRGFS (1&4) IRGFS (4) CEFAS (4)	70	20	10	
WC	Cod, haddock, whiting, plaice, sole, mackerel	Anglers, herring, lemon sole, cephalopods, elasmobranchs, pelagic species	CEFAS (1&4) EVHOE (4)	10	20	50	20
CN	Cod, haddock, whiting, hake, megrim, plaice, sole, anglers	*Nephrops*, turbot, Pollack, ling, elasmobranches, lemon sole, pelagic species	CEFAS (1), IRGFS (4) EVHOE (4)	10	30	50	10
CS	Cod, haddock, whiting, hake, megrims, anglers, sole	*Nephrops*, Pollack, elasmobranches, ling, lemon sole, pelagic species	CEFAS (1&4) IRGFS (4) EVHOE (4)	60	30	10	
BB	Hake, megrims, anglers, whiting, horse mackerel, blue whiting, sole	*Nephrops*, elasmobranchs	EVHOE (4) RESGASC (2&4)	70	20	10	
NS	Hake, megrims, anglers, *Nephrops*, horse mackerel, blue whiting	Mackerel	SPGFN (4)	70	10		20
PT	Hake, horse mackerel, blue whiting, rose & red shrimps, mackerel, Spanish mackerel	Megrim, anglers, *Nephrops*	PGFS (3&4)	20	40	20	20
CA	Hake, horse mackerel, rose & red shrimps, *Nephrops*, Wedge sole	Mackerel, sea breams, cephalopods	PGFS (3&4) SPGFS (2&4)	80	10		10

Bottom Trawl Surveys in Lake Ontario

In 1978, the U.S. Geological Survey and the New York Department of Environmental Conservation began conducting annual surveys of fish stocks in the U.S. waters of Lake Ontario, the easternmost of the North American Great Lakes (Owens *et al.* 2003). Surveys are conducted in early spring, late spring, mid summer, and fall. Each survey has a different target species and is timed to coincide with the peak availability of that species to the bottom trawl. Two vessels, one from each agency, participate in the surveys and each vessel uses identical trawling gear. Procedures for shooting and retrieving the trawl are standardized as are procedures for handling the catch and obtaining biological subsamples. The vessels have been intercalibrated and are used interchangeably. Scientific staff receives training on each vessel to ensure consistent data collection. Since the beginning of the survey program, only one major change in trawling gear occurred and it prompted a multi-year intercalibration study to determine a correction coefficient.

Bottom trawl surveys in Lake Ontario use a fixed station design. Abundance indices for prey fishes are stratified means whereas total catch is used as an abundance index for juvenile lake trout (*Salvelinus namaycush*). Strata are species dependent and are either depth ranges or a combination of depth ranges and geographic areas. Areas within strata are used as weighting factors. Depths sampled vary between surveys depending on the bathymetric distribution of the target species but the intent of all surveys is to sample through the depth range occupied by the target species.

3.1.2 Acoustic surveys

Line transect designs currently used are predominantly parallel designs, with either systematic or random spacing. All designs incorporate the possibility for prior stratification of effort before transects are laid out.

Systematic parallel transect surveys

Systematic parallel designs are carried out with either a fixed or random start point. A random start point for a systematic design is implemented though a one dimensional uniform random shift in transect location on the scale of the transect spacing. There are some advantages in the use of a randomised start point, allowing for analytical techniques that need the properties of a random design. In many situations, the randomness may not have great advantage unless there are links between the spatial distribution of the stock and fixed geographical locations. Where inclusion of a random start point is simple to implement, there appears to be advantages in including it and little advantage in a fixed design. A fixed starting point infers that the survey will be providing an index of abundance, a random start point allows for the results of the survey to be expressed as an absolute estimate. The advantage of a systematic design is improved precision or improved efficiency over more randomised designs, with the disadvantage of more complex methods for variance estimation.

Within a systematic framework there is a possibility for zigzag or triangular survey designs. Rivoirard *et al.* (2000) considered this issue. Although zigzag designs give reduced information near each apex due to closely spaced transects, parallel designs expend time at the transect ends which provide little or no information for estimating the mean. Rivoirard *et al*. (2000) concluded that when transect lengths are more than twice the width of the transect interval, the parallel designs are more efficient, otherwise zigzag designs are more efficient.

Examples of surveys that use systematic designs are: North Sea herring, Iberian Sardine (*Sardina pilchardus*), Alaskan walleye pollock, Peruvian anchovy (*Engraulis ringens*), Baltic clupeiods, and Western US hake.

Stratified random parallel transect surveys

Random parallel designs are carried out with a random start point in one dimension using a uniform random shift in transect location on the scale of the stratum. This process is augmented by additional rules. When a location is selected too near to an already selected transect, it is replaced by another randomly drawn location. Larger strata can be split (arbitrarily) into two or more strata to ensure that sampling effort is distributed more evenly through the sampled area.

The advantages in the use of stratified random designs are the simplicity in estimation and the unbiasedness of the variance. The disadvantage is that the precision is not maximised for autocorrelated spatial distributions. Examples of surveys that use stratified random designs are: Antarctic krill; South African anchovy; New Zealand hoki (*Macruronus novaezelandiae*); and Orange roughy (*Hoplostethus atlanticus*).

Biological data collection

Pelagic fish concentrations are generally very patchy, yielding a high proportion of zero EDSUs. Thus, predetermined trawl station designs are likely to result in a large number of zero catch values and a few high values. The acoustic data is, therefore, used to decide on the location of biological samples. This is usually done through an informal selection process.

3.1.3 Other surveys

Egg surveys

Egg surveys are used to assess the biomass of mackerel and horse mackerel stocks along the continental shelf west of the British Isles. These egg surveys have been conducted triennially since 1977 by various marine research institutes representing a consortium of European nations. The objective of the surveys is to cover the entire spawning area in space and time. If this is done successfully, the total spawning production for any year can be estimated. Then, by making assumptions about female fecundity, atresia, and sex ratio, the spawning stock biomass can be inferred. Various statistical 'estimators' can be used to estimate total annual egg production, which requires a degree of spatial and temporal interpolation. At present a simple design-based method known as the 'Traditional Estimator' is used. However, other more complex model-based methods (generalised additive models, geostatistics) have been suggested as superior alternatives. A recent European Union funded project (Combining Geostatistical and Bayesian Methods for the Management of Atlantic Mackerel Fishery - GBMAF) was, therefore, set up to compare the relative accuracies and precisions of these different analytical protocols using simulated datasets with a known total annual egg production. The overall conclusion was that the simple 'Traditional Method' is more accurate but less precise than the model-based methods which tend to be more precise but less accurate.

Larval sea lamprey surveys

The St. Marys River is the connecting channel between Lake Superior and Lake Huron, and is a major nursery area for sea lamprey (*Petromyzon marinus*) larvae. Because of the river's size, discharge, and depth, compared to other regularly sampled Great Lakes tributaries, different (more costly) sampling gear is required (a deepwater electrofisher, Bergstedt and Genovese 1994) and a new survey was designed particularly for it by the Great Lakes Fishery Commission.

The survey design in the St. Marys River is stratified, systematic, and adaptive. The river is divided into strata according to geographic region and larval density (areas of high and low density – based on four years of intensive transect sampling, 1993-1996). Stratification (introduced in 1999) improves the cost-effectiveness of the survey (improving the precision of the population estimate given the cost) and allows for flexibility in allocating samples annually.

Initial samples are regularly spaced along a grid using systematic sampling (introduced in 2004) with a random starting point. Sampling intensity is higher for the high density strata, and lower for the low density strata. The systematic design gives better spatial coverage of the river than simple random sampling which yielded some gaps.

Any initial sample that yields one or more larvae induces adaptive sampling in the four adjacent grid cells. Adaptive sampling continues until a network of positive catches is surrounded by a series of zero catches. The introduction of (stratified) adaptive cluster sampling to this survey (in 2000) dramatically improved the precision of the population estimate (Figure 9).

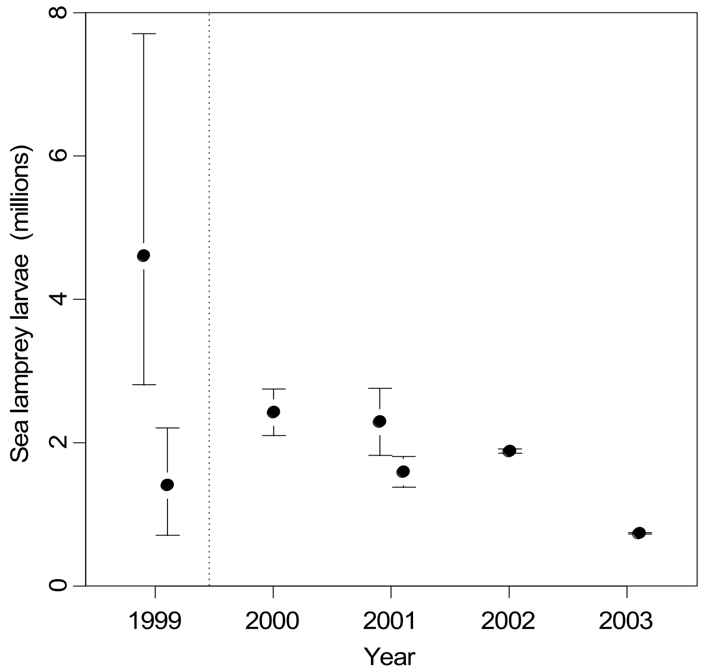


Figure 9. Larval sea lamprey population estimates and 95% confidence limits from the St. Marys River, 1999-2003. Vertical line denotes timing of incorporation of adaptive cluster sampling in the survey design.

Scallop drag surveys

Annual surveys of the scallop grounds off Digby, Nova Scotia have been conducted since 1981 (Smith and Lundy 2002). These surveys have been conducted every June, but vessels, gear and stratification have changed over time. From 1981 to 1988, the survey was conducted on board a commercial scallop vessel using 7-gang Digby drag gear. Since 1989, the government vessel **CCG J. L. Hart** has been used with 4-gang Digby drag gear but estimates have been expressed in terms of 7-gang gear throughout the whole time series. The only change that has been made to gear was the introduction of rubber washers in 1983.

These surveys use a stratified random design with strata based on historical fishing patterns. In surveys from 1981 to 1989, stations were allocated to strata defined to encompass of similar commercial effort areas (low, medium and high) in the recent fishery, based on log book information. Strata were defined to be fixed areas in 1991, and these boundaries have remained the same since then. The area covered by the survey has remained fairly constant since at least 1982, although the sample size has increased with time. Each year, two (three when 7-gang gear used) of the survey dredges were lined with 38 mm polypropylene stretch mesh. Catches in the lined gear were used to estimate the abundance of scallops with shell height less than 80 mm, while the catches from the unlined gear were used to estimate the abundance of scallops with shell heights greater than or equal to 80 mm. Catches of scallops with shell heights less than 40 mm are thought to give qualitative indications of abundance only, due to uncertainties about catchability of the small animals. The number and size distribution of clappers or paired empty shells caught by the survey gear are also recorded each year. These data are used to develop a proxy measure for non-fishing mortality.

Nephrops survey
Video coverage of bottom transects using a towed sled provides a useful methodology in some circumstances. We concentrate here on Fisheries Research Services' (FRS) *Nephrops* surveys, but note that very similar techniques may be appropriate for some other species, such as scallops. *Nephrops* are particularly difficult to survey by direct means. In trawl surveys, the proportion that is caught depends on the proportion that are out of their burrows when the trawl passes, and on habitat. Using video transects, the number of burrows within a strip of a given width can be counted; this indirect method should be more precise than a direct trawl survey, and should generate data from which absolute abundance may be more easily estimated.

Scottish surveys of *Nephrops* use a strip of width 1 m and Its length corresponding to the distance that can be covered in 20 minutes, usually 200-250 m. Strips are chosen according to a stratified random sampling scheme. A single animal has a system of burrows, so an assessment is made of which burrows belong to a single system, such that the count (of systems of burrows) approximates the number of animals in the strip. Empty burrows occur mostly when a trawl has recently passed over the area, catching the animals from those burrows. In that case, the burrows are often destroyed by the trawl; if the animal is not caught, it reopens its burrows. Thus approximating animals by number of burrow systems is believed to be fairly accurate.

3.2 Estimating abundance

3.2.1 Trawl surveys

In general, bottom trawl surveys are used to generate indices of relative or minimum abundance rather than absolute abundance. In particular, the expected catch per standardised tow [a CPUE index] is assumed proportional to absolute abundance. A stratified random (usually with a degree of 'buffering,' see Kingsley *et al.* 2004) or a stratified systematic design is usually employed for trawl surveys and the sample mean within a stratum is used to estimate the mean CPUE for that stratum. Then, based on the area of each stratum, the weighted average of the strata sample means is used to estimate the mean CPUE for the entire survey area or for a subset of strata.
The survey index of northern shrimp biomass (Figure 6) is calculated from the swept area biomass estimates for the different strata. The index might be slightly biased for the recent years in which buffered random sampling were used as this design does neither correspond to an independent random nor an entirely systematic design.

3.2.2 Acoustic surveys

Acoustic data
The analysis method used for acoustic data is to estimate the mean abundance of the strata by the arithmetic mean of the estimated fish density observations from the survey track. The track designs ensure that the samples provide an unbiased estimate. There may be alterations to this general approach, particularly when dealing with transect ends. For example, in areas where parallel transect designs are used and homogeneous density is expected, transects may terminate one half the transect spacing before the boundary and all values are used. This gives equal effort across the area sampled. Near coasts, transect ends are normally excluded, but if the transect stops well clear of the coast (for safety reasons), the part of the transect end equivalent to the distance from the shore is sometimes used as an estimate of near shore abundance.

Biological data
When estimating the abundance from acoustic survey data, fish density (numbers per unit area) is proportional to the integrated acoustic density attributed to a fish species (area scattering per unit area), divided by the mean backscattering cross-section of the fish (area). The latter term is determined from the application of target strength functions which require an estimate of fish length distribution (see MacLennan and Simmonds 1992): this is determined from biological samples taken during the survey. Biological samples also provide age length keys to disaggregate the estimate by age. Often it is difficult to define good biological strata in advance but trends in biological data are common. Thus, samples often contain good information about the local biology, but may not be applicable to the whole survey area. Also, because the samples are not without error, using them in a piece-wise approximation to a surface may be rather noisy. There are three general methods currently used for analysing biological data:

1) Global mean. The estimated biological parameter is the mean of the observations from all the hauls. This leads to local bias and does not reflect or honour the local biological variability.
2) Nearest neighbour. This provides a piece-wise biological surface estimated at each point by the nearest haul. This results in local noise due to sampling errors and local variability in the biological characteristics of the population.
3) Mean of samples in a series of post stratified strata. In an attempt to reduce the local surface noise but to reflect some of the perceived spatial variability in biological data, regions of relative homogeneity are defined either by eye or with the aid of statistical tests, e.g., Kolmogorov-Smirnov tests, to define clusters of hauls. These clusters are then used in a broader nearest neighbour approach to assign means of biological data.

This process is not well founded statistically and estimation of precision is difficult. Biological characteristics of the population are obtained by weighting the information from the piece-wise biological strata by the abundance from the acoustic survey. There is a need to improve the estimation of the spatial distribution of biological data. Current methods result in difficulties in estimating the variance and provide piece-wise surfaces where smooth surfaces might be preferred.

3.2.3 Other surveys

Larval lamprey survey
A single whole-river population (design-based) estimate is derived using a modification of the Horvitz-Thompson estimator for stratified adaptive cluster sampling based on initial intersection probabilities (Thompson 1991, Thompson and Seber 1996). Individual abundances within treatment plots are also estimated using this unbiased estimator.

Scallop drag survey
Standard design-based estimates are used in the scallop survey for all number and weight indices. The exception was in 1999 when the adaptive allocation experiment was run. Rao-Blackwell estimates were used for these data (Thompson and Seber 1996). The mapping of the survey results is conducted using Delauney triangles with an exponential decay function.

3.3 Estimating variance

3.3.1 Trawl surveys

If the survey design is stratified random, then the variance of a survey index can be easily estimated using standard techniques provided that there at least two samples per strata (see Section 5.1). However, it is not clear what the best way is to estimate the variance of systematic surveys. For some systematic surveys, the usual estimate of the variance is used assuming a random and independent location of samples, but this may be an overestimate in the presence of positive spatial correlation. To account for spatial correlation, geostatistical techniques are sometimes drawn on to estimate the precision, but often the model variogram is imprecise (either because there are no stations close together, or because the skewed nature of the statistical distributions renders a poor model fit to the experimental variogram) and thus the resulting estimate of the variance may also be imprecise. Finally, for some systematic surveys no attempt is made to estimate the variance.

In the northern shrimp survey, overall error coefficients of variation (in %) were calculated as relative standard errors:

$$OECV = \sqrt{\sum \frac{STD^2{}_i}{n_i}} \;/ \sum B_i * 100$$

where STD^2, n, and B are variance , number of hauls and biomass stimate for stratum i.

Values of OECV fluctuated erratically between 12 and 20% in the early years of the survey and decreased consistently below 14% in the recent years when buffered random sampling has been used and the higher proportion of short tows allowed a larger total number of hauls (Figure 6). The OECV remained at his low level even when the total number of hauls decreased due to restrictions in survey time and depth stratification was extended to two survey regions when reliable depth information had become available in 2003 (Figure 6). However, in reality the survey appears to be more precise as measured by its conventional standard error as catches at stations close to one other were positively correlated and buffered random sampling reduced the number of stations close to each other thus improving the precision of the survey estimates of biomass.

3.3.2 Acoustic surveys

The precision of acoustic surveys can be obtained through a wide variety of methods. Whole survey variance has been estimated by Rose *et al.* (2000) using predominantly bootstrap resampling methods where all the errors are assumed to be independent. Demer (2004) has proposed a more integrated estimate of variance by inferring spatial variability from the transect data and all other sources of variability from observations by different frequency sounders. Gimona and Fernandes (2003), have applied geostatistics in a conditional simulation to obtain fields that mimic the detail of the spatial distribution. Most users concentrate on the variance of the spatial sampling design rather than the complete survey. Random designs employ the design-based estimators, e.g., Antarctic krill (Demer 2004). Geostatistical estimation variance is used for a number of systematic acoustic surveys designs to estimate the precision of the mean density for acoustic data. A variety of different techniques have been employed but the main methods used routinely are: one dimensional transitive method, e.g., Williamson *et al.* (1996) Bering Sea walleye pollock; and two dimensional global estimation variance, e.g., of acoustic data of North Sea herring (Rivoirard *et al.* 2000).

Variability in age, weight and maturity data has been included through the bootstrap method, resampling at the trawl level and then resampling trawls at length and age. The sampling unit of the trawl is thought to describe most of the variability. Simple methods for combining acoustic and biological data variance are described in Simmonds (2003) for the estimate of precision of acoustic, trawl and larval surveys to obtain comparative variances for weighting in a stock assessment. This method allows correlation among ages in a single survey to be estimated. As the survey designs are not suited to the use of design-based estimators for variance, the bootstrap variance is scaled by the autocorrelation though the relationship between sample variance and geostatistical estimation variance.

3.3.3 Other surveys

Larval sea lamprey survey

An unbiased (design-based) estimate of the variance of the mean density for stratified adaptive cluster sampling is derived based on initial intersection probabilities (Thompson 1991, Thompson and Seber 1996). The use of adaptive sampling in this survey dramatically improved the precision of the population estimates (Figure 9).

Scallop drag survey

Standard design-based formulae are used for variance estimates in the scallop survey. Confidence intervals for the estimates are based on bootstrap resampling (Smith 1997). Analysis of design efficiency indicates that the strata used in the scallop survey usually provide a reasonable gain in efficiency over simple random sampling. However, the allocation of tow stations to strata may result in lower overall efficiency than simple random sampling. One year's experience with adaptive allocation (Thompson and Seber 1996) indicates that this in-survey method of determining optimum allocation may provide better performance for the design. The relative standard errors of the estimates over the survey series tend to be less than 0.15.

Egg surveys

When considering populations with diffuse limits such as fish eggs, a transitive geostatistical approach can be used (Bez 2002). This provides an estimation variance for systematic sampling designs with a random origin or regular stratified sampling. The approach solves the problem of zeroes and of delineation of a domain. In addition, the estimation of the transitive covariogram is much easier and more robust, particularly to high values, than the classical variogram.

3.4 Incorporation of survey results into assessment

3.4.1 Trawl surveys

If there is no commercial catch data for a stock, then trawl surveys are used to track abundance trends. When the total weight of the commercial catch is known, then survey series are often incorporated in a surplus production model assessment. If the commercial catch-at-age is available, then the survey series, along with, perhaps, several other index series, is used to tune a VPA type of model.

3.4.2 Acoustic surveys

Acoustic survey data are usually incorporated into assessments as age disaggregated tuning indices. The Integrated Catch-at-age assessment model (Patterson and Melvin 1996), for example, is used for many herring stocks (e.g., ICES 2004e). Simmonds (2003) describes how such data are weighted in North Sea herring assessment and indicates that the acoustic survey is the most consistent of the age disaggregated surveys used (which include a larvae, juvenile and trawl survey index). Similar tuning procedures are used for the largest single herring stock in the world - the Atlanto-Scandian herring (Toresen and Østvedt 2002). In other cases the survey results are used as absolute estimates of abundance, either to determine stock size (e.g., Barents Sea capelin, Gjøsæter *et al.* 2002) or to set catch limits (e.g., Antarctic krill, Hewitt *et al.* 2002).

3.4.3 Other surveys

Larval sea lamprey survey

The whole-river population estimate of larval sea lampreys in the St. Marys River is used as an ongoing assessment of sea lamprey control efforts on the St. Marys River. Estimates of larval density in individual plots within the river are used to determine the allocation of treatments the following year, i.e., plots with density estimates above some annually-defined trigger are selected for treatment with lampricide. The pattern of larval networks that results from the adaptive sampling, is compared to the boundaries of the treatment plots to determine if the plots fully contain larval hot spots.

Scallop drag survey
The stock assessment for scallops uses a version of the Delay-difference biomass dynamic model (Deriso 1980; Hilborn and Walters 1992) with the survey indices for live and dead (clappers) as the only population indices. The model is structured as a full Bayesian state-space model using Monte-Carlo Markov Chain derived estimates of the posterior distributions (Smith and Lundy 2002; Smith *et al*. 2003). Hyperpriors for variance terms are based on survey estimates. The performance of the model is evaluated through comparing the one-year ahead prediction of the survey biomass with the actual observation from the survey in the following year. To date, this performance has been satisfactory. However, spatial patterns and recent strong annual changes in growth have resulted in increased deviations between the predicted and observed survey estimates (Ouellet *et al*. 2003; Smith and Rago 2004).

4 Design and analysis issues and problems

4.1 Survey specific issues

4.1.1 Tow duration in trawl surveys

Based on many tow duration experiments (both published and unpublished), it appears that tows of short duration are more efficient for surveying a stock than long tows (see, e.g., Godø *et al*. 1990; Pennington and Vølstad 1991; Goddard 1997; Kingsley *et al*. 2002). There are many advantages to taking a short tow at a station; less wear on the gear, time saved can be used to increase the number of stations surveyed, less need to subsample large catches, ability to sample more locations in a survey area with short tows, etc.

Another important reason for using short tows is that fish caught together tend to be more similar than those in the entire population. That is, at a station a 'cluster' of fish is caught and when there is positive intra-cluster correlation for some biological characteristic, such as length, age or stomach contents, the amount of information contained in the sample is greatly reduced (Cochran 1977). One measure of the effect of intra-cluster correlation is the effective sample size (Kish 1965), which is defined as the number of fish that would be needed if sampled randomly to get the same precision for some population parameter as was obtained by the sample of clusters (see also Section 5.4). For many surveys the effective sample size is much lower than the total number of fish sampled. For example, the effective sample size for estimating a fish population's mean length is often approximately one fish per tow (Pennington and Vølstad 1994; Pennington *et al*. 2002). Therefore, just saying 10,000 fish were measured during a survey provides little information. A better way is to give the effective sample size. For example, 10,000 fish were measured and the effective sample size was approximately 70.

In the presence of positive intra-cluster correlation, it is generally best to sample a few fish from as many stations as possible. Thus, shortening tow duration increases the number of locations sampled and reduces the number of fish needlessly sampled.

4.1.2 Gear size in trawl surveys

Large trawl nets with otter boards and sweeps catch fish efficiently but, for survey purposes, have the disadvantage that they take a long time to shoot, stabilise in configuration on the sea floor, to haul, and finally to clean and prepare for the next shot. Good arguments exist for maximising the number of stations fished during a survey. This can be achieved by shortening towing times as discussed elsewhere, or by using a trawl which requires relatively little time before and after towing. A relatively small net that sinks and can be retrieved quickly appears to be what is required for surveys. It is not necessary for it to catch a lot of fish provided it could be used more quickly than the usual commercial-type otter trawl. Beam trawl surveys already use gear with these properties. Many of these issues were discussed by the Study Group on Survey Trawl Gear (SGSTG, ICES 2003a; ICES 2004b).

4.1.3 Edge effects and detectability in sled surveys

The use of towed sleds, bodies and submersibles for video surveys for either characterising the sampling environment, or as a primary observation tool, is increasing. The former application provided information for definition of strata, similar to using multibeam sounders or echosounders with Roxanne or Questar Tangent equipment. The actual survey of the fish populations will be carried out using the more traditional surface platforms. There may be survey design issues concerning the collection of the habitat information but these were not discussed here.

Video surveys can also be used as the primary observing tool. For towed bodies or submersibles, monitoring of the actual size of the sampling unit is necessary, as it is difficult to keep these vehicles at a set depth. The size of the sampling unit is determined by the distance of the video gear from the bottom, and changes in the height off the bottom will have an impact on the detectability of the target species off of the camera centreline. Videos and photographs may be interpreted manually or through the use of image analysis software. In either case, it may be useful to consider standards for dealing with edge effects and detectability issues, e.g., distance sampling methods. In many cases it is difficult to get size composition information from the video or photographs necessary to convert counts into biomass.

Discussion of the Fisheries Research Services *Nephrops* survey was used to focus on the requirements for using a towed video sled as the primary observing tool. Sled surveys are used to count *Nephrops* burrows around Scotland, assuming that all burrow systems within a strip of width 1 m are detected. Although this approach overestimates *Nephrops* density by around 30%, using the counts as a relative measure of abundance hasn't proved problematic. However, it should be fairly straightforward to estimate absolute abundance using distance sampling methods, which would be more informative for stock management and would allow greater comparability across surveys conducted by different organisations.

Assuming that all burrow systems within a given strip width are detected introduces three potential sources of bias that could be significant. First, burrows at or just beyond the edge of the strip tend to be included by observers conducting strip transect counts: because the strip is so narrow for *Nephrops* surveys (1 m), the percentage bias from this source can be large. Second, this bias can be exacerbated because the units recorded are the systems of burrows. If a system is mostly outside the survey strip, but one or more of its burrows falls within the strip, it will be included as inside the strip. Thus in effect, a strip appreciably wider than the nominal 1 m might be surveyed. Both of these sources of bias would tend to inflate the density estimate. The third source of bias arises if some burrow systems inside the strip but close to the edge of visibility are not detected. This may be due to bottom topography, poor visibility, or simply the small size of some of the burrows. This would generate downward bias in the density estimate.

These biases could be avoided by applying distance sampling methods (Buckland *et al*. 2001) to surveys of individual burrows. By using individual burrows as the units, bias arising by including burrow systems in the strip whose centres lie outside the strip is avoided. A subsample of burrows can be examined to estimate the mean number of burrows per system, together with a standard error. This allows conversion of burrow density to system density. Further, by measuring the distance of each detected burrow from the transect centreline, bias arising from including edge burrows in the strip count is avoided. Recording of these distances also allows a detection function to be modelled, representing the probability that a burrow is detected from the video, as a function of distance from the transect centreline. Typically, this probability is assumed to be one at or near the centreline, but is allowed to decrease with increasing distance from the line. Existing video could be reanalysed to generate suitable line transect data, to test the method.

Additional survey work is needed to monitor *Nephrops* on sample plots. This would allow the estimation of the proportion of burrows that are occupied. It would also provide data for assessing the relationship between burrow size and animal length, and for estimating the size at which *Nephrops* become large enough to survey using burrow counts. *Nephrops* below a certain size cannot be reliably surveyed by counting burrows. Their burrows may be impossible to identify as *Nephrops* burrows, and they may occupy burrows off a larger burrow. Also, size distribution cannot at this stage be reliably estimated from observed burrow sizes.

A possible candidate for line or strip transect methods is the monkfish. Increasing demand for monkfish, together with the absence of assessments in European waters, have led to calls for surveys, to allow precautionary TACs to be replaced by TACs that reflect population size. Traditional VPA-type assessments are not feasible for monkfish, and reliable fishery-independent survey methods are needed. Line transects surveys, using a submersible or a towed video camera, may therefore be useful, if the strip width can be made sufficiently wide. If standard distance sampling methods are to work, monkfish on or near the centreline of the transect should be almost certain to be detected, whereas fish further from the line do not need to be detected with certainty. Distances of detected fish from the line would be measured, and probability of detection as a function of distance from the line would be modelled using these distances (Buckland *et al*. 2001).

4.1.4 Biological parameters in acoustic surveys

In most acoustic surveys, species identification is based on *ad hoc* tows. More objective ways for allocating tows in space would support the computation of overall estimation variance for those surveys. Meanwhile, the biological parameters provided by these tows must be attributed in some way to the locations of acoustic measurements. This step is open to improvement, particularly in the interpolation (mapping) of length frequency distributions; currently these are mapped using a nearest neighbour approach or an ad hoc block mapping method (see Section 3.2.2 and ICES 2004a for examples).

4.2 Common design and analysis issues and problems

4.2.1 Adaptive sampling

Adaptive sampling was used in the larval sea lamprey, scallop, and *Nephrops* surveys. Adaptive cluster sampling was used for the sea lamprey survey, while the other two surveys used the adaptive allocation approach. All of these surveys used stratified random designs, for which adaptive estimates are available in the literature (Francis 1984; Thompson and Seber 1996). Experience with these surveys suggests that adaptive sampling was successful for increasing precision. However, it was noted that in each case we were dealing with a single sedentary species that was the primary target of its respective survey. In addition, the geographical limits of the distribution of the species were relatively well known.

Sea lampreys provide an excellent example of when adaptive sampling can be effective. However, as control efforts in the Great Lakes continue to contribute to the declining abundance of sea lampreys, sea lamprey larvae in stream sediments will become harder to find, and ongoing assessments of the sea lamprey population will become more challenging. Increasing the first phase of sampling effort to find the larval concentrations so that the network sampling can be initiated is not cost effective because of the large expanse of habitat. Yet, accurate assessment of their numbers and distribution are essential in applying effective control measures. Adaptive sampling could become progressively less efficient than single phase stratified sampling in this situation.

4.3 Analytical methods

4.3.1 Case studies

Six presentations given on analytical methods which although did not cover the entire range of analytical approaches used in survey data analysis were sufficient to discuss common problems.

Two presentations combined the use of simulations and data truncation (or abundance class “binning”) to deal with extreme values. A geostatistical conditional simulation (Lantuéjoul 2002) of Bering Sea walleye pollock rendered the full distribution of the abundance estimator that was compatible with both the sample data and a function characterizing the autocorrelation. The latter study employed an indicator simulation which solved some problems related to the simulation of the two extremities of the distribution (zero’s and extreme values). This study also highlighted the need to improve the interpolation of biological parameters (mainly length frequency). In another type of simulation, herring schools were mapped using observed schools sizes and internal densities placed on an underlying probability distribution derived from several years of survey data. The objective was to reproduce both the short term random movement of schools and their long term migration, and to examine the impact of the fish motion on the global biomass estimate. In the latter case fish motion does not induce significant bias in the global estimation.

A non-linear approach using General Additive Model (GAM) with longitude, latitude and depth as explanatory variables was presented for estimating sardine biomass from an acoustic survey off Portugal. Zero data happen to be clustered providing information on the border of the population (structural zeroes). A two steps approach is then used. First, a presence/absence GAM model generates a map of the probability to enter the population. A second model aims at modeling the sardine inner distribution.

A review of geostatistical concepts involved in global estimation variance was presented. The main steps of a geostatistical analysis are the delineation of the field of positive values and the definition of a variogram model to characterize the autocorrelation present in the data as a function of distance. This model allows for the quantification of how well the arithmetic mean of samples represents the mean fish concentration of the delineated field. Major difficulties remain in lowering the variance significantly, because of the extreme values that occur in surveys of both pelagic and demersal fish populations.

To address the question of the extreme values, a non linear approach based on the spatial correlation structure between pairs of cutoffs, conditionally to one of them was suggested. The conditioning idea is helpful in the analysis of adaptive sampling because adaptive sampling is based on adding extra samples conditionally to already sampled ones. This method also provides means to post-stratified survey data.

Taking a sample at random in a quadrate allows an unbiased estimation of the quadrate density. In practice, that means that over a large number of quadrates, those which are underestimated are compensated by the overestimated ones. However, selecting the quadrates whose inner sample is larger than a given cut-off breaks down this unbiasedness: the mean density of these selected quadrates is by no means equal or smaller than the sample values (conditional bias, as opposed to the overall bias). Attention is then drawn on the fact when the decision to make a sample rely on the fish density itself (adaptive sampling, trawl location in acoustic surveys), this should be based in practice not on the sample values but on an estimate of the density over the quadrate used in the survey.

Comparison of survey designs highlighted the usefulness of Kish’s design effect (the ratio of the variance of an estimate obtained from a complex design to that obtained from a simple random sample of the same size, Kish 1965) and of the effective sample size, which can be used to weight estimates from several surveys in order to combine them (see also Section 5.4). It is also important to take the cost of the survey into account when considering survey designs.

4.3.2 Design-based and model-based approaches

Design-based methods guarantee unbiased estimation (and unbiased variance estimators) without any model assumptions. Unbiased estimation of means, or totals, simply require that the probability that any given sampling unit selected is known (and non-zero), while unbiased variance estimation has the additional requirement that the joint selection probability of any pair of units be known, and greater than zero[3]. This can be guaranteed by design. Inference is based on the variability that would occur from one realisation of the design to another. The methods remain valid whatever the properties of the data. For example, the observations may exhibit autocorrelation or non-normality without invalidating estimates or variances, although in extreme cases, estimates of variance may be very imprecise, for example when observations are dominated by one large value. Confidence intervals are usually, but not always, based on an assumption of normality, but, for reasonable sample sizes, the Central Limit Theorem ensures that this assumption is satisfactory.

Systematic surveys tend to give better precision for abundance estimates than does simple random sampling. However, a single systematic sample with a random start does not provide the replication needed to estimate variance using design-based methods because the joint inclusion probabilities of pairs of units are not all greater than zero. Hence the variance is often estimated assuming the systematic sample was selected using simple random sampling. This procedure typically gives a positively biased estimate of the true systematic sampling variance for autocorrelated populations (Cochran 1977, p. 220). In marine abundance surveys, pairs of observations that are close together are likely to be more similar than observations distanced further apart, resulting in positive autocorrelation. The bias introduced by treating a systematic sample as a simple random sample might be substantial in the presence of strong positive autocorrelation.

Model-based approaches aim at using this autocorrelation in an explicit manner in order to estimate estimation variance for many different survey designs. Geostatistical techniques, for instance, provide a means of estimating the variance for systematic survey designs (Wolter 1985, pp 250-253, provide alternative methods). Two situations can be considered. The so-called transitive approach is based on the random uniform characteristic of the starting point of the grid. This ensures unbiasedness of the global estimate and of estimation of the autocorrelation function (covariogram). Estimates of estimation variance can then be fully undertaken. When the systematic design covers a predefined survey area, one can estimate the mean concentration over the survey domain by the arithmetic mean of the inner samples. In this case, the variogram model is needed to determine the estimation variance. The methods also remain valid whatever the properties of the data. The distribution of the error is then not known, and the estimation variance is usually simply turned into a coefficient of variation (CV) for the estimation. In practice, a structural model has to be defined, and the quality of the overall procedure depends on the quality of its estimation.

4.3.3 Simulations of spatial process

One may distinguish between different kinds of simulations. Two different kinds of simulations have been considered during the workshop: the simulation of a probabilistic model defined by an autocorrelation function (with an implicit assumption on the probability distribution) and the simulation of a deterministic spatio-temporal model. In the first instance, simulations consist of producing numerical fields that honour a given autocorrelation function. Many algorithms exist; some are based on approximations; some others, more demanding in terms of computer power, provide exact simulations. Conditional simulations correspond to simulations that also honour sets of known georeferenced data. In the second case, simulations represent the time evolution of a given state according to the spatio-temporal equations defining the dynamics of the system.

Such simulations can be used to estimate the full distribution of the estimation of the global mean abundance and to derive risk analyses. They also allow hypothesis testing (potential gains of adaptive sampling, effects of fish motion) as they can be surveyed while knowing the true underlying fish concentration. The results of the simulation should not be the sole criterion for choosing a particular survey design. As a matter of fact, the reasons for one particular survey design performing well could be related to some characteristic of the simulations that may not correspond to the true field of fish distribution (simulations are only mimicking models, not reality).

4.3.4 Non-linear approaches

Dividing the data in groups of abundances classes (sometimes referred to “binning” the data due to the analogy of allocating data into various “bins”) can improve model-based estimation procedures particularly with regard to extreme values. Using cutoffs or “bins” is by nature a non-linear transformation of the data which was demonstrated on several occasions during the workshop: simulation the deciles of the distribution of walleye Pollock densities through geostatistical indicator simulations; modelling of the presence-absence of sardine using a GAM with a ‘logit’ link function; and post-stratification through disjunctive kriging. In this non-linear framework, the question of spatial structure concerns the analysis of the correlation between pairs of cut-offs, conditionally on one of them. In particular, one can address the question of whether areas of high fish density are, or are not, randomly located in areas of medium

---

[3] This is a requirement for the Horvitz-Thompson estimator. If the joint inclusion probabilities of pairs is not known (as for a systematic sample) then there is no unbiased variance estimator.

density (edge effects). Using a zero cut-off amounts to an analysis of the spatial correlation between areas of presence and absence (structural zeroes).

Occasional occurrence of extreme values can imply that survey stations are too sparse in regions of aggregation. Unfortunately, increasing the number of stations locally is likely to be very costly and, furthermore, the discovered aggregations are not necessarily consistently located in one place from year to year. However, if means can be found to gain further definition of zones of aggregation, and precautions are taken during the analysis to guard against possible *a posteriori* bias due to adjustment of sampling in the light of results, the precision of the survey estimates could be improved. One possible low cost method is to ask fishers to find or advise on the whereabouts of aggregations of fish; this is part of their usual work. Bias can also be avoided by using adaptive allocation (Thompson and Seber 1996), or by redefining sampling strata for future surveys.

4.3.5 Calculation protocols

Several participants had encountered problems in recreating series of abundance indices given the original data, e.g., as found in the Evaluation of Research Surveys (EVARES) study (Beare *et al.* 2003). There appeared to be a problem that documentation for the calculation procedures was not, in all cases, well known or widely available. This makes any critical appraisal of the procedures time consuming or impossible for those not immediately involved in collection of the data. It also raises concerns that the procedures may be changing over time, leading to drifting bias in the abundance indices. The group agreed that all publicly funded surveys, particularly those benefiting from EC funding, should make their estimation procedures readily available as protocols, preferably on websites.

5 Areas of agreement and specific areas of work where progress could be made

5.1 Choice of survey design: simple random, stratified random or systematic

This section considers the performance of three different sampling designs for estimating the population mean. In particular, it compares:

- simple random sampling
- stratified random sampling, where the survey area is divided into equally sized strata and an equal number of samples are taken in each stratum
- systematic sampling with a random start point

Moving from simple random sampling, through stratified random sampling, to systematic sampling, corresponds to increasing the regularity with which samples are spread across the survey area.

The sample mean is an unbiased estimate of the population mean for all three sampling designs. But how does the variance of the sample mean depend on the sampling design; and how well can that variance be estimated ? These issues are discussed by Cochran (1977; chapters 5, 5A, 8); Wolter (1985; chapter 7); Thompson (1992; chapter 21); Ripley (1981; chapter 3); and Matheron (1971). Unfortunately, there are no clear-cut conclusions, but some of the salient results are given here. We use the notation V_{random}, $V_{stratified}$, and $V_{systematic}$ to denote the variance of the sample mean under simple random sampling, stratified random sampling and systematic sampling respectively.

5.1.1 Variance of sample mean

- Under any conditions:

  $$V_{stratified} \leq \frac{N-1}{N-H} V_{random}$$

  where N is the number of possible sampling units in the population and H is the number of strata (see Cochran 1977, Section 5.6). In the worst case scenario, $V_{stratified}$ is only marginally greater than V_{random}.

- If the number of sampling units is large compared to the number of strata (as is usually the case for fish or shellfish populations):

  $$V_{stratified} \leq V_{random}$$

  (Matheron 1971; Cochran 1977, Section 5.6). There is never any real loss in using a stratified random survey and when there are large differences between strata means Vstratified is much smaller than Vrandom.

- For particular types of positive local autocorrelation:

  $V_{systematic} < V_{stratified} < V_{random}$

  (Cochran 1977, p220). However, it is easy to construct examples where Vsystematic > Vstratified (e.g., Cochran 1977, p215).
- Cochran (1977, Section 8.10) compares systematic sampling with stratified random sampling for several natural populations and shows that systematic sampling was more precise than stratified random sampling in all but one case.
- Usually, stratified random and systematic sampling will do well compared to simple random sampling.

5.1.2 Estimators of the variance of the sample mean

The design-based estimator of the variance (e.g., Cochran 1977; Thompson 1992) is unbiased for simple random surveys and for stratified random surveys with at least two samples per stratum. There are $n - H$ degrees of freedom available for estimating the variance, where n is the total sample size and H is the number of strata (H=1 for the simple random survey).

For stratified random surveys with one sample per stratum and systematic surveys, the variance can be estimated by 'combining' adjacent points (as if they had come from a stratified random survey with two samples per stratum) and using the design-based estimator (Thompson 1992, p119; Cochran 1977, Section 5A.12). This process is also known as 'collapsing the stratification'. The estimator is usually positively biased. Alternatively, a model-based estimator can be used, but this requires assumptions to be made about the population.

The properties of the different sampling designs are illustrated below using two examples.

5.1.3 Simulated data

Figure 10 shows six populations, each with the same 10,000 randomly selected numbers distributed over a 100 × 100 grid. Each population has the same mean and variance. However, the 10,000 numbers have been reorganised to have increasing spatial autocorrelation (Figure 10). Each population was sampled many times by a random survey and by a systematic survey (with a random starting point). Both the random survey and the systematic survey had a sample size of 64. Figure 11 shows that for these populations:

- the variance of the sample mean under random sampling is independent of the autocorrelation in the data;
- the variance of the sample mean under systematic sampling decreases as the autocorrelation increases;
- the variance of the sample mean under systematic sampling is smaller than that under random sampling whenever there is autocorrelation;

5.1.4 Simulations based on North Sea herring acoustic survey data

Simmonds and Fryer (1996) investigated survey strategies for a variety of populations with properties based on North Sea herring. The populations had different mixtures of local positive correlation, a short-scale random component, and a non-stationary or trend component. Several survey strategies were employed, each taking n = 40 samples:

- random sampling;
- stratified random sampling where the survey area was divided into:
- 2 strata and 20 samples were taken in each stratum;
- 4 strata and 10 samples were taken in each;
- 8 strata and 5 samples…;
- 20 strata and 2 samples…;
- 40 strata and 1 sample…;

(Note that random sampling is equivalent to stratified random sampling with 40 samples in 1 stratum)

- systematic sampling with 40 samples and a random starting point;
- systematic sampling with 40 samples and a centred starting point.

The performance of each survey strategy for each population was investigated by simulating many surveys of many realisations of the population. The results from all the populations were broadly similar and are illustrated in Figure 12. The variance of the sample mean always decreased as the amount of stratification increased, with the systematic surveys giving the most precise estimate of the population mean.

Simmonds and Fryer (1996) also considered the precision and bias of estimators of the variance of the sample mean.

- the usual design based estimator of the variance (e.g., Thompson 1992) was used for all the sampling designs (with the strata collapsed in pairs for the stratified random design with one sample per stratum and for the systematic designs);
- a geostatistical estimator was used for the stratified random designs with one or two samples per stratum and for the systematic designs.

Again, the results were broadly similar across all the populations, although there was some differences that depended on the amount of local autocorrelation. For example, Figure 12 shows the median and 90% intervals of each estimator for each sampling design when the range of the autocorrelation was about 5% of the survey area. Figure 13 shows the width of the 90% intervals across all the populations.

- the design based estimator of the variance was unbiased when there were at least two samples per stratum (this is always true);
- the design based estimator of the variance was positively biased (by about 50%) for the stratified random design with one sample per stratum and for the systematic designs;
- the geostatistical estimator of the variance was approximately unbiased;
- the shortest 90% interval was obtained for the stratified random survey with two samples per stratum (both the design based estimator and the geostatistical estimator had similar 90% intervals for this design).

In general, increasing the stratification decreases the variance of the sample mean, but also reduces the degrees of freedom available for estimating that variance. In terms of the variance estimators, this means that the mean variance decreases, but the width of the 90% interval relative to the mean increases. Thus, the best sampling strategy for variance estimation depends on the balance between these two effects.

Using the design based estimator of variance for the stratified random sampling design with one sample per stratum or for the systematic surveys provides a relatively precise but biased estimate of the variance. Combining more than two adjacent strata is also an option (particularly if degrees of freedom are limited), but would increase the bias in the estimate of variance still further. A danger then would be that the survey is perceived to be less precise than it really is and that efforts to improve the survey might be misdirected.

In practice, the choice of survey will be a trade-off between getting the most precise estimate of the population mean and the best estimate of variance. For North Sea herring, an abundance estimate is required annually, but a variance estimate is less important (the time series of surveys are modelled to provide estimates of variance), so a systematic survey with a random start is used.

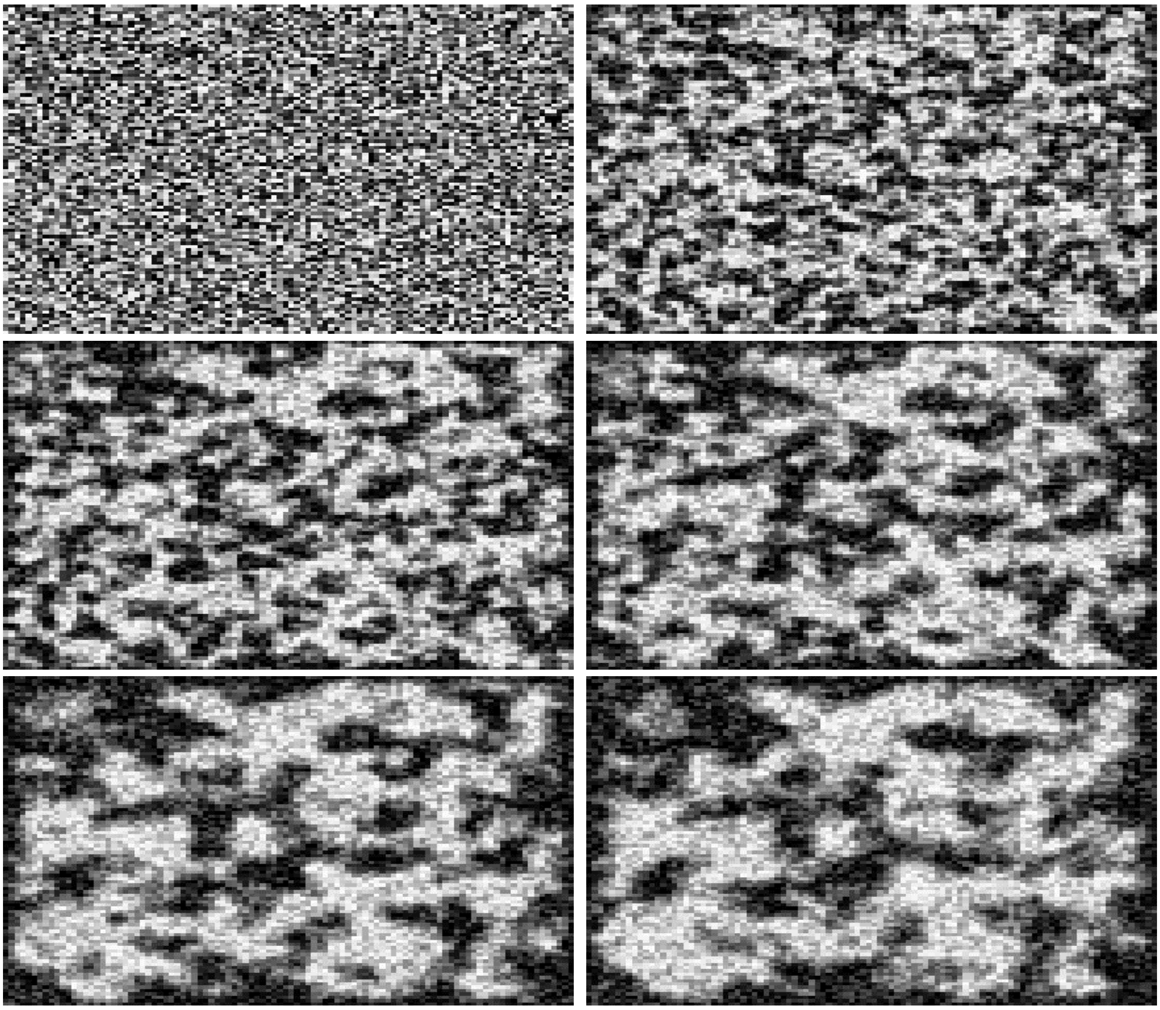


Figure 10. 10,000 random numbers organized in space with increasing spatial autocorrelation from top left to bottom right.

5.1.5 Conclusions on the choice of survey design

Consider the three design options:

- simple random sampling
- stratified random sampling, where the sample area is divided into equally sized strata and an equal number of samples are taken in each stratum,
- systematic sampling with a random starting point

Assume that the total sample size is the same for each design, and that the sample mean is used to estimate the population mean.

In the presence of positive local autocorrelation, a more precise estimate of the population mean will usually be obtained by stratified random sampling or systematic sampling than by simple random sampling. The optimal sampling design will depend on the population under study and the relative importance attached to getting the most precise estimate of the population mean and to getting a good estimate of that precision. A wide range of real and simulated examples suggest that systematic sampling will often be optimal if getting the most precise estimate of the sample mean is the dominant objective. However, stratified random sampling will often be preferable if getting a good estimate of the precision is also important.

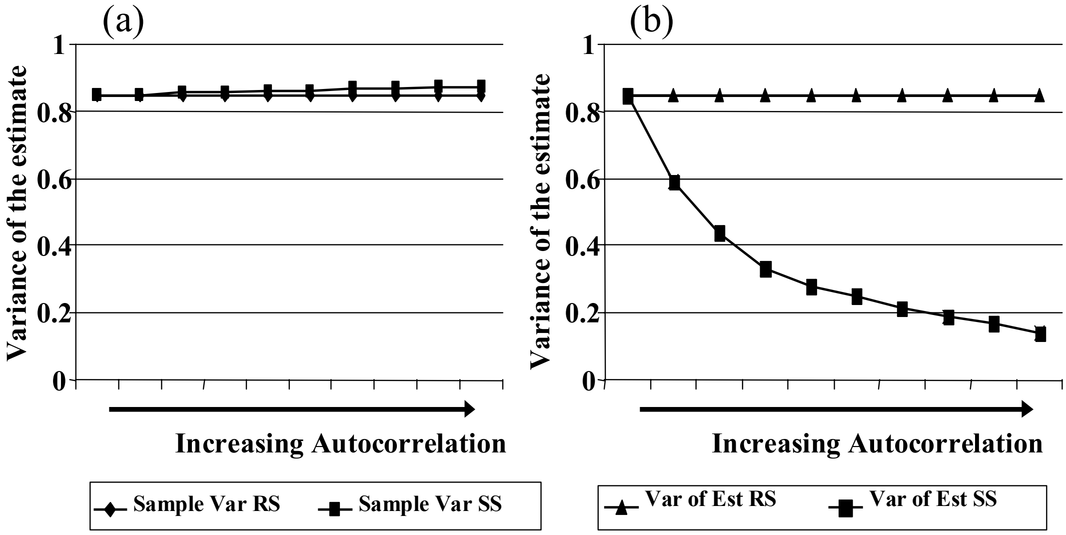


Figure 11. The precision of a random and a systematic survey for the spatial distributions shown in Figure 9: a) the sample variance /n; b) the variance of the sample mean. The random survey estimates the population mean with a variance that is independent of the autocorrelation; the sample variance can be used to obtain an unbiased estimate of the variance of the sample mean. The systematic survey estimates the population mean more precisely as the spatial correlation increases; the sample variance no longer provides an unbiased estimate of the precision of the sample mean.

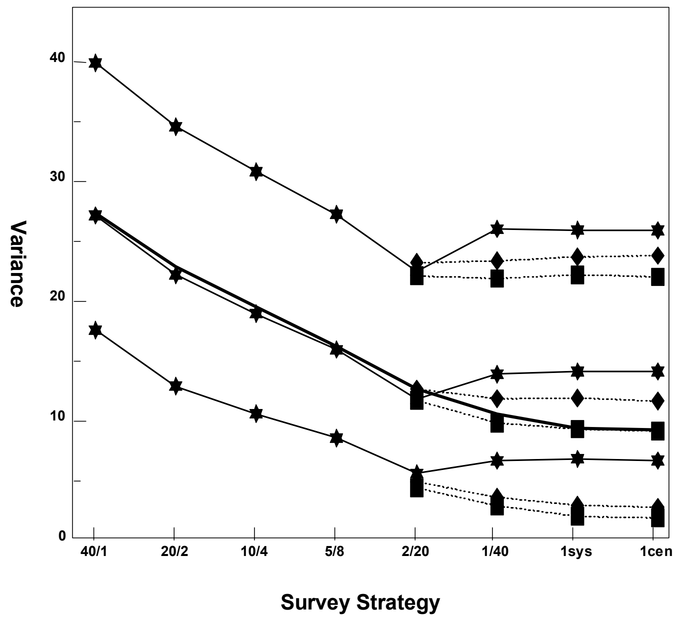


Figure 12. The variance of the sample mean (thick line) for a series of different survey strategies applied to a simulated population with properties based on North Sea herring. The strategies are 40 samples in 1 stratum (40/1), 20 samples per strata in 2 strata (20/2), …, 1 sample in each of 40 strata (1/40), a systematic survey with 40 samples and a random start (1sys) and a systematic survey with 40 samples and a centred start (1cen). The 5, 50 and 95 percentiles of the design based variance estimator (*) and the geostatistical variance estimator (■) are also shown. The design based variance estimator is unbiased for 40/1 to 2/20 samples/strata. The geostatistical estimator is approximately unbiased for the 2/20 and 1/40 designs. The sample mean is most precise for systematic designs; the variance estimator is most precise for 2 samples per strata in 20 strata.

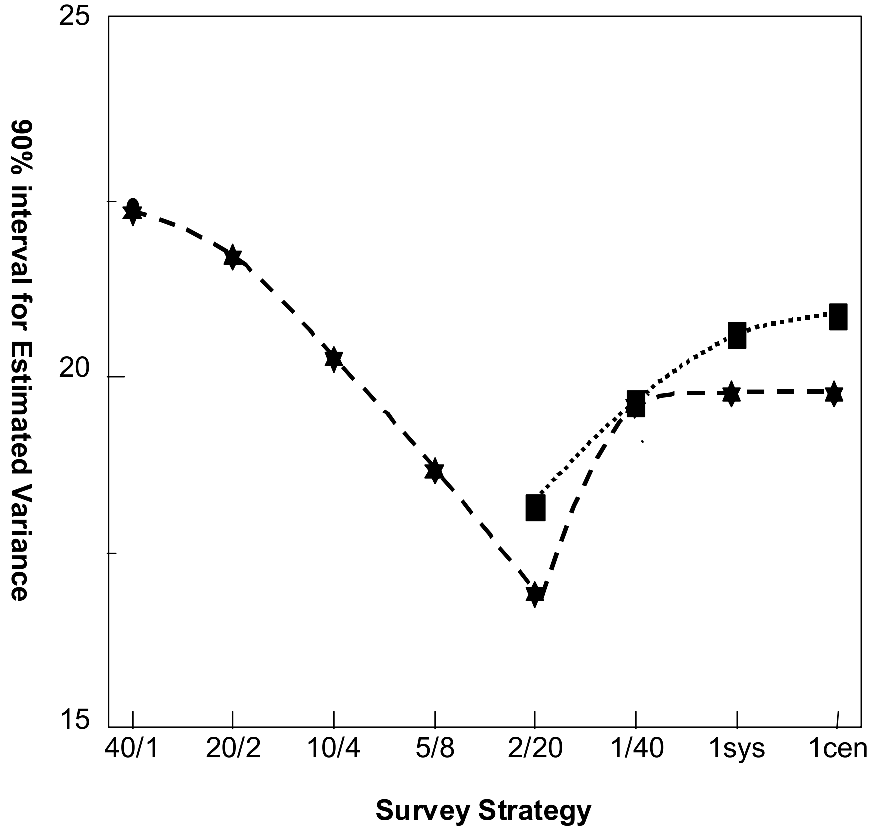


Figure 13. The width of the 90% interval for the design based variance estimator (*) and the geostatistical variance estimator (■) for various survey designs applied to a simulated North Sea herring population (the survey designs are defined in Figure 11). The shortest 90% interval is obtained for the survey with 2 samples per strata in 20 strata. Although the geostatistical estimator and the design based variance estimator have similar 90% intervals for the 1/40 design, the design based variance estimator is biased for this design.

5.2 Fixed survey designs

Designs that keep an initial probability-based selection of units fixed over time can be effective for detecting trends when the spatial distribution of the population being surveyed is persistent (Warren, in ICES 1992). Such designs can be cost-effective for bottom trawls surveys in areas with significant un-trawlable bottom habitat, for example, because visits to unfavourable locations are eliminated. One drawback with a completely fixed design is that the condition for obtaining unbiased estimates of the variance of means or totals is not met (after the initial selection, all other units have a zero probability of being selected). A combination of fixed and random stations, with a sub-set of stations being matched from one survey to the next, is an alternative strategy that works well for estimating both status and trends. This method, which often is referred to as sampling with partial replacement, was first examined by Jessen (1942), and is discussed in Cochran (1977) and Jessen (1978). Cochran (1977, p. 346) shows how matched and unmatched stations can be combined to provide overall estimates of the mean and sample variance. Unless the correlations of matched samples are high, the fixed design is only marginally better than the matching of 25% to 50% of stations for estimating change (Cochran 1977).

5.3 Reporting precision

The precision of fish surveys is commonly described using the coefficient of variation (CV) as a parameter, and the CV may be specified in the plan as a design criteria. Most statisticians know CV as the sample standard deviation divided by the mean (e.g., Tietjen 1986). Kish (1995) describes the *element coefficient of variation*, denoted as C_x and defined as the standard deviation divided by the mean; the *coefficient of variation of the mean* is denoted as $CV_{\bar{y}}$ and defined as the standard error divided by the mean. Reporting a CV without defining it is, therefore, dangerously ambiguous. Furthermore, its statistical properties are not very good and, as a result, its use is diminishing and should be discouraged (Pagano and Gauvreau 1993).

Participants recommend that past use of CV in ICES reports should be closely scrutinised to ensure that its use is consistent between reports and adequately explained. In future, the use of CV as notation should be avoided and replaced by the *relative standard error* (RSE) reported as a percentage. RSE is defined as: 100% × standard error / estimate (Jessen 1978).

5.4 Analytical evaluation of design efficiency

Given the expense of abundance surveys, it is important to optimise the survey design so that the precision of key parameters such as relative abundance indices is maximised for a fixed total survey cost. One way to evaluate the efficiency of a probability-based survey is to compare the variance of the estimated mean CPUE ($\bar{y}$) for the actual design (and variance estimator employed) with the expected variance obtained under simple random sampling. Kish (1965; 1995; 2003) defined the *design effect* as the ratio of the two variances:

$$deff = Var_c(\bar{y}_c) / Var_{srs}(\bar{y}_{srs})$$

where $Var_c(\bar{y}_c)$ is the variance estimate based on the actual (*c* = complex) survey design using an appropriate estimator, and $Var_{srs}(\bar{y}_{srs})$ is the expected design-based estimate under simple random sampling (*srs*) for a sample of equal size. The design effect, in contrast to the relative standard error (Jessen 1978), removes the effect of sample size. Kish (1995) provides a general discussion on the calculation of design effects. Using a stratified random survey with proportional allocation as an example, the expected $Var_{srs}(\bar{y}_{srs})$ can be estimated straightforwardly by treating the stations as a simple random sample from the total survey area. This is justified because stations across strata have equal inclusion probabilities. SUDAAN (RTI 2001), specialised software for the analysis of complex surveys and cluster-correlated data (Brogan 1998; Carlson 1998), is useful for estimating the design-based variance of an estimated mean, and also provides estimates of design effects for a wide range of probability-based survey designs.

The *effective sample size* for estimation of mean CPUE ($\bar{y}$) from the complex survey design is defined as the number of samples selected by simple random sampling that would be required to achieve the same precision obtained with *n* samples under the actual complex sampling design,

$$n_c^* = n / deff$$

If, for example, the design effect equals two for the estimated mean CPUE for a transect survey with 60 stations, then a simple random sample of 30 stations (the effective sample size) would have been expected to achieve the same precision. The effective sample size depends on the survey design as well as the variance estimator. For acoustic surveys of populations with strong and consistent auto-correlation in space (e.g., North Sea herring) the use of model-based estimators appears to increase the effective sample size significantly by reducing the variance in the estimated mean abundance (see Section 5.1).

Another approach for evaluating stratified random designs is to decompose the difference between the stratified random variance and the simple random variance for a sample of the same total sample size. Formulae are available for estimating the simple random variance from a stratified random sample (Cochran 1977; Gavaris and Smith 1987). The difference between the variances can be characterised as a component due to the stratification scheme and a component due to the allocation of sampling stations to strata. In particular, the allocation component could help identify problems with lack of precision in the survey. The allocation scheme will lead to equal or increased precision if the allocation of stations is proportional to strata size or strata variance, respectively (Smith and Gavaris 1993a). An arbitrary allocation scheme could actually result in a stratified random design being less precise than a simple random sample no matter how good the stratification scheme is.

Evaluation methods for surveys should explicitly include the design in the methodology. Examples for evaluating influential observations with respect to abundance and confidence interval estimates for stratified random survey designs are given in Smith (1996). The advantage of comparing the estimation variance to what it would be for independent samples (σ^2/N), is that it is easy to compute and quickly indicates the order of magnitude of any problems. However, the sampling unit over which the data are gathered (the 'support') considered must be explicitly added, since σ^2/N depends on the support (except when there is no autocorrelation). This is particularly important for variables like the Nautical Area Scattering Coefficient in acoustic surveys, that are recorded (essentially) continuously, where the support (Equivalent Distance Sampling Unit, MacLennan and Simmonds 1992) can vary. The geostatistical variance is consistent when support varies, but not σ^2/N (see e.g., Rivoirard *et al.* 2000, p. 135).

5.5 Combining two surveys

When two independent probability-based surveys are conducted over the same population, a composite estimator can be used to take a weighted average of the two survey indices (e.g., Korn and Graubard 1999; Rao 2003). The estimator for the combined mean for a stratified random survey and a survey with trawling along transects, for example, is:

$$\bar{y}_{comb} = \phi\bar{y}_{str} + (1-\phi)\bar{y}_t$$

with the weight, ϕ ($0 \le \phi \le 1$), chosen to minimise the variance of $\bar{y}_{comb}$,

$$Var(\bar{y}_{comb}) = \phi^2 Var(\bar{y}_{str}) + (1-\phi)^2 Var(\bar{y}_t)$$

where $Var(\bar{y}_{str})$ and $Var(\bar{y}_t)$ are appropriate estimates of the variances of the mean CPUE estimators for the stratified random and transect surveys, respectively. The optimum weight, expressed as a function of the effective sample sizes for each survey (n^*_{str}, n^*_t), is:

$$\phi_{opt} = \frac{n^*_{str}}{n^*_{str} + n^*_t}.$$

5.6 Ecosystem monitoring

There was general agreement that any one survey design or survey gear could not monitor enough of the ecosystem to provide data required in support of advice on ecosystem management. Instead, multiple surveys using a variety of gears and designs may be necessary. Collaborative monitoring of seafloor, fish and epibenthic species over a large marine area was successfully demonstrated by Callaway *et al.* (2002) and Zuhlke *et al.* (2002). They used standardised sampling protocols, which involved towing a lightweight, 2-metre beam trawl briefly between the primary trawling stations, by several countries participating in the North Sea IBTS. However, methods for combining information from different surveys and gears will need to be developed when standardised sampling protocols cannot be arranged. For example, the EU-CATEFA project (Combining Acoustic and Trawl data for Estimating Fish Abundance) aims at using both acoustic and trawl data from bottom trawl surveys as a cost-effective improvement of biomass estimation. This is in recognition that bottom trawl surveys are the most important fisheries-independent data source used in stock assessment of commercial groundfish in European waters. The inclusion of simultaneously collected acoustic data could potentially improve the precision and accuracy of these surveys at little extra cost.

Good consistency has been observed at the study scale, between acoustic data recorded both during trawling and during steaming. However, correlation between acoustic observations and catch data is weak in two (North Sea, Irish Sea) of the three areas used in the study precluding the development of combined tools at an operational level. In the Barents Sea surveys, combined models can be developed leading to improved interpolated maps that provide finer details of the spatial distribution and lower estimation variance.

The merits of the project rely on the relevancy and accuracy of data collection and preparation. Algorithms (manual, semi-automatic and automatic) have been developed to avoid confusion between bottom and fish detection. However, it is suggested that the actual integration of backscattering energies in regular depth layers is not optimal with regards to the objective of combining acoustic and trawl data. Preliminary analyses indicate that identifying the proportions of backscattering energies by species could lead to improved correlations between species-specific acoustic measurements and trawl catches.

5.7 Incorporating Environmental and Habitat Covariates to Improve Survey Results

5.7.1 Incorporating Environmental and Habitat Covariates - Post-Survey Analyses

Many studies have shown that there are relationships between catches and environmental variables (e.g., Perry and Smith 1994; Kostylev *et al*. 2003). In many cases these relationships appear to be clear and straightforward enough to use in increasing the precision of the survey estimates.

For design-based surveys there are currently three methods for incorporating these relationships into the survey estimates. The most straightforward approach has been to either design strata that correspond to homogeneous areas with respect to the environmental variable or in the case of existing surveys, post-stratify the survey stations based upon the homogeneous areas. For post-stratification, the new strata boundaries need to be known. Sample size becomes a random variable and adds additional variance to the estimates. Generally, this additional variance tends to be small relative to the total variance. In addition, not all of the new strata may have tow locations in them. An example of post-stratification for scallops using sediment type is given in Smith and Robert (1998).

A number of authors have promoted the prediction approach (Valliant 2000; Smith 1990). For this approach the survey domain is partitioned into those stations observed during the survey and those that were not in the survey. If there is an auxiliary variable available that exhibits a relationship with catch and this variable is known for all sites, then this relationship can be used to predict catch for the unsampled sites. Design-based estimates and associated variances that include the component due to estimating this relationship are available. Applications of this approach were presented by Smith and Robert (1998) on scallops with sediment type and by Adams on spawning sea lampreys with drainage area and treatment history (Mullett *et al.* 2003).

Chen *et al*. (2004) have developed a model-assisted approach based on empirical likelihood. Similar to the previous two approaches this method requires complete knowledge of an auxiliary variable(s) over the whole area. The advantage of this method is that a wider range of models appears to be available in addition to time series model. Chen *et al*. (2004) present an example for yellowtail flounder using a variety of covariates on the Grand Banks.

Recently, attention on the Scotian Shelf has focused on using surficial geology information interpreted from side-scan and multibeam data to improve survey design for scallops (e.g., Kostylev *et al*. 2003). This study is ongoing and will include comparisons of the performance of the estimators discussed here.

5.7.2 Incorporation of Environmental and Habitat Data: Survey Design Considerations

Many current surveys employ relatively simple sampling designs that are based on limited design parameters (e.g., depth, region). Information on environmental (e.g., temperature, salinity, currents) and habitat (e.g., surficial substrate) are becoming more available The temporal scale upon which these parameters can vary on a continuum of temporal scales ranging from tidal cycle and daily scales to parameters that are likely fixed through the lifespan of most surveys. Examples of parameters that are likely to vary on tidal cycle or daily temporal scales include salinity, temperature and light. Parameters that are likely fixed over the lifespan of a survey include latitude and depth.

As information on environmental and habitat data become both available and accessible on real and near-real time scales, considerable potential exists for greater incorporation of these variables into fixed and hybrid fixed adaptive survey designs. A hybrid fixed adaptive design might include an underlying survey design with some components of stratification or sampling effort allocation that can be adaptively applied based on dynamic environmental conditions. There is a potential for considerable improvements in survey performance if parameters that influence the target organisms' distribution are accounted for in the sampling design.

The ability to include a greater number of parameters in survey design considerations is dependent upon a number of factors including the survey objectives, the number of target species in the survey, and the availability and quality of environmental or habitat parameters. Surveys that have broad survey objectives or high numbers of target species (e.g., multi-species surveys) are likely not strong candidates for greater refinement in survey design. The ability to address multiple objectives including objectives that may be conceived in the future is likely enhanced by relatively simple survey designs. On the other end of the continuum, surveys that target single species and have limited objectives represent prime candidates for incorporation of greater numbers of parameters during the survey design phase, especially if these parameters have been identified to influence the distribution of the target organism.

Strategies for incorporating additional covariates into survey designs depend on the temporal scale of variability of habitat and environmental variables. Variables that demonstrate little variability over time (e.g., surficial substrate) have the potential to be incorporated directly into fixed aspects of a sampling design. An example might be represented by a stratification design that involved stratification by both depth and substrate type. Parameters that demonstrate considerable variability over shorter time scales (daylight, temperature and salinity) may be candidates for consideration in designing adaptive aspects of survey design. An example of this approach for a bottom trawl survey might involve an underlying fixed stratification design, with a certain number of stations to be allocated adaptively to strata based on real or near real time information on an auxiliary habitat or environmental variable. For acoustic surveys, this might involve altering the density of transects in response to prevailing environmental variables. A more complex application might involve alteration of the actual stratification in response to dynamic environmental or habitat parameters.

Increased emphasis on ecosystem monitoring and management and the advent of plans for "Ocean Observing Systems" emphasise the importance of integrating fisheries data collection systems with other ocean environment monitoring systems. Fisheries data collection systems are increasingly being asked to support ecosystem-level information needs. Many of these data collection systems were not originally conceived for this purpose, but data is being utilised in this manner because it is often the only data available for such purposes. If fisheries resource experts fail to recognise these trends, it is likely that decision makers from outside the scientific discipline will play a greater role in framing scientific questions and data collection systems in the future.

5.7.3 Incorporation of Stakeholder Data and Knowledge to Improve Fisheries Surveys

Incorporating stakeholder input into research survey programs includes a continuum of existing and potential interactions. These interactions range from consultation regarding surveys conducted on dedicated research vessels; commercial fishing vessels being utilised as research platforms; utilising or generating fishery information in advance of surveys to target sampling allocation within existing fishery independent survey designs; and targeted cooperative survey efforts to address information needs that cannot or can only be insufficiently addressed by existing fishery independent survey programs.

Most survey programs have incorporated some form of consultation with stakeholders during the establishment of fishery independent surveys that are reliant on fish capture methodologies. Technical expertise related to the design, construction and performance of fishing gear (e.g., bottom and midwater trawls) is more prevalent within the commercial fishing community than within the fisheries scientific community. It is critical that these interactions focus on the adoption of survey gear and deployment practices that address survey objectives (e.g., representative sampling of species and size distributions including pre-recruits) rather than gear that has been designed to achieve commercial fisheries objectives (e.g., maximising catch of marketable species and sizes). Effective communication during these consultations is critical to assure that all parties understand the objectives to be addressed by sampling gear, and to gain confidence among stakeholder groups that adopted gear is capable of producing a representative picture of the targeted population(s). At least one survey program (northeast United States) has established a formal industry and academic advisory panel to provide regular and ongoing input in current and future survey operations.

Conducting scientific surveys aboard commercial fishery vessels (i.e., leasing) without stakeholder involvement in the activities to be conducted is not considered "cooperative" in the sense that the stakeholder group has not had significant input into the survey objectives, data collection and analysis procedures and products to be generated. Use of commercial fishery vessels to conduct surveys has limitations including vessel noise that may influence survey results (particularly in the case of acoustics surveys), limited ability to accommodate scientific personnel, and limited ability to conduct interdisciplinary studies. The use of commercial fishery vessels to conduct surveys and its inherent limitations has been considered by an ICES study group dedicated for this purpose (ICES Study Group on Collection of Acoustic Data from Fishing Vessels - SGAFV).

Considerable potential exists for stakeholder provided information to be used to alter sampling allocation to optimise survey performance. As was the case with the incorporation of environmental and habitat information into survey designs, the desirability to do this is contingent upon the complexity of survey objectives and number of target organisms. In cases where survey objectives or target organisms are narrowly defined, considerable potential exists for utilising auxiliary information on resource distribution to optimise sampling effort allocation. This auxiliary information could be generated by a number of means including the spatial distribution of commercial fishing effort obtained through the vessel monitoring system (VMS), observed commercial trips, or pre-surveys conducted by industry vessels. This information could then be used to adaptively allocate some portion of sampling effort to sample in expected areas of high abundance or high variability provided that statistical estimators are available to prevent varying bias from year to year (survey to survey) as a result. Maintenance of an underlying sampling design remains critical to ensure that the range of the target organism/stock is sampled consistently through time.

5.7.4 Dedicated Cooperative Surveys to Augment Fishery Independent Surveys

In some cases, information needs or stakeholder concerns cannot be directly addressed through design or adaptive sampling effort allocations within existing fishery independent survey frameworks. These information needs and concerns are prime candidates for dedicated cooperative survey ventures between scientists and stakeholders. It is critical for scientists and decision-makers to recognise that such efforts are designed to augment, rather than to replace existing fishery independent survey efforts. Such efforts should be carefully evaluated and planned in advance to ensure that information needs are addressed in a manner that can be incorporated into existing stock assessment and management frameworks. Inclusion of a cross section of stakeholders (scientists, commercial fishery industry, processors, recreational fishery stakeholders, conservation organisations, academic interests) during these efforts often results in greater acceptance of study results.

To be successful, cooperative survey efforts must clearly define the study objectives, sampling gear, sampling protocols, data availability, analyses to be conducted, data products, and uses and management applications in advance of data collection. The establishment of reasonable expectations for how data will be handled and how it is likely to be incorporated into stock assessment and management applications in advance of the survey is critical for establishing and maintaining credibility in the process. Issues related to standardisation and the implications of failing to adequately address standardisation issues should be addressed during the planning phases. During data collection activities, scientific personnel should always be present and in control of the execution of the sampling design. If pre-survey planning has been adequately executed, there should be little or no ambiguity among scientists and stakeholders in terms of executing the survey design.

Stakeholders should continue to be involved in the process once data has been collected to ensure that this phase of the study is transparent. Normally, collected raw data is made available to involved stakeholders once quality assurance and control procedures have been completed. Stakeholders can be kept informed as analyses are completed and it is preferable if the generalised format for these analyses and resulting data products have been specified and agreed to in advance of the survey. Finally, scientists need to ensure that data that have been generated through effectively designed and executed survey designs are evaluated and incorporated into relevant stock assessment and management processes. Failure to do so usually results in damage to the credibility of the scientific community and its relationship with stakeholders.

5.7.5 Guidelines for the Conduct of Cooperative Research Surveys with Stakeholders

1) Select projects with appropriate objectives, levels of complexity, and temporal scale (shorter duration projects). Projects that are of high stakeholder interest require a lesser degree of standardisation and have shorter execution times are preferable.
2) Involve a cross section of stakeholders rather than a targeted sector if possible.
3) Institute a planning process in advance of survey execution to gain consensus relative to study objectives, sampling gear, sampling protocols, data availability, analyses to be conducted, data products and uses/management applications.
4) Establishment of reasonable expectations for how data will be handled and how it is likely to be incorporated into stock assessment and management applications in advance of the survey is critical for establishing and maintaining credibility in the process.
5) Issues related to standardisation and the implications of failing to adequately address standardisation issues should be addressed during the planning phases.
6) During data collection activities, scientific personnel should always be present and in control of the execution of the sampling design.
7) Stakeholders should continue to be involved in the process once data has been collected to ensure that this phase of the study is transparent.

8) Scientists need to ensure that data that have been generated through effectively designed and executed survey designs are evaluated and incorporated into relevant stock assessment and management processes. Failure to do so will damage both the credibility of the scientists involved and that of the cooperative process.

5.7.6 Incorporating additional information: conclusions

Incorporation of appropriate habitat or environment covariates into survey designs or in post-survey analyses has the potential to improve survey performance. Incorporation of multiple covariates into survey designs is most appropriate for surveys with limited and specific objectives or target species. Surveys with broad objectives or multiple target species may be better served by more generalised survey designs.

Cooperative surveys involving stakeholders may be appropriate to address information needs and concerns that cannot are less effectively addressed within existing frameworks for fishery independent surveys. Caveats related to design and execution of these types of surveys (Sections 5.7.3 – 5.7.4) and guidelines (Section 5.7.5) are provided.

6 Workplans for identified areas of development

6.1 Simulated surveys

Participants agreed that a limited simulation exercise would provide a greater and shared understanding of analytical methods and an appreciation of the effects of deviations from certain assumptions of the methods. A common set of simulated fish population data will be provided to all parties for a series of comparative analyses. The data will consist of two, simulated, two-dimensional fish density fields with the following properties:

1) Field 1: Low autocorrelation: high nugget and short range;
2) Field 2: High autocorrelation: low nugget and long range;

Both fields will:

3) be of a square area grid of 120 by 120 n.mi., discretised into points representing potential trawl sampling units of 0.25 n.mi.2 (57600 points);
4) contain an unknown proportion of structural zeros, representing areas where fish do not occur beyond a certain boundary;
5) never have been sampled before (i.e., the fish population is unknown);
6) and have either a trend, representing a gradual reduction in abundance in one particular direction, or some other component of variability.

The following rules apply:

1) The fields will be generated using geostatistical techniques (Lantuéjoul 2002) by a simulator at CDG France.
2) The properties of the population (abundance and distribution) will remain unknown to all participants, until the next meeting.
3) The fields will be prepared by 31 October 2004.
4) Participants will be given the opportunity to locate samples in each field using a survey design of their choice. Participants may choose up to 3 designs (i.e., 3 surveys) for each field, but must submit their designs at the same time (i.e., designs cannot be submitted after an analysis of a previously submitted design).
5) The assumed sampling tool is a bottom trawl, delivering fish densities in number per square nautical mile.
6) Each survey must be completed in 9 whole days (216 hours).
7) Each survey must start and end at the origin (coordinates 0,0).
8) Travel speed during the survey will not exceed 10 knots at any time.
9) Each 0.25 squared n.mi. pixel takes 0.5 hours to sample. The sampling point is defined as the midpoint of any pixel(s) sampled. The cruise track proceeds from the midpoint of each sampling point, such that there is no travel through the pixel(s) being sampled, just the relevant time penalty for each sampled pixel plus the one hour trawl station time (e.g., 1.5 hours to sample 1 pixel at a station, 2 hours to sample 2 pixels at a station, and 2.5 hours to sample 3 pixels at a station). Where there is more than one pixel to be taken for a sample, the simulator will decide which pixels comprise the sample based on the sample midpoint location.
10) Any sample design and any sample size may be chosen as long as the survey is completed, and the vessel is returned to port within the 9 days.

11) The 9 days is based on a rounding up of the time taken to collect 64×1.5 hour samples in a systematic grid, sampling the midpoint of 64 evenly-spaced geographical strata, and returning home. A random sample taking 0.5 hour samples should, therefore, manage a few more samples; or a different configuration might give you fewer but longer (2 hours = 2 pixels) samples[4].
12) Submissions should be sent to the simulator by 31 October 2004 and consist of:
    13) Survey designs as sets of coordinates (x, y in n.mi.) of the midpoints of sample locations (trawl stations).
    14) For each sample, the sample size (1, 2, or more pixels).
    15) The total time (travel time + sampling time < 216 hours).
16) The simulator will deliver values of fish density corresponding to the grid coordinates of the submitted design by 30 November 2004.
17) Participants will then be at liberty to analyse their surveys and report the results at the next meeting.
18) Specific outputs required:
    19) Global abundance expressed as the total number of fish.
    20) An estimate of the precision of the abundance estimate.
    21) A map of the fish distribution.
    22) The cruise track length
    23) Some interpretation of the results.

6.2 Comparative analyses of existing datasets

Participants agreed that attempts should be made to apply more than one analytical method to datasets of their choice. The objective is to determine the global, or mean abundance, with an associated measure of precision. The methods applied should include alternative estimation concepts, ideally:

- a design-based estimator (alone and with 'collapsing the stratification' see Section 5.1)
- a model-based estimator and;
- another robust or alternative estimator.

At least one of the estimators should be suited to dealing with extreme values. The objective of this exercise is to give participants an appreciation of the different methods and to examine the effect of applying the methods to a variety of datasets. The following datasets have been proposed as candidates for the comparative analyses:

1) Canadian snow crab (*Chionoecetes opilio*, slow moving, patchy, demersal, extreme values): S. Smith and J. Choi
2) North Sea herring (fast moving, patchy, pelagic, extreme values): P. Fernandes
3) Lake Ontario alewife (*Alosa pseudoharengus,* semi-pelagic, patchy, extreme values): B. O'Gorman and J. Adams
4) Bering Sea snow crab (slow moving, patchy, demersal, extreme values): D. Somerton
5) Bering Sea walleye pollock (fast moving, large patches, pelagic, extreme values): P. Walline
6) Portuguese sardine (fast moving, patchy, pelagic, extreme values): J. Zwolinski
7) South-west England monkfish (slow moving, demersal, trend with hotspots): J. Cotter

The results will be presented and discussed at the next workshop. Participants' report should include:

- a brief description of the survey in terms of Cochran's 11 steps (Section 3);
- a post plot (bubble plot) of the survey data;
- a description of the estimators and their assumptions;
- an estimate of abundance and its precision; and
- a discussion of the results

[4] Methods to optimise travel time can be found in Harbitz and Pennington (2004). In the shrimp survey they analysed, they came to the tentative conclusion that even though more stations could be sampled using a random design than a systematic design (143 versus 118), the systematic design was better. Their conclusion was tentative because the estimate of variance for random sampling was based on geostatistics and they were not sure what effect the nugget had on the estimate.

6.3 Trawl tow length

Participants agreed that if it could be demonstrated that trawl tows of shorter length have no adverse affect on the catch rate of fish, then there may be merits in taking shorter tows to provide more time to take more trawl samples elsewhere. A review of the literature on this topic should be conducted to verify existing and past knowledge of this issue. Any new experiments should also be evaluated along with any additional supporting data. The review will be conducted by M. Pennington and presented at the following meeting. It should provide suggestions for a proposed tow length and include mitigation methods to account for the change (intercalibration). It should also include a measure of the benefits of reduced tow length in terms of the number of extra tows that could be conducted in a variety of surveys. On the basis of this, participants will decide on whether to take forward a recommendation to reduce tow length at the next meeting.

6.4 Incorporating additional information as covariates to improve precision and accuracy of surveys

Advances in technology have resulted in a number of tools which can provide additional information during fish surveys (Section 5.4). Such tools include environmental monitoring instruments (for temperature, salinity and fluorescence) and acoustic devices such as quantitative echosounders to sample the whole water column; multibeam sonars to derive bathymetric maps or whole water column detections of fish schools; and gear mensuration devices. This information may, in the ideal case, be available over the whole survey area, but should be available over a greater area than the measurement of fish density. Methods for incorporating these data as covariates to improve the estimation of fish abundance need to be developed. In other cases, different surveys of the same target population could be combined to provide an improved estimate. Participants agreed that suitable case studies should be carried out to develop and evaluate such methods. The following datasets were identified:

1) Barents Sea cod (acoustic and trawl): IMR Bergen K. Korsbrekke
2) Bering sea walleye pollock (acoustic and trawl): D. Somerton
3) South west Nova Scotian Scallop (multibeam and drag): S. Smith
4) North Sea mackerel (multibeam and echo sounder): P. Fernandes

The results will be presented and discussed at the next workshop. Participants' report should include:

- a brief description of the survey in terms of Cochran's (1977) 11 steps (Section 3);
- a post plot (bubble plot) of the survey data;
- an appropriate graphical summary of the covariate;
- a description of the estimators and their assumptions;
- an estimate of abundance and its precision, with and without the associated covariate, including the estimation of covariate relationship; and
- a discussion of the results.

6.5 Biological data

Biological data consisting of more than one variable some of which have distributions estimated from a single location (e.g., fish age, length distributions, weights etc.), present certain difficulties in the estimation process. Participants agreed that it would be useful to review methods currently in use for estimating the population level biological parameters such as proportion at age, proportion mature, growth, and their spatial distribution. A variety of different methods are currently used and associated documentation is currently difficult to obtain. There is a need to understand the sensitivity of the estimates to the methods currently used. Participants should supply a report describing:

- The objectives of collecting the data;
- how the parameters are measured and sampled;
- a description of the methods; and
- an exploration of the sensitivity of the method to its assumptions and to errors in the measurements.

Specific example datasets include:

1) North Sea herring proportion at age and proportion mature: J. Simmonds
2) Bering Sea walleye pollock proportion at age: D. Somerton
3) North Sea haddock age distribution: D. Beare
4) Lake Ontario alewife age distribution: B. O'Gorman and J. Adams

6.6 Methods of combining survey data

Conventional surveys are expensive and may be affected by external effects that may not have been anticipated in the design. There is, therefore, a constant need to explore alternative sources of data that may have a role in providing additional survey results or indicating the need for design modifications. Given the current stage of knowledge, it would be useful to draw up a list of survey methods and attempts to merge them. It is not expected as yet that alternatives to direct surveys will provide equivalent stock assessments.

The group agreed to conduct a review of methods for combining surveys of the same resource using different methods.

7 Methods to deal with intercalibration studies of fishing gears and survey vessels

7.1 Introduction

A subgroup considered intercalibration of trawl and acoustic surveys. This is the estimation of a factor that allows the catch per unit effort (trawl surveys) or the estimated biomass (acoustic surveys) found by one survey vessel and gear combination to be related to that estimated by another. Different intercalibration factors are likely to be needed for each species and, possibly, for each length or age group.

Trawl surveys provide fishery-independent indices of stock abundance, given the primary assumption that individual fish in that stock have the same probability of being caught from one survey to another. This assumption will be open to question if there is any alteration of:

- the trawl gear;
- the method of trawling;
- the geographic locations of fishing stations;
- the season or timing of fishing; and
- the survey vessel;

The first four factors are obviously important. The fifth is important because of so-called 'ship effects' on catchability. Every vessel has its own sound signature (Mitson 1995) and the effects of these vary with type of fish (demersal/pelagic), species, and depth (Godø 1994). Also, one vessel may be less powerful than another resulting in different towing speeds through the water depending on tide. Intercalibration of a trawl survey should be considered whenever any of the listed factors change. Intercalibration is also desirable in multi-vessel surveys, particularly if the vessels operate in different sub areas of the total survey area. A well-estimated factor would allow the different abundance indices obtained to be compared without worrying about possible ship effects (ICES 1998).

Acoustic surveys provide fishery-independent measures of stock abundance based on the return of acoustic energy from an echo-sounder. Each different vessel may receive a different acoustic signal strength from a given biomass of fish due to different responses of the fish to different vessel sounds. The possible responses include tilting, diving, and spreading to either side of the track of the vessel, any of which would affect the acoustic signal. Also, the sound recorded from the returned echo may depend on the vessel's own noise, and on the degree of retention of bubbles under the hull. Intercalibration factors are intended to allow for these possible differences between vessels.

The subgroup heard presentations on seven intercalibration exercises. In the light of the reported experience, options available for intercalibration are suggested and guidance offered for choosing between them. Trawl surveys are considered first, acoustic surveys second. Lastly, suggestions for good intercalibration practice are listed. Formulation of specific recommendations was not thought appropriate at this stage because of the wide variety of practical, geographic and biological factors that can bear upon decisions about intercalibration.

7.2 Intercalibration studies

7.2.1 Intercalibration of Baltic survey trawls (R. Oeberst)

An initiative was started by the ICES working group on Baltic International Trawl Surveys to coordinate the national surveys in the Baltic Sea in 1995. Four years later an EU project started to develop and introduce a new standard survey gear. In order to combine the time series based on the national gears with those based on the new standard gears, it was necessary to estimate conversion factors for converting the CPUE values of the national gears into units for the new standard gear. The paired hauls method was used for the intercalibration with the requirement that the second haul should be carried out in the track of the first haul immediately afterwards and in the same direction. Two types of experiments were carried out during these first studies. Type 1 experiments used the sequence: one haul with the old gear followed by one haul with the new gear. Type 2 experiments used the opposite sequence: one haul with the new gear followed by one haul with the old gear. The studies have shown that fish density found by the second haul is affected by catch of the first haul. This effect was noted as disturbance effect (ICES 2002c; Lewy *et al.* 2004; Oeberst *et al.* 2000). The estimation of the conversion factors based on the Type 1 and Type 2 experiments requires that the disturbance effect for the gears compared are the same (Oeberst *et al.* 2000). Because it could be expected that the disturbance effect of gears with different characteristics is different, two additional type of intercalibration experiments were defined to estimate them (ICES 2002c; Lewy *et al.* 2004). During Type 0 experiments the old gear was applied twice and during Type 3 experiments the new gear was applied twice. Using at least one of these additional types of experiments, different disturbance effects could be estimated for the gears compared.

Analyses of the German intercalibration experiments have shown that it is difficult for a small fishing vessel to fulfil the requirement that the second haul use the track line of the first haul due to the weather conditions and the limited technical equipment on board. The differences between the start positions and the end positions of the paired hauls were larger than the expected mean door spread in many cases.

Because it can be assumed that the disturbance effect decreases with increasing distance between the paired hauls with probably an S-curve structure, variations of the distance between the hauls influence the disturbance effect. Based on these results, a changed design of paired hauls was proposed for when it was not possible to use exactly the same track of the first haul during the second haul, i.e., the second haul should be carried out in the same depth layer and in the same direction of the first haul, and the distance between both the paired hauls should be at least larger than twice the expected door spread of the gears used. With this procedure it can be expected that variations of the distance between the hauls cause only small variations of the disturbance effect. The distance must be altered when a significant noise effect of the vessels can be expected that is larger than the proposed distance. These trials resulted in satisfactory estimation of conversion factors that have been used in stock assessments by the WGBFAS (ICES 2004c).

7.2.2 Intercalibration of new vessel on Icelandic groundfish survey (B. Steinarsson)

The Icelandic Groundfish Survey (IGFS) commenced in 1985 and has been conducted in March every year since then (Pálsson *et al.* 1989). About 600 towing stations are taken on the continental shelf within the 500 m depth zone. The same standardised gear and same identical commercial trawlers built in Japan in 1972 have been used over the last 20 years. The survey provides the most important datasets used for the assessment of over 15 different species in Icelandic waters and is practically the only source of data for several species. Recently, the Marine Research Institute of Iceland (MRI) acquired a new research vessel, Árni Friðriksson, and the question was raised about the feasibility of replacing one or two of the Japanese trawlers by the new vessel. Furthermore it is considered unavoidable that the old Japanese trawlers will have to be replaced by newer vessels in the near future when they will no longer be operative. From 2001 – 2004, four intercalibration experiments were conducted during the annual survey:

2001	Árni Friðriksson vs. the Japanese trawler Jón Vídalin, 101 valid paired hauls and 10 pairs for each vessel same ship same track, NW off Iceland;
2002	Árni Friðriksson vs. the Japanese trawler Páll Pálsson, 111 valid paired hauls and 10 pairs for each vessel same ship same track, NW off Iceland;
2003	Japanese trawler Páll Pálsson vs. Japanese trawler Breki, 93 valid paired hauls, NW off Iceland; and
2004	Estimation of disturbance factor of Japanese trawlers Páll Pálsson and Brettingur, 50 pairs each trawler, NW and East of Iceland.

Paired hauls were conducted where successive hauls were taken along the same track line about 1.5 hour later. A disturbance factor was estimated by letting the same the vessel carry out repeated hauls. The catch at station *st* by vessel *i* when taking the tow first was defined as:

$$C_{st,i} = \delta_{st}\alpha_i$$

and catch at station *st* by vessel *j* when vessel *i* has taken the tow earlier was defined as:

$$C_{st,j,i} = \delta_{st}\alpha_j\beta_i$$

where δ_{st} is the underlying density at station *st*, α_x is the catchability for vessel *x*, and β_i is a measure of the disturbance caused by vessel *i* by executing the tow about one and a half hours earlier.

A generalised linear model using quasi family with the log link and different variance functions was used for evaluation of the results. Three variance functions were looked at: normal, poisson with dispersion estimated, and Gamma or approximately lognormal. Large hauls control the outcome in the normal model but the small hauls can affect the outcome of the Gamma model too much so, in that case, the smallest tows need to be excluded.

A considerably higher catch rate for most species was observed for the research vessel Árni Friðriksson than the two Japanese trawlers, although the differences were more pronounced versus the Jón Vídalín than the Páll Pálsson (Table 4). In order to avoid affecting the survey results in 2001, the Árni Friðriksson always followed the Jón Vídalín. The disturbance factor, indicating that the Jón Vídalín is fishing about 10-30% less than the Árni Friðriksson is not corrected for here.

Regardless of the variance model used, the observed catch rate differences between the two Japanese trawlers, Páll Pálsson and Brettingur, seem to be insignificant (Table 5). The Poisson model including all tows can be rejected on the basis of the non-normal distribution of the residuals of the model. The other three models indicate 12-18 % higher catch rate in the first tow. Results of 2004 experiment (not reported here) indicate 10-30% higher catches in the first tow compared to the second tow.

Table 4. Preliminary results of Icelandic groundfish survey intercalibration studies conducted in 2001 and 2002, including number of stations where catch by both vessels was greater than 1 kg (n), percent of stations with the catch of one vessel greater than the other, and average proportion of catches of the two vessels, where vessels are abbreviated as ÁF = new research vessel Árni Friðriksson, JV = Japanese trawler Jón Vídalín, and PP = Japanese trawler Páll Pálsson.

	ÁF vs. JV, 2001			ÁF vs. PP, 2002		
Species	n	ÁF > JV (%)	JV/ÁF	n	ÁF > PP (%)	PP/ÁF
Cod	100	68	0.87	109	42	0.97
Haddock	78	81	0.53	99	61	0.78
Saithe	12	67	4.19	19	58	0.70
Redfish	94	76	0.55	107	69	0.76
Plaice	28	82	0.55	36	61	0.95
Catfish (*Anarhichas* spp.)	92	84	0.60	106	70	0.76
Long rough dab *Hippoglossoides platessoides*	96	80	0.60	103	67	0.80

Table 5. Preliminary results of Icelandic groundfish survey intercalibration studies conducted in 2003, comparing the estimated vessel effects (B/PP) and disturbance factors (T2/T1) for cod from Japanese trawlers Páll Pálsson (PP) and Brettingur (B) based on 50 pairs of tows for each vessel.

	Vessel effects			Disturbance factors		
Variance model (limits)	B/PP	SE	t	T2/T1	SE	t
Poisson	0.16	0.10	1.59	0.30	0.10	2.97
Poisson (< 400 kg)	0.07	0.07	0.99	0.18	0.07	2.62
Gamma (> 10 g)	0.10	0.07	1.38	0.13	0.07	1.80
Lognormal (> 0 kg)	0.09	0.08	1.13	0.12	0.08	1.48

7.2.3 Intercalibration of Survey Vessels and Gear: An Emerging Issue on the Great Lakes (R. O'Gorman and J. Adams)

In the Great Lakes, much of the information on fish stocks is collected by boats operated by state, provincial, tribal, or federal agencies (hereafter called agencies). Surveys are conducted annually, mostly with bottom trawls, but gill nets and hydroacoustics are also used. The longest running surveys started in the 1960s, but most began in the 1970s and 1980s. Boats used to conduct the surveys are at or beyond their useful working life and are being replaced with new vessels. Fishing power of the new vessels must be determined to maintain continuity of the data from bottom trawl surveys. Gill nets used on most surveys are constructed with multi-filament netting which was widely available when the surveys were instituted. Now, however, multi-filament netting is expensive and difficult to obtain, whereas monofilament netting is cheap and easy to obtain. We anticipate that many agencies will eventually switch to monofilament netting, a change that will require gear intercalibration. Moreover, we sense that in the future there will be a general movement towards standardisation of sampling gear (nets and acoustic) and coordinated multi-agency, multi-vessel surveys. The conduct of multi-agency surveys will require intercalibration of vessels and gear.

The changing invertebrate fauna of the Great Lakes has forced modifications to fishing gear that require intercalibration studies. In the 1990s, zebra and quagga mussels (*Dreissena* spp.) spread across the Great Lakes, occupying previously smooth bottom. The mussels became so numerous in southern Lake Ontario that standard survey bottom trawls could no longer be towed along the bottom without fouling. To solve this problem, bottom trawls will have to be modified to fish lighter on the bottom and intercalibration studies will have to be conducted. We suspect that agencies in other lakes will eventually encounter problems with *Dreissena* similar to those encountered in Lake Ontario.

The U.S. Geological Survey (USGS) and the New York State Department of Environmental Conservation (NYSDEC) have been conducting cooperative bottom trawl surveys with two vessels in the US waters of Lake Ontario since 1978. In the early 1980s, the NYSDEC replaced their research vessel, and an intercalibration study was conducted. The two vessels conducted side-by-side trawling (distance between the vessels while towing was no more than 300 ft) during regularly scheduled surveys from 1984 to 1989. Nets and trawl doors were identical and fishing procedures were standardised as much as equipment would allow. A total of 56 tows were successfully completed. Difference in catch rates were first evaluated by *t*-tests of the difference in log-transformed catch by species and life stage (adult or yearling). The *P* values ranged from 0.07 for yearling alewife to 0.29 for yearling lake trout. The marginally significant result for yearling alewife prompted us to conduct a simulation to evaluate the effect of applying a fishing power correction factor (FPC) to catches of yearling alewife.

Munro (1998) developed a decision rule for applying fishing power corrections to trawl survey data. The objective was to balance the trade-off between the reduction in bias and the increase in variance that comes with applying correction factors. We developed a new decision rule (to be published) based on improvements made to the one proposed by Munro (1998). The new decision rule is based on the root mean square error (RMSE) of a change in CPUE (e.g., a change in abundance from one year to the next), which can be fixed in the simulation. The RMSEs of the estimated change are estimated separately for each of three options: (1) no correction factor applied to either vessel, (2) a correction factor applied to vessel A, and (3) a correction factor applied to the vessel B. If the RMSE of the estimated change in CPUE is smaller when a correction factor is applied regardless of vessel (2 and 3) than when no correction factor is applied (1) then (and only then) we recommend applying an FPC. The RMSEs are calculated from simulated data, based on observed distributions of catch in paired trawl hauls over a wide range of fishing power differences. Using this method, we found that the FPC for yearling alewife calculated from the side-by-side trawling fell would not reduce error in tracking CPUE over time and concluded that an FPC was not needed for these data.

7.2.4 Intercalibration of trawl surveys off Alaska. (D. Somerton)

Trawl surveys in which more than one vessel participates require that the vessels be calibrated so that the CPUE from one vessel can be expressed in units that are equivalent to those of the other. An attempt was made to join the annual bottom trawl survey conducted by the Alaska Department of Fish and Game (ADFG) annually to that conducted by the National Marine Fisheries Service (NMFS). The ADFG trawl survey utilised a low net (< 2 m) and is conducted on soft bottoms, while the NMFS trawl survey utilised a high net (> 9 m) and is conducted in rock areas as well as areas with rock and cobble bottoms. Although both surveys are conducted during summer months, the ADFG survey covers the Gulf of Alaska from west to east while the NMFS survey is conducted from east to west. A calibration experiment was conducted where the two vessels completed 33 nearby parallel tows near the geographic centre of the survey area (von Szalay and Brown 2001). A calibration function was then estimated. New survey strata were defined in the areas in which both vessels fished and the biomass for the four target species was then completed. For two flatfish species the estimated biomasses were quite similar between the two vessels, however, for walleye pollock the biomass predicted from the ADFG data was more than five times larger than that for the NMFS data (von Szalay 2003). The apparent reason for this is that the vertical distribution of pollock changed over the course of the summer, which in turn, changed the availability to the two trawls. Because of this, the appropriate calibration between the NMFS and AFDG trawl changes temporally, and perhaps spatially, which negates its use to join the two surveys. The lesson from this experiment is to recognise that inter-vessel correction factors may vary in space and time.

7.2.5 Modelling results of intercalibration of new Scottish research vessel (R. Fryer)

A method was described (Fryer *et al*. 2003) for modelling the catch rates-at-length of one vessel relative to another, as used to calibrate Scotia III when it recently replaced Scotia II as the Fisheries Research Services (FRS) research vessel. The new vessel used half-hour tows instead of the 1-hour tows used by the old vessel. Smooth nonparametric curves were used to estimate the relative catch rate for each paired tow. These were then combined over all paired tows to estimate the average relative catch rate. Results were presented for haddock. The data were consistent with a relative catch rate of 0.5 (the expected value given the differences in tow duration) for most lengths. However, there was some suggestion that the new vessel using the short tows caught more large haddock than expected. The pragmatic decision was taken to use the factor of 0.5 to calibrate abundances between the two vessels because, with only 24 paired tows, the relative catch rates were not estimated sufficiently precisely to provide acceptable alternative conversion coefficients.

7.2.6 Intercalibration of North Sea IBTS (J. Cotter)

A linear modelling method (Cotter 2001) was applied to intercalibrate results for cod, haddock, whiting, and Norway pout (*Trisopterus esmarkii*) for nine national bottom trawl surveys which form part of the annual suite of surveys coordinated by ICES IBTS. No sea time was required for the intercalibration. Whole-survey population abundance indices-at-age were transformed to logarithms and modelled as functions of year-class recruits and total mortality, *Z*. An age-related factor was also included to allow for apparently lower catchabilities of young fish. The models fitted satisfactorily, permitting intercalibration factors to be estimated with standard errors. No indications of changes in Z were found over the period or over different year-classes for any of the species. Residual degrees of freedom were adjusted for correlations of abundances across ages within years. A method for comparing the relative precisions of the different surveys given the fitted model is also available.

7.2.7 Intership Comparison in Acoustic-Trawl Surveys (Author: N. Williamson; presented by D. Somerton).

Acoustic surveys of fish abundance measure a quantity known as total nautical area scattering coefficient (NASC) by integrating acoustic energy reflected by fish using an echosounder. Density is then estimated by dividing NASC by the backscatter associated with a single fish of a known length, where the length distribution within an area is estimated by sampling the fish with a trawl. Before conducting surveys, the echosounder is calibrated with an object (standard sphere) of known reflective properties. In situations in which one survey vessel must be calibrated against another, for example, when a new vessel is to replace the current survey vessel, a calibration experiment is typically conducted. This is done because, although the echosounders of each vessel are calibrated against a standard, either the fish may behave differently to the noise or other stimuli produced by each vessel or the acoustic environment such as the presence of bubbles about each vessel may be sufficiently different to cause a difference in NASC from the same density of fish. At the Alaska Fisheries Science Center, calibration experiments consist of the two vessels following nearly the same trackline with one keeping some distance to the side and slightly behind the other. Since the vessels may affect the distribution of the fish, and thereby affect the NASC, the lead position is alternated between vessels. A calibration coefficient is estimated by calculating the functional regression of the NASC values, integrated over appropriate length segments of the cruise track, on the NASCs of the other vessel. In practice, the calibration coefficients calculated for the 13 calibration experiments conducted with Canadian, Russian and Japanese vessels have not been used to correct the time series of acoustic survey data. In addition, none of the coefficients consider any uncertainty in the collection of the fish length data needed to convert NASC to biomass.

7.3 Intercalibration options for trawl surveys

The papers heard by the subgroup provide good examples of various ways to intercalibrate different surveys. This section discusses several options, their applicability, and some advantages and disadvantages in more general terms.

7.3.1 The precision of intercalibration factors

If intercalibration factors are estimated with poor precision, then it may be sensible to simply ignore the possible effects of a change of survey vessel or gear. This is because the bias induced by using a poorly estimated intercalibration factor might be greater than the true difference (bias) between the two vessels (gears). Munro (1998) describes a simulation based method for deciding when precision is too poor to risk correcting a time series of abundance indices affected by a change of fishing practice but the group had little experience of applying it (with the exception of O'Gorman and Adams, who found problems with the procedure and made improvements to it, see Section 7.2.3). A particular problem for age-based assessments is that a different factor may be needed for each age group.

Where no intercalibration has been done, or where the precision of the intercalibration factors is low, a survey with a new vessel or gear might be treated as a new CPUE series by a WG and, typically for ICES, not used until at least 5 years of data were available. At that time, estimation of the constant of proportionality, the 'catchability' q, between CPUE and stock size would provide an intercalibration factor relative to other tuning fleets.

The subgroup suggested the following things should be considered before investing in an intercalibration exercise.

- Decide if changes to the gear, vessel, or fishing technique lead to a prior expectation of changed catchabilities.
- Decide on the purpose of the intercalibration exercise. Unless large resources are available (e.g., for many tows), intercalibration exercises do not usually provide conversion factors that are sufficiently precise. Intercalibration exercises of this sort are a comfort blanket to hide behind – they don't show up anything other than gross differences between two vessels or gears – and should be regarded as such.
- Decide on the required precision of the conversion factors (and hence the required resources). This can be derived by simulating stock assessments that use the survey, and by considering the effect of: a) not adjusting the survey time series, and b) adjusting the survey time series with conversion factors estimated with particular levels of precision. The required precision will depend on the assessment method, the other indices that are used to tune the assessment, and the attitude of the stock WG to rejection of unreliable tuning series

7.3.2 Comparative fishing trials

Comparative fishing between one vessel and another may be carried out to estimate an intercalibration factor for vessel or gear effects. Ideally this will involve blocking off pairs of trawling trials so as to reduce the geographic and temporal separation between the tows of the two vessels. The purpose of this principle of experimental design is to reduce the variation of abundance that will be encountered by each of the vessels, and thus to reduce the number of trawl tows necessary to obtain an acceptably small standard error for the estimated factor (Pelletier 1998). However, the vessels should not be so close together that they could be influencing the catch of the other vessel, e.g., if the noisy vessel frightens fish into the path of the quiet vessel. Tows may be paired by trawling side by side at approximately the same time, by trawling one after the other down the same track with an interval between tows, or by trawling along parallel tows at an interval. The first method is likely to give better homogeneity of fish populations in the absence of a disturbance effect, if that may be assumed. The second and third can allow the disturbance effect to be estimated (see Section 7.2.1). The disturbance effect may be different for the different vessels due to noise, and in this case it is important to alternate or randomise the lead vessel in accordance with the principles of experimental design.

To be effective and reasonably efficient, comparative fishing trials may only be carried out where the fish species of interest are known to occur reliably in moderate or large numbers (Pelletier 1998). A paired trial resulting in a zero catch in either or both hauls provides no information about the factor and is wasted effort. Low catch numbers in either haul are not much better because then the ratio of catches on each pair of hauls by the two vessels depends on at least one low and variable number of fish giving a higher variance for the ratio than will be the case when moderate or high catch numbers are being taken in both hauls. It is very unlikely that the testing of intercalibration factors for differences with age will be possible when catch numbers are mostly low.

Parallel trawling is of course not possible when a factor is to be estimated for a change of trawl gear only, since there is only one vessel available for the trials. The same situation arises if the older vessel has been scrapped. In these cases, the gear must be changed over repeatedly during trials with the currently used vessel. For otter trawls, this is usually a time-consuming operation at sea. As a result, several hauls with each gear are likely to be used between each change-over, and comparisons between the sequences with each gear will be hindered by the additional variability of fish abundance over these larger areas and longer time periods. Changing weather and sea state may affect catching efficiencies and add to the extraneous variance of the estimated factors.

Comparative fishing trials with two vessels may be carried out either on a special paired vessel cruise, or by one vessel shadowing another at selected stations during the usual survey. For multi-vessel surveys, the most economical arrangement is likely to be for pairs of vessels to undertake parallel trawling trials at stations near the boundary between their respective sub-areas. There are several disadvantages to such comparative fishing trials:

- Organising for two, fully-staffed vessels to be in the same place at about the same time is costly and operationally difficult to achieve, particularly if the vessels come from different countries as in some multi-vessel surveys.
- There is a high risk of failure due to lack of fish or poor weather.
- Experience in the literature suggests that there is a risk of very poor precision for the estimated factors unless hundreds of parallel trawls can be achieved.

However well the factors are estimated, they will relate to the conditions of the trials (Pelletier 1998). Ship effects may vary with ground type and weather if towing the trawl at the standard speed requires the full power of the vessel. Gear of a certain design may fish differently at different depths, on different ground types, and in different weathers. Season, the presence of certain year-classes, size, and migrational factors may also be relevant (see Section 7.2.4). There is evidently a risk associated with assuming that a factor estimated in one set of conditions will be applicable to another and this risk may be greater than assuming that the factor is equal to one. Ideally, the comparative fishing trials will be broadened to include a wide range of conditions but this is likely to increase costs. Fisheries agencies in the USA are considering the need for using two vessels for the whole survey over two successive years when the vessels must be intercalibrated.

7.3.3 Modelling

Intercalibration factors can be estimated theoretically by modelling a fish population using available survey data to estimate expected catch, then by estimating a factor to align the actual catches of the two vessels (or gears) with expectations. Modelling can be done without costly comparative fishing trials at sea and is a sensible option if such trials cannot be made. However, faith must be placed in the model. Of course, modelling may also be necessary to analyse the results of intercalibration trials at sea (see Section 7.2.5).

Modelling at the catch level was reported by Cotter (1993); ICES (1992); Munro (1998); Sparholt (1990), and Pelletier (1998). A problem with this method is that many factors may serve to predict catch sizes, e.g., year, region, depth, time of day, etc. aside from the ship- and gear-related factors. A suitable model is therefore hard to identify satisfactorily. A further problem is that observed numbers of fish tend to vary greatly from catch to catch causing uncertainty about the statistical distribution. The log transformation is commonly applied.

Modelling at the level of whole-survey abundance index was described by (Cotter 2001). Since the indices are average CPUEs from the whole survey they are less variable than the individual catch data and the estimated intercalibration factors are directly applicable to the whole survey index without reservations about the special circumstances of trawling trials. Much of the variation in the indices can be explained by fitting recruitments and mortality coefficients (Z), so identification of a suitable model is easier. A change in the survey design that might cause bias is represented by fitting a constant that causes a step change in the trajectory of the decline in log numbers in each year-class (cohort). This method was used to estimate intercalibration constants for several national surveys within the North Sea IBTS covering changes of gear, vessel, and season (see Section 7.2.6).

7.4 Intercalibration options for acoustic surveys

The range of experimental designs in use for intercalibration of acoustic surveys is limited. Generally, one vessel closely follows another, but stays far enough away to avoid disturbances to the fish distribution caused by the leading ship's noise and prop wash (sailing loss). In some studies, the ships have run side by side (Arrhenius *et al.* 2000; Kieser *et al*. 1987,), and, in at least one study, the entire survey was replicated by having the two participating ships start at opposite ends of the survey area (Monstad *et al.* 1992) Usually the vessels follow in the same track (Alaska Fisheries Science Center, Torstensen and Toresen 2000; Rottingen *et al.* 1994; Wyeth *et al.* 2000), but in many cases the following vessel is to the side (Dorchenkov and Hansen 1992; Simmonds *et al.* 1998) to avoid sailing loss, as recommended in the ICES calibration manual (Foote *et al*. 1987). The vessels must trade places in order to remove these effects from the comparison of the relative acoustic performances (or alternatively, to evaluate them).

The procedures used to analyse the data collected during an intercalibration exercise are not well developed for acoustic surveys. Analysis has focused on the best way to fit the relation between acoustic backscatter pairs. MacLennan and Pope (1983) suggested using maximum likelihood methods for estimating the parameters in the functional regression between integrator values, instead of the geometric mean method that Ricker (1973) proposed and Monstad *et al.* (1992) used. Kieser *et al*. (1987) recommended a 'ratio of logs' approach. Simmonds *et al*. (1998) averaged regressions of x vs y and y vs x.

The distance over which NASC is to be averaged to build the data pairs varies in different studies. Some use complete surveys (Monstad *et al.* 1992), some use transects of various lengths (Rottingen *et al.* 1994; Kieser *et al.* 1987; Wyeth *et al.* 2000), and some use short distance units such as EDSU's (Alaska Fisheries Science Center, Arrhenius *et al.* 2000; Dorchenkov and Hansen 1992; Simmonds *et al.* 1998).

During acoustic intercalibration exercises between the US Research Vessel Miller Freeman and vessels from Japan, Russia, and Korea in the Bering Sea, replicate passes are made over aggregations, with mean NASC as a data point. Values obtained by each ship in the exercise are reported to the other ship by radio upon the conclusion of each pass, to allow modification of experimental design in real time if gross differences are detected. The data are analysed by fitting a functional regression to the set of data pairs. Complete data sets are exchanged shortly after the surveys are completed, and detailed comparison of high resolution NASC data is conducted. By plotting one vessel's data against the other's, it is possible to determine whether the difference observed is systematic or the result of chance due to patchy fish distribution.

In 2005 the NOAA ship Oscar Dyson, which meets the ICES noise standard, will join the Alaska Fisheries Science Centre (AFSC) "acoustic fleet" along with RV Miller Freeman. Over a period that will last more than a year, the acoustic performance of these ships will be compared. The data will be collected in the same way as was done in past intercalibration exercises. Data with resolution down to the level of individual pings will be available from both ships. The procedure that will be used to analyse the data sets has not yet been specified, but is likely to involve comparisons on short length scales. Using data averages from smaller sampling units than passes (short transects) might enhance ability to detect differences between ships. This might be more difficult in areas with fish distributions characterised by the presence of schools small in relation to the distance between ships, but since the relative performance of the two ships may depend on the characteristics of the fish distribution, comparisons need to be made in all types of fish distribution, and in as many other different circumstances (season, fish densities, length composition, water depths, etc.) as possible.

One factor limiting the ability of an intercalibration exercise to detect differences in acoustic system performance is the precision attainable with current sphere calibration procedures. This precision is thought to be about 0.2 dB. Thus, differences must be greater than this to be detectable.

7.5 Additional advice on intercalibration

Proposals for intercalibration trials should preferably be discussed with ICES colleagues outside the marine laboratory directly involved. It is likely that they will have additional experience that will help reduce the risk of obtaining poor estimates of the factors due to overlooking an important point in the design of the trials.

The expense and staffing difficulties caused by the need for intercalibration trials at sea imply that precautions should be taken to minimise the need for them. Good maintenance of the vessel to permit long life could be a good investment, as well as choice of standard designs of hull, propeller, and engine that are likely to be replaceable with minimal changes in relation to their possible effects on underwater noise and fish. Trawl gear also should be carefully maintained (and mended).

On multi-vessel surveys, e.g., IBTS, allocation of the same vessel to the same geographic region so far as possible assists with standardisation of results from one year to the next (ICES 1998). Paired trawling should preferably be included every year at a few locations near where the regions trawled by two vessels join. This will build up a series of parallel tow results that can be used to adjust results when one of the vessels must be replaced. Such parallel trawling should observe principles of experimental design such as alternation of the lead vessel. Incorporating paired trawling during routine surveys would also help to prevent drift in survey catchabilities over time due to any gradual modernisation of gear and procedures, e.g., new twine, better echosounder, etc.

Procedures for catch handling and subsampling for biological measures should be identical among tows and vessels during intercalibration trials. Vessel crew and biological staff should be given written protocols. Furthermore, any aspect of the study that might have a chance of influencing catch rates should be standardised. Detailed written records of each trial should be made for each trawl haul.

An important dimension to the fishing power of a vessel is the captain. Two captains fishing at the same coordinates with the same vessel and gear may achieve different average catch rates due to different approaches to the tide, weather, and different speeds of shooting and hauling, and different responses to variations of gear geometry. Some of this variability might be standardisable with well-written protocols.

Intercalibration trials that do not succeed in producing a precisely estimated intercalibration factor may, nevertheless, be of value to a stock assessment if the factor is included with a Bayesian prior to allow for the uncertainty.

The sub-group did not have a member with experience of intercalibrating acoustic surveys. Nevertheless, it was noted that trawl hauls used intermittently to assess species and size composition of the fish registered by the echosounder could be affected by a change in vessel noise or gear. These factors may also need intercalibration in addition to acoustic back scatter when two acoustic survey vessels are being compared.

8 Recommendations

The Workshop on Survey Design and Analysis recommends that:

The **Workshop on Survey Design and Analysis** should meet in Seté, France, on 9–13 May 2005 under the co-chairmanship of P. Fernandes (U.K., Scotland) and M. Pennington (Norway), to:

a) Evaluate analyses of estimates of the abundance, associated variance, and density maps, from surveys of a simulated fish population whose abundance is known.
b) Evaluate alternative analyses of several survey datasets.
c) Review the state of knowledge regarding the effect of trawl duration on fish catch rate with a view to considering a reduction in sample trawl duration.
d) Evaluate analyses of covariate data which could provide improved precision of abundance estimates.
e) Review methods for combining surveys of the same resource using different methods.
f) Evaluate the sensitivity of methods to estimate biological parameters in terms of analytical assumptions and measurement error.

Additionally the group recommends:

1) ***Inclusion of systematic sampling (with stratification) or stratified random sampling should be considered in the designing of a fish survey.*** In the presence of positive local autocorrelation, a more precise estimate of the population mean will usually be obtained by systematic sampling or stratified random sampling than by simple random sampling. The optimal sampling design will depend on the population under study and the relative importance attached to getting the most precise estimate of the population mean and to getting a good estimate of that precision. A wide range of real and simulated examples suggest that systematic sampling will often be optimal if getting the most precise estimate of the sample mean is the dominant objective. However, stratified random sampling will often be preferable if getting a good estimate of the precision is also important.
2) ***Information from the commercial fishing industry should be considered***, where appropriate, to provide guidance on survey design (e.g., in the definition of strata).
3) ***Efforts should be made to maximise the number of samples taken***, if survey precision needs to be enhanced. This may be achieved by shortening towing times or by using instruments in as efficient a manner as possible. Consideration should be given to the effect of shortened tow times to establish if this is a practical and effective course of action.
4) ***Information additional to that of fish density should be collected*** on surveys, particularly when that information is related (covariate) and can be collected more extensively. Incorporation of appropriate covariates (habitat, environment) can lead to improved precision of the abundance estimate, provided that a good relationship exists, and that the covariate is known at more sample locations than the fish density. Ideally, the covariate should be known at all locations where the fish density is interpolated to (i.e., the whole survey area).

5) ***Means to provide direct estimates of abundance from surveys should be investigated***. Calibrating a survey time series using historical catch data may generate more robust abundance estimates (in recent time periods) than a catch-at-age analysis due to problems associated with the accuracy of catch data.
6) ***All publicly funded surveys should include a description of their estimation procedures*** in their reports, particularly those benefiting from EC funding and those carried out under the auspices of ICES. Survey reporting practises vary considerably and, in some cases, the methods used to estimate abundance are not described.
7) ***The design effect and the effective sample size should be reported*** whenever possible to give a measure of the efficiency of a survey design, and the sampling unit over which the data were gathered (the 'support') should be explicitly stated. The design effect is a measure of the efficiency of a survey. It is calculated as the ratio of the variance of the estimated mean for the actual design (and variance estimator employed) and the expected variance obtained under simple random sampling. The effective sample size is the number of samples selected by simple random sampling that would be required to achieve the same precision obtained with *n* samples under the actual complex sampling design.
8) ***Survey precision should be reported as the relative standard error*** (100% × standard error / estimate). The term coefficient of variation (CV) is ambiguous and should be avoided.

9 References

Aglen, A. (1989). Empirical results on precision - effort relationships for acoustic surveys. *ICES CM 1989/**B**:30.* 28 pp.

Armstrong, M., Peel, J., McAliskey, M., McCurdy, W., McCorriston, P., and Briggs, R. (2003). Survey indices of abundance for cod, haddock and whiting in the Irish Sea (Area VIIaN): 1992 – 2003. *ICES CM 2003/ACFM:01*; WD: 3.

Arrhenius, F., Bethke, E., Cardinale, M., and Hakansson, N. (2000). Intercalibration of nautical area scattering coefficients between research vessels. *Arch. Fish. Mar. Res.* 48: 31-42.

Augustin, N.H., Borchers, D.L., Clarke, E.D., Buckland, S.T., and Walsh, M. (1998). Spatiotemporal modelling for the annual egg production methods of stock assessment using generalised additive models. *Canadian Journal of Fisheries and Aquatic Sciences.* **55**: 2608-2621.

Azarovitz, T.R. (1981). A brief historical review of the Woods Hole laboratory trawl survey time series. *Bottom trawl surveys*, Canadian Special Publication of Fisheries and Aquatic Science. **58:** 62-81.

Bailey, M.C., Maravelias, C.D., and Simmonds, E.J. (1998). Changes in the distribution of autumn spawning herring (*Cupea harengus* L.) derived from annual acoustic surveys during the period 1984-1996. *ICES Journal of Marine Science.* **55**: 545-555.

Beare, D., Castro, J., Cotter, J., van Keeken, O., Kell, L., Laurec, A., Mahé, J-C, Moura, O., Munch-Petersen, S., Nielsen, J. R., Piet, G., Simmonds, J., Skagen, D., and Sparre, P. J. (2003). Evaluation of research surveys in relation to management advice (EVARES - FISH/2001/02 - Lot 1) Final Report to European Commission Director-General Fisheries.

Bergstedt, R. A., and Genovese, J. H. (1994). New technique for sampling sea lamprey larvae in deepwater habitats. *North American Journal of Fisheries Management* **14**: 449-52.

Bez, N. (2002). Global fish abundance estimation from regular sampling: the geostatistical transitive method. *Canadian Journal of Fisheries and Aquatic Sciences* **59**: 1921-1931.

Brierley, A., Gull, S.F., and Wafy, M.H. (2003). A Bayesian maximum entropy reconstruction of stock distribution and inference of stock density from line-transect acoustic survey data. *ICES Journal of Marine Science.* **60**: 446-452.

Brogan, D. (1998). Software for sample survey data, misuse of standard packages. In Armitage, P., and Colton, T.(Eds) *Encyclopedia of Biostatistics*, Volume 5. John Wiley and Sons, New York, NY pp. 4167-4174.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., and Thomas, L. (2001). *Introduction to Distance Sampling.* Oxford University Press, Oxford.

Caers, J., Vynckier, P., Beirlant, J., and Rombouts, L. (1996). Extreme Value Analysis of Diamond-Size Distributions. *Mathematical Geology* **28**(1): 25-43.

Callaway, R., Alvsvåg, J., de Boois, I., Cotter, J., Ford, A., Hinz, H., Jennings, S., Kröncke, I., Lancaster, J., Piet, G., Prince, P., and Ehrich, S. (2002). Diversity and community structure of epibenthic invertebrates and fish in the North Sea. ICES J. mar. Sci., 59, 1199-1214.

Cardador, F., Sanchéz, F., Pereiro, F.J., Borges, M.F., Caramelo, A.M., Azevedo, M., A., S., Pérez, N., Martins, M.M., Olaso, I., Pestana, G., Trujillo, V., and Fernandez, A. (1997). Groundfish surveys in the Atlantic Iberian waters (ICES Divisions VIIIc and IXa): history and perspectives. *ICES CM 1997/**Y**:08.* 30 pp.

Carlson, B.D. (1998). Software for Statistical Analysis of Sample Survey Data. In Armitage, P., and Colton, T.(Eds) *Encyclopedia of Biostatistics*, Volume 5. John Wiley and Sons, New York, NY: pp. 4160-4167.

Carlsson, D, Folmer, O., Kanneworff, P., Kingsley, M., and Pennington, M. (2000). Improving the West Greenland Trawl Survey for shrimp (Pandalus borealis). *J. Northw. Atl. Fish. Sci*., **27**: 151-160.

Chen, J., Thompson, M.E., and Wu, C. (2004). Estimation of fish abundance indices based on scientific research trawl surveys. *Biometrics* **60**: 116–123.

Chen, Y., and Jackson, D.A. (1995). Robust estimation of mean and variance in Fisheries. *Transactions of the American Fisheries Society* **124**: 401-412.

Chiles, J.-P., and Delfiner, P. (1999). *Geostatistics: modelling spatial uncertainty*. Wiley, New York, 695 pp.

Clarke, S.H. (1981). Use of trawl survey data in assessments. *Bottom Trawl Surveys*, Canadian Special Publication of Fisheries and Aquatic Science. **58:** 82-92.

Cochran, W.G. (1977). Sampling Techniques (third edition). John Wiley and Sons, New York, 428 pp.

Cotter, A. J. R. (1993). Intercalibration of groundfish surveys using regression analysis of yearclass mortalities., International Council for the Exploration of the Sea, Copenhagen

Cotter, A. J. R. (2001). Intercalibration of North Sea International Bottom Trawl Surveys by fitting year-class curves. *ICES Journal of Marine Science* **58**: 622-632 [Erratum, Ibid. 658:1340]

Dalen, J., Hylen, A., Jakobsen, T., Nakken, O., Randa, K., and Smedstad, O. (1982). Norwegian investigations on young cod and haddock in the Barents Sea during the winter 1982. *ICES CM 1982/**G:** 41*, 20 pp.

Demer, D. A. (2004). "An estimate of error for the CCAMLR 2000 estimate of krill biomass." CCAMLR 2000 Special Issue: XX.

Deriso, R. B. (1980). Harvesting strategies and parameter estimation for an age-structured model. Canadian Journal of Fisheries and Aquatic Sciences. **37**: 268–282.

Despres-Patanjo, L.I., Azarovitz, T.R., and Byrne, C. J. (1988). Twenty-five years of fish surveys in the northwest Atlantic: The NMFS Northeast Fisheries Center's Bottom Trawl Survey Program. *Marine Fisheries Review* **50**(4):69-71.

Dickie, L.M. (1981). Historical review and impact of surveys on management advice: Chairman's remarks. *Bottom Trawl Surveys*, Canadian Special Publication of Fisheries and Aquatic Science. **58:** 8-10.

Dorchenkov, A., and Hansen, K.A. (1992). Intercalibration of acoustic systems onboard R/V Johan Hjort and R/V Pinro. ICES Council Meeting Papers, ICES, Copenhagen. 10 p.

Doubleday, W.G., and Rivard, D. (1981). Bottom trawl surveys. *Canadian Special Publications on Fisheries and Aquatic Science* **58**.

FAO (Food and Agriculture Organization) (1995). Precautionary approach to fisheries. Part 1: Guidelines on the precautionary approach to capture fisheries and species introductions. *FAO Fisheries Technical Paper* **350**(1): 52.

Fernandes, P.G., Gerlotto, F., Holliday, D.V., Nakken, O., and Simmonds, E.J. (2002). Acoustic applications in fisheries science: the ICES contribution. *ICES Marine Science Symposia* **215**: 483-492.

Folmer, O., and Pennington, M. (2000). A statistical evaluation of the design and precision of the shrimp trawl survey off West Greenland. *Fisheries Research* **49**:165-178.

Foote, K. G., Knudsen, H. P., Vestnes, G., MacLennan, D. N., and Simmonds, E. J. (1987). Calibration of acoustic instruments for fish density estimation: a practical guide. ICES Coop. Res. Rep. 144. 57p.

Foote, K.G. (1993). Abundance estimation of herring hibernating in a fjord. *ICES CM 1993/**D:**45.*

Forest, A., and Minet, J.P. (1981). Abundance estimates of the trawlable resources around the islands of Saint-Pierre and Miquelon (NAFO Subdiv. 3Ps): methods used during the French research surveys and discussion of some results. *Bottom trawl surveys*, Canadian Special Publication of Fisheries and Aquatic Science. **58:** 68-81.

Francis, R.I.C.C. (1984). An adaptive strategy for stratified random trawl surveys. *New Zealand Journal of Marine and Freshwater Research* **18**: 59-71.

Francis, R.I.C.C. (1991). Statistical properties of two-phase surveys: comment. *Can. J. Fish. Aquat. Sci.* **48**: 1128.

Fryer, R. J., Zuur, A. F., Graham, N. (2003). Using mixed models to combine smooth size-selection and catch-comparison curves over hauls. *Canadian Journal of Fisheries and Aquatic Sciences* **60**: 448-459.

Gavaris, S., and Smith, S. J. (1987). Effect of allocation and stratification strategies on precision of survey abundance estimates for Atlantic cod (*Gadus morhua*) on the Eastern Scotian Shelf. *Journal of Northwest Atlantic Fisheries Science* **8**: 137–145.

Gimona, A., and Fernandes, P.G. (2003). A conditional simulation of acoustic survey data. *Journal of Aquatic Living Resources* **16**(3): 123-129.

Gjøsæter, H., Bogstad, B., and Tjelmeland, S. (2002). Assessment methodology for Barents Sea capelin, *Mallotus villosus* (Müller). *ICES Journal of Marine Sicence*, **59**:1086-1095.

Goddard, P.D. (1997). The effects of tow duration and subsampling on CPUE, species composition and length distributions of bottom trawl survey catches. M.S. Thesis. University of Washington, Seattle, Washington, 119 pp.

Godø, O. R. (1994). Factors affecting the reliability of groundfish abundance estimates from bottom trawl surveys. In Ferno, A., and Olsen, S. (eds) *Marine fish behaviour in capture and abundance estimation*. pp. 166-199. Farnham, Surry, England: Fishing News Books

Godø, O. R., Pennington, M., and Vølstad, J.H. (1990). Effect of tow duration on length composition of trawl catches. *Fisheries Research* **9**:165-179.

Gohin, F. (1985). Planification des expériences et interprétation par la théorie des variables régionalisées: application à l'estimation de la biomasse d'une plage. *ICES CM 1985/**D:**3.* 11 pp.

Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation. Oxford University Press. 483 pp.

Gunderson, D.R. (1993). *Surveys of fisheries resources*. Wiley, New York.
Halliday, R.G., and Koeller, P.A. (1981). A history of Canadian groundfish trawling surveys and data usage in ICNAF divisions 4TVWX. *Bottom trawl surveys*, Canadian Special Publication of Fisheries and Aquatic Science. **58:** 27-41.
Hanselman, D.H., Quinn, T.J., Lunsford, C., Heifetz, J., and Clausen, D. (2003). Applications in adaptive cluster sampling of Gulf of Alsaska rockfish. *Fishery Bulletin* **101**: 501-513.
Harbitz, A., and Pennington, M. (2004). Comparison of shortest sailing distance though random and regular sampling points. *ICES Journal of Marine Science* **61**:140-147.
Heessen, H.J.L., Dalskov, J., and Cook, R.M. (1997). The International Bottom Trawl Survey in the North Sea, Skagerrak and Kattegat. *ICES CM 1997/**Y:**31.* pp.
Hewitt, R.P., Watkins, J.L., Naganobu, M., Tshernyshkov, P., Brierley, A.S., Demer, D.A., Kasatkina, S., Takao, Y., Goss, C., Malyshko, A., Brandon, M.A., Kawaguchi, S., Siegel, V., Trathan, P.N., Emery, J.H., Everson, I., and Miller, D.G.M. (2002) Setting a precautionary catch limit for Antarctic krill. *Oceanography* **15**:26-33.
Hilborn, R., and Walters, C.J. (1992). *Quantitative fisheries stock assessment: choice, dynamics and uncertainty*. Chapman and Hall, New York, 570 pp.
Hughes, S.E. (1976). System for sampling large trawl catches of research vessels. *Journal of the Fisheries Research Board of Canada* **33**: 833-839.
Hutchings, J.A., and Myers, R.A. (1994). What can be learned from the collapse of a renewable resource ? Atlantic Cod, *Gadus morhua*, of Newfoundland and Labrador. *Canadian Journal of Fisheries and Aquatic Sciences* **51**(9): 2126-2146.
Hylen, A., Nakken, O., and Sunnanå, K. (1986). The use of acoustic and bottom trawl surveys in the assessment of North-East Arctic cod and haddock stocks. In: M. Alton (ed.), A Workshop on Comparative Biology, Assessment and Management of Gadoids from the North Pacific and Atlantic Oceans. Northwest and Alaska Fisheries Center, Seattle, WA, pp. 473-498.
ICES (1989). Report of the workshop on spatial statistical techniques. *ICES CM 1989/**K:**38.* 42 pp.
ICES (1990a). Report of the working group on methods of fish stock assessment. *ICES CM 1990/**Assess:**15.* pp.
ICES (1990b). Report of the Study Group on the Applicability of Spatial Statistical Techniques to Acoustic Survey Data. *ICES CM 1990/**D:**34.* 103 pp.
ICES (1992). Report of the workshop on the analysis of trawl survey data. *ICES CM 1992/**D:**6.* 96 pp.
ICES (1993). Report of the workshop on the Applicability of Spatial Statistical Techniques to Acoustic Survey Data. *ICES Cooperative Research Report* **195**: 87.
ICES (1998). Report of the study group on the evaluation of the quarterly IBTS surveys. *ICES CM 1998/**D**:4.*
ICES (1999). Manual for the International Bottom Trawl Surveys, Revision VI. *ICES CM 1999/D*:2; Addendum 2: 1-49.
ICES (2000). Report Of The Study Group On Life Histories Of Nephrops. *ICES CM 2000/**G:**06.* pp.
ICES (2002a). Report of the Working Group on Mackerel and Horse Mackerel Egg Surveys. *ICES CM 2002/**G:**06.* 106 pp.
ICES (2002b). Manual for The International Bottom Trawl Surveys in the Western And Southern Areas. Revision II. Addendum to ICES CM 2002/D:03. Dublin, Ireland 8–11 April 2002.
ICES (2002c). Report of the Baltic International Fish Survey Working Group. ICES CM/G:05
ICES (2003a). Report of the International Bottom Trawl Survey Working Group. *ICES CM 2003/**D:**5.* 75 pp.
ICES (2003b). Report of the planning group for herring surveys. *ICES CM 2003/**G:**03.* 180 pp.
ICES (2003c). Report of the planning group on surveys on pelagic fish in the Norwegian Sea. *ICES CM 2003/**D:**10.* 60 pp.
ICES (2003d). Report of the Working Group on Beam Trawl Surveys. *ICES CM 2003/**G:**14.* 49 pp.
ICES (2003e). Study Group on the Review of the Structure of the Fisheries Technology Committee. *ICES CM 2003/**B:**08.* 34 pp.
ICES (2004a). Report of the planning group for herring surveys. *ICES CM 2004/**G:**05.* 191 pp.
ICES (2004b). Report of the International Bottom Trawl Survey Working Group, *ICES CM 2004/**D**:05.*
ICES (2004c). Report of the Baltic Fisheries Assessment working group. ACFM:22
ICES (2004d). Report of the Baltic International Fish Survey Working Group, *ICES CM 2004/**G**:08,* 162 pp.
ICES (2004e). Report of the Herring Assessment Working Group for the Area South of 62ºN (HAWG). *ICES CM 2004/ACFM:18,* 538 pp.
IPROSTS - International Program of Standardised Trawl Surveys. (2001). Final report to the Commission of European Communities. DG XIV Study contract reference: 98-057.
Jakobsen, T., Korsbrekke, K., Mehl, S., and Nakken, O. (1997). Norwegian combined acoustic and bottom trawl surveys for demersal fish in the Barents Sea during winter. *ICES CM 1997/**Y**:17*, 26 pp.
Jakobssen, J. (1983). Echo surveying of the Icelandic summer spawning herring 1973-1982. *FAO Fish Report* **300**: 240-248.
Jessen, R.J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa. Agr. Exp. Sta. Res. Bull.* **304**.
Jessen, R.J. (1978). *Statistical Survey Techniques.* John Wiley and Sons, New York, NY.

Jolly, G.M., and Smith, S.J. (1989). A note on the analysis of marine survey data. *Proceedings of the Institute of Acoustics* **11**(3): 195-201.
Jolly, G.M., and Hampton, I. (1990). A stratified random transect design for acoustic surveys of fish stocks. *Canadian Journal of Fisheries and Aquatic Sciences*. **47**: 1282-1291.
Kanneworff, P., and Wieland, K. (2003). Stratified-random trawl survey for northern shrimp (Pandalus borealis) in NAFO Subareas 0+1 in 2003. NAFO SCR Doc., No. 71, Serial No. N4910, 26 p.
Kappenman, R.F. (1999). Trawl survey based abundance estimation using data sets with unusually large catches. *ICES Journal of Marine Science.* **56**: 28-35.
Kieser, R., Mulligan, T. J., Williamson, N. J., and Nelson, M. O. (1987). Intercalibration of two echo integration systems based on acoustic backscattering measurements. Can. J. Fish. Aquat. Sci. 44: 562-572.
Kimura, D.K., and Balsinger, J.W. (1985). Bootstrap methods for evaluating sablefish pot index surveys. *North American Journal of Fisheries Management* **5**: 47-56.
Kingsley, M.C.S., Carlsson, D. M., Kanneworff P., and Pennington, M. (2002). Spatial structure of the resource of *Pandalus borealis* and some implications for trawl survey design. *Fisheries Research* **58**:171-183.
Kingsley, M.C.S., Kanneworff, P., and Carlsson, D. M. (2004). Buffered random sampling: a sequential inhibited spatial point process applied to sampling in a trawl survey for northern shrimp *Pandalus borealis* in West Greenland waters. *ICES Journal of Marine Science* **61**:12-24.
Kish, L. (1965). *Survey Sampling*. John Wiley and Sons, New York, NY. available as 1995 reprint.
Kish, L. (1995). Methods for design effects. *Journal of Official Statistics* **11**:55-77.
Kish, L. (2003). *Selected Papers*. Edited by G. Kalton, and S. Heeringa. John Wileys and Sons. Wileys Series in Survey Methodology.
Korn, E.L., and Graubard, B.I. (1999). *Analysis of Health Surveys*. John Wiley and Sons, New York, NY.
Korsbrekke, K., Mehl, S., Nakken, O., Pennington, M. (2001). A survey-based assessment of the Northeast Arctic cod stock. *ICES Journal of Marine Science* **58**:763-769.
Kostylev, V. E., Courtney, R. C., Robert, G., and Todd, B. J. (2003). Stock evaluation of giant scallop (*Placopecten magellanicus*) using high-resolution acoustics for seabed mapping. *Fisheries Research* **60**:479–492.
Krebs, C.J. (1989). Ecological methodology. Harper and Row, New York. 654 pp.
Lantuéjoul C. (2002). *Geostatistical simulation. Models and Algorithms*. Springer, Berlin. 256 pp.
Legendre, P., and M.-J. Fortin (1989). Spatial pattern and ecological analysis. *Vegetatio* **80**:107-38.
Lenarz, W.H., and Adams, P.B. (1980). Some statistical considerations for the design of trawl surveys for rockfish (Scorpaenidae). *Fishery Bulletin* **78**: 659-674.
Lewy, P., Nielsen, J. R., and Hovgard, H. (2004). Survey gear calibration independent of spatial fish distribution. Canadian Journal of Fisheries and Aquatic Sciences 61: 636-647
MacLennan, D.N., and Pope, J.A. (1983). Analysis procedure for the inter-ship calibration of echo integrators. ICES CM 1983/B:22.
MacLennan, D.N., and MacKenzie, I.G. (1988). Precision of acoustic fish stock estimates. *Canadian Journal of Fisheries and Aquatic Sciences* **45**(4): 605-616.
MacLennan, D.N., and Simmonds, E.J. (1992). *Fisheries acoustics*. Chapman and Hall, London, 325 pp.
Matheron, G. (1971). The theory of regionalised variables and its application. *Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau*, Fasc. 5, Ecole Nat. Sup. des Mines de Paris.
Matheron, G. (1989). *Estimating and choosing*. Springer-Verlag, Berlin, 141 pp.
McConnaughey, R.A., and Conquest, L.L. (1992). Trawl survey estimation using a comparative approach based on lognormal theory. *Fishery Bulletin* **91**: 107-118.
McQuinn, I. (1997). Metapopulations and the Atlantic herring. *Reviews in Fish Biology and Fisheries* **7**: 297-329.
Mitson, R. B. (1995). Underwater noise of research vessels - review and recommendations. *ICES Cooperative Research Report* **209.**
Mohn, R. (1999). The Retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. *ICES Journal of Marine Science* **56**:473-488.
Mohn, R.K., Robert, G., and Roddick, D.L. (1987). Research sampling and survey design of Georges Bank scallops (*Placopecten magellanicus*). *Journal of Northwest Atlantic Fisheries Science* **8**: 117-122.
Monstad, T., Borkin, I., and Ermolchev, V. (1992). Report of the joint Norwegian-Russian acoustic survey on blue whiting, spring 1992. ICES C.M. 1992/H:6.
Mullett, K. M., Heinrich, J. W., Adams, J. V., Young, R. J., Henson, M. P., McDonald, R. B., and Fodale, M. F. (2003). Estimating lake-wide abundance of spawning-phase sea lampreys (*Petromyzon marinus*) in the Great Lakes: extrapolating from sampled streams using regression models. *Journal of Great Lakes Research* **29**: 240-252.
Munro, P. T. (1998). A decision rule based on the mean square error for correcting relative fishing power differences in trawl survey data. *Fishery Bulletin* **96**: 538-546.
Myers, R.A., and Pepin, P. (1990). The robustness of log-normal based estimators of abundance. *Biometrics* **46**: 1185-1192.
Nakken, O. (1998). Past, present and future exploitation and management of marine resources in the Barents Sea and adjacent areas. *Fisheries Research* **37**:23-35.
NRC (National Research Council) (1998). *Improving fish stock assessments*. National Academic Press, Washington D.C., 188 pp.

O'Gorman, R., and Schneider, C. P. (1986). Dynamics of alewives in Lake Ontario following a mass mortality. *Transactions of the American Fisheries Society* **115**:1-14.

Oeberst, R., Ernst, P., and Friess, C. C. (2000). Intercalibrations between German demersal gears HG 20/25 and TV3 520 as well as between the gears TV3 520 and TV3 930. ICES C.M. 2000/K:20

Ohman, M.D., and Venrick, E.L. (2003). CalCOFI in a Changing Ocean. *Oceanography* **16**: 76-85.

Ouellet, P., Grégoire, F., Harvey, M., Head, E., Morin, B., Robert, G., Savard, L., Smith, S. J., and Starr, M. (2003). Exceptional environmental conditions in 1999 in eastern Canadian waters and the possible consequences for some fish and invertebrate stocks. AZMP Bulletin. **3**: 21–27.

Owens, R. W., O'Gorman, R., Eckert, T. H. and Lantry, B. F. (2003). The offshore fish community in southern Lake Ontario, 1972-1998. Pages 407-442, in M. Munawar [ed.], The State of Lake Ontario: Past, Present, and Future. Ecovision World Monograph Series, Aquatic Ecosystem Health and Management Society, Burlington, Ontario, Canada.

Ouellet, P., Grégoire, F., Harvey, M., Head, E., Morin, B., Robert, G., Savard, L., Smith, S.J., and Starr, M. (2003). Exceptional environmental conditions in 1999 in eastern Canadian waters and the possible consequences for some fish and invertebrate stocks. AZMP Bulletin. 3: 21–27.

Pagano, M., and Gauvreau, K. (1993). *Principles of Biostatistics*. Duxbury.

Pálsson, O. K., Jonsson, E., Schopka, S. A., Stefansson, G., and Steinarsson, B. (1989). Icelandic groundfish survey data used to improve precision in stock assessments. *Journal of Northwest Atlantic Fisheries Science* **9**: 53-72.

Parma, A.M. (1993). Retrospective catch-at-age analysis of Pacific halibut: implications on assessments of harvesting policies. In G. Kruse, D.M. Eggers, R.J. Marasco, C. Pautzke and T.J. Quinn II (Eds.). Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations, Alaska Sea Grant College Program Report No. 93-02, Fairbanks, University of Alaska.

Patterson, K. (1998). Assessing fish stocks when catches are misreported: model simulation tests and application to cod, haddock, and whiting in the ICES area. *ICES Journal of Marine Science.* **55**: 878-891.

Patterson, K.R., and Melvin, G.D. (1996). Integrated Catch-at-age Analysis Version 1.2. *Scottish Fisheries Research Report*, **58**.

Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., Punt, A., Restrepo, V., Skagen, D.W., and Stefansson, G. (2001). Estimating uncertainty in fish stock assessment and forecasting. *Fish and Fisheries* **2**: 125-157.

Pauly, D., V., C., Guénette, S., Pitcher, T., Sumaila, R., Walters, C.J., Watson, R., and Zeller, D. (2002). Towards sustainability in world fisheries. *Nature* **418**: 689-695.

Pelletier, D. (1998). Intercalibration of research survey vessels in fisheries: a review and an application. *Canadian Journal of Fisheries and Aquatic Sciences.* **55**: 2672-2690.

Pennington, M. (1983). Efficient estimators for fish and plankton surveys. *Biometrics* **39**: 281-286.

Pennington, M. (1985). Estimating the relative abundance of fish from a series of trawl surveys. *Biometrics* **41**:197-202.

Pennington, M. (1991). Comments on the robustness of lognormal-based abundance estimators. *Biometrics* **47**:1623

Pennington, M. (1996). Estimating the mean and variance from highly skewed marine survey data. *Fishery Bulletin* **94**: 498-505.

Pennington, M., and Volstad, J.H. (1991). Optimum size of sampling unit for estimating the density of marine populations. *Biometrics* **47**: 717-723.

Pennington, M., and Vølstad, J. H. (1994). Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. *Biometrics* **50**:725-732

Pennington, M., and Godø, O. R. (1995). Measuring the effect of changes in catchability on the variance of marine survey abundance indices. *Fisheries Research* **23**:301-310.

Pennington, M., and Strømme, T. (1998). Surveys as a research tool for managing dynamic stocks. *Fisheries Research* **37**:97-106.

Pennington, M., Burmeister L.M., and Hjellvik, V. (2002). Assessing the precision of frequency distributions estimated from trawl-survey samples. *Fishery Bulletin* **100**:74-81.

Perry, R.I., and Smith, S.J. (1994). Identifying habitat associations of marine fishes using survey data: an application to the NW Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences* **51**: 589–602.

Petitgas, P. (1999). A review of linear geostatistics for fisheries survey design and stock assessment. *Quantitative geology and geostatistics* **10**: 1-12.

Power, J.H., and Moser, E.B. (1999). Linear model analysis of net catch data using the negative binomial distribution. *Canadian Journal of Fisheries and Aquatic Sciences.* **56**: 191-200.

Rao, J.N.K. (2003). *Small Area Estimation*. John Wiley and Sons, New York, NY.

Ricker, W.E. (1973). Linear regressions in fishery research.. Journal of the Fisheries Research Board of Canada, **30**: 409-434.

Ripley, B.D. (1981). Spatial Statistics. John Wiley and Sons, Inc

Rivoirard, J., Simmonds, E.J., Foote, K.F., Fernandes, P.G., and Bez, N. (2000). *Geostatistics for estimating fish abundance*. Blackwell Science Ltd., Oxford, 206 pp.

Robotham, H., and Castillo, J. (1990). The bootstrap method: An alternative for estimating confidence intervals of resources surveyed by hydroacoustic techniques. *Rapp. P.-v. Reun. Cons. perm. int Explor. Mer.* **189**: 421-424.

Rose, G., Gauthier, S., and Lawson, G. (2000). Acoustic surveys in the full Monte:simulating uncertainty. *Aquat. Living Resour.* **13**: 367-372.

Rottingen, I., Foote, K.G., Huse, I., and Ona, E. (1994). Acoustic abundance estimation of wintering Norwegian spring spawning herring, with emphasis on methodological aspects. ICES CM 1994/(B+D+G+H):1.

RTI (Research Triangle Institute) (2001). SUDAAN User's Manual, Release 8.0. Research Triangle Park, NC: Research Triangle Institute.

Sainsbury, K.J., Punt, A.E., and Smith, A.D.M. (2000). Design of operational management strategies for achieving fishery ecosystem objectives. *ICES Journal of Marine Science.* **57**: 731-741.

Salthaug, A., and Aanes, S. (2003). Catchability and the spatial distribution of fishing vessels. *Canadian Journal of Fisheries and Aquatic Sciences* **60**:259-268.

SESITS (Evaluation of Demersal Resources of Southwestern Europe from Standardised Groundfish Surveys). D.G. XIV Study contract reference: 96-029.

Shotton, R., and Bazigos, G.P. (1984). Techniques and considerations in the design of acoustic surveys. *Rapp. P.-v. Reun. Cons. int Explor. Mer.* **184**: 34-57.

Sigler, M.F., and Fujioka, J.T. (1988). Evaluation of variability in sablefish, Anoplopoma fimbria, abundance indices in the Gulf of Alaska using the bootstrap method. *Fishery Bulletin* **86**: 445-452.

Simard, Y., Marcotte, D., and Naraghi, K. (2003). Three-dimensional acoustic mapping and simulation of krill distribution in the Saguenay - St. Lawrence marine park whale feeding ground. *Aquat. Living Resour.* **16**: 137-144.

Simmonds, E.J. (2003). Weighting of acoustic- and trawl-survey indices for the assessment of North Sea herring. *ICES Journal of Marine Science.* **60**: 463-471.

Simmonds, E.J., Williamson, N.J., Gerlotto, F., and Aglen, A. (1992). Acoustic survey design and analysis procedures: a comprehensive review of current practice. *ICES Cooperative Research Report* **187**: 127.

Simmonds, E.J., and Fryer, R.J. (1996). Which are better, random or systematic acoustic surveys ? A simulation using North Sea herring as an example. *ICES Journal of Marine Science.* **53**: 39-50.

Simmonds, J., Toresen, R., Pedersen, J., and Gotze, E. (1998). Intercalibration of participating vessels in the ICES Coordinated Surveys of North Sea Herring. ICES CM 1998/OPEN:23 Poster.

Sinclair, A., Gascon, D., O'Boyle, R., Rivard, D., and Gavaris, S. (1991). Consistency of some northwest Atlantic groundfish stock assessments. *Northwest Atlantic fisheries Organization Scientific Council Studies* **16**:59-77.

Sinclair, A.F. (1998). Estimating trends in fishing mortality at age and length directly from research survey and commercial catch data. *Canadian Journal of Fisheries and Aquatic Sciences* **55**:1248-1263.

Skagen, D.W., and Hauge, K.H. (2002). Recent development of methods for analytical fish stock assessment within ICES. *ICES Marine Science Symposia* **215**: 523-531.

Smith, S.J. (1981). A comparison of estimators of location for skewed populations - with applications to groundfish trawl surveys. *Bottom Trawl Surveys*, Canadian Special Publication of Fisheries and Aquatic Science. **58:** 8-10.

Smith, S.J. (1990). Use of statistical models for the estimation of abundance from groundfish trawl survey data. *Canadian Journal of Fisheries and Aquatic Sciences.* **47**: 894-903.

Smith, S. J. (1996). Analysis of data from bottom trawl surveys. in H. Lassen. ed. Assessment of groundfish stocks based on bottom trawl surveys. *NAFO Scientific Council Studies* **28**: 25–53.

Smith, S.J. (1997). Bootstrap confidence limits for groundfish trawl survey estimates of mean abundance. *Canadian Journal of Fisheries and Aquatic Sciences.* **54**: 616-630.

Smith, S.J. (1999). Survey method for fisheries: challenges and statistical issues. *Statistical Society of Canada Annual Meeting*, June 1999, Regina, Canada.

Smith, S.J., and Gavaris, S. (1993a). Improving the precision of abundance estimates of eastern Scotian Shelf Atlantic cod from bottom trawl surveys. *North American Journal of Fisheries Management* **13**: 35-47.

Smith, S. J., and Gavaris, S. (1993b). Evaluating the accuracy of projected catch estimates from sequential population analysis and trawl survey abundance estimates. in S. J. Smith, J. J. Hunt and D. Rivard. eds. Risk evaluation and biological reference points for fisheries management. Canadian Special Publication of Fisheries and Aquatic Science. 120: 163–172.

Smith, S. J., and Robert, G. (1998). Getting more out of your survey designs: an application to Georges Bank scallops (*Placopecten magellanicus*), in G. S. Jamieson and A. Campbell, eds, 'Proceedings of the North Pacific Symposium on Invertebrate Stock Assessment and Management', Special Publication of Canadian Fisheries and Aquatic Science, 125, pp. 3–13.

Smith, S. J., and Lundy, M. J. (2002). Scallop production area 4 in the Bay of Fundy: stock status and forecast. Canadian Science Advisory Secretariat Research Document. **2002/018**, 86 pp. http://www.dfo-mpo.gc.ca/csas/Csas/English/Research_Years/2002/R_D2002_e.htm

Smith, S. J., Lundy, M. J., Roddick, D., Pezzack, D., and Frail, C. (2003). Scallop production areas in the Bay of Fundy and scallop fishing area 29 in 2002: stock status and forecast. Canadian Science Advisory Secretariat Research Document**. 2003/010**, 103 pp.

Smith, S. J., and Rago, P. (2004). Biological reference points for sea scallops (*Placopecten* magellanicus): the benefits and costs of being nearly sessile. Canadian Journal of Fisheries and Aquatic Sciences. **61**: (in press).

Smith, T.D. (1994). Scaling fisheries: the science of measuring the effects of fishing, 1855-1955. Cambridge University Press, Cambridge, 392 pp.

Smith, T.D. (2002). The Woods Hole bottom-trawl resource survey: development of fisheries-independent multipsecies monitoring. *ICES Marine Science Symposia* **215**: 474-482.

Sparholt, H. (1990). Using GLM analysis on the IYFS herring data for the North Sea. *ICES CM 1990/**H**:6.*

Stefansson, G. (1996). Analysis of groundfish survey abundance data: combining the GLM and delta approaches. *ICES Journal of Marine Science.* **53**: 577-588.

Syrjala, S.E. (2000). Critique on the use of the delta distribution for the analysis of trawl survey data. *ICES Journal of Marine Science.* **57**: 831-842.

Taylor, C.C. (1953). Nature of variability in trawl catches. *Fisheries Bulletin* **54**: 145-166.

Thompson, S. K. (1991). Stratified adaptive cluster sampling. *Biometrika* **78**:389-397.

Thompson, S. K. (1992). *Sampling*. John Wiley and Sons, New York, N.Y.

Thompson, S.K., and Seber, G.A.F. (1996). *Adaptive sampling*. John Wiley and Sons Inc., New York, 265 pp.

Tietjen G. L. (1986). *A Topical Dictionary of Statistics*. Chapman and Hall.

Toresen, R., and Østvedt, O.J. (2002). Stock structure of Norwegian spring-spawning herring: historical background and recent apprehension. *ICES Marine Science Symposia* **215**: 532-542.

Torstensen, E., and Toresen, R. (2000). Survey report RV "G.O. Sars" 30.June – 18.July 2000. Fisken Og Havet 8 – 2000. Havforsknings Instituttet.

Valliant, R., Dorfman, A.H., and Royall, R.M. (2000). Finite population sampling and inference: A prediction approach. John Wiley and Sons.

Vølstad, J.H., Christman, M., and Miller, T.J. (in prep). Design efficiencies of transect and stratified random trawl surveys.

von Szalay, P. G. (2003). The feasibility of reducing the variance of fish relative abundance estimates by integrating CPUE data from two demersal trawl surveys in the Gulf of Alaska. *Alaska fishery bulletin* **10**: 1-13

von Szalay, P. G., and Brown, E. (2001). Trawl comparisons of fishing power differences and their applicability to National Marine Fisheries Service and Alaska Department of Fish and Game trawl survey gear. *Alaska fishery bulletin* **8**: 85-95

Walsh, S.J. (1996). Efficiency of Bottom Sampling Trawls in Deriving Survey Abundance Indices. *NAFO Sci.Coun.Studies* **28**: 9-24.

Walsh, S.J. (1997). Performance of mobile and static gears used in single and multi-species resource surveys: a review. *ICES CM 1997/**W:**02.* 22 pp.

Walsh, S.J., Engås, A., Ferro, R.S.T., Fontaine, R., and Van Marlen, B. (2002). To catch or conserve more fish: the evolution of fishing technology in fisheries science. *ICES Marine Science Symposia* **215**: 493-503.

Walters, C., and Maguire, J.J. (1996). Lessons for stock assessment from the northern cod collapse. *Reviews in Fish Biology and Fisheries* **6**(2): 125-137.

Westrheim, S.J. (1967). Sampling research trawl catches at sea. *Journal of the Fisheries Research Board of Canada* **24**: 1187-1202.

Westrheim, S.J. (1976). A further note on sampling research trawl catches at sea. *Journal of the Fisheries Research Board of Canada* **33**: 1196.

Wieland, K. (2003). Abundance of young (Age 1, 2 and 3) northern shrimp (Pandalus borealis) off West Greenland (NAFO Subareas 0+1) in 1993-2003, and changes in mean size-at-age related to temperature and stock size. *NAFO SCR Doc.,* No. 76, Serial No. N4916, 22 p.

Williamson , N.J., and Traynor, J.J. (1996). Application of a one-dimensional geostatistical procedure to fisheries acoustic surveys of Alaskan pollock. *ICES Journal of Marine Science* **53**: 423-428.

Wolter, K.M. (1985). Introduction to Variance Estimation. Springer-Verlag. New York, NY.

Wyeth, M. R., Stanley, R. D., Kieser, R., and Cooke, K. (2000). Use and calibration of a quantitative acoustic system on a commercial fishing vessel. Can. Tech. Rep. Of Fisheries and Aquatic Sciences 2324. 46 p.

Zar, J. H. (1984). Biostatistical analysis. Second edition. Prentice-Hall, Englewood Cliffs, New Jersey, USA. 718 pp.

Zühlke, R, Alvsvåg, J., de Boois, I., Cotter, J., Ehrich, S., Ford, A., Hinze, H., Jarre-Teichmann, A., Jennings, S., Kröncke, I., Lancaster, J., Piet, G., Prince, P. (2002). Epibenthic diversity in the North Sea. *Senckenbergiana maritima* **31**(2): 269-281.

10 Annexes

Annex 1 Participant contact details

Name	Affiliation, location	E-mail address
Jean Adams	USGS, Marquette, U.S.A.	jvadams@usgs.gov
Doug Beare	FRS, Aberdeen, Scotland	bearedj@marlab.ac.uk
Nicola Bez	CDG, Fontainebleau, France	bez@cg.ensmp.fr
Russell Brown	NMFS, Woods Hole, USA	russell.brown@noaa.gov
Steve Buckland	St. Andrews Uni., Scotland	steve@mcs.st-andrews.ac.uk
John Cotter	CEFAS, Lowestoft, England	A.J.Cotter@cefas.co.uk
Paul Fernandes	FRS, Aberdeen, Scotland	fernandespg@marlab.ac.uk
Rob Fryer	FRS, Aberdeen, Scotland	fryerr@marlab.ac.uk
Marco Kienzle	FRS, Aberdeen, Scotland	m.kienzle@marlab.ac.uk
Knut Korsbrekke	IMR, Bergen, Norway	knut.korsbrekke@imr.no
Bart Maertens	DVZ, Ostend, Belgium	bart.maertens@dvz.be
Bob O'Gorman	USGS, Oswego, USA	robert_o'gorman@usgs.gov
Rainer Oeberst	IOR, Rostock, Germany	rainer.oeberst@ior.bfa-fisch.de
Michael Pennington	IMR, Bergen, Norway	Michael.pennington@imr.no
Allan Reese	CEFAS, Lowestoft, England	r.a.reese@cefas.co.uk
John Simmonds	FRS, Aberdeen, Scotland	Simmondsej@marlab.ac.uk
Stephen Smith	DFO, Dartmouth, Canada	Smithsj@mar.dfo-mpo.gc.ca
Dave Somerton	NMFS, Seattle, USA	David.somerton@noaa.gov
Bjorn Steinarsson	MRI, Reykjavik, Iceland	Bjorn@hafro.is
David Stokes	MI, Galway, Ireland	David.stokes@marine.ie
Jon Volstad	Versar, Maryland, USA	jvolstad@versar.com
Paul Walline	NMFS, Seattle, USA	Paul.Walline@noaa.gov
Kai Wieland	GINR, Nuuk,Greenland	wieland@natur.gl
Juan Zwolinski	IPIMAR, Lisbon, Portugal	juan@ipimar.pt

Annex 2 Working Documents

WD1. Fernandes, P.G. Design and analysis of fish surveys: a draft review. Reproduced as section 2 of the report.
WD2. Pennington M.and Nakken O. Timely Evaluation of Stock Status Based on Scientific Surveys
WD3. Chaves, C. and Cardador, F. Portuguese groundfish surveys: an overview.
WD4. Buckland, S. Line transect sampling.
WD5. Cotter, J. Can fishers teach scientists how to improve fish surveys? Selected results from spatially intense, commercial fv surveys of nine English fisheries in 2003-4.
WD6. Zwolinski, J.P. The use of generalised additive models in sardine acoustic estimates.
WD7. Vølstad, J.H., Christman, M. and Miller, T.J. Design efficiencies of transect and stratified random trawl surveys.
WD8. Imrie, C., Mosqueira, I., Beare, D. Reid, D. Korre, A. and McAllister, M. Comparing survey designs and estimators: an example using icthyoplankton data.
WD9. Bez N., Rivoirard J. and Petitgas, P. Geostatistical approach to biomass estimations from survey data.
WD10. Petitgas P. Geostatistics in fisheries survey design and analysis: a history of ideas, 1990-2004.
WD11. Petitgas P. About non linear geostatistics and adaptative sampling.

Timely Evaluation of Stock Status Based on Scientific Surveys

Michael Pennington and Odd Nakken
Institute of Marine Research
PO Box 1870 Nordnes, N-5817 Bergen, Norway

Abstract

The usual method for assessing a fish stock for which there exist scientific surveys and commercial catch statistics is to use the survey series to 'tune' a VPA- type model. Such assessments are often subject to rather large revisions as more catch data for cohorts remaining in the fishery become available. It is conjectured that one reason VPA abundance estimates for cohorts still in the fishery tend to be variable and often biased is that the relation between the composition of the commercial catch and the actual population is unknown, and this relation likely varies from year to year. It is suggested that a more stable method for assessing the current condition of a stock would be to reverse the roles played by surveys and catch data. That is use abundance estimates based on historical catch data (*i.e.* catch statistics for cohorts that are no longer in the fishery) to tune the survey series. As an example, converged VPA-type abundance estimates of Northeast Arctic cod (*Gadus morhua*) during a calibration period were used to 'tune' a yearly bottom trawl survey of this stock. For the two age groups considered in this paper, the survey-based procedure generated estimates of subsequent converged VPA estimates that were usually more precise than the annual estimates. Since survey-based estimates will not be revised and would be available as soon as the survey is completed, it is concluded that they would form a timely basis for developing and implementing a stable management strategy.

1. Introduction

For fish stocks that are monitored by scientific surveys and for which commercial catch statistics are collected, the generally accepted method for assessing these stocks is to combine the survey estimates with the catch data (Figure 1). Before a particular cohort leaves the fishery, the cohort's estimated abundance, based on these virtual population analysis (VPA) type assessments, tends to vary from year to year. Not only are the annual estimates of a cohort's abundance quite variable (see, *e.g.*, Nakken, 1998; Pennington and Strømme, 1998; Korsbreckke *et al.*, 2001), there is a tendency for the catch-based estimates to decrease as more catch data becomes available, which is the so called "retrospective problem" (Sinclair *et al.*, 1991; Parma, 1993; Sinclair, 1998; Mohn, 1999).

One reason that VPA estimates of current stock size are subject to large revisions is that the relation between the commercial catch during recent years and the actual population structure is usually unknown (Figure 1). Many factors may cause this relationship to vary from year to year. One obvious factor is a change in the spatial distribution of fishing effort over time (Salthaug and Aanes, 2003). If the commercial catch data are correct, then for a cohort no longer in the fishery the estimate of its historical abundance (*i.e.* the converged estimate) may be fairly accurate.

Abundance indices based on scientific surveys often track converged VPA estimates fairly closely, while the non-converged estimates and the survey-based indices tend to diverge (Pennington and Godø, 1995; Pennington and Strømme, 1998; Korsbrekke, *et al.*, 2001). Since recent VPA estimates will be revised, while the survey estimates will stay the same, this implies that the information contained in the survey data is not being effectively used to assess the stock.

An alternative to a VPA-type assessment of the current condition of a stock would be to base the assessment only on known, at least in theory, relations. A survey ideally covers the entire stock while converged VPA-type estimates, based on accurate commercial catch data, should provide fairly accurate historical estimates of a cohort's size. Therefore for some stocks it may be sensible to reverse the roles currently played by surveys and commercial catch data. That is instead of using survey data to tune the current catch data, use historical catch data to calibrate the survey indices (Figure 3).

As an example of this alternative assessment technique, converged VPA abundance estimates for Northeast Arctic cod during an initial time period are used to calibrate abundance indices generated for this stock by the winter surveys in the Barents Sea. The survey-based estimates are compared to subsequent converged estimates of cohort size and to the annual assessments.

2. Calibrating Survey Abundance Indices

Suppose for a particular stock there are two abundance estimates; a fishery independent index of relative abundance generated by research surveys and estimates of absolute abundance based on commercial catch data (*e.g.*, a VPA-type estimate). For the converged VPA estimates, which only depend on the commercial data and thus are independent of the survey index, it is assumed that

$$E[N_i] = P_i, \tag{1}$$

where N_i is the estimated stock number for a particular age or group of ages in year i from the VPA and P_i is the true number in the population.

Furthermore, assume that the expected value of the survey index, I_i, is also proportional to P_i. That is

$$E[\beta I_i] = P_i. \tag{2}$$

Then it follows that

$$E[N_i] = \beta E[I_i]. \tag{3}$$

The estimates, N_i and I_i, can be expressed as

$$N_i = E[N_i] + \varepsilon_i \qquad \text{and} \qquad I_i = E[I_i] + \delta_i, \tag{4}$$

respectively, where ε_i and δ_i are assumed to be random errors. Then it follows from equations (1) through (4) that

$$N_i = \beta I_i + \xi_i, \tag{5}$$

where $\xi_i = \varepsilon_i - \beta\delta_i$. Though equation (5) is in the form of a standard regression equation, it differs since I_i and ξ_i are generally not independent, and, therefore, the standard regression estimator of β is usually biased (for more details, see Draper and Smith, 1981). If the variance over time of the expected survey index, $E[I_i]$, is large with respect to the variance of δ_i, then the bias of the standard regression estimator of β will be small and can be safely ignored (Draper and Smith, 1981).

If the relation between the survey index and the VPA estimates differs significantly from equation (5), then this implies that either assumption (2) or (3) or both are not valid. When this is the case, other information needs to be employed to select the index that is most likely proportional to P_i.

For more variable surveys, often using time series techniques will generate a survey index that more closely tracks the converged VPA estimates (Pennington, 1985; Fogarty *et al.*, 1986; Pennington and Godø, 1995). Using a smoothed survey index in Equation (5) will in general cause the error term to be autocorrelated and this should be taken into account in the fitting procedure (Brockwell and Davis, 1996).

3. An Illustrative Example: Northeast Arctic Cod (*Gadus morhua*)

The fishery for Northeast Arctic cod is the largest cod fishery in the world. In the 1950's, which was the peak period for the fishery, the yearly catches averaged

800 000 t (Nakken, 1994). More recently the catches have ranged from more than 700 000 t in 1997 down to 400 000 t in 2000. Fluctuating stock size is the main reason that commercial catches have varied over time, and failure to timely detect changing trends in abundance has been a problem for managing this stock. The next section is an overview of past assessments and in the subsequent sections it is shown that scientific surveys have generally provided a more robust and timely assessment of the condition of the Northeast Arctic cod stock.

3.1 A brief overview of the assessments of Northeast Arctic cod

The International Commission for the Exploration of the Sea's (ICES) annual assessments of the Northeast Arctic cod stock usually underestimated fishing mortality rates and, therefore, overestimated stock numbers (Nakken, 1998; Korsbrekke *et al.*, 2001). The general tendency was that fishing mortality rates for a given year were revised upward and thus stock numbers were reduced as more catch at age data became available, which is an example of the common retrospective problem. An examination of the ICES assessments during the period from 1982 through 1995 indicated that the annual estimates of fishing mortality rates ranged from 55 to 110% of the converged value and was, on average, 80% of the final "true" value. Furthermore, it took four to five years before the estimates converged (Nakken, 1999).

It is not known if this bias was caused by biases in the input data (commercial catch and survey data) or to inadequacies in the applied assessment methodology (XSA, Extended Survivor Analysis). Whatever the case, this fairly consistent overestimation of stock size had the unfortunate effect that often management measures to reduce fishing mortality were ineffective.

3.2 Calibrated survey estimates of stock size

Since 1981 the Institute of Marine research has conducted an extensive bottom-trawl survey (approximately 250 stations per survey) in the Barents Sea from mid-February to mid-March (for more details, see Aglen *et al.*, 2003). The area covered by the winter survey was expanded in 1993, and the survey abundance indices (Table 1) prior to 1993 were adjusted accordingly (Korsbrekke *et al.*, 2001). We included in the analysis only cod age 4 and older since few cod younger than 4 are in this fishery. Because of sampling variability and aging errors, both the survey indices and the VPA estimates at age were rather imprecise (Pennington *et al.*, 2002; Aanes and Pennington, 2003). Therefore, to lessen the effect of aging errors, our primary focus was on two age groupings: ages 4, 5 and 6, which for the most part were pre-spawners; and ages 7+, which were mainly spawners. The annual ICES abundance estimates for the two age groups during the period 1995 through 2003 are in Table 2.

For the calibration period, 1981 through 1995, the VPA abundance estimates for ages 4 through 6 and 7+ were not proportional to the survey indices, but were linearly related (*i.e.* $N_t = \hat{\beta} I_t + \hat{\alpha}$) with a significantly positive intercept (Figure 3). Since the estimated standard error of I_t was small compared with the range of the index, it is unlikely that a bias in the estimate of the slope caused the intercept to be positive (Draper and Smith, 1981). Therefore it follows that either the converged VPA estimates or the survey indices (or both) were not proportional to the true population.

Since one of our objectives was to use the surveys to predict the final, converged VPA estimates, we first included an intercept when calibrating the surveys and then, assuming that the survey abundance index was proportional to the actual population (*i.e.* the intercept equaled 0), we generated estimates of the proportionality constant in equation (5).

The calibrated survey estimates of abundance for 1995 though 2004 for the two age groupings, generated by the regressions shown in Figure 3, are in Table 3, column 4. Figures 4 and 5 are plots of the calibrated survey estimates and the latest ICES (2003) abundance estimates, along with the annual ICES estimates during this period.

Even though including an intercept in the calibration procedure resulted in rather accurate predictions of the ICES converged abundance estimates for the two age groups, there is no reason to assume that the ICES converged estimates were unbiased. The implication of a positive intercept is that it will tend to ‘maintain’ stock numbers as the survey index decreases. In particular, when abundance is low the stock may be declining (or increasing) at a much faster rate than indicated by the survey-based estimates.

Aging errors tend to cause the sizes of small cohorts to be greatly overestimated and large cohorts to be slightly underestimated. To reduce this source of bias, the survey was calibrated, using Equation (5), based only on those years for which the survey index was greater than its average over the calibration period (Nakken and Pennington, 2001). These calibration lines are shown in Figure 3 and the associated survey-based abundance estimates are in Table 3, columns 5 and 9.

3.3 Discussion

The calibrated survey estimates of the abundance of ages 4 through 6 cod closely tracked the 2003 ICES estimates during the years 1995 through 1998, a period in which the ICES estimates have converged, whereas for recent years the calibrated estimates were below the 2003 ICES estimates (Figure 4). It should be noted that prior to 1998, the tendency was for the annual ICES estimates to decline over time as more catch data became available, while in the last few years the annual ICES abundance estimates have increased.

For ages 7+, which form the bulk of the spawning stock, the survey-based abundance estimates and the 2003 ICES estimates are fairly similar (Figure 5). An advantage of the survey-based estimates for both age groups is that they accurately predicted the converged 2003 ICES estimates for the period 1995 – 1998 years before the annual assessments converged to their present levels (Figures 4 and 5). In particular, while the 1996 through 1998 ICES assessments were significantly overestimating spawning stock size, the survey-based estimates were apparently providing accurate and timely estimates of the true size of the spawning stock (Figure 5).

It is not clear at this time whether or not to include an intercept when calibrating the winter survey. Though aging errors may be a problem, there is no other apparent reason for the surveys to be biased, while there are many possible factors that may cause converged VPA estimates to be biased, such as misreporting of catches, discarding at sea, *etc.* Helle *et al.* (2000) found that the estimates of the relative abundance of Northeast Arctic cod as three-year-olds generated by the winter survey were proportional to survey-based estimates of the cohort's abundance at earlier life stages. In contrast, the VPA estimates of the abundance of age 3 cod were not proportional to survey indices at any stage, and in particular, the intercepts were all significantly positive. The consistency of the survey indices is an indication that it may be the converged VPA estimates that are not proportional to 'true' stock numbers.

4. Conclusions

There are at least two likely reasons that simply calibrating a survey series using historical catch data may generate more robust abundance estimates than a VPA-type analysis. The first is that the calibration procedure is based on known relations: the total catch of each cohort that has gone through the fishery, and on a scientific survey that monitors the cohorts still in the fishery. The second reason that calibrated survey estimates may be more robust is that the calibration procedure is based on a much simpler model than a VPA-type analysis. It has been observed in many fields that predictions based on complicated structural models are often less accurate than those based on simpler models (see, *e.g.*, Nerlove *et al.*, 1979; Wheelwright and Makridakis, 1985; Newbold *et al.*, 1993; Stergiou *et al.*, 1993). Jenkins (1976, page 132) gives a nice summary of some of the problems associated with predictions based on complicated models.

Because of inaccurate or incomplete commercial catch data; converged VPA-type estimates of abundance may be biased. Therefore, it is important to choose a calibration period for which the catch data are judged to be fairly accurate, or use ancillary information on known or likely sources of errors to adjust the historical catch data.

For stocks that are adaptively managed, one advantage of using survey-based estimates is that they would be available as soon as the survey is finished. It often takes considerable time to collect and collate commercial fishery data, and, therefore, VPA-type estimates are usually not available until several months after the survey is completed. As demonstrated by the Northeast Arctic cod example, another appealing feature of survey-based abundance estimates is that they are not subject to frequent revisions as are the VPA-type estimates. It is difficult to see how stable and effective management strategies can be agreed on and implemented if the assessment of the condition of the stock varies significantly from year to year (Figures 4 and 5, Table 2).

References

Aanes, S., Pennington, M., 2003. On estimating the age composition of the commercial catch of Northeast Arctic cod from a sample of clusters. ICES Journal of Marine Science. 60:297-303.

Aglen, A., Alvsvåg, J. Halland, T.I., Høines, Å., Nakken, O., Russkikh, A., Smirnov, O., 2003. Investigations on demersal fish in the Barents Sea winter. 2003. IMR/PINRO Joint Report Series, No.1/2003.

Brockwell, P.J., Davis, R.A., 1996. Introduction to time series and forecasting. Springer-Verlag, New York, 420pp.

Draper, N.R., Smith, H. 1981. Applied Regression Analysis. 2nd ed. John Wiley and Sons, New York, 709 pp.

Fogarty, M.J., Idoine, J.S., Almeida, F.P., Pennington, M., 1986. Modeling trends in abundance based on research vessel surveys. International Council for the Exploration of the Sea (ICES) C. M. 1986/G:92.

Helle, K., Bogstad, B., Marshall, C., Michalsen, K., Ottersen, G., Pennington, M. 2000. An evaluation of recruitment indices for Arcto-Norwegian cod (*Gadus morhua* L.). Fisheries Research, 48:55-67.

Jenkins, J.M., 1976. Practical experiences with modeling and forecasting time series. In: Anderson, O.D. (Ed.), Forecasting. North-Holland, New York, pp. 43-166.

Korsbrekke, K., Mehl, S., Nakken, O., Pennington, M., 2001. A survey-based assessment of the Northeast Arctic cod stock. ICES Journal of Marine Science, 58:763-769.

Mohn, R., 1999. The Retrospective problem in sequential population analysis: an Investigation using cod fishery and simulated data. ICES J. Mar. Sci. 56 :473-488.

Nakken, O. 1994. Causes of trends and fluctuations in the Arcto-Norwegian cod stock. ICES Marine Science Symposia, 198: 212-228.

Nakken, O., 1999. Retrospective review of management advice and TAC's for some stocks. In: Jakobsen, T. (Ed.) 2002.Management strategies for the fish stocks in the Barents Sea. Proceedings of the 8th Norwegian-Russian Symposium.

Nakken, O., 1998. Past, present and future exploitation and management of marine resources in the Barents Sea and adjacent areas. Fisheries Research. 37:23-35.

Nakken, O., Pennington, M., 2001. On the relation between 'true' numbers of Northeast Arctic cod and VPA- or survey-based abundance estimates. ICES CM 2001/Q:15.

Nerlove, M., Grether, D.M., Carvalho, J.H., 1979. Analysis of Economic Time Series. Academic Press, New York, NY, 468 pp.

Newbold, P., Agiakloglou, C., Miller, J., 1993. Long-term inference based on short-term forecasting models. In: T. Subba Rao (Ed.), Developments in Time Series Analysis. Chapman and Hall, London, pp. 9-25.

Parma, A.M., 1993. Retrospective catch-at-age analysis of Pacific halibut: implications on assessments of harvesting policies. G. Kruse, D.M. In: Eggers, R.J. Marasco, C. Pautzke and T.J. Quinn II (Eds.). Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations, Alaska Sea Grant College Program Report No. 93-02, Fairbanks, University of Alaska.

Pennington, M. 1985. Estimating the relative abundance of fish from a series of trawl surveys. Biometrics 41:197-202.

Pennington, M., L.-M. Burmeister and V. Hjellvik. 2002. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fishery Bulletin. 100:74-81.

Pennington, M., Godø, O. R. 1995. Measuring the effect of changes in catchability on the variance of marine survey abundance indices. Fisheries Research 23:301-310.

Pennington, M., Strømme, T. 1998. Surveys as a research tool for managing dynamic stocks. Fisheries Research 37:97-106.

Salthaug, A., Aanes, S. 2003. Catchability and the spatial distribution of fishing vessels. Canadian Journal of Fisheries and Aquatic Sciences 60: 259-268.

Sinclair, A., Gascon, D., O'Boyle, R., Rivard, D., Gavaris, S., 1991. Consistency of some northwest Atlantic groundfish stock assessments. Northwest Atlantic fisheries Organization Scientific Council Studies 16: 59-77.

Sinclair, A.F., 1998. Estimating trends in fishing mortality at age and length directly from research survey and commercial catch data. Canadian Journal of Fisheries and Aquatic Sciences 55:1248-1263.

Stergiou, K.I., Christou, E.D., Petrakis, G., 1997. Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods. Fish. Res., 29:55-95.

Wheelwright, S.C., Makridakis, S., 1985. Forecasting Methods for Management (4^{th} ed.). John Wiley and Sons, New York, NY, 404 pp.

Table 1**.** The winter survey indices of abundance for Northeast Arctic cod adjusted for the expansion of the survey area in 1993.

	Age			
Year	4	5	6	7+
1981	29.1	47.6	45.7	7.3
1982	34.6	28.1	18.5	21.0
1983	65.4	51.6	20.2	12.0
1984	35.4	25.5	14.0	5.5
1985	157.5	23.7	9.1	4.4
1986	179.9	76.4	9.9	2.8
1987	488.9	64.7	18.7	3.0
1988	100.6	206.3	24.4	4.9
1989	94.8	45.0	107.4	13.0
1990	43.6	41.2	24.5	34.8
1991	42.1	30.5	25.6	30.3
1992	72.1	21.2	15.3	17.7
1993	140.1	72.5	15.8	14.7
1994	310.2	147.4	50.6	14.6
1995	241.4	255.9	76.7	22.9
1996	115.4	137.2	106.1	27.7
1997	64.0	70.4	52.7	35.4
1998	181.3	36.5	25.9	27.9
1999	173.2	58.1	13.4	13.1
2000	132.1	108.3	26.9	7.9
2001	182.8	83.4	38.2	10.6
2002	135.0	109.6	42.5	18.0
2003	129.7	91.1	67.3	24.4
2004	172.5	56.9	44.7	37.0

Table 2. Annual ICES estimates of the abundance of the total number (in millions) of Northeast Artic cod ages 4 through 6 (a) and ages 7+ (b). The data are from the annual reports of the Arctic Fisheries Working Group, ICES, 1995 – 2003.

(a) Ages 4 through 6

Abundance Year	Assessment Year 95	96	97	98	99	00	01	02	03
1995	1069	1646	1418	1363	1257	1240	1254	1233	1233
1996		1219	1038	1011	927	905	914	899	899
1997			610	637	607	583	581	574	574
1998				669	611	648	634	620	632
1999					768	792	758	748	776
2000						831	801	776	850
2001							884	858	964
2002								834	973
2003									902

(b) Ages 7+

Abundance Year	95	96	97	98	99	00	01	02	03
1995	140	108	94	89	86	85	86	85	85
1996		223	176	157	141	141	141	136	137
1997			268	240	186	179	181	178	177
1998				190	138	130	131	126	125
1999					87	77	76	72	72
2000						50	45	45	46
2001							56	58	63
2002								93	107
2003									154

Table 3. Calibrated survey abundance estimates (with and without an intercept) of the number (in millions) of Northeast Arctic cod in the two age groupings and for comparison, the 2003 ICES estimates and the annual ICES estimates of abundance.

	Estimated total number of ages 4 - 6				Estimated total number of ages 7+			
Year	ICES 2003	ICES Annual	Calibrated Survey (intercept)	Calibrated Survey (no intercept)	ICES 2003	ICES Annual	Calibrated survey (intercept)	Calibrated survey (no intercept)
1995	1233	1069	1269	1249	85	140	127	128
1996	899	1219	910	827	137	223	150	157
1997	574	610	574	430	177	268	182	196
1998	632	669	634	500	125	190	150	157
1999	776	768	630	495	72	86	87	78
2000	850	831	703	564	46	50	60	45
2001	964	884	779	672	63	55	83	73
2002	937	834	739	624	107	93	105	101
2003	902	902	754	644	154	154	134	137
2004	NA	NA	703	585	NA	NA	191	207

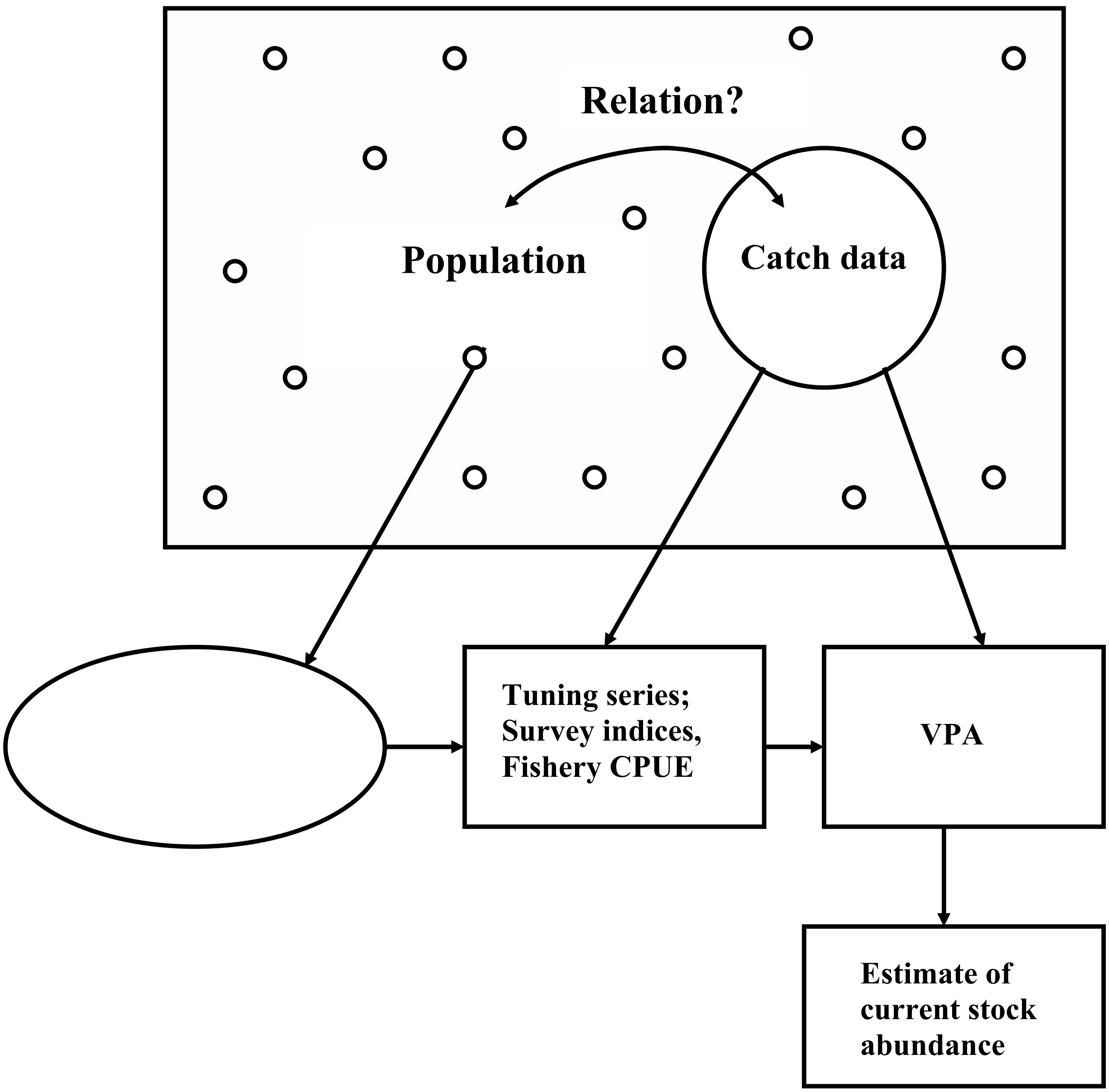


Figure 1. Diagram of the data flow for the standard VPA-type assessment of a stock for which both fishery independent survey data and commercial catch data are available.

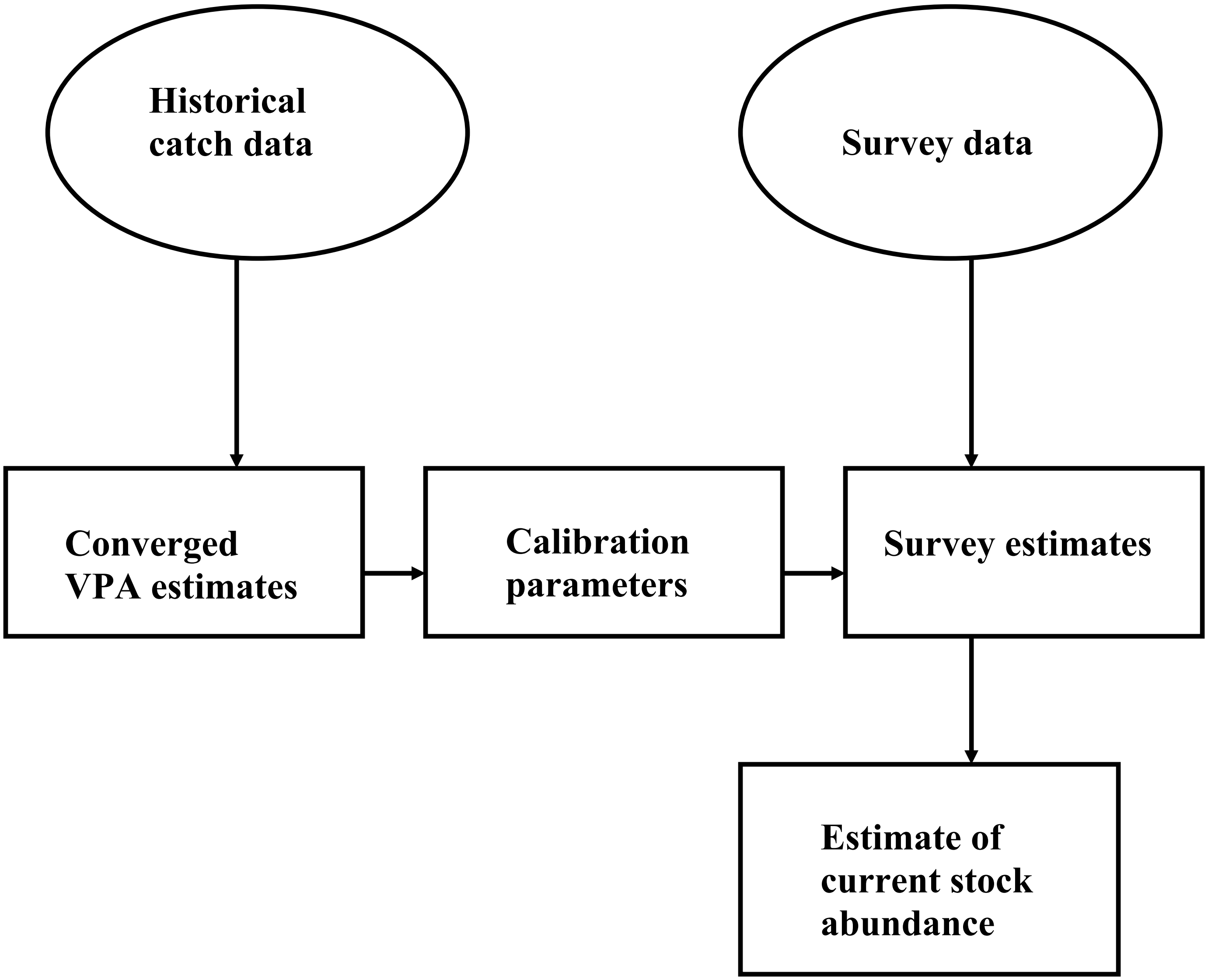
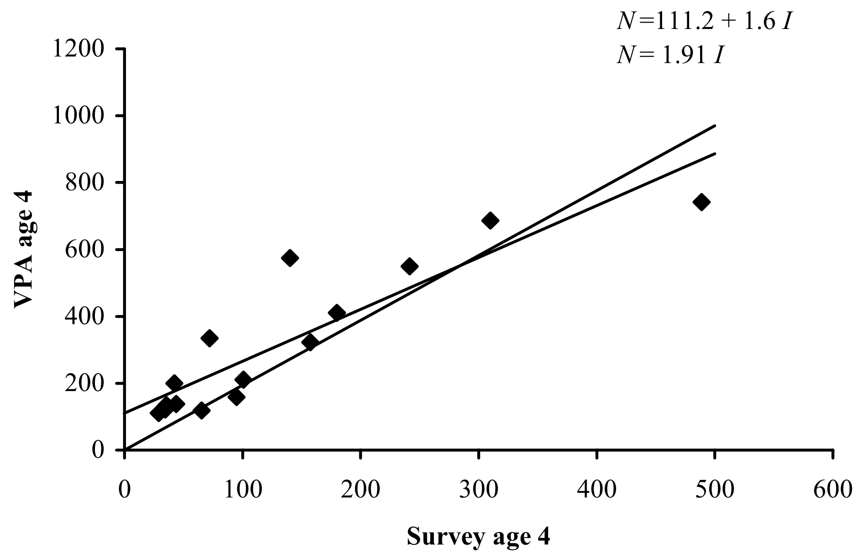
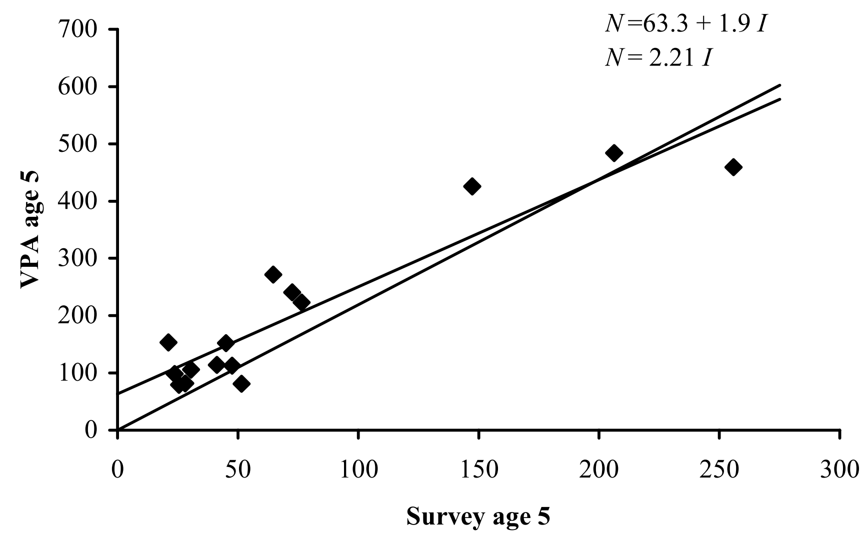
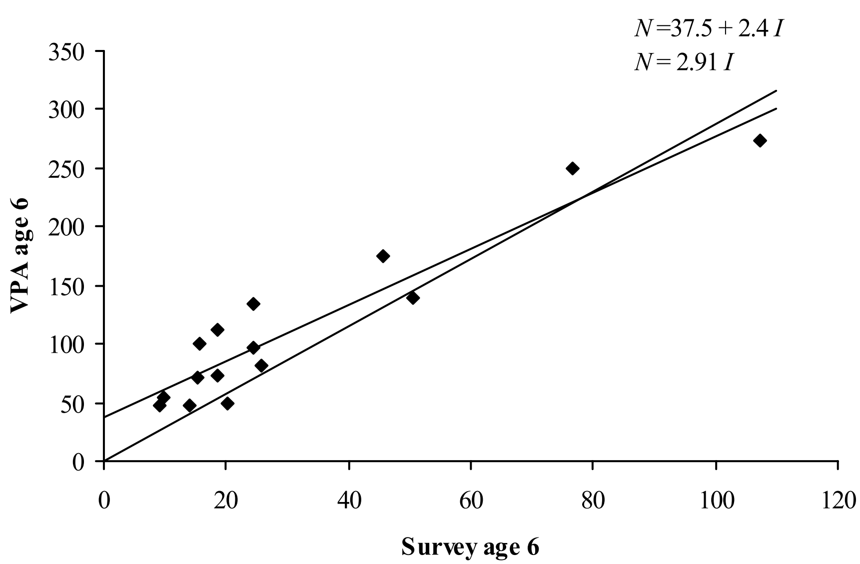


Figure 2. Diagram of the assessment procedure when historical catch data are used to calibrate the survey data.







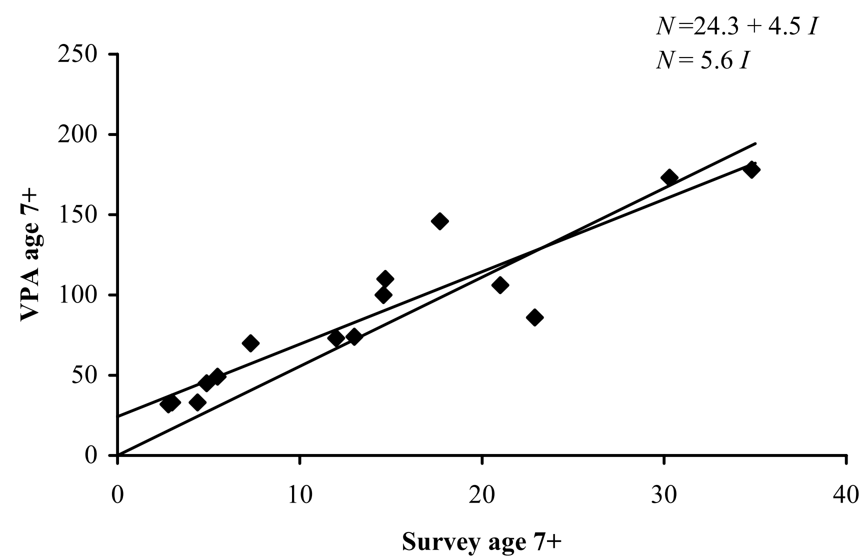


Figure 3. VPA estimates of the abundance of Northeast Arctic cod versus the winter survey index of abundance for the period 1981 through 1995. The regression line with an intercept is based on all the data and the line through the origin is based only on those years for which the survey index is greater than its mean value

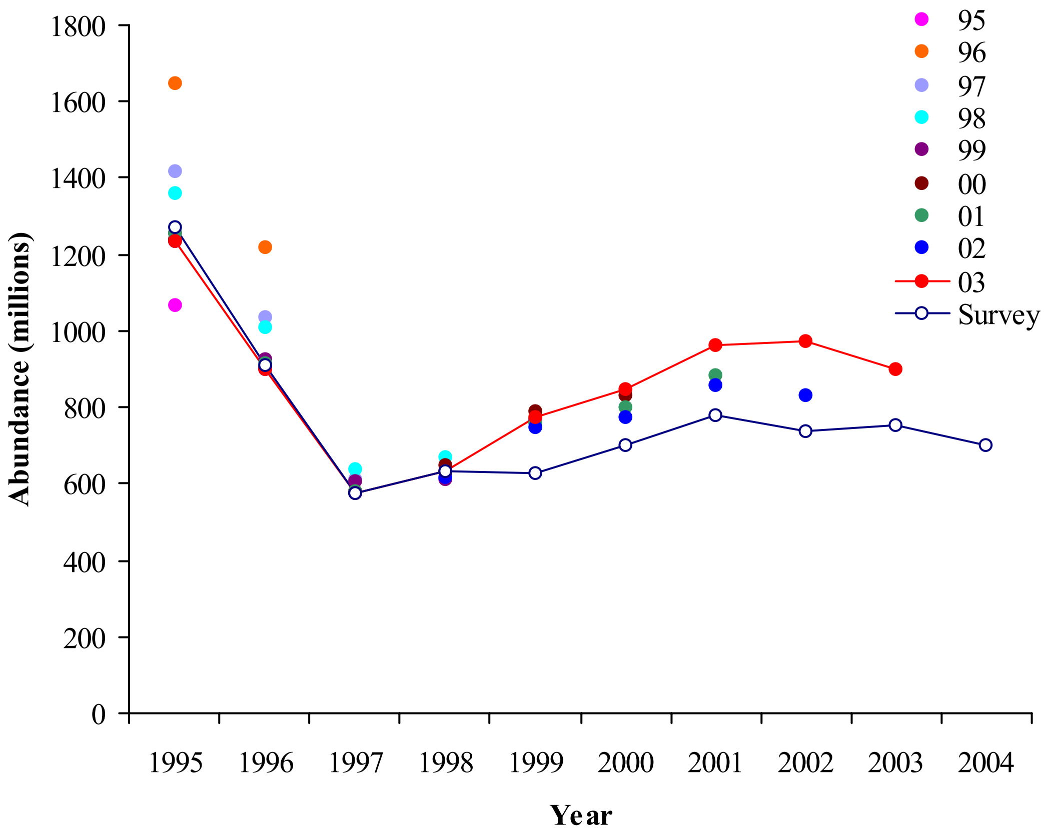


Figure 4. Calibrated survey estimates (connected open circles), ICES 2003 estimates (connected solid circles) and the 1995- 2002 ICES annual assessments (unconnected solid circles) of the total number of Northeast Arctic cod ages 4 through 6.

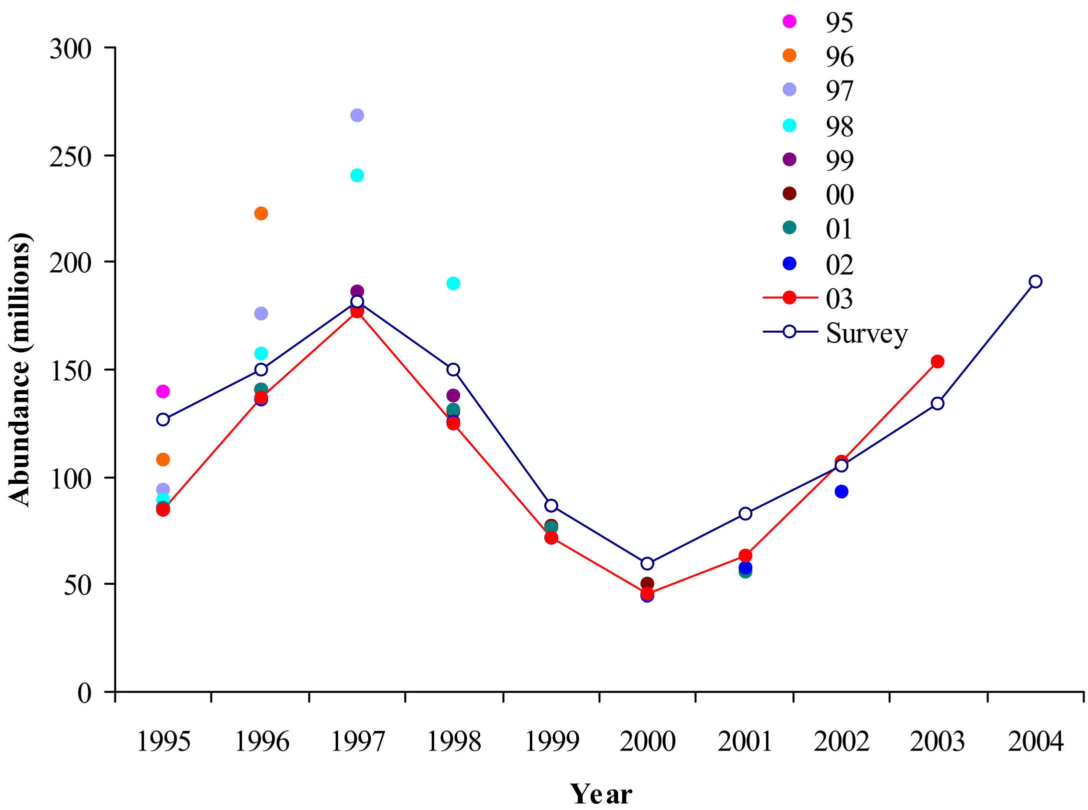


Figure 5. Calibrated survey estimates (connected open circles), ICES 2003 estimates (connected solid circles) and the 1995- 2002 ICES annual assessments (unconnected solid circles) of the total number of Northeast Arctic cod ages 7 and older.

Working Document to be presented at the Workshop on Survey Design and Data Analysis, 21st to 25th June 2004, Aberdeen.

PORTUGUESE GROUNDFISH SURVEYS: AN OVERVIEW.

Corina Chaves and Fátima Cardador
INIAP/IPIMAR, Av. Brasília, 1449-006 Lisbon, Portugal
corina@ipimar.pt, cardador@ipimar.pt

INTRODUCTION

Portuguese groundfish surveys are conducted along the Portuguese continental waters since June 1979, twice a year in Summer and Autumn, on board of the R/V *Noruega*.

The data collected has been the main source for several biological studies and stock parameters estimation. These studies include those related to the species distribution by area and depth, recruitment estimation, abundance indices length-weight relationship, age determination, maturity, fecundity and food habits (Table 1). Portuguese groundfish survey series have been providing the basic information to perform the analytical assessments of several stocks of the Northeast Atlantic currently assessed within ICES Assessment Working Groups (Table 2), the results have been reported to the relevant ICES Working Groups, communicated to ICES Annual Conferences and/or published in journals of biology and fisheries. Data collected from groundfish surveys were also the basis to carry out assemblage studies and to several advices to the Portuguese and European fisheries Administrations concerning the implementation of technical measures for fish stock management.

OBJECTIVES

The main objectives of the surveys are to:

- estimate indices of abundance and biomass of the most important commercial species off the Portuguese continental waters;
- study the distribution pattern and estimate indices of abundance for recruits;
- estimate biological parameters, maturity evolution, sex-ratio, weight, food habits;.
- build length and/or age compositions for the target species.

The target species are hake (*Merluccius merluccius*), horse mackerel (*Trachurus trachurus*), mackerel (*Scomber scombrus*), blue whiting (*Micromessistius poutassou*), megrims (*Lepidorhombus boscii* and *L. whiffiagonis*), monkfish (*Lophius budegassa* and *L. piscatorius*) and Norway lobster (*Nephrops norvegicus*).

GROUNDFISH SURVEY HISTORY

A stratified random sampling design was adopted during 1979-1989. The number of strata changed during this period: from 1979 to 1980 the surveyed area was divided into 15 strata and since 1981 into 36 strata. The design was revised in order to decrease the size of the strata. The new strata are smaller than the previous ones and can be combined to get the older ones. The aim of increasing the number of strata was to increase the probability of spreading the random sampled units in order to decrease the total variance of the mean abundance indices by species. The boundaries of each stratum are based on depth and geographical areas. The depth ranges during 1979-1988 surveys were 20-100m, 101-200m and 201-500m. Each stratum was divided into units of approximately 25 nm^2, sequentially numbered. During 1979-1980 the number of random hauls per stratum was based on the previous information of the relative abundance of the target species in each geographical area and on the vessel time available. During 1981-1989, when the number of strata was 36, two random units were sampled by stratum whenever possible, to become possible to estimate the standard error of the stratified mean by stratum. In Autumn 1989 a fixed stations plan was established as a result of an extensive discussion on the scope of ICES Methods Working Group (ICES, 1990) about the trade on biased estimations with low variance (fixed design) or unbiased estimations with large variance (stratified design). The fixed design is more appropriate for a time series obtained for the purpose of tuning the commercial catch-at-age time series. As a result it was considered that the fixed station design is more appropriate for VPA tuning than the random allocation design. Simultaneously the survey area was extended to the 750 m bathymetric in order to sample the adult hake, and the lower distribution bound of Norway lobster and monkfish.

The tow duration in Summer surveys has always been 60 minutes. The Autumn surveys varied in tow duration: it was 60 minutes during 1979-1980, changing to 30 minutes during 1981-1989, changed to 60 minutes in 1990-2001, and back to 30 minutes since 2002, at a trawling speed of 3.5 knots. The first decrease from 60 to 30 minutes in 1981 was based on an analysis which has indicate that a 30 minutes tow was enough to get abundance indices for the target species recruits which was validated by Cardador (1983). However in the 1989 Summer survey, experiments with the two durations at the trawling speed of 3.5 Knots have been performed indicating that 60 minutes tow was more adequate to sample all the structure of the horse mackerel population. The large adults of horse mackerel were not caught at a trawling speed of 3.5 knots with duration of 30 minutes because the large pelagic fish can swim at higher speeds in front of the trawl net. It is by maintaining the trawl pursuing the fish during a longer period than 30 minutes that the larger horse mackerel looses its *stamina* and enters into the trawl net. The juveniles were well sampled with 30 minutes trawling at 3.5 knots (Cardador *et al.*, 1997). In 2002, the decrease in tow duration was mainly due to the lack of time in survey duration and to increase the number of hauls. The experiments performed in Summer 2002 survey indicate that 30 minutes are enough to sample recruits of hake and horse mackerel which are the main objective of autumn surveys.

SAMPLING DESIGN

Since 1989, the surveys cover the Portuguese continental waters from 20 m to 750 m, following a fixed station sampling scheme. A total of 97 fixed stations are planned, spread over 12 sectors. Each sector is subdivided into 4 depth ranges: 20-100m, 101-200m, 201-500m and 501-750 m, with a total of 48 strata (Figure 1). The positions of the 97 fixed stations were selected based on common stations made during 1981-1989 surveys and taking into account that two stations should be made by stratum.

Fishing stations take place during daylight, with an average duration tow of 60 minutes till 2001 and 30 minutes since 2002, with a mean trawl speed of 3.5 knots.

Oceanographic stations take place at the final of each fishing station using a CTD equipment in order to get temperature and salinity data by depth to be used in biological studies.

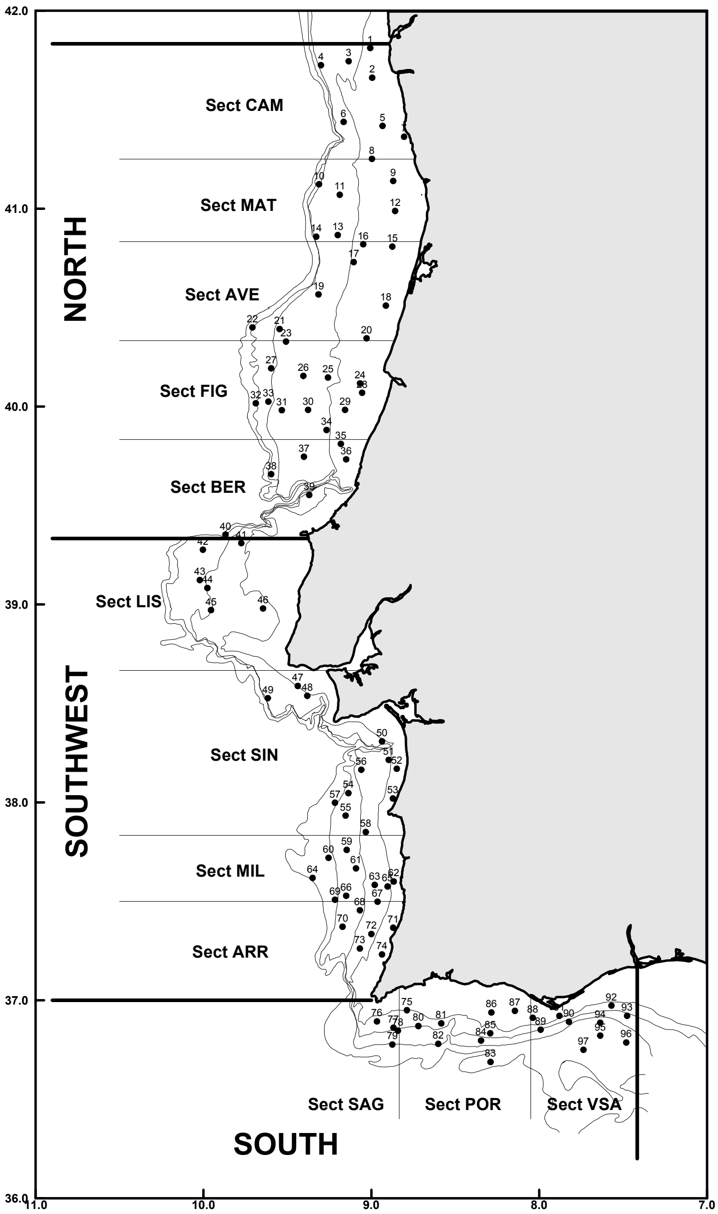


Figure 1 – Sampling scheme used in Portuguese Groundfish Surveys.

During the period 1979-2003 a total of 49 surveys were carried out. The season, total fishing days and valid hauls by survey and depth are shown in the Table 3. In average 2 surveys per year were carried out, with 20 effective fishing days and 86 valid hauls per survey.

SHIP AND GEAR CHARACTERISTICS

Portuguese surveys are carried out on board of the Portuguese R/V *Noruega*, a stern trawler with 47.5 m length overall, 495 tons GRT and 1500 HP, built in 1978 in Bergen, Norway.

The trawl gear used is denominated Norwegian Campelen Trawl 1800/96 (NCT). The main characteristics of this gear are the ground rope with bobbins, 9 m sweep and three bridles, lower 40 m, and upper and middle 20 m long. During 1979-1980 a codend of 40 mm mesh size was used. Selectivity studies conducted during 1981 groundfish surveys were made with a cover of 20 mm and a codend of 40 mm mesh. Since 1982 a single codend of 20 mm mesh size is adopted. The mean vertical opening is 4.6 m and the mean horizontal opening between wings and doors are 15.1 m and 45.7 m, respectively. These gear parameters were obtained with Scanmar Equipment. The polyvalent trawl doors used are rectangular (2.7 m x 1.58 m) with an area of 3.75 m^2 and weighting 650 Kg (Borges *et al*, 1999).

INFORMATION COLLECTED

The catch from each haul is sorted, counted and weighed by species. For the target species and for some other commercial species (fishes, cephalopods and crustaceans) length measurements, as well as other biological information, e.g., weight by length group, sex, maturity stages, stomach contents, are undertaken. Furthermore, complete species list is provided and information on the length distribution of other commercial species are available. Table 1 summarizes the referred studies by species. Most of the results of these studies are reported or published.

DATA PROCESSING AND ANALYSIS

All data collected, log sheets, catch data, biological data, etc, are stored into a database. Apart from biological studies, the main concern about the data collected is to obtain abundance indices disaggregated by age to tune VPA in analytical assessments, especially for hake and horse mackerel. Attempts have been made to take profit of the data series to other species stock assessments like blue whiting, mackerel, megrim and four-spot-megrim. Nevertheless the Portuguese gear used (Norwegian Campell Trawl) is not efficient to catch monkfish and megrims and cannot appropriately sample these stocks.

The estimation of the abundance and biomass indices is based on the methodology presented by Cochran (1960) for calculation of estimators for the stratified random sampling:

Sample stratified mean: $$\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N}$$

Sample stratified variance: $$s^2(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h (N_h - n_h) \frac{s_h^2}{n_h}$$

Where:

N – Total number of units in all strata, $N = N_1 + N_2 + \ldots + N_L$

N_h – Total number of units in stratum *h*

n_h – Number of samples in stratum *h*

y_{ha} – Catch in number (or weight) in haul *a* in stratum *h*

$\bar{y}_h$ – sample mean of catch in number (or weight) in stratum *h*, $\bar{y}_h = \frac{1}{n_h} \sum_{a=1}^{n_h} y_{ha}$

s_h^2 – sample variance in stratum *h*, $s_h^2 = \frac{1}{n_h - 1} \sum_{a=1}^{n_h} (y_{ha} - y_h)^2$

For abundance indices disaggregated by age, estimators of variance by age are obtained by the methodology presented by Cochran (1960) and Goodman (1960):

Sample mean by age *i*: $$\bar{n}_i = \sum_{l} (\bar{n}_l \times p_{il})$$

Sample variance by age *i*:

$$\mathrm{var}(\overline{n}_i) = \sum_l \left[n_l^2 \times \mathrm{var}(p_{il})\right] + \sum_l \left[\mathrm{var}(n_l) \times p_{il}^2\right] + \sum_l \left[\mathrm{var}(n_l) \times \mathrm{var}(p_{il})\right]$$ [1]:

Where:

$p_{il} = \dfrac{o_{il}}{o_l}$; proportion of individuals with age i in length class l,

o_{il} – number of otoliths read in age i and length class l

o_l - number of otoliths read in length class l

Assuming that the number of otoliths read in each length class is much lower then the total number of individuals in the length class in the population, the associated variance is:

$$\mathrm{var}(p_{il}) = \frac{p_{il}(1 - p_{il})}{o_l - 1}$$

$\overline{n}_{il} = \overline{n}_l \times p_{il}$; mean number per hour of individuals with age i in length class l

$\mathrm{var}(\overline{n}_{il}) = \overline{n}_l^2 \times \mathrm{var}(p_{il}) + \mathrm{var}(n_l) \times p_{il}^2 + \mathrm{var}(n_l) \times \mathrm{var}(p_{il})$

$\overline{n}_l$ - mean number per hour of individuals in length class l

$\overline{n}_i$ - mean number per hour of individuals in age i (frequency by age)

POSSIBILITIES OF IMPROVEMENT AND REVISION OF SAMPLING DESIGN

The ICES Southern Shelf Demersal Stocks Working Group pointed out, in 1996 (ICES, 1996a), some difficulties to use the abundance indices in the case of stocks distributed on the area covered by more than one country (hake for example), due to some discrepancies in the indices estimated from the distinct surveys. The International Bottom Trawl Surveys Working Group (ICES, 1996b) pointed out the lack of coordination and standardization of these surveys. To respond to these problems the study project SESITS titled "Evaluation of demersal resources of Southwestern Europe from standardized groundfish" took place in 1997-1998. The main objectives of the SESITS project were to standardize the methodology of the bottom trawl surveys among areas and countries and to maintain and standardize the surveys databases. This implied calibration of gears and vessels to estimate catchability conversion factors so indices by species (ANNEX).

Selection of a homogeneous criteria concerning sampling design (depth strata, station grid, haul monitoring, hydrographic stations scheme, etc.) is necessary. The precision of the species abundance indices is dependent of the degree of homogeneity of the strata. It is evident that in the whole of the ICES area there are different biological communities and target species, which implies that a design of strata suitable for one area is not necessarily adequate for another. A first approximation has been made by analysing the data of recent years from Portuguese historical series in an attempt to determine whether the stratification used to date is suitable, or if other strata should be adopted which provide an improvement in the precision of indices. With multi-species surveys, a stratification suitable for one species is not necessarily so for others. This becomes even more complicated when the same species, in the different stages of its life, occupies different niches (juveniles usually

---

1 Assuming independence between n_i: $z=x*y$; $\mathrm{var}(z)=x^2*\mathrm{var}(y)+\mathrm{var}(x)*y^2+\mathrm{var}(x)*\mathrm{var}(y)$ (Goodman, 1960)

live in more superficial waters than adults). One way of dealing with the problem is to apply multi-variant analyses to a matrix of data to determine the degree of similarity between sampling stations and to establish the main groupings or associations between species. Previous works of this type have been reported for the Portuguese waters (Cardador, 1997; ICES, 1997)

To improve the abundance indices estimation, based on the fixed sampling design in use in the Portuguese surveys application of generalized additive methods and generalized linear models and other statistical studies are expected to be further developed.

REFERENCES

Borges, L., Cardador, F., Fernández, A., Gil, J., Moguedet, P., Panterne, P., Poulard, J.C., Sánchez, F., Sánchez, R., Sobrino, I. 1999. Evaluation of Demersal Resources of Southwestern Europe from standardized groundfish surveys. *Final Report Study Contract 96-029, 195 pp.*

Cardador, F., 1983. Contribuição para aumentar a precisão dos indices de abundância obtidos nas campanhas portuguesas de investigação "Tipo Demersal". *Bol.INIP* **9**:17-67.

Cardador, F., Sanchéz, F., Pereiro, F.J., Borges, M.F., Caramelo, A.M., Azevedo, M., Silva A., Pérez, N., Martins, M.M., Olaso, I., Pestana, G., Trujillo, V. and Fernandez, A., 1997. Groundfish surveys in the Atlantic iberian waters (ICES Divisions VIIIc and IXa): history and perspectives. ICES CM 1997/Y:08 30 pp.

Cardador, F., 1997. Portuguese groundfish surveys - is depth stratification adequate?. Working Document to International Bottom Trawl Surveys W.G. 1997, Santander, 4pp.

Cochran, W.G., 1960. Sampling Techniques. John Wiley and Sons, inc., 1st edition

Goodman, 1960. On the exact distribution of variance of products. *J. Am. Stat. Assoc.* **55**, 708-713 [9]

ICES, 1990. Report of the working group on Methods of fish stock assessments. ICES C.M. 1990/Assess:15

ICES,1996a. Report of the Working Group on the assessment of the Southern Shelf Demersal Stocks. ICES CM 1996/Assess:5.

ICES, 1996b. Report of the International Bottom Trawl Survey Working Group. ICES CM 1996/H:1.

ICES, 1997. Report of the International Bottom Trawl Survey Working Group. ICES, CM 1997/H:6, 50 pp.

Table 1 - Type of biological studies carried out by species using information collected from the Portuguese groundfish surveys series.

Biological studies	Species
Distribution	hake, horse-mackerel, mackerel, Spanish mackerel, blue whiting, four-spot-megrim, megrim, black and white monkfish, John dory (*Zeus faber*), rock fish (*Helicolenus dactylopterus*), Pouting (*Trisopterus luscus*) European squid (*Loligo vulgaris*) and Veined squid (*Loligo forbesi*)
Length-weight relationship	hake, horse-mackerel, mackerel, Spanish mackerel, blue whiting, European squid and Veined squid.
Age determination	hake, horse-mackerel, mackerel, Spanish mackerel, blue whiting, four-spot-megrim, European squid and Veined squid
Maturity	hake, horse-mackerel, mackerel, Spanish mackerel, blue whiting, four-spot-megrim, black monkfish, European squid and Veined squid
Fecundity	horse-mackerel
Food habits	hake, horse-mackerel, mackerel, blue whiting, black monkfish, John dory , European squid and Veined squid
Recruitment	hake, horse-mackerel, mackerel, blue whiting, megrim, four-spot-megrim
Parasitism	horse mackerel, hake, mackerel
Communities, diversity	demersal fish and invertebrates
Trawl selectivity	hake, horse mackerel

Table 2 - Groundfish surveys series used in ICES stock assessment Working Groups (in VPA tuning), by species.

Species	Portuguese Summer	Portuguese Autumn	Source
Hake	since 1989	since 1989	WGHMM
Horse-mackerel	since 1985	since 1985	WGMHSA
Mackerel		since 1986	WGMHSA
Four-spot-megrim	since 1989	since 1990	WGHMM
Megrim	since 1989 (*)		WGHMM
Blue whiting		since 1985	WGNPBW
Norway lobster		since 1990 (*)	WGNEPH

(*) used only for cpue trend analysis

Table 3 – Portuguese Surveys: List of Surveys, fishing days and valid by survey.

Year	Late Winter / Spring						Summer						Autumn					
	Hauls by Depth				Valid Hauls	Fishing Days	Hauls by Depth				Valid Hauls	Fishing Days	Hauls by Depth				Valid Hauls	Fishing Days
	20-100 m	101-200 m	201-500 m	501-750 m			20-100 m	101-200 m	201-500 m	501-750 m			20-100 m	101-200 m	201-500 m	501-750 m		
1979							20	24	13		57	12	12	29	14		55	14
1980	14	11	11		36	9	26	23	14		63	16	17	28	17		62	18
1981	22	23	22		67	19	24	23	22		69	18	38	49	25		112	18
1982	23	24	22		69	14	22	25	23		70	17	72	79	39		190	33
1983	22	24	23		69	18	21	24	23		68	17	53	49	21		123	22
1984																		
1985							39	36	26		101	22	79	48	23		150	24
1986							47	45	26		118	17	46	47	24		117	21
1987													28	30	23		81	14
1988													44	31	23		98	17
1989							32	33	27	22	114	25	50	48	32	8	138	23
1990							27	33	24	14	98	22	41	42	23	17	123	35
1991							31	39	26	23	119	30	23	34	22	14	93	30
1992	27	31	19	11	88	27	24	32	16	9	81	22	18	21	12	8	59	16
1993	21	27	17	10	75	23	20	24	12	10	66	21	20	23	14	8	65	20
1994													23	31	22	13	89	25
1995							24	25	22	10	81	21	25	35	19	9	88	24
1996*													25	24	14	7	70	21
1997							23	29	21	13	86	23	17	26	8	7	58	16
1998							24	34	18	11	87	21	18	27	18	11	74	19
1999*							18	27	12	8	65	17	20	33	17	9	79	20
2000							23	31	21	13	88	21	19	29	17	13	78	22
2001							26	33	17	7	83	21	20	24	14		58	14
2002							19	46	28		93	20	21	28	17		66	19
2003*													23	30	17	10	80	21

* - R/V "Capricórnio", trawl net without rollers.

Annex

ADITIONAL RESEARCH ACTIVITIES PERFORMED IN THE PORTUGUESE GROUNDFISH SURVEYS:

- **CALIBRATION EXPERIMENTS BETWEEN GEARS AND**
- **CALIBRATION EXPERIMENTS BETWEEN VESSELS / GEARS**

Introduction

During 1997-1998 a Study project designated as SESITS - Evaluation of demersal resources of Southwestern Europe from standardised groundfish surveys, co-financed by DGXIV/EC (Contract 96/029), involved IEO (Spain), IPIMAR (Portugal) and IFREMER (France).

<u>The main tasks of the project were:</u>

- Co-ordination and standardisation of the methodology of the bottom trawl surveys carried out by the three institutes in Autumn;
- To analyse the data of those surveys.

The area surveyed extends from 52ºN to 36º N in latitude, in depths between 15-750 m, French, Spanish and Portuguese Atlantic waters (ICES Divisions VIIf, j, g, h, VIIIa, b, c and IXa).

The target species considered was hake, blue whiting, horse mackerel, megrims, anglerfishes and Norway lobster for the whole area and also mackerel, Spanish mackerel, red shrimp (*Aristeus antennatus*) and rose shrimp (*Parapenaeus longirostris*) for the Portuguese and Spanish Southern areas.

1. Calibration between gears (one vessel-two gears)

The first task developed by this project comprised calibrations between the gears used by the IEO (Baka gear) and IPIMAR (NCT- Norwegian Campell Trawl) with the GOV gear adopted by the IBTS (International Bottom Trawl Surveys) Working group as the standard gear for the ICES areas (ICES, 1996). The objectives of these experiments were to estimate conversion coefficients in order to estimate standard indices of abundance for the GOV gear and to evaluate the possible adoption of this gear in the groundfish surveys in Iberian Atlantic waters.

In Portuguese waters the area selected was the South shelf (Algarve) based on the homogenous distribution of the target species and on the bottom characteristics appropriate for trawling. These experiments took place in 1997 and 1998.

Methodology

All the hauls were carried out during the daylight period. A haul carried out in one day with one of the gears was repeated in the following day, on the same place and time (or as close as possible) with the other gear. This method was used because the extra time needed to change the gears was not compatible with the ship time available. A total of 21 paired valid half hour hauls and 18 paired valid hour hauls have been performed, respectively in 1997 and 1998.

Scanmar equipment was used with four sensors providing information on horizontal and vertical openings of the net, distance between doors and towing depth.

Statistical analysis

The length distribution by species was checked for polimodal shape to evaluate the appropriate way to test mean differences. Differences between mean catch rates, in number and weight per haul, and mean individual length from GOV and NCT hauls were tested with a significant level of 0.1. Paired-sample t-test was used when the assumption of normality of differences between pairs were fulfilled, according to Kolmogorov-Smirnov test, and the non-parametric Wilcoxon matched pairs test was used when this assumption was not met ($p<0.05$).

Pair hauls showing zero catches for one of the gears were included in the estimation of mean catch rates but excluded when testing for differences between mean lengths. In the case of the pelagic species, blue whiting, horse mackerel and Spanish mackerel, special attention was given to the pair hauls presenting high catch rates for only one of the gear. In these cases, in order to inspect if these high catch rates could represent "school effects" a further mean comparison was performed excluding those hauls from the analysis. Effect size was also determined, taking unto account the correlation between pair hauls, and power analysis carried out to analyse the probability of correctly rejecting a false null hypothesis.

The conversion factors were calculated by the ratio of the numbers or weight between nets when the mean lengths were not statistically different. When the mean lengths were different a second approach was used, the analysis was carried out by length class according to the method described by Warren *et al.* (1997) *in* Borges *et al* (1999). A log-linear model applied to the ratio of numbers caught by each net by length class describes the conversion factor.

The statistical analysis for 1997 suggests that changing from NCT to GOV gear has no effects for hake, horse mackerel, Spanish mackerel and Norway lobster. The experiments in 1998 showed that there were no significant differences between both gears, except in number/hour and weight/hour for hake, mean length for blue whiting and in all three variables for rose shrimp. For these two species the calibration coefficients were determined by the log-linear model (Borges, *et al*, 1999).

2. Calibration between vessels/gears - overlapping experiments with *Noruega* and *Cornide de Saavedra*

A second task of the SESITS project has comprised overlapping experiments with the research vessels and gears in order to estimate conversion coefficients for the target species. These experiments were conducted in 1997 and 1998 in the Portuguese South shelf and in the Southern Spanish shelf, respectively. These experiments took place during the autumn surveys involving the Portuguese RV *Noruega* with NCT gear and the Spanish RV *Cornide de Saavedra* with the baka gear.

A total of 10 and 22 hour hauls have been performed, respectively in 1997 and 1998.

Methodology

The trawling speed was 3.5 knots for *Noruega*, 3.0 knots for *Cornide de Saavedra* which are the usual speeds used in these surveys. The two vessels worked in tandem, with *Cornide de Saavedra* at the front, at geographical coordinates provided by IPIMAR in 1997, with the opposite done in 1998. The tow duration was 60 minutes for both vessels and the hauls were performed during the daylight period.

Statistical analysis

The statistical analysis has involved several stages before the estimation of the conversion coefficients by species.

First, an analysis of variance was carried out for catches in number and weight by species considering the vessel/gear and depth stratum factors. The results indicate difference of catchability between vessels for rose shrimp.

Secondly, to examine if the hauls from the same stratum comprise samples of the same fish population the rates of the length distributions by stratum were compared. A test of normality was carried out by mean of Kolmogorov-Smirnov test to decide which test could be used for the comparisons. If the difference was significant a paired Wilcoxon test was applied and if not a paired t-test was applied. The results indicate for both experiments that both vessels sampled the same populations.

The last stage of the analysis consisted in the comparisons between years and strata.

The conversion coefficients were estimate by means of the quasi-likelihood method. Quasi-likelihood approaches may be considered as a generalisation of likelihood approaches in that they do not require a full specification of the distribution of the observations. However Efron (1982) *in* Borges, *et al*, 1999 shows that according to relatively small sample sizes, one way to increase the robustness of the estimators and to evaluate their bias and variance is to use the bootstrap procedure. The estimator of the conversion coefficient, $\hat{x}$ obtained by a mathematical development of the quasi-likelihood approach when the abundances sampled by both vessels are assumed to have a common constant expectation is:

$$\hat{x} = C_{2T} / C_{1T}$$

where C_{1T} and C_{2T} are the total catch of the two considered vessels/gears.

Results

The coefficients estimated to convert the catches from R/V *Noruega/*NCT gear to R/V *Cornide de Saavedra/* Baka gear were 1 for hake, blue whiting, horse mackerel, mackerel, red shrimp and Norway lobster and 3.1 for rose shrimp.

References

Borges, L., Cardador, F., Fernández, A., Gil, J., Moguedet, P., Panterne, P., Poulard, J.C., Sánchez, F., Sánchez, R. and Sobrino, I. 1999. Evaluation of Demersal Resources of Southwestern Europe from standardized groundfish surveys. *Final Report Study Contract 96-029*, 195 pp.

ICES, 1996. Manual for the IBTS, Revision V, Addendum to ICES CM 1996/H: 1, 58 pp.

Line transect sampling

Line transect sampling (Buckland et al. 2001) is widely used for assessing animal abundance, but has seen relatively little use for fish surveys. It is essentially a plot sampling method, in which the plots are long, narrow strips, and for which it is not necessary to detect all animals on the plot. An observer travels down the centreline of each strip, and records distances of detected animals (or animal clusters, e.g. shoals of fish) from the line. These distances are used to estimate how detectability falls off with distance from the line (assuming certain detection on the line). This in turn allows estimation of the proportion of animals in the strip that are detected, so that abundance in the strip can be estimated. Provided the strips randomly sample the survey region, this allows estimation of abundance in the wider region.

Line transect sampling is widely used for estimating marine mammal abundance, using both ships and aircraft (e.g. Buckland et al. 2001, 2004). Marine mammals must surface to breathe, making them more suitable than most fish populations for sightings survey methods. Ship or aircraft sightings surveys of fish are thus rare. One example is the aerial surveys of bluefin tuna off the south coast of Australia (Chen and Cowling 2001). However, line transect surveys of fish populations are becoming more widespread, using submersibles (O'Connell and Carlile 1993) or SCUBA divers (Letourneur et al. 2000).

The advantages of line transect sampling are that they provide fishery-independent absolute estimates of abundance; it is easy to verify whether assumptions hold; and there are only three assumptions that matter. The first of these is that all fish on or near the transect line are detected; the second is that the fish move slow relative to the speed of the observation platform (around half the speed on average or less); and the third is that distances from the line are accurately measured.

There are several problems associated with line transect surveys for fish. The first is that fish move in a three-dimensional environment, whereas standard distance sampling methods are designed for two dimensions (though theory exists for three-dimensional surveys); sightings surveys from ships or aircraft are unworkable for most fish species; many fish occur in large schools, and even if these schools are treated as the sampling unit, it can be difficult to define the limits of the school, record the distance of its centre of mass from the line, or estimate its size; fish may move away from the observation platform before they are detected, leading to underestimation of abundance; accurate estimation of distance is problematic underwater (though not insoluble); and detection on the centreline may not be certain if a poor observation platform is used, or for species that can hide.

Possible uses of line transect sampling in the North Sea fisheries include the following.

Monkfish are an important part of the demersal fisheries, yet no stock assessments are conducted. Line transect surveys should be perfectly feasible for monkfish, provided that a small submersible is available as the observation platform. Alternatively, a towed video unit might be used as the observation platform.

Nephrops burrows are surveyed by towing a 'sledge', with a camera mounted on it. These are currently analysed as strip transects – line transect surveys in which everything within a specified distance of the line is counted. Line transect sampling, possibly in conjunction with revised video equipment, might allow a wider strip to be surveyed, and hence might yield great precision.

Pelagic stocks are surveyed using acoustic methods. It may be possible to integrate acoustic surveys with line transect methods to improve the acoustic stock size estimates.

Demersal stocks such as cod and haddock, like monkfish, could be surveyed using a submersible or towed video.

References

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. and Thomas, L. (2001) Introduction to distance sampling. Oxford University Press, Oxford.
Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. and Thomas, L. eds (2004) Advanced distance sampling. Oxford University Press, Oxford.
Chen, S.X. and Cowling, A. (2001) Measurement errors in line transect surveys where detectability varies with distance and size. Biometrics 57, 732-742.
O'Connell, V.M. and Carlile, D.W. (1993) Habitat-specific density of adult yelloweye rockfish *Sebastes ruberrimus* in the eastern Gulf of Alaska. Fishery Bulletin 91, 304-309.

Letourneur, Y., Kulbicki, M. and Labrosse, P. (2000) Fish stock assessment of the northern New Caledonian lagoons: 1 - Structure and stocks of coral reef fish communities. Aquatic Living Resources 13, 65-76.

Working Paper for ICES Study Group on Survey Design and Analysis, Aberdeen, June 2004

CAN FISHERS TEACH SCIENTISTS HOW TO IMPROVE FISH SURVEYS? SELECTED RESULTS FROM SPATIALLY INTENSE, COMMERCIAL FV SURVEYS OF NINE ENGLISH FISHERIES IN 2003-4.

John Cotter
Centre for Environment, Fisheries and Aquaculture Science,
Lowestoft, England NR33 0HT
email: *a.j.cotter@cefas.co.uk*

Summary
Selected results from 9 surveys undertaken by English commercial fishing vessels (FVs) with a trained scientist on board to identify and measure catch are reported. The surveys covered trawl fisheries for saithe in the northern North Sea, and for cod off the NE of England and in the Irish and Celtic Sea, beam trawl fisheries for flatfish and monk off the SW of England, and a trawl fishery for deep sea species in the Atlantic west of Scotland. Fishing areas and gears used were nominated by the industry. The most striking result was the spatial coherence of catch rates. Variously located, small-scale aggregations of fish were found for most commercial species. Comparisons between FV catches and those by CEFAS research vessels (RVs) in the same season indicated reasonable agreement between geographically close stations after standardisation of effort. Spatial coherence of catch rates was more evident for these surveys by FVs than for surveys by RVs because of the closeness of the fishing stations. Spatially coherent variations in the length compositions of the catches were also observed. Results are presented that show the importance of matching trawl gear to ground type. They raise the possibility that the fish distributions illustrated may have resulted from the gear used as well as from the local environment. Implications of the results for survey design are discussed. It is suggested that FVs could be used more in surveys, either to generate additional, regional tuning (CPUE) series, or to locate aggregations of fish, knowledge of which could be used to improve the efficiency of the RV surveys.

Introduction

English commercial fishers have often asserted that commercial fishing with fishing vessels (FVs) can give a more favourable impression of the status of a fish stock than surveys with research vessels (RVs). In 2003, a 'Fisheries-Science Partnership' (FSP) was formed between the National Federation of Fishermen's Organisations (NFFO) and the Centre for Environment, Fisheries and Aquaculture Science (CEFAS) in order to conduct surveys in a variety of fisheries using commercial FVs with a CEFAS scientist on board to record and measure the catches. Eleven of these FSP surveys were conducted between autumn 2003 and spring 2004. One particularly valuable feature of them was that fishing stations were mostly much closer together than is possible on RV surveys which, typically, are constrained to cover large areas in limited periods of time. The FSP surveys therefore provided unusually detailed snapshots of the spatial variations in abundance and length composition of certain commercial fish species within these English fisheries.

The present working paper presents selected results from nine of the FSP surveys that appear relevant to the design of fish surveys. [The two omitted surveys had more relevance to fishery management than to survey design.] In some cases, results from CEFAS RV surveys conducted in the same localities and seasons are also presented for comparison. The discussion at the end of the paper puts forward implications of the results for design and analysis of fish surveys.

Methods

Priority areas for fishing were nominated by the NFFO based on national and regional fisheries management and assessment interests. Fishing vessels were chartered to fish commercially in order to obtain new data on the catch rates and size distributions of commercial species. Subsidiary objectives were sometimes also included. A survey grid and work plan were developed for each survey between NFFO, CEFAS, and the vessel skippers. Details of the surveys, fishing vessels, fishing methods, gear, etc. are given for each survey # in table 1. Locations of the survey areas are shown in fig. 1.

Results

The full set of FSP survey results is extensive and was reported to NFFO and to the funding body, the UK Dept. of Environment, Food and Rural Affairs (DEFRA). This presentation is only a subset.

Spatial variability of catch rates
The closeness of most of the fishing stations clearly revealed spatial coherence of fish catch rates. Fig. 2a-h shows examples for a variety of commercial species from eight of the FSP surveys.

Saithe, fig. 2a, survey #2, were repeatedly found in number using an otter trawl in a small locality east of Shetland (at 2 deg E, 61 deg N). There appears to have been a large shoal there. The highest catch rate was 6300 saithe.h^{-1}. Elsewhere, catch rates were mostly much lower.

Cod, fig. 2b, survey #3, were found aggregated in a band off the English NE coast using otter trawls rigged for hard ground. Both FVs found this even though their fishing trips were about two weeks apart, see table 1, implying stability of the aggregations over that period at least. The observer on this trip, R. Mainprize, noted that "*the better hauls of fish were found . . . on the harder/rocky ground and in amongst the herring marks where the herring were heavy with spawn.*" The highest catch rate was 450 cod.h^{-1}. Also shown in fig. 2b (as solid triangles) are the catch rates of the English groundfish survey using RV CEFAS Endeavour in August 2003.

There were five stations located in the offshore part of the area. None was located in the aggregations of cod. Catch rates were low at four of the stations and zero at one but these results were consistent with the low catch rates obtained by the FVs at stations nearby and outside the main aggregation.

Monk (*Lophius piscatorius*), fig. 2c, surveys #4 and 5, were found by four beam trawlers fishing in the W Channel and Celtic Sea to be relatively aggregated to the NW of Cornwall, around the Scilly Isles, and in deeper water in the central, western Channel. The rates shown are standardised per metre of beam length (M^{-1}) since the different vessels fished different sizes and numbers of beam trawls. The highest catch rate was 5.5 monk.$M^{-1}.h^{-1}$. All of the catches shown in fig. 2c were obtained with trawls fitted with chain mats but otherwise unstandardised so there may have been differences in the fishing power of the different vessels, particularly as two were targetting sole (FSP survey #4) and the other two were targetting monk (FSP survey #5). Nevertheless, a reasonably coherent picture of the geographic distribution of monk appears to be presented by combining results of the two surveys as in fig. 2c. Also shown in fig. 2c (as hollow triangles, base down) are the catch rates of the CEFAS beam trawl survey using RV Corystes in autumn 2003. A 4-metre beam fitted with chain mat and a fine mesh codend liner (40mm mesh) was used. Catch rates appear reasonably comparable to those of the FVs fishing close by, except that, sometimes, Corystes appeared to outfish the FVs, e.g. off Cornish N coast, possibly due to the small mesh net in use.

Fig. 2d shows catch rates of sole also from surveys #4 and 5. The highest rates were obtained by Corystes in the Bristol Channel although this partly reflects large numbers of small sole caught by the 40 mm mesh in a well-known nursery area. Up to 34.5 sole.$M^{-1}.h^{-1}$ were taken there. Otherwise, the higher catch rates of sole were taken in grounds south of Devon around 3.5 deg W, 50.2 deg N. The comments made for monk about unstandardised gear and differing target species also apply to sole.

Fig. 2e, survey #7, shows an aggregation of whiting in the central N Irish Sea, where catch rates with an otter trawler reached 591 whiting.h^{-1}. Catch rates can be seen to be uniformly lower in another region a few miles to the east. Fig. 2f, survey #8, shows a similar, localised aggregation of haddock off the Cornish north coast, and sparsity of fish around about. These catches were taken with one net of a twin rig trawler. The two nets were fitted with slightly different ground gear, and the same side was not used consistently so this lack of standardisation may have caused some of the diversity of catch rates in fig. 2f.

Fig. 2g, survey #10, shows catch rates for *Nephrops* in the Farn Deeps ground off NE England using two different sized trawlers. This survey included parallel towing at 10 stations to permit the relative catching powers of the two vessels to be estimated. Catch rates shown in fig. 2g for the larger trawler, FV1, were divided by the average factor, value 1.72, estimated as shown in table 2. The highest catch rates of Nephrops were found in a band in the offshore part of the survey area, bending out from the coast further northwards. CEFAS undertook similar surveys using the same survey grid in 1994, 1995, 1996, 1998, and 1999 but none was in the same month as this FSP survey. The survey results, not illustrated, exhibited different patches to those shown in fig. 2g, and no attempt has been made to clarify whether the patchy distributions are stable features dependent, for example, on sediment characteristics, or whether they reflect between-year differences in the pattern of settlement or harvesting of successive yerar-classes, or varying catchability caused by well-known tidal and seasonal effects. Time of day can also affect catch rates for *Nephrops* but probably was not a factor in FSP survey #10, see fig. 3. Both the FSP and the CEFAS surveys restricted fishing to the first part of the day. Maximum catch rates also differed between the FSP and CEFAS surveys. The CEFAS surveys in November and December repeatedly found catch rates of 20 000 *Nephrops*.h^{-1}, whereas the highest catch rate on this FSP survey in March by FV1, unstandardised, was 9 000 Nephrops.h^{-1}. The Nephrops season in the Farn Deeps is generally considered to end in spring so this low result may simply reflect the end of the season.

Fig. 2h, survey #11, shows catch rates for cod in the central and western Irish Sea using a semi-pelagic trawl fitted with a 100 mm codend. As for whiting and haddock in the Irish Sea, figs. 2e, f, the best catch rates were highly localised. Up to 28 cod.h^{-1} were taken. Haddock were taken at much higher rates on this survey, up to 14 000 haddock.h^{-1} but in a different locality to cod, off Dublin (not illustrated).

Spatial variability of length compositions

Samples of the catches of selected commercial species were measured on all of the FSP surveys and length frequency distributions (LFDs) prepared. This section presents examples that are instructive about spatial variability of length.

LFDs for North Sea cod from two surveys, #2 (targetting saithe) in the north with a 110 mm mesh codend, and #3 (targetting cod) from the NE coast of England with 80 mm, are compared in fig. 4. Note the different catch rate ordinates. Whereas the bulk of the northern catch was made up of fish over 50 cm, the bulk of the catch off the NE coast was made up of fish under 50 cm. The different target species and mesh sizes used for the two surveys does not satisfactorily explain this result since one would expect large cod to have been taken with 80 mm had they been present off the NE coast. Also, when 120 cm mesh was fished off the NE coast in a restricted area (results not illustrated), approximately 3 cod.h^{-1} were taken at the minimum landing size, 35 cm, by both vessels. Therefore 110 mm should have caught fish less than 50 cm in the northern North Sea had they been present there. This result illustrates the heterogeneous distribution of cod by size around the North Sea, the NE coast being an example of an adolescent area supporting a recruitment fishery. A comparable size fractionation of cod by geographic area was also found in the Irish and eastern Celtic Sea when LFDs from surveys #7, #8, and #11 were compared. See fig. 5. Codend mesh sizes were 80, 85, and 100 mm respectively (table 1) but these differences do not satisfactorily explain the relative lack of 60 to 70 cm fish in survey #8 in the eastern Celtic Sea.

Geographic variation of LFDs were found for several other species including those from the deep sea obtained on survey #9 from three widely separated areas of the Atlantic west of Scotland. Variation of the shapes of the distributions (not illustrated) between sites was considerable.

Another way to look at changing size of fish with location is given in fig. 6a for monk (*Lophius piscatorius*) off the SW of England. Here, mean length of fish in each catch is plotted geographically. A constant was subtracted from each result to accentuate the differences between the scaled symbols. Spatial coherence of mean length can be seen in several places, e.g. for small fish around the Scilly Isles (~ -6.3 W, 49.8 N), NW of Cornwall (~ -5.2 W, 50.8 N), and in the central Channel (~ -3.3 W, 49.7 N); and for larger fish south of Devon (~ -3.1 W, 50.2 N) and in the SW approaches (~ -7.3 W, 48.7 N). Fig. 6b shows variation of mean length for plaice from these surveys for comparison with monk. The largest fish were found in the SW approaches (~ -7.0 W, 48.7 N).

Effects of trawling technique and gear variations
The examples of geographic variation of catch rates and length shown in figs. 2 and 6 suggest matching variations of abundance and size distribution in the populations. However, there may also be an element of illusion due to geographic variations of trawling methods and gear during the surveys. The common scientific view is that controllable factors should be held constant in surveys to improve the comparability of results across times and locations. A different view held by several fishers is that certain factors should be deliberately varied so as to reveal what can be caught in different circumstances. A few examples of what was revealed by this different approach follow.

A frequently heard statement is that longer towing times yield bigger fish in the catch (because big fish can maintain a position in front of a trawl for a longer time before tiring). Several of the surveys used widely varying towing times. Plots of mean length in catch against towing time for saithe, cod and haddock (survey #2) are shown in fig. 7a, b. Towing times varied between 30 and 400 minutes. Surprisingly, no clear relationship is seen for saithe which is a powerful swimming fish, and no effect was seen for haddock which is not so surprising, but a significant linear increase of mean length with towing time was found for cod. This last result is only suggestive of an effect since there may have been a correlation between the choice of towing time and the likelihood of large fish being present especially as acoustic studies of the shoals of fish were being made during this survey. Of possible interest is another finding: mean length of saithe increased with depth. See fig. 8.

The catch of cod with an otter trawl in coastal waters at least is likely to depend on whether the ground gear is rigged for hard or soft ground. During FSP survey #3, R. Mainprize noted:

> *The skipper did express concern, that when towing between hauls 17 to 33 he did not expect to catch a lot of cod, as the ground we were working was soft sandy bottom. Our gear was rigged for hard bottom and would not be effective on this type of seabed. This was confirmed when we finished all the sub areas in the soft ground as there was a definite reduction in cod catches relative to fishing on the rocky bottom. Also fishing vessels of the same size and horsepower, fishing alongside us working the correct fishing gear for this type of bottom, were reporting two or three times as much fish including two to three times as much cod as we were catching.*

The importance of using the right gear for hard and soft ground was demonstrated for beam trawls on FSP survey #4. One of the participating FVs rigged a chain mat beam trawl on one side, and a V-trawl (without the chain mat but with more tickler chains) of the same beam length on the other. The chain mat serves to keep rocks out of the catch making it suitable for hard ground. The V-trawl can only be used on soft ground but is thought to be more efficient there. Codend mesh was 80 mm in both cases. Sixteen tows with the two gears side by side were made on soft ground known to yield good catches of sole. Table 3 lists the catch per hour for each haul and trawl and shows that the V-trawl substantially outfished the chain-mat trawl for flatfish on this type of ground.

That RV Corystes outfished the commercial FVs on some occasions, e.g. see fig. 2c, d, may also be a result of trawling technique. Corystes generally fished for half an hour whereas the FVs mostly fished for an hour or more. A short tow that hits an abundance of fish will show a high catch rate, while a long tow will probably also hit areas of low abundance so that overall catch rate is lower. Short tows can therefore be expected to be more variable when fish are present.

Discussion

These FSP surveys have served to demonstrate (probably not for the first time) what fishers have known for years: that several species of commercial fish tend to be found in localised aggregations on the sea floor. It is also instructive to see that the geographic variations of abundance can be quite clearly discerned despite haul to haul variance. Possibly, some scientists, like the author, have held an impression that haul to haul variance is larger than it actually is because they are conditioned to looking at RV survey catches taken at stations which, although neighbouring within the survey, are 30 to 50 or more miles apart and typically show radically different catch compositions and sizes. Such pairs of stations, it appears, are likely to be sited in substantially different habitats and fish communities, and the differences between them may therefore depend more on spatial variation than on haul to haul uncertainty. Some RV surveys take replicate tows at the same station and still find high haul to haul variance. This may be an entirely different effect due to disturbance.

An important question now is whether the observed aggregations of fish have any permanence or periodicity. One might expect aggregations related to sea floor conditions to be permanent or seasonally periodic, and aggregations related to hydrographic or planktonic conditions to be more flexibly located over space and time. Illusory aggregations related to conditions of fishing, as may have been found on the Nephrops survey (#10), unfortunately complicate this simple picture. A survey that successfully targets real aggregations is likely to be more efficient (in terms of precision of mean CPUE per unit cost) than one that consistently takes low or zero catches. On the other hand, searching for aggregations, as in adaptive sampling, would detract from efficiency.

A further problem associated with patchy abundance concerns whether any of the patchiness is an artifact of the gear being used. The results with two types of beam trawl, table 3, and the anecdotal evidence quoted above about otter trawls both indicate that one gear design can outfish another on the appropriate ground by factors around three times. A survey that covers an area of mixed ground types with one type of net is vulnerable to variance arising from fish moving from one ground type to the other, possibly as a response to changing population size. Also, aggregations of fish on the poorly fished ground type could be missed altogether.

Use of both hard and soft ground gear on a survey could ease these problems but creates logistical problems on a single survey vessel associated with changing the trawls over, especially if this is done frequently as is desirable for good experimental control of external factors such as weather. Use of two gear types would be more feasible when more than one vessel is available, as in some of the co-ordinated European surveys. The two survey series resulting might either be treated as separate tuning indices, or some kind of intercalibration might be attempted to prepare a joint index. The best course would depend on the eventual utility of the two series and on whether the trends diverged or not.

Small-scale, patchy variation of fish abundances, assuming that they really exist and are not just a product of fishing technique, poses major problems for surveys which have to cover large areas with few stations. The 2-dimensional grid favoured for many surveys may serve well when stock numbers are high and fish are widely distributed but, when stocks are low and have contracted to localised aggregations, are vulnerable to high variance of mean CPUE depending on whether a small number of stations fall on aggregations, or just miss them. Stratified designs could be more efficient provided they are based on reliable knowledge of where the fish will be found. If not, efficiency could plummet. Unfortunately, optimising a survey design for one species could seriously impair efficiency for any others being targetted, and single objective fish surveys are a rare extravagance in these cost-conscious times.

The reported FSP surveys have shown that LFDs are patchily distributed as well as abundances. This suggests that age distributions would also be. [Otoliths were collected on the surveys but have not all been read.] ICES considers that North Sea cod belong to one stock for management purposes, yet the LFDs in fig. 4 were strikingly different even though gear-related and seasonal effects were not thought to have been responsible. LFDs for Irish Sea cod, surveys #7 and 11, appeared more similar than LFDs for cod in the E Celtic Sea and Bristol Channel, survey #8. See fig. 5. In this case, the distinction made by ICES between northern (VIIa) and southern (VIIf and g) stocks of cod in the Irish/Celtic Sea is supported. Geographic variation of length, age, and probably length-at-age implies that LFDs and age-length keys derived from surveys should be prepared for the smallest possible areas.

Tow duration could also be affecting length and age-frequency distributions. Weak evidence for this effect in cod is seen in fig. 7. Better evidence for this could be obtained with parallel trawling trials in which one vessel makes several short tows alongside long tows by the other in a region where large fish are known to be present. Since a lack of large fish is sometimes taken as a sign of over-fishing without other analysis, and since many survey scientists have no information about the effects of varying tow length on length frequency distributions because tow length is deliberately held constant at all stations, research on the effects of varying tow lengths on catches of species that swim well could improve the credibility of RV surveys.

These FV surveys were undertaken on commercial FVs. Should they have a continuing role in scientific surveys of fish stocks? FVs have been used successfully for surveys in Iceland, Canada, and in England previously (for VIIe sole), and indications from the results presented here are that their catch rates are similar to those of CEFAS RVs when different towing times and beam lengths have been allowed for. FVs tend to be substantially cheaper to operate than RVs but have limited accommodation and facilities, and may be limited in range. FVs are often modified for their primary commercial role over the years, and this could generate intercalibration problems. Furthermore, even minor aspects of fishing can be quite hard to standardise on FVs, as was found in the limited time available to organise these FSP surveys.

ICES and several sectors of the European fishing industry are at present actively seeking more involvement of fishers in the stock assessment process (See reports of the ICES/NSCFP Study Group on the Incorporation of Additional Information from the Fishing Industry into Fish Stock Assessment(SGFI), 2003, 2004). One way to achieve this would be to use fishers to survey their own fishery with the help of a scientist on board, as on the FSP surveys. The results would then form a supplementary index that could either be kept as a separate, regional tuning series, or intercalibrated with other RV series. Another way would be to use fishers to locate aggregations of fish that could be assigned special strata for intensive surveying by the RV survey in order to boost overall efficiency of the survey, as discussed above. Fishers are particularly likely to welcome the second idea which fits in well with their normal professional approach. Also, long-term standardisation of FVs is much less of a critical issue for searches than for generating tuning series.

Acknowledgements

The FSP surveys in 2003/4 were made possible by a grant from the UK Department of Environment, Food and Rural Affairs (DEFRA). CEFAS is grateful for the good co-operation shown by the NFFO, skippers, and owners. Thanks are also due to C. Bannister and T. Boon who managed and organised the surveys with NFFO, and to the numerous CEFAS staff who took part and advised the author.

Table 1 FSP and CEFAS surveys: details of fishing activities.
FV = commercial fishing vessel; RV = research vessel.

Id. #	Locality	1^o Target species	Gear	Codend mesh (mm)	Vessels (length m)	Survey dates	# of stations	Towing times (h)
2	N North Sea	Saithe	Otter trawl	110	FV (36)	18.9.03 - 5.10.03	56	0.5 - 7.5
3	English NE coast	Cod	Otter trawl	80 and 120	FV1 (16) FV2 (16) RV (73)	26.8.03 - 19.9.03 22.9.03 - 31.10.03 14.8.03 - 21.8.03	66 68 5	1.3 - 5 4 - 5.3 0.5
4	W Channel	Sole	Beam trawl	80	FV1 (29) FV2 (23) RV (52)	18.8.03 - 30.8.03 25.9.03 - 8.10.03 24.9.03 - 3.10.03	111 71 81	0.25 - 1.2 0.4 - 2.2 0.5
5	W Channel + Celtic Sea	Monk	Beam trawl	80	FV1 (27) FV2 (27) RV(52)	23.9.03 - 9.10.03 4.11.03 - 21.11.03 6.9.03 - 2.10.03	106 105 63	0.25 - 3 1.2 - 3 0.5
7	NE Irish Sea	Cod	Otter trawl	80	FV (21)	10.2.04 - 7.3.04	51	0.7 - 7.5
8	E Celtic Sea & Bristol Channel	Cod	Twin rig trawl	85	FV (14)	13.2.04 - 1.4.04	77	0.7 - 4.7
9	Deep water	Blue ling	Otter trawl		FV (40)	28.2.04 - 12.3.04	37	1 - 7.3
10	Farn Deeps (English NE coast)	Nephrops	Nephrops trawl	20	FV1 (17) FV2 (9)	1.3.04 - 9.3.04 1.3.04 - 9.3.04	35 34	0.5 0.5
11	W Irish Sea	Cod	Semi-pelagic trawl	100	FV (20)	16.2.04 - 9.3.04	42	0.8 - 12.5

Table 2. FSP survey #10: Farn Deeps Nephrops, March 2004: details of parallel trawling trials.

Parallel trawl #	FV1: Nephrops.h^{-1}	FV2: Nephrops.h^{-1}	FV1/FV2
1	616	374	1.65
2	158	110	1.43
3	10	14	0.71
4	822	596	1.38
5	320	106	3.02
6	338	152	2.22
7	238	254	0.94
8	3192	1375	2.32
9	980	637	1.54
10	344	172	2.00
Average +/- st. err.			1.72 +/- 0.22

Table 3. FSP survey #4: Comparison of numbers caught per hour during 16 tows with two types of trawl rigged on different sides of one commercial fishing vessel.

	Port: 12m BT+chain mat					Stbd: 12m BT (V trawl)			
Haul	Plaice	Lemon sole	Monk	Sole		Plaice	Lemon sole	Monk	Sole
96	8.00	0.00	0.00	13.33		18.67	0.00	0.00	26.67
97	4.00	0.00	0.00	13.33		10.67	0.00	0.00	45.33
98	0.00	0.00	0.00	13.33		2.67	0.00	0.00	34.67
99	1.33	0.00	0.00	4.00		2.67	0.00	0.00	38.67
100	8.00	0.00	0.00	10.67		21.33	0.00	0.00	41.33
101	29.33	0.00	0.00	36.00		46.67	0.00	0.00	117.33
102	21.33	0.00	0.00	17.33		34.67	1.33	0.00	97.33
103	7.83	1.30	0.00	39.13		7.83	1.30	0.00	56.09
104	32.00	5.33	5.33	18.67		50.67	12.00	1.33	46.67
105	54.67	1.33	0.00	26.67		104.00	13.33	4.00	56.00
106	45.00	11.67	5.00	8.33		116.67	1.67	1.67	95.00
107	16.00	0.00	1.33	10.67		21.33	0.00	0.00	66.67
108	14.63	0.00	0.00	13.17		67.32	2.93	0.00	61.46
109	36.00	0.00	1.33	28.00		77.33	0.00	1.33	108.00
110	22.98	0.00	0.00	25.53		49.79	0.00	0.00	114.89
111	17.33	0.00	0.00	36.00		37.33	1.33	0.00	133.33
Avge	19.90	1.23	0.81	19.64		41.85	2.12	0.52	71.22

Fig. 1 FSP surveys: map of the British Isles showing the locations of panels in fig. 2. The survey numbers (#) refer to rows in table 1.

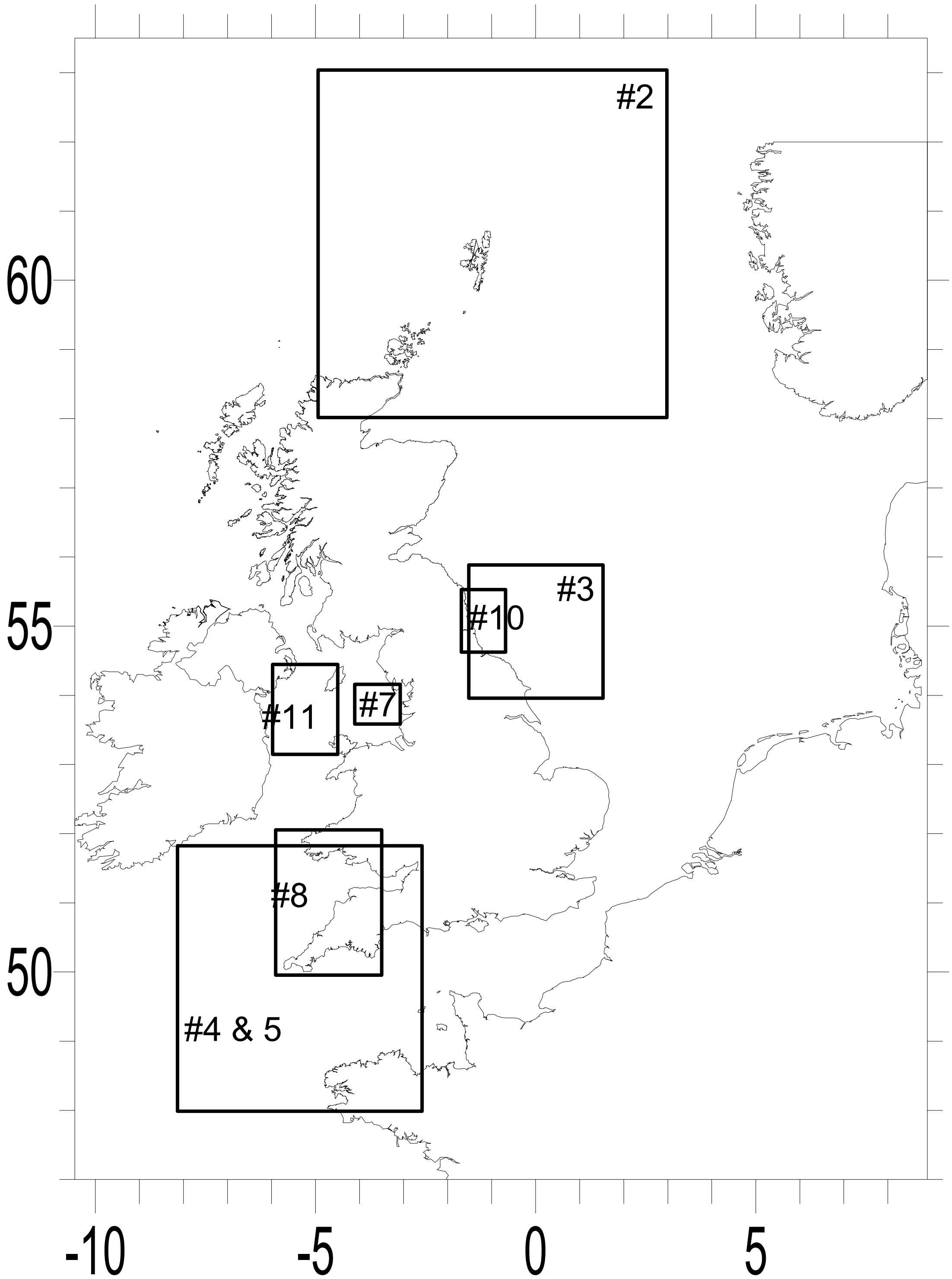
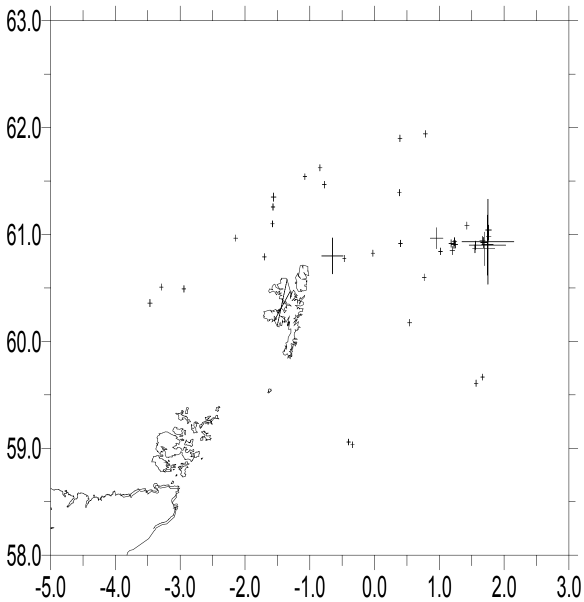
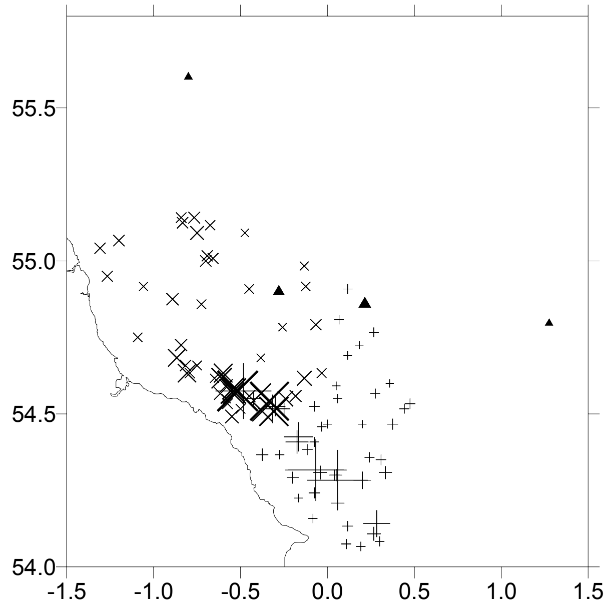


Fig. 2 FSP surveys: Examples of spatial coherence of catch rates. Symbols scaled linearly to catch.h^{-1} (a, b, e, h), catch.M^{-1}.h^{-1} for beam trawlers (M = unit beam length) (c, d), catch.h^{-1} per net for the twin rig trawler (f), and catch.h^{-1} standardised to the smaller FV for Nephrops (g). Zero catches are shown as dots; largest catch rate given as Msh= maximum symbol height. Fishing details for each survey # shown in table 1. Panel locations in fig. 1. FV= fishing vessel; RV=research vessel.

a) N North Sea (#2), saithe;
Msh = 6300 fish.h^{-1}
x = FV



b) NE coast (#3), cod;
Msh = 450 fish.h^{-1};
x = FV1, + = FV2, ▲ = RV.



c) W Channel & Celtic Sea (#4 & 5), monk;
Msh = 5.5 fish.M^{-1}.h^{-1};
x = FV1, + = FV2, ✂ = FV3, ▤ = FV4, Δ = RV

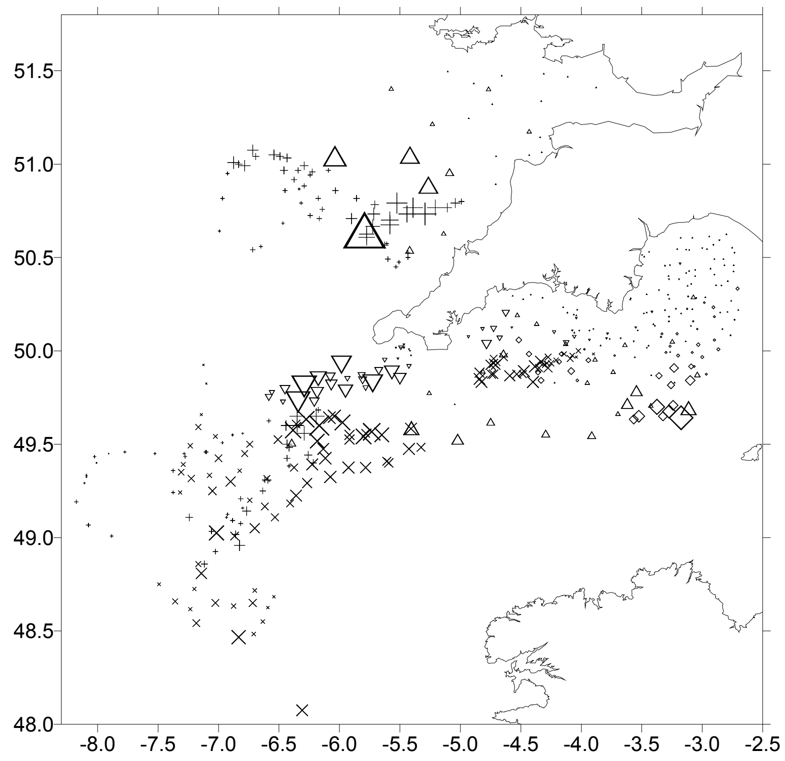
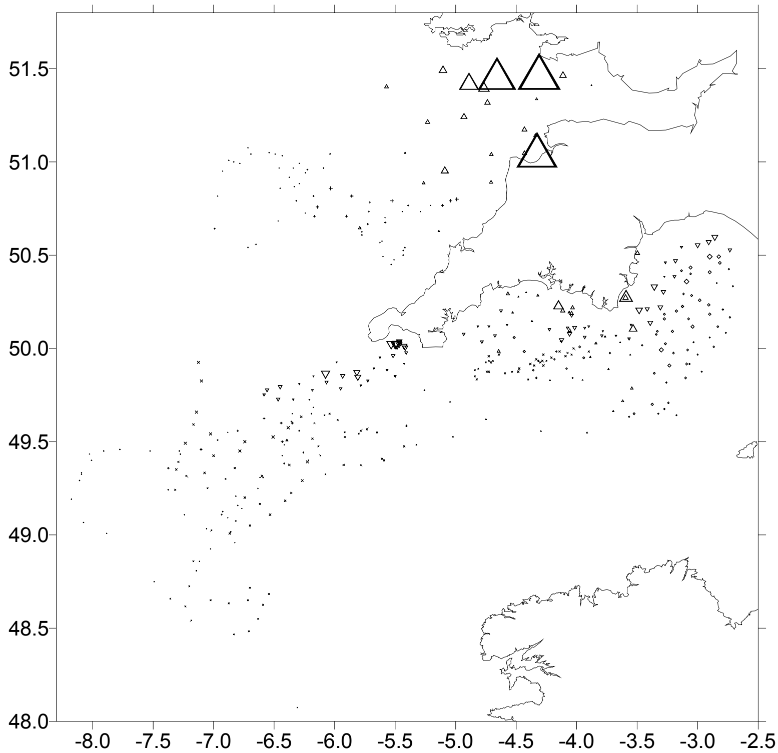
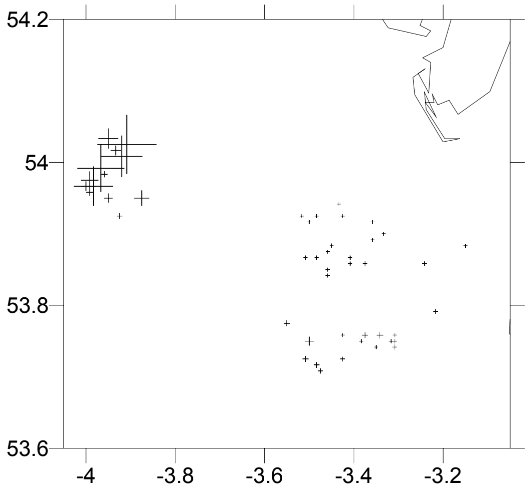


Fig. 2 cont.
d) W Channel & Celtic Sea (#4 & 5), sole;
Msh = 34.5 fish.$M^{-1}.h^{-1}$;
x = FV1, + = FV2, ✂ = FV3, ▤ = FV4, Δ = RV



e) NE Irish Sea (#7), whiting;
Msh = 591 fish.h^{-1}
+ FV



f) E Celtic Sea & Bristol Channel (#8), haddock;
Msh = 370 fish.h^{-1} per net
+ = FV

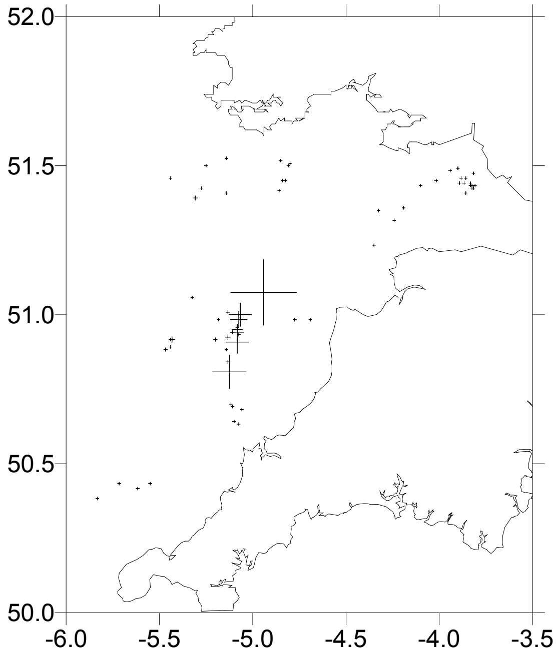
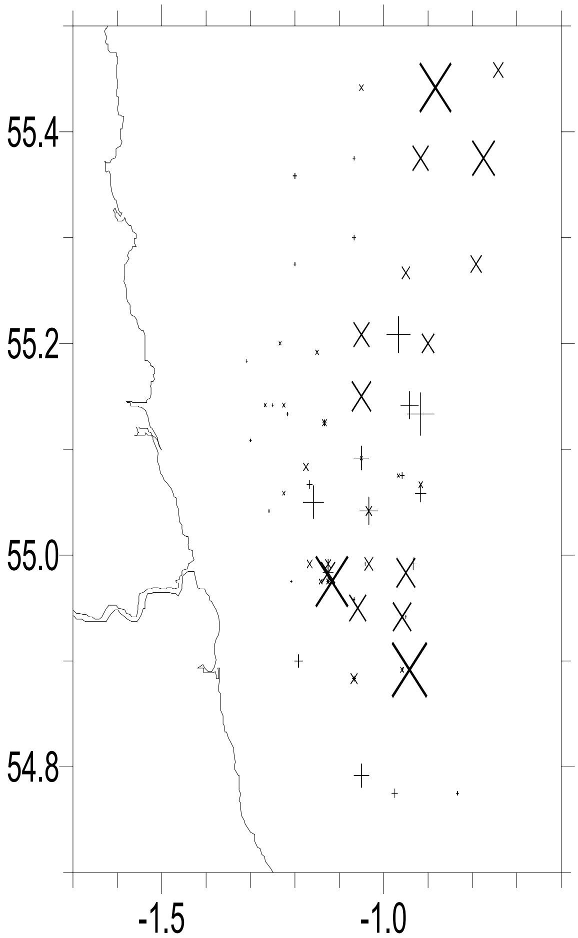


Fig. 2 cont.

g) Farn Deeps (#10), Nephrops;
Msh = 5228 fish.h^{-1} (standardised)
x = FV1, + = FV2, ✂ = FV3, ▤ = FV4, Δ = RV



h) W Irish Sea (#11), cod;
Msh = 28 fish.h^{-1}
x = FV

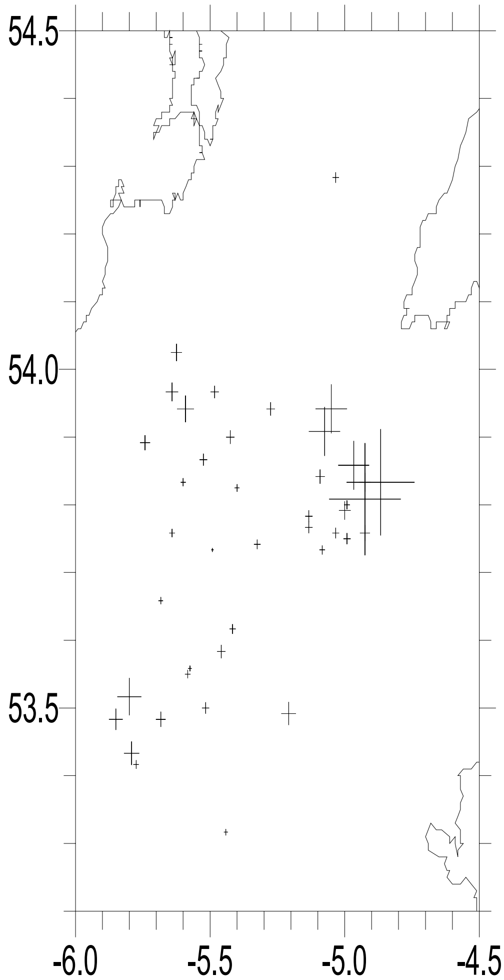


Fig. 3 FSP Survey #10: Farn Deeps Nephrops, spring 2004: Nephrops.h^{-1} (standardised to catching power of FV2) against time of day.

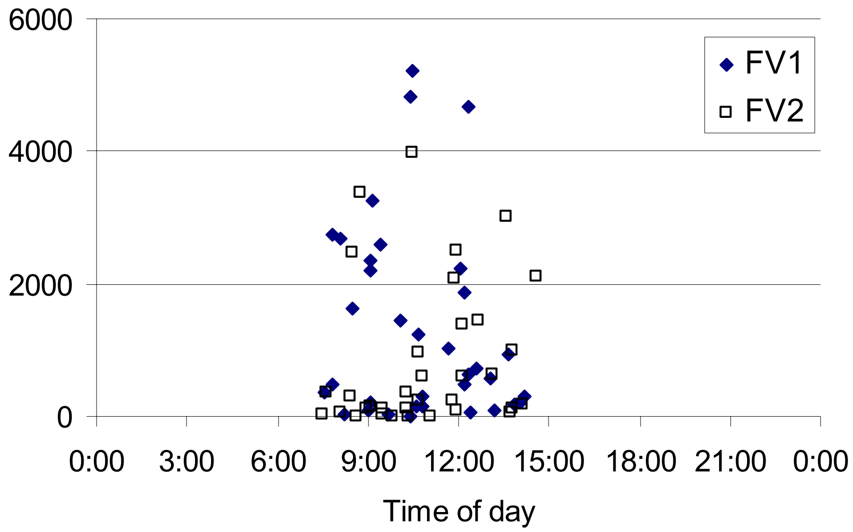


Fig. 4 FSP surveys #2 and 3: North Sea cod, autumn 2003: length frequency distributions as obtained by three commercial fishing vessels (FV, FV1, FV2). Note different ordinate axes.

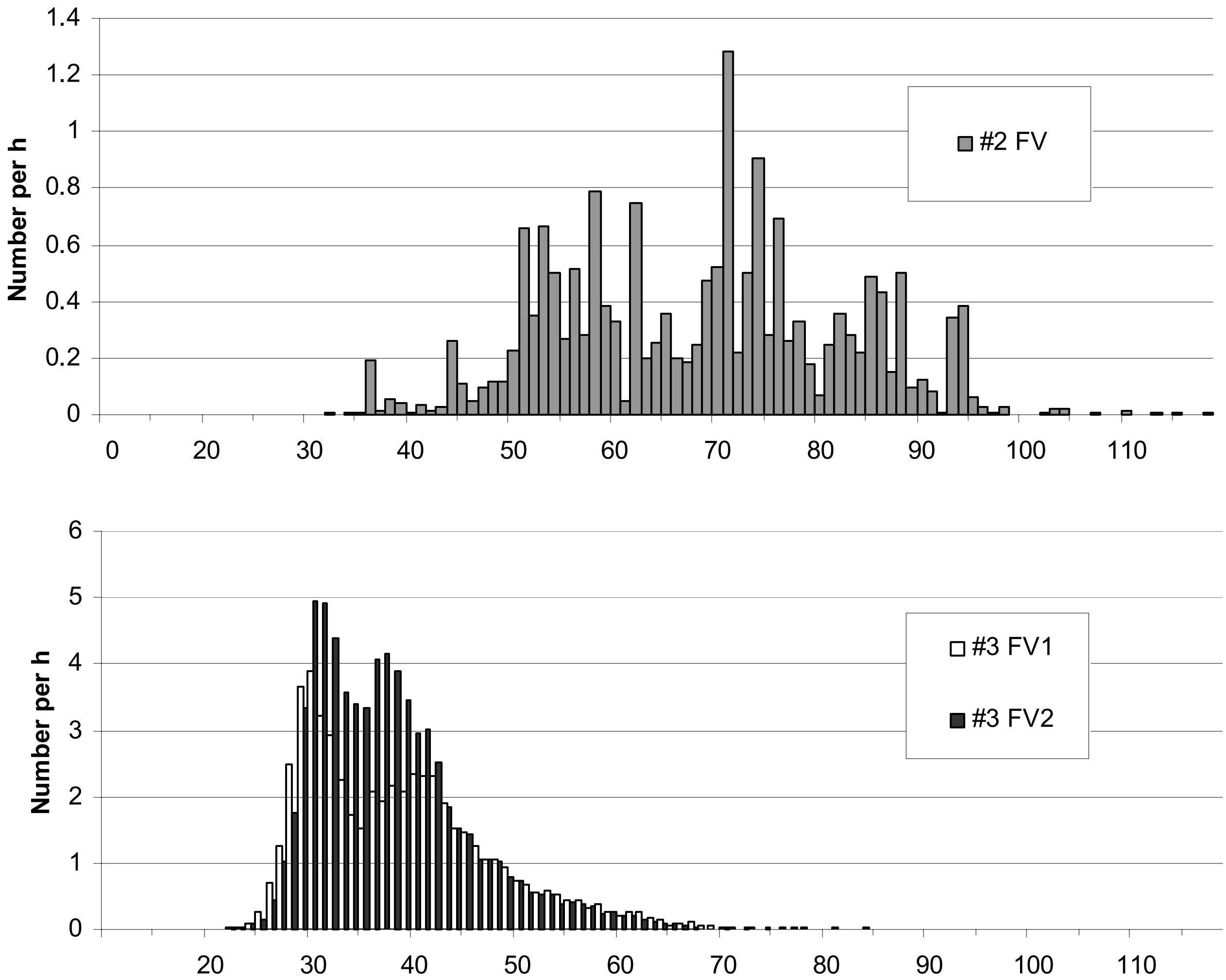


Fig. 5 FSP surveys #7, 8 and 11: Irish and Celtic Sea cod, spring 2004: length frequency distributions as obtained by three different commercial fishing vessels (FV). #7.1 and #7.2 correspond to E and W parts respectively of FSP survey #7, as illustrated for whiting in fig. 2e. Note different ordinate axes.

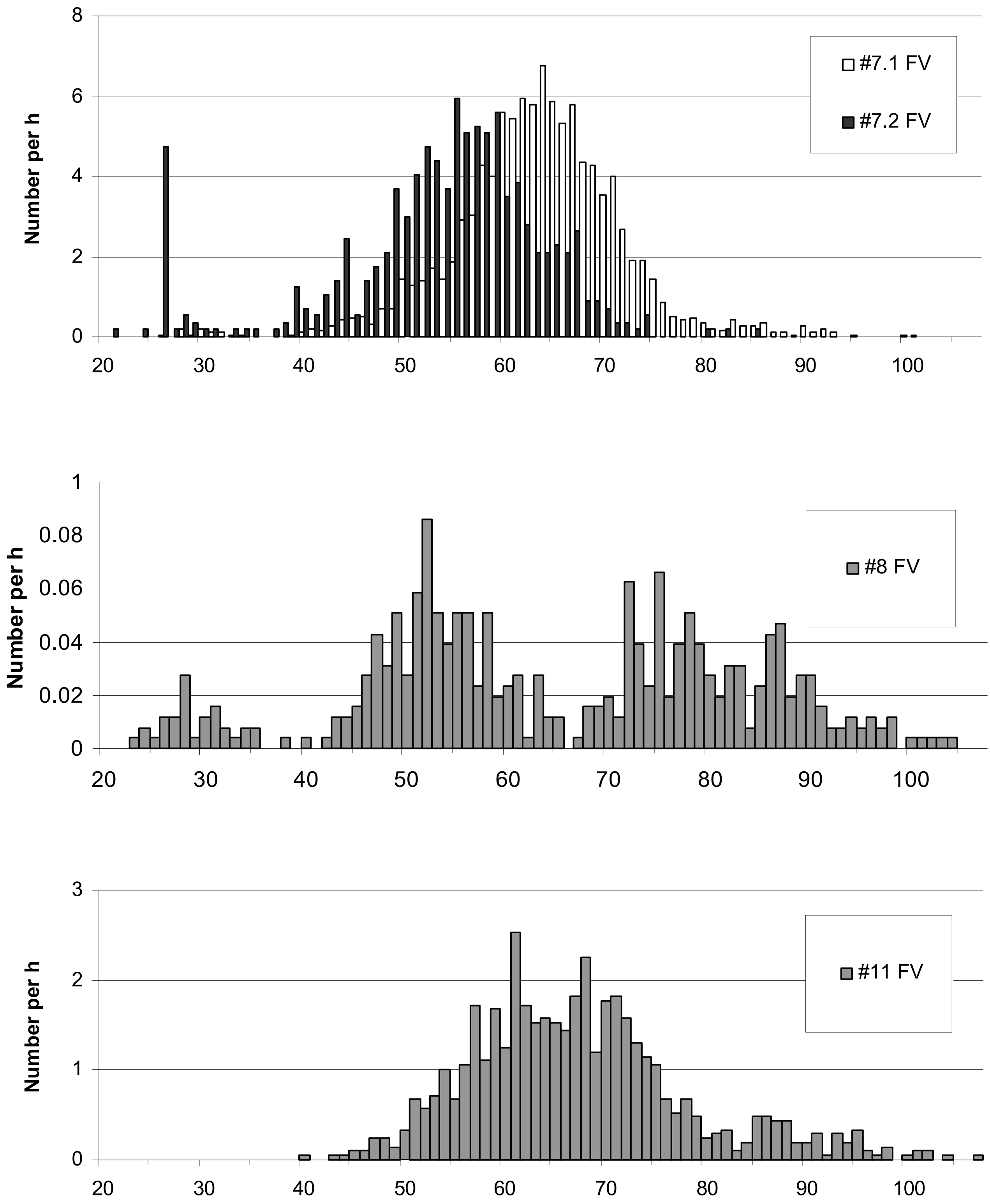
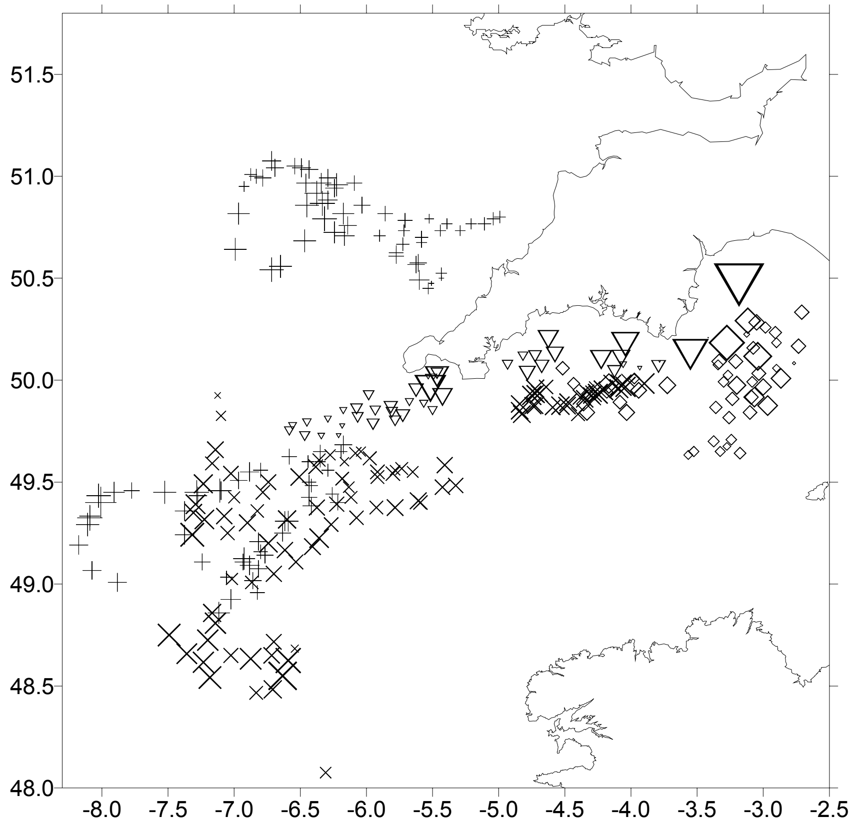


Fig. 6 FSP surveys #4 and 5: Spatial coherence of mean length in catch. Symbol heights scaled linearly to (mean length - 20 cm). x = FV1, + = FV2, ✄ = FV3, ▤ = FV4. FV= fishing vessel.

a)Monk, *Lophius piscatorius*



b) Plaice

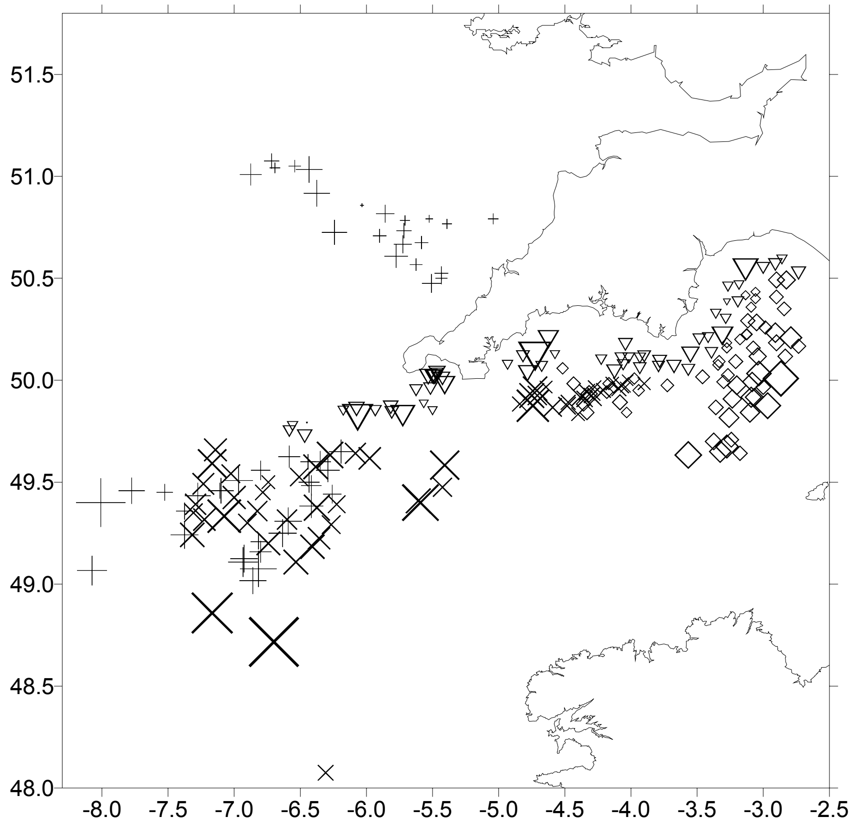
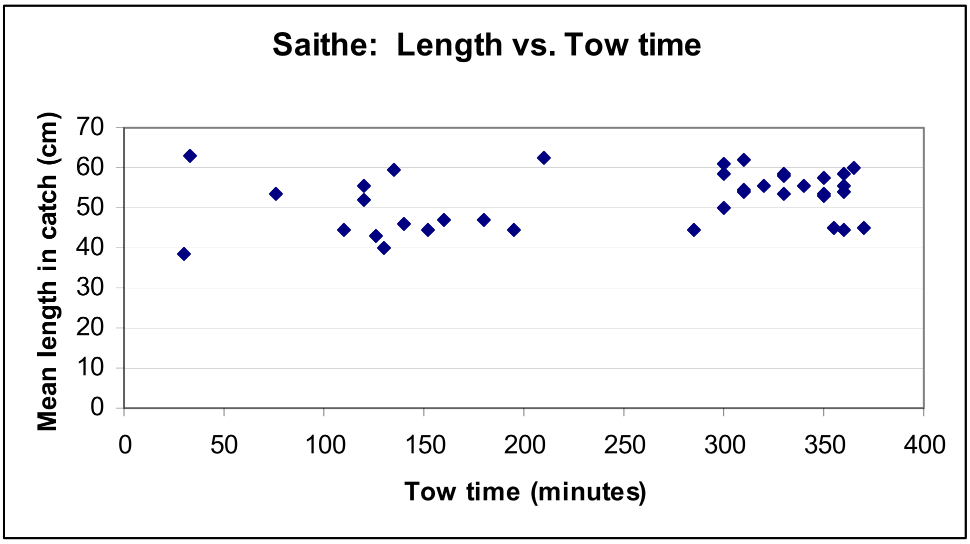
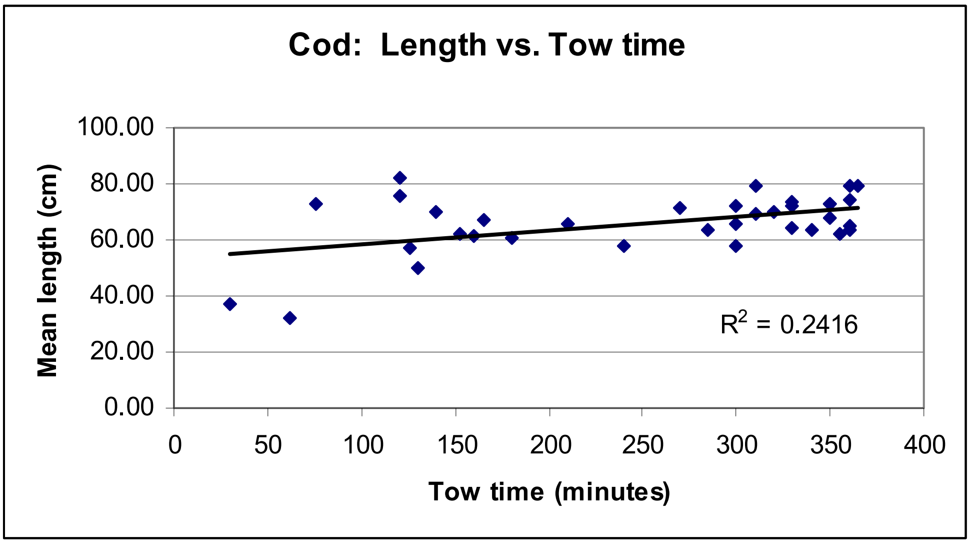


Fig. 7 FSP survey #2: mean length of saithe and cod in catches as a function of towing time. No significant linear relationship for saithe or haddock.





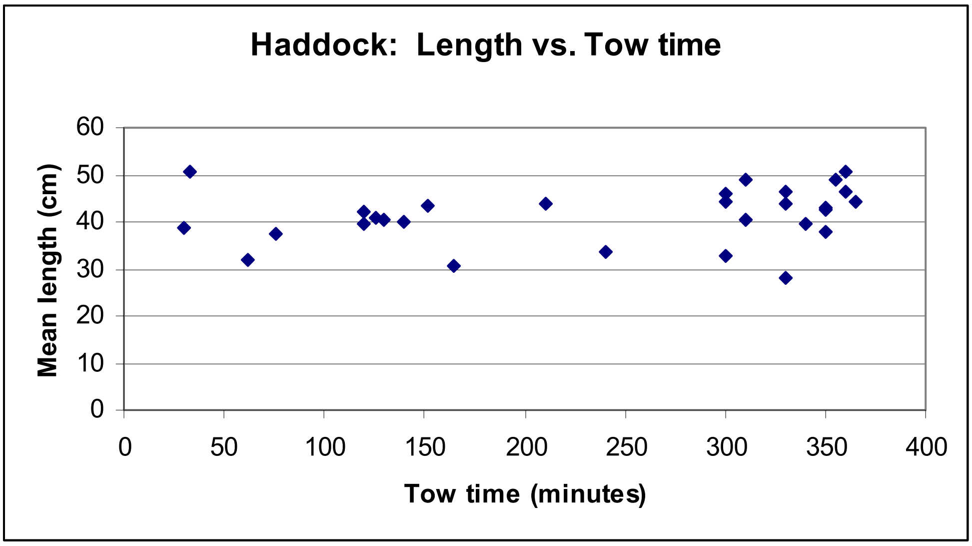
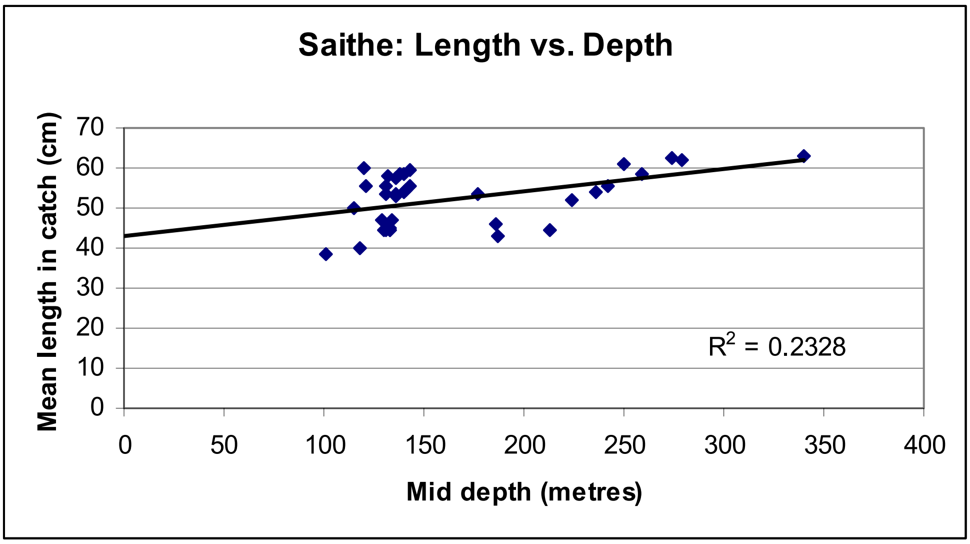


Fig. 8 FSP survey #2: mean length of saithe in catches as a function of depth.



Working Document to be presented at the Workshop on Survey Design and Data Analysis, 21-25 June 2004, Aberdeen

The use of generalised additive models in sardine acoustic estimates.

Juan Pablo Zwolinski
INIAP/IPIMAR, Av. de Brasília, 1449-006 Lisbon, Portugal
Juan@ipimar.pt

Introduction

Acoustic sampling is one of the basic tools used by many fisheries research institutes through out the world to quantify stocks of pelagic fish (MacLennan & Simmonds, 1992). Acoustic surveys are based on the principle that fish can reflect acoustic signals produced by a scientific echo-sounder and that those echoes can be quantified. Acoustic surveys with a regular sampling design are thus carried in order to estimate the total amount of backscattered acoustic energy of the target species. As the survey is conducted acoustics readings are averaged at regular intervals called "Elementary Distance Sampling Unit" - EDSU, usually ranging from 1 to 2.5 nautical mile long intervals. The summed or integrated backsattered energy per EDSU is proportional to fish density and inversely proportional to its size. A particular problem in acoustics abundance estimation is the statistical combination of line-transect measurements of fish density to estimate abundance over the survey region and it's variance (Foote & Stefánson, 1993). This is a major concern for the International Committee for Exploration of the Seas – ICES, that has since the early 90's sponsored meetings and workshops dedicated to this particular matter.

Acoustic data (in the form of Nautical Area Scattering Coefficient - NASC) is typically "contaminated" by a large proportion of zero observations and has a highly skewed positive distribution for non-zero values. Post-survey stratification of the region is a common practice to estimate overall abundance and to lower its variance. But this method assumes homogeneity and independence within each stratum that it's likely to be false in the case of acoustic surveys. Stefánson (1996) made use of generalised linear models - GLM (McCullagh & Nelder, 1989) in the analysis of ground groundfish survey data. This author extended the concept of the delta distribution (Syrjala, 2000), which is characterised by a positive probability of a zero observation and a lognormal distribution for the positive ones, to a GLM framework. GLM's are a powerful tool describing some biological phenomena as they are capable of handling non-linear relations of the predictors and response via a link function and the error distribution is not restricted to Gaussian distribution. Nevertheless they still keep the parametric form of the common linear models on the linear predictor. For that reason GLM's are not well suited in a spatial context where the relation between the response and predictors is complexly non-linear.

Borchers *et al* (1997) took the last approach one step ahead and instead of the GLM structure, they modelled egg abundance with Generalised additive models - GAM (Hastie & Tibshirani, 1990). As Stéfanson (1996) stated in his GLM approach the probability of a positive sample and the density of the positive ones can be fitted separately and then multiplied to obtain the final fitted density at a given set of predictors. In an acoustical application, the form of Borchers *et al* (1996) model can be defined as

$$\mathrm{E}\left[\mathrm{y}\right]= f\left(\beta_{01} + \sum_i s_i\left(x_i\right)\right)\times g\left(\beta_{02} + \sum_j s_k\left(x_k\right)\right)$$

Where the left-hand term is the expected value of the acoustic energy attributed to the target species. The $f(.)$ function is the inverse of the link function for the binomial model, typically the logit link function; β_{01} is a constant and $\sum_i s_i\left(x_i\right)$ is a sum of smooth functions of covariates. The function $g\left(.\right)$ is the inverse link of the function used for the positive observations and it's argument is again a constant and a sum of smooth functions of the predictors (see Wood & Augustin, 2002). Approximate confidence intervals for a given set of predictor variables can be obtained by simulation of model parameters. These parameters have an asymptotically Gaussian multivariate distribution.

Application to sardine acoustic data

The method described above has been applied to sardine NASC data form Portuguese surveys. Figure 1 represents the acoustic energy per 1 nautical mile EDSU in the Portuguese northern waters in November 2000. As it is easily recognisable in the figure, non-zero samples are encountered in shallower waters and the highest abundances appear at the centre and southern part of the region. The mgcv package (Wood, 2001) for R software was used to fit the GAMs. Both the presence/absence and positive NASC models were satisfactory fitted with only latitude, longitude and depth as predictors. A Binomial distribution with a logit link function was used in the presence/absence model. A Poisson distribution allowing overdispersion with a log link function was used to model non-zero NASC values. A final density surface was constructed as the product of both models predictions on a 1x1 nautical mile grid covering all of the sampled area. The reconstructed NASC surface matches quite well the observed data (Figure 2). Integration of the fitted surface with nearest neighbour application of mean sardine length (not shown) to each grid node, in order to convert NASC to fish number, provided an overall abundance estimate of $26*10^3$ millions of individuals. This value is slightly lower to the one obtained in the traditional estimate - $29.4*10^3$. Simulation of the model parameters provided 95% confidence limits of $23.8*10^3$ and $31.6*10^3$. The 95% confidence limits for the abundance obtained treating the data as independent samples were $19.4*10^3$ and $36.8*10^3$.

Discussion and conclusions

The method described above has already been applied successfully to data with the same characteristics as acoustic data. The first applications to Portuguese sardine acoustic data revealed good model properties such as consistency with traditional estimates and a reduction of the confidence interval bounds. A descriptive analysis suggest that there could be room for improvement of model fit and consequently in precision of the final abundance estimate, with the inclusion of environmental parameters such as chlorophyll.

References:

Borchers, D.L., Buckland, S.T., Priede, I.G., Ahmadi, S. 1997. Improving the precision of the daily egg production method using generalized additive models. Can. J. Fish. Aquat. Sci. 54: 2727-2742

Foote, K.G., Stefánson, G. 1993. Definition of the problem of estimating fish abundance over an area from acoustic line-transect measurments of density. ICES J. Mar. Sci 50: 369-381

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

MacLennan, D.N., Simmonds, E. J. 1992. Fisheries Acoustics. Chapman & Hall, London.

McCullagh, P., Nelder, J.A. 1997. Generalized Linear Models. Chapman & Hall, London

Stefánson, G. 1996. Analysis of groundfish survey abundance data: combining the GLM and delta aproaches. ICES J. mar. Sci. 53: 577-588

Syrjala, S.E. 2000 Critique on the use of the delta distribution for the analysis of trawl survey data. ICES J. Mar. Sci. 57: 831-842

Wood, S. 2001. mgcv: GAMs and generalized Ridge Regression for R. R News 1(2): 20-25

Wood, S., Augustin, N. 2002 GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. Ecological Modelling 157:157-177

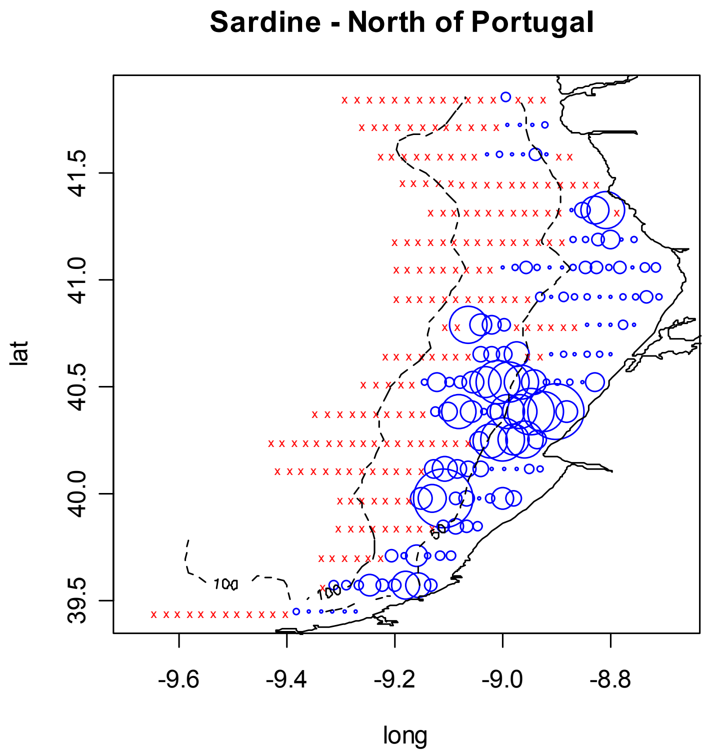


Figure 1: Sardine backscattered energy in space. Circle diameter are proportional to the squared root of acoustic energy. Red crosses are samples with no sardine. The black line represents the coast and the broken ones follow the 50 and 100 meter depth contour.

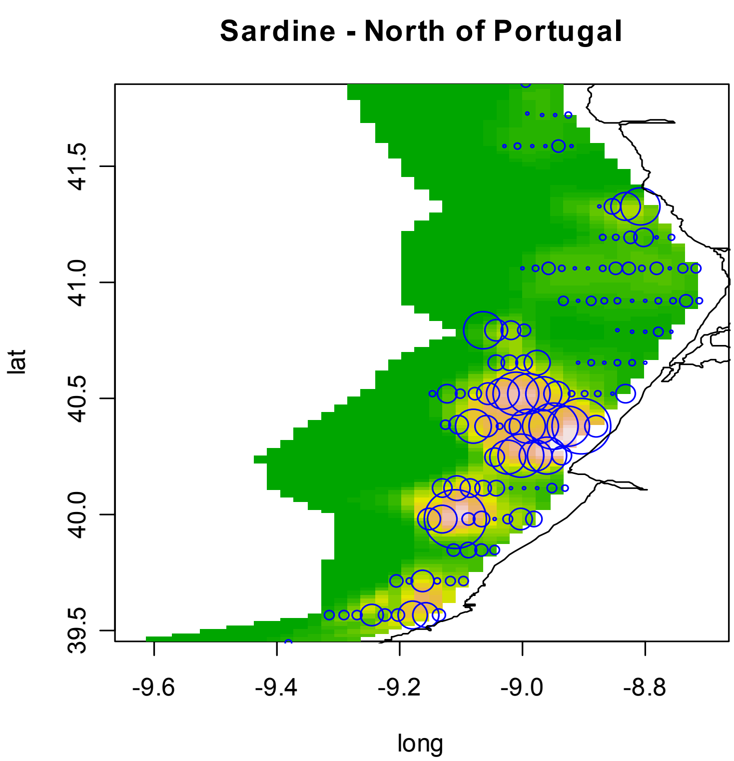


Figure 2 : Fitted sardine energy - on the background; and observed data in circles - the same as above.

Woking Document 7

Design efficiencies of transect and stratified random trawl surveys

By

Jon H. Vølstad[1], Mary Christman[2], and Thomas J. Miller[3]

[1]Versar, Inc., 9200 Rumsey Rd., Columbia, MD 21045, USA

[2]Department of Animal and Avian Sciences, University of Maryland,

College Park, MD 20742, USA

[3] University of Maryland Center for Environmental Science,

Chesapeake Biological Laboratory, Solomons, MD 20688, USA.

Abstract

Spatially overlapping transect and stratified random trawl surveys were conducted simultaneously in the Chesapeake Bay during spring, summer, and fall during 2002 and 2003 to quantify the relative abundance, diversity, distribution and trophic interactions of economically and ecologically important fish species. We evaluated the design efficiency of each survey with respect to the precision of relative abundance estimates for a given sampling effort. We used Kish's design effects and effective sample sizes to assess the efficiency and cost-effectiveness of each survey with respect to estimates of relative abundance for selected species and overall. The design effect is the ratio of the actual variance in mean CPUE for the complex design to the expected variance of that statistic under a simple random sample of the same size. The empirical design effects suggested that the stratified random survey was more effective than the transect sampling in estimating mean CPUE across species, and for most target species. However, for weakfish trawling along transects appeared to be more effective than the stratified random survey. The stratified random survey with proportional allocation was generally more effective than simple random sampling, suggesting that the stratification was effective for most species. We applied a composite estimator to combine the estimates of mean CPUE from the two independent surveys within season for each year, using weights that are based on their respective effective sample sizes. The weights are optimized with respect to achieving minimum variance. The method presented here for combining estimates across survey is robust to heterogeneous variances such as might be found when the means and variances are correlated, which is often the case for CPUE from marine trawl surveys. A similar approach, with possible modifications for gear differences, could also be used to combine multiple surveys for use in tuning VPA in stock assessments.

Keywords: Marine trawl surveys; Survey design; Design effect; Transect surveys; Effective sample size; Cluster-sampling; Intra-cluster correlation; Composite estimator

Fisheries-independent trawl surveys are widely used to provide estimates of relative abundance and other population parameters crucial for effective fisheries management (Gunderson, 1993). The average number of fish caught per area or volume swept has long been used as an index of relative abundance (Grosslein, 1969; Pennington, 1985, 1986; Smith, 2002). Standardized trawl surveys, coupled with appropriate allocation of stations, can provide reliable estimates of changes in abundance, if the catch efficiency of the survey gear remains approximately constant by depth and over time. However, given the expense of trawl programs, it is important to optimize the survey design so that the precision of key parameters is maximized for a fixed total survey cost.

The distribution of marine resources is generally highly patchy (cf. Seber, 1986), and often results in strong short-range spatial autocorrelation in measures of relative abundance (Polacheck and Vølstad, 1993; Kingsley et al., 2002). A considerable body of research has sought to determine survey designs appropriate for temporal and spatial variability. A classic example includes the application of stratification and sample allocation schemes that seek to increase inter-stratum variability while minimizing intra-stratum variability, thereby increasing the precision of parameter estimates for single species (Cochran, 1977; Gavaris and Smith, 1987; Harbitz et al., 1998). Designs for surveys that target multiple species are more complicated as distributions of the individual species are likely to vary at differing spatial and temporal scales. The recent interest in multispecies fisheries management will require multispecies, fishery-independent surveys. However, guidance on optimization of multispecies surveys is lacking.

The presence of biological and technical interactions among Chesapeake Bay fisheries (Miller et al., 1996) has motivated management agencies to establish a goal of implementing multispecies management in the Chesapeake Bay by 2007 (C2K agreement). To support the development of multispecies management in the Chesapeake Bay an international workshop recommended the development of coordinated, baywide surveys to estimate key species abundances and to provide biological data on both economically and ecologically important species that are currently lacking (Houde et al., 1998).

We conducted seasonal midwater trawl surveys in the Chesapeake Bay to evaluate the efficiency of different survey designs for a range of finfish species as a component of the Chesapeake Bay Fishery Independent Multispecies Survey (CHESFIMS). Information for the initial survey design came from an earlier research program funded by the National Science Foundation to quantify 'Trophic Interactions in Estuarine Systems' (TIES). The TIES program included baywide midwater trawl surveys, conducted from 1995 - 2000 during spring, summer and fall (Jung, 2002; Jung and Houde, 2003). It demonstrated that baywide seasonal surveys are required to characterize the communities in the Bay because of the spatial and temporal variability in fish abundance and community structure (Jung and Houde, 2003).

Our goals for the design of CHESFIMS were to: (1) maintain the time series from the TIES program while improving survey efficiency, (2) complement existing fishery-independent surveys and (3) expand survey coverage to ecologically important forage species to aid development of multispecies fisheries management in the Chesapeake Bay. We sought to develop a cost-effective design for the monitoring of abundance, diversity, distribution and trophic interactions of economically and ecologically important fish species in the Chesapeake Bay over broad spatial and temporal scales. Accordingly, we conducted spatially and temporally overlapping transects and stratified random trawl surveys in the Chesapeake Bay during spring, summer, and fall in 2002 and 2003. Here we evaluate the efficiency of both survey designs for estimating relative abundance using design effects and effective sample sizes (Lehtonen and Pakinen, 1994; Kish, 2003). We also demonstrate how two survey indices of abundance can be combined to yield an overall index with minimum variance.

Material and Methods

Study Area

The Chesapeake Bay is the largest estuary in the United States, with an area of approximately 6,500 km^2, a length of over 300 km, a mean depth of 8.4 meters, and a volume of 52,112 km^3. Bay anchovy (*Anchoa mitchilli*) is the most abundant and ubiquitous fish in the Bay (Jung and Houde, 2004). Although bay anchovy is not harvested, it has an important role in the ecosystem because it is a major prey of piscivores, including several economically important fishes (Jung and Houde, 2004; Baird and Ulanowicz, 1989; Luo and Brandt, 1993; Hartman and Brandt, 1995). Fisheries in Chesapeake Bay contribute significantly to U.S. catches at the national and regional levels. Fisheries statistics from the National Marine Fisheries Service (NMFS) shows that around 220,000 metric tons (t) of fish and shellfish were harvested in 2002 from Chesapeake Bay waters, with a dockside value of more than $172 million. All exploited species in the Chesapeake Bay are currently managed and regulated on a single species basis.

Survey methods

Standardized trawl surveys have been conducted during spring, summer, and fall in the mainstem of the Chesapeake Bay since 1995 using an 18-m^2 midwater trawl (MWT) with 3-mm codend mesh towed from the University of Maryland Center for

Estuarine and Environmental Science R/V 'Aquarius' (Jung, 2002). The MWT samples small fish (30-256 mm total length) of most species effectively, but appears to be less effective for Atlantic menhaden and striped bass (*Marone saxatilis*) of all sizes (Jung and Houde, 2003). The trawl stations in the TIES program were located along 15 fixed transects spaced approximately 18.5 km (10 nm) apart from the head of the Bay to the Bay mouth (Jung and Houde, 2003). Within each season, 11 of the 15 transects were occupied. The survey area was stratified by dividing the Bay into upper, middle, and lower regions (Figure 1) that each has distinctive characteristics, with strata boundaries broadly corresponding to ecologically relevant salinity regimes and depths above 5 m. The upper Bay is generally shallow, with substantial areas with depth less than 5 m, and has well mixed waters with high nutrient concentrations. The bottom topography in the mid Bay includes a narrow channel in the middle of the Bay with a stratified water column and broad flanking shoals. This region has relatively clear waters and experiences seasonally high nutrient concentrations. The lower Bay has the clearest waters, greatest depths and lowest nutrient concentrations (Kemp et al., 1999). The strata volumes are 26,608 km^3 (Lower), 16,840 km3 (Mid) and 8,664 km^3 (Upper).

CHESFIMS was initiated in 2001, employing the TIES trawling procedures, transect design and stratification, with 2-4 trawl stations sampled within 11 transects during spring, summer, and fall. The m_i stations within each transect i were selected by restricted random sampling. Each transect was divided into m_i segments of equal size, with one station allocated randomly within each segment. Starting in 2002, the transect sampling was augmented with an independent stratified random trawl survey in an effort to optimize the monitoring design. In the stratified random surveys, conducted during the same three seasons as the transect surveys, 20 stations were allocated proportional to the volume of each stratum during each sampling period. Latitude and longitude of stations within strata were randomly generated. For logistical reasons, a number of the 20 stratified random stations were not completed during some surveys. An example of the allocation of stations in the 2002 stratified random survey is shown in Figure 1. For the transect surveys, the clustering of stations and variable transect lengths resulted in heterogeneous selection probabilities for stations within strata.

The transect and stratified random surveys followed the TIES trawling procedures, with standardized 20-minute oblique, stepped tows conducted at each station using the same midwater trawl with 18 m^2 opening and 3-mm cod end mesh. The trawl was towed for two minutes in each of ten depth zones distributed throughout the water column from the surface to the bottom, with minimum trawlable depth being 5 m. The section of the tow conducted in the deepest zone sampled epibenthic fishes close to or on the bottom. The remaining portion of the tow sampled pelagic and neustonic fishes. All tows were conducted between 19:00 and 7:00 Eastern Standard Time to minimize gear avoidance and to take advantage of the reduced patchiness of multiple target species at night. Catches were identified, enumerated, measured and weighed onboard.

Estimating mean catch per unit effort

To estimate the mean catch per unit effort (CPUE) and the associated variance, we treated the transect survey as a stratified two-stage design, with primary sampling units of unequal size (e.g., Cochran, 1977; Wolter, 1985) to account for the variable transect lengths. For estimating purposes, we assumed that the n_h primary sampling units (transects) were selected randomly with equal probability from each stratum h ($h = 1, 2, 3$); the m_{hi} stations (2 - 4) within each transect were selected with equal probability. Let $\overline{y}_{hi}$ denote the mean CPUE for the m_{hi} stations within transect i in stratum h, and let l_{hi} denote the transect length. We applied a combined (across strata) ratio estimator for a two-stage survey (Cochran, 1977) to estimate the overall mean CPUE

$$\overline{y}_T = \frac{\sum_{h=1}^{3} w_h \overline{l}_h \overline{y}_h}{\sum_{h=1}^{3} w_h \overline{l}_h} \tag{0.1}$$

where, for stratum h,

$$\overline{y}_h = \frac{\sum_{i=1}^{n_h} l_{hi} \overline{y}_{hi}}{\sum_{i=1}^{n_h} l_{hi}} \tag{0.2}$$

is the weighted mean CPUE across transects,

$$\bar{l}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} l_{hi} \tag{0.3}$$

is the mean transect length within a stratum, and

$$w_h = \frac{V_h}{\sum_{h=1}^{3} V_h} \tag{0.4}$$

is the stratum weight, with V_h being the stratum size (volume). The transect-wise variation of mean CPUE as well as the variation in estimated mean transect length by stratum contributes the variance of (0.1). An approximate estimator of the variance of (0.1) is (see Sukhatme and Sukhatme, 1970, p. 307)

$$\mathrm{var}(\bar{y}_T) = \sum_{h=1}^{3} \lambda_h^2 \left\{ \frac{s_{hb}^2}{n_h} + \frac{1}{n_h^2}\sum_{i=1}^{n_h}\left(\frac{l_{hi}}{\bar{l}_h}\right)^2 \frac{s_{hi}^2}{m_{hi}} \right\}, \tag{0.5}$$

where

$$\lambda_h = \frac{w_h \bar{l}_h}{\sum_{h=1}^{3} w_h \bar{l}_h}$$

is the proportion of second-stage units in stratum h,

$$s_{hb}^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(\bar{y}_{hi} - \bar{y}_h)^2$$

is the between transect variance in CPUE for stratum h, and

$$s_{hi}^2 = \frac{1}{m_{hi}-1}\sum_{j=1}^{m_{hi}}(y_{hij}-\overline{y}_{hi})^2$$

is the within transect variance in CPUE for transect i in stratum h. We used SUDAAN (RTI 2001), a specialized software for the analysis of complex surveys and cluster-correlated data (Brogan, 1998; Carlson, 1998), to estimate the variance of $\overline{y}_T$ (eq. 1.1). SUDAAN adjusts for the within-transect correlation in CPUE by using the between cluster variance estimator for cluster-correlated data commonly used in multi-stage sample surveys (Hansen et al., 1953; Cochran, 1977; Särnal et al., 1992) in combination with a Taylor series linearization approach (Woodruff, 1971; Binder, 1983; Wolter, 1985; Williams, 2000).

In the stratified random surveys (STR), stations were allocated proportional to the volume of each stratum. We applied the standard estimators for the stratified mean and its variance (Cochran, 1977)

$$\overline{y}_{str} = \sum_{h=1}^{3} w_h \overline{y}_{h,srs} \tag{0.6}$$

and

$$Var(\overline{y}_{str}) = \sum_{h=1}^{3} w_h^2 Var(\overline{y}_{h,srs}) \tag{0.7}$$

where the weight for stratum h is based on its fraction of the total volume as in equation (0.4), $\overline{y}_{h,srs}$ is the ordinary mean CPUE for simple random sampling within stratum h, and $Var(\overline{y}_{h,srs})$ is the corresponding variance of the stratum mean CPUE. The spring survey in 2002, and the fall survey in 2003 had incomplete sampling coverage for logistical reasons. For these surveys we collapsed the strata and treated all stations as a simple random sample (SRS), and then used the ordinary estimators of the mean and variance.

Analytical evaluation of design efficiency

The efficiency of each survey design was evaluated by comparing the respective design-based variance of the estimated mean CPUE ($\bar{y}$) with the expected variance obtained under simple random sampling. Kish (1965; 1995; 2003) defined the “design effect” as the ratio of the two variances,

$$deff = Var_c(\bar{y}_c) / Var_{srs}(\bar{y}_{srs}) \tag{0.8}$$

where $Var_c(\bar{y}_c)$ is the variance estimate based on the actual (complex) survey design, and $Var_{srs}(\bar{y}_{srs})$ is the estimate based on simple random sampling (SRS) for a sample of equal size. The design-based variance, $Var_c(\bar{y}_c)$, reflects the effects of stratification and, for the transect survey, clustering of stations. Kish (1995) and Potthoff et al. (1992) provide a general discussion on the calculation of design effects and effective sample sizes. We estimated $Var_c(\bar{y}_c)$ for mean CPUE for the stratified random survey from equation (0.7), while $Var_{srs}(\bar{y}_{srs})$ was estimated by treating the stations as a simple random sample from the total survey area. This estimate of the variance that would have been obtained under a simple random sample of the same size is justified because stations in general were allocated proportionally to the strata sizes.

The "effective sample size" for estimation of mean CPUE ($\overline{y}$) using data from the complex survey design C is defined as

$$n_C^* = n / deff \ . \tag{0.9}$$

The effective sample size n_C^* is the number of stations selected by simple random sampling that would be required to achieve the same precision obtained with n stations under the actual complex sampling design. If, for example, the design effect equals two for the estimated mean CPUE for a transect survey with 30 stations, then a simple random sample of 15 stations (the effective sample size) would have achieved the same precision.

We used a composite estimator to take a weighted average of the mean CPUE for the independent stratified random and transect surveys (e.g., Korn and Graubard, 1999; Rao, 2003). The estimator for the combined mean is given by:

$$\overline{y}_{comb} = \phi \overline{y}_{STR} + (1-\phi)\overline{y}_T \tag{0.10}$$

with the weight ϕ $(0 \le \phi \le 1)$ being chosen to minimize the variance of $\overline{y}_{comb}$

$$Var(\overline{y}_{comb}) = \phi^2 Var(\overline{y}_{STR}) + (1-\phi)^2 Var(\overline{y}_T) \tag{0.11}$$

where $Var(\overline{Y}_{STR})$ and $Var(\overline{Y}_T)$ are the design-based variances of the mean CPUE estimators for the stratified random and transect surveys, respectively. The optimum weight (ϕ_{opt}), expressed as a function of the effective sample sizes for each survey (n^*_{STR}, n^*_T), is

$$\phi_{opt} = \frac{n^*_{STR}}{n^*_{STR} + n^*_T}$$

and, hence,

$$1 - \phi_{opt} = \frac{n^*_T}{n^*_{STR} + n^*_T}$$

(See appendix A for derivation). Thus, the optimal weight depends only on the ratio of the effective sample sizes ($R = n^*_T / n^*_{STR}$). We used the sample data to estimate the optimal weight ϕ_{opt}.

3. Results

Our comparisons of the transect and stratified random survey designs focused on mean catch per unit efforts for all species combined, and for Bay anchovy (*Anchoa mitchilli*), white perch (*Morone americana*), Atlantic croaker (*Micropogonias undulatus*), and weakfish (*Cynoscion regalis*) (Tables 1-6). The relative standard error (RSE), defined as the ratio of the standard error (SE) to the survey estimate ($\overline{y}$), is used as a measure of precision (Jessen, 1978). The two independent surveys produced comparable estimates of mean catch per unit effort (CPUE) across species (Figure 2) but differed for the individual species that were considered (compare Tables 1-2, and Tables 4-5). The combined estimates shows that bay anchovy (*Anchoa mitchilli*) was most abundant and widely distributed during all seasons in both years, but with significantly lower ($p<0.5$) summer abundance in 2003 as compared to 2002 (Tables 3 and 6). White perch (*Morone americana*) dominated catches in the upper Bay, but was relatively rare in the mid and lower bay strata. In the mid and lower bay regions, weakfish (*Cynoscion regalis*) and croaker (*Micropogonias undulatus*) were common. Overall, diversity of catches was highest at the northern-most and southern-most stations.

The respective design effects suggest that the stratified random survey consistently is more efficient than the transect survey for estimating mean CPUE in every season for all species combined (Figure 3), and for bay anchovy and white perch specifically (Tables 1-2, and 4-5). The relative standard errors of these estimates from the stratified random survey were on par, or lower than those from the transect survey, although the latter survey occupied from 55 % to 81% more stations. This supports the conclusion that transect sampling generally is less efficient than stratified random sampling for the system studied here. The effective sample sizes for estimating overall mean CPUE for the transect surveys were lower than the number of stations during all seasons, and similar to the number of transects during summer and fall. In contrast, the design effects suggested that transect sampling were more effective than stratified random sampling for estimating the abundance of weakfish during all three seasons (*deff* ≤ 1) (Tables 1-2 and 4-5; Figure 4). Here the effective sample sizes were much larger than the actual number of stations for some seasons, especially for spring and fall. These design effects may be biased downwards because of the high frequency of zero catches within some strata. However, an inspection of the distribution of weakfish and the allocation of stations (Figure 1) confirms that transects effectively captured the spatial variability in abundance. For bay anchovy, in contrast, the stations within transects had similar CPUE and thus did not capture the overall variance. White perch were patchily distributed, but some locations appear to have consistent high density over time. Trawl stations along transect had significantly higher mean CPUE than the stratified random stations in both years (Tables 1-2, and 4-5). This suggests that the fixed transects covered patchy habitats that are favored by white perch, while the random stations missed these high-density areas by chance. For this species, hence, transects can be more effective than random sampling for detecting trends in abundance over time. The composite estimator produced precise estimates of mean CPUE for all species combined, with relative standard errors (RSE) between 13% and 22% (Tables 3 and 6). The combined estimates for white perch shows that the use of effective sample sizes to determine weights can yield more accurate results than the use of traditional weights based on the variances of each survey estimate. In the 2002 fall surveys, the stratified random trawl stations caught zero white perch (Table 1), compared to a mean of 7.9 fish per tow for the transects. Weighting based on the variances would have assigned all the weight to the stratified random survey, resulting in a poor combined CPUE estimate of zero. The combined estimate based on effective sample sizes (2.3 fish per tow) is more reliable than the former.

4. Discussion and Conclusions

The comparison of design effects shows that stratified random sampling with proportional allocation generally is more effective than transect sampling for estimating the mean CPUE across all species in the mainstem of Chesapeake Bay. The transect survey resulted in less precise estimates of mean CPUE across species compared to the stratified random survey in each season, even though the former survey completed 55% to 81% more trawling stations.

The design effect is superior to the use of relative standard error (RSE) for evaluating survey efficiency, as it is independent of the sample size *n,* unlike RSE, which is a function of $\sqrt{n}$. When determining the required sample size to achieve an adequate level of precision in key estimates it is essential to account for the design effect resulting from a given design. The efficiency of sampling along transect depends on the homogeneity in catch per tow for clusters of stations within transects in addition to the stratification. Taking stations along transects would be an efficient design if catch per tow within each transect were as variable as those in the general population of trawling stations since then the intra-cluster correlation would be low or even zero. In the transect survey analyzed here, selecting an additional station from the same transect generally adds less new information than would a completely independent selection. With exception of weakfish and croaker, results from the fixed transect surveys show that the design effect tends to be higher than unity. Hence, a simple random survey would generally be expected to produce more precise estimates for a similar survey effort.

We treated transects and stations within transects as a two-stage cluster sampling design. In fact, the actual design was slightly more structured in the sense that transects were chosen so that spacing would be somewhat even. Such a design does not involve replication, and thus there is no analytical method for determining either the potential bias or the variance of the estimator of mean CPUE. One reason for this allocation of stations in the TIES program was to ensure maximal spatial coverage over the bay. This is the same reasoning often used for taking systematic samples. Another possible advantage of stratified random sampling over systematic sampling is that the former tends to result in a shorter sailing distance to occupy all stations (Harbitz and Pennington, 2004), and thus reduces survey cost for a fixed number of stations. Depending on the range of positive autocorrelation for the target species, a systematic allocation of stations that maximizes their separation may increase the effective sample size relative to random sampling by reducing the spatial autocorrelation for CPUE. However, even when this is the case, the increase in effective sample size would have to be sufficiently large to offset the increased cost (cruise-track length) for collecting the samples.

Effective estimation of means using complex survey designs requires estimators that fully account for the sampling design. Data collected from clustered samples often result in a reduction in the effective sample size for estimating a statistic such as mean CPUE because of the tendency of measurements collected within clusters to respond more similarly than measurements taken between clusters (e.g., Pennington and Vølstad 1994; Williams 2000). For complex survey designs, the estimation of the variance of the mean CPUE under the assumption of independence between all observations will generally underestimate the true variance in the presence of positive intra-cluster correlation. For complex surveys, we concur with Brogan (1998) and recommend that specialized sample survey software such as SUDAAN be used for estimation of population parameters, descriptive analyses and analytical analyses.

The effective sample size for estimating CPUE for all species combined for the transect survey is closer to the number of transects than to the overall number of stations. This again suggests that CPUE from stations within the same transect tend to be correlated, while any two observations from different transects are independent. Thus, the between cluster variance estimator is justified under the assumption of independence between observations (Williams, 2000). The designated strata and proportional allocation of stations did not substantially reduce variability in CPUE between stations in the 2002 surveys, but generally resulted in higher precision than for simple random sampling in the 2003 surveys. Thus, the relative efficiency of the stratified random and transect sampling vary across years. However, the stratified random design appears to be robust to variations in the spatial distribution of different species, and is recommended over simple random sampling.

The theoretical optimum allocation (Neyman allocation) of a given number of hauls among strata is to sample each stratum in proportion to its standard deviation multiplied by the stratum size (Cochran, 1977). However, in practice it may be better to allocate stations proportional to stratum size since the standard deviation only can be approximated from previous surveys, and the spatial distribution of fish exhibits temporal variability. Stratification with proportional sampling nearly always leads to gain in precision (Cochran, 1977), with the largest gains being achieved when the strata means exhibit large variation. We are continuing overlapping stratified random and transects surveys during 2004 to evaluate if the relative efficiency of the two survey designs remains approximately constant, and to provide more data for optimizing the survey design. Design effects related to individual components such as stratification and clustering will be used to guide the choice of an efficient survey design for multispecies trawl surveys.

The composite estimates of mean CPUE across surveys were more precise than the individual survey estimates, with relative standard errors 18% to 27% lower than the most precise individual component estimate. This study suggests that a composite estimator with weights based on the effective sample sizes of individual survey estimates can yield more accurate results than weights based on variances. For trawl surveys, patchy distributions can result in zero catch even though some habitats have high densities. If an estimate from a simple random survey with zero catch is combined with another survey estimate with mean and variance greater than zero, then the commonly used weighting based on their respective variances will produce a poor combined estimate of zero. Our method of assigning weights based on their respective effective sample sizes, in contrast, will yield a more equitable combined estimate because the number of stations will determine the weight of the simple random survey. The composite estimator is unbiased as long as the variances are independent of the means. For trawl surveys, catch per tow tend to have a skewed distribution because of patchiness of fish populations, and the variance thus is possibly related to the mean (Seber, 1986; Pennington and Vølstad, 1991). When this is the case, a time series of combined estimates would be biased downward if the means differ appreciably because the survey with the lowest mean would tend to be weighted more. For large effective sample sizes, the distribution of the means of either survey would be normalized, thus eliminating such bias. Our analysis suggest that the variances of the mean CPUE for the transect survey were relatively high because the stations were clustered, and not because the means were higher than for the stratified random. For a series of combined surveys, we recommended that the weight be fixed to eliminate bias caused by the dependence between the variance and the mean. The maximum reduction of 50 % in the variance of the combined mean CPUE is achieved when the two surveys have equal effective sample sizes. However, the precision of the combined estimate is robust to deviation from the optimal ratio (R) of effective sample sizes; a change R from unity to 6 does not significantly increase the variance of the composite estimator (Rao, 2003, p. 58). In practice, hence, the use of a fixed weight across years should not appreciably increase the variance of combined estimates.

Acknowledgements

This study (No XXX of the University of Maryland Center for Environmental Science) was funded by the National Marine Fisheries Service, through a grant from the Chesapeake Bay Stock Assessment Committee. This research would not have been possible without the help of the students and technicians at CBL who participated in the research cruises. In particular, we thank Kiersten Curti, Chris Heyer, and Dave Loewensteiner who served as chief scientists on CHESFIMS cruises. We appreciate comments from John Hoenig (VIMS) that helped improve this manuscript.

Literature cited

Baird, D., and R.E. Ulanwitz.

1989. The seasonal dynamics of the Chesapeake Bay ecosystem. Ecol. Monogr. 59:329-364.

Brogan, D.

1998. Software for sample survey data, misuse of standard packages. *In*: Encyclopedia of Biostatistics, Volume 5 (P. Armitage and T. Colton, Eds.). John Wiley & Sons, New York, NY pp. 4167-4174.

Carlson, B.D.

1998. Software for Statistical Analysis of Sample Survey Data. *In:* Encyclopedia of Biostatistics, Volume 5 (P. Armitage and T. Colton, Eds.). John Wiley & Sons, New York, NY: pp. 4160-4167.

Cochran, W.G.

1977. Sampling Techniques 3rd ed. John Wiley & Sons, New York, NY.

Gavaris, S., and S.J. Smith.

1987. Effect of allocation and stratification strategies on precision of survey abundance estimates for Atlantic cod (*Gadus morhua*) on the eastern Scotian Shelf. J. Northw. Atl. Fish. Sci. 7:137-144.

Grosslein, M.D.

1969. Groundfish survey program of BFC Woods Hole. Comm. Fish. Rev. 31:22-30.

Gundersen, D.R.

1993. Surveys of Fisheries Resources. John Wiley & Sons, New York, NY.

Hansen, M.H., W.N. Hurwitz, and W.H. Madow.

1953. Sample Survey Methods and Theory. Volume 1. Methods and Applications. John Wiley & Sons, New York, NY.

Harbitz, A., M. Aschan, and K. Sunnanå.

1998. Optimal effort allocation in stratified, large area trawl surveys, with application to shrimp surveys in the Barents Sea. Fish. Res. 37:107-113.

Harbitz, A., and M. Pennington.

2004. Comparison of shortest sailing distance through random and regular sampling points. ICES J. Mar. Sci. 61:140-147.

Hartman, K.J. and S.B. Brandt.

1995. Predatory demand and impact of striped bass, bluefish, and weakfish in the Chesapeake Bay: applications of bioenergetics models. Can. J. Fish. Aquat. Sci. 52: 1667-1687.

Houde, E.D., M.J. Fogarty, and T.J. Miller.

1998. Prospects for multispecies management in Chesapeake Bay: A workshop. STAC Publication 98-002. Chesapeake Research Consortium, Edgewater, MD.

Jessen, RJ.

1978. Statistical Survey Techniques. John Wiley & Sons, New York, NY.

Jung, S.

2002. Fish community structure and the spatial and temporal variability in recruitment and biomass production in Chesapeake Bay. Ph.D. Dissertation, University of Maryland College Park, MD.

Jung, S., and E.D. Houde.

2003. Spatial and temporal variabilities of pelagic fish community structure and distribution in Chesapeake Bay, USA. Fish. Bull. 102:63-77.

Jung, S., and E.D. Houde.

2004. Recruitment and spawning-stock distribution of bay anchovy in Chesapeake Bay. Fish. Bull. 102:63-77.

Kalton, G.

1979. Ultimate cluster sampling. J. Royal. Stat. Soc. A 142, 210-222.

Kemp, W.M., J. Faganelli, S. Ouscaric, E.M. Smith, and W.R.Boyton.

1999. Pelagic-benthic coupling of nutrient cycling. *In*: Malone, T.C., Malej, A., Harding jr., L.W., Smodlaka, N., Turnber, R.E. (eds.). 1999. *Coastal and estuarine Studies: Ecosystems and the land-sea Margin Drainage Basin to Coastal Sea*. American Geophysical Union, Washington D.C. pp. 295-340

Kish, L.

1965. Survey Sampling. John Wiley & Sons, New York, NY.

Kish, L.

1995. Methods for design effects. Journal of Official Statistics 11:55-77.

Kish, L.

2003. Selected Papers. Edited by G. Kalton, and S. Heeringa. John Wileys & Sons. Wileys Series in Survey Methodology.

Kingsley, M.C.S., D.M. Carlsson, P. Kanneworff, and M. Pennington.

2002. Spatial structure of the resources of *Pandalus borealis* and some implications for trawl survey. Fish. Res. 58:171-183.

Korn, E.L., and B.I. Graubard.

1999. Analysis of Health Surveys. John Wiley & Sons, New York, NY.

Lehtonen, R., and E.J. Pahkinen.

1994. Practical Methods for Design and Analysis of Complex Surveys. John Wiley & Sons, Chichester.

Luo, J. and S.B. Brandt.

1993. Bay anchovy *Anchoa mitchilli* production and consumption in mid-Chesapeake bay based on a bioenergetics model and acoustic measurements of fish abundance. Mar. Ecol. Prog. Ser. 98:223-236.

Miller, T. J., E.D. Houde, and E.A. Watkins.

1996. Chesapeake Bay Fisheries: Prospects for multispecies management and sustainability. CRC Publication 154B. Chesapeake Research Consortium, Edgewater MD.

Pennington, M.

1985. Estimating the relative abundance of fish from a series of trawl surveys. Biometrics 41:197-202.

1986. Some statistical techniques for estimating abundance indices from trawl surveys. Fish. Bull. 84(3):519-525.

Pennington, M., and J.H. Vølstad.

1991. Optimum size of sampling unit for estimating the density of marine populations. Biometrics 47:717-723.

1994. The effect of intra-haul correlation and variable density on estimates of population characteristics from trawl surveys. Biometrics 50:725-732.

Polacheck, T., and J.H. Vølstad.

1993. Analysis of spatial variability of George Bank haddock (*Melanogrammus aeglefinus*) from trawl survey data using a linear regression model with spatial interaction. ICES Journal of Marine Science 50:1-8.

Potthoff, R.F., M.A. Woodbury, and K.G. Manton, K.G.

1992. "Equivalent sample size" and "equivalent degrees of freedom" refinements for inference using survey weights under superpopulation models." Journal of the American Statistical Association 87:383-396.

Rao, J.N.K.

2003. Small Area Estimation. John Wiley & Sons, New York, NY.

Research Triangle Institute (RTI).

2001. SUDAAN User's Manual, Release 8.0. Research Triangle Park, NC: Research Triangle Institute.

Seber, G.A.F.

1986. A review of estimating animal abundance. Biometrics 42:267-292.

Skinner, C.J.

1986. Design effects of two-stage sampling. J.R. Statist. Soc. **B** 48:89-99.

Skinner, C.J., D. Holt, and T.M.F. Smith.

1989. Analysis of Complex Surveys. John Wiley & Sons, New York, NY.

Snedecor, G.W., and W.G. Cochran.

1980. Statistical Methods. 7th ed. The Iowa State University Press. Ames, Iowa, U.S.A.

Särndal, C.E., B. Swensson, and J. Wretman.

1992. Model Assisted Survey Sampling. Springer-Verlag, New York, NY.

Smith,T.

2002. The Woods Hole bottom-trawl resource survey: development of fisheries-independent multispecies monitoring. ICES Marine Science Symposium 215:474-482.

Sukhatme, P.V. and B.V. Sukhatme.

1970. Sampling Theory of Surveys With Applications. 2nd edition. Iowa State University Press. Ames, Iowa.

Taylor, L.R.,

1961. Aggregation, variance and the mean. Nature 189,:732-735.

Williams, R.L.

2000. A note on robust variance estimation for cluster-correlated data. Biometrics 56:645-646.

Wolter, K.M.

1985. Introduction to Variance Estimation. Springer-Verlag. New York, NY.

Woodruff, R.S.

1971. A simple method for approximating the variance of a complicated estimate. J. Am. Stat. Assoc. 66:411-414.

Table 1

Mean catch-per-unit-effort (CPUE) and measures of precision and design effects for the 2002 stratified random surveys.

Season	**Species**	***deff***	***n***	n_{eff}	$\overline{y}_{STR}$	**SE**	**RSE**
Spring	Bay Anchovy	1.0	16	16	84.2	19.4	0.23
Summer	Bay Anchovy	1.0	19	19	570.8	143.6	0.25
Fall	Bay Anchovy	1.0	20	20	1541.8	445.6	0.29
Spring	Croaker	1.0	16	16	4.8	2.0	0.42
Summer	Croaker	1.0	19	19	1.6	0.7	0.44
Fall	Croaker	1.0	20	20	3.2	1.4	0.44
Spring	White Perch	1.0	16	16	0.6	0.5	0.83
Summer	White Perch	1.0	19	19	0.7	0.6	0.86
Fall	White Perch	1.0	20	20	0	0.0	
Spring	Weakfish	1.0	16	16	1.6	1.0	0.63
Summer	Weakfish	1.0	19	19	2.6	1.2	0.46
Fall	Weakfish	1.1	20	18	8.3	2.7	0.33
Spring	All Species	1.0	16	16	103.4	18.8	0.18
Summer	All Species	1.1	19	17	733.0	161.9	0.22
Fall	All Species	0.9	20	22	1643.6	442.0	0.27

Table 2
Mean catch-per-unit-effort and measures of precision and design effects for the 2002 transect surveys. The design effects (*deff*) are obtained from SUDAAN.

Season	Species	*deff*	*n*	n_{eff}	$\overline{y}_T$	SE	RSE
Spring	Bay Anchovy	1.3	29	22	98.0	27.1	0.28
Summer	Bay Anchovy	2.2	31	14	482.6	157.4	0.33
Fall	Bay Anchovy	3.0	31	10	1452	422.8	0.29
Spring	Croaker	1.5	29	19	3.3	1.1	0.33
Summer	Croaker	3.9	31	8	2.6	1.6	0.62
Fall	Croaker	0.7	31	44	7.8	2.2	0.28
Spring	White Perch	3.5	29	8	14.1	9.6	0.68
Summer	White Perch	4.2	31	7	4.8	4.1	0.85
Fall	White Perch	3.7	31	8	7.9	7.0	0.89
Spring	Weakfish	0.6	29	48	1.6	0.7	0.44
Summer	Weakfish	1.0	31	31	1.1	0.5	0.45
Fall	Weakfish	0.4	31	78	13.4	1.3	0.10
Spring	All Species	1.3	29	22	145.2	25.5	0.18
Summer	All Species	2.2	31	14	565.0	177.6	0.31
Fall	All Species	3.1	31	10	1557.8	419.6	0.27

Table 3. Mean catch-per-unit-effort and measures of precision and design effects for the 2002 combined stratified random and transect surveys. The estimates are obtained by using the optimal weights (ϕ_{opt}) in eqs. 1.10 and 1.11. R is the RSE for the composite estimate divided by the smaller RSE for the individual survey estimates.

Season	Species	*deff*	*n*	n_{eff}	$\overline{y}_C$	SE	RSE	ϕ_{opt}	R
Spring	Bay Anchovy	1.2	45	38	92.2	17.7	0.19	0.42	0.83
Summer	Bay Anchovy	1.5	50	33	533.4	106.2	0.20	0.58	0.79
Fall	Bay Anchovy	1.7	51	30	1511.9	335.0	0.22	0.67	0.77
Spring	Croaker	1.3	45	35	4.0	1.1	0.27	0.46	0.83
Summer	Croaker	1.9	50	27	1.9	0.7	0.36	0.70	0.82
Fall	Croaker	0.8	51	64	6.4	1.6	0.25	0.31	0.88
Spring	White Perch	1.9	45	24	5.1	3.2	0.63	0.67	0.93
Summer	White Perch	1.9	50	26	1.8	1.2	0.66	0.73	0.77
Fall	White Perch	1.8	51	28	2.3	2.0	0.89	0.71	
Spring	Weakfish	0.7	45	64	1.6	0.6	0.36	0.25	0.83
Summer	Weakfish	1.0	50	50	1.7	0.6	0.33	0.38	0.73
Fall	Weakfish	0.5	51	96	12.4	1.2	0.09	0.19	0.94
Spring	All Species	1.2	45	38	127.6	16.7	0.13	0.42	0.73
Summer	All Species	1.6	50	31	657.1	119.8	0.18	0.55	0.82
Fall	All Species	1.6	51	32	1616.8	332.0	0.21	0.69	0.76

Table 4. Mean catch-per-unit-effort (CPUE) and measures of precision and design effects for the 2003 stratified random surveys.

Season	Species	*deff*	*n*	n_{eff}	$\overline{y}_{STR}$	SE	RSE
Spring	Bay Anchovy	0.7	20	29	91.8	17.6	0.19
Summer	Bay Anchovy	0.6	20	32	570.8	143.6	0.25
Fall	Bay Anchovy	1.0	9	9	1113.6	456.5	0.41
Spring	Croaker	0.8	20	27	1.6	0.5	0.31
Summer	Croaker	0.9	20	22	4.2	2.0	0.47
Fall	Croaker	1.0	9	9	0.1	0.1	1.00
Spring	White Perch	0.7	20	27	0.7	0.4	0.60
Summer	White Perch	1.1	20	18	1.0	1.0	1.00
Fall	White Perch	1.0	9	9	0.2	0.2	1.00
Spring	Weakfish	0.9	20	22	0.2	0.1	0.55
Summer	Weakfish	1.0	20	20	0.2	0.1	0.53
Fall	Weakfish	1.0	9	9	1.3	1.2	0.91
Spring	All Species	0.6	20	31	59.3	11.2	0.19
Summer	All Species	0.2	20	91	123.0	13.0	0.11
Fall	All Species	1.0	9	9	1122.9	455.4	0.41

Table 5. Mean catch-per-unit-effort and measures of precision and design effects for the 2003 transect surveys. The design effects (*deff*) are obtained from SUDAAN.

Season	Species	*deff*	*n*	n_{eff}	$\overline{y}_T$	SE	RSE
Spring	Bay Anchovy	1.1	31	30	73.7	21.3	0.29
Summer	Bay Anchovy	1.9	31	16	64.7	20.3	0.31
Fall	Bay Anchovy	2.1	20	9	1228.9	378.5	0.31
Spring	Croaker	1.0	31	32	3.0	1.5	0.51
Summer	Croaker	1.0	31	33	1.7	1.0	0.57
Fall	Croaker	1.7	20	12	0.2	0.2	0.94
Spring	White Perch	2.7	31	11	17.3	13.4	0.78
Summer	White Perch	6.3	31	5	51.7	6.3	0.12
Fall	White Perch	1.9	20	11	12.3	10.1	0.82
Spring	Weakfish	0.6	31	53	0.3	0.2	0.61
Summer	Weakfish	1.0	31	32	0.8	0.5	0.60
Fall	Weakfish	0.8	20	24	1.2	0.6	0.52
Spring	All Species	1.1	31	29	103.2	22.2	0.22
Summer	All Species	4.8	31	6	223.2	108.9	0.49
Fall	All Species	2.4	20	8	1412.7	529.8	0.38

Table 6. Mean catch-per-unit-effort and measures of precision and design effects for the 2003 combined stratified random and transect surveys. The estimates are obtained by using the optimal weights (ϕ_{opt}) in eqs. 1.10 and 1.11. R is the RSE for the composite estimate divided by the smaller RSE for the individual survey estimates.

Season	Species	*deff*	*n*	n_{eff}	$\overline{y}_C$	SE	RSE	ϕ_{opt}	R
Spring	Bay Anchovy	0.9	51	59	82.7	13.8	0.17	0.50	0.9
Summer	Bay Anchovy	1.1	51	48	402.6	96.1	0.24	0.67	0.9
Fall	Bay Anchovy	1.6	29	18	1172.8	295.2	0.25	0.49	0.8
Spring	Croaker	0.9	51	59	2.4	0.9	0.37	0.45	1.2
Summer	Croaker	0.9	51	55	2.7	1.0	0.36	0.41	0.8
Fall	Croaker	1.4	29	21	0.1	0.1	0.72	0.43	0.8
Spring	White Perch	1.3	51	39	5.6	4.0	0.71	0.71	1.2
Summer	White Perch	2.2	51	23	11.9	1.6	0.13	0.78	1.1
Fall	White Perch	1.5	29	20	6.8	5.5	0.81	0.46	1.0
Spring	Weakfish	0.7	51	75	0.3	0.1	0.49	0.29	0.9
Summer	Weakfish	1.0	51	51	0.5	0.3	0.54	0.38	1.0
Fall	Weakfish	0.9	29	33	1.2	0.6	0.46	0.27	0.9
Spring	All Species	0.9	51	60	80.3	12.1	0.15	0.52	0.8
Summer	All Species	0.5	51	97	129.6	14.1	0.11	0.93	1.0
Fall	All Species	1.7	29	17	1263.1	347.8	0.28	0.52	0.7

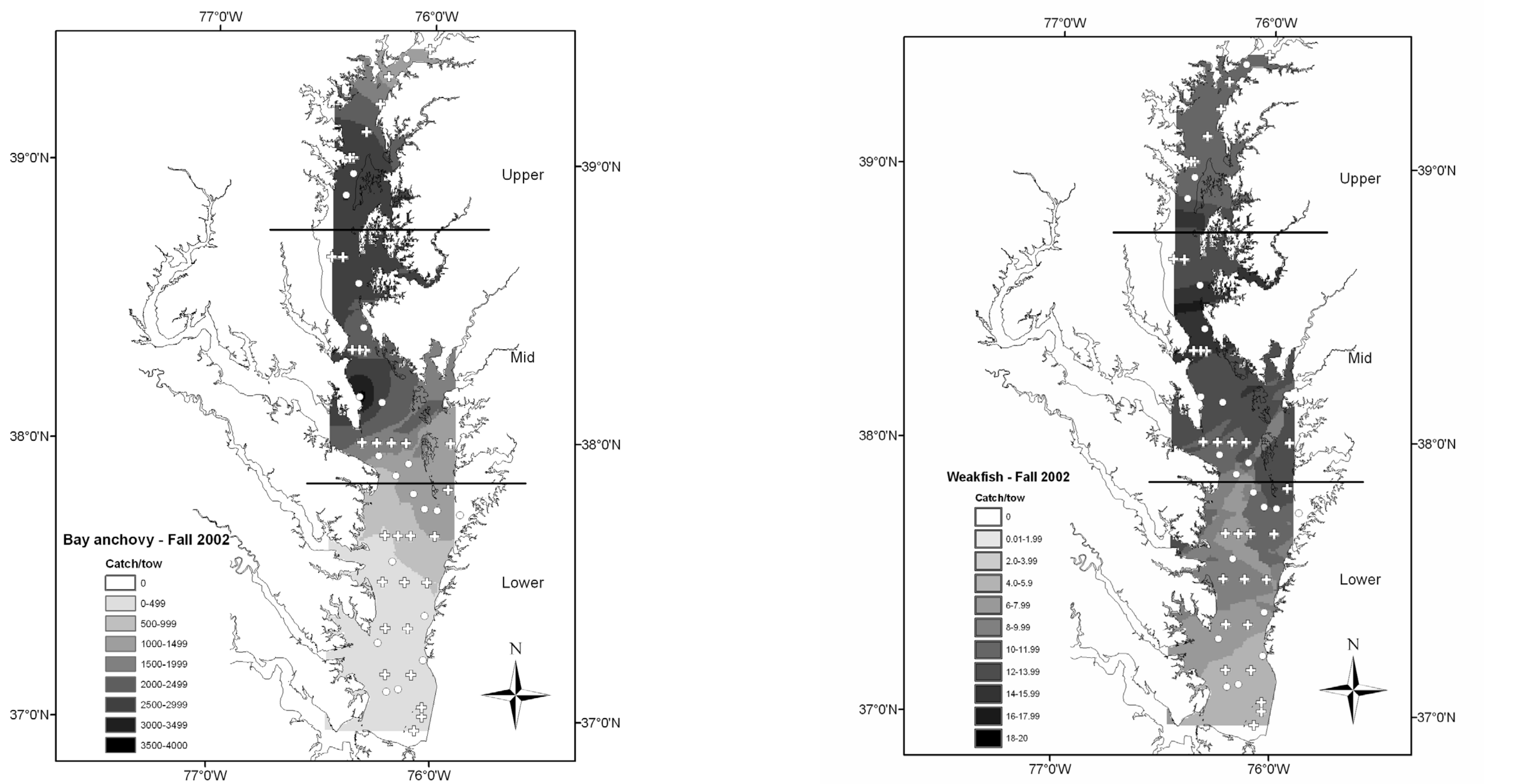
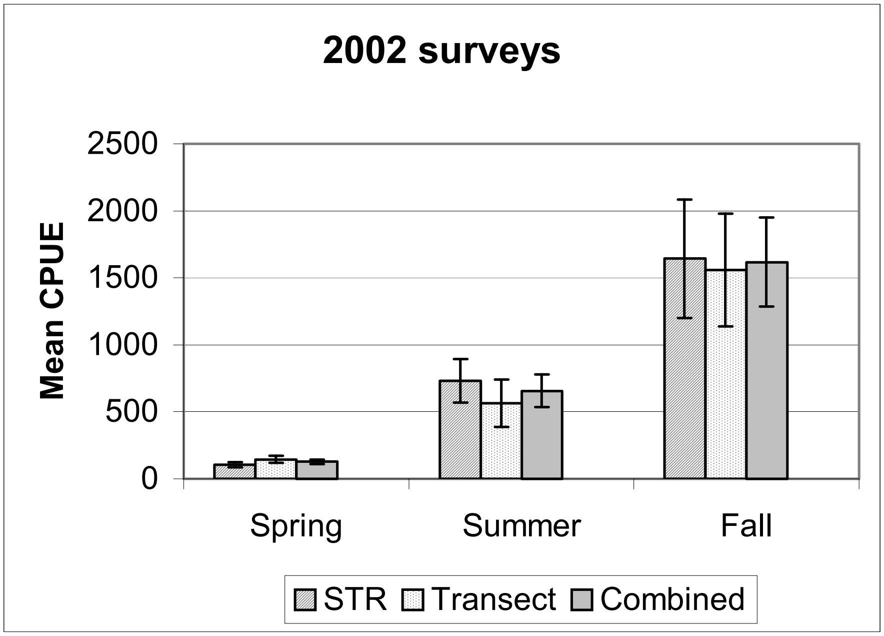


Figure 1. Stratification of the Chesapeake Bay mainstem (Lower, Mid, and Upper Bay), and an example of station allocation for the transect survey (╬) and stratified random survey (o) and the density distribution of (a) Bay anchovy and (b) weakfish during fall 2002.



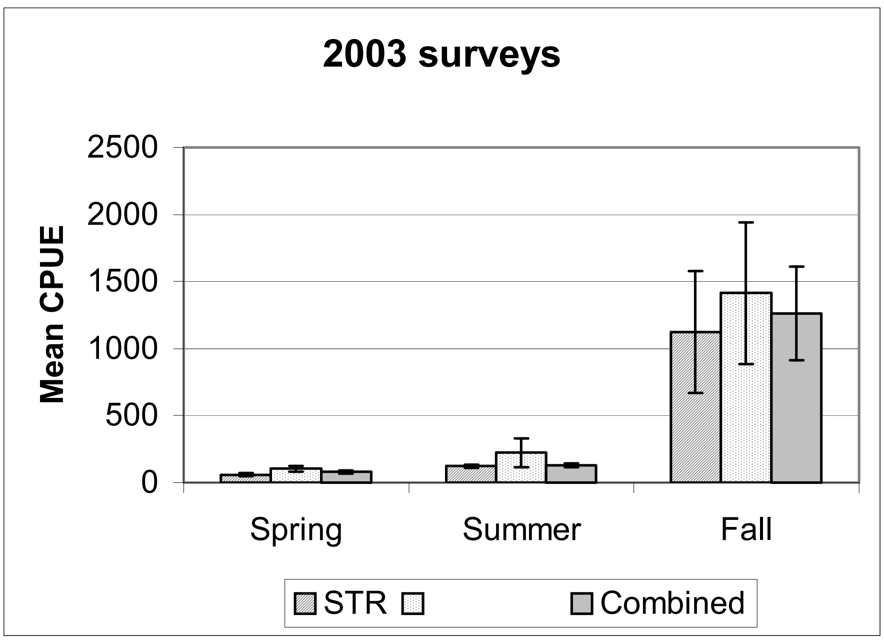


Figure 2. Seasonal design-based estimates of mean catch per unit effort (cpue) across all species, and the combined estimates across surveys based on the composite estimator. Error bars represent ± SE.

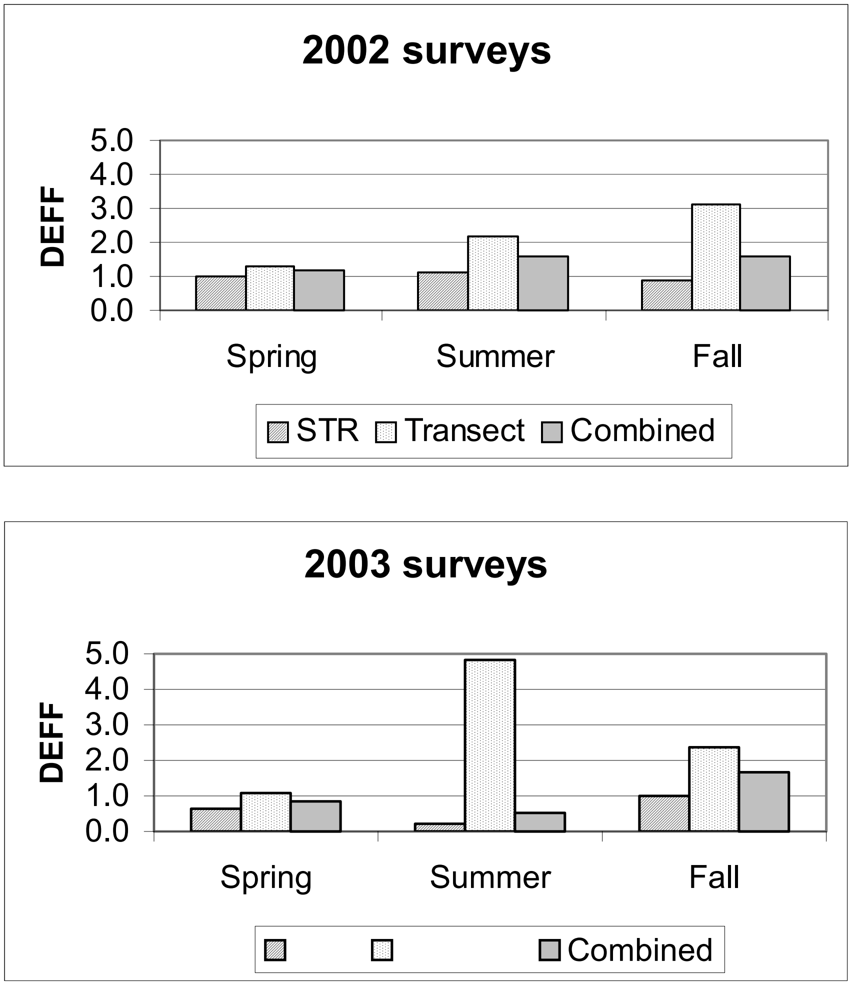
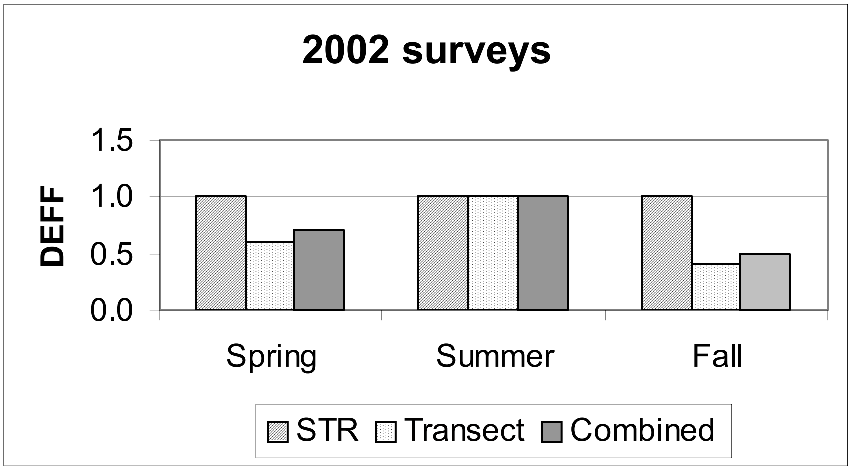


Figure 3. Design effects for estimating seasonal mean catch per unit effort (CPUE) across all species by survey. The combined estimate is based on the composite estimator.



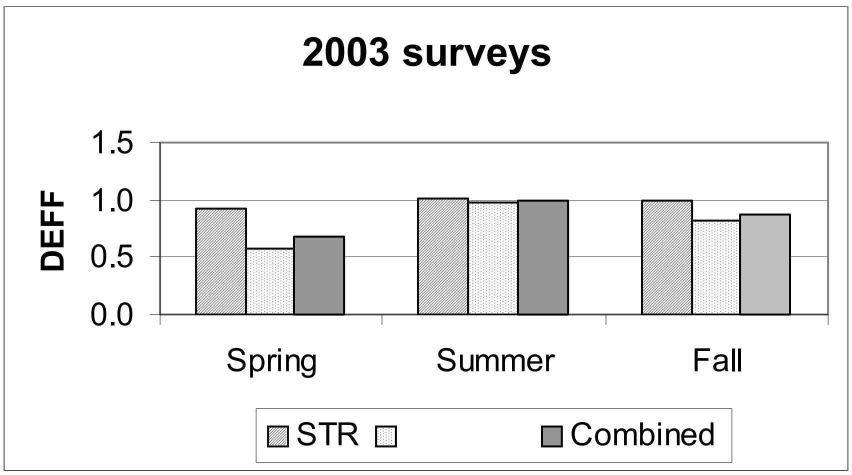


Figure 4. Design effects for estimating seasonal mean catch per unit effort (CPUE) of weakfish by survey. The combined estimate is based on the composite estimator.

Appendix A

The optimum weight, obtained by minimizing equation (0.11) with respect to ϕ (Rao 2003) is

$$\phi_{opt} = \frac{V(\overline{Y}_T)}{V(\overline{Y}_{STR}) + V(\overline{Y}_T)}$$

when the two surveys are independent. By definition, the variance of the estimated mean CPUE from each survey can be expressed by dividing the population variance of CPUE under simple random sampling, σ_{SRS}^2, with the effective sample size,

$$V(\overline{Y}_T) = \frac{\sigma_{SRS}^2}{n_T^*} \quad \text{(A.1)}$$

and

$$V(\overline{Y}_{STR}) = \frac{\sigma_{SRS}^2}{n_{STR}^*} \quad \text{(A.2)}$$

where n_T^* and n_{STR}^* are the effective sample sizes for the transect and stratified random surveys, respectively. Replacing $V(\overline{Y}_T)$ and $V(\overline{Y}_{STR})$ in ϕ_{opt} with their equivalents from (A.1) and (A.2) yield the desired result.

Working document for Workshop on Survey Design and Analysis (WKSAD) 21-25 June 2004, FRS, Aberdeen, Scotland, UK.

Comparing survey designs and estimators: an example using icthyplankton data.

Claire Imrie[1], Iago Mosqueira, Doug Beare[2], Dave Reid, Anna Korre, Murdoch McAllister.

1. **Imperial College, Department of Environmental Science and Technology, Royal School of Mines, Prince Consort Road, London, SW7 2BP, London. Email: c.imrie@ic.ac.uk.**
2. **Fisheries Research Services, Marine Laboratory, Victoria Road, Torry, Aberdeen. Email: d.beare@marlab.ac.uk.**

INTRODUCTION

Overview and objectives

Atlantic mackerel fisheries are economically important in EU waters with annual gross landed values of approximately 50 million Euros. A major tool for the assessment of these stocks, which forms the main scientific basis for these fisheries management, is a triennial survey of pelagic mackerel eggs. From these surveys, absolute estimates of spawner abundance are obtained. However, recent research (ICES, 1999) has suggested that there may be large biases in survey estimates of abundance due partly to incomplete survey coverage of the spatial and temporal extent of pelagic eggs in the spawning season. It is also unclear whether the estimation methods applied account adequately for uncertainties in the estimates of spawner abundance and whether the current management approach is sufficiently robust to address these uncertainties.

The pelagic egg survey is to date the only fisheries independent data source available for the mackerel and horse mackerel stocks. The main objective of these surveys is the establishment of egg abundance, which can then be used on its own as an index of stock abundance. In this case, the egg abundance is entered into a stock assessment model that applies an age-structured population estimation method, VPA. Therefore, the egg survey data is used as one of the input parameters, together with fecundity and sex ratio, to estimate biomass providing an absolute stock abundance. The main objectives of the project were:

- to combine geostatistical and Bayesian statistical methods to improve the scientific basis for the management of Atlantic mackerel stocks,
- to apply Bayesian decision theory to evaluate the potential consequences for fishery management of applying both total allowable catch (TAC) and spatial controls, and to assess the information gathering requirements of these controls.

The fundamental research carried out in this project dealt with the sequential survey data on pelagic egg densities and aimed to improve the estimates of egg production from this data and better account for uncertainty. The research did not specifically aim to improve the stock biomass estimation methods and did not consider parameters such as fecundity and sex ratio. However, an improved estimate of egg production would potentially have a positive impact on the output of models currently used to estimate stock abundance. Based on this principle, this project also aimed at quantifying the impact of improved egg production estimates on the stock biomass estimates obtained through conventional modelling techniques which use the egg abundance estimates as one of the input parameters.

In this project, geostatistical and Bayesian estimation methods were combined in attempt to reduce bias in estimates of egg abundance and to better account for uncertainties in these estimates. subsequently, Bayesian decision analysis methods were applied to identify fishery control measures and information gathering and estimation methods that should ensure that the management methods applied are adequately robust to deal with the uncertainties and conform to the recently adopted precautionary guidelines for fishery management in the CFP. This project is unique because it is the first to combine Bayesian and geostatistical estimation methods to improve the scientific basis for fisheries management. Furthermore, this project is the first one to apply, explicitly, the Bayesian decision theory to evaluate the potential consequences for fisheries management of applying both TAC and spatial controls as well as assessing the information gathering requirements of these controls.

Background

During the mid 1960s there was a rapid increase in catches of mackerel in European waters due to advances in fishing technology, a crash in herring stocks, and a switch in fishing effort. In the UK alone, by 1979, mackerel accounted for 42% of the total fin-fish catch (Priede and Watson, 1993). In the Northeast Atlantic, two major stocks have been distinguished: the Western stock which spawns west of the British Isles and the North Sea stock which spawns in the Northern North sea. The North Sea stock was severely depleted as of 1990, whereas the western stock had a potential yield of about 500,000 tonnes (Priede and Watson, 1993). Mackerel (*Scomber scombrus*), and horse mackerel (*Trachurus trachurus*) of the wide ranging western stocks account for a total annual catch of 500,000 tonnes and 100,000 tonnes respectively. The fishery is managed by the setting of a total allowable catch (TAC). This TAC is agreed by International Council for Exploration of the Seas (ICES) and the EU for fisheries in national waters. The TAC is set according to a stock assessment that applies an age-structured population estimation method, VPA, which incorporates catch-age data and triennial estimates of spawner abundance from annual pelagic egg surveys. An area closure Southwest of the UK, the Cornish Box, protects juveniles and a closure of the North Sea during the first quarter of each year protects over-wintering adults. However, an important nursery area to the Northwest of Ireland still remains open.

The assessment of fish population abundance from hydroacoustic or ichthyoplankton survey data is often complicated by the spatial and temporal correlation among observations. Several approaches have been taken to deal with this problem, including classical statistics (stratified random sampling) and those based on time series analysis. Geostatistical estimates have been used to assess fish population abundance from hydroacoustic survey data (Sullivan, 1991) and trawl measurements of density at distinct stations (Armstrong *et al*., 1992; Petitgas, 1993; EU project FAIR21007). The research carried out examined the spatial structure of the data aiming to reduce the uncertainty in abundance estimation and consequently improve the estimates.

The primary tool for assessment of mackerel abundance is the pelagic egg survey, together with histological estimates of fecundity by weight, from which estimates of spawner abundance are obtained. This survey is very costly and is hence done once every three years. During the spawning season, successive comprehensive spatial surveys of the density of pelagic eggs are carried out. From these observations, absolute estimates of spawner biomass are obtained (ICES, 1996). The methods applied include a traditional estimator that assumes a relatively simple model to approximate the abundance of eggs over the season. More recently, a general additive model (GAM), which can allow more sophisticated models of the abundance of eggs over the season and account for the effects of environmental variables on observations of egg densities, has been explored (Anon., 1999, EU project CFP 970097). Earlier studies have shown that due to a variety of reasons such as incomplete spatial coverage of the extent of the pelagic eggs, both of these estimation methods can result in significant biases in estimates of spawner biomass (Augustin *et al*., 1998).

The problems encountered with traditional estimates lie in the fact that some of the areas and times are incompletely covered, and as a result the produced estimates may be significantly biased. In addition to this the traditional estimators use quite long periods, in the order of one month so we lose a lot of fine scale variation. Generalised Additive Models (GAMs) were introduced to try solving these problems. However, the dependence of the GAM estimators on so-called structural zeroes (imaginary zero values set beyond the expected spatio temporal boundaries) tends to introduce negative bias.

In this project the aim was to produce unbiased estimates of TAEP(total annual egg production) with accurate measures of precision by combining the geostatistical and Bayesian analysis approaches. Additionally, the estimators were developed taking the temporal-time series aspect into consideration. The geostatistical approach offered the additional benefit of facilitating the development of a methodology to design the egg survey. Finally, decision theoretic methods have been developed and applied to evaluate the potential consequences of alternative western and horse mackerel fishery management options that can potentially help fishery managers to more effectively achieve their fishery management objectives. A detailed discussion of the methodology that was adopted and developed is presented in the following section. The objectives of the project were achieved through the following tasks:

1. Data collection and compilation - GIS database design.
2. Geostatistical estimator development for the spatio-temporal modelling of egg survey data.
3. Incorporation of geostatistical and Bayesian analysis techniques for egg survey data modelling.
4. Comparison of new geostatistical and combined geostatistical - Bayesian estimators with conventional design based techniques.

5. Optimal egg survey design - Evaluation of the expected changes in fishery yield and management options from improved survey estimates.
6. System review and final synthesis.

Background to geostatistics and its use in fisheries

In brief, geostatistics makes use of the spatial autocorrelation structure within a dataset to make unbiased estimates of the variable of interest. Its use has been particularly interesting to stock assessment scientists in recent years due to its generation of corresponding estimation variances (e.g., Petitgas and Lafont, 1997; Rivoirard *et al*., 2000).

Geostatistics is concerned with the study of phenomena that vary in space and/or time. An unsampled value *z* is treated as a random variable *Z*, the uncertainty of which is modelled by a probability distribution. Most of the information regarding the value of *Z* at a specific location is derived from the sample values at neighbouring locations. The extent of the dependence of *Z* on these sampled values can be obtained through the use of a tool called the semivariogram (termed variogram from now on), which describes the spatial autocorrelation present in the data. A model is fitted to the experimental variogram, and this is used to allocate weights to the surrounding data points that are used to estimate the value of a point or block. The experimental variogram is calculated as follows:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \left[z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h}) \right]^2$$

where $\mathbf{h}$ is the separation vector and $\mathbf{x}_i$ is a point in space. An omnidirectional variogram assumes that the spatial correlation of the variable is isotropic. Multidimensional directional variograms can be modelled to represent any anisotropy. Variograms may be modelled using a positive linear combination of certain structures, such as a nugget effect, which accounts for measurement error and unobserved small-scale variation, and/or any spherical, exponential or Gaussian model components.

The most commonly used geostatistical estimation technique is called ordinary kriging (e.g., Deutsch and Journel, 1998). This aims to provide a 'best linear unbiased estimate' by seeking a residual error of zero and minimising the error variance. Ordinary kriging is generally undertaken using a 'moving neighbourhood' so that local means of the variable are used in the estimations, thereby allowing the prerequisite assumptions of stationarity to be relaxed. Block kriging can be used to obtain mean values of a variable in cells in a regular grid, or the mean of a variable within a polygon (Petitgas and Lafont, 1997).

The paragraphs above briefly describe the most commonly applied geostatistical tools, which can be termed 'intrinsic' methods. However, an alternative branch of geostatistics referred to as transitive methods has been used in a number of fisheries applications, including the analysis of mackerel egg data. In this process each sample is given a weight proportional to its density. The relative abundance per class of parameter is used to provide the probability distribution of the parameter per individual, with its mean (centre of gravity) and variance (inertia). A level and an index of aggregation have been proposed to characterise statistically the population while the inertia of locations and the transitive covariogram were used to define as the spatial characteristics of the population (Bez and Rivoirard, 2000). One limitation of this approach is that the mean and the variance is estimated experimentally, weighted by the density values, based on the assumption of a regular sampling design. Such limitations, due to the assumptions required for this type of modelling, necessitate that the estimates are calculated in one dimension, east-west, enhancing the importance of the edge of the continental shelf for the spatial pattern, and north-south separately. These two estimates are subsequently averaged to compute the mean (centre of gravity) of the abundance per class of parameter.

Transitive geostatistical methods have been applied to mackerel ichthyoplankton in order to characterise their spatial aggregation patterns (Bez and Rivoirard, 2001). A global estimation for mackerel egg production was calculated by Bez (2002) for the 2nd period of the 1989 survey. However, this assumed that the data points were synoptic and temporal variability within the 22-day period was not considered. Williamson and Traynor (1996) applied transitive geostatistical theory to acoustic surveys of Alaskan pollock. Doonan *et al*. (2003) found that applying a polar version of transitive kriging gave better results than ordinary kriging when assessing the benefits of using star acoustic surveys of localised fish aggregations.

The transitive method can be attractive because it applies to a finite domain beyond which the variable is equal to zero (Rivoirard *et al*., 2000). However, its application is restricted to datasets which are based on a regular or almost-regular sampling pattern, and may only be applied to two-dimensional datasets (Bez, 2002).

More numerous applications of geostatistics in fisheries research have been undertaken using intrinsic geostatistics. For example, Rivoirard *et al*. (2000) describe a number of case studies, including the abundance of cod in the Barents Sea from trawl survey data. Romaine *et al*. (2002) estimated euphausiid population size, although Murray (1996) found a similar application to Antarctic krill was thwarted by the high skewness of the dataset. Maynou *et al*. (1998) evaluated biomass populations of Norwegian lobster in the northwestern Mediterranean. Páramo and Roa (2003) assessed habitat abundance relationships of small pelagic fish from the Colombian Caribbean, while stock estimates of scallop in Uruguayan waters were made by Gutiérrez and Defeo (2003) using ordinary block kriging. Petitgas *et al*. (2003) compared commercial and research survey catch per unit of effort with regard to megrim in the Celtic Sea.

When the data are considered to be strictly non-stationary, an 'external drift' can be included in the analysis (Matheron, 1971). Petitgas (1997) modelled spatio-temporal sole egg distributions by removing the non-stationary mean (trend) and calculating variograms from the residuals. Rivoirard and Weiland (2001) performed ordinary kriging for the estimation of juvenile haddock, while including an external drift to account for the effect of daylight on the trawl survey data.

For the majority of applications listed above, an estimate of uncertainty has been calculated to accompany the stock abundance estimates. In all intrinsic cases this has been based on the kriging variance. However, this is no straightforward procedure, mainly due to the complex shapes of the extent of the species and non-regular survey designs. A special software, EVA, was developed by Petigas and Lafont (Version 2: 1997) to enable fisheries scientists to compute variances along with global estimations, a feature that is unavailable in general commercial geostatistical software.

One major disadvantage of the above mentioned techniques is that time is not taken into account during the estimation process. As noted above, a global estimation for mackerel egg production calculated by Bez (2002) for the 2nd period of the 1989 survey was based on the assumption that the data points were synoptic and temporal variability within the 22-day period was not considered. While the methodology could be extended to a 3D system with time as the third dimension, the spatio-temporal sampling pattern could not be considered regular.

Background to Bayesian geostatistics

The science of fisheries stock assessment has a long history in providing quantitative advice on the health and exploitation possibilities of fish stocks. In doing so its practitioners have had to deal with considerable uncertainty regarding our knowledge of the dynamics of fish populations (Hilborn and Walters, 1992). In recent years, an increasing trend can be detected in the use of Bayesian methods in fisheries assessment and management as an approach for incorporating uncertainty in both models and their parameters. Other methods have also attempted to deal with uncertainty in fisheries models (McAllister and Kirkwood, 1998), most notably: sensitivity analyses, confidence bounds, and data resampling procedures such as jackknife and bootstrap.

The advantages of Bayesian methods span several levels. Firstly, the Bayesian paradigm is theoretically consistent and logically coherent (Smith and Bernardo, 2000) as it makes explicit use of probability theory. Secondly, it is a rigorous method to provide weights to alternative parameter values in a model, both in continuous and discrete settings. The interpretation of its output is more congruent with common-sense; for example a Bayesian (probability) interval for an unknown quantity can be considered as having high probability of containing the unknown quantity, in contrast to the convoluted interpretation of frequentist (confidence) intervals (Gelman *et al*., 1995). Other grounds for favouring the use of Bayesian methods include the fact that previous information, from other studies or similar settings, should be used in the analysis, something that can be accomplished through the construction of prior probability distributions, and not simply discarded (Berger, 1985). It also provides decision analytic models with coherent inputs. The use of decision analysis has grown steadily in natural resources management (Ellison, 1996) and Bayesian analysis provides the tool needed for its efficient implementation.

Bayesian analysis generally starts with the construction of prior probability distributions (priors), as summaries of the information held or gathered by the researcher, previous to the analysis, on the possible states of nature. A wide range of information sources can be integrated, from surveys or experiments conducted in previous years, to expert opinions and beliefs (O'Hagan, 1998). Once probabilities have been assigned to the possible values of the parameters involved, point estimates are substituted by discrete or continuous probability distributions into the model, and data fitted by means of a likelihood function. A joint posterior probability distribution can then be generated, its values sampled by Markov Chain Monte Carlo, Sampling Importance Resampling algorithms, or grid methods (Gelman *et al*., 1995). By integrating over other estimated parameters, a probability distribution for parameters of interest can be estimated.

In the last decade, Bayesian methods have been increasingly applied in fisheries assessment and management (see, for example, the reviews by Punt and Hilborn, 1997 and McAllister and Kirkwood, 1998). Achieving the main objective of any stock assessment, to ensure sustainable production over time from fish stocks (Hilborn and Walters, 1992), typically requires the setting of quantitative limits to the fishing activity. Those management decisions should be based on quantitative predictions about the potential biological and economic consequences of alternative management options. In doing so, attention has focused increasingly on Bayesian methods as a particularly suitable approach for providing such advice to fishery managers. Although Bayesian methods have been in use for many years (Berger, 1985), it was only as computing power became widely available that scientists turned to Bayesian methods and other computer-intensive methodologies.

Different topics in stock assessment have been analysed applying the Bayesian approach. Examples of implementations relate to the choice of survey design in trawl surveys (McAllister and Pikitch, 1997), the fitting of age-structured models for New Zealand Hoki (Hilborn *et al*., 1994), analysing the role of environmental factors in the effect of management options on Baltic cod (Kuikka *et al*., 1999), and calculation of the probability of depensation, over-reaction of the population to excessive fishing pressure, by gathering information from many stocks and taxons (Liermann and Hilborn, 1997).

Prior to GBMAF, Bayesian methods had never been applied to the analysis of pelagic egg surveys, despite their wide use for stock assessment, the large variability and uncertainty in the present estimation procedures, and the economic importance of pelagic stocks (particularly sardine, anchovy, mackerel and horse mackerel) worldwide.

Despite the frequent use of the method-of-moment variogram estimator (Matheron, 1965), a precise quantification of the uncertainty of the estimator is rarely provided (Bogaert, 1999). Typically, variogram selection and modelling are done on a subjective basis without considering the variability of the estimators. However, points on an experimental variogram are often calculated on the basis of a small number of data points, particularly when the sampling pattern is irregular, resulting in a high variability.

While the variograms presented in this project have all been modelled manually, many researchers favour the use of a weighted least squares procedure (Cressie, 1993). In the weighted least squares method, the weights are derived from the number of pairs in each variogram distance class. A number of studies on characterising the uncertainty associated with the fitting procedure have been reported. For example, Bogaert (1999) used an analytical approach based on the properties of the characteristic functions in the frequency domain to derive probability density functions for each distance class of the experimental variogram. He then used a Taylor expansion to obtain the covariance matrix of the parameter estimators for the variogram model.

Pilz *et al*. (1996) attempted to take into account uncertainty in spatial covariance estimation by modelling a whole class of plausible variogram functions instead of fitting a single model. They used spectral representations to specify the variograms then proposed a new kriging method to find the linear spatial interpolator which minimised the maximum possible kriging variance with respect to all the plausible variograms.

Safai-Naraghu and Marcotte (1996) proposed a bootstrap procedure for assessing uncertainties associated with variogram parameters. The experimental variogram is calculated as normal, and then sampling with replacement is performed for each distance class to obtain a bootstrap variogram. The parameters of the model (nugget, range and sill) are then fitted by weighted least squares. This procedure is then repeated many times to obtain a bootstrap distribution for the model parameters. Conditional simulations were performed to assess the practical impact of accounting for the uncertainty in the variogram parameters.

Pardo Iguzquiza (1999) compared a number of combinations of estimators and priors of a covariance function under the Bayesian paradigm. The range of the variogram is estimated under different scenarios and for small sample sizes (15 to 30). The results were in general highly dependent on the choice of prior, although good results could be obtained even by the use of uniform prior probabilities. Greater improvements are anticipated if objective prior information is available.

Geostatistical kriging methods make the implicit assumption of a Gaussian model of stochastic variation in the data (Diggle *et al*., 1998). This might not be reasonable in many situations, and various attempts have been made to overcome this limitation by incorporating uncertainty in the estimation of both the systematic and stochastic components of the model (e.g., Le & Zidek, 1992; Handcock & Stein, 1993; Diggle *et al*., 1998). Two main approaches can be identified: one that is limited to Gaussian-based models (Omre, 1987; Omre & Halvorsen, 1989; Pardo Iguzquiza, 1999; Ribeiro & Diggle, 2000); and one which is model-based (Diggle *et al*., 1998).

A simple Bayesian analysis of kriging is proposed by Handcock & Stein (1993). A parametric representation of the covariance structure can be modelled under a Bayesian framework to account for error in the estimation, by interpreting ordinary kriging as Bayesian "with the non-informative prior for the mean parameter". The assumption of a known covariance structure is then relaxed, although it is considered to be a member of a parametric class. The optimality of the kriging procedure, based on a known covariance structure, can then be tested in a situation whereupon that structure is in fact estimated. Because the probability distributions of all parameters are members of the normal family, analytical integration is possible. This makes the procedure relatively fast and straightforward.

As noted above, various sources of uncertainty occur in any geostatistical modelling exercise. The most important source resides in the estimation of the variogram model parameters. Although a weighted least squares procedure may be used to fit a variogram model, the modelling process still relies on a series of subjective decisions, which are based on experimental or ancillary information (Goovaerts, 1997):

- Whether to fit an isotropic (omnidirectional) or anisotropic variogram;
- Which number and type (i.e., nugget, spherical, exponential, etc.) of basic variogram model structures to use;
- Which parameters (i.e., sill, anisotropy, range, etc.) to use for each model component.

Background to simulation studies for comparisons of estimators and survey designs

A number of previous publications describe research undertaken to compare population abundance estimators and sampling strategies. For example, Defeo and Rueda (2002) analysed the issues of spatial structure, sampling design and abundance estimation in sandy beach macroinfauna. They defined two different types of sampling: species driven, whereby the samples are taken regularly along samples; and environmentally driven, whereby the samples are taken according to the position of beach features such as tidal marks. The abundance of clams and isopods was tested using linear interpolation and block kriging. They noted that the use of geostatistics represented an important advance in overcoming pervasive shortcomings in sampling design. However, they found that surveys that follow random sampling schemes among *a priori* fixed strata, while they are devised to satisfy the assumptions of random sampling theory, are suboptimal for a geostatistical approach.

Simmonds and Fryer (1996) considered the design of acoustic surveys for estimating the mean abundance of spatially correlated populations, using North Sea herring as an example. They first analysed survey data to determine the spatial structure, then used this to develop a number of population models. They then generated 1000 realisations of each model, with differing statistical properties. These surfaces possessed local positive correlation, a short-scale random component, and a non-stationary or trend component. Each realisation was subsampled using eight different sampling strategies, and each of the resulting datasets were used to obtain mean population estimates and the corresponding variance. They first tested a simple random survey, followed by five increasingly stratified surveys, then a systematic survey with a random starting point, and finally a systematic survey with a centred starting point. The mean abundance was calculated using the overall sample mean. However, the error variance was estimated in three ways: pooled within strata variance; geostatistical estimation variance using a spherical model with nugget; and geostatistical estimation variance using an exponential model with nugget. The authors found that precision generally increased with increasing stratification, which agreed with theory. According to Matheron (1971), for an infinite population, a stratified random survey with equally sized strata and an equal sampling allocation to strata always has a smaller (or equal) error variance than a simple random survey. Both geostatistical estimators gave similar variance values.

Doonan *et al*., 2003 investigated the use of 'stars' as an alternative design for acoustically surveying isolated fish aggregations and concluded that they were a robust and effective way of estimating biomass while minimising vessel time and yielding good precision. A number of simulated surfaces were generated with varying complexities. A number of abundance and variance estimators were tested, and although abundances estimates were generally consistent regardless of the method, the estimates of variance differed considerably. It was supposed that if the estimated variances accurately reflected the true errors, they would be close to the true RMSE. The 'best' variance estimations were made using a polar-coordinate version of transitive geostatistics.

Brus *et al*. (2002) used a procedure called simulated annealing to find designs with minimum sampling variance for a fixed budget. Simulated annealing is a search technique that starts off with random values (e.g. of sampling units) and iteratively alters these numbers until a target function is minimised. This approach can be used when prior information on the spatial variation of the target variable is available.

In the field of soil quality, van Groenigen (2000) used spatial simulated annealing to investigate the influence of variogram parameters on optimal sampling schemes for mapping by kriging. He defined a grid with a small number of randomly located observations and discretised it finely for point kriging. He specified two optimisation criteria: minimising the average kriging variance; and minimising the maximum kriging variance. The simulated annealing algorithm was used to optimise the location of 10 additional samples with regard to each criterion. To minimise the average variance, the extra samples were generally placed in the largest unsampled areas. To minimise the maximum variance, the samples were added preferentially towards the edges of the area. The author also found that different variogram shapes (e.g. Gaussian, exponential, spherical) gave rise to different optimal sampling schemes. However, although this approach is simple and effective, it focuses on minimising variances that are independent of the actual values of the variable.

Unfortunately, most of the publications do not provide a detailed description of how the data was simulated for their experiments. An exception is Simmonds and Fryer (1996), who used the following method. The nugget effect was simulated using a normal distribution with mean 0 and variance 1. The local autocorrelation was reproduced using an autoregressive series with a variable range parameter. A non-stationary component was generated by three different methods: a random walk; a linear trend; and a cosine trend. The coefficients of the nugget, autocorrelation and non-stationary components were varied to allow simulation of a number of different variographic scenarios.

In general, the reported studies involved comparisons of abundance and variance estimates using two-dimensional population distributions. Furthermore, they were able to use survey designs which were able to cover the full extent of the species under consideration, and which were assumed to be reasonably regular (systematic or random stratified, whereby there is a degree of regularity). The problem considered here concerned a spatio-temporally distributed variable which is typically sampled in pseudo-regular patches across space and time.

MATERIALS AND METHODS

Development of the geostatistical estimator

The following pages describe the progress made towards developing a methodology for the estimation of TAEP for mackerel and horse mackerel, with corresponding measures of estimation uncertainty. The work started with a comprehensive review of the measurement and surveying techniques, as well as a review of the related literature. This was followed by an initial geostatistical analysis which served to highlight some of the potential difficulties and solutions. A three-dimensional co-kriging methodology was developed so that egg production surfaces could be estimated on a finer and more appropriate temporal resolution, while being conditioned on bathymetric data. However, while this methodology produced reasonable estimates of TAEP, the problem of calculating the global estimation variance remained. This was finally solved by employing a conditional simulation procedure.

It should be noted that much of the development work presented has been done using the mackerel egg production data in preference to the horse mackerel data. This is due in part to the fact that the mackerel data appear to be more spatio-temporally correlated and therefore easier to model. However, midway through the project it emerged that there was some doubt as to whether horse mackerel were determinate or indeterminate spawners, and therefore whether the methodologies developed would be appropriate for their stock assessment. For this reason the results presented in this section focus on the western mackerel. However, the methodology developed should be applicable for the determination of the TAEP for horse mackerel data if this is appropriate.

Review of technique for calculating egg density

The work on the geostatistical estimator commenced with a familiarisation with the techniques used to sample the mackerel and horse mackerel eggs, the sampling strategies employed for each of the triennial surveys, and the techniques used to classify the eggs into the different stages (Lockwood *et al.*, 1981). In recent campaigns, the experience gained over the years has been utilised to define the main spawning area for the various stocks. The survey aims to collect data from all over the spawning area during a number of ‘periods’, whereby two samples are collected in each ICES rectangle. However, due to financial and technical constraints, repeated complete coverage is impossible and the available resources are allocated to surveying the areas where the fish are assumed to be spawning in the highest numbers. For example, the sampling effort is generally concentrated in the southernmost part of the spawning area at the start of the season, and further north towards the end of the season. This aims to capture the northwards migration of the spawning population. Furthermore, the vicinity of the shelf break (around a depth of 200 m) may be sampled in preference to the area boundaries as this area is more likely to contain high numbers of eggs (Lockwood, 1988). All data records consist of sample position (in degrees longitude and latitude), date and time, number of eggs in each development stage for mackerel and horse mackerel, calibration factor pertaining to the sampler, depth sampled, temperature at 20 m depth and salinity.

The volume of water filtered by the sampler is first calculated using equation (1). The egg densities in units of eggs per metre squared are then obtained using (2).

$$Volume\ filtered\ (m^3) = \frac{flowm. - revs. \times aperture}{flowm. - calib.} \times efficiency\ factor \tag{1}$$

$$Eggs / m^2 = \frac{Eggs\ counted \times factor}{volume\ filtered\ (m^2)} \times depth\ sampled \tag{2}$$

where:
flowm.-revs. = number of rotations of the flowmeter during tow
aperture = the area of the mouth opening of the sampler in m^2
flowm.-calib. = the number of flowmeter revolutions per metre towed, obtained from the flume or sea calibration in free flow
eggs counted = number of eggs in sub-sample
factor = raising factor from the sub-sample to the whole sample
depth sampled = the maximum depth of the sampler during the tow in metres
efficiency factor = the sampler efficiency from flume or towing tank calibration

However, these densities are not directly comparable since eggs are more likely to be observed at a given stage if the duration of this stage is longer. Lockwood *et al.* (1981) suggested that the densities are corrected using the sea surface temperature measured at sample locations to represent daily productions expressed in number of eggs per m^2 per day using the following development equations for both species:

For stage I mackerel eggs:

$$Eggs / m^2 / day = 24 \times Eggs / m^2 / \exp\left(-1.61 \log_e \left(T°C\right) + 7.76\right) \tag{3}$$

For stage I horse mackerel eggs:

$$Eggs / m^2 / day = 24 \times Eggs / m^2 / \exp\left(-1.608 \log_e \left(T°C\right) + 7.713\right) \tag{4}$$

As part of the Traditional estimator the *eggs/m²/day* are then raised to the area of the rectangle they represent. The rectangle values are summed to give numbers of eggs per day in each stage over the survey area for each sampling period. Rectangle areas are calculated by each ½º row of latitude using the formula:

$$Area\ (m^2) = \left(\cos(latitude) \times 30 \times 1853.2\right) \times \left(30 \times 1853.2\right) \tag{5}$$

The daily egg production is estimated for each survey period in turn and the resulting daily egg production curve (shown for 1998 in Figure 2) is integrated to give the total annual egg production (TAEP).

In order to become familiarised with the data, and to develop ideas regarding how to treat the spatial and temporal components of the mackerel and horse mackerel egg distributions, it was first necessary to undertake a visual inspection of the data for each survey campaign. This analysis was presented in more detail in the 12-month report. The general findings were consistent with the established theory that the western mackerel migrate north from the Bay of Biscay towards the feeding grounds as they spawn (Lockwood, 1988). However, it was concluded from this comparison of density/latitude/date plots between the years that it is difficult if not impossible to make generalisations regarding the spatial and temporal distribution of mackerel and horse mackerel egg density. It was therefore considered necessary to discover the environmental, and perhaps biological, factors that influence the migration of spawning mackerel, and use these to develop a multivariate geostatistical model.

<u>*Initial geostatistical analysis*</u>

An initial geostatistical study was undertaken using a simple two-dimensional procedure. In order to try to minimise the temporal variability in the data, it was decided to first select a suitable 'snapshot' from the time series. The snapshots were selected from each time series based on the general criteria that: a wide range of latitudes was covered over a reasonably short time period; and there were sufficient data points to obtain a good geostatistical representation. An example of a suitable time-window is shown in Figure 4, where the latitude of each sample taken has been plotted against the date for the 1998 survey.

The selected snapshot (between the vertical lines) represents a full areal coverage between 42° and 59° N over a period of 12 days (13/6/98 to 24/6/98). It was possible to select such snapshots for each of the survey campaigns, and these were normally from a 10 to 14 day period around the end of May to the middle of June, reflecting the increased intensity of the survey effort around this time.

To undertake a geostatistical analysis of the egg survey data, it was first necessary to transform the coordinates of the dataset from degrees longitude and latitude into a coordinate system with absolute distances independent of the curvature of the earth. It was decided to create a reference system based on nautical miles, centred on 35° N, 7° W. This origin was selected to allow for the subsequent inclusion of additional data relating to the Southern stock. The coordinates were transformed using the following relations:

$$x = 60 * \cos(latitude) * (longitude + 7)$$
$$y = 60 * (latitude - 35)$$

The basic geostatistical tool used for this initial part of the geostatistical analysis was the variogram (technically termed the semivariogram), defined as follows:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \left[z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h}) \right]^2 \qquad (6)$$

where **h** is the separation vector and **x** is a point in space.

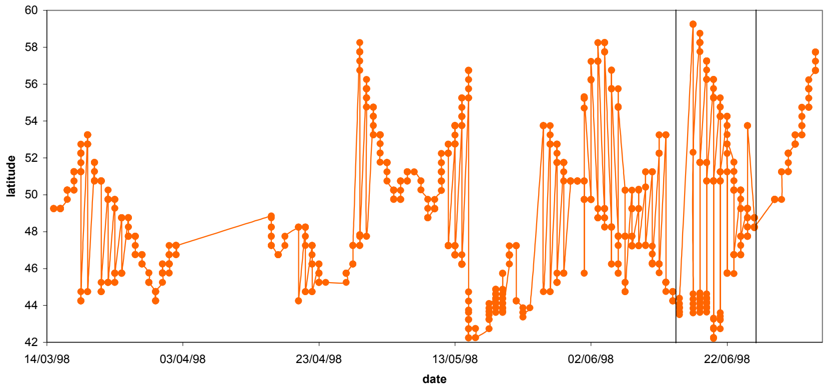


Figure 1. Time series of sample latitude against time for the 1998 campaign. The vertical lines surround the 'snapshot' selected for analysis.

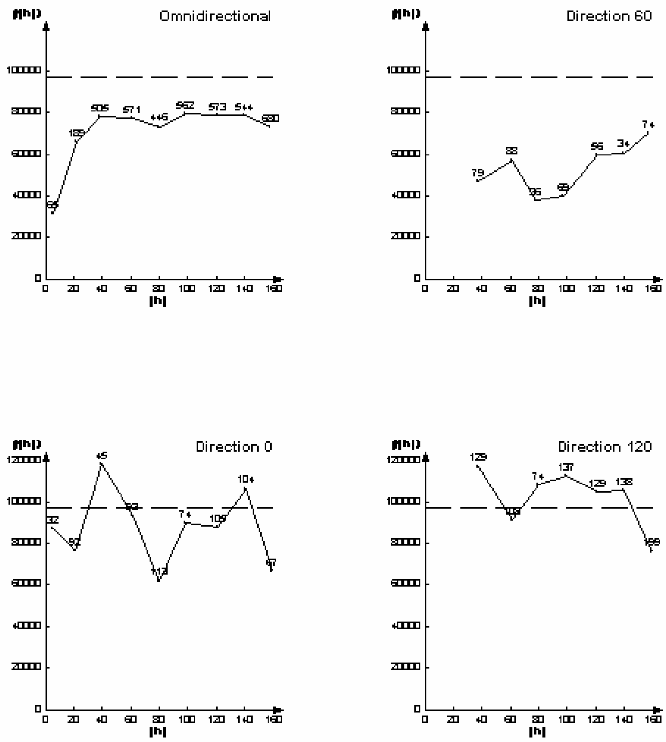


Figure 2. Example showing omnidirectional and directional variograms calculated for 1983 mackerel egg density data.

Variogram models were obtained for each of the surveys using the geostatistical software Variowin 2.21 (Pannatier, 1996). The program allows for the creation of directional and omnidirectional variograms, and so a number of variograms were calculated for each snapshot to assess whether there was any anisotropy that should be accounted for in the spatial model. The lag spacing was specified to be 20 nautical miles. An example of the results for the 1983 mackerel egg density data is provided in Figure 6. The figure shows the omnidirectional variogram and three directional variograms oriented at 0°, 60° and 120°, calculated with an angular tolerance of 30°. The numbers above each point on the graphs refer to the number of data pairs available to make the calculations for each separation distance.

It can be seen from Figure 5 that there is no obvious anisotropy in the directional variograms. There are two likely reasons for this:

- There are smaller numbers of data pairs available to calculate the directional variograms;
- The mackerel generally follow the 200 m contour line, the orientation of which changes in all directions.

Omnidirectional variograms were therefore considered to be more suitable for the initial geostatistical study.
The variogram models fitted for mackerel and horse mackerel for each of the triennial surveys were provided in APPENDIX I of the 12-month report, alongside GIS plots of the entire dataset. Figures 6 and 7 show the variograms and GIS maps of the 1998 egg survey results for mackerel and horse mackerel respectively.
In summary, the majority of fitted variogram models for both mackerel and horse mackerel were spherical. For horse mackerel, three years (1977, 1983 and 1992) are Gaussian, and for mackerel 1983 and 1992 were exponential while 1986 was Gaussian. Most of the models had a significant nugget component. The magnitude of the sill for each model is related to the variance in the egg densities observed during the snapshot period.

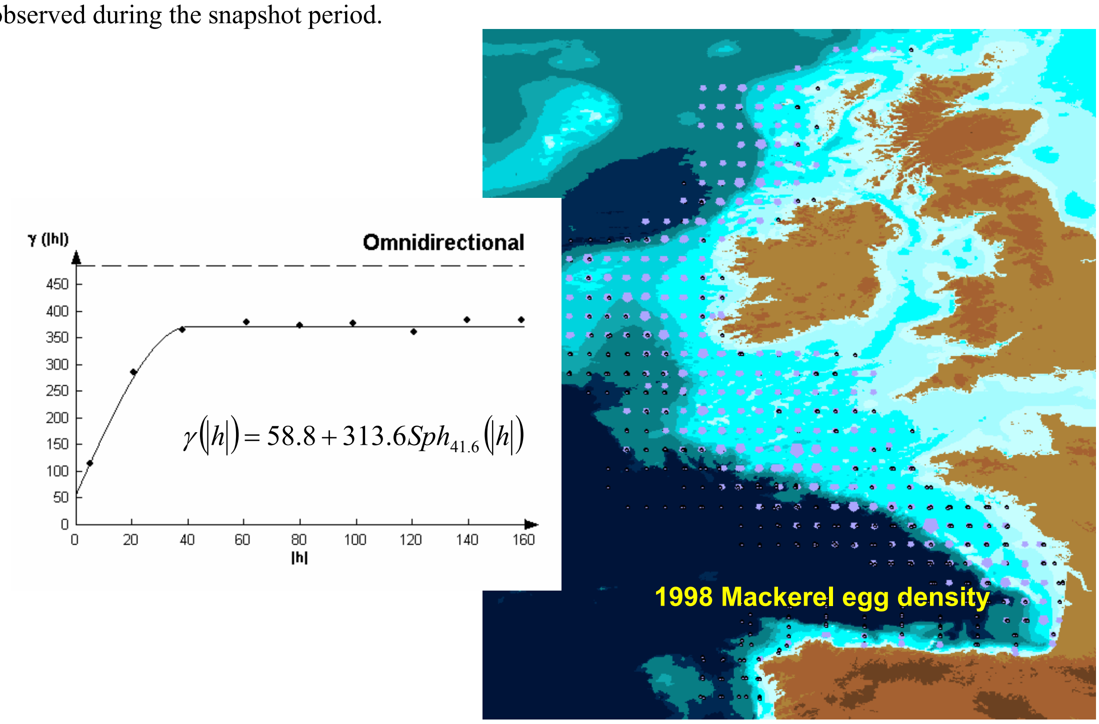


Figure 3. Omnidirectional variogram model fitted to the 1998 mackerel egg densities and GIS map of the respective values.

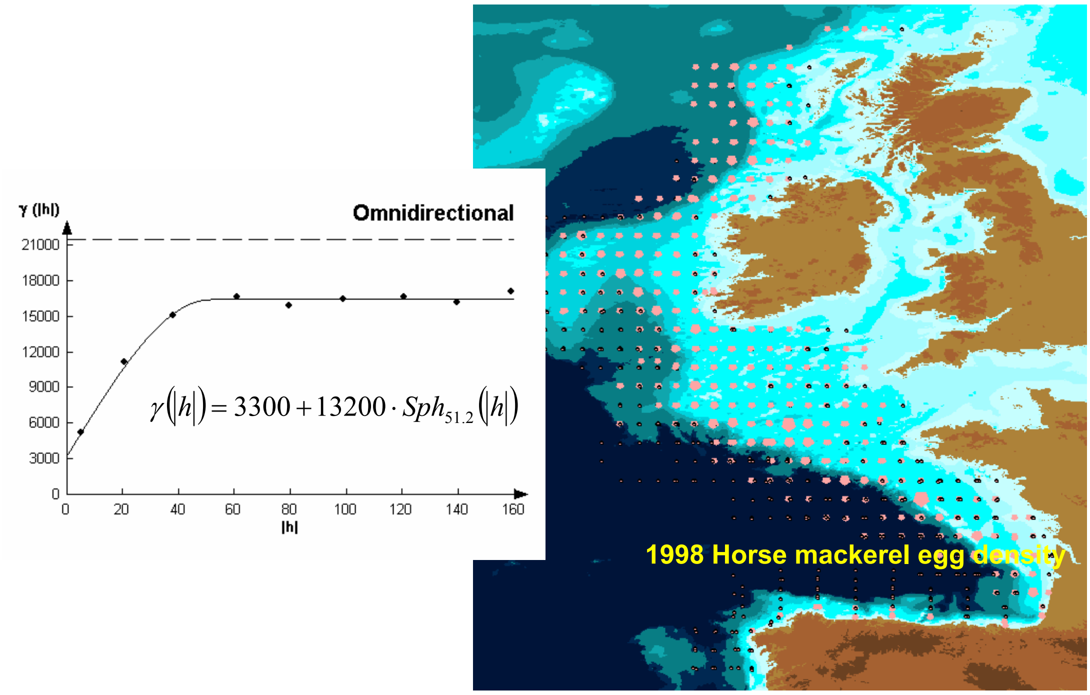


Figure 4. Omnidirectional variogram model fitted to the 1998 horse mackerel egg densities and GIS map of the respective values.

The ranges of the models relate to the spatial extent of conformity in the data. In order to facilitate a comparison between the surveys, the ranges have been plotted against the year in Figure 9. An inspection of the graph revealed that, despite a peak range for mackerel in 1992, the ranges tend to remain within 25 to 60 nautical miles, and if this peak is treated as an outlier, the mean ranges for mackerel and horse mackerel are 38.5 and 42.3 nautical miles respectively.

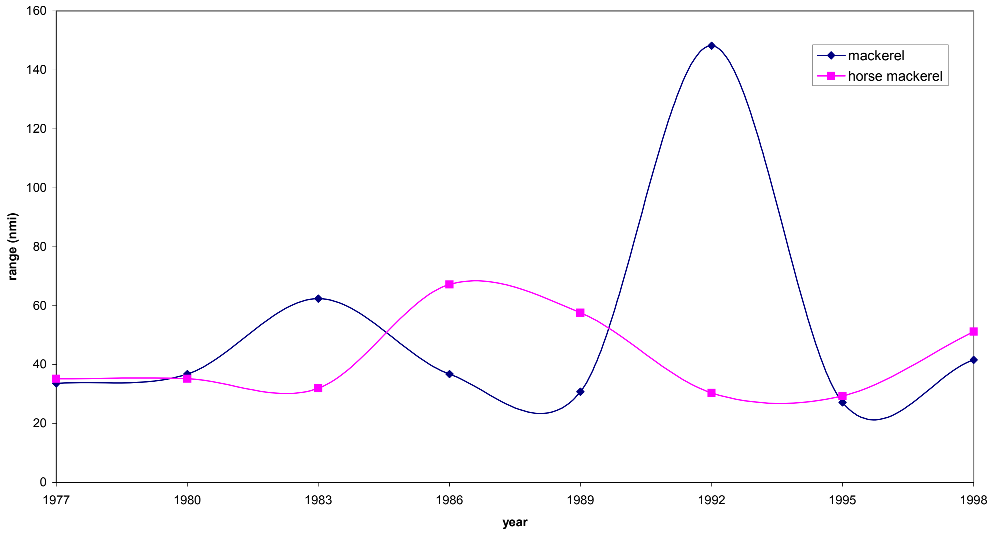


Figure 5. Graph of range against year for mackerel and horse mackerel egg density

The ranges of the models depend on the spatial correlation of the schools of mackerel over the survey area at a point in time, but their determination is subject to a number of sources of error: the locations and numbers of data pairs used to make the calculations; and the manual model-fitting procedure. It is not possible to increase the number of data points without increasing the length of time series included, thereby introducing an additional error source due to the temporal variability.

Development of geostatistical modelling procedure

Using the information gained in the initial study, a more intensive campaign to develop the geostatistical estimator was commenced. One of the suggestions made at the end of the second reporting period was that the knowledge that the spawning fish tend to migrate along the 200 m depth

contour should be used to try to improve the estimates. Using depth as a covariate could potentially provide the estimator with information regarding the preferred longitudinal location of the fish, and at the same time provide a form of directional component that could not be included using the egg densities alone.
Depth data for the survey area were obtained from NOAA (www.noaa.gov), as mentioned in the previous section. The dataset has a fine spatial resolution (1 x 1 nautical miles). Figure 9 shows the bathymetry of the spawning area as plotted using ArcView. The map has been overlaid with the yellow dots representing the mackerel egg densities measured in 1998. It can be seen that the larger densities tend to be observed around the 200 m contour, which is marked by black dots.
The use of a depth-related variable was assessed as a suitable covariate for both mackerel and horse mackerel egg density estimation. Although temperature has been widely toted as a potential covariate, it has been noted by MLA in Beare and Reid (2002) that SST is strongly correlated with time, and so the date (e.g. Julian day), as well as latitude, can be used instead to infer trends in SST. Furthermore, the sea water temperature is already used in the calculation of egg density from the stage I egg count data. For these reasons, SST was rejected as a potential covariate in this study.
In the following sections, the initial development work on the geostatistical framework for estimation of TAEP is presented. In brief, the work focused on the following:

1. Data analysis and preparation;
2. Calculation of experimental variograms and fitting variogram models;
3. Cross validation;
4. Kriging for block estimates of egg density and associated kriging variance;
5. Comparison of results using egg density data alone and with depth covariate;
6. Calculation of AEP and variance.

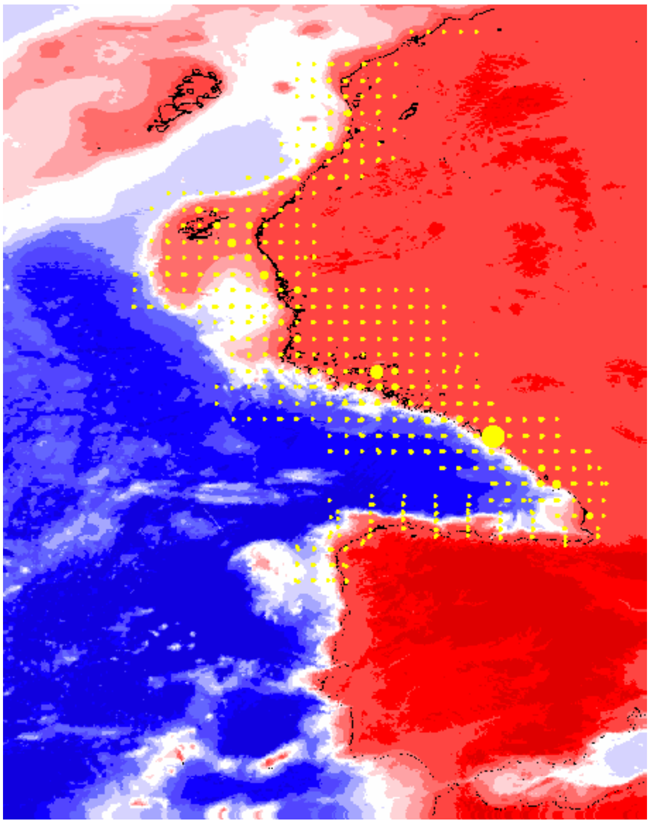


Figure 6. Bathymetry of NE Atlantic mackerel spawning area with 1998 mackerel egg densities.

Data Analysis and Preparation

The egg density data for each time period are highly positively skewed. This can be demonstrated by an example given in Figure 10(a), which shows the raw histogram of the data from one of the survey periods. Since the underlying theory of geostatistics assumes that the data to be modelled are normally distributed, it is common to use the logarithm of highly skewed data instead. The egg density values were therefore transformed as follows:

$$D_{\ln} = \ln(D_{orig} + 1) \tag{7}$$

where D denotes the egg density. The implications of this transformation will be considered in more detail later. The histogram of the transformed data is shown in Figure 10(b). While it can be seen that the distribution is still skewed, it is noticeably closer to the normal distribution desired, and will be assumed to be suitable for geostatistical modelling.

(a) (b)

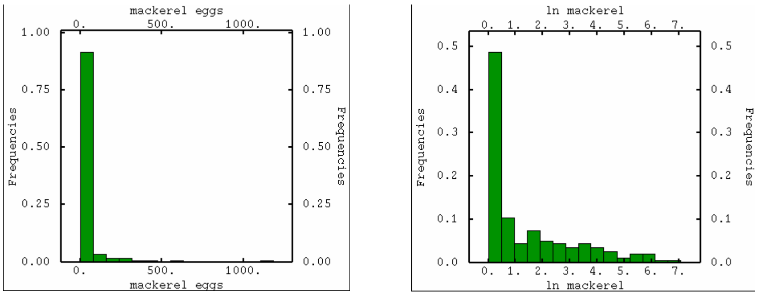
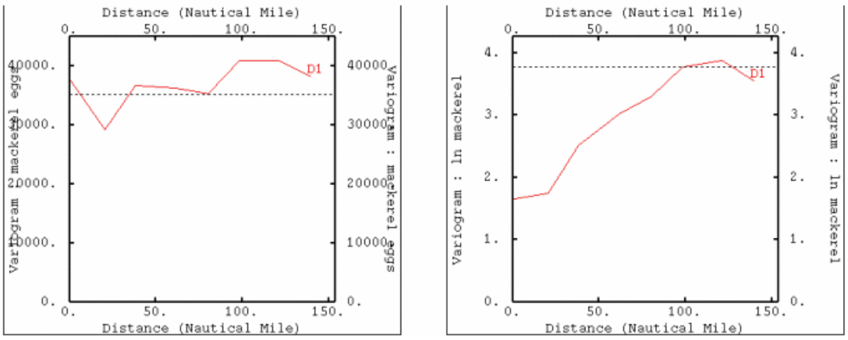


Figure 7. (a) histogram of raw egg density data. (b) histogram of log-transformed data.

The benefits gained in transforming the data can be appreciated at the outset in calculating the experimental variograms. Figure 11 shows the experimental variograms obtained using the original and transformed mackerel egg density data from Period 2 in 1992. It can be seen that transforming the data reveals a more coherent spatial structure, and that the variogram is more amenable to model fitting. It was found that, overall, the log-transformed data also tended to be correlated to a greater spatial extent than the original values.



(a) (b)

Figure 8. Experimental variograms of (a) original data and (b) log-transformed data.

As mentioned previously, mackerel are assumed to follow the 200 m contour as they migrate north. This means that the 200 m depth contour can be considered as the 'mean depth', that is, the bottom depth above which most of the eggs will be spawned. To create a suitable depth covariate, a logical step would be to measure the distance away from this contour. The simplest approach is to calculate the vertical (one-dimensional) difference between the mean depth and the actual depth. Assuming the mean depth is indeed 200m, the depth variable V_{depth} is calculated as follows:

$$V_{depth} = -|-200-A|$$

where A is the actual bottom depth in metres corresponding to the sampled location. The larger, or less negative, the value of V_{depth}, the higher its correlation should be with the egg density.

It was observed, however, that high egg densities could often be detected at locations where the bottom depth was greater (and occasionally less) than 200 m. It was therefore decided to study this in more detail before treating the 200 m contour as a constant mean depth. The actual mean depths were therefore calculated for each time period, as follows. First, the log-transformed egg densities were multiplied by their corresponding bottom depths. The resulting values were then summed over the dataset, and divided by the sum of transformed egg densities to give the mean depth. The mean depths obtained were plotted against the mid point (expressed as the day of the year) of the time period for each year, as shown in Figure 12.

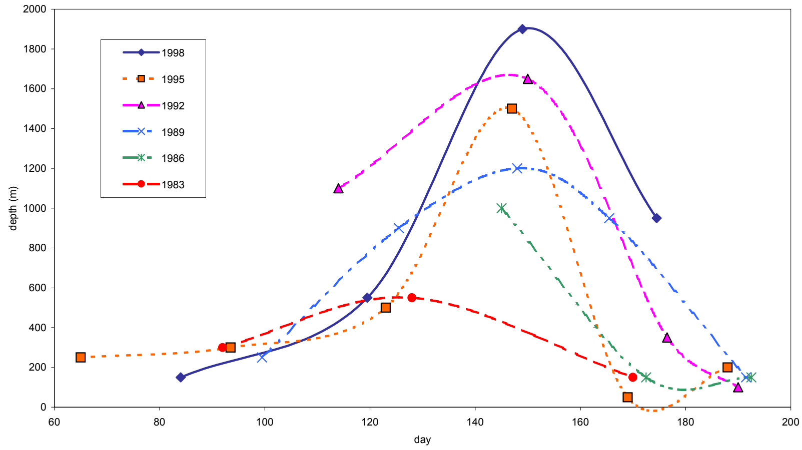


Figure 9. Mean bottom depths corresponding to mackerel egg density.

It is clear from Figure 12 that 200 m does not seem to be an appropriate optimal depth indicator for the whole of the spawning period. The following can be observed:

- Mean depth tends to be around 200 m at the start and end of the spawning period;
- Mean depth tends to increase and reach a peak towards the end of May;

As a further demonstration, the Pearson correlation coefficient between the egg density and the value of V_{depth} is higher when V_{depth} is calculated using the actual mean depth than when it is measured from the 200 m contour. Two examples are given below:

- Example 1: 1977 Period 4.
  Correlation coefficient between $V_{\text{depth-200}}$ and egg density: -0.139
  Correlation coefficient between $V_{\text{depth-1750}}$ and egg density: 0.457
- Example 2: 1989 Period 3.
  Correlation coefficient between $V_{\text{depth-200}}$ and egg density: -0.132
  Correlation coefficient between $V_{\text{depth-1850}}$ and egg density: 0.354

The spawning area was then divided into three zones: south of 48° N (subarea 1); between 48° N and 52.5° N (subarea 2); and north of 52.5° N (subarea 3); to investigate whether the mean bottom depth also varied with latitude. The results for 1977 are shown in Figure 13. It can be seen that the biggest deviations from the 200 m contour are observed below 48° N, and are less marked further north. These findings are similar for all years.

This phenomenon could be related to a number of factors, such as temperature and salinity, and the availability of food. A more detailed analysis should be left to the experts. It is noted, however, that the timing of the increase in optimum bottom depth seems to be coincident with the formation of the thermocline in late May and June (e.g. Coombs *et al*., 2001). This suggests that, despite the overall depth of the water column, the mackerel eggs will be concentrated above the thermocline in the upper mixed layer. Similar results were obtained for horse mackerel egg density.

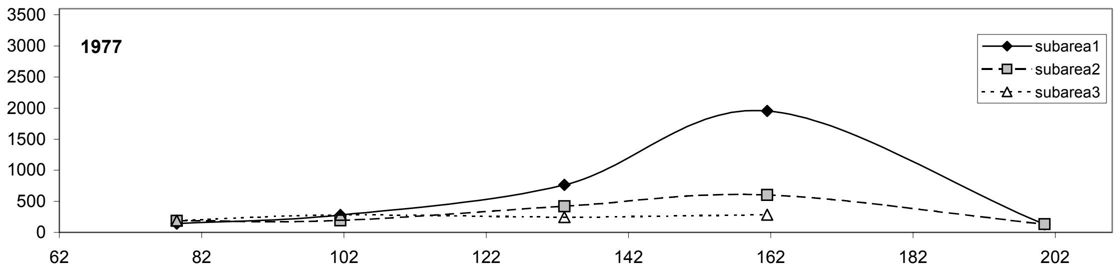


Figure 10. Latitude-dependent mean bottom depth for the 1977 mackerel egg survey.

Calculation of experimental variograms and fitting variogram models

Experimental covariograms were calculated for each of the time periods, using the log-transformed mackerel and horse mackerel egg density and the latitude-dependent V_{depth} as the covariates. Variograms using egg density on its own were also produced, to facilitate a comparison and to evaluate

the benefits of using depth information. Variogram models were then fitted to these experimental covariograms and variograms.
The lower right box in Figure 14 shows the log-transformed egg density variogram, which is modelled using a nugget effect, a spherical term with a range of around 60 nautical miles, and a smaller Gaussian term with a longer range. The upper left box shows the depth variable variogram, which is modelled mainly with a long-range Gaussian term, and a smaller spherical term. The depth-density covariogram is modelled using elements of the egg density and V_{depth} variogram model.

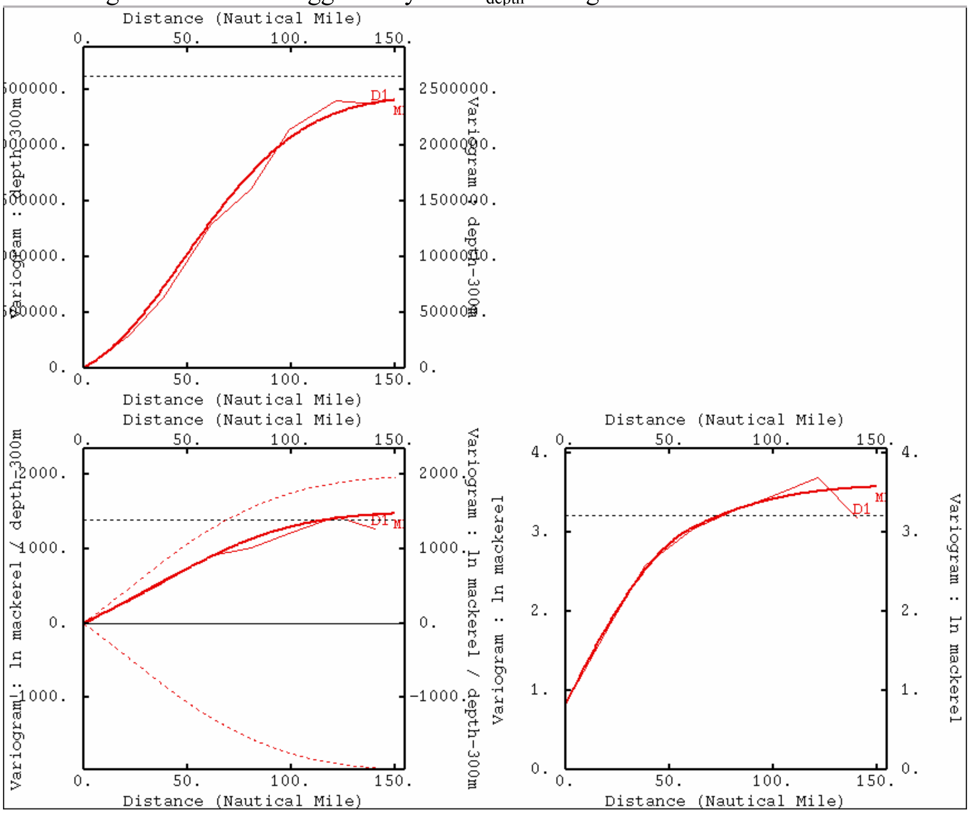


Figure 11. Fitted covariogram models for 1998, Period 4.

Cross Validation

The variogram and covariogram models for each data period can be cross-validated to help evaluate and compare the suitability of different models. The approach employed here was to remove each of the egg density values in turn and estimate its value using the remaining values and the variogram model. Since the
objective of the study was to evaluate the use of a depth-related variable as a covariate, these values were known at each of the points to be estimated. The differences between the resulting estimates and the actual values (errors) were summarised to form a number of readily comparable statistics, such as the mean error and mean standardised error. The results for the univariate and bivariate cases, for all of the time periods and for both mackerel egg density and horse mackerel egg density, are compared in Table 1. The statistic used for comparison here is the mean standardised error (*SE*), defined as:

$$SE = \frac{1}{N}\sum_{N}\frac{D - D^{*}}{\sigma} \tag{8}$$

where N is the number of samples, D is the egg density, D^* is the estimated egg density and σ is the standard deviation of the egg density values. The shaded cells are those which contain the lower mean value of *SE* between the univariate and bivariate cases.
It can be seen from Table 1 that for the case of mackerel egg density, using a depth-related variable as a covariate reduces the *SE* for the vast majority of data periods. The improvement can also be observed for the case of horse mackerel, although it is less pronounced.

Table 1. Mean standardised errors calculated after univariate and bivariate cross-validation.

	mackerel		horse mackerel	
period	univariate	bivariate	univariate	bivariate
1977p1	0.0085	0.0048	0.0006	0.0017
1977p2	0.0069	0.0057	-0.0032	-0.0005
1977p3	-0.0001	-0.0007	-0.0118	-0.0118
1977p4	0.0194	0.0065	0.0054	0.0168
1977p5	-0.0032	-0.0085	-0.0080	-0.0116
1980p1	0.0099	0.0062	0.0008	0.0000
1980p2	0.0039	0.0010	-0.0105	-0.0155
1980p3	0.0038	0.0002	0.0024	-0.0001
1980p4	0.0031	0.0011	0.0045	0.0009
1980p5	-0.0003	0.0003	0.0005	0.0018
1983p1	0.0021	0.0006	0.0050	0.0015
1983p2	0.0112	0.0103	0.0003	-0.0015
1983p3	0.0033	0.0033	0.0010	0.0008
1986p2	0.0027	0.0019	0.0022	0.0026
1986p3	0.0023	-0.0011	0.0132	0.0087
1986p4	0.0215	0.0067	0.0048	0.0026
1989p1	0.0040	0.0018		
1989p2	0.0109	0.0072	0.0060	0.0070
1989p3	0.0117	0.0094	0.0057	0.0054
1989p4	0.0084	0.0076	0.0164	0.0084
1989p5	0.0071	0.0052	0.0044	0.0016
1992p2	0.0094	0.0041	0.0060	0.0021
1992p3	-0.0002	0.0005	-0.0020	-0.0015
1992p4	0.0187	0.0144	0.0148	0.0090
1992p6	0.0041	0.0034	0.0025	0.0021
1995p2	0.0161	0.0039	0.0084	0.0124
1995p3	0.0017	0.0017	-0.0070	-0.0047
1995p4	0.0031	0.0016	0.0012	0.0006
1995p5	0.0074	0.0062	-0.0001	-0.0005
1995p6	0.0038	0.0018	0.0015	-0.0002
1995p7	0.0085	0.0049	0.0096	0.0047
1998p3	0.0073	0.0057	0.0014	0.0003
1998p4	0.0056	0.0037	0.0117	0.0202
1998p5	0.0033	0.0023	0.0031	0.0008
1998p6	0.0016	0.0006	0.0001	-0.0007

Kriging for block estimates of egg density and associated kriging variance

The variogram models obtained above were then used to obtain block estimates of mackerel and horse mackerel egg density, and their associated variance. The first requirement was to select an appropriate grid size. It was decided that a grid size of 15 x 15 nautical miles would provide a good resolution, and would be in rough accordance with the spacing of the latitudinal transects. The next step was to define a kriging 'neighbourhood'. This involved the specification of a number of parameters, as follows. The 'number of sectors' and the 'minimum number of samples' were set to 3, meaning that for each cell to be estimated, the surrounding area would be split into 3 equiangular sectors and at least one sample would have to be present in each. The maximum sample distance was set to 60 nautical miles, after a consideration of the predominant range of the variogram models.

An approach termed 'collocated kriging' was employed for the bivariate kriging. This enabled the depth variable, which is known all over the spawning area, to be used at all locations whether there was an egg density sample present or not.

Block kriging was done for each of the data periods. Geostatistical theory assumes that the value of each cell will lie within a normal distribution, which is calculated using the available data points in conjunction with the variogram models. These distributions are described by their mean value, which provides an estimate for the value of the mean egg density within the grid cell, and the variance, which provides a measure of the expected accuracy of the estimate.

Since the estimates produced are of the log-transformed variable, these need to be back-transformed into original space. The most straightforward back-transformation of the grid estimates from log space to original space results in the median of the kriged normal distribution, rather than the mean:

$$\text{Median}_{\text{orig\_space}} = e^{\text{Mean}_{\text{log\_space}}} - 1 \tag{10}$$

To obtain an estimate of the mean density within a grid cell in original space, the following back-transformation is required:

$$\text{Mean}_{\text{orig\_space}} = \text{Median}_{\text{orig\_space}} \times \left(e^{\frac{\text{Var}_{\text{log\_space}}}{2}} - 1 \right) \tag{11}$$

The kriging variance is back-transformed as follows:

$$\mathrm{Var}_{\mathrm{orig\_space}} = \left(\mathrm{Mean}_{\mathrm{orig\_space}}{}^{2}\right) \times \left(e^{\mathrm{Var}_{\mathrm{orig\_space}}} - 1\right) \quad (12)$$

As discussed above, the density data are positively skewed, and hence the median of the distributions is less than the mean. The median is less sensitive to extreme values and sampling error, and for this reason it is often preferred to the mean as an estimate under these circumstances (Kitandis and Shen, 1996). Therefore, although it would tend to underestimate the true value of egg density, the median has been adopted here as the 'best' estimate.

The kriging procedure provides us with egg density estimates over each valid grid cell with an area of 15 x 15 nautical miles. Since the egg densities are measured in metres squared, the estimates need to be scaled up to the size of the grid cell. The densities can then be summed over the kriged surface to give a measure in units of eggs per day. To obtain an estimate of the egg production for the data period, that value can then be multiplied by the number of days over which the samples were taken. Since, however, the data periods are often separated by a number of days and do not cover the supposed start and end dates of the spawning season, a simple summation of the egg productions per period would not provide an adequate approximation of TAEP. A slightly more sophisticated method was therefore employed. The estimate of eggs per day for each period was plotted against the period's mid-point. Zeroes were plotted at the start day (assumed to be 10th February, or day 40) and the end day (31st July, or day 210). An estimate of TAEP was then obtained by calculating the area under the graph.

Figure 15 shows such a graph for the 1998 horse mackerel data. Each period's estimate of egg production (*EP*) has 95% confidence interval associated with it. These were calculated from the kriging variance values as follows.

$$\text{Lower confidence limit} = EP_{period} \times e^{-1.96\sigma} \quad (13)$$

$$\text{Upper confidence limit} = EP_{period} \times e^{+1.96\sigma} \quad (14)$$

where

$$\sigma = \sqrt{\ln\left[1 + CV^{2}\right]} \quad (15)$$

and

$$CV = \sqrt{\frac{Var_{EP}}{EP^{2}}} \quad (16)$$

The variance associated with the egg production estimates for each period, Var_{EP}, was calculated as follows.

$$Var_{EP} = A^{2} \times \left(\left[\frac{1}{n^{2}}\right]\sum_{b=1}^{n} s_{b}^{2}\right) \quad (17)$$

where A is the total kriged area, n is the number of kriged cells and s_b is the kriging standard deviation associated with each grid cell b.

Table 2 lists the egg productions calculated for each data period for both mackerel and horse mackerel, and for the estimates made with and without the depth as a covariate. The coefficient of variation (*CV*) is useful in comparing the univariate and bivariate geostatistical models. In general, the lower the ratio of the variance to the mean, the more accurate the estimate. It can be seen from Table 2 that in the case of mackerel, the CV is unanimously lower for the bivariate models, which, in conjunction with the results in Table 1, suggests that the depth variable is a very beneficial covariate. The improvement is less clear in the case of horse mackerel. For example, the mean CV in the univariate case is less than that that obtained for the bivariate case. However, when the extreme value observed in 1980 Period 4 is removed, the mean CVs become 0.194 and 0.167 for the univariate and bivariate cases respectively. Regarding the egg productions (EP), it can be seen that for both mackerel and horse mackerel, the bivariate models generate lower density estimates.

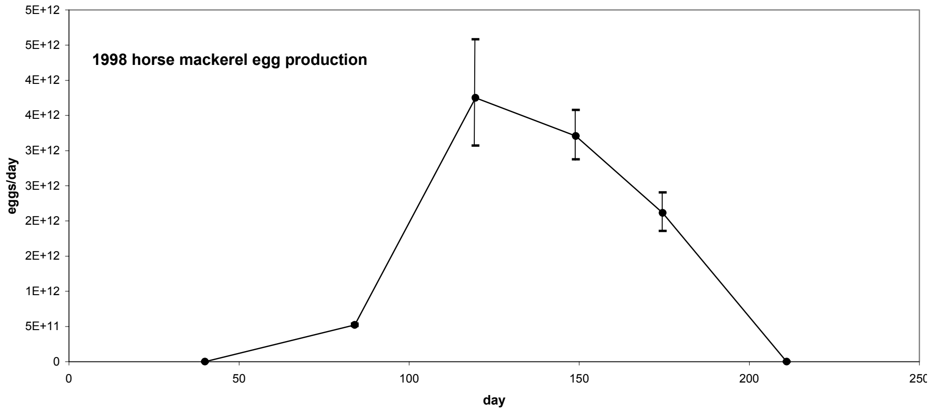


Figure 12. Horse mackerel egg production curve for 1998.

Table 2. Egg productions and associated CV for univariate and bivariate geostatistical models.

	mackerel				horse mackerel			
	univariate		bivariate		univariate		bivariate	
period	EP	CV	EP	CV	EP	CV	EP	CV
1977p1	1.759E+12	0.732	1.395E+12	0.182	2.908E+11	0.036	2.77E+11	0.041
1977p2	3.094E+12	1.627	3.13E+12	0.260	2.799E+11	0.050	2.547E+11	0.036
1977p3	3.058E+13	1.420	2.114E+13	0.126	6.642E+12	0.724	5.612E+12	0.534
1977p4	1.954E+13	0.487	1.916E+13	0.172	1.913E+12	0.189	3.285E+12	0.243
1977p5	1.635E+11	0.069	1.724E+11	0.047	3.116E+12	0.440	3.418E+12	0.422
1980p1	1.668E+12	0.565	1.56E+12	0.155	1.701E+11	0.081	1.639E+11	0.076
1980p2	5.133E+12	1.093	4.153E+12	0.177	7.171E+11	0.131	7.252E+11	0.118
1980p3	5.348E+12	0.849	5.031E+12	0.115	7.885E+11	0.071	7.442E+11	0.050
1980p4	1.041E+13	3.622	9.898E+12	0.105	3.745E+12	1.550	2.638E+12	8.061
1980p5	2.51E+11	0.108	2.595E+11	0.066	1.752E+12	0.203	1.948E+12	0.202
1983p1	1.901E+12	1.232	1.643E+12	0.190	4.167E+11	0.041	4.403E+11	0.048
1983p2	1.024E+13	0.420	9.83E+12	0.121	1.491E+12	0.045	1.587E+12	0.043
1983p3	1.386E+13	0.196	1.393E+13	0.041	3.576E+12	0.071	3.616E+12	0.069
1986p2	1.59E+13	0.347	1.508E+13	0.064	2.604E+13	0.074	3.991E+12	0.230
1986p3	1.471E+13	0.246	1.386E+13	0.080	7.594E+12	0.164	7.461E+12	0.188
1986p4	3.661E+12	0.158	3.517E+12	0.080	3.137E+12	0.253	3.003E+12	0.276
1989p1	1.864E+13	42.046	1.073E+13	0.291				
1989p2	9.696E+12	1.106	8.309E+12	0.114	7.021E+12	0.006	6.813E+12	0.665
1989p3	3.395E+13	0.810	3.101E+13	0.172	8.942E+12	0.159	8.797E+12	0.105
1989p4	1.403E+13	0.223	1.378E+13	0.059	2.087E+13	0.252	1.978E+13	0.222
1989p5	2.434E+12	0.688	2.313E+12	0.138	1.604E+12	0.197	1.509E+12	0.108
1992p2	2.081E+13	1.815	1.552E+13	0.121	3.126E+12	0.360	1.952E+12	0.090
1992p3	6.636E+12	0.024	7.075E+12	0.012	1.959E+12	0.063	2.057E+12	0.082
1992p4	2.728E+12	0.154	2.579E+12	0.052	6.856E+12	0.615	6.725E+12	0.225
1992p6	3.645E+11	0.090	3.626E+11	0.055	8.546E+11	0.248	8.42E+11	0.169
1995p2	2.781E+11	2.762	2.395E+11	0.147	1.214E+11	0.069	1.011E+11	0.169
1995p3	2.379E+12	1.181	1.891E+12	0.088	2.287E+12	0.348	1.721E+12	0.201
1995p4	8.32E+12	3.707	8.557E+12	0.195	9.202E+11	0.055	9.689E+11	0.073
1995p5	5.821E+13	2.195	5.11E+13	0.959	1.177E+13	0.719	1.054E+13	0.413
1995p6	4.511E+12	0.293	4.245E+12	0.039	8.528E+12	0.157	8.205E+11	0.068
1995p7	2.382E+12	0.805	2.414E+12	0.142	1.804E+13	0.258	1.764E+13	0.096
1998p3	3.754E+12	0.559	3.546E+12	0.119	5.063E+11	0.024	5.221E+11	0.017
1998p4	5.654E+12	0.492	5.288E+12	0.067	3.718E+12	0.092	3.751E+12	0.102
1998p5	5.092E+12	0.155	4.6E+12	0.075	3.469E+12	0.114	3.207E+12	0.056
1998p6	2.669E+12	0.316	2.915E+12	0.068	2.099E+12	0.084	2.113E+12	0.066
Mean	9.736E+12	2.074	8.578E+12	0.140	4.834E+12	0.234	3.795E+12	0.399

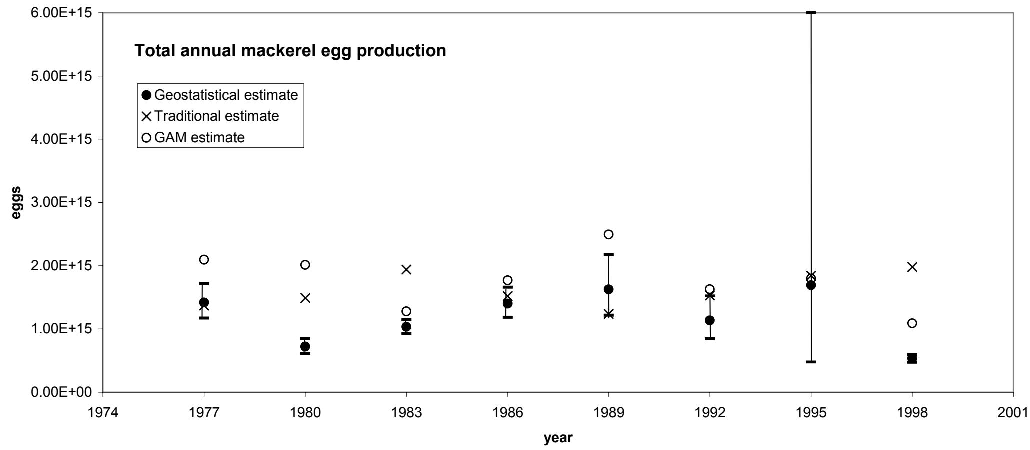


Figure 13. Co-kriged estimates of mackerel egg production with 95% confidence intervals.

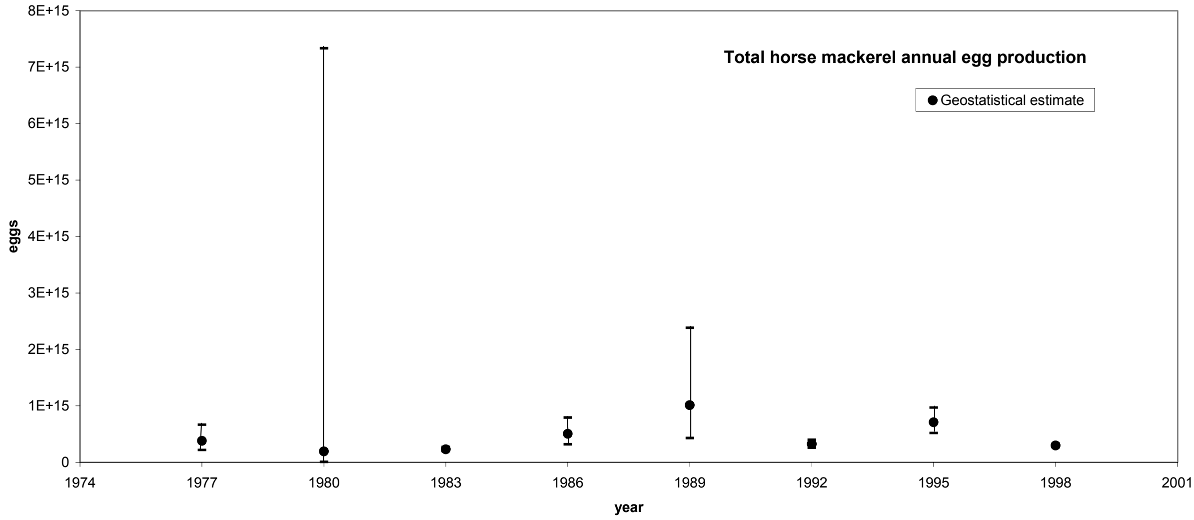


Figure 14. Co-kriged estimates of horse mackerel egg production with 95% confidence intervals.

Calculation of the variance of the value obtained for TAEP is complicated due to the discontinuous and incomplete nature of the sampling coverage. If the coverage were continuous, then it could be estimated from multiplying the variance calculated for each period by the square of the number of days in that period, and then summing over all the data periods. Here, however, there are often gaps between sampling periods, and at the start and end of the spawning season. Nevertheless, an effort to estimate it has been made here by extrapolating the variance estimates of each data period so that they meet at some intermediate point, and extend to the ends of the spawning season. The variance was then multiplied by the square of the extended number of days for each period, and these values were then summed together to provide an 'index' for the variance of the TAEP. The bivariate estimates of TAEP obtained for each triennial survey, with their associated confidence limits, are plotted in Figures 16 and 17 for mackerel and horse mackerel respectively. Figure 16 also includes the traditional and GAM estimates of TAEP for comparison.

It is important to note that the results presented here are first estimates and should not be considered as anything else. An inspection of Figure 16 will indicate that the geostatistical estimates are in general lower than the estimates than the traditional and GAM methods. There are a number of sources of error affecting the present geostatistical estimates that are not accounted for in the variance estimates, and these will be discussed below. An explanation for the large error bars obtained for 1995, and for the 1980 horse mackerel TAEP in 1980, will be presented in the following paragraphs.

The co-kriged estimates of mackerel and horse mackerel egg density were displayed in APPENDIX I.2 of the mid-term report. An example is provided in Figure 18 below showing the estimates of the log-transformed mackerel egg density from 1998. The actual log-transformed data points are superimposed in black. The blue squares signify high egg densities, while red denotes zero egg density. It can be seen that, particularly in Periods 5 and 6, the northwestern edges of the kriged area seem incomplete. In an ideal world, all of the cells on the edge of the kriged area should be zero. What this suggests is that the cells adjacent to these northwestern edges may in fact have significant egg concentrations, and that the overall egg productions for these periods may therefore be seriously underestimated. Similar edges and holes in the kriged surface are observed for the majority of the data periods, and are in general due to insufficient survey coverage.

Another point that should be made is that there are a number of cells, particularly in Period 4 towards the west of the Bay of Biscay, that have significant density estimates in cells adjacent to data points with low or zero egg densities. This is an artefact of the use of a depth variable as a covariate. The model has used the depth variable to create non-zero estimates at locations where the depth has been specified as optimum. This represents a potential source of overestimation of the overall egg production, and should not be ignored.

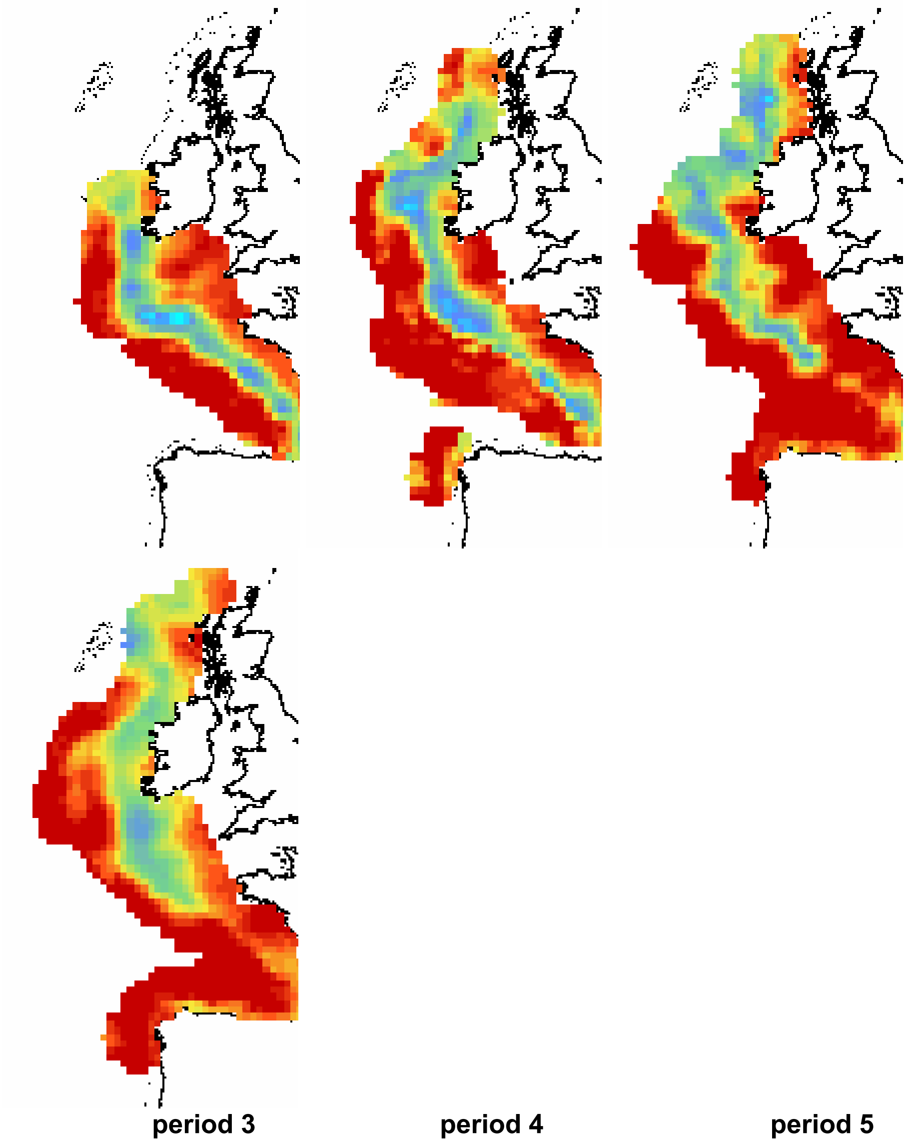
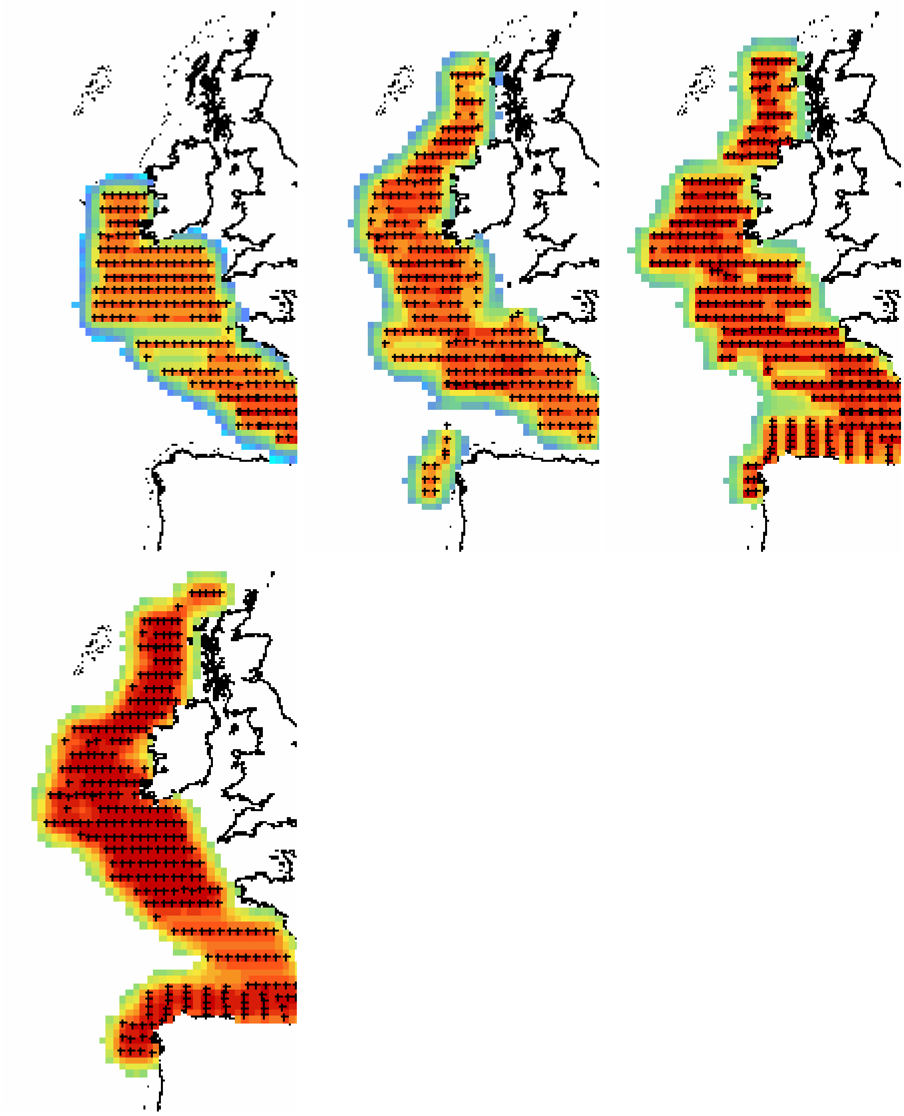


Figure 15. Co-kriged estimates of log-transformed mackerel egg density for 1998.

It is also informative to inspect the spatial distribution of the geostatistical variance. Figure 19 shows the log-transformed standard deviations obtained for the 1998 survey, corresponding to the estimates shown in Figure 18. The locations of the samples are superimposed. As before, the high values are blue while the low numbers are red. On inspection of Figure 19 it is clear that the highest standard deviations are observed around the edges of the kriged area, which corresponds to a lower number of data samples used to calculate the estimate. It can also be seen that the highest standard deviations are associated with the kriging estimates made for Period 3. This may be due to the slightly lower range of the variogram model that was used to calculate the estimates, as well as the overall variability of the data. The kriging variance in Periods 5 and 6 is lower to the intensive sampling campaign.

It was emphasised in the mid-term report that the estimates of variance and 95% confidence limits must not be taken as perfect. There were additional sources of error and other considerations that remained to be addressed. These are summarised below.

- Accounting for inter and intra-period temporal variability;
- Use of kriging variance to obtain global variance estimates;
- Improving the spatio-temporal resolution of the depth-related covariate.



period 3 period 4 period 5
period 6

Figure 16. Log-transformed co-kriging standard deviations for mackerel egg density for 1998.

Development of 3-d spatio-temporal geostatistical methodology

Following the submission of the mid-term report, further progress was made towards the development of the geostatistical estimator. The work resulted in a robust methodology which provided adequate estimates of TAEP. A description of the various aspects of the estimator is provided below.

Specification of 3-d Grid

A three dimensional grid was defined, with cells of dimension 15 nautical miles x 15 nautical miles x 7 days. This allowed a much finer spatio-temporal resolution than can be achieved with the Traditional method. The grid was large enough in all directions to cover the spatio-temporal extent of the spawning season, and is shown in Figure 20.

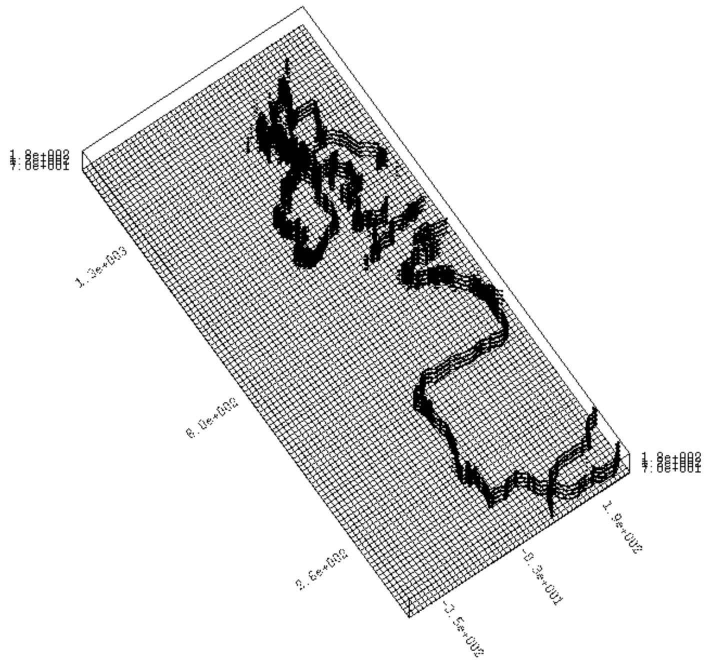


Figure 17. 3-d grid showing western European coastline.

Improved depth-related covariate

As mentioned above, the depth covariate provided a useful means of improving the kriging estimates. It was observed that the optimal depth for egg density appeared to be a function of latitude and time. To undertake 3-d bivariate modelling the optimal depth covariate had to be defined over the entire grid. This was done by compiling the entire set of log-transformed egg survey data, and sorting and discretising it into sequential slices of 100 nautical miles (measured along a line of longitude) x 20 days. This partitioning of the data allowed a reasonable resolution to be obtained while ensuring that there were sufficient data points in each section. The mean depth corresponding to each of the blocks was then calculated.

3-d Covariogram Modelling

Three-dimensional experimental covariograms are calculated using the entire datasets for each survey campaign, along with the latitude and time dependent V_{depth} covariate. The spatial plane is treated as omnidirectional as before. An example of the experimental and modelled covariogram for 1998 is provided in Figure 21.

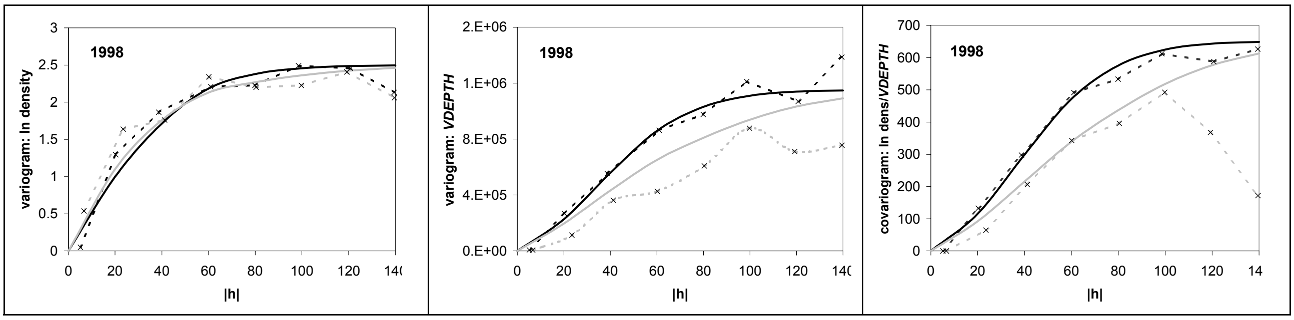


Figure 18. Experimental and modelled covariogram for the 1998 dataset. The black line represents the omnidirectional spatial variogram while the grey line represents the temporal variogram.

Neighbourhood specification and addition of zeroes

Kriging estimates are more accurate if only neighbouring data points located within the range of the variogram are used (the range is typically in the region of 75 nautical miles and 60 days). However, it

was observed that limiting the neighbourhood in this way often led to holes in the estimation surface. This meant that the coverage was incomplete, suggesting that the TAEP estimates were negatively biased. The kriging neighbourhood was therefore specified with a search radius of 100 nautical miles in the spatial plane and 100 days in the time dimension. It was also stipulated that there would have to be at least 3 data points used in each estimation, with at least one located within each of three equi-angular sectors, and a maximum of 15 data points.
It may be noted that, for any grid cell estimated using only data points located outside the variogram range, the estimation consists of a simple averaging of the data values. The corresponding estimation variance will be high to reflect the increased uncertainty.
In addition to the occurrence of holes in the estimation surface, a further problem associated with the estimation of egg densities involves the extrapolation of high values at the edges of the survey. The specification of a large search radius allowed false zeroes to be added to the datasets. These zeroes can help to tie the egg densities down towards the spatial limits of the spawning area, and also help to generate estimations in areas where the data are sparse. The zeroes are placed outside the limits of the survey area, where we are confident there will be no eggs. They are also added throughout the entire survey area prior to the start and after the end of the spawning season.
Structural zeroes are also necessary for the implementation of the GAM method. Unfortunately, they have been found to give rise to a negative bias in the GAM's estimation of TAEP. However, the zeroes should not give rise to a negative bias in the geostatistical method, as the kriging is done using a relatively small moving neighbourhood, and the zeroes are only included at the edges of the spatio-temporal distribution where minor quantities of eggs are generally found.

Block Co-kriging

Collocated co-kriging is performed over the grid to obtain egg density estimates and corresponding kriging variance for each cell in the spawning area. The spawning area was defined similarly to that used for the Traditional and GAM methods.

Back-transformation

As before, the egg density data were subjected to a log transformation prior to kriging, and therefore the estimates and associated kriging variances need to be back-transformed into original space before they are used to calculate the TAEP. The back-transformation formulae for the kriging median, mean and variance were given in equations (10), (11) and (12) respectively.
However, back-transformed estimates do not always fulfil the non-bias condition (Journel and Huijbregts, 1978). A solution to this was found by multiplying the $\text{Median}_{\text{orig\_space}}$ values by a corrective factor *k*. This correction factor is estimated from the bias observed in the back-transformed cross-validation results. The maximum value in the dataset is removed prior to the calculation of the bias to make the calculation more robust.

Calculation of TAEP and Confidence Intervals

The estimates, in units of metres squared per day, are scaled up to the size of the grid cell. The resulting densities can then be summed over the kriged surface to give a measure in units of eggs per week:

$$\text{WEP}_{week} = k \times \sum_{c=1}^{No\_cells_{week}} \left(\text{Median}_{orig\_space,week,c} \times 7 \times raising\_factor \right) \tag{18}$$

where *raising_factor* and the additional factor 7 account for the area of the kriged surface and the number of days in the week, *k* is the correcting factor obtained by cross-validation analysis and WEP_{week} is the weekly egg production for a specific week. The estimates of TAEP are then obtained by summing over the egg production values for the 25 weeks of the spawning season:

$$\text{TAEP} = \sum_{week=1}^{25} \text{WEP}_{week} \tag{19}$$

where week 1 begins on the 40th day of the year and week 25 ends on the 215th day. To calculate the weekly confidence intervals, the values of kriging variance are first back-transformed as in equation (12). Then the values are raised to the dimension of each grid cell and summed over the kriged area to give the variance associated with the WEP_{week} estimation as follows:

$$\mathrm{Var}_{week} = \sum_{c=1}^{No\_cells_{week}} \left(\mathrm{Var}_{\mathrm{orig\_space},week,c} \times (7 \times raising\_factor)^2 \right) \tag{20}$$

The coefficient of variation (CV_{week}) is then calculated:

$$\mathrm{CV}_{week} = \sqrt{\frac{\mathrm{Var}_{week}}{(\mathrm{WEP}_{week})^2}} \tag{21}$$

and the error bars are then obtained as follows:

$$\text{Lower confidence limit}_{week} = \mathrm{WEP}_{week} \times e^{-1.96\sigma_{week}} \tag{22}$$

$$\text{Upper confidence limit}_{week} = \mathrm{WEP}_{week} \times e^{+1.96\sigma_{week}} \tag{23}$$

where

$$\sigma_{week} = \sqrt{\ln\left[1 + CV_{week}{}^2\right]} \tag{24}$$

The confidence intervals for the TAEP estimates are then obtained by summing over the weekly confidence limits:

$$\text{Lower confidence limit} = \sum_{week=1}^{25} \text{Lower confidence limit}_{\mathrm{week}} \tag{25}$$

$$\text{Upper confidence limit} = \sum_{week=1}^{25} \text{Upper confidence limit}_{\mathrm{week}} \tag{26}$$

An example of the kriged weekly egg production estimates obtained for the 2001 survey is provided in Figure 22. For clarity the values have been log-transformed prior to plotting.

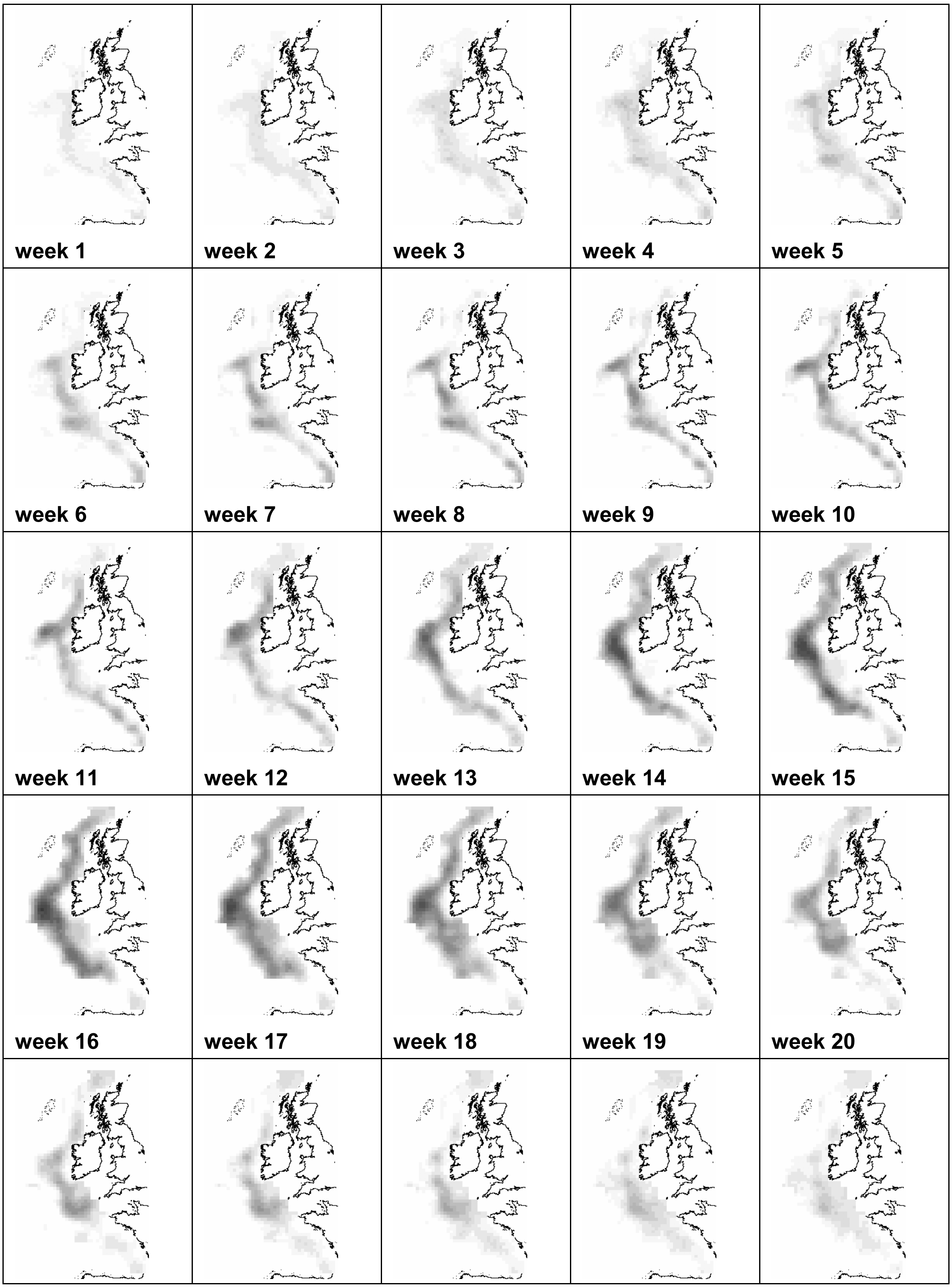
Table 3 lists the triennial TAEP estimates and CV values that were calculated using the method presented above. The correction factor *k* is also provided, and it can be seen that this varies between 1.35 for 1983 and 2.39 for 2001. Figure 23 plots the TAEPs and their error bars and compares them with the estimates made by the GAM and Traditional methods.

The results in Figure 23 show that the geostatistical method provides estimates of TAEP that are comparable with those obtained using the GAM and Traditional methods. The geostatistical estimates are generally closer to those of the GAM method, and on almost all occasions are higher than the Traditional estimates.

Table 3. TAEP estimates for each survey, adjusted according to corrective factor *k*.

Year	Kriged TAEP	k	Adjusted TAEP	CV
1977	1.21×10^{15}	1.88	2.28×10^{15}	27%
1980	1.03×10^{15}	1.72	1.77×10^{15}	29%
1983	1.08×10^{15}	1.35	1.46×10^{15}	20%
1986	1.45×10^{15}	1.48	2.15×10^{15}	22%
1989	1.67×10^{15}	1.44	2.41×10^{15}	22%
1992	1.33×10^{15}	1.44	1.91×10^{15}	17%
1995	1.03×10^{15}	1.88	1.93×10^{15}	11%
1998	0.62×10^{15}	2.37	1.47×10^{15}	12%
2001	0.57×10^{15}	2.39	1.36×10^{15}	9%

However, it was noted that method of calculating the error bars detailed above does not theoretically provide an accurate estimate of the true confidence intervals. Firstly, it is assumed that the lognormal back-transformation of the kriging variances gives the correct value in original space. Secondly, to sum the individual kriging variances is incorrect as this assumes that they are independent (Armstrong, 1998, p. 128). Confidence intervals obtained in this way would theoretically be overestimated. Furthermore, many researchers maintain that since the local kriging variance is independent of the data values, it should not be used as a measure of uncertainty (Deutsch & Journel, 1998). The error bars plotted in Figure 23 should therefore not be treated as absolute measures of confidence, but rather simply used for purposes of comparison. Since the determination of confidence intervals is an important aspect of this project, much of the research undertaken subsequently was devoted towards improving the variance estimates.



week 21	week 22	week 23	week 24	week 25

Figure 19. Kriged weekly egg density surfaces for the 2001 survey. The values were log transformed prior to plotting. Darker patches indicate higher egg numbers

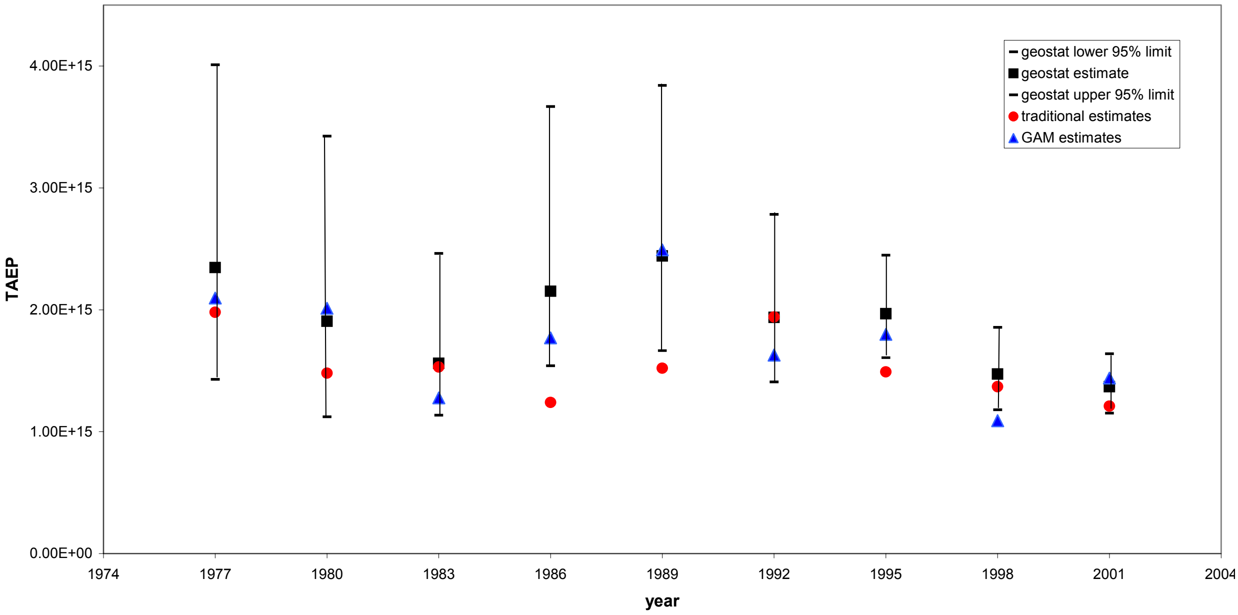


Figure 20. Geostatistical TAEP estimates with 95% confidence intervals and corresponding Traditional and GAM estimates.

Methodology for TAEP and Variance Estimation

As mentioned above, the CVs associated with the TAEP estimates presented so far were calculated using false assumptions. The accurate calculation of global estimation variance is of primary importance for good management of the mackerel stocks. Although geostatistical applications such as this were not widely reported in the literature, two contrasting potential methods for estimating global variance were identified:

(1) Combination of error terms;

(2) Conditional Simulation;

The two methods and their application to the mackerel egg density data are described in turn below:

Combination of error terms

This method for variance calculation was introduced by Journel and Huijbregts (1978) to be used in conjunction with global estimation on a domain of known geometry. The method assumes that the procedure for estimating the mean value of a variable over the domain can be considered as consisting of three steps, each of which are associated with an independent estimation error. The steps involved in the estimation of TAEP could be thought of as follows:

a. The estimation of the quantities of eggs aligned along latitudinal transects, which have the greatest sampling density. The estimation variance corresponding to this step is referred to as the "line term", T_l ;

b. The estimation of the spatial egg surface by the latitudinal transect quantities, which are assumed to be perfectly known. The estimation variance corresponding to this step is the "section term", T_s;

c. The estimation of the egg production for each period by their median levels, which are assumed to be perfectly known. The estimation variance corresponding to this step is the "slice term", T_v.

Figure 21. 1998 sampling locations. The 200 m depth contour is indicated.

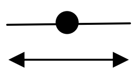
The individual error terms are then summed to give the overall relative estimation variance. An additional term that can be added is the "border term", which would account for uncertainties in the estimation of the extents of the spawning season and spawning area. For the present time it will be assumed that the spatio-temporal boundaries of the spawning season are perfectly known.
An example of the implementation of the combining of errors method will be presented using the dataset collected in 1998. The spatial locations of the 1998 data are shown on Figure 24.
Firstly, a three-dimensional relative variogram is modelled. To create a general relative variogram, the variogram is simply standardised by the mean *m* of the data values that belong to the particular interval *h*, that is:

$$\gamma_R(h) = \frac{\gamma(h)}{\left(\frac{m_{+h} + m_{-h}}{2}\right)^2} \tag{27}$$

For 1998, the relative variogram had a nugget of 0.2, and a spherical component with a sill of 17.8, and ranges a_s = 76 nautical miles in the isotropic spatial plane and a_t 65 days in the temporal direction.

1. The line term

The data are generally sampled along latitudinal transects separated by half a degree. In the northern most transects, this corresponds to a distance of around 15 nautical miles, whereas in the southern part of the spawning area the samples are separated by around 20 nautical miles. However, the transects do not always reach the edges of the designated spawning area, and this has to be taken into account in the uncertainty calculations. This is done in a crude fashion here by 'stretching' the samples across the width of their transect and then calculating the average separation distance. This separation distance can be thought of as the 'segment of influence'.



l

As can be seen in Figure 24, the latitudinal transects have different lengths (for example the spawning area is widest around 47.5°N and narrowest around 44°N and 57°N). In order to account for their individual significance, the average separation distances were calculated for each transect and then weighted by their length. The line term T_l quantifies the extension error of the point samples to their segments of influence, and is calculated as follows:

$$T_l = \frac{1}{nL}\sum_i l_i \sigma^2_{Bb1i} \tag{28}$$

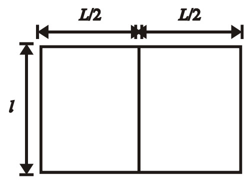
where L is the sum of the lengths l of each transect i, n is the number of samples and the value of the elementary extension for each transect is obtained from Chart no. 7 (in Mining Geostatistics, Journel and Huijbregts, 1978, p131) using $l_i/a_s = l_i/76$ for the 1998 data. This gives:

$$T_l = \frac{1}{nL}\sum_i l_i\left(C_0 + C \times \sigma^2_{E1i}\right)$$

$$T_l = \frac{1}{951 \times 6685}\sum_i l_i\left(0.2 + 17.8 \times \sigma^2_{E1i}\right) = 0.0019$$

The section term

This term measures the extensions of the latitudinal egg quantities to their sections of influence, which in this case corresponds to the 30 nautical mile distance between transects.



As above, the term is calculated after weighting each line according to its relative contribution. The section term is given by:

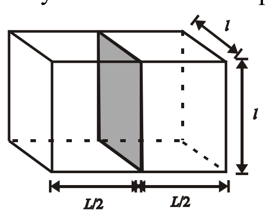
$$T_s = \frac{\sum_i l_i^2 \sigma^2_{El_ib_2}}{\left(\sum_i l_i\right)^2} \tag{29}$$

The extension variances corresponding to each of the 33 transects were obtained from Chart no. 8 (in Mining Geostatistics, Journel and Huijbregts, 1978, p132) using a constant $L/a_s = 30/76 = 0.395$. The variances were then multiplied by the square of the length of the transect. The section term was then obtained as follows:

$$T_s = \frac{175020}{(6855)^2} = 0.0037$$

The slice term

The most significant simplifications in applying this method of calculating the estimation variance were made here. In order to use the extension variance chart (Chart no. 9, in Mining Geostatistics, Journel and Huijbregts, 1978, p133), it was first necessary to discretise the spawning area into square cells. Because the kriging was undertaken on a 15 x 15 nautical mile grid, these dimensions were used in the calculation of the variance. The error term is then based on the variance of extending these square planes, which are assumed to be perfectly known, to their time periods of influence. The data were analysed to work out the number of times each 15 x 15 square was revisited during the 1998 survey campaign. The total length of the spawning period is around 175 days, so that the width L of each slice is simply given by 175 divided by the number of samples.



The slice term is obtained as follows:

$$T_v = \frac{1}{V^2}\sum_j v_j^2 \sigma^2_{Ej} \tag{30}$$

where V is the total volume and $v_j = v$ are the volumes of the slices. As before, the individual error contributions for each slice were weighted according to the square of their volume. Using variable $L/a_t = L/65$ and constant $l/a_s = 15/76 = 0.197$ the slice term is calculated as:

$$T_v = \frac{1.877 \times 10^{12}}{1.295 \times 10^{15}} = 0.0015$$

The three error terms obtained above can then be summed to give the relative estimation variance:

$$\sigma_E^2 = T_l + T_s + T_v = 0.0019 + 0.0037 + 0.0015 = 0.0071$$

The CV is then given by the square root of the relative estimation: $\sqrt{0.0071} = 0.0843$, or 8.43%.

It must be emphasised that due to the spatio-temporal complexity of the actual survey paths, a number of assumptions and simplifications have had to be made while implementing this method. Care must therefore be taken while analysing its results.

Table 4 presents the results obtained for the remaining surveys. Relative variograms were calculated for each of the datasets, and their spatio-temporal locations were analysed to calculate the average segments of influence for use in the calculation of the three error terms. The variogram models were restricted to a nugget component and a spherical component in order to use the above-mentioned charts. In Table 4, the ranges a_s, a_t and Sill refer to the parameters of the spherical component.

The CVs listed in Table 4 range between 15.9% for 1977 and 6.5% for 1989. The large CV obtained for 1977 is mainly derived from the line term. The line term is the only component which accounts for the nugget effect, which is a relatively significant structure in the 1977 variogram. The line term T_l is also inversely proportional to the square of the number of samples in the dataset, which is relatively small in the first triennial survey. The slice term T_v also contributes significantly towards the high CV for 1977. This can be attributed to a poor temporal coverage. On average, each latitudinal transect was visited only 1.9 times during the spawning season. The section term, in contrast, is relatively small, due to the very large range a_s of the variogram in the spatial plane.

Table 4. CVs for TAEP estimations calculated using method of combination of error terms

Year	TAEP	Nugget	a_s	a_t	Sill	No. Data	T_l	T_s	T_v	CV
1977	2.28×10^{15}	7.65	180	65	15.35	436	0.0200	0.0015	0.0037	15.9%
1980	1.77×10^{15}	0.00	65	45	23.00	619	0.0075	0.0039	0.0029	12.6%
1983	1.46×10^{15}	0.07	110	60	12.93	727	0.0024	0.0021	0.0020	8.0%
1986	2.15×10^{15}	0.31	110	85	13.69	690	0.0017	0.0022	0.0014	7.3%
1989	2.41×10^{15}	0.66	160	65	10.34	838	0.0013	0.0012	0.0017	6.5%
1992	1.91×10^{15}	0.00	190	50	18.00	788	0.0011	0.0019	0.0037	8.2%
1995	1.93×10^{15}	6.39	90	50	5.61	1007	0.0069	0.0010	0.0008	9.3%
1998	1.47×10^{15}	0.19	76	65	17.81	951	0.0019	0.0037	0.0014	8.4%
2001	1.36×10^{15}	1.82	100	40	14.18	1165	0.0022	0.0024	0.0019	8.1%

The smallest CV is obtained for 1989. This seems a little surprising, as this survey does not have the greatest spatio-temporal coverage. While the relative variogram has a similar spherical component in terms of its range, the nugget and sill are considerably smaller, and the number of samples in the dataset is almost twice as much as in 1977. It would appear that the low CV for 1989 is mainly due to the low sill of the relative variogram.

Table 4 revealed that, in using the combination of errors method of variance calculation, there were many factors affecting the uncertainty of the TAEP estimates. This made it difficult to assess the strengths of the different sampling campaigns. An experiment was therefore conducted to compare the different surveys by using identical relative variograms. This allows us to assume that the properties of the egg density surface are identical throughout time, and so the CV is dependent only on the nature of the sampling campaign. Figure 25 plots the CVs calculated for each of the surveys when a selection of the relative variograms are used.

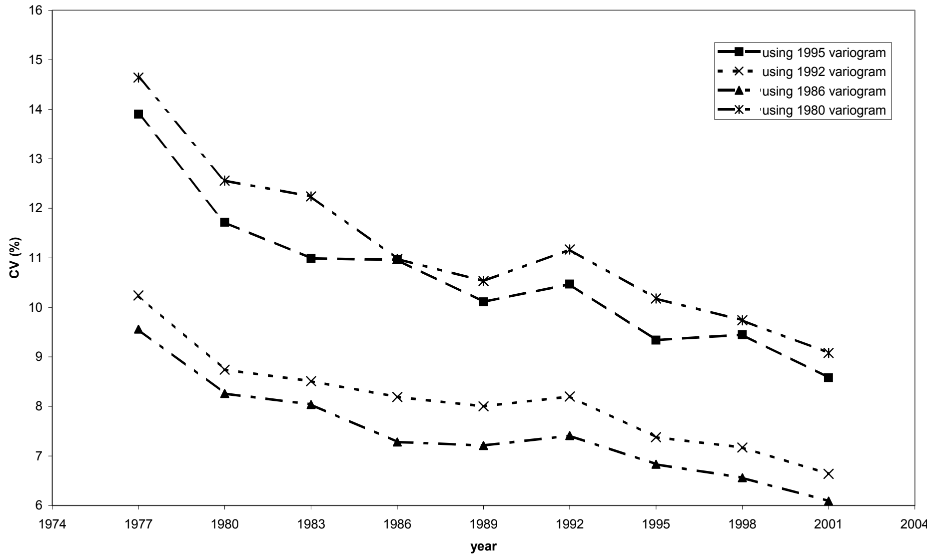


Figure 22. CVs calculated using identical variogram models for each survey campaign.

Each of the curves in Figure 25 displays a decreasing trend between the first survey in 1977 and the most recent in 2001. The trends suggest an overall improvement in the survey design over time, resulting in increased confidence in the TAEP estimates. The anomalous peak for the 1992 campaign displayed for each of the curves is due to the fact there are fewer samples in the dataset than there are for 1989 and 1995. This effect is most pronounced when the variogram has short ranges (e.g. the 1980 variogram), and less pronounced when the ranges are longer (e.g. the 1986 variogram). When the 1995 variogram is used, the CV calculated for the 1998 survey campaign is slightly greater than that obtained for the 1995 campaign. This can also be attributed to the greater number of samples available for 1995, which is particularly important when there is a large nugget component. The overall reduction in CV gained by improving the spatio-temporal coverage between 1977 and 2001 is in the region of 4.3%.

The results presented above indicate that the method of combination of errors can be used to provide a measure of uncertainty corresponding to the TAEP estimates. However, considerable simplifications and assumptions have been made in order to implement the method with the irregularly spaced data. For example, linear averaging has been done to allocate temporal segments of influence to individual transects, so that it is assumed that the data are regularly spaced in time. In many cases this is far from the truth, and there are often large gaps in the sampling campaign. On these occasions, the simple averaging of the data may lead to large underestimations of the uncertainty. A major disadvantage of the current methodology was that the procedure is manual, and would therefore be difficult to implement within a Bayesian framework in WP3. Furthermore, errors from misreading the charts would be easily made.

Conditional Simulation

In general, conditional simulation algorithms model uncertainty by creating a set of R realisations of the spatial distribution of a variable. Each realisation is conditioned on the original data, and approximately reproduces the sample histogram and variogram. When a large number of realisations have been created, the variance of their mean value allows us to calculate the estimation variance. This method is therefore akin to bootstrap resampling methods for modelling uncertainty (Thayer *et al.*, 2000).

There are a number of different simulation algorithms available, such as sequential simulation, turning bands and simulated annealing. Sequential simulation is a popular and computationally efficient method for generating simulations that aim to honour the histogram and variogram of the dataset (Caers, 2000). Sequential Gaussian Simulation (SGS) is often employed as it allows the simulated values to be drawn from Gaussian distributions whose parameters are determined by the solution of a simple kriging system. An investigation into the potential of using simulation techniques to measure the uncertainty in TAEP estimation was therefore undertaken using the SGS algorithm.

Conditional SGS proceeds as follows (e.g. Goovaerts, 1997 p. 380; Deutsch & Journel, 1998):

Data transformation.

When using SGS, the data must be normally distributed. If the raw data are skewed, a normalisation can be achieved by performing a normal score transformation. The *n* sample data are ordered according to their value, and allocated a rank *k*. The normal score transformation is given by the *k*/*n* quantile of the standard normal cumulative density function (cdf). It is usually necessary to specify a maximum possible value and the rate at which it can be approached.

Variogram modelling

A variogram model is created using the new Gaussian variable.

3. Sequential simulation

The simulations are performed on a two or three-dimensional grid with a size and resolution specified by the user. In this application the grid shown in Figure 21 was used.

i. The data points are assigned to a grid node closest to their location (this is an option designed to speed up the simulation procedure), and these nodes remain fixed.
ii. A random path, which visits each remaining node of the grid once, is defined.
iii. At each node, the parameters (mean and variance) of the Gaussian conditional cdf (ccdf) of the estimate are determined using SK (simple kriging) with the normal score variogram model. Alternatively, if a collocated secondary variable is available (in our case depth), cokriging is used rather than SK. The algorithm uses real samples and previously simulated values close to the node to define the ccdf. A neighbourhood, which limits the number and distance of these values, can be specified. In this application a neighbourhood similar to that used above was specified.
iv. A simulated value is then drawn from that ccdf and added to the grid.
v. We then proceed to the next node on the random path and repeat the two previous steps. This process is continued until all nodes have been simulated.

Back-transformation

The cdf of the original data is used to back-transform the simulated values. The values can then be used to make an estimate of the variable of interest. For the calculation of TAEP, the egg density estimates are scaled up to the size of the grid cells and summed over the spawning area and the spawning season.

This procedure may be used to generate a number of different realisations by changing the random path. The variance of the different TAEP estimates can then be used to calculate the CV, and hence develop confidence intervals.

However, there are a number of drawbacks associated with SGS (Caers, 2000; Soares, 2001):

- The assumption of multi-Gaussianity leads to simulated realisations that have maximum entropy, so that the extreme values are intentionally disconnected. This often conflicts with reality.
- SGS reproduces only the normal score histogram and variogram. When the original data are highly skewed, the reproduction of the histogram and variogram after back-transformation are not guaranteed.

The SGS algorithm in GSLIB (**sgsim**, Deutsch & Journel, 1998) was used to make stochastic variance estimates for the TAEP using the data from each of the triennial surveys. The software is provided with its source code in Fortran 77. In order to speed up the simulation process, the program was modified so that it simulates only those nodes that are inside the designated spawning area. This was achieved by creating a file that allocates a ‘1’ or ‘-99’ flag to each node, depending upon whether it is to be simulated or not. The program reads the file at the beginning of the algorithm, and the value corresponding to each node along the random path is checked before the simulation is done. This has reduced the processing time considerably. A further modification enabled the program to calculate the TAEP for each egg production surface, and after the required number of simulations had been done, the program outputs the mean TAEP and the associated CV.

Examples of the variograms modelled for the Gaussian-transformed data of 1989, 1998 and 2001 are provided in Figure 26.

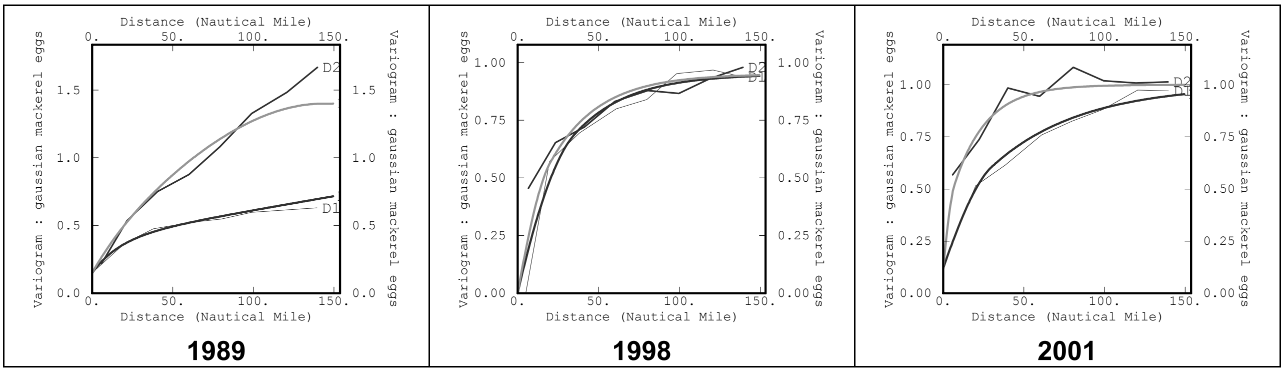
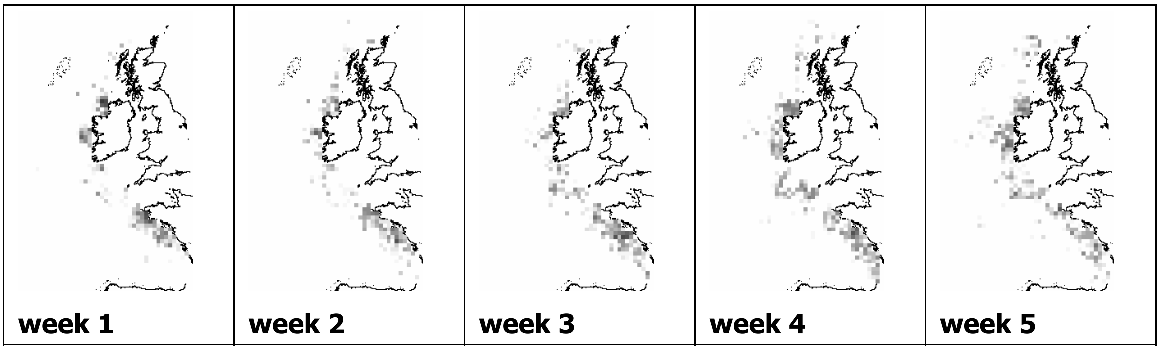


Figure 23. Variograms modelled for 1989, 1998 and 2001 data after normal scores transformation. The grey model represents the temporal direction (units of days), while the black model represents the omnidirectional spatial variogram.

For each of the surveys, 100 realisations were generated. The standard deviations of the estimates were calculated and used to determine the CVs. The results are presented in Table 5. For comparison, the TAEP estimates obtained using Ordinary Kriging (OK) and the Traditional and GAM methods are also included.

Table 5. Comparison of TAEPs calculated using Traditional, GAM, collocated kriging and SGS methods, with the CVs calculated for the two geostatistical methods

	Traditional	GAM	OK		SGS	
Survey	TAEP	TAEP	TAEP	CV	TAEP	CV
1977	1.98×10^{15}	2.10×10^{15}	2.28×10^{15}	15.9%	1.64×10^{15}	13.0 %
1980	1.84×10^{15}	2.01×10^{15}	1.77×10^{15}	12.6%	2.14×10^{15}	9.8 %
1983	1.53×10^{15}	1.28×10^{15}	1.46×10^{15}	8.0%	1.45×10^{15}	7.8 %
1986	1.24×10^{15}	1.77×10^{15}	2.15×10^{15}	7.3%	2.09×10^{15}	11.2 %
1989	1.52×10^{15}	2.49×10^{15}	2.41×10^{15}	6.5%	1.72×10^{15}	8.7 %
1992	1.94×10^{15}	1.63×10^{15}	1.91×10^{15}	8.2%	1.78×10^{15}	14.8 %
1995	1.49×10^{15}	1.80×10^{15}	1.93×10^{15}	9.3%	2.26×10^{15}	8.0 %
1998	1.37×10^{15}	1.18×10^{15}	1.47×10^{15}	8.4%	1.44×10^{15}	8.3 %
2001	1.21×10^{15}	1.44×10^{15}	1.36×10^{15}	8.1%	1.18×10^{15}	7.5 %



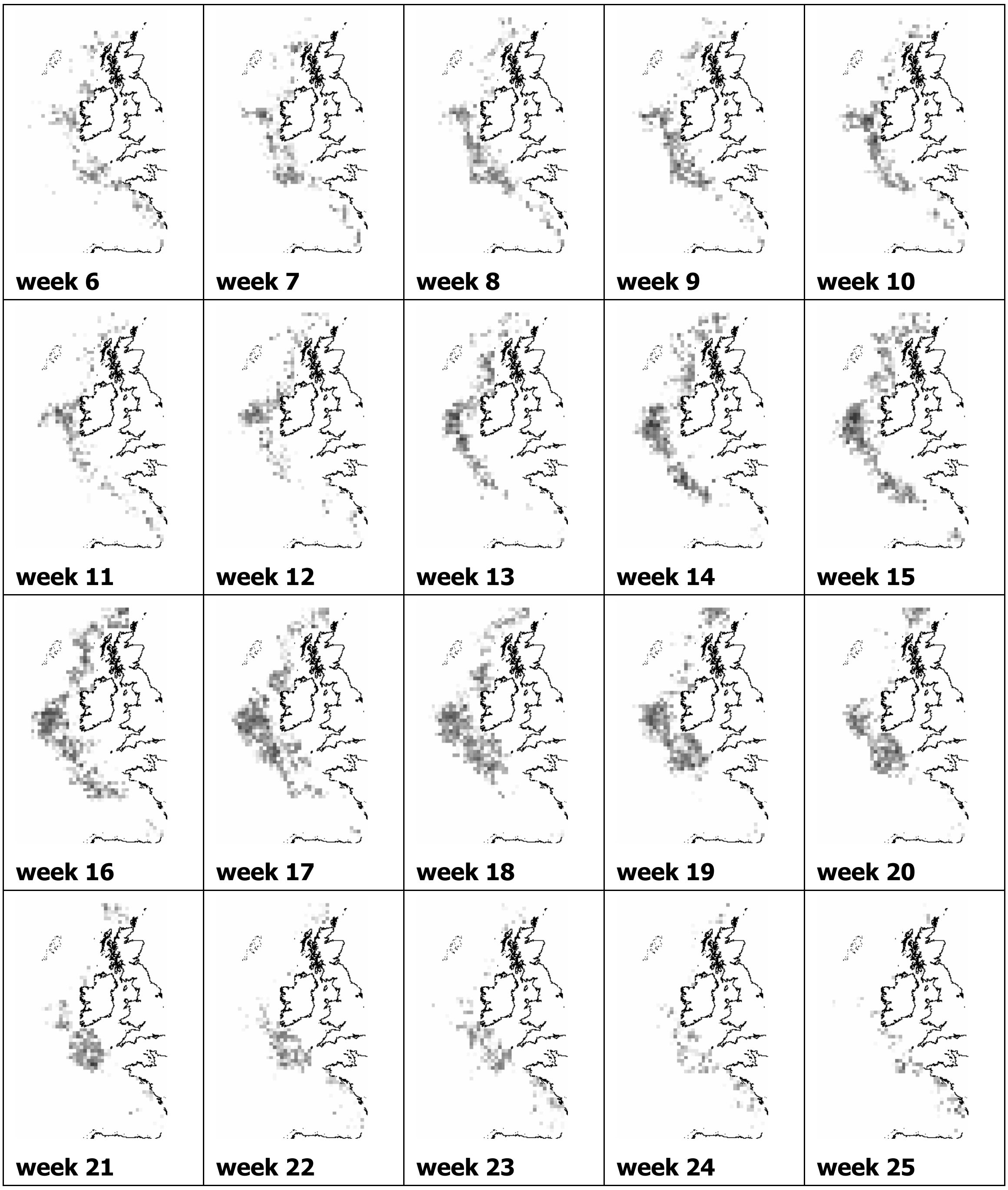


Figure 24. Example of an SGS realisation based on the 2001 dataset.

As can be seen from Table 5, the TAEP estimates made by SGS are in general comparable with those obtained using ordinary kriging, although significant differences occur for 1977 and 1989. One characteristic of the 1989 dataset is that it contains a number (5) of particularly high egg densities. This could have resulted in a higher correction factor than would have been obtained if more than one of these values had been omitted from the bias calculation.

There are considerable differences in TAEP between the estimators for a number of the triennial surveys. For example, while the estimates for 1977, 1980 and 1983 are reasonably consistent, the geostatistical estimates of TAEP are considerably higher than the Traditional estimate for 1986. The 1986 dataset is characterised by a delayed start to the survey campaign, with no egg densities available prior to mid-May. The traditional method has dealt with this in a far less generous way than the geostatistical methods, which are able to interpolate over this period in a more realistic way.

In general, the CVs calculated using the combination of error terms and SGS methods are similar. Both methods find that the lowest CV is obtained for 2001, as expected. Both methods also result in a somewhat unexpectedly low CV for 1989. The methods differ however in the calculation of CV for 1992, with the combination of error terms CV (8.2%) far lower than that of the SGS method (14.8%). An explanation for this may reside in the nature of the 1992 survey. Although there were a high number of samples collected in 1992, many of these were concentrated in a particular part of the spawning area, and there were no samples collected before week 10 of the spawning season. The simplifications necessary for the implementation of the combination of error terms method may have caused the samples to be represented as more regular than they actually were, leading to an underestimation of the uncertainty.

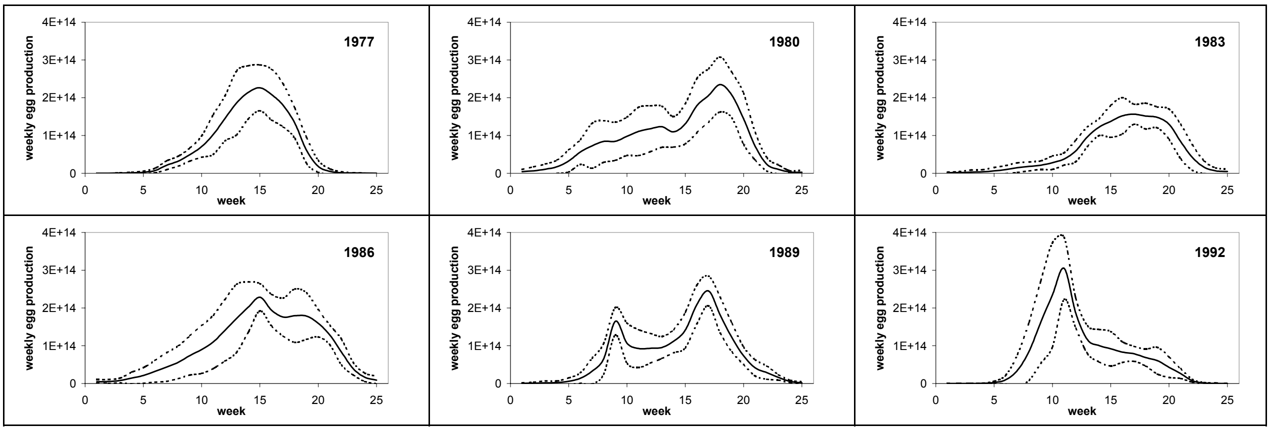
An example of a realisation based on the 2001 dataset is provided in Figure 27. It can be seen that the simulation retains many of the features of the spawning season, with reduced egg production at the start and end of the season, and towards the edges of the spawning area. It may also be noted that the images are less smooth than those obtained by kriging (see Figure 22).

The results suggest that SGS is an appropriate methodology for the estimation of TAEP and its uncertainty. The method is preferable to the ordinary kriging method due to the simplicity with which the uncertainty is obtained, and that fact that there are less simplifications and assumptions necessary. The methodology is straightforward to implement within a Bayesian framework. This will be described in more detail in the following section. A further potential advantage is the ability to generate equiprobable surfaces which keep the histographic and variographic properties of the original dataset. This may offer the potential to generate simulated datasets from which to test survey designs. An attempt to do this is described in WP4.

Figure 28 plots the TAEP curves for each of the triennial surveys calculated using the SGS method. The dotted lines give indications of the 95% confidence intervals. These have been calculated using the standard deviations around the estimations of egg production for each week over all the realisations, and assume that these are independent. However, they do offer some useful information. For example, the curves for 1986 and 1992 have wider confidence intervals at the start of the spawning season, where the start of sampling has been delayed. The early spawning peak of 1992 suggests that a large number of eggs may have been missed, thereby contributing to the large uncertainty associated with the TAEP estimate.

The shapes of the curves in Figure 28 vary markedly between the years. The curves of 1980, 1983 and 1995 suggest a relatively late spawning peak, whereas 1992 displays an early spawning peak. The curves of 1986 and 1989 are bimodal. Although these interannual differences may be due to climatic fluctuations, it may also be feasible that they are due to the age structure of the spawning mackerel. For example, it is known that the older and larger fish tend to start spawning earlier and migrate faster than the younger fish (Lockwood, 1988). Data on the annual stock components were obtained from ICES (2002b). The data are based on catch in numbers-at-age, which have been corrected for selectivity.

The numbers-at-age curves for 1980, 1983, 1995 and 2001 suggest that there are a relatively large number of younger fish, which could help to explain the late peak in egg production. Conversely, the age curve for 1992 shows a relatively low number of young fish. The curves for 1986 and 1989 are distinctly bimodal.



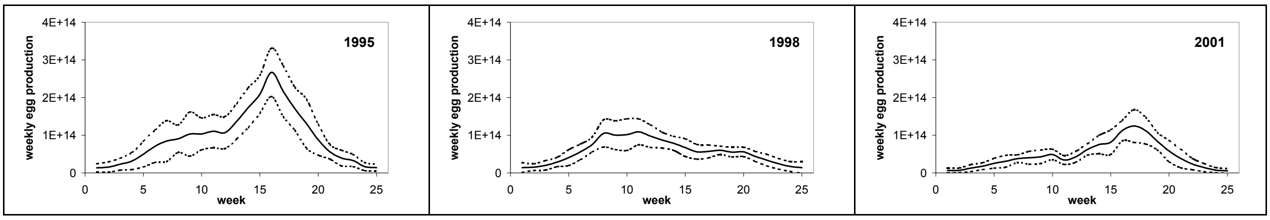


Figure 25. Weekly egg productions for each of the triennial surveys. The dotted lines indicate the upper and lower 95% confidence intervals.

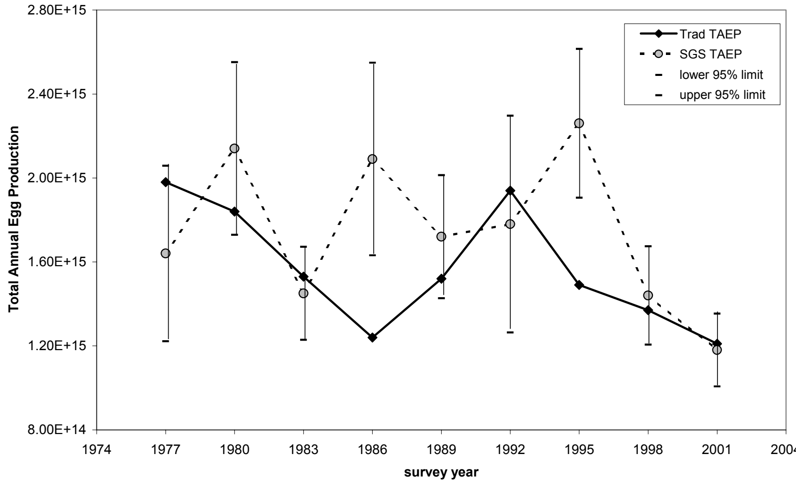


Figure 26. Plot of TAEP estimated by SGS and the Traditional method.

If information on the age structure of mackerel is known prior to the commencement of a triennial survey, it may be possible to adapt the design of the sampling campaign accordingly.

It should be noted that the use of conditional simulation in fisheries is relatively new, and there is a need for more research to be undertaken before the methodology could be used in practice. The procedure has a number of parameters, besides those of the variogram model, which are assumed to be specified correctly. While the Bayesian analysis will consider the uncertainty due to a number of these parameters, more research will need to be undertaken to appreciate the robustness of the algorithm.

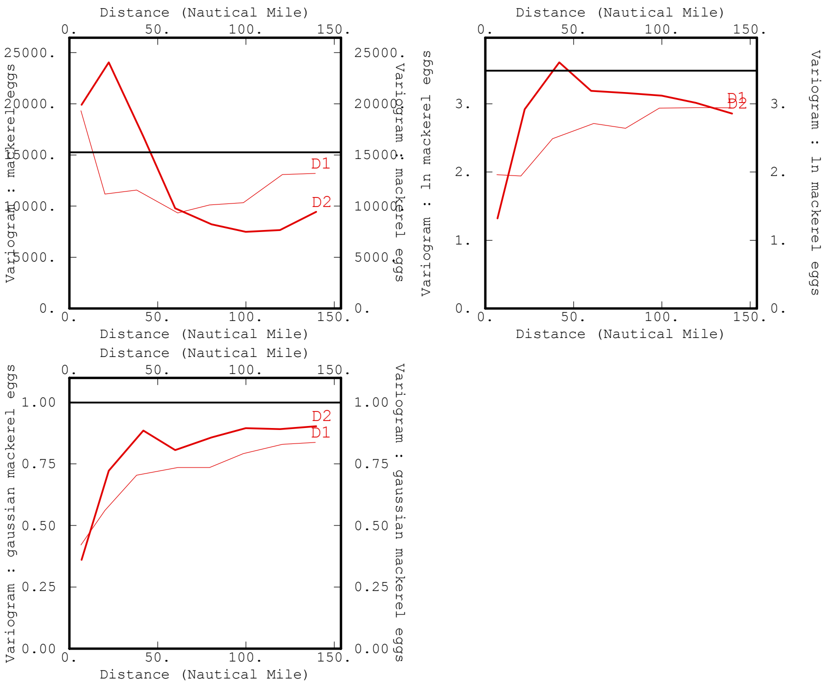
The final TAEP estimations obtained using SGS are plotted with their 95% confidence intervals for each of the triennial surveys in Figure 29. The Traditional estimates are also provided. The procedure and the results presented satisfy deliverables D4 and D5.
The research presented towards the end of this section was used to prepare a paper which has been sent to the Canadian Journal of Fisheries and Aquatic Sciences. This is provided in APPENDIX I., and is currently in the hands of reviewers.

Incorporation of geostatistical and Bayesian analysis techniques for egg survey data modelling

The work undertaken towards the Bayesian-geostatistical estimator is described in the following pages. An initial analysis of the egg abundance datasets is first presented, with a subsequent Bayesian assessment of the use of various parameters in the Traditional TAEP estimator. This is followed by a description of the work towards developing the Bayesian Geostatistical estimator. The work involved an initial study into Bayesian estimation of variogram model parameters, then a progression towards a more advanced methodology based on hierarchical modelling for incorporation into the geostatistical procedure developed in the previous section.

Bayesian analysis of variogram model parameters

Throughout the research in GBMAF a variety of different variogram models for each of the triennial surveys have been presented. Figure 33 shows how the experimental variogram shape changes according to the transformation applied to the 1995 mackerel egg data. The experimental variogram obtained for the raw data (Figure 33 (a)) would appear to be a pure nugget effect. Log-transforming the data prior to calculating the experimental variogram reveals a much clearer spatial correlation structure among the egg density data (Figure 33 (b)), while the clearest structure is obtained when the data are normalised (Figure 33 (c)). These differences indicate how experimental variograms calculated using highly skewed raw data can mask existing spatial correlation features. Some practitioners recommend that the experimental variogram be first calculated using the log-transformed data, so that the values can be back-transformed and used to model a variogram in raw space (e.g., Rivoirard *et al*., 2000). However, it is evident that the variogram fitting procedure may represent a considerable source of uncertainty.



(a) (b)

(c)

Figure 27. (a) Raw experimental variogram for 1995 mackerel egg survey data; (b) experimental variogram for log-transformed 1995 data; (c) experimental variogram for normal scores transformed 1995 survey data. The solid line represents the omnidirectional spatial variogram, while the dashed represents the variogram for the temporal direction.

A further note on variogram fitting regards generalisation. It is important to avoid over-fitting, as the objective is to capture the main spatial correlation characteristics of the variable. Therefore, it is recommended that the number of structures used to model the variogram is kept to a minimum (Goovaerts, 1997).

Initial investigation into Bayesian estimation of variogram model parameters

The impact of both model and parameter choice in the final estimate was initially investigated by a simple Bayesian procedure based on Monte Carlo techniques (Gelman *et al*., 1995; Gamerman, 1997; Chen *et al*., 2000). Variogram parameters were randomly chosen from a uniform prior distribution with proper limits, or a more informative prior distribution, and then plugged into the kriging procedure. The question of model choice was tackled with the use of the Matérn distribution (Stein, 1999), whose shape parameter allows it to change between the spherical and exponential models. A likelihood function linking each set of parameter values and their probability given the data was developed. This procedure allows for estimation of what is probably the largest source of uncertainty, and has the flexibility to accommodate different prior probability distributions without the constraints of a conjugate formulation (Handcock & Stein, 1993). The assumption of normality (or log-normality) in the model deviates present in the conjugate formulation might not be correct, and a negative binomial distribution could be a better description of the sampling error.
In contrast, conjugate Bayesian methods make use of the properties of mixtures of probability distributions belonging to the same family. When this is the case, posterior probabilities can be described algebraically, sparing the use of numerically intensive methods of integration. However, they are limited to a number of families of probability distributions, which limits greatly the range of prior distributions that can be used.

Initial work was focused on developing posterior probabilities for the parameters of the following simple model structures:

- nugget + spherical

$$\gamma(h) = C_0 + C\left[\frac{3}{2}\frac{h}{a} - \frac{1}{2}\frac{h^3}{a^3}\right]$$

- nugget + exponential

$$\lambda(h) = C_0 + C\left[1 - \exp\left(-\frac{|h|}{a}\right)\right]$$

- nugget + Matérn function

$$\lambda(h) = C_0 + C\left[\frac{1}{2^{k-1}\Gamma(k)}\left(\frac{h}{a}\right)^k K_k\left(\frac{h}{a}\right)\right]$$

where $C_0 + C$ is the total sill and a is the range. In the Matérn finction, K is the modified Bessel function of the third kind, k is the shape parameter and Γ is the gamma function. Although the shape of the Matérn function can approximate that of the exponential and spherical functions, it is horizontal at |h| = 0, which makes it appear similar to the Gaussian function.

A maximum likelihood approach was used to obtain posterior distributions for the parameters of each of the models when fitted to the log-transformed experimental variograms. The width of the posteriors was used as a criterion to try to identify the model that fitted best. It was observed that, which the Matérn function offered flexibility with regard to its shape, the 'Gaussian' nature at |h| = 0 meant that it fitted the experimental variograms poorly at short lag distances. There was little to choose between the spherical and exponential models.

A posterior distribution of exponential model parameters was obtained for the 1998 data, and the kriging procedure applied using a large number of variograms from this joint distribution. The correction factor k (see WP2) was calculated using the maximum likelihood variogram and then kept fixed. The posterior distribution of 1998 TAEP is shown in Figure 34. The CV was calculated to be 3.3%.

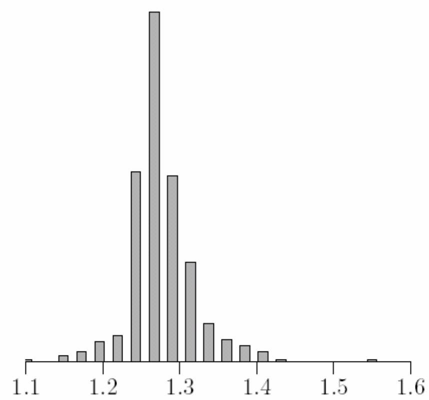


Figure 28. Posterior distribution of TAEP for 1998 obtained using Bayesian-Geostatistical kriging.

While the initial work described above provided a basic procedure for the Bayesian fitting of variogram model parameters, more work needed to be done to allow it to be compatible with the Geostatistical TAEP estimation process. The further development work is described in the following paragraphs.

Hierarchical variogram modelling

The geostatistical methodology developed in the Section 2.2.6.2 involved using Sequential Gaussian Simulation (SGS) to obtain estimates of TAEP and its associated uncertainty. A simple methodology for the Bayesian-Geostatistical estimator would therefore be to compile a set of variogram model parameters from a joint probability distribution for each survey, and then to use this as input into the SGS algorithm. Since the algorithm is somewhat slow with the available computing facilities, SGS will generate 50 equiprobable egg production surfaces for each of 500 variogram models in turn. This will create a distribution of 25000 TAEP values from which to calculate the mean TAEP and associated CV.

The use of SGS involves transforming the data using a normal scores procedure prior to calculating and modelling the variogram. Therefore, the variograms modelled in the current study were also based on the normalised data.

While the variogram models described in the pervious section were manually fitted, many researchers use automatic fitting procedures such as weighted least squares (Cressie, 1993). The Bayesian approach employs a likelihood function to generate joint posterior probability distributions for the variogram model parameters. Although it is recommended to keep the number of parameters to a minimum, it was considered that limiting the variogram to a strictly exponential or spherical shape could lead to a distribution that did not cover all the possibilities. The variograms previously modelled were often nested, combining a nugget with both spherical and exponential components. The likelihood function was therefore set up to estimate 7 parameters, as follows:

1. nugget, C_0;
2. sill corresponding to spherical component, C_{sph};
3. range corresponding to spatial spherical component (omnidirectional), a_{sph_xy};
4. range corresponding to temporal spherical component, a_{sph_z} ;
5. sill corresponding to exponential component, C_{exp};
6. range corresponding to spatial exponential component (omnidirectional), a_{exp_xy};
7. range corresponding to temporal exponential component, a_{exp_z}.

The spatial and temporal variogram models are expressed as:

$$\gamma(h)_{\mathrm{xy}} = C_0 + C_{\mathrm{sph}}\left[\frac{3}{2}\frac{h}{a_{\mathrm{sph\_xy}}} - \frac{1}{2}\frac{h^3}{a_{\mathrm{sph\_xy}}{}^3}\right] + C_{\mathrm{exp}}\left[1-\exp\left(-\frac{|h|}{a_{\mathrm{exp\_xy}}}\right)\right]$$

$$\gamma(h)_{\mathrm{z}} = C_0 + C_{\mathrm{sph}}\left[\frac{3}{2}\frac{h}{a_{\mathrm{sph\_z}}} - \frac{1}{2}\frac{h^3}{a_{\mathrm{sph\_z}}{}^3}\right] + C_{\mathrm{exp}}\left[1-\exp\left(-\frac{|h|}{a_{\mathrm{exp\_z}}}\right)\right]$$

Since there are 9 sets of survey data, there were 9 individual experimental variograms to be modelled. These were generated after normal scores transformation with a lag spacing of 10 nautical miles/days. In order to place a degree of emphasis on the part of the variogram closest to the origin, as is done in practice, the lag spacing was thinned to 20 nautical miles/days beyond a total separation distance of 100. This meant that the models were to be fitted to 13 data points. This is rather low when considering the number of parameters to be estimated (7). However, extra information can potentially be gleaned by pooling together the entire set of experimental variograms in a procedure called hierarchical modelling.

Models for a certain process in separate populations of the same species, or any other situation in which separate estimates can be considered to be related to each other, can be elegantly accommodated in a Bayesian hierarchical model (Gelman *et al.*, 1995). Parameter values for each single realisation of the process, in our case spatio-temporal variograms of egg abundance in different seasons of mackerel spawning, θ_j, are viewed as samples from a common population distribution. This is achieved by setting prior probability distributions to the mean and variance of θ. These hyperparameters provide a measure of the degree of relatedness among the individual realisations, and improve the estimate of θ_j when data are scarce or uninformative. Non-hierarchical models might also overfit to the present data, limiting the usefulness of the model when applied to subsequent years or new experiments.

As an example, an exponential model for both temporal and spatial variograms could be formulated as a hierarchical model, whereby the model parameters (nugget, sill, range in space, and range in time) would be considered to be normally distributed, with mean and variance drawn themselves from a normal and a lognormal distribution.

The advantages of this approach are mainly twofold. The hyperparameter posterior probability indicates how similar or otherwise the variograms for each year really are, and how much can we expect them to change in future surveys. Secondly, it helps the model fitting procedure by using the information gathered from all years in those with more complex or poorly behaved experimental variograms.

The experimental variograms calculated for the normal scores transformed mackerel egg survey data are plotted in Figure 35. Since the overall covariance in the transformed data is 1, it can be expected that the sills of the variograms will approximate unity.

Hierarchical models assume complete exchangeability of the data. If one year of data was to be removed from the analysis, results should not be significantly affected (Gelman *et al.*, 1995). However, an inspection of the sample variograms in Figure 35 reveals two distinct patterns. Some years (e.g. 1980, 1992 and 1998) display a clearly asymptotic behaviour, whereas others (1983, 1986 and 1989) appear to be more linear. To force both types into one single group would be counterproductive. The years were therefore divided into two groups. The idea was to first model the ‘good’ set of variograms (1977, 1980, 1992, 1995, 1998 and 2001) using uninformative priors, and then use the posteriors obtained to help model the ‘bad’ years.

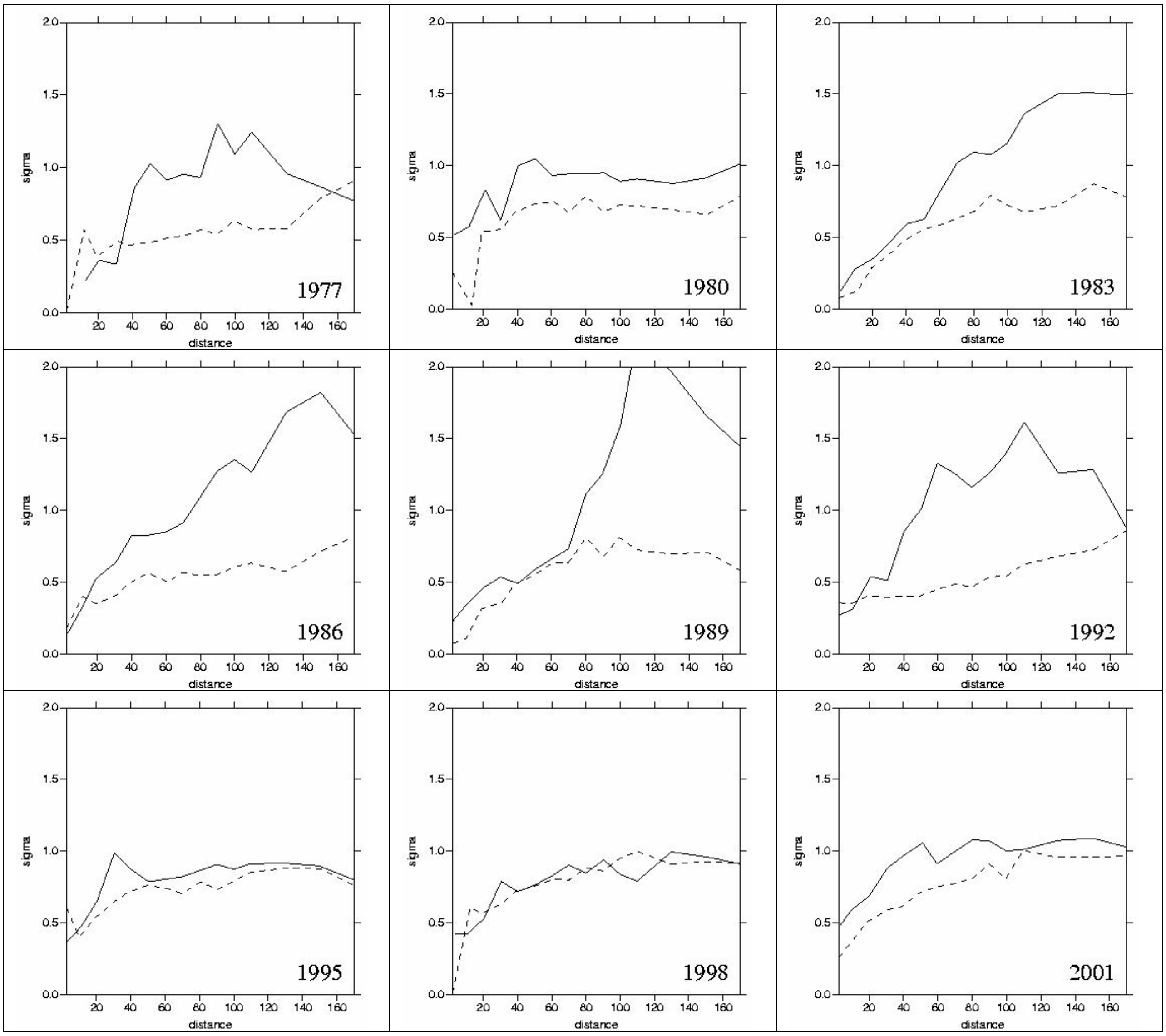


Figure 29. Spatio-temporal experimental variograms for each of the triennial surveys.

The likelihood function was set up with a normal pdf, whereby the variance was estimated using the CV of the residuals multiplied by a correction factor of 1.2. The priors for all the parameters were specified with very wide, although not completely uninformative, distributions. The means were given normal distributions whereas the variances were given lognormal distributions. The distributions for all the parameters were truncated at just greater than zero. The distributions of the hyperparameters are given in Table 8.

Table 6. Prior distributions for the hyperparameters

Parameter	Distribution	Lower limit	Upper limit
Sill mean	N(1, 10^2)	1 x 10^{-8}	1 x 10^{28}
Range mean	N(100, 10^2)	1 x 10^{-8}	1 x 10^{28}
Nugget mean	N(1, 50^2)	1 x 10^{-8}	1 x 10^{28}
Sill variance	L(10^2, 200^2)	1 x 10^{-8}	1 x 10^{28}
Range variance	L(10^2, 200^2)	1 x 10^{-8}	1 x 10^{28}

Nugget variance	L(50^2,20^2)	1 x 10^{-8}	1 x 10^{28}

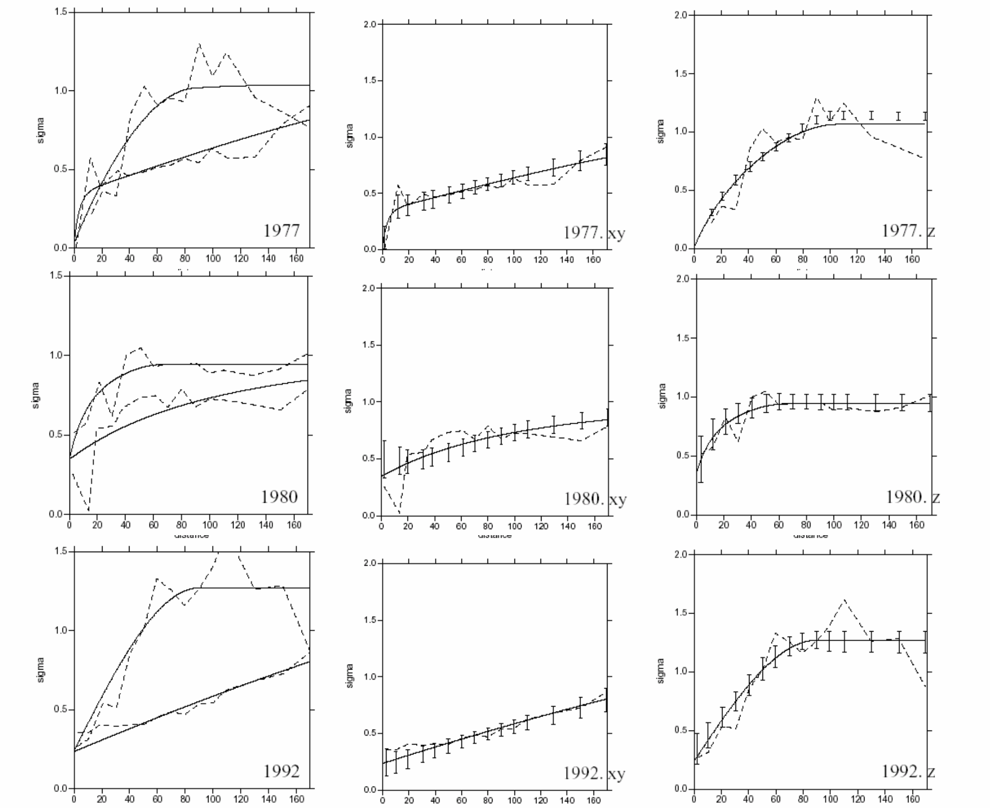
The parameters of the variogram models, sill, range and nugget, are then distributed according to N(sill mean, sill variance), N(range mean, range variance) and N(nugget mean, nugget variance) respectively. The posterior distributions obtained were used as priors for the abnormal years (1983, 1986 and 1989). The Bayesian variogram modelling procedure was implemented in MCSIM, an open source C-language based package that uses a common Metropolis-Hastings Markov Chain Monte Carlo process. The code was tailored for the specific needs of the current study.

The Monte-Carlo procedure was run 3,000,000 times, so that the chains of variogram model parameters for each year included in the hierarchical analysis were fully stabilised. Convergence was tested by inspection of both the chain and posterior distributions, and by the methods incorporated in the CODA package (Best *et al*., 1996). The sets of 500 nested variogram models to be passed to the SGS procedure were then extracted by thinning the final 1,000,000 runs. Figure 36 shows the maximum likelihood variograms with their confidence intervals.

It was observed that the 7-parameter nested models were in general better fitted to the experimental variograms than those obtained using singular structures based on the spherical, exponential and Matérn functions.

As indicated in Figure 36, the confidence intervals tend to be widest at small separation distances. This is because there is considerable variability amongst the years regarding the size of the nugget component. This variability may be explained by the lack of fine scale (<10 nautical miles) sampling resolution. The largest error bars appear to apply to the 1998 variogram, whereas the shortest error bars are associated with the 1977 and 1992 variograms.

The GSLIB SGS procedure (**sgsim**, Deutsch & Journel, 1998) was modified to automatically read the variogram model parameters in turn from a separate file. The algorithm reads the first set of parameters and simulates 50 equiprobable egg density surfaces. The individual TAEPs and the summary statistics (mean TAEP and CV) are output to a file. The algorithm then reads the parameters of the next variogram model and a further 50 simulations are generated. This procedure continues until each of the variogram models has been used. Current limitations in processing speed meant that the results for each set of survey data were obtained in around 5 days.



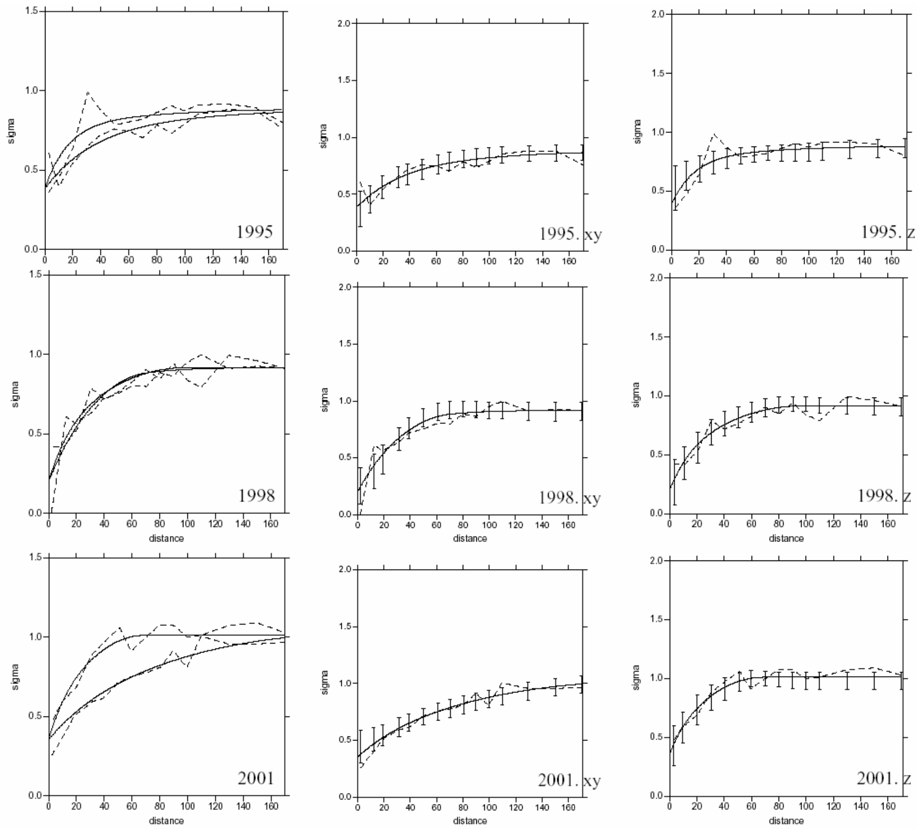


Figure 30. Maximum likelihood variogram models (left) with the confidence intervals for the spatial (centre) and temporal (right) components.

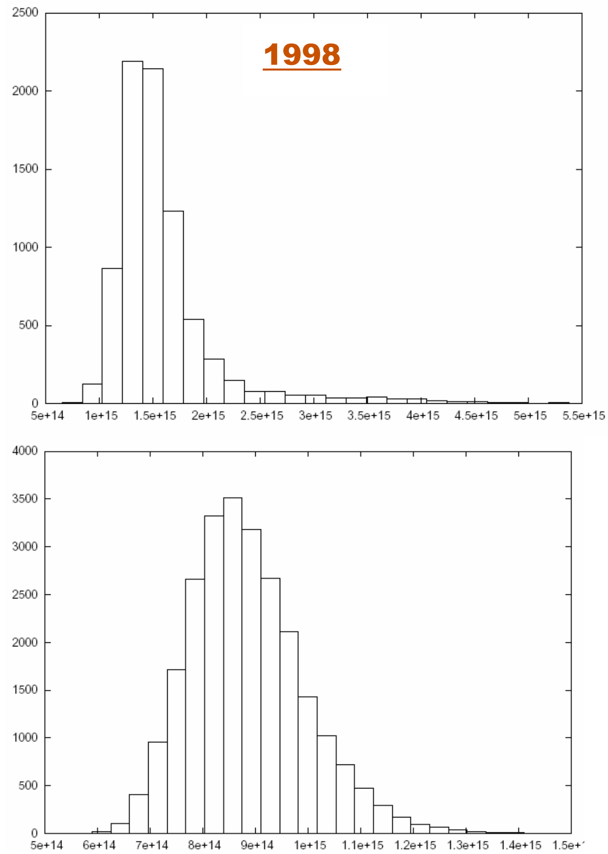


Figure 31. Examples of histograms of 25000 TAEP estimates for 1998 and 2001.

A summary of the results is provided in Table 9, along with the TAEP estimates achieved using the Traditional, GAM and SGS methods. Unfortunately it was not possible to obtain results for the years 1983, 1986 and 1989, due to problems with the simulation algorithm associated with the very high sills and long ranges of the variograms. Since the TAEP distributions tended to be highly positively skewed, as shown for 1998 and 2001 in Figure 37, the median, rather then the mean is provided in Table 9. The CVs are given to allow a very rough comparison, but due to the skewness these values cannot be used to provide estimates of the uncertainty. The 95% confidence intervals were instead obtained by ranking all the 25000 TAEP values and extracting those corresponding to the 2.5th and 97.5th percentiles. The median TAEPs and their 95% confidence intervals are plotted in Figure 38, alongside those originally obtained with constant variogram model parameters.

Table 7. Comparison of TAEPs and CVs for the different estimators.

	Traditional	GAM	SGS		BG	
Survey	TAEP	TAEP	TAEP	CV	TAEP	CV
1977	1.98×10^{15}	2.10×10^{15}	1.64×10^{15}	13.0 %	1.91×10^{15}	13 %
1980	1.84×10^{15}	2.01×10^{15}	2.14×10^{15}	9.8 %	1.85×10^{15}	23 %
1983	1.53×10^{15}	1.28×10^{15}	1.45×10^{15}	7.8 %	-	-
1986	1.24×10^{15}	1.77×10^{15}	2.09×10^{15}	11.2 %	-	-
1989	1.52×10^{15}	2.49×10^{15}	1.72×10^{15}	8.7 %	-	-
1992	1.94×10^{15}	1.63×10^{15}	1.78×10^{15}	14.8 %	1.33×10^{15}	16 %
1995	1.49×10^{15}	1.80×10^{15}	2.26×10^{15}	8.0 %	2.15×10^{15}	18 %
1998	1.37×10^{15}	1.18×10^{15}	1.44×10^{15}	8.3 %	1.47×10^{15}	32 %
2001	1.21×10^{15}	1.44×10^{15}	1.18×10^{15}	7.5 %	0.87×10^{15}	13 %

Table 9 shows that there are significant differences between the SGS and BG estimates of TAEP for 1977, 1980, 1992 and 2001. This demonstrates two points: the differences in the variogram parameters found subjectively and calculated using the hierarchical Bayesian method; and the sensitivity of the SGS algorithm to these variogram parameters. A sensitivity analysis revealed that the most influential parameter was the range (both spatial and temporal), and that TAEP decreases exponentially as range increases.

In the case of 1977 and 1980, the median TAEPs are very close to those obtained using the Traditional method. However, 1995 remains considerably higher, and the value for 2001 is significantly lower. The new 1977 estimate is higher than the original SGS value because the hierarchical maximum likelihood method selected ranges that were generally shorter than those modelled subjectively.

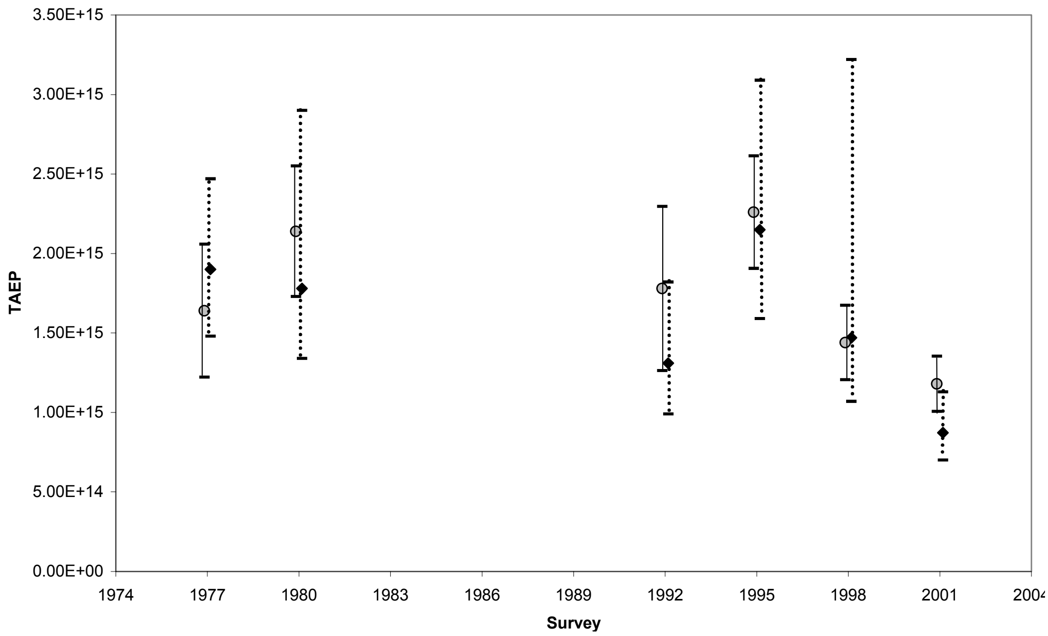


Figure 32. TAEPs and their 95% confidence intervals calculated by the SGS algorithm (grey circles, solid lines) and the BG method (black diamonds, dotted lines).

The confidence intervals obtained for 1995 and 1998 (Figure 38) using the Bayesian method are significantly larger than those obtained in WP2, where the uncertainty in the variogram model parameters was not taken into consideration. However, there is little difference between those obtained for 1977 and 1992. This can be largely explained by the size of the error bars allocated to the maximum likelihood variogram models in Figure 36. As noted above, the error bars for the variograms of 1977 and 1992 were smaller than those attributed to 1998. In each case, however, the lower 95% confidence limit is significantly lower than that obtained using the straightforward SGS method. This suggests that, if the original figures had been passed to a management model, significant and potentially dangerous errors could have been made.

The preceding paragraphs presented the findings of a novel approach towards accounting for variogram uncertainty. A framework was developed to then incorporate this uncertainty into the geostatistical procedure presented in Section 2.2.6.2. The methodology developed is theoretically sound, and has led to some reasonable results, thus satisfying deliverables D6 and D7. However, there are a number of factors that would benefit from further research. Many of these issues concern the use of SGS as an estimation algorithm. The work undertaken in this workpackage demonstrated the sensitivity of the simulated TAEP values to the variogram parameters. It was noted above that the TAEP decreased exponentially with increasing variogram range. This helps to explain the positive skewness of the resulting distributions of TAEP values obtained using the Bayesian-geostatistics methodology. However, why the variograms with high sills and long ranges modelled for 1983, 1986 and 1989 caused problems for the SGS algorithm is not yet known.

The algorithm is potentially sensitive to a number of additional parameters which have not been considered here, such as:

- Neighbourhood parameters
- Correlation coefficient with depth covariate (and the uncertainty within that parameter itself)
- The effect of migrating the data to the nearest grid nodes
- Variance reduction factor, added to reduce the overestimation of the co-kriging variance

Additional work on the Bayesian variogram modelling procedure could involve taking into account the number of samples available for each of the surveys, which would perhaps increase the error bars allocated to the surveys with fewer samples such as 1977 and 1992.

A possible extension of the application of SGS within the Bayesian Paradigm could be achieved by the use of Bayesian transformed Gaussian prediction methods (De Oliveira *et al*., 1997). Research would have to be carried out for this methodology to be applied in the kind of situation considered here, and tests should be focused on its suitability for spatio-temporally distributed egg data. However, they could allow for a more complete consideration of the sources of uncertainty present in geostatistics-based egg production estimation methods.

The source code of MCSIM was adapted for the modelling of variograms. The source code of the GSLIB (Deutsch & Journel, 1998) algorithm **sgsim** was modified to create an inclusive program for the Bayesian-geostatistical estimation of TAEP, thus completing deliverable D8.

Comparison of new geostatistical estimators with conventional design based techniques

Overview

The research in this workpackage was first focused on developing a realistic egg density simulator in order to provide a means of comparing the estimators and different survey designs. An ideal simulator would be able to generate multiple realisations from a ‘known’ underlying egg production surface. Using specified survey designs, it would then be possible to extract egg density datasets to be used as input for the various estimators. From a geostatistical point of view, the data sets should have similar variographic properties to those observed in the actual survey data. Furthermore, the density values should also be similarly correlated with the local bathymetry.

The initial work involved assessing the use of a GAM-based simulator, which was adapted from a simulator developed as part of a previous EU project. This enabled an initial comparison of the estimators to be made. However, as will be described below, the simulated data lacked the spatio-temporal auto-correlation that is observed in the survey data, and this meant that it was difficult to properly appraise the performance of the geostatistical estimator.

Following the development of the geostatistical TAEP and variance estimator in the later stages of WP2, it was decided to assess the use of conditional simulation to generate feasible egg production surfaces from which to sample. A number of survey designs were devised, and corresponding data extracted from the simulated egg surfaces.
The methods and results obtained as the research progressed towards the fulfilment of deliverables D9 to D12 are presented in the following pages.

Initial work using a GAM-based egg production simulator

The basis for the simulator which was used in this preliminary study was based on a Generalized Additive Model [GAM (Hastie & Tibshirani, 1990)] developed by researchers at the University of St. Andrews (Augustin *et al*., 1998) as part of an EC-funded project (Study No. 97/0097). This involved fitting a GAM surface to mackerel data obtained in one of the ICES pelagic egg surveys using locally-weighted regression smoothers within the GAM regression framework (Beare & Reid, 2002).
The original simulator works by fitting a GAM surface to mackerel data obtained in one of the ICES pelagic egg surveys using a splining GAM approach. The value given by a point on the surface is assumed to be the true egg density at that point. The measurement error term for each sample location is given by a negative binomial (NB) density function, and these NB errors are assumed to be independent in time and space. The parameters of this density function are estimated from survey data and set to be a function of covariates such as date, distance perpendicular from the 200 m depth contour, distance along the 200 m contour, and bottom depth. Therefore, it is assumed that there is no systematic error at any of the sample locations and that the sample obtained at each location could be expected to be unbiased.
The study involved a comparison of the three TAEP estimators developed to date, namely the Traditional Estimator, the GAM estimator and the Geostatistical Estimator. The various development stages of the geostatistical estimator were described in WP2. While this work was being undertaken, the current geostatistical estimator was based on a 3-d collocated cokriging approach. A brief review of the Traditional and GAM estimators is provided below.

Traditional Estimator

The Traditional Estimator uses a highly stratified design in which ICES squares define spatial data. In the spatial dimension interpolation is done by using the average of adjacent squares, while in the time dimension, a piece-wise linear trend between sampling points is assumed. The main advantage of this type of method is that the properties of the estimator do not depend on the unknown true egg distribution, and it is the assumed randomness of the sampled points within squares which is the basis for drawing inferences about un-sampled parts of the survey. In the current context its main disadvantage is that it involves estimating many parameters, and the more parameters that are estimated, the higher the variance of the resulting egg production estimate.

GAM Estimator

Inferences from the GAM-based TAEP Estimator are model based. Raw egg density measurements are modelled as smooth functions of space and time and TAEP is then estimated by integrating under the fitted curve (Borchers *et al*., 1997). GAMs provide a flexible framework for accommodating a wide range of trends and random fluctuation in egg distributions, although accurate estimation of egg production depends on models fitting well, which is difficult to judge when data are so sparse. Most aspects of the GAM selection process can be automated although there are subjective elements to the process.
The critical steps involved in obtaining an adequate GAM for egg production are:

(1) deciding on the form (ie. loess, spline), dimension (i.e. 1-D, multi-D), and degree (span, degrees of freedom) of smoothing in the GAM;
(2) deciding which covariates are to be used in the GAM;
(3) deciding on an appropriate error distribution.

The simulator

Simulated data were generated using the 1995 data and the "true" TAEP was calculated to be 1.715 x 10^{15} (see Figure 39). The 2001 survey locations were then used to sample from the simulated spatio-temporal dataset and 1000 simulated datasets were created. Noise was added using the negative binomial distribution (see Figures 39 and 40). In Figure 40 all 1000 simulated datasets are plotted against two possible temporal predictors: Week Number and Julian Day. A close inspection of the

sample points shown in Figure 40 reveals that there are spatio-temporal gaps in the sampling campaign that may have resulted in the occurrence of large egg abundances being unsampled.
Figure 41 is another illustration of the simulated data, and was plotted by first identifying arbitrary points along the 200m bathymetric contour (see Figure 40 top left). At these locations, average egg densities were extracted from the "true" dataset (Figure 39) and random noise added. Figure 41 demonstrates the seasonality of "typical" simulated data at arbitrary point locations.
There was some concern over the fact that the negative binomial error for each observation generated from the simulated pelagic egg survey and underlying "true GAM surface" was statistically independent of all other random errors generated at other sampling points, and it was decided to investigate and verify this assumption. The unadjusted residual error terms (i.e., the difference between the observed egg density and the GAM surface egg density) at each sample point from the 1995 pelagic egg survey were analysed for possible spatial and temporal autocorrelation by constructing a 3-D variogram from the residuals.

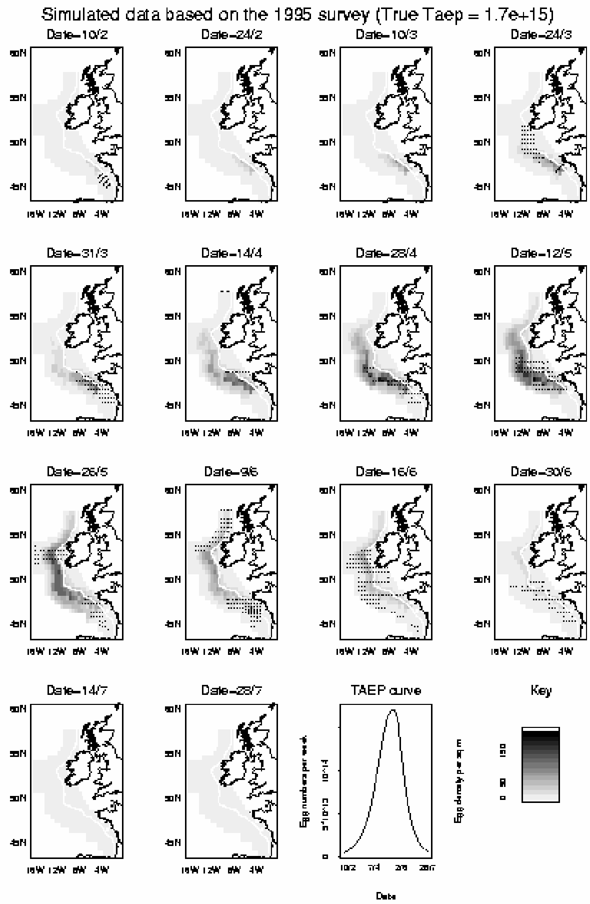
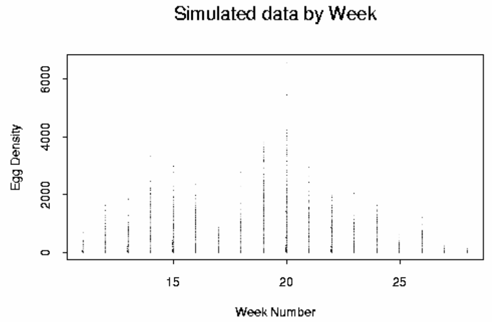


Figure 33. Simulated spatio-temporal egg production surface based on data collected during the 1995 survey.



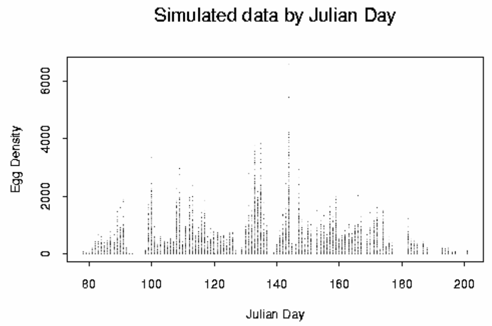


Figure 34. Example of simulated dataset with noise added displayed by week number and Julian day.

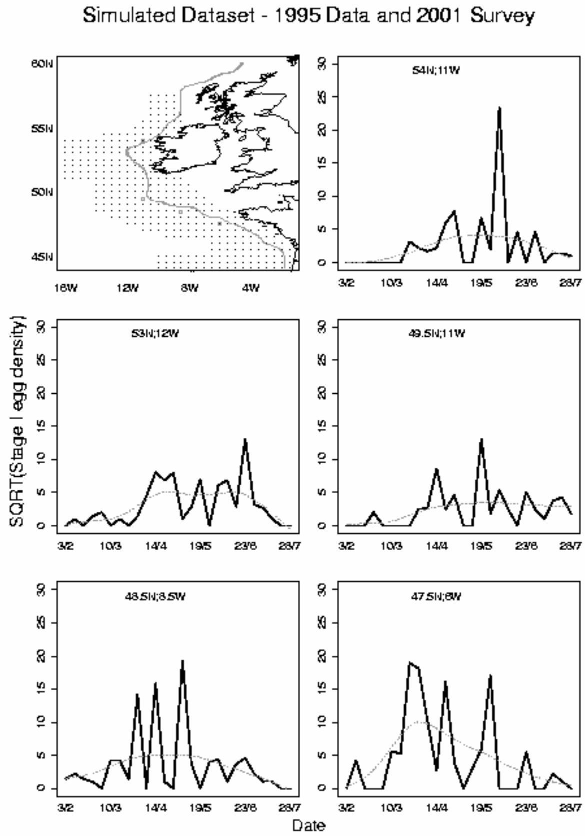


Figure 35. Example of simulated data at intervals along the 200 m bathymetric contour. The dotted line shows the smoothed data to summarise seasonal dependence.

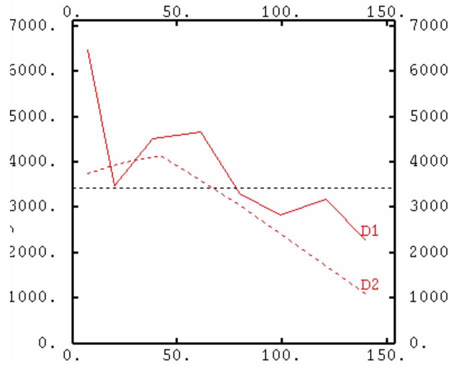


Figure 36. 3-D variogram of residuals arising from negative binomial error distribution.

Figure 42 shows the variograms of the residuals calculated with respect to the omnidirectional spatial plane (solid line) and time direction (dashed line). These variograms suggest that there is no autocorrelation between the residuals in space or time, as required.

The simulated data were then used to construct estimates of TAEP using the Traditional, GAM and Geostatistical Estimators. The principal advantage of this approach is that the “true” TAEP is known, so the biases of the respective estimators can easily be ascertained. It should be noted that the main purpose of the current phase is to harmonise the respective software and identify the number of simulations and scenarios that can be realistically done given current time constraints and computing power. Once this has been done a large range of different “true” datasets and sampling designs can be simulated in order to investigate the performance of the four estimators.

Initial Results

Traditional TAEP Estimator

In the past the Traditional Estimator used an estimate of variance based on a rather *ad hoc* procedure that only uses the positive part of the data. This was considered inadequate for the current project and variances have instead been estimated using bootstrap re-sampling. The Traditional Estimator runs quickly in FORTRAN and 1000 simulations with 1000 bootstrap estimates of variance can be done in about 5 hours. Clearly there is considerable scope for exploring a wide range of scenarios using this estimator.

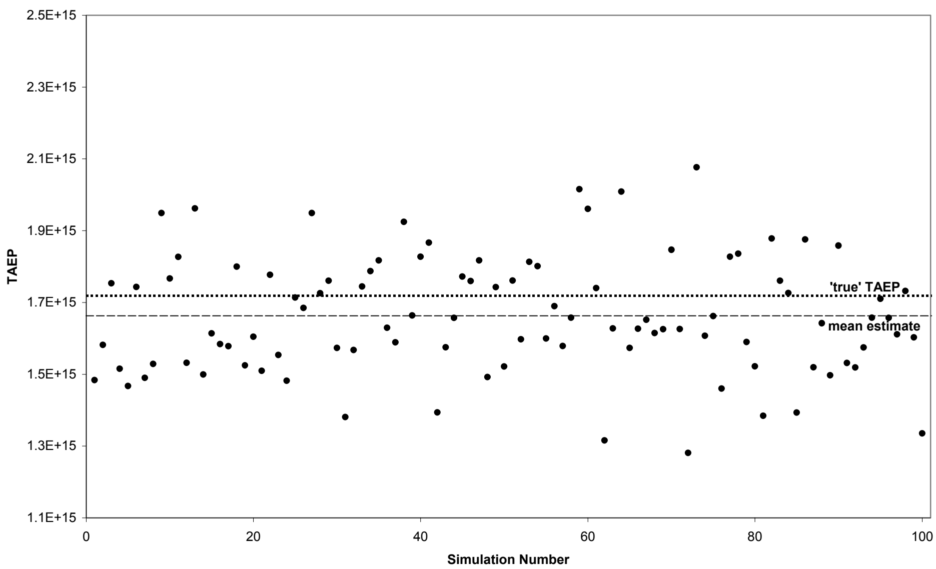


Figure 37. TAEPS calculated by the Traditional Estimator for the first 100 simulations. The dashed line represents the ‘true’ TAEP.

Figure 43 shows the TAEPs calculated by MLA for the first 100 simulations. The average for the first 100 Traditional TAEP estimates was 1.661×10^{15}, giving a small negative bias of –3.13% [i.e. $((1.661 \times 10^{15} - 1.715 \times 10^{15}) / 1.715 \times 10^{15}) \times 100$].

Variances for the estimations were calculated as follows. Firstly, 1000 bootstrapped point estimates are obtained for each of the 100 simulation models. The standard deviation σ and mean m of these point estimates are then calculated. The coefficient of variation (CV) is then obtained in the form of a percentage as follows:

$$\mathrm{CV} = \sqrt{\frac{m^2}{\sigma^2}} \times 100 \qquad (1)$$

To obtain the confidence limits the point estimates are ranked in ascending order and the 2.5% and 97.5% quantiles extracted. As expected, the Traditional Estimator has very high variances (see Figure 44) and average coefficients of variation for 1000 simulations were 22.1% (see Figure 45). As can be seen from Figure 44, the 95% confidence intervals are wide, but appear to encompass the 'true' TAEP for each simulation.

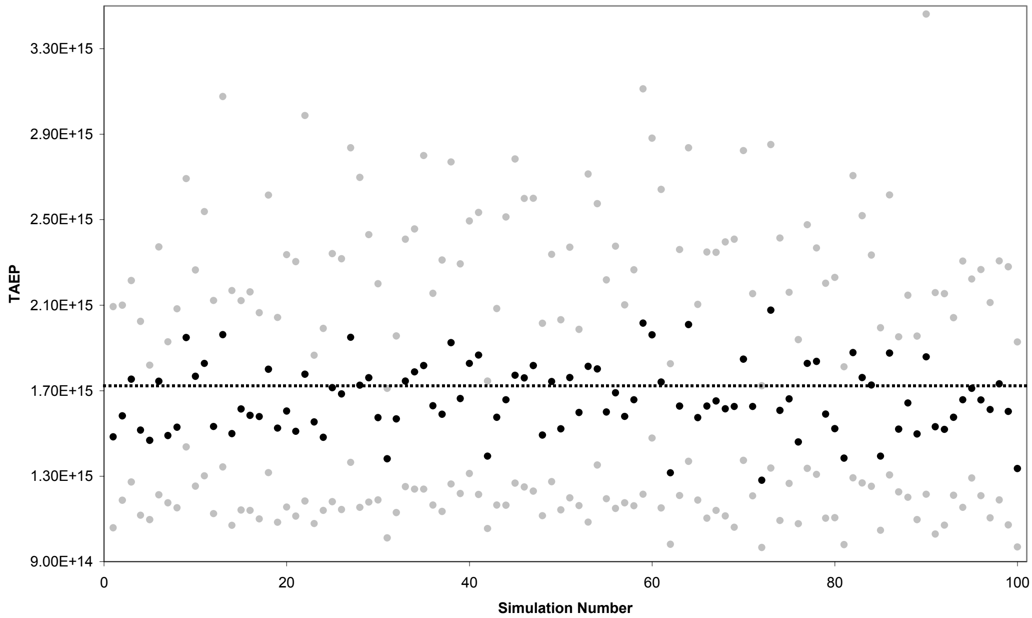


Figure 38. First 100 Traditional estimates of TAEP (black dots) with lower (2.5%) and upper (97.5%) confidence intervals represented by grey dots.

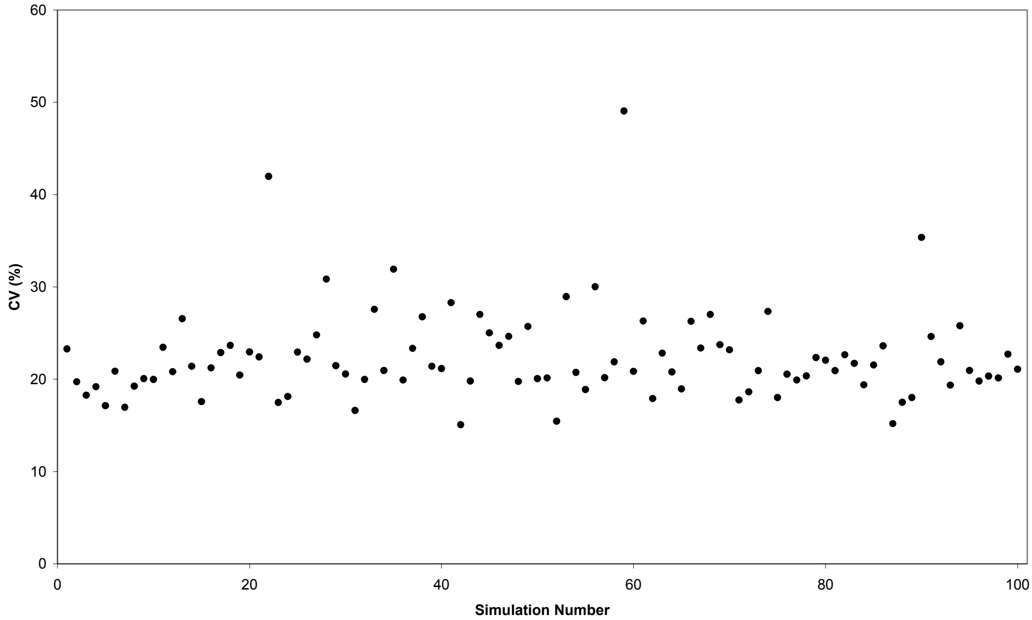


Figure 39. First 100 coefficients of variation for Traditional estimates of TAEP.

The weekly estimates for each of the first 100 TAEPs estimated by the Traditional Estimator are shown in Figure 46. This has been prepared by summing the daily estimates for each week. The graph shows a low variability at the start and end of the spawning season, and a high variability at the spawning peak.

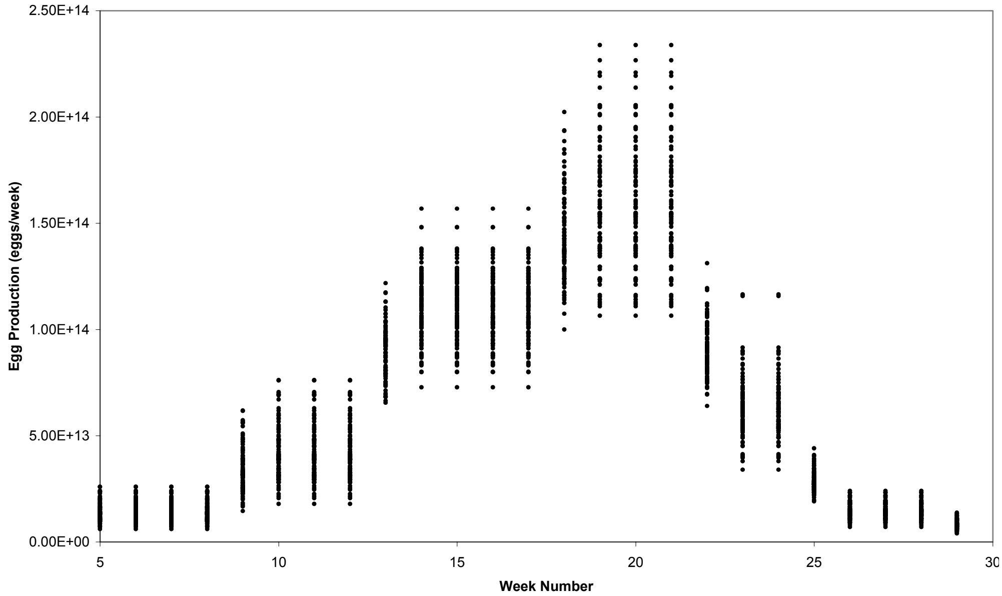


Figure 40. Weekly EP estimates calculated by Traditional Estimator.

GAM TAEP Estimator

The GAM estimation software is written in S-plus, which is very slow, and it currently takes over a week to do 1000 GAM simulations with relevant 1000 bootstrapped variance estimates. The average for the first 100 TAEP estimates by GAM was 1.548 x 10^{15}, indicating a significant bias of -0.715 % [(1.548 x 10^{15} – 1.715 x 10^{15}) / 1.715 x 10^{15}) x 100]. The TAEPS calculated for the first 100 simulations are plotted in Figure 47.

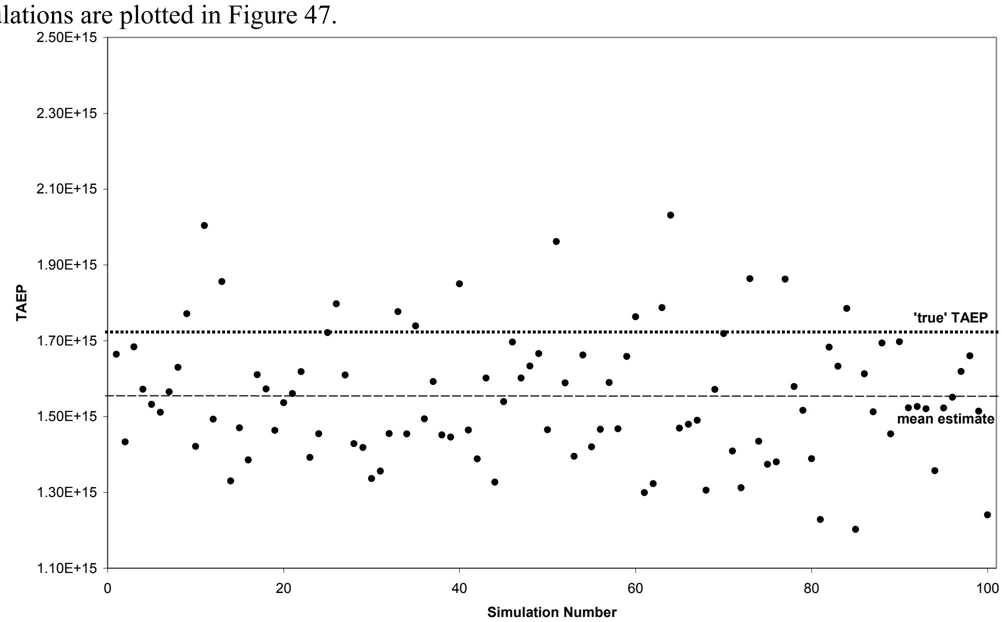


Figure 41. TAEPS calculated by the GAM Estimator for the first 100 simulations. The dashed line represents the 'true' TAEP.

The variances and confidence intervals for the GAM estimates are calculated in the same manner as for the Traditional Estimator described above. The GAM estimator had much less variance than that of the Traditional Estimator (see Figure 45), with an average coefficient of variation of 9.4% (see Figure 49). However, the 95% confidence intervals do not enclose the 'true' TAEP for 59% of the first 100 simulations (Figure 48).

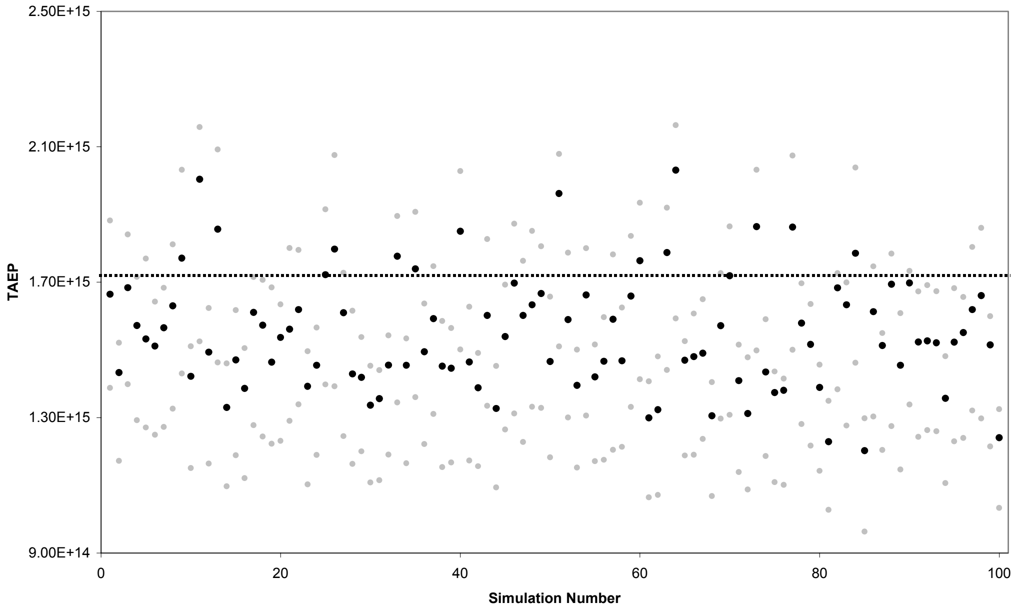


Figure 42. The first 100 GAM estimates of TAEP (black dots) with lower (2.5%) and upper (97.5%) confidence intervals represented by grey dots.

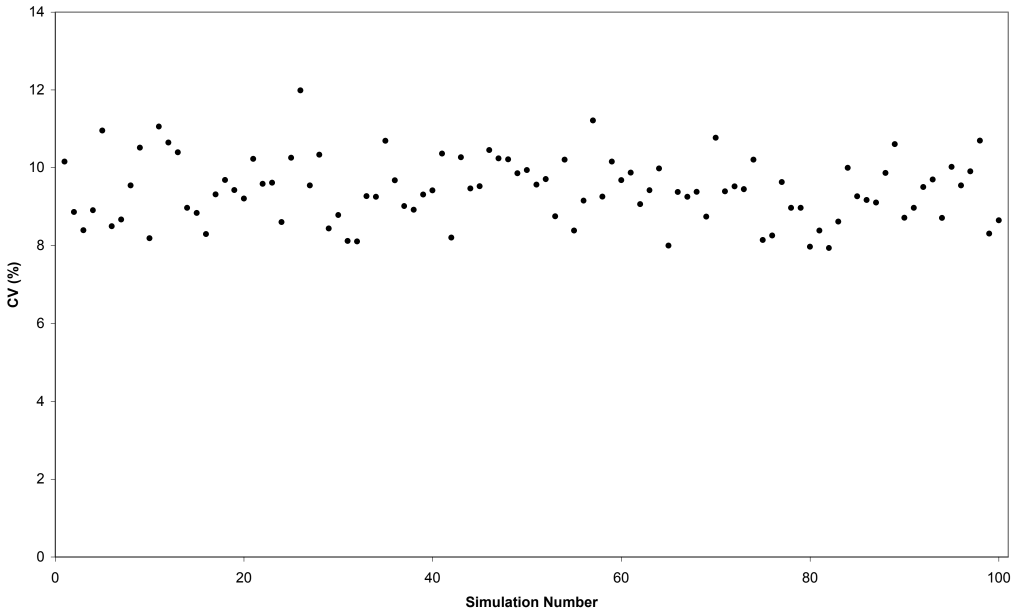


Figure 43. The first 100 coefficients of variation for GAM estimates of TAEP (solid dots).

The weekly estimates for each of the first 100 TAEPs estimated by the GAM Estimator are shown in Figure 50. The graph shows a very low variability at the start and end of the spawning season, and a high variability at the spawning peak around week 20.

It should be noted that the GAM Estimator has calculated its estimates after the addition of structural zeroes. However, as pointed out by MLA, the simulated dataset has been created in such a way that the edges of the simulated data behave better at the edges than raw data. This means that the values tend to come down at the edges in space and time, which is not necessarily true for real data. Therefore, structural zeroes are not in fact necessary for the simulated data, and have caused a significant bias in the GAM estimates. In fact, when the structural zeroes are omitted from the dataset, the mean TAEP is around 1.8×10^{15}, signifying a slight positive bias.

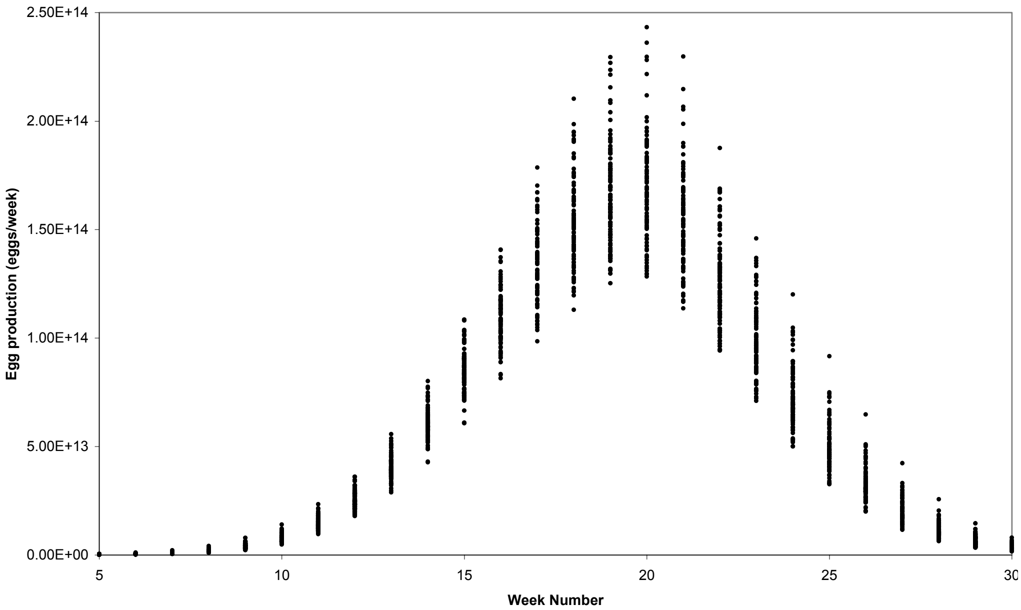


Figure 44. Weekly EP estimates calculated by GAM Estimator.

Geostatistical Estimator

Due to the degree of manual input involved in the geostatistical estimation procedure, it has only been possible to obtain estimates for the first 100 of the simulated datasets so far. It is hoped that it will be possible to automate at least part of the procedure and hence obtain a greater number of estimates for future simulation exercises.

An example of a 3-D variogram (experimental and fitted) calculated from the simulated datasets is provided in Figure 51. The optimal depth parameter has been included as a covariate, and border zeroes were added to each dataset prior to analysis.

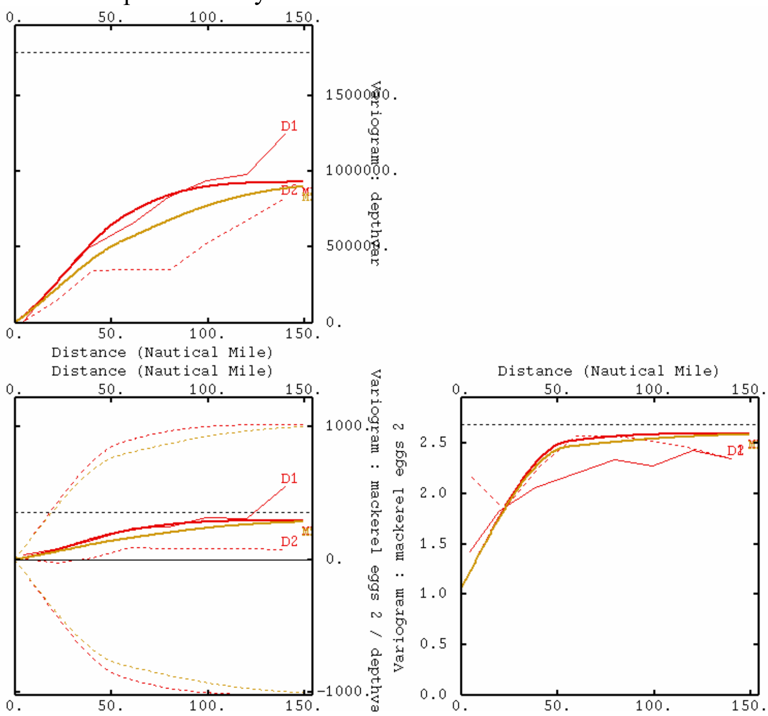


Figure 45. Example of a 3-D variogram obtained from a simulated dataset.

Figure 51 shows that the size of the nugget effect in the 3-D egg density variogram is considerable, particularly when compared to the variograms of the actual log-transformed egg density data. In fact, this was one of the clearer covariance structures identified in each of the 100 simulated datasets modelled. This suggests that the simulated data does not fully capture the spatial and temporal autocorrelation of the egg densities found in nature. For the first 100 simulations, the corrective factor *k* (see WP2) ranged between 2.5 and 3.5, with a mean value of 3.0. This is higher than those values calculated for each of the triennial surveys, as presented in Table 3. Over the first 100 datasets, the mean TAEP estimated using the geostatistical method was 1.808 x 10^{15}, which represents a positive bias of 5.40%. The TAEPs are plotted in Figure 52.

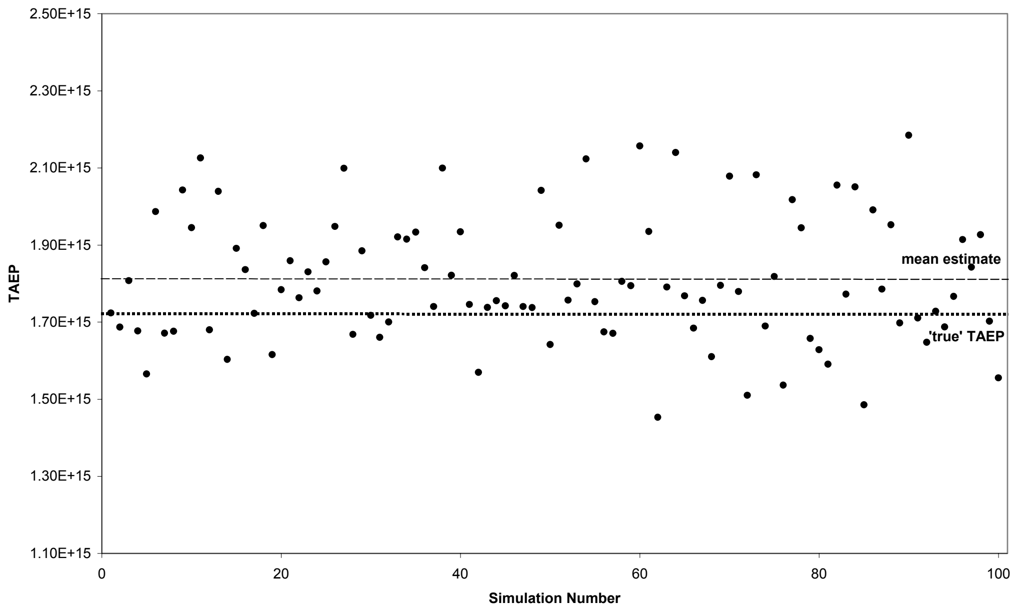


Figure 46. Plot of the first 100 Geostatistical estimates of TAEP from simulated dataset. The dashed line represents the 'true' TAEP of 1.715 x 10^{15}.

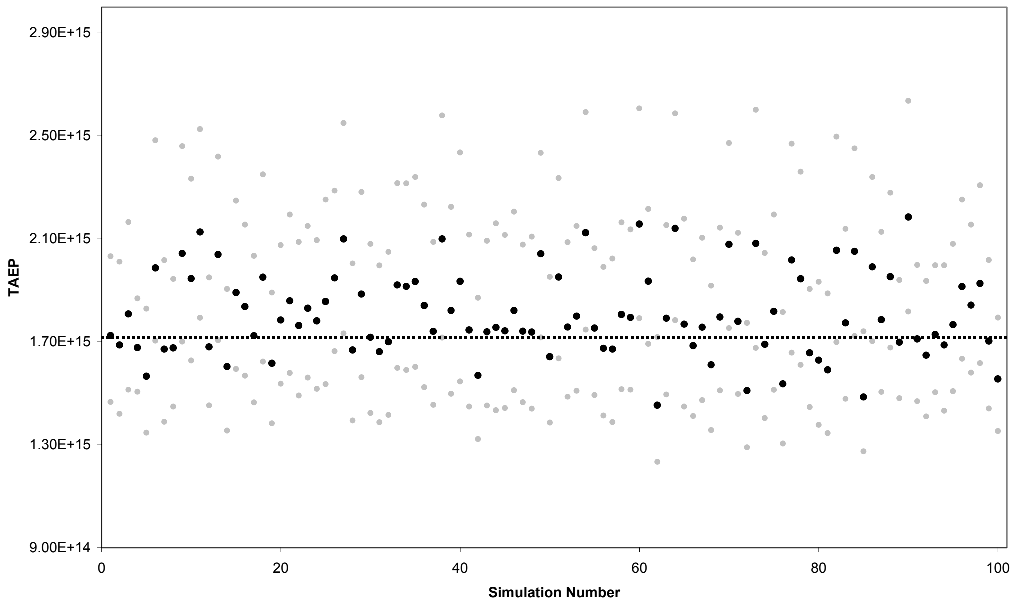


Figure 47. The first 100 Geostatistical estimates of TAEP (black dots) with lower (2.5%) and upper (97.5%) confidence intervals represented by grey dots.

There is a large spread of TAEP estimates around the 'true' figure. The CV of the estimates is 9.13%. Figure 53 shows the TAEP estimates with their upper and lower 95% confidence limits, as calculated from the kriging variance. The error bars enclose the 'true' TAEP in 90 of the first 100 simulated cases. However, if the bias is removed from the estimates, this figure increases to 95.
The CVs of the individual estimates are plotted in Figure 54. The mean value for the CV is 8.91%.

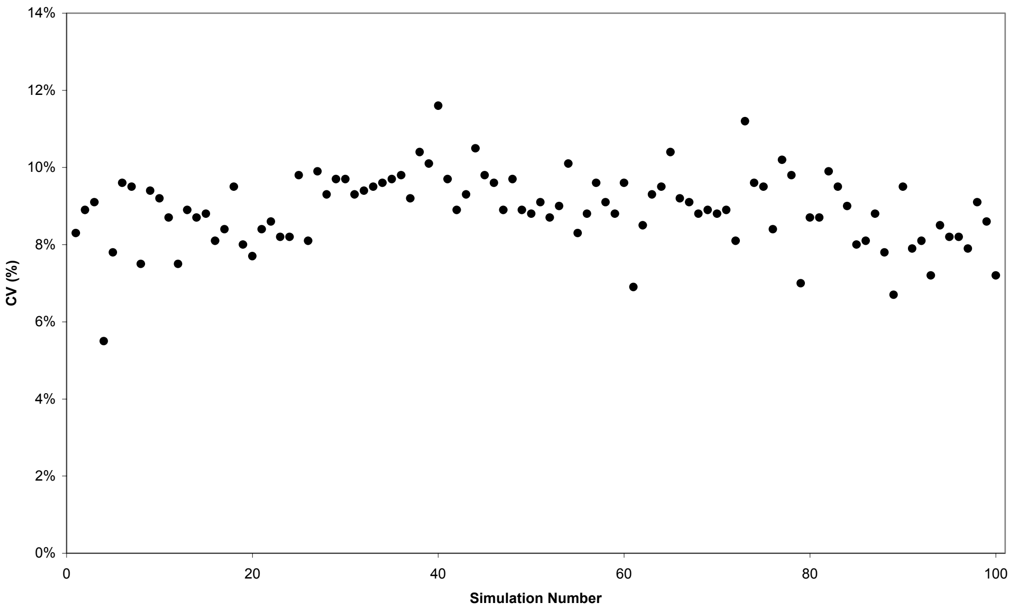


Figure 48. Coefficients of variation for geostatistical estimates of TAEP.

The values of egg production calculated for each week have been plotted in Figure 55. As can be seen, the spread of values is lowest at the start and end of the spawning season, and highest around week 20. There is no sign of bimodality in the weekly estimates.

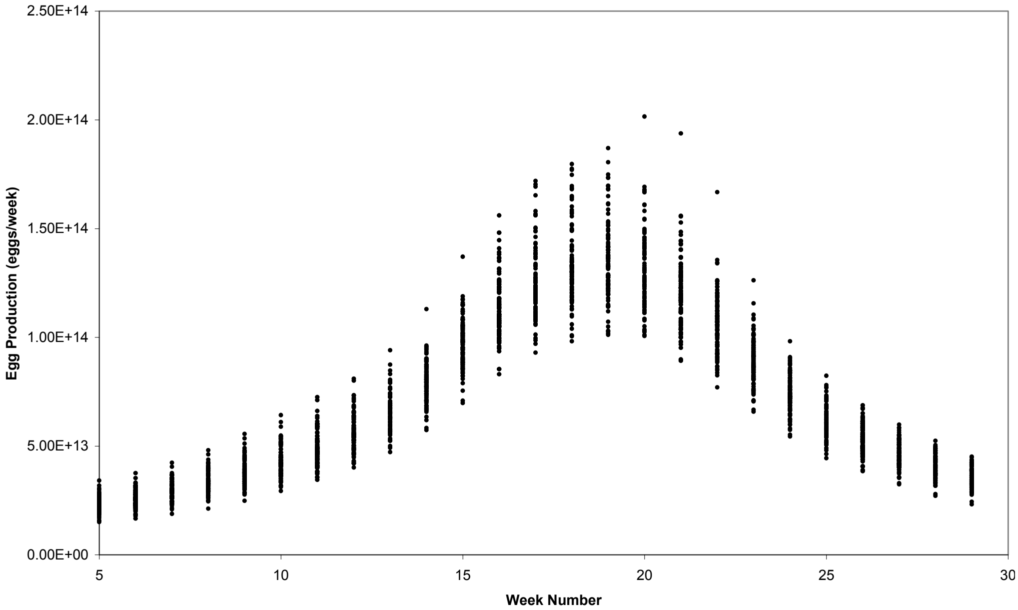


Figure 49. Weekly EP estimates produced by the Geostatistical Estimator for the first 100 simulations.

Summary and Discussion

Table 10 displays basic statistics calculated for each of the three estimators prepared so far over the first 100 simulations.

Table 8. Summary statistics relating to the performance of the three estimators over the first 100 simulations.

	Trad. Estimator	GAM Estimator	Geo. Estimator
Mean TAEP	1.661×10^{15}	1.548×10^{15}	1.808×10^{15}
Max TAEP	2.077×10^{15}	2.031×10^{15}	2.185×10^{15}
Min TAEP	1.281×10^{15}	1.202×10^{15}	1.453×10^{15}
TAEP Standard Deviation	1.635×10^{14}	1.690×10^{14}	1.650×10^{14}
TAEP CV	9.84%	10.92%	9.13%
Bias	-3.13%	-9.72%	+5.40%
Mean Estimate CV	22.10%	9.42%	8.91%
% Enclosed by Error Bars	99%	41%	90%

The TAEP estimates obtained with the Traditional Estimator tend to be less biased than those of the GAM and Geostatistical methods. Furthermore, its estimates are least spread around the 'true' value of 1.715×10^{15}. It is also noted that the inclusion of structural zeroes in the GAM Estimator resulted in the significant negative bias of –9.72%. It is interesting to see that the Geostatistical Estimator, which also included structural zeroes, is positively biased. This is because, while the GAM Estimator uses all of the data points (of which around 25% are structural zeroes) for its estimations, the Geostatistical Estimator bases its estimates on only those values that lie within local neighbourhoods.

The values of CV calculated for the TAEP estimations vary widely between each Estimator. The CVs associated with the Traditional Estimator are large, with the error bars covering a wide range of values. These error bars ensured that the 95% confidence interval associated with each TAEP estimate encompassed the true value 99% of the time. The CVs calculated for the Geostatistical Estimator are similar to those calculated for the GAM Estimator, although the GAM Estimator was less successful at encompassing the 'true' TAEP with its error bars, even when the bias was removed. Although the bias-free Geostatistical confidence intervals enclosed the target value in 95 of the cases, it should be remembered that the proper treatment of kriging variance was still under investigation at that stage.

Table 9. Correlation matrix between the three TAEP Estimators and the mean and maximum values of the first 100 simulated datasets.

	Trad	GAM	Geostat	Mean	Max
Trad	1	0.62	0.82	0.42	0.41
GAM	0.62	1	0.67	0.45	0.15
Geostat	0.82	0.67	1	0.55	0.17
Mean	0.42	0.45	0.55	1	0.00
Max	0.41	0.15	0.17	0.00	1

Finally, it is interesting to consider the correlation between the TAEP estimates of the different Estimators. Table 11 shows the correlation matrix calculated between the TAEP estimations from each estimator, and the mean and maximum values of the first 100 simulated datasets. The geostatistical estimator was most highly correlated with the Traditional Estimator, with a correlation coefficient of 0.82. The GAM Estimator is similarly correlated with both the Traditional (0.62) and Geostatistical (0.67) methods. While the Geostatistical Estimator seems to have the highest dependence on the mean value of the dataset, the Traditional Estimator is most highly correlated with the maximum value. This could suggest that Geostatistical and GAM methods may be more robust with respect to extreme values in the dataset than the Traditional Method, although it is possible that this is due to the inclusion of structural zeroes in the latter methods.

Preliminary Conclusions

The preliminary results provided above suggest that in terms of the TAEP estimations, the Traditional Estimator is the most accurate. The simulated dataset, however, was found to be potentially less suited towards the GAM and Geostatistical methods than the actual survey data for the following reasons:

- There was no need for structural zeroes, although these were included in the GAM estimation procedure for the sake of comparison;
- The spatio-temporal correlation structure found in nature was not recreated in the simulated dataset, rendering it less suitable for geostatistical modelling.

The release of the first set of simulated data, however, did provide a useful means for comparison amongst the different estimation methods. It was suggested, for example, that the Traditional method overestimated the variance, with the target TAEP falling within the 95% confidence intervals 99% of the time. With the bias removed, the target TAEP fell within the 95% confidence intervals of the GAM estimates 90% of the time, suggesting that the uncertainty was underestimated. The CVs for the geostatistical estimator could not be considered as legitimate measures of uncertainty, and so could not be compared in the study.

The above research was presented as a Working Document at the WGMEGS meeting in Dublin, 2002 (Beare *et al*., 2002).

Further work using an SGS-based egg production simulator

In Section 2.2.6.2, a methodology for estimating TAEP with associated variance was presented. This was based on a conditional simulation algorithm called Sequential Gaussian Simulation (SGS). The algorithm proceeds by generating a number of equiprobable egg density surfaces, so that a distribution of possible TAEP values can be obtained. Each surface generated retains the geostatistical properties of the original dataset; that is, the variogram and histogram are reproduced. The correlation present between the modelled variable and a secondary collocated variable is also retained. These features were found to be lacking in the surfaces produced by GAM-based simulator described above, rendering it unsuitable for use with the SGS TAEP estimator.

The advantage of the GAM-based method was, however, that a true underlying surface was available as a target for the estimators to aim for. Nevertheless, it was decided to assume that the SGS algorithm would allow an approximation of the mean TAEP to serve as a 'true' value to aim for. The algorithm was used to generate 500 representations of the 1998 egg production surface on a grid with the following resolution: 5 nautical miles (along a line of latitude) x 15 nautical miles (along a line of longitude) x 7 days. This resolution was thought to be fine enough to adequately test a wide range of possible survey designs, without creating prohibitively large files. The mean TAEP was calculated to be 1.41×10^{15}, with a CV of 8%. An example of one of the simulations is provided in Figure 56.

The 1998 dataset was selected for a number of reasons:

- there was considerable agreement between all the methodologies on the TAEP (Trad 1.37×10^{15}, GAM 1.18×10^{15}, SGS 1.44×10^{15}, OK 1.47×10^{15});
- there was a good coverage over the spawning area and throughout the season;
- the data were well behaved (i.e. there were no high values on the edges of the survey);
- the spatio-temporal autocorrelation structure is well-defined.

However, the 1998 egg production curve is particularly flat, unlike the other years which have well defined spawning peaks or some bimodality (see Figure 28). A more detailed analysis would involve considering spawning seasons which are characterised by early or late spawning peaks, as is often observed in nature.

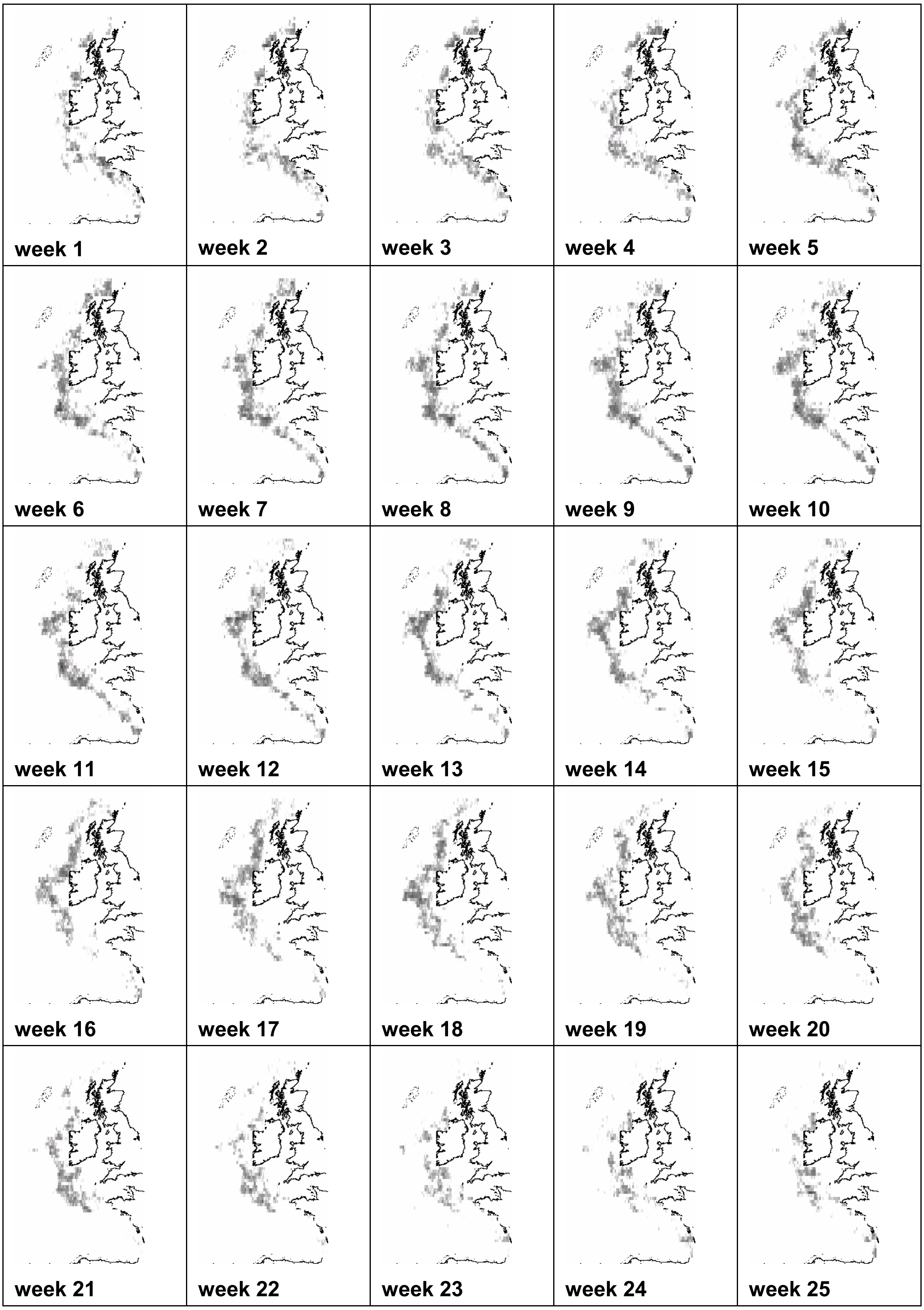


Figure 50. Example of a simulated egg production surface based on the 1998 survey data. The values have been log-transformed prior to plotting.

The objectives of the study were twofold: firstly to compare the performances of the different estimators; and secondly to compare the merits of different survey designs. The Bayesian-Geostatistical

estimator, as described in 2.3.2.2 was not used in the study due to the prohibitively long processing times involved. However, inferences can be made as to how the estimator would have performed.
The datasets were provided to MLA, who used S-plus to analyse the data and prepare a number of survey designs. The first set of designs involved an imaginary random representative (RR) sampling campaign. While this would be impossible to achieve in real life, such a campaign would be theoretically ideal for the implementation of the Traditional and GAM TAEP estimators.
The 277500 grid node values in each simulation were listed in a single column (the standard output of GSLIB [Deutsch & Journel, 1998], with the coordinates increasing first along x, then y and then z). It was therefore possible to generate a random vector of integers between 1 and 277500 and extract the grid node values corresponding to those positions in the column. Since the grid was a simple cuboid shape, this meant that many points fell outside the designated spawning area. This problem was surmounted by specifying 4 times the number of data points required. Those which fell outside the spawning area, which occupied approximately a quarter of the volume of the grid, were subsequently removed. A number of RR datasets were thus generated with varying numbers of data points: RR500, RR1000, RR2000, RR4000, RR8000, which contained approximately 500, 1000, 2000, 4000 and 8000 samples respectively.
The datasets were then tested to make sure that they had similar variographic properties to the original 1998 data. The experimental and modelled variograms for the original data and an example from RR8000 are provided for comparison in Figure 57.

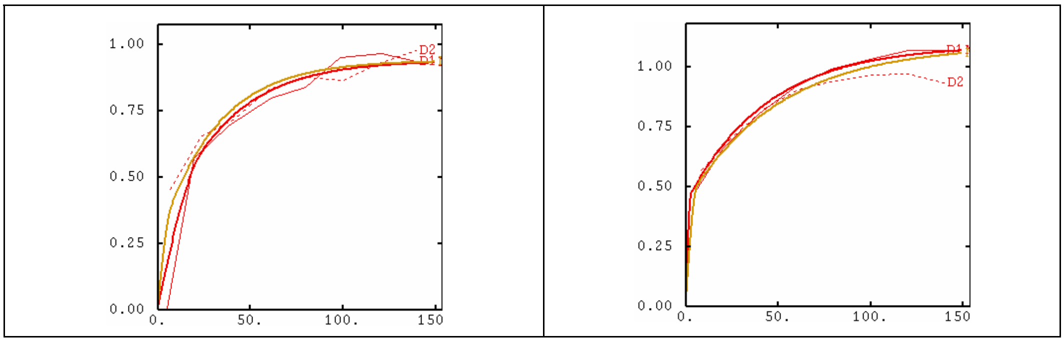


Figure 51. Original 1998 variogram using normal scores transformed data (left) and variogram obtained using RR8000 data. D1 is the omnidirectional spatial variogram while D2 is the time variogram.

A comparison of the variograms in Figure 57 reveals that at a first glance, the original and simulated datasets have similar spatio-temporal autocorrelation structures. However, there are some minor differences. In particular, the sill for the simulated data set is slightly higher than that for the original dataset, and the very short-range structure now appears more like a nugget effect.
Examples of the random representative surveys from RR500 and RR1000 are shown in Figure 58. These plots also include the artificial zeroes that are used in the geostatistical estimation procedure.
Besides the random representative datasets, MLA prepared a number of more realistic surveys based on the surveys previously undertaken in 1992, 1995, 1998 and 2001. For each year, three vectors were generated: one with half the number of samples (H1992, H1995, H1998, H2001); one with the same number of samples (O1992, O1995, O1998, O2001); and another with twice the amount of samples (T1992, T1995, T1998, T2001). The vectors were generated by isolating the areas and times at which parts of the spawning area were sampling, and randomly selecting a specified number of points within those windows. The survey designs were therefore similar in terms of spatio-temporal coverage to the original surveys, but not exactly the same.

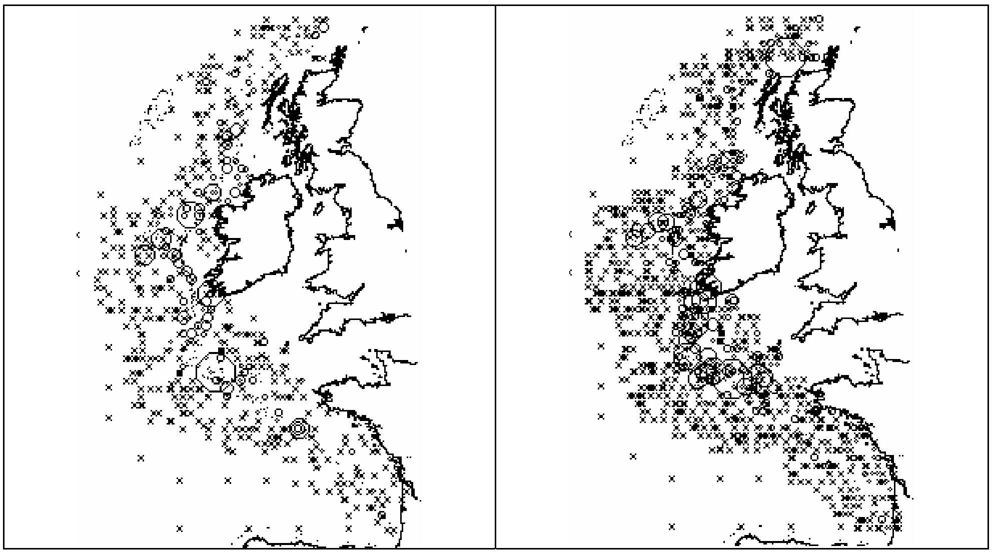


Figure 52. Examples of the simulated surveys RR500 (left) and RR1000 (right). The circles are sized in proportion to egg density. x denotes zero egg density.

Examples of the O1992, O1995, O1998 and O2001 surveys are provided in APPENDIX II. Their characteristics are briefly described below:

- **1992**: No samples collected until week 9 (mid-April), and only in the south. The vast majority of samples were collected in weeks 16 and 17. This survey may be characterised as one with a poor temporal coverage.
- **1995**: Samples first collected in week 7 and every week sampled from then until week 24. Good spatio-temporal coverage.
- **1998**: Early sampling, beginning in week 6. No sampling in week 10. Better spatio-temporal coverage in the second half of the spawning season.
- **2001**: first samples collected in north in week 7. Good spatio-temporal coverage throughout, but slightly better in the first half of the spawning season.

Using each of the 18 survey designs, the Traditional, GAM and Geostatistical (SGS) methods were used to estimate the TAEP and associated uncertainty for as many of the simulated surfaces as possible. The Traditional method is quick to implement, and so it was possible to cover all 500 surfaces. However, the GAM and SGS methods involve many more calculations and it was only feasible to cover the first 250 surfaces for the GAM estimator and 100 for the SGS method. The CVs for each of the Traditional and GAM TAEP estimates were obtained by bootstrapping.

The geostatistical simulation was done with the same parameters as those used to generate the data. This assumed that the original variogram model was valid for use with the simulated data.

The mean TAEPs, mean CVs and overall biases are listed in Table 12 for each of the estimators and each of the survey designs. Figure 59 plots the percentage bias observed for each of the estimators. An inspection of the results shows that for the Traditional and SGS methods, there is always a negative bias. This bias reduces in general when the number of samples is increased (see Figure 60), and this effect is most evident for the Traditional method. In general, the SGS method performs better than the Traditional method when there are fewer than 1000 samples. The only survey year for which the SGS method performs better than the Traditional method despite the number of samples is 1992. This may be due to the fact that the survey campaign in that year was delayed, and the geostatistical method is more suited towards extrapolation.

However, it is not known why the negative bias persists throughout all the survey strategies. There are a number of possibilities, such as:

- The simulated surfaces do not adequately represent 'realistic' egg production surfaces (known problems associated with the SGS algorithm include discontinuity of high values);
- In the case of the SGS estimates, the decision to retain the variogram model parameters may have been detrimental;
- The survey designs, even those related to the actual sampling campaigns, are relatively random and may contain less samples from the known high density areas than would have been collected in practice.

In contrast, the mean GAM TAEPs are positively biased for many of the 'realistic' survey designs. A look at the histogram of the individual TAEPs for each survey shows that there are a number of extreme estimates which are contributing towards the overall positive bias and large CVs. It is not known why the GAM estimator is displaying this instability. Although there was no need to add structural zeroes due to the well-behaved nature of the data, MLA found that adding them did not prevent the occurrence of the high TAEPs. One suggestion was that the GAMs were overestimating the seasonality of the spawning season, as 1998 is relatively flat.

Table 10. Summary of results obtained using the three estimators.

		TRAD			GAM			SGS		
Survey	No. Data	TAEP ($x10^{15}$)	Mean CV (%)	BIAS (%)	TAEP ($x10^{15}$)	Mean CV (%)	BIAS (%)	TAEP ($x10^{15}$)	Mean CV (%)	BIAS (%)
RR500	512	0.52	20.2	-63.4	1.29	48.1	-9.5	0.55	14.0	-61.4
RR1000	1017	1.06	16.8	-25.4	1.22	6.6	-14.3	0.71	10.3	-50.3
RR2000	2066	1.22	11.0	-14.4	1.23	4.9	-13.6	0.89	8.1	-37.2
RR4000	4149	1.33	7.7	-6.4	1.25	3.27	-12.1	1.11	5.5	-22.0
RR8000	8258	1.30	4.6	-8.2	1.22	2.27	-14.8	1.27	3.7	-10.6
H1992	581	0.67	17.03	-52.7	1.64	63.6	+15.2	1.00	13.6	-29.8
O1992	1173	0.93	14.4	-34.3	1.97	45.4	+38.9	1.09	11.7	-23.2
T1992	2346	1.02	10.2	-28.4	1.65	10.2	+15.9	1.04	10.4	-27.0
H1995	773	1.05	19.5	-25.8	1.33	12.1	-12.7	1.07	11.9	-24.8
O1995	1559	1.26	12.3	-11.4	1.44	10.4	+1.1	1.07	10.4	-24.8
T1995	3118	1.36	8.34	-4.0	1.32	9.3	-7.1	1.22	9.4	-14.2
H1998	536	0.74	17.1	-47.6	1.33	17.6	-6.2	0.95	13.0	-33.1
O1998	1080	1.21	14.1	-14.4	1.63	12.0	+14.5	1.07	10.6	-24.3
T1998	2160	1.32	10.5	-7.1	1.49	5.5	+4.63	1.13	8.9	-20.4
H2001	701	1.03	14.6	-27.4	2.89	15.8	+100.01	1.14	10.0	-19.7
O2001	1409	1.24	11.4	-12.8	2.51	23.1	+76.3	1.15	8.4	-19.3
T2001	2818	1.36	9.2	-4.0	2.48	8.6	+74.5	1.21	6.9	-15.0

Figures 60 and 61 show how the bias changes with number of samples for the Traditional and SGS methods respectively. Figure 60 shows how the bias drops dramatically between 500 and 1000 for the Traditional method, and less significantly beyond 1000 samples. The biases are most significant for the 1992 survey as noted above. The bias observed for the SGS method in Figure 61 indicates a more gradual reduction with increase of samples, and is greatest for the random representative survey designs. For most of the ‘realistic’ surveys, increasing the number of samples does not significantly reduce the bias. In 1992, increasing the number of samples to twice the amount of the original survey leads to an increase in bias. This may have been caused by the large concentration of samples collected towards the south of the spawning area coinciding with a low occurrence of eggs. Although it is difficult to make generalisations regarding the results, it would appear that the finer-scale surveys that are actually undertaken, as opposed to the random representative surveys, tend to suit the Geostatistical method significantly more than the Traditional method.

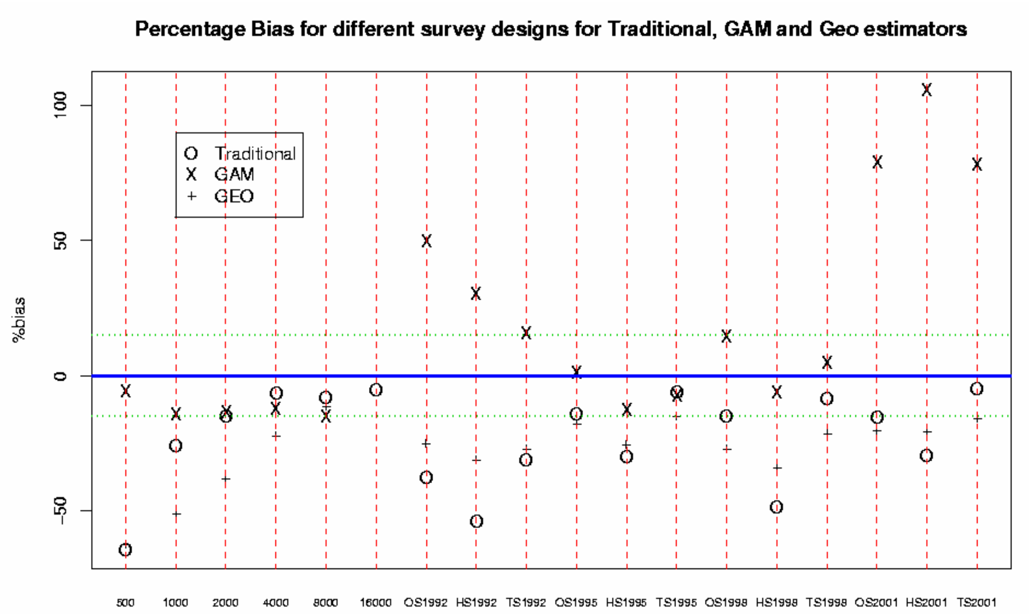


Figure 53. Bias in the Traditional, GAM and SGS estimates for each of the survey designs.

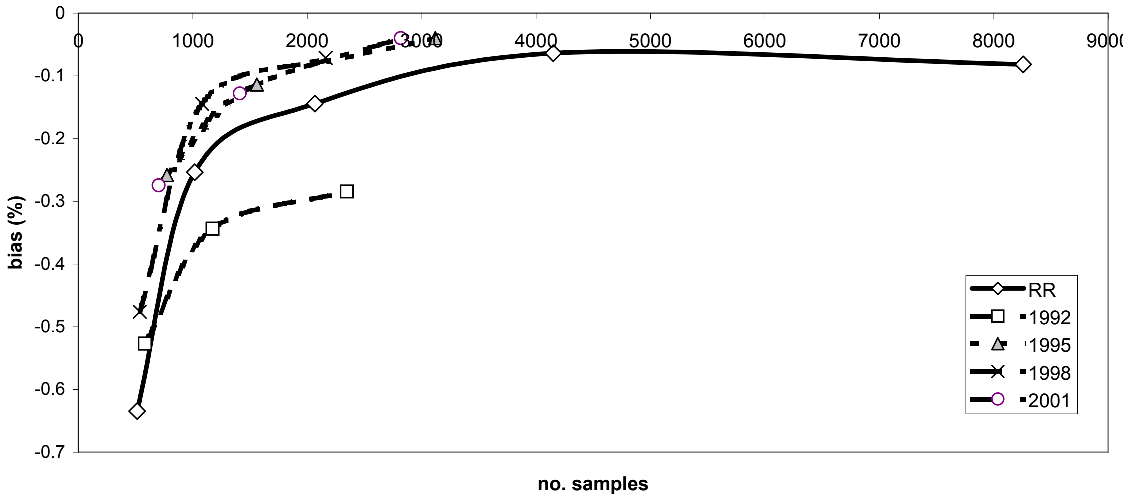


Figure 54. Change in bias with number of samples for the Traditional Estimator for each subset of surveys.

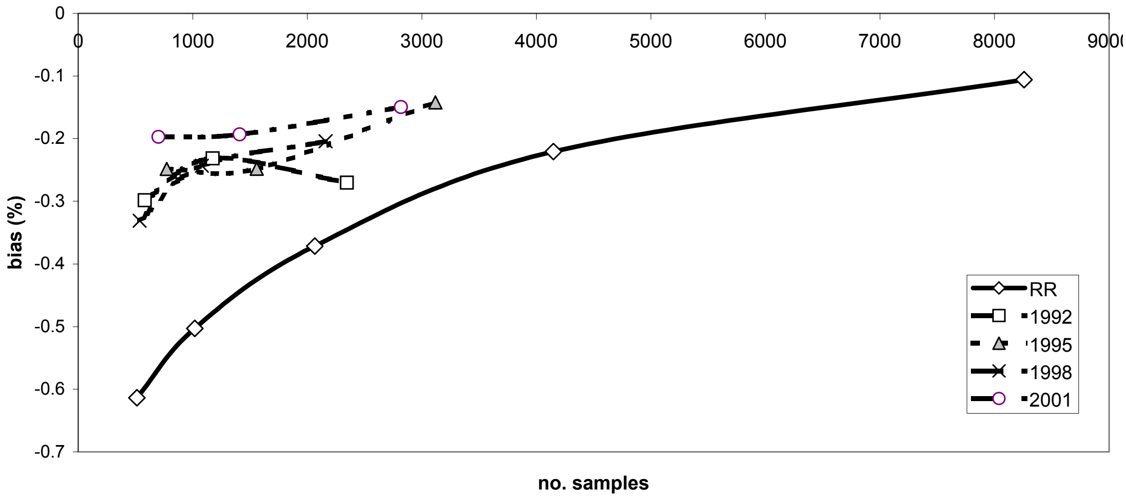


Figure 55. Change in bias with number of samples for the SGS Estimator for each subset of surveys.

Table 12 also lists the mean CVs associated with the TAEP estimates obtained using each estimator with each different survey design. The CVs are plotted against the number of samples for each subset of survey designs in Figures 62 and 63 for the Traditional and SGS methods respectively. As expected, the mean CV reduces logarithmically with the number of samples for both estimators. The lowest mean CVs are obtained with the RR8000 survey design at 3.7% and 4.6% for the SGS and Traditional estimators respectively. In each case except T1995, the mean CV of SGS is lower than that of the Traditional method. However, it should be noted that the CVs presented for the SGS method neglect any uncertainty in the modelling parameters. The use of the Bayesian-Geostatistical estimator developed in WP3 would probably increase the CVs by between 2 and 10%, with the additional variance likely to reduce with the number of samples.

Figures 62 and 63 indicate that the two different methods behave differently when presented with different survey designs. For example, the curves of reduction in mean CV observed for the Traditional method tend to converge around 10% when there are 2000 samples, regardless of the survey design. This convergence is not observed with the SGS method, whereby the curves tend to decrease in parallel. Up until 2500 samples, the 2001 survey appears to have the most suitable design. Beyond 2500 samples, the random representative survey offers more precision. This would suggest that when there are 2500 samples, the sampling density is more able to describe the egg production surface than the existing survey types. It is therefore possible that the precision in the estimation of TAEP using the SGS method could be improved by increasing the spatio-temporal extent of the survey coverage instead of increasing the sampling resolution. For the implementation of the Traditional method, increasing the resolution of the sampling campaign would appear to have a similar benefit to increasing the spatio-temporal coverage.

However, there are some important points to consider before placing any significance on these results. In this study, we have been limited to one set of simulated egg production surfaces, which are based on the original 1998 survey data. The study would not be complete without assessing the survey designs on surfaces with different characteristics, such as those with early or late spawning peaks. Furthermore, the nature of the survey designs tested has been limited to two types: completely random; and confined to the area and time windows of the original surveys. A more comprehensive study would explore realistic designs which were able to traverse those boundaries.

The current results suggest that, of the realistic surveys, the 2001 sampling campaign is best in terms of both accuracy and precision for both Traditional and SGS methods. It may also be noted that reducing the number of samples collected by 50% led to an increase in CV of only 1.6% for the geostatistical estimator, suggesting that good results could be achieved with a significantly reduced effort. However, the 2001 sampling campaign may have been more suited towards the 1998 spawning season due to the increased effort during the earlier months, capturing the slightly early peak detectable in Figure 28.

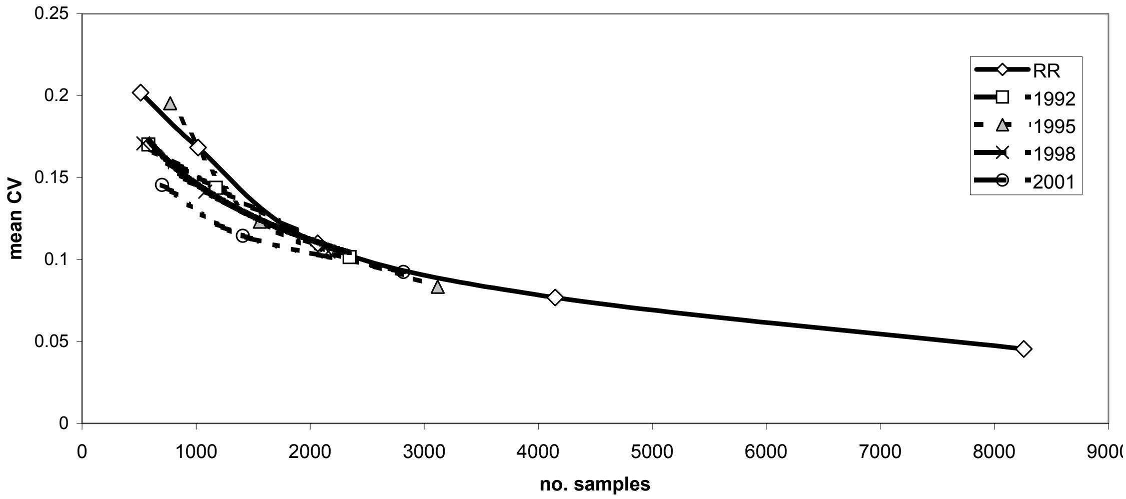


Figure 56. Change in CV with number of samples for the Traditional Estimator for each subset of surveys.

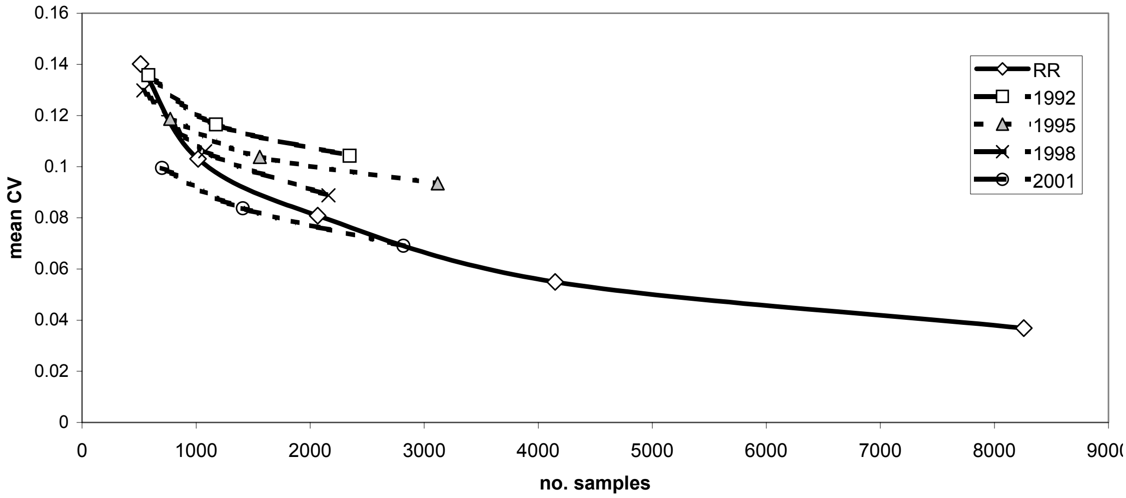


Figure 57. Change in CV with number of samples for the Geostatistical Estimator for each subset of surveys.

<u>*Summary*</u>

The research undertaken in this workpackage culminated in the achievement of the deliverables D9 to D12, as will be described in the following paragraphs.

A simulator of pelagic egg density observations was developed using conditional simulation (SGS), satisfying D9. This technique allows the generation of many equiprobable egg production surfaces based on a set of survey data. While there is no 'true' underlying surface, the data generated retain the geostatistical and statistical properties of the original dataset. However, application of the TAEP estimators to a number of different survey designs revealed a prevailing negative bias. This is a novel approach to data generation with considerable potential, and further research should be done to improve it.

An assessment of each estimator's performance was made across a number of different sampling scenarios. Two studies were made, initially using a GAM-based estimator, and then using an SGS-based simulator. The first study involved generating a number of simulated surfaces based on the 1995 data and sampled using the 2001 survey campaign. In terms of bias among the estimators, the Traditional estimator performed best (-3.1%), followed by the kriging-based Geostatistical estimator (+5.4%). The GAM estimator was most significantly biased (-9.7%), probably due to the additional of structural zeroes. In terms of variance, the Traditional method (calculated using a bootstrapping technique rather than the ICES method) was found to overestimate the size of the confidence intervals, whereas the GAM slightly underestimated the uncertainty. The kriging-based geostatistical method was not able to produce valid estimates of variance.

The SGS method, which replaced the kriging-based Geostatistical estimator in Section 2.2.6.2, was found to work poorly with the GAM-based simulated datasets, as these did not reproduce the spatio-temporal autocorrelation in the egg surfaces, nor did they possess the same correlation with depth. For this reason, an SGS-based simulator was developed and used to generate surfaces with which to compare the simulators over a number of different survey designs. This then formed part of an improved methodology for testing the relative estimation performance of different estimators and diagnosing the causes of estimation failure for pelagic egg estimators of spawners abundance (D11). The methodology is summarised below:

1. A large number of equiprobable egg production surfaces are generated using SGS, based on actual survey data. The surfaces retain the variographic and environmental (depth-related) features of the original data. The mean TAEP calculated over the surfaces becomes the 'target' TAEP. Different scenarios can be simulated using years displaying contrasting characteristics.
2. The surfaces are then sampled using a number of survey designs. This is done using vectors which select the samples from the column of grid values generated by SGS. Survey designs mimicking those actually undertaken can be extracted very quickly by isolating survey snapshots in space and time (that a ship would be able to cover in a 3-week outing, for example) and sampling randomly within them.

3. The different estimators calculate the TAEP and associated uncertainty for as many of the simulated surfaces as is reasonable. The statistics are then calculated, compiled and compared.

Deliverables D10 and D12 involved assessing and evaluating the individual estimators. An initial assessment was made above, using the GAM-based simulator. A further evaluation was made using the improved simulator.

The GAM method was found to be somewhat unstable, for unknown reasons, and detailed assessment of its performance was therefore not possible. In terms of bias, both the Traditional and SGS estimators were persistently negatively biased. This may have been due to inadequacies in the simulated surfaces, or by using relatively less samples from high-density areas. As expected, the bias decreased with the number of samples, with the curve tending to be steeper for the Traditional method. In general, the finer scale pseudo-regular surveys (similar to those actually undertaken) seemed to work better for the SGS method than for the Traditional method, which preferred the random representative surveys. The bias was less significant for the SGS method in comparison with the Traditional method when the number of samples was less than 1000.

In terms of CV, the uncertainty decreased logarithmically with the number of samples for both the Traditional and SGS methods. The CVs were more sensitive to the survey design for the SGS method, whereas beyond 2000 samples, the differences in CV for the Traditional method were negligible. For the SGS method, the 2001 survey-based TAEPs were generally the most precise, although beyond 2500 samples, the random representative survey performed better. This suggests that the overall spatio-temporal coverage of the random sampling strategy improves on the transect-based surveys when the sampling resolution is sufficiently fine. A tentative conclusion would therefore be that future surveys should concentrate on increasing the spatio-temporal coverage, rather than increasing sampling resolution.

In almost all cases the CV was less for the SGS estimator than for the Traditional method. However, it was noted that the analysis was done without considering the variance in the modelling parameters, as the use of the Bayesian-Geostatistical estimator developed in WP3 would have been prohibitively time-consuming. Considering the uncertainty in the model parameters would have potentially increased the CV of the Geostatistical estimates by between 2 and 10%, which would render it greater than the variance in the Traditional estimates on most occasions. However, it is not known how reliable the Traditional variance estimates are in terms of reflecting the true uncertainty. The method of variance calculation used by ICES was rejected in the current study for various reasons, such as its neglect of zero values. The Bayesian-geostatistical method would appear to provide the most theoretically reasonable measures of uncertainty, although the method in general requires more research into its optimal implementation.

Results and Discussion

The previous chapter on methods and materials was written in such a way that it included the results that were obtained throughout the research. This allowed the development of the algorithms to be described, justified and presented in a more logical fashion. This chapter will provide a summary and discussion of the main results achieved.

GIS Database

In order to facilitate a means of compiling and sharing data, and for presenting the graphical results, Imperial College first devoted its efforts towards developing a GIS database. This initially involved collecting and compiling the ICES triennial egg survey data. The collective dataset revealed a number of inconsistencies which were subsequently amended by MLA. Two researchers from Imperial College joined the research vessel ‘Scotia’ during the 2001 triennial egg survey to learn more about the egg collection and identification and therefore gain more experience with regard to potential error sources. The database was supplemented with data pertaining to temperature profiles and the bathymetry of the Northeast Atlantic. The GIS system was used to enable compatibility within two coordinate systems: based on longitude and latitude, as used by the Traditional TAEP estimation method; and in nautical miles as necessitated by the Geostatistical estimator.

In addition, a standard graphical framework was developed. This was used to display the output from computer estimations and simulations, in coordinates in terms of both degrees and nautical miles. Examples of these outputs were provided throughout this report.

Development of the geostatistical estimator

Section 2.2 described the progress made towards developing a methodology for the estimation of TAEP for mackerel and horse mackerel, with corresponding measures of estimation uncertainty.
An initial geostatistical analysis and literature review served to highlight some of the potential difficulties and solutions related to geostatistical modelling. In particular, it was found that:

- The egg density data are very highly positively skewed, with a large proportion of zero values.
- The coordinate system should be based on nautical miles rather than degrees longitude and latitude.
- Spatial autocorrelation in the egg density was observed, thus confirming that geostatistical analysis and estimation was appropriate.
- No anisotropy was observed in the spatial variograms. This was attributed to the fact that the mackerel tend to spawn in the region of the shelf break, the orientation of which changes considerably over the spawning area. Omnidirectional spatial variograms were therefore adopted for further work.
- Previous researchers had observed that sea surface temperature did not appear to be a useful covariate.

An initial geostatistical modelling procedure was then developed. This was based on the two-dimensional approach, and estimates of egg production per data period were obtained. The use of bathymetric information for improving the accuracy of the estimates was assessed by comparing the results of a cross-validation procedure. The following results and observations were obtained:

- Variograms were easier to model after the raw egg data were log-transformed. This served to reduce the skewness of the data, and the spatial autocorrelation structure became more evident as a result.
- An analysis of the mean horizontal location of eggs with respect to bottom depth revealed that the mean depth generally begins at around the 200 m contour at the start of the spawning season. However, as the season progresses, the eggs tend to be spawned further out into deeper waters. After the start of June (coincident with the formation of the thermocline), the mean depth returns to around 200 m.
- A new depth-related covariate was calculated over the spawning area, measuring the difference between the actual bottom depth and the mean depth. The correlation coefficients between the mean depth and egg densities were found to be significant.
- The cross-validation procedure showed that using a depth-related variable as a covariate reduced the mean standardised error for the vast majority of data periods, suggesting better estimation accuracy.
- A subsequent analysis comparing kriging variances with co-kriging variances demonstrated that the depth covariate could also be successfully used to improve the precision of the estimates. This effect was more pronounced for mackerel than for horse mackerel.
- However, it was noted that the temporal resolution in the data was insufficient for a two-dimensional geostatistical modelling procedure that aimed to take into account all sources of uncertainty.
- Furthermore, the current incomplete spatio-temporal data coverage led to two problems: extrapolation leading to extreme values of egg density at the edges; and large gaps where no estimation can be done.

A three-dimensional co-kriging methodology was subsequently developed so that egg production surfaces could be estimated on a finer and more appropriate temporal resolution, while being conditioned on bathymetric data. This part of the development procedure also involved the introduction of a set of border zeroes which would help to reduce the problems mentioned in the final bullet point above. These zeroes were minimal, and were placed just outside the limits of the designated spawning area.
A 3-d grid was defined, with a resolution of 15 x 15 nautical miles x 7 days. Although this resolution is considerably finer than that used with the traditional method, its use was supported by the spatio-temporal resolution of the data. It was therefore possible to obtain weekly egg production estimates in addition to TAEP. Variograms were modelled for the spatial plane and temporal direction simultaneously. A kriging neighbourhood was specified that would be large enough to allow complete spatio-temporal coverage.
Co-kriging was undertaken using the log-transformed egg density values. The subsequent backtransformation to the kriging median (which is more robust than the mean) was negatively biased. A method for obtaining a correction factor to remove this bias was found by comparing the backtransformed cross validation estimates with the original data. The kriging estimates for each cell

were then raised to units in nautical miles and weeks, and summed over the grid to give the TAEP. The values of TAEP were in general comparable with those calculated using the Traditional and GAM methods.

However, while this methodology produced reasonable estimates of TAEP, the problem of calculating the global estimation variance remained. The individual kriging variances for each cell were not independent and therefore could not simply be summed. An alternative method involved using a procedure called 'combination of error terms' was assessed. This involved calculating the contributions of error variance from first estimating the value of the variable in one direction, then raising this to two directions, and finally raising it to three dimensions. The size of the contributions depends on the sampling frequency and the variogram model. However, the method was found to be very difficult to implement, due to the spatio-temporal complexity of the egg sampling campaigns and also the need to read from a number of complex charts. Furthermore, it has been pointed out by some respected geostatisticians (e.g., Deutsch and Journel, 1998) that since the local kriging variance is independent of the data values, it should not be used to provide a measure of uncertainty.

The problem of calculating both estimates of TAEP and the associated global error variance was finally solved by employing a conditional simulation procedure. This involved generating a number of equiprobable egg production surfaces on the 3-d grid, each based on the original data and retaining its statistical and geostatistical features. Each egg production surface could be integrated over to obtain individual estimates of TAEP. Generating a large number of simulations then led to the calculation of a distribution of TAEPs, from which the mean TAEP and associated CV could be easily calculated.

The simulations were done using sequential Gaussian simulation (SGS), which first involved a normal scores transformation of the raw egg density data. Variograms were then calculated and modelled using the transformed data. After each simulation was complete, the values were backtransformed according to the original cumulative density function. The TAEP estimates made by SGS were in general comparable with those obtained earlier using ordinary kriging.

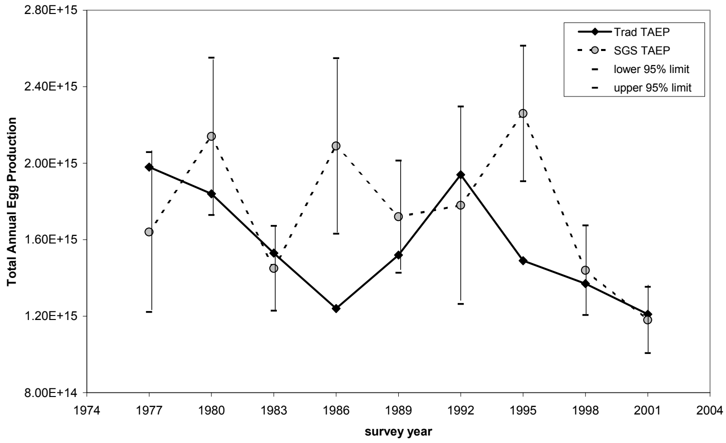


Figure 58. Plot of TAEP estimated by SGS and the Traditional method.

For convenience, Figure 29 has been reproduced in Figure 82, as it summarises the final results of the geostatistical estimator. There were considerable differences in TAEP between the geostatistical and Traditional estimators for a number of the triennial surveys. For example, while the estimates for 1977, 1980 and 1983 were reasonably consistent, the geostatistical estimates of TAEP were considerably higher than the Traditional estimate for 1986. The 1986 dataset is characterised by a delayed start to the survey campaign, with no egg densities available prior to mid-May. The traditional method had dealt with this in a far less generous way than the geostatistical methods, which were able to interpolate over this period in a more realistic way.

The lowest estimation CV (7.5%) was obtained for 2001, as expected. However, there was a somewhat unexpectedly low CV for 1989 (8.7%), and a particularly high value for 1992 (14.8%). An explanation for the high CV in 1992 may reside in the nature of its survey. Although there were a high number of samples collected in 1992, many of these were concentrated in a particular part of the spawning area, and there were no samples collected before week 10 of the spawning season.

The results suggested that SGS is an appropriate methodology for the estimation of TAEP and its uncertainty. The methodology was also considered to be straightforward to implement within a Bayesian framework. A further potential advantage was the ability to generate equiprobable surfaces which keep the histographic and variographic properties of the original dataset. This offered the potential to generate simulated datasets from which to test survey designs.

However, it was be noted that the use of conditional simulation in fisheries is relatively new, and there is a need for more research to be undertaken before the methodology could be used in practice. The procedure has a number of parameters, besides those of the variogram model, which are assumed to be specified correctly. While the subsequent Bayesian analysis considered the uncertainty due to a number of these parameters, more research will need to be undertaken to appreciate the robustness of the algorithm.

Finally, it was noted that much of the development work presented was done using the mackerel egg production data in preference to the horse mackerel data. This was due in part to the fact that the mackerel data appeared to be more spatio-temporally correlated and therefore easier to model. However, midway through the project it emerged that there was some doubt as to whether horse mackerel were determinate or indeterminate spawners, and therefore whether the methodologies developed would be appropriate for their stock assessment. For this reason the results presented in this section focus on the western mackerel. However, the methodology developed should be applicable for the determination of the TAEP for horse mackerel data if this is appropriate.

Incorporation of geostatistical and Bayesian analysis techniques for egg survey data modelling

In Section 2.3, an initial analysis of the egg abundance datasets was first presented, with a subsequent Bayesian assessment of the use of various parameters in the Traditional TAEP estimator. This is followed by a description of the work towards developing the Bayesian Geostatistical estimator. The work involved an initial study into Bayesian estimation of variogram model parameters, then a progression towards a more advanced methodology based on hierarchical modelling for incorporation into the geostatistical procedure developed in the previous section.

As an initial study, a Bayesian implementation of the traditional estimator was constructed and tested. Information was compiled on all the variables that constitute the estimator, from egg counts to final biomass calculation. On each of them, the point estimates presently used were substituted with probability distributions constructed from the distribution of original values or through Monte Carlo simulations, if and when the analysis of the error associated with them revealed that they are significant sources of uncertainty. Otherwise the point estimates were left unchanged. The uncertainty was considered to be significant if the coefficient of variation is greater than 5%.

The first factor to be analysed was the use of the temperature at a depth of 20 m in the development equation. This involved the assumption that it was representative of the entire spawning length of the water column. Egg and temperature profiles were obtained for the months of April, May and June. The daily egg productions were calculated using random egg and temperature profiles. After 10,000 draws, the distributions of values obtained for each month revealed significant differences: the results suggested that the Traditional method might be consistently overestimating daily egg production. By using a higher temperature than the actual temperatures at which most of the eggs are spawned (84% are spawned below 20 m), the length of stage I is underestimated, and therefore a positive bias is introduced in the final estimate.

After it was found that net speed at various depths had a negligible effect on egg density estimation, the analysis was applied to the assumption that eggs are only found within the upper 200 m of the water column. This was expected to be most significant at the start of the spawning season, prior to the formation of the thermocline. It was observed that in a small number of egg profile samples, up to 80% of the eggs were concentrated beneath a depth of 200 m.

A bias model was created, incorporating the two sources identified above in conjunction with their relationships with Julian day and bottom depth. The model was applied to the 1998 dataset, and included the southern as well as the western mackerel component. While the two sources of error contributed in different ways (the use of temperature at 20 m caused a positive bias, the under-sampling of the water column caused a negative bias), the traditional method appeared to overestimate the 1998 daily egg production by an average of 3%. However, it was noted that if higher percentages of the spawning were to take place in deep waters and during the first half of the spawning season, the TAEP estimates could be significantly biased.

It should be noted that these biases are applicable to the egg production calculations, and are not directly applicable to the SSB estimates in Section 2.5.3. A detailed analysis of the calculation of the SSB was beyond the scope of the project, although a less detailed study was undertaken to provide biases and variances for the assessment of management options.

The next part of the Bayesian analysis was devoted to evaluating the uncertainty in the variogram model parameters. This is considered to be the most significant source of uncertainty in the geostatistical modelling procedure. A Bayesian variogram estimation procedure was devised during an initial experimentation with spherical, exponential and Matérn function variogram models. The variogram parameters and their posterior probability distributions were jointly estimated using a maximum likelihood function within a Monte Carlo framework. The resulting set of exponential variogram models was then used to obtain kriging estimates of TAEP, based on the log-transformed data. The resulting CV associated with the mean backtransformed TAEP was 3.3%.

In order to develop the Bayesian-geostatistical estimator, it was necessary to model the variograms obtained with normal scores transformed data. It was considered that limiting the variogram to a strictly exponential or spherical shape could lead to a distribution that did not cover all the possibilities. The variograms modelled in Section 2.2 were often nested, combining a nugget with both spherical and exponential components. The likelihood function was therefore set up to estimate a total of seven parameters. This was somewhat high when the number of data points on the experimental variogram (thirteen) was considered. This problem was surmounted by adopting a Bayesian hierarchical modelling approach. Using a hierarchical approach enabled information to be extracted and shared between all of the variogram models.

The analysis revealed that three of the triennial variograms (1983, 1986 and 1989) appeared to be more linear than the other variograms, which displayed an asymptotic behaviour. These were therefore set aside from the initial hierarchical modelling procedure, and modelled later using priors determined from the posteriors obtained. The priors for the first set of variograms were specified with very wide, although not entirely uninformative, distributions.

The Monte Carlo procedure generated chains of variogram model parameters for each of the survey years. From the ends of each of these chains, 500 models were extracted. These models were listed in a file and read sequentially by the specially modified SGS algorithm. For each set of model parameters, 50 realisations of the egg production surface were generated, thus giving rise to 500 probability distributions for the TAEP. The distributions tended to be significantly positively skewed, which meant that the CV was not an entirely appropriate statistic for reporting the uncertainty. However, it was possible to obtain 95% confidence intervals by ranking the 25,000 TAEP values. The median TAEP was calculated in preference to the mean. The positive skewness was attributed to the response of the SGS algorithm to changes in the variogram range. The TAEP was observed to decrease exponentially as spatial and temporal range increased.

The sensitivity of SGS to variogram parameters led to the failure of the procedure to generate results for the three separated years, 1983, 1986 and 1989. The variogram models obtained for these years were characterised by very high sills and long ranges. For an unknown reason, this caused the SGS program to become unstable and generate vast logs of error messages. While it was not possible to solve this problem within the time scale of this project, it will be subject to future research.

The results for the ‘good’ years were variable. In the case of 1977 and 1980, the median TAEPs were very close to those obtained using the Traditional method. However, 1995 remained considerably higher, while the value for 2001 was significantly lower. The new 1977 estimate is higher than the original SGS value because the hierarchical maximum likelihood method selected ranges that were generally shorter than those modelled subjectively.

The confidence intervals obtained for 1995 and 1998 using the Bayesian method are significantly larger than those obtained in Section 2.2, where the uncertainty in the variogram model parameters was not taken into consideration. However, there was little difference between those obtained for 1977 and 1992. This was attributed to the size of the error bars allocated to the maximum likelihood variogram models (see Figure 36). The error bars for the variograms of 1977 and 1992 were smaller than those attributed to 1998. In each case, however, the lower 95% confidence limit is significantly lower than that obtained using the straightforward SGS method. This suggests that, if the original figures had been passed to a management model, significant and potentially dangerous errors could have been made.

The lowest CV was obtained for the 2001 survey, at 13%. This is higher that the figure passed to the Decision Analysis model used in Section 2.5.3 as a means to compare pre- and post GBMAF management options (11%). The CV used in the Decision Analysis was based on earlier results using ordinary kriging, which suggested that the uncertainty in the variogram parameters would contribute around 3% to the overall estimation variance. Due to the large processing times necessary for both the Bayesian-geostatistical modelling and the management modelling, it was necessary to base the latter on earlier results. In retrospect, the actual post-GBMAF CV for SSB estimates would be greater than 13%. However, the use of 11% in the Decision Analysis, while overoptimistic, should not have significantly affected the results and conclusions in Section 2.5.3.

Comparison of new geostatistical estimators with conventional design based techniques

The research was first focused on developing a realistic egg density simulator in order to provide a means of comparing the estimators and different survey designs. An ideal simulator would be able to generate multiple realisations from a 'known' underlying egg production surface. Using specified survey designs, it would then be possible to extract egg density datasets to be used as input for the various estimators. From a geostatistical point of view, the data sets should have similar variographic properties to those observed in the actual survey data. Furthermore, the density values should also be similarly correlated with the local bathymetry.

The initial work involved assessing the use of a GAM-based simulator, which was adapted from a simulator developed as part of a previous EU project. This enabled an initial comparison of the estimators to be made. However, the simulated data lacked the spatio-temporal auto-correlation that is observed in the survey data, and this meant that it was difficult to properly appraise the performance of the geostatistical estimator.

Following the development of the geostatistical TAEP and variance estimator in the later stages of Section 2.2, it was decided to assess the use of conditional simulation to generate feasible egg production surfaces from which to sample. A number of survey designs were devised, and corresponding data extracted from the simulated egg surfaces. It was then possible to make more comparison between the performances of the different estimators.

The GAM based simulator involved fitting a GAM surface to mackerel egg data, which was then assumed to be the 'true' surface. A measurement error term for each point on the surface was given a negative binomial density function, whereby the error were assumed to be independent in space and time. The surface was based on the 1995 dataset, and was sampled using the 2001 survey design. 1000 simulated datasets were generated and used to make estimates of TAEP by the Traditional, GAM and co-kriging based geostatistical estimators. The variance estimates for the Traditional and GAM methods were calculated using bootstrap techniques.

The TAEP estimates obtained with the Traditional Estimator tended to be less biased than those of the GAM and Geostatistical methods. Furthermore, its estimates were least spread around the 'true' value of 1.715×10^{15}. It was also noted that the inclusion of structural zeroes in the GAM Estimator resulted in a significant negative bias. It was interesting to see that the Geostatistical Estimator, which also included structural zeroes, was positively biased. This contradiction occured because, while the GAM Estimator uses all of the data points (of which around 25% are structural zeroes) for its estimations, the Geostatistical Estimator bases its estimates on only those values that lie within local neighbourhoods.

The values of CV calculated for the TAEP estimations varied widely between each Estimator. The CVs associated with the Traditional Estimator are large, with the error bars covering a wide range of values. These error bars ensured that the 95% confidence interval associated with each TAEP estimate encompassed the true value 99% of the time. The CVs calculated for the Geostatistical Estimator are similar to those calculated for the GAM Estimator, although the GAM Estimator was less successful at encompassing the 'true' TAEP with its error bars, even when the bias was removed. Although the bias-free Geostatistical confidence intervals enclosed the target value in 95 of the cases, it was noted that the proper treatment of kriging variance was still under investigation at that stage.

These preliminary results suggested that in terms of the TAEP estimations, the Traditional Estimator was the most accurate. The simulated dataset, however, was found to be potentially less suited towards the GAM and Geostatistical methods than the actual survey data for the following reasons:

- There was no need for structural zeroes, although these were included in the GAM estimation procedure for the sake of comparison;
- The spatio-temporal correlation structure found in nature was not recreated in the simulated dataset, rendering it less suitable for geostatistical modelling.

The latter drawback led to the consideration of the sequential Gaussian simulation (SGS) algorithm as an alternative means of generating realistic egg production surfaces. Here, it was necessary to assume that the mean TAEP over all the realisations could be treated as the 'true' TAEP. This time the surfaces were based on the 1998 dataset, and were sampled using a variety of different sampling designs. The surfaces were found to possess the appropriate geostatistical characteristics as required, with only slight differences in the resulting variograms.

The survey designs fell into two categories: "random representative" and "real life". The random representative surveys involved sampling a specified number of grid cells randomly in space and time, which would be impossible in real life. However, such a campaign would be theoretically ideal for the implementation of the Traditional and GAM estimators. Datasets containing 500, 1000, 2000, 4000 and 8000 samples were extracted from the set of simulated surfaces.

The "real life" surveys were based on the actual survey campaigns undertaken in 1992, 1995, 1998 and 2001, although the number of samples were also halved and doubled. A total of 18 survey designs were tested with each of the three estimators (Traditional, GAM and SGS). Due to the very long processing times involved, it was not possible to do the study with the Bayesian-geostatistical estimator developed in Section 2.3.

The mean TAEPs, mean CVs and overall biases were calculated for each of the estimators and each of the survey designs. For the Traditional and SGS methods, there was a persistent negative bias. This bias reduced in general when the number of samples was increased, and this effect was most evident for the Traditional method. In general, the SGS method performed better than the Traditional method when there were fewer than 1000 samples. The only survey year for which the SGS method performed better than the Traditional method despite the number of samples was 1992. This may be due to the fact that the survey campaign in that year was delayed, and the geostatistical method is more suited towards extrapolation.

However, it is not known why the negative bias persisted throughout all the survey strategies. There are a number of possibilities, such as:

- The simulated surfaces did not adequately represent 'realistic' egg production surfaces (known problems associated with the SGS algorithm include discontinuity of high values);
- In the case of the SGS estimates, the decision to retain the original variogram model parameters may have been detrimental;
- The survey designs, even those related to the real life sampling campaigns, were relatively random and may contain less samples from the known high density areas than would have been collected in practice.

In contrast, the mean GAM TAEPs were positively biased for many of the 'realistic' survey designs. A look at the histogram of the individual TAEPs for each survey showed that there were a number of extreme estimates which contributed towards the overall positive bias and large CVs. It is not known why the GAM estimator displayed this instability. Although there was no need to add structural zeroes due to the well-behaved nature of the data, MLA found that adding them did not prevent the occurrence of the high TAEPs. One suggestion was that the GAMs were overestimating the seasonality of the spawning season, as 1998 is relatively flat.

A plot of bias against number of samples (Figure 60) showed how the bias dropped dramatically between 500 and 1000 for the Traditional method, and less significantly beyond 1000 samples. The biases were most significant for the 1992 survey. The bias observed for the SGS method (Figure 61) indicated a more gradual reduction with increase of samples, and was greatest for the random representative survey designs. For most of the 'realistic' surveys, increasing the number of samples did not significantly reduce the bias for the SGS method. In 1992, increasing the number of samples to twice the amount of the original survey actually led to an increase in bias. This may have been caused by the large concentration of samples collected towards the south of the spawning area coinciding with a low occurrence of eggs. Although it was difficult to make generalisations regarding the results, it appeared that the finer-scale surveys that are actually undertaken, as opposed to the random representative surveys, tended to suit the Geostatistical method significantly more than the Traditional method.

Table 12 also lists the mean CVs associated with the TAEP estimates obtained using each estimator with each different survey design. The CVs are plotted against the number of samples for each subset of survey designs in Figures 62 and 63 for the Traditional and SGS methods respectively. As expected, the mean CV reduces logarithmically with the number of samples for both estimators. The lowest mean CVs are obtained with the RR8000 survey design at 3.7% and 4.6% for the SGS and Traditional estimators respectively. In each case except T1995, the mean CV of SGS is lower than that of the Traditional method. However, it should be noted that the CVs presented for the SGS method neglect any uncertainty in the modelling parameters. The use of the Bayesian-Geostatistical estimator developed in WP3 would probably increase the CVs by between 2 and 10%, with the additional variance likely to reduce with the number of samples.

The two different methods were observed to behave differently when presented with different survey designs. For example, the curves of reduction in mean CV observed for the Traditional method (Figure 62) tended to converge around 10% when there were 2000 samples, regardless of the survey design. This convergence was not observed with the SGS method (Figure 63), whereby the curves tended to decrease in parallel. Up until 2500 samples, the 2001 survey appeared to have the most suitable design, while beyond 2500 samples, the random representative survey offered more precision. This suggested that when there were 2500 samples, the sampling density was more able to describe the egg production surface than the existing survey types. It is therefore possible that the precision in the estimation of TAEP using the SGS method could be improved by increasing the spatio-temporal extent of the survey coverage instead of increasing the sampling resolution. For the implementation of the Traditional method, increasing the resolution of the sampling campaign would appear to have a similar benefit to increasing the spatio-temporal coverage.

However, in this study, we were limited to one set of simulated egg production surfaces, which were based on the original 1998 survey data. The study would not be complete without assessing the survey designs on surfaces with different characteristics, such as those with early or late spawning peaks. Furthermore, the nature of the survey designs tested has been limited to two types: completely random; and confined to the area and time windows of the original surveys. A more comprehensive study would explore realistic designs which were able to traverse those boundaries.

The results suggested that, of the realistic surveys, the 2001 sampling campaign was best in terms of both accuracy and precision for both Traditional and SGS methods. It was noted that reducing the number of samples collected by 50% led to an increase in CV of only 1.6% for the geostatistical estimator, suggesting that sufficiently good results could be achieved with a significantly reduced effort. However, the 2001 sampling campaign may have been more suited towards the 1998 spawning season due to the increased effort during the earlier months, capturing the slightly early peak detectable in Figure 28.

As a final note, the mean CV for the Traditional estimator obtained using the 2001 design varied depending on the simulator used to generate the data. The mean CV for the 1995-based datasets generated by the GAM method resulted in a mean CV of 22%. Conversely, the 1998-based datasets generated using SGS generated a far lower mean CV of 11.4%. However, due to the lack of knowledge regarding the applicability of the bootstrapping technique for calculating estimation variance, a value close to the original CV was used in the Decision Analysis model to represent pre-GBMAF precision.

Conclusions

The aim of GBMAF was to combine geostatistical and Bayesian methods to improve the scientific basis for the management of Atlantic mackerel fisheries. Until now, a traditional estimator has been applied to produce estimates of Atlantic mackerel SSB based on the data collected in the triennial pelagic egg surveys of Atlantic mackerel. Concerns raised about potential biases and imprecision in this traditional estimator led to the consideration of alternatives to the traditional estimator. In an earlier EC funded project a GAM estimator was developed as an alternative and the statistical properties of the traditional and GAM estimators were evaluated and compared. The simulation evaluations of these estimators revealed some potential biases and substantial imprecision in the total annual egg production in both of these.

The GBMAF study was initiated to consider a third and fourth alternative estimator, namely a geostatistical and Bayesian-geostatistical estimator. The geostatistical estimator was proposed to establish the temporal and spatial autocorrelation patterns in the pelagic egg survey data and to exploit these in the estimation of total annual egg production. It was expected that a geostatistical estimator could improve the precision and potentially the accuracy in the total annual egg estimates by statistically taking into account the spatial-temporal autocorrelation patterns in the data. However, due to the common assumption that the variogram parameters in geostatistical estimators are known without error, it was recognised that geostatistical estimates of TAEP might be overly precise. This is a concern in fisheries modelling because statistical estimates of CVs that are overly precise might give the resulting abundance estimates too much weight in a stock assessment, apparent trends might be given too much weight, and estimates of risks can be under estimates because the fully variability in the data are not accounted for. These problems are compounded by the sparseness of the SSB time series which results from the pelagic egg survey being conducted once every three years due to the large costs and logistical effort required to implement the survey over such a vast area of the Atlantic.

To address the problem of over-precision in geostatistical estimates of TAEP, the fourth alternative, a Bayesian-geostatistical estimator of TAEP was proposed. Unlike classical estimation methods, Bayesian methods treat the values of parameters as random variables and hence Bayesian probability distributions can be assigned to model parameters and using Bayes' theorem, updated with new datasets. Parameters that are otherwise treated as fixed in classical analyses can thereby be considered to be uncertain random variables in Bayesian applications. A common result is that Bayesian estimates more readily account for uncertainties in statistical modelling than do classical methods. At the same time, Bayesian methods can allow the initial prior probability distributions applied to be either non-informative or informative based on previous data analyses or expert judgment. For these reasons, Bayesian methods have been seen to be appealing particularly in fields in which it is important to take into account uncertainty in estimation and the application of results in risk assessment.

Bayesian geostatistical estimators have been developed within the geostatistical literature for over a decade (Omre, 1987; Handcock and Stein, 1993) but until now have not yet been applied in geostatistical analyses of fisheries data. These methods have been developed within Gaussian (Handcock and Stein, 1993) and non-Gaussian frameworks (Diggle *et al*., 1998). GBMAF adopted a Gaussian geostatistical Bayesian modelling approach because of its initial development of a Gaussian stochastic geostatistical estimator of TAEP and its associated uncertainty, and the suitability of the algorithm for application within a Bayesian framework.

The geostatistical modelling methods developed and applied in Section 2.2 advanced the state of the art of application in fisheries geostatistical estimation with a variety of innovations that are highly appropriate to fisheries data. Pelagic egg survey data are characterised by extremely high variability. Logarithmic and Gaussian transformations of the data are often necessary to calculate experimental variograms that reveal the autocorrelation structures in highly skewed datasets. In previous fisheries research, experimental variograms were calculated for log-transformed data, then backtransformed to give a smoother experimental variogram to model for the untransformed variable (Rivoirard *et al*., 2000). However, in this research a more robust method was found to involve co-kriging the log-transformed variable, then applying a backtransformation with a subsequent bias correction determined using cross-validation.

The development of a three-dimensional spatio-temporal geostatistical estimation system was a novel and successful approach towards modelling the egg production surfaces. The improved temporal resolution made it possible to produce estimates of egg production on a weekly basis, enabling the progress of the spawning period to be viewed as a time series. Most importantly, it was possible to account for the variability in time, as opposed to the Traditional method, and previous geostatistical applications which have assumed that the data are representative of the duration of the period they are allocated to (e.g. Bez, 2002). This therefore helped to improve the precision and accuracy of the TAEP. Co-kriging and simulation with the use of transformed depth data also helped considerably to improve precision and accuracy in estimates because of the tendency for the highest egg densities to be found in the region of the shelf break.

The Bayesian-geostatistical modelling methods applied to TAEP in Section 2.3 further advanced the state of the art in fisheries geostatistical estimation for a variety of reasons. Even before the Bayesian-geostatistical modelling methods were developed for TAEP, it was recognised that the parameters used to transform egg counts into egg densities were assumed to be known without error. These parameters included the use of mean temperature at 20 m depth to calibrate egg development, the velocity of the net and the restriction of sampling only the first 200 m of the water column. Because a task of the project was to evaluate the potential long-term benefits of implementing new SSB estimators that used Bayesian-geostatistical methods, the uncertainty and potential biases in these parameters and the impacts of these on TAEP estimates were also evaluated within Section 2.3.

A variety of data, including experimental survey data, were compiled to evaluate the potential error variability in the parameters used to convert egg counts into egg densities. It was found that biases could enter the calculations of TAEP, and hence SSB, because of under-sampling of eggs below 200 m depth (a negative bias) and the assumption that the temperature observed at 20 m depth was representative of the water column in which eggs occurred (a positive bias). These biases, however, largely cancelled each other out and resulted in a net bias of 3%.

Section 2.3 also described the development of a Bayesian-geostatistical approach to TAEP production. Existing methods (e.g., Handcock and Stein 1993) were applied to develop posterior pdfs for variogram parameters. In fitting a parametric variogram model to the experimental variogram (variogram data points based on the observations) conventional geostatistical methods typically treat the parameter estimates obtained as known without error. This will tend to lead to overestimation of the precision in TAEP estimates obtained from the use of the obtained variogram model in the block kriging and sequential Gaussian simulation (SGS) methods applied to estimate TAEP. The joint posterior probability distribution for variogram parameters was applied in a Monte Carlo simulation in which 500 draws of variogram parameter values were taken from the joint posterior pdf. For each of these draws, SGS was implemented to generate 50 equiprobable egg production surfaces, and hence 50 TAEP values. Thereby 500 different pdfs were computed for the TAEP. These were all weighted equally to provide a final Bayesian posterior pdf for TAEP.

While this latter protocol for obtaining a marginal posterior pdf for TAEP is not strictly Bayesian since it applies a procedure similar to bootstrapping to compute the pdf for TAEP given each experimental variogram, we have adopted it because developing a MCMC or SIR approach for the Bayesian integration that would be required in kriging would be exceedingly difficult and computationally very demanding. The general procedure that we present as an approximation for Bayesian kriging is quite straightforward and easily implemented; hence it should be accessible to a wide audience of practitioners.

The computation of a joint posterior probability distribution for the variogram model parameters will tend to prevent such overestimation of precision in TAEP. This was indeed found when the 95% confidence intervals from the 3-D geostatistical estimator were compared with an analogous Bayesian-geostatistical estimator. The CVs in the Bayesian-geostatistical TAEP estimates for most years were in the order of 15-20% compared to about 10-15% for the non-Bayesian geostatistical estimates. The more accurate estimates of precision in Bayesian-geostatistical estimates compared to the non-Bayesian estimates will permit the uncertainty in the geostatistical estimates to be taken into account in stock assessments and projection modelling. Therefore, risks of stock size falling below reference points such as BPA can be more accurately assessed.

A further innovation in Section 2.3 was to apply Bayesian hierarchical modelling of the experimental variogram data from different years. Due to the variability in the experimental variogram points about best fit variogram models and the sparseness of observations near the origin, one of the most critical parts of the variogram, the estimates of variogram parameters for individual years tended to be quite imprecise. Under such conditions, when data from several different similar populations are available, a common approach to Bayesian estimation is to develop a hierarchical model to model the parameter estimates for each individual population on one level, and on a second level model the distribution of parameter values across the different populations (e.g., a mean and variance, assuming some parametric distribution) (Gelman *et al*. 1995). This approach helps to improve the reliability of parameter estimates for individual populations when data in some of the population are rather sparse or relatively uninformative. It is also an approach that is often applied to construct a prior pdf for some "new" population.

A hierarchical modelling approach was applied to the set of experimental variograms from the triennial surveys from 1977 to 2001 to evaluate the potential improvements in precision obtainable by using the combined information available in the joint set of data. It was not possible to obtain satisfactory fits of the hierarchical model to the data when all years were included in the same analysis. Upon inspection of the experimental variograms, it was found that two different patterns exist. One is a well-behaved asymptotic pattern. A second, less well-behaved, is a monotonically increasing non-asymptotic pattern over the range of experimental variogram data points. In the latter set there was a trend to decreasing rates of increase with increasing values on the x-axis but the range of observations on the x-axis was too short for an asymptotic pattern to emerge. Therefore, a hierarchical model was initiated using the set of years with asymptotic experimental variograms. The posterior pdf obtained from this was applied as a prior for the non-asymptotic variograms. In both cases, acceptable fits to the data were obtained and the results were suitable for kriging to produce TAEP estimates for all different years. However, results were not obtained for the non-asymptotic years using SGS due to the inordinately high variograms sills, suggesting that there is more research to be done in this area. Nevertheless, this two-step hierarchical modelling approach was another innovative methodology applied in the development of the Bayesian-geostatistical modelling methodology.

In Section 2.4, the alternative estimators including the traditional, GAM, and geostatistical estimators were tested using two sets of simulated data with known and assumed underlying TAEP values. The Bayesian-geostatistical estimator was not included in the comparison because of its relatively long computing time.

An assessment of each estimator's performance was made across a number of different sampling scenarios. Two studies were made, initially using a GAM-based estimator, and then using a novel SGS-based simulator. The first study involved generating a number of simulated surfaces based on the 1995 data and sampled using the 2001 survey campaign. In terms of bias among the estimators, the Traditional estimator performed best (-3.1%), followed by the kriging-based Geostatistical estimator (+5.4%). The GAM estimator was most significantly biased (-9.7%), probably due to the additional of structural zeroes. In terms of variance, the Traditional method (calculated using a bootstrapping technique rather than the ICES method) was found to overestimate the size of the confidence intervals, whereas the GAM slightly underestimated the uncertainty. The co-kriging-based geostatistical method was not suitable for producing valid estimates of variance.

However, it was noted that the GAM-based datasets did not reproduce the spatio-temporal autocorrelation in the egg surfaces, nor did they possess the observed correlation with depth. For this reason, a novel SGS-based simulator was developed and used to generate surfaces with which to compare the simulators over a number of different survey designs. This then formed part of an improved methodology for testing the relative estimation performance of different estimators and diagnosing the causes of estimation failure for pelagic egg estimators of spawners abundance. The technique allows the generation of many equiprobable egg production surfaces based on a set of survey data. While there is no 'true' underlying surface, the data generated retain the geostatistical and statistical properties of the original dataset.

The GAM method was found to be somewhat unstable, for unknown reasons, and detailed assessment of its performance was therefore not possible. It was suggested that the flatness of the 1998 egg production curve may have caused problems for the smoothing algorithm. However, the results obtained for the original 1998 data are in line with those obtained with the Traditional, Geostatistical and Bayesian-geostatistical estimators. More research would be required to identify the root of the problem.

In terms of bias, both the Traditional and SGS estimators were persistently negatively biased. This may have been due to inadequacies in the simulated surfaces, or by using relatively less samples from high-density areas. As expected, the bias decreased with the number of samples, with the curve tending to be steeper for the Traditional method. In general, the finer scale pseudo-regular surveys (similar to those actually undertaken) seemed to work better for the SGS method than for the Traditional method, which preferred the random representative surveys. The bias was less significant for the SGS method in comparison with the Traditional method when the number of samples was less than 1000.

In terms of CV, the uncertainty decreased logarithmically with the number of samples for both the Traditional and SGS methods. The CVs were more sensitive to the survey design for the SGS method, whereas beyond 2000 samples, the differences in CV for the Traditional method were negligible. For the SGS method, the 2001 survey-based TAEPs were generally the most precise, although beyond 2500 samples, the random representative survey performed better. This suggested that the overall spatio-temporal coverage of the random sampling strategy improves on the transect-based surveys when the sampling resolution is sufficiently fine. A tentative conclusion would therefore be that future surveys should concentrate on increasing the spatio-temporal coverage, rather than increasing sampling resolution.

In almost all cases the CV was less for the SGS estimator than for the Traditional method. However, it was noted that the analysis was done without considering the variance in the modelling parameters, as the use of the Bayesian-geostatistical estimator would have been prohibitively time-consuming. Considering the uncertainty in the model parameters would have potentially increased the CV of the Geostatistical estimates by between 2 and 10%, which would render it greater than the variance in the Traditional estimates on most occasions. However, it is not known how reliable the Traditional variance estimates are in terms of reflecting the true uncertainty. The method of variance calculation used by ICES was rejected in the current study for various reasons, such as its neglect of zero values. The Bayesian-geostatistical method would appear to provide the most theoretically reasonable measures of uncertainty, although the method in general requires more research into its optimal implementation.

In summary the GB estimators developed in this study could provide an improved scientific basis for fisheries management of Atlantic mackerel for the following reasons:

1) Unlike other alternative estimators, geostatistical estimators statistically account for the temporal and spatial autocorrelation patterns in the pelagic egg survey data. This can improve the reliability of the TAEP and SSB estimates as this study found marked spatial and temporal autocorrelations in the pelagic egg survey data that could be exploited.
2) The development of geostatistical Bayesian estimators of TAEP and SSB can more effectively account for the spatial temporal existing patterns in the data by providing a hierarchical modelling framework to combine data from different years. The improved ability to account for parameter uncertainty and structural uncertainty in the formulation of the geostatistical estimators can permit more reliable accounting of uncertainty in SSB estimates and in the stock assessments and projection modelling done to provide fisheries management advice.

REFERENCES

Armstrong, M., Renard, D., Rivoirard, J. and Petitgas, P. 1992. Geostatistics for fish survey data, course published by ICES, Centré de Géostatistique, Fontainebleau.

Armstrong, M., 1998. *Basic Linear Geostatistics*. Springer-Verlag, Berlin, Heidleberg, New York.

Augustin, N. H., D. L. Borchers, et al. 1998. Spatiotemporal modelling for the annual egg production methods of stock assessment using generalized additive models. *Canadian Journal Fisheries & Aquatic Sciences* 55: 2608-2621.

Berger, J. O. 1985. *Statistical decision theory and Bayesian analysis*. Springer-Verlag, New York ; London.

Best, N., Cowles, M. K. and Vines, K., 1996: *CODA: Convergence diagnostics and output analysis software for Gibbs sampling output. Version 0.30*. Technical Report, MRC Biostatistics Unit.

Bogaert, P., 1999: Assessing the variability of the variogram estimator. *Proceedings: geoENV II – Geostatistics for Environmental Applications*, pp. 479 – 490. Kluwer Academic Publishers.

Bez, N., Rivoirard, J., Guiblin, P. H. & Walsh, M., 1996. Covariogram and related tools for structural analysis of fish survey data. In *Geostatistics Wollongong '96*, vol. 2 (ed. E. Y. Baafi and N. A. Schofield), pp. 1316-1327. Kluwer Academic Publishers.

Bez, N. and Rivoirard, J., 2000. On the role of sea surface temperature on the spatial distribution of early stages of mackerel using inertiograms. *ICES Journal of Marine Science*, **57**: 383-392.

Borchers, D. L., S. T. Buckland, et al. 1997. Improving the precision of the daily egg production method using generalized additive models. *Canadian Journal Fisheries & Aquatic Sciences* 54: 2727-2742.

Borchers, D. L., Richardson, A. & Motos, L., 1997. Modelling the spatial distribution of fish eggs using generalized additive models. *Ozeanografika* **1,** 103-120.

Brus, D. J., Jansen, M. J. W. and De Gruiter, J. J., 2002: Optimizing two- and three-stage designs for spatial inventories of natural resources by simulated annealing. *Environmental and Ecological Statistics*, **9**, 71-88.

Caers, J., 2000. Adding local accuracy to direct sequential simulation. *Mathematical Geology*, 32, 815-850.

Casey J., Nicholson, M. D., Warnes S. 1992. Selectivity of square mesh cod-ends on pelagic trawls for Atlantic mackerel (Scomber-Scombrus L). *Fish. Res*. 13: 267-279.

Chen, M.-H., Q.-M. Shao and J. G. Ibrahim. 2000. *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag, New York.

Clark, W.G. 1999. Effects of an erroneous natural mortality rate on a simple age-structured stock assessment. *Can. J. Fish. Aquat. Sci*. **56**:1721-1731.

Coombs, S. H., Morgans, D. and Halliday, N. C., 1996. The vertical distribution of eggs and larvae of mackerel (Scomber scombrus). Copenhagen, ICES.

Defeo, O. and Rueda, M., 2002: Spatial structure, sampling design and abundance estimates in sandy beach macroinfauna: some warnings and new perspectives. *Marine Biology*, **140**, 1215-1225.

De Oliveira, V., Kedem, B. and Short, D. A., 1997: Bayesian prediction of transformed Gaussian random fields. *J. Am. Stat. Ass.*, **92**, 1422-1433.

Deriso, R.B., Quinn II, T.J., and Neal, P.R. 1985. Catch-age Analysis with Auxiliary Information. *Can. J. Fish. Aquat. Sci*. **42**:815-824.

Deutsch, C. V. and Journel, A. G., 1998. *GSLIB Geostatistical Software Library and Users Guide*. 2nd Edition. Oxford University Press, New York.

Doonan, I. J., Bull, B. and Coombs, R. F., 2003: Star acoustic surveys of localised fish aggregations. *ICES J. Mar. Sci*., **60**: 132-146.

Doubleday, W.G. 1976. A Least Squared Approach to Analysing Catch at Age Data. *Res. Bull. Int. Comm. NW Atl. Fish*. **12**:69-81.

Ellison, A. M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecological Applications* 6: 1036-1046.

Fletcher, W. J. and Sumner, N. R., 1999: Spatial distribution of sardine (Sardinops sagax) eggs and larvae: an application of geostatistics and resampling to survey data. *Can. J. Fish. Aquat. Sci*., **56**, 907-914.

Fournier, D., and Archibald, C.P. 1982. A General Theory for Analyzing Catch at Age Data. *Can. J. Fish. Aquat. Sci*. **39**:1195-1207.

Francis, R.I.C.C. 1992. Use of Risk Analysis to Assess Fishery Management Strategies: A Case Study using Orange Roughy (*Hoplostethus atlanticus*) on the Chatham Rise, New Zealand. *Can. J. Fish. Aquat. Sci*. **49**:922-930

Francis, R.I.C.C., Robertson, D.A., Clark, M.R., and Coburn, R.P. 1992. Assessment of the ORH 3B orange roughy fishery for the 1992/93 fishing year. New Zealand Fisheries Assessment Research Document 92/4. pp. 45.

Gamerman, D. 1997. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall, London.

Gelman, A. & Rubin, D.B. (1992) Inference from Iterative Simulation Using Multiple Sequences. Statistical Science **7**(4):457- 511

Gelman, A., J. B. Carlin, et al. 1995. *Bayesian Data Analysis*. Chapman & Hall, London.

Geromont, H.F., and Butterworth, D.S. 2000. Possible extensions to the ADAPT VPA model applied to Western North Atlantic Bluefin Tuna, addressing in particular the need to account for "additional variance".

Geweke, J. 1989. Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica*. **57**(6):1317-1339.

Geyer, C.J. 1992. Practical Markov Chain Monte Carlo. *Statistical Sciences* **7**(4):473-503.

Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, Oxford.

Hamre, J. 1978. The effect of recent changes in the North Sea mackerel fishery on stock and yield. Rapp. P. –v. Reun. Cons. Int. Explor. Mer. 172: 197-210.

Handcock, M.S. and M.L. Stein. 1993. A Bayesian-Analysis of Kriging. *Technometrics* 35: 403-410.

Hastie, T. & Tibshirani, R., 1990. *Generalized Additive Models*, 1st edition. Chapman & Hall, London.

Hastings, W.K. 1970. Monte Carol sampling methods using Markov chains and their applications. *Biometrika* **57**(1):97-109

Hilborn, R. & Walters, C. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Chapman & Hall, Inc., New York.

Hilborn, R. 1992. Current and Future Trends in Fisheries Stock Assessment and Management. *S. Afr. J. mar. Sci*. **12**:975-988

Hilborn, R., E. K. Pikitch, et al. 1994. A Bayesian-Estimation and Decision-Analysis For an Age-Structured Model Using Biomass Survey Data. *Fisheries Research* 19: 17-30.

Hilborn, R., Pikitch, E.K. & McAllister, M.K. 1994. A Bayesian estimation and decision analysis for an age-structured model using biomass survey data. *Fisheries Research* **19**:17-30

Houghton, R., 1987. The consistency of the spatial distributions of young gadoids with time. *ICES CM1987/D:15*.

ICES, 1999. Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy. ICES, C.M. 2000/ACFM:5.

Journel, A. and Hiujbregts, C., 1978. *Mining Geostatistics*, Academic Press, London.

Kimura, D.K. 1989. Variability, tuning, and simulation for the Doubleday-Deriso catch-at-age model. *Can. J. Fish. Aquat. Sci*. **46**:941-949.

Kinas, P.G. 1996. Bayesian fishery stock assessment and decision making using adaptive importance sampling. *Can. J. Fish. Aquat. Sci*. **53**:414-423

Kolody, D. & Patterson, K. 1999. Evaluation of NE Atlantic mackerel stock assessment models on the basis of simulated long-term management performance. *ICES CM/S:01*.

Kolody, D. & Patterson, K. 1999. Evaluation of NE Atlantic Mackerel Stock Assessment Models on the Basis of Simulated Long-Term Management Performance. Working Document for the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy, September, 1999

Kuikka, S., Hildén, M., Gislason, H., Hansson, S., Sparholt, H. and Varis, O. 1999. Environmentally Driven Uncertainties in Baltic Cod Management - Modelling by Bayesian Influence Diagrams. *Can. J. Fish. Aqua. Sci*., 56: 629-641.

Kuikka, S., M. Hilden, et al. 1999. Modeling environmentally driven uncertainties in Baltic cod (Gadus morhua) management by Bayesian influence diagrams. *Canadian Journal of Fisheries and Aquatic Sciences* 56: 629-641.

Kvalsvik, K., Misund, O. A., Gamst, K., Holst, R., and Galbraith, D. Selectivity experiments using sorting grid in pelagic mackerel trawl. Unpubl. manus.

Liermann, M. and R. Hilborn. 1997. Depensation in fish stocks: a hierarchic Bayesian meta-analysis. *Canadian Journal of Fisheries and Aquatic Sciences* 54: 1976-1984.

Lockwood, S. J. and J. G. Shepherd. 1984. An assessment of the western Mackerel stock. *J. Cons. Ciem*. 41: 167-180.

Lockwood, S. J., 1988. *The Mackerel – its biology, assessment and the management of a fishery*. Fishing News Books Ltd, Farnham, UK.

Lockwood, S. J., J. H. Nichols, et al. 1977. The development rates of mackerel (Scomber scoumbrus, L.) eggs over a range of temperature. Copenhagen, ICES.

Lockwood, S., Nichols, J. and Dawson, W., 1981. The estimation of mackerel (*scomber scombrus L.*) spawning stock size by plankton survey. *J. Plankton Research*, No 3 and 2.

Matheron, G., 1971. The theory of regionalised variables and its applications. 212p.

McAllister, M. K. and E. K. Pikitch. 1997. A Bayesian approach to choosing a design for surveying fishery resources: application to the eastern Bering Sea trawl survey. *Canadian Journal of Fisheries & Aquatic Sciences* 54: 301-311.

McAllister, M. K. and G. P. Kirkwood. 1998. Bayesian stock assessment: a review and example application using the logistic model. *ICES Journal of Marine Science* 55: 1031-1060.

McAllister, M., and Kirchner, C. 2000. Accounting for structural uncertainty to facilitate precautionary fishery management: illustration with Namibian orange roughy. Draft.

McAllister, M., Babcock, E.A. & Pikitch, E.K. 2000a. Using Bayesian Methods to Provide an Empirical Basis to Weight Conflicting Stock Assessment Results on Stock Rebuilding when there is Uncertainty over Processes Affecting Future Recruitment. *International Commission for the Conservation of Atlantic Tunas, Madrid. Document SCRS/00/103*.

McAllister, M., Babcock, E.A. & Pikitch, E.K. 2000b. Using Bayesian Methods and Decision Analysis as a Rational Basis to Dealing with Conflicting Stock Assessment Results while Providing Management Advice on Stock Rebuilding. *International Commission for the Conservation of Atlantic Tunas, Madrid. Document SCRS/00/34*.

McAllister, M.K. & Ianelli, J.N. 1997. Bayesian stock assessment using catch-age data and the sampling – importance resampling algorithm. *Can. J. Fish. Aquat. Sci*. **54**(2):284-300).

McAllister, M.K. & Kirkwood, G.P. 1998a. Using Bayesian decision analysis to help achieve a precautionary approach for managing developing fisheries. *Can. J. Fish. Aquat. Sci*. **55**(12):2642-2661.

McAllister, M.K. & Kirkwood, G.P. 1998b. Bayesian stock assessment: a review and example application using the logistic model. *ICES J. Mar. Sci*. **55**:1031-1060.

McAllister, M.K. & Pikitch, E.K. 1997. A Bayesian approach to choosing a design for surveying fishery resources: application to the eastern Bering Sea trawl survey. *Can. J. Fish. Aquat. Sci*. **54**:301-311.

McAllister, M.K., Pikitch, E.K., Punt, A.E. & Hilborn, R. 1994. A Bayesian Approach to Stock Assessment and Harvest Decisions Using the Sampling/Importance Resampling Algorithm. *Can. J. Fish. Aquat. Sci*. **51**:2673-2687.

Megrey, B.A. 1989. Review and Comparison of Age-structured Stock Assessment Models from Theoretical and Applied Points of View. *American Fisheries Society Symposium*. **6**:8-48.

Misund, O. A. and Beltestad, A. K. 2000. Survival of mackerel and saithe that escape through sorting grids in purse seines. *Fish. Res*. 48: 31 – 41.

Myers, R. and Stokes, T., 1989. *Density-dependent habitat utilisation of groundfish and the improvement of research surveys*. ICES CM1989/D15.

O'Hagan, A. 1998. Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society Series D-the Statistician* 47: 21-35.

Overholtz, W.J. 1993. Harvesting strategies and fishing mortality reference point comparisons for the northwest Atlantic stock of Atlantic Mackerel (Scomber-Scombrus). *Can. J. Fish. Aqua*. Sci.50: 1749-1756.

Pannatier, Y., 1996. *VARIOWIN: Software for spatial data analysis*. Springer-Verlag, New York.

Patterson, K.R. 1999. Evaluating uncertainty in harvest control law catches using Bayesian Markov chain Monte Carlo virtual population analysis with adaptive rejection sampling and including structural uncertainty. *Can. J. Fish. Aquat. Sci*. **56**(2):208-221

Petitgas, P., 1993. Geostatistics for fish stock assessment: a review and an acoustic application. *ICES Journal of Marine Science*, 50: 285-298.

Petitgas, P., 1997. Sole egg distributions in space and time characterised by a geostatistical model and its estimation variance. *ICES Journal of Marine Science*, 54:213-225.

Petitgas, P., 1998. A review of linear geostatistics for fisheries survey design and stock assessment. *Proc. 2nd European Conference on Geostatistics for Environmental Applications*, Valencia, Spain. J. Gomez-Hernandez, A. Soares and R. Froidenaux eds., Kluwer Academic Publishers.

Pilz, J., Spoeck, G. and Schimek, M. G., 1996: Taking account of uncertainty in spatial covariance estimation. In *Geostatistics Wollongong '96*, vol. 1 (ed. E. Y. Baafi and N. A. Schofield), pp. 302-313. Kluwer Academic Publishers.

Pope, J.G. 1977. Estimation of fishing mortality, its precision and implications for the management of fisheries. In *Fisheries Mathematics*. *Edited by* J.H. Steel. Academic press, New York. pp 63-76.

Powers, J.E., and Restrepo, V.R. 1993. Evaluation of stock assessment research for Gulf of Mexico king mackerel: benefits and costs to management. *N. Am. J. Fish. Manage*. **13**:15-26.

Punt, A. E. and R. Hilborn. 1997. Fisheries stock assessment and decision analysis: The Bayesian approach. *Reviews in Fish Biology and Fisheries,* 7: 35-63.

Punt, A.E. & Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. *Reviews in Fish Biology and Fisheries* **7**:35-63

Punt, A.E., Pikitch, E.K., McAllister, M.K. & Hilborn, R. (1993) Stock assessment and decision analysis for the western stock of hoki (*Macruronus novaezelandiae*) for 1993. New Zealand Fisheries Assessment Research Document 93/13.

Quinn, T.J. & Deriso, R.B. 1999. *Quantitative Fish Dynamics*. Oxford University Press. New York. Oxford. 542pp

Raftery, A.E. Givens, G.H. & Zeh, J.E. 1995. Inference from a Deterministic Population Dynamics Model for Bowhead Whales. *Journal of the American Statistical Association*, **90**(430):402-430.

Ribeiro, P.J. and Diggle, P.J. 1999. Bayesian inference in Gaussian model-based geostatistics. *Technical Report ST-99-08*. Department of Mathematics and Statistics, University of Lancaster.

Richards, L.J. 1991. Use of contradictory data sources in stock assessments. *Fisheries Research*. **11**:225-238.

Rivoirard, J., Simmonds, J., Foote, K. G., Fernandes, P. and Bez, N., 2000. *Geostatistics for Estimating Fish Abundance*. Blackwell Science, Oxford.

Rubin, D.B. 1987. Comment: A Noniterative Sampling/Importance Resampling Alternative to the Data Augmentation Algorithm for Creating a Few Imputations when Fractions of Missing Information are Models: The SIR Algorithm. *Journal of the American Statistical Association*. **82**(398):543-546.

Safai-Naraghi, K. and Marcotte, D., 1996: Bootstrapping variograms. In *Geostatistics Wollongong '96*, vol. 1 (ed. E. Y. Baafi and N. A. Schofield), pp. 188-199. Kluwer Academic Publishers.

Schnute, J.T., and Richards, L.J. 1995. The influence of error on population estimates from catch-age models. *Can. J. Fish. Aquat. Sci*. **52**:2063-2077.

Seber, G.A.F. 1982. *Estimation of animal abundance and related parameters*. Macmillan, New York.

Shelton, P.A. 1992. The Shape of Recruitment Distributions. *Can. J. Fish. Aquat. Sci*. **49**:1754-1761.

Simmonds, E. J. and Fryer, R. J., 1996: Which are better, random or systematic acoustic surveys? A simulation using North Sea herring as an example. *ICES J. Mar. Sci*., **53**, 39-50.

Smith, A. F. M. and J. M. Bernardo. 2000. *Bayesian Theory*. Wiley, Chichester.

Soares, A., 2001. Direct sequential simulation and cosimulation. *Mathematical Geology*, **33**, 911-926.

Stein, M.L. 1999. *Interpolation of spatial data: Some theory for Kriging*. Springer, New York.

Sullivan, P.J., 1991. Stock Abundance Estimation using depth-dependent trends and spatially correlated variation. . *Canadian Journal of Fisheries and Aquatic Sciences*. 48: 1691-1703.

Thayer, W. C., Goodrum, P. G., Diamond, G. L. and Hassett, J. M., 2000. *Application of Geostatistics to Risk Assessment*. Syracuse Research Corporation, New York.

Thompson, B. M., Guegen, J. C., Schoefer, W., Eltink, A., Walsh, M. & Coombs, S. H., 1984. *The western mackerel spawning stock estimate for 1983*. ICES, Copenhagen (Denmark).

Van Dijk, H.K. & Kloek, T. 1983. Monte Carlo Analysis of Skew Posterior Distributions: an Illustrative Econometric Example. *The Statistician* **32**:216-223.

Van Dijk, H.K., Hop, J.P. & Louter, A.S. 1987. An algorithm for the computation of posterior moments and densities using simple importance sampling. *The Statistician* **36**:83-90.

Van Groenigen, J. W., 2000: The influence of variogram parameters on optimal sampling schemes for mapping by kriging. *Geoderma*, **97**, 223-236.

Walters, C. & Ludwig, D. 1994. Calculation of Bayes Posterior Probability Distributions for Key Population Parameters. *Can. J. Fish. Aquat. Sci*., **51**:713-722.

West, M. 1993. Approximating Posterior Distributions by Mixtures. *J. R. Statist. Soc. B*. **55**(2):409-422.

WGMEG. 1999. Report of the Working Group on Mackerel and Horse Mackerel Egg Surveys. Copenhagen, ICES: 83.

WGMEG. 2000. Report of the Working Group on Mackerel and Horse Mackerel Egg Surveys. Copenhagen, ICES: 54.

ICES / CIEM
Workshop on Survey Design and Analysis Working Document
21-25 June 2004, Aberdeen

Geostatistical approach to biomass estimations from survey data

Bez N.*, Rivoirard J.* and Petitgas P.**

* Centre de Géostatistique, 35 rue St Honoré, 77305 Fontainebleau, France, tel : +33164694700, Fax :+33164694705 ; nicolas.bez@ensmp.fr, jacques.rivoirard@ensmp.fr
** IFREMER, rue de l'Île d'Yeu, 44311 cedex 3 Nantes, France. tel: +33240374163; Fax: 33240374075; pierre.petitgas@ifremer.fr

1 Introduction

The ICES workshop held in Reykjavik in 1991 contributed to explain the difference between the estimation variance used a classical and a geostatistical approach. However, the concept of estimation variance is sometimes misleading and we believe useful to start the present review by defining the estimation variance (formulae and meaning). This allows providing the basic concepts of geostatistics useful for estimating fish abundance from survey data. This document is far from a technical note on the method. It is rather a point of view (feed back) based on the various study areas in which we have been involved in the past 10 years.

The first section is devoted to the classical approach where variograms are used to compute estimation variances or to interpolate the fish concentration between data points. We then introduce the transitive approach which appeared to be an operational and robust approach for global estimations. The last section is devoted to list the major problematic aspects of geostatistics applied to fisheries data. The conclusion section amounts to a pros & cons summary table based on the previous sections.

2 Classical geostatistical approach

2.1 Domain

The first step of any spatial analysis is the definition of the domain to be used for the analyses. Despite the apparent facility, this is a strategic step where one decides for the sub set of active sample points, chooses the zero data to be considered as inner or outer zeroes, etc. This happens to be particularly difficult for populations with diffuse limits (eggs & larvae for instance) and for long series of surveys (one domain to be defined per survey). Biological knowledge and stability observed over series of surveys can help deciding for study area or habitats.

2.2 Spatial structure

After a domain is defined, (geo)statistical analyses can proceed further on, aiming at the definition of a covariance type function. This function describes the spatial autocorrelation (spatial structure) present in the data. The most used tool is the variogram:

$$\gamma(h) = C(0) - C(h)$$

where

$C(h)$ is the covariance at distance h

$C(0)$ is the variance

An empirical variogram can always be computed. In favourable cases, it helps quantifying the key scales of the autocorrelation of the phenomenon. However, it is of no further use as long as it is not modelled. Defining a variogram model implicitly assumes that increments of fish concentrations are stationary, i.e., that their statistical characteristic, and more specifically their mean and variance, do not depend on the location the points considered but only on the distance between them. Defining a variogram model also means that the data values $z(x_i), i \in [1, N]$ are considered as the outcomes of a Random Function $Z(x), x \in R^d$. This process is comparable with the classical statistical approach which is based on the definition of a distribution for the random variable Y from the empirical histogram of the data $y_i, i \in [1, N]$. The parameters of the model can be inferred, estimated or more simply chosen. The structural analysis, that is the computation of an empirical variogram and its modelisation, represents the central part of any geostatistical analysis. This crucial step is often difficult in fisheries (outliers, locations of zero data, stationarity, spatial distribution dependent on the proximity to the edge or the hart of the population habitat, etc). Cautious is required but its forces one to look at the variability of the data before using it.

To speed up multi survey analyses or to reduce the subjectivity of the variogram fit, automatic (and semi-automatic) fitting procedure have been developed based on various minimization criteria (sum of square errors, relative mean square errors, etc).

2.3 Global estimation variance

2.3.1 Formulae

When estimating the mean concentration over a domain V by the mean of its samples $Z(x_i)$

$$Z^*(V) = \frac{1}{N}\sum_{i=1}^{N} Z(x_i)$$

the estimation variance needs only the variogram model:

$$\sigma^2_E = \mathrm{var}\left(Z(V) - Z^*(V)\right) = 2\overline{\gamma}(V,x) - \overline{\gamma}(V,V) - \overline{\gamma}(x,x)$$

It is in general different from the classical estimation variance

$$\sigma^2_{iid} = \frac{\sigma^2}{N}$$

It can be either smaller or larger depending on the sample spacing, on the geometry of the target and of the spatial structure. So knowing the spatial structure allows quantifying the precision of estimation. Coefficient of variation (often expressed in %), without dimension, can be preferred

$$CV = \frac{100 \cdot \sigma_E}{Z^*(V)}$$

2.3.2 Estimation variance and variability of the estimate

The estimation variance $\mathrm{var}\left(Z(V) - Z^*(V)\right)$ is conceptually different from the variability of the estimate $\mathrm{var}\left(Z^*(V)\right)$ often used in statistics to quantify uncertainties.

The former expression has the advantage to depend on the target of the estimation, i.e. $Z(V)$, and not only on the sample points. Quantifying only the variability of the estimate is not sufficient to evaluate the quality of the estimation. Let us consider for instance an acoustic line of size L divided in N segments (typically the segments correspond to ESDU for acoustic surveys). The estimate of the biomass along that line $Z(L)$ is error free as we know exhaustively the line. In general, estimators are expected to converge to the true value when the number of samples increases homogeneously. In the particular case of exhaustive sampling, the estimate equals the truth and the estimation variance is zero. However, the estimate of the biomass over an area including the line is no longer error free. Evaluating how well the area is estimated by the fish concentration over the line requires knowing the statistical links between the fish concentration in two points, one on the line and the other one in the area to estimate. The quality of an estimate must consider the experimental material available for the estimation and the target.

The convergence of some estimators is sometimes questionable. Applying a geostatistical approach to the first case (i.e. one line made of exhaustively know segments) would lead to $\sigma^2_E = 0$; which is consistent. However, a bootstrap procedure whose advantage is to quantify the fluctuation of estimators, would end up with some variability for $Z^*(L)$. In this regards, bootstrap estimators fail at converging. This also stands for GLM approaches when modelling fish concentrations by the sum of smooth functions of explanatory (environmental) variables plus a residual. As a matter of fact, even with exhaustive information, fish concentration can probably be modelled by a set of smooth functions having residuals. The bootstrap procedure based on these residuals will then provide, contrary to the convergence property, a positive estimation variance.

2.3.3 Consistency with sampling scale

The classical estimation variance σ^2 / N is not robust against change of sampling scale, that is against change of ESDU for acoustic surveys, change of trawl duration for trawl survey, or also against change of dimensions (2D to 1D by integration over acoustic transects for instance).
It has been observed for instance that NASC values appear spatially independent when integrated over intermediate ESDU size but show some spatial structure at another scale.
Geostatistical models are consistent in this regard.

2.3.4 Model-based estimate but not design-based

The method can be applied irrespective of the sampling design (except for inhomogeneous sampling schemes). As a matter of fact, the method can be applied as long as (i) the sample location, (ii) the target location(s) and (iii) the spatial structure are known (the locations can be fixed or random). The structure can be either based on real data (the usual case) or defined by external means. Optimization of sampling design can thus be undertaken. In particular one can show that a random stratified sampling (one sample located at random in each of N statistical squares) is always better than a purely random sampling, and that a systematic design is often better than a random stratified survey, but not always.

Inhomogeneous designs (clusters of samples, oversampling around samples with large values, etc) raise specific problems.

2.4 Mapping

Kriging can be applied to weight neighbours for optimal interpolation. Estimation variance is then also computed. The estimation variance is not a conditional variance, i.e. sensitive to the relative locations of the known and known points but independent to the particular values of the known points. The estimation variance does not, for instance, account for proportional effects (heteroscedasticity).

3 Transitive

The above mentioned techniques, as most of the statistical tools (histogram, regressions, correlation, etc), are domain dependent. They change when the domain changes. This happens to be a strong limitation when considering populations with diffuse limits (e.g. eggs). In these cases, a transitive approach can be advocated, replacing the variogram by the (transitive) covariogram. Roughly speaking, moving to transitive approach amounts to replacing “means” by “sums”. While the previous method is appropriate for estimating mean concentrations, the transitive approach is thus designed to estimate total biomasses.

For regular sampling designs (or to some extend to regular stratified sampling), global estimations with estimation variance can be fully undertaken since a covariogram model *g(h)* is defined:

$$\sigma^2{}_E = a\sum_k g(k\cdot a) - \int g(h)dh \quad \text{where } a \text{ is the grid mesh}$$

Kriging map can also be made (simple interpolation; no estimation variance).

This approach, which solves the problem of the zeroes, is also far less sensitive to large values than the classical one. However, it is, in practice, restricted to regular samplings (design-based).

4 Problematic aspects of geostatistics applied to fisheries

4.1 Large values

Extreme values remain problematic. In most cases, they correspond to outliers and not to measurement errors, so that the tiny proportion of the samples is responsible for a large proportion of the total abundance.

4.1.1 Robust tools

To correct for the destructuration induced by outliers, some tools are available though not a panacea:

- "Robust" variograms (only robust for some distributions; not justified otherwise)
- Variograms along the vessel track for acoustic surveys for instance
- Back transformation after log transformation of the data
- Weighted variograms (with a weight related to the surface of influence of each sample for instance)

4.1.2 Disjunctive kriging & poststratification

More sophisticated non linear techniques consist in splitting the study variable in classes of values and to consider the spatial relationship between the regions associated to each thresholds. Large values no longer contribute through their particular value but only through to their membership to a class of large values. Such models can be used, for instance, to delineate rich zones (Disjunctive kriging) but have not yet been used to provide estimation variance for estimation of index of abundance. They usually require more assumptions than the previous linear approaches.

Non linear approaches can serve as tools to delineate post-strata, allowing different but more appropriate models in each of the strata. Typically, assuming that spatial structure vanishes with increasing concentrations, there could be a geostatistical approach for the stratum with low concentrations and a statistical approach for the rich strata.

4.2 Inhomogeneous sampling

Re-sampling around large value is an attractive approach to reduce estimation variance. As a matter of fact, this allows having more samples where variability is larger. However, the statistical community has not endorsed complete analyses of these approaches.
From a geostatistical perspective, given that an observation is larger than a prespecified threshold, it is expected (and observed) that the average value obtained by a repeated sampling is lower than the first measure. The leads to re-visit the decision rule for such type of sampling scheme suggesting that a repeated sampling should be performed only when the local *mean* concentration around the sampling location exceeds a pre-specified threshold (block kriging).

4.3 Multivariate approaches

4.3.1 Full process variance

Estimation variances are usually computed for one particular variable, say number of fish per hour trawling. When the final estimate is the combination of several variables like for instance in acoustic surveys (fish length, fish abundance an NASC), estimation variance should quantify the uncertainty of entire process. However, relationships between variables are often difficult to identify and to model. Moreover, these links are not linear rending the classical multivariate geostatistical model inappropriate. The question of full process variance is not straightforward and simulations could be used to this end (e.g. Scottish herring).

4.3.2 Auxiliary variables

The estimation variance provided by a geostatistical analysis refers to the error made when fill in the gaps in between sample data. It includes unsystematic measurement errors (nugget effect) but does not include other source of the uncertainties (selectivity, availability, vessel effects, etc). The coefficients of variation associated to these estimates are generally low or medium (from 5% for very accurate estimations to 15 or 20% for medium ones). This leads to suggest that the errors coming from the spatial interpolation of observations are generally rather low compared to the inter annual fluctuations of biomass indices. In other words, satisfactory methods exist to spatially fill in the gaps but the parameters controlling the fluctuations of the biomass are not taken into account.

4.3.3 Time and others…

Taking time, fish movements, or capturability (gear, vessel) into account increases the dimension of work meanwhile the number of observation is stable. This induces larger estimation variances which is logical but disappointing. This supports the idea of studying alternative sampling designs.

4.4 Confidence interval

When assumptions are made on the errors' distributions, estimation variances can be used to define confidence intervals. In essence, geostatistics does not prevent to compute such confidence intervals even though it is usually not done. The fact is that the additional required assumption is sometimes considered fragile and not enough supported by data. Adding assumptions "allows more" but also reduces the overall robustness of the result.
Conditional simulations can be an alternative here to generate a full distribution for the estimator.

Pros	Cons
• concepts well identified (scale, support, variance) • includes and complement the traditional statistics • models are consistent • structural analysis precedes the estimation • transitive approach well designed for global estimation based on regular sampling scheme	• variography is often difficult because of extreme data or/and of too irregular sampling designs • human time consumming • multivariate models quickly impractical (except in some particular cases where simplifications occur)

ICES / CIEM
Workshop on Survey Design and Analysis Working Document
21-25 June 2004, Aberdeen

Geostatistics in fisheries survey design and analysis: a history of ideas, 1990-2004

Pierre Petitgas

IFREMER, rue de l'Île d'Yeu, 44311 cedex 3 Nantes, France. tel: +33240374163; Fax: 33240374075; pierre.petitgas@ifremer.fr

Linear geostatistics has solved the question concerning survey design and provided a solution for not randomising sampling locations. The application of linear geostatistics has been confronted to accute problems of extreme values and temporal changes. Three approaches have been developed. Robust techniques have been developed to estimate the linear variogram (data weighting and transformation) with some success. Variation in fish density has also been analysed with multivariate techniques demonstrating the impact of time of day and particular environmental parameters (e.g., bottom shape) depending on the ecosystem. Multivariate techniques have not allowed to decrease significantly the estimation variance, probably because the dimension of the estimation space was mathematically increased and because of high residual variance. The other approach was to analyse the spatial statistical characteristics of the surroundings of high values, using non-linear geostatistics. Areas of high value occurrence have been estimated by disjunctive kriging. This approach has not lead to significant increase in precision or change in the design, probably because the probability of occurrence of high values was high over large areas. Related to this approach is that of post-stratification and heterogeneous sampling. It is felt that much is to be done still in this field. Last, geostatistical simulations have been used to test survey design bias and precision as well as to estimate full scale estimation variance when the abundance estimate is a combination of primary variables.

1. Estimates of abundance and variance for homogeneous survey designs

In the early 1990s, geostatistics was introduced in fisheries surveys as a tool to estimate model-based estimates of variance for population abundance of acoustic surveys. Acoustic surveys were made of parallel regularly spaced transects and this was critisised by statisticians (Jolly, 1990). Because geostatistics estimates the correlation structure in the process values with a variogram model, it provides a methodological solution to derive model-based estimates of variance (Matheron, 1971) whatever the sampling design, provided that sampling is independent of process values (homogeneous design). Software was developed to allow for these computations, e.g. EVA (Petitgas and Lafont, 1997), Splus/R libraries (review in Rivoirard et al., 2000). ICES organised various meetings on the topic (ICES, 1989; ICES, 1992a; ICES, 1992b; ICES, 1993). Rivoirard et al. (2000) compiled theory and demonstrative applications on case studies. Petitgas (2001) provided a review on geostatistical applications and modelling fisheries data.

When the abundance estimate is a combination of primary variables that are sampled, the full estimation variance has been accessed by conditional geostatistical simulations. Each primary variable is simulated over the surveyed domain and simulated fields are combined (Gimona and Fernandes, 2003).

2. Practical difficulties when estimating variograms of survey data

Rivoirard et al. (2000) provide practical examples and Petitgas (2001) a review on the topic. The delineation of the estimation domain may be problematic when there are many diffuse zero values and this will affect the variogram. Bez and Rivoirard (2001) suggest the use of transitive geostatistics. Guiblin et al. (1995) proposed an estimate of the variogram based on the non-centred covariance that is robust to transitions between zero and non-zero values.

Extreme values values tend to destructure the variogram range and make the estimation of the sill imprecise. ICES (1992b) had proposed power transformation of the arithmetic scale. But Rivoirard et al. (2000) propose to estimate the variogram on logtransformed data, deduce the corresponding arithmetic variogram (under the hypothesis of lognormality) and perform the esitmation on the arithmetic scale. Matheron (1971) had suggested to calibrate the sill of the variogram using the fact that the average variogram for all distances in the domain equal the (dispersion) variance.

Survey design may affect the estimate of the variogram. In zig-zag acoustic surveys, the along transect variogram may differ from the planar estimated variogram. Rivoirard et al. (2000) suggest a weighting of the values, so as to "decluster" the sample locations.

The variogram was automated using a criteria of goodness of fit (see, Rivoirard et al. 2000).

3. Dealing with extreme values

Extreme values of fish density per mile square in acoustic surveys are caused by a large dense schools and are the major source of imprecision in surveying. Their probability of occurrence is low but their potential domain of occurrence is big. Research followed two pathways: biological understanding and modelling.

Biological understanding has been devoted mainly to the characterisation of what is a school, how does it aggregate and disaggregate and in which circumstances (Fréon and Misund, 1999, Chap.4 Schooling behaviour). Fish density was known to vary with time of day (Massé, 1989). Occurrence of schools at particular locations was shown to be related with environmental variables (e.g., bottom depth, Reid et al., 2001) in particular ecosystems.

The school occurrence process was separated from that of the school internal density (MacLennan and MacKenzie, 1988; Marchal and Petitgas, 1993). The school occurrence process was clustered and precisely modelled while the school density was highly variable. In order to predict the occurrence of a rich school, the relationship between a rich school and its surrounding schools was studied but there was no obvious rule: Petitgas and Lévénez (1996) found no relationship between a rich school and the characteristics of its surrounding schools; Beare et al (2002) found that cluster of schools with a rich schools contained a smaller number of schools.

The relation between the schooling aggregation and the population biomass was analysed. Schools clusters were identified using a point process approach (ICES, 2000; Petitgas, 2003). The cluster characteristics (e.g., dimension, nb of schools) were correlated with total population school number but not with total population biomass (Petitgas et al. 2001), at least not at intermediate population biomass levels. This meant that the clustering of schools depended for a large part on the local environmental conditions.

None of the characterisation of rich schools was transferred to estimation procedures and survey design.

Non-linear geostatistics (Rivoirard, 1994) allowed for the analysis of the spatial continuity between classes of fish density values, i.e., between areas. It was found that areas that contained rich values (they can be defined by thresholding) were spatially independent from other areas (Petitgas, 1993). This lead to envisage post-stratification of the data based on the spatial structural characteristics (Petitgas, 1997) and potentially opens the door to adaptative sampling methodology.

4. Dealing with large scale drifts

In bottom trawl and ichtyoplankton surveys, variation in fish density with environmental parameters was modelled using GLMs and GAMs (e.g., Stefanson, 1996; Alderstein and Ehrich, 2003; Borchers et al., 1997). Petitgas (2001) reviewed underlying assumptions between GLM and geostatistics: spatial variation is modelled in the drift in GLM while it is modelled as spatial correlation structure in geostatistics. Intrinsic functions of order-k (Chilès and Delfiner, 1999, Chap.4) and mixed effects models (Pinheiro and Bates, 2000) stand in between as they allow the estimation of a correlation structure with that of the drift. The geostatistical extrernal drift method was used to incorporate in the estimation process a functional relationship (either parametric or not) between fish density and its environment (e.g., time of day, Rivoirard and Wieland, 2001; bottom depth, Rivoirard et al., 2000, Petitgas et al., 2003). The multivariate models did not result in major decrease of estimation variance, probably because residual variability stayed high. Also, the consideration of multivariate relationships increases the mathematical dimensionality of the estimation problem which means that reduction in variance will be difficult to obtain with such modelling (discussion at ICES Theme Session K, 2000 Annual Science Conference).

5. Dealing with temporal changes

Variation in fish density with time concerns (i) variation during the survey because of a biological process hapenning at a smaller scale than the survey and (ii) variation between yearly surveys. Variation during the survey is due to mouvements (Rivoirard et al., 2000), aggregation and environment dynamics (review in Petitgas and Williamson, 1997). Impact of particular school mouvements was tested for northern North Sea herring by simulation (Rivoirard et al., 2000): there was little impact on the final population estimate. Variation in time during the survey is in fact a space-time interaction because time cannot be considered as a third dimension independent of space. Rivoirard (in Petitgas and Williamson, 1997) suggested particular variogram models to account for random motion of schools.

Inter-annual variation in variographic structure has been analysed as well as in spatial distribution. Rivoirard et al. (2000) show inter-annual variation in variographic structure on demersal fish. Mixed effect models could be used to estimate yearly variogram parameters in a multi-annual modelling. Consistency in spatial distribution across the years has lead to estimate large scale drifts invariant of time, either through a relationship with bottom depth and the use of external drift methods (Rivoirard et al., 2000; Petitgas et al., 2003) or through the construction of multiplicative geostatistical models (Petitgas, 1997a). Multi-survey geostatistics models have been of interest in two situations: when population abundance (e.g., egg production during the season) requires an integration over space and time; when surveying for particular years lacked consistent sampling of a variable requiring other years to interpolate the lacking information.

6. Heterogeneous / homogeneous sampling schemes.

Much of the biological variability originates from biological processes occurring at different scales. Therefore, survey design could be conceived to sample processes at different scales in order feed a multivariate model that would then be spatially integrated. In contrast, fisheries surveys sample mean values over large areas with evenly located samples over the domain. Scale in the design of surveys and analysis procedures has received some attention. Variance in a small scale intense survey was found as large as that in a large scale survey (IBTS in the North Sea, Petitgas, 2001). Comparing research surveys with commercial catches in the same areas, Petitgas et al. (2003) showed that there was good coherence between the two in the areas where the probability of high catches was low. Should survey designs incorporate over sampling schemes with a rule for intensifying sampling that is guided by some parameter ?

Typically, an adaptative design is made of level 1 samples which are located according to a homogeneous sampling scheme independent of process values, e.g., a systematic design. Then level 2 samples are added in the vincinity of level 1 samples conditionally on the values observed at those level 1 samples. Adaptative sampling suffers the risk of bias in the design. Because additional samples are positioned in the vincinity of rich encountered values, the result suffers the risk of a systematic underestimation. This because lower values are additionally sampled close to rich ones when richer values are not additionally sampled close to low ones. The bias depends on the rule adopted to allocate addtional samples. Intuitively, the bias should be small when the level 1 design is such that the probability to traverse rich patches is high and when the level 2 samples cover the entire patches. Thompson (1992) proposed an adaptative design which prevents bias by covering the entire cluster around the rich value that triggers the addition of samples. Though applied by Lo et al. (1997) Thompson's adaptative rule (1992) seems irrealistic for fisheries surveys (Petitgas, 2001) because the value of one sample is not the value of a block mean. Motos et al. (1997) showed by geostatistical simulation and resampling that adding level 2 samples based on the mean of level 1 samples (not on that particular level 1 value) lowered the bias and increased the precision.

It seems that to make progress on accounting for the formation of high values, heterogeneous sampling could be a way forward. some precautions will be necessary. As in the early 90s, when variographic structure modelling allowed to free data analysis from survey design; non-linear spatial structure modelling has the potential to allow for the estimation of the mean and variance in the case of adaptative sampling design. Petitgas (1997b) provide case study using post-stratification based on a non-linear spatial structure analysis. Geostatistical simulations have the potential to allow for testing bias in the adaptative sampling rule (ex: Motos et al., 1997).

References:

Adlerstein, S. and Ehrich, S. 2003. Patterns in diel variation of cod catches in North Sea bottom trawl surveys. Fisheries Research, 63(2): 169-178.

Beare D., Reid D. & Petitgas P. 2002. Spatio-temporal patterns in herring school abundance and size in the northwest North Sea: modelling space-time dependencies to allow examination of the impact of local school abundance on schools size. *ICES Journal of Marine Science* 59: 469-479.

Bez N. & Rivoirard J. 2001. Transitive geostatistics to characterise spatial aggregations with diffuse limits: an application on mackerel ichtyoplankton. *Fisheries Research* 50: 41-58.

Borchers, D., Buckland, S., Priede, I., Ahlamadi, S. 1997. Improving the precision of the daily egg production method using generalised additive models. Canadian Journal of Fisheries and Aquatic Science, 54: 1205-1233.

Chilès, J.-P. and Delfiner, P. 1999. Geostatistics: modelling spatial uncertainty. John Wiley and Sons, New York, 695p.

Freon, P. and Misund, O. 1999. Dynamics of pelagic fish distribution and behaviour: effects on fisheries and stock assessment. Fishing News Books, Blackwell Science, Oxford.

Gimona A. & Fernandes P. 2003. A conditional simulation of acoustic survey data: advantages and pitfalls. *Aquatic Living Resources* 16: 123-129.

Guiblin, P., Rivoirard, J. and Simmonds, E. 1995. Analyse structurale de données à distribution dissymétrique, exemple du hareng écossais. Cahiers de Géostatistique, Fasc. 5: 137-159. ENSMP, Paris.

ICES 1989. Report of the workshop on spatial statistical techniques. ICES CM 1989/K:38

ICES 1992a. Report of the workshop on the analysis of bottom trawl survey data. ICES CM 1992/D:6

ICES 1992b. Acoustic survey design and analysis procedures: a comprehensive review of current practice. ICES Cooperative Research Report, 187, 127p. Ed. by Simmonds et al.

ICES 1993. Report of the workshop on the applicability of spatial statistical techniques to acoustic survey data. ICES Cooperative Research Report, 195, 87p.

ICES 2000. Echotrace classification. ICES Cooperative Research Report, 238. Ed. by D. Reid.

Jolly G. & Hampton I. 1990. Some problems in the statistical design and analysis of acoustic surveys to assess fish biomass. *Rapp. P.-v. Réun. Cons. Int. Explor. Mer* 189: 415-420.

MacLennan D. & Mackenzie I. 1988. Precision of acoustic fish stock estimates. *Can. J. Fish. Aquat. Sci.* 45: 605-616.

Marchal E. & Petitgas P. 1993. Precision of acoustic fish stock estimates: separating the number of schools from the biomass in the schools. *Aquatic Living Resources*: 211- 219.

Petitgas P. 2003. A method for the identification and characterization of clusters of schools along the transects lines of fisheries-acoustic surveys. *ICES Journal of Marine Science* 60: 872-884.

Petitgas P. 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. *Fish and Fisheries* 2: 231-249.

Petitgas P. (2001) Allocation of survey effort between small scale and large scale and the precision of fisheries survey-based abundance estimates. *ICES CM 2001/P:17*.

Petitgas P. 1997a. Sole egg distributions in space and time characterized by a geostatistical model and its estimation variance. *ICES Journal of Marine Science* 54: 213-225.

Petitgas, P. 1997b. Use of disjunctive kriging to analyse an adaptative survey design for anchovy eggs in Biscay. Ozeanografika, 2: 121-132.

Petitgas P. 1993. Use of disjunctive kriging to model areas of high pelagic fish density in acoustic fisheries surveys. *Aquatic Living Resources* 6: 201-209.

Petitgas P., Massé J., Grellier P. & Beillois P. 2003. Variation in the spatial distribution of fish length: a multi-annual geostatistics approach on anchovy in Biscay, 1983-2002. *ICES CM 2003/Q:15*.

Petitgas P., Poulard J.-C. & Biseau A. 2003. Comparing commercial and research survey catch per unit effort: megrim in the Celtic Sea. *ICES Journal of Marine Science* 60: 66-76.

Petitgas P., Reid D., Carrera P., Iglesias M., Georgakarakos S., Liorzou B. & Massé J. 2001. On the relation between schools, clusters of schools, and abundance in pelagic fish. *ICES Journal of Marine Science* 58: 1150-1160.

Petitgas P. & Lafont T. (1997) EVA2: estimation variance version 2 - a geostatistical software on windows for the precision of fish stock surveys. *ICES CM 1997/Y:22*.

Petitgas, P. and Williamson, N. 1997. Report of the workshop on time variability and space-time interaction in fisheries acoustic surveys. pp. 35-43 *In*: Report of the working group on Fisheries Acoustics Science and Technology. ICES CM 1997/8 5

Petitgas P. & Levenez J. J. 1996. Spatial organisation of pelagic fish: echogram structure, spatio-temporal condition and biomass in Senegalese waters. *ICES Journal of Marine Science* 53: 147-153.

Pinheiro, J. and Bates, D. 2000. Mixed effects models in S and Splus. Springer-Verlag, Berlin.

Lo, N., Griffith, D. and Hunter, J. 1997. Using a restricted adaptative cluster sampling to estimate pacific hake larval abundance. CalCOFI report, 38:103-113.

MacLennan D. & Mackenzie I. 1988. Precision of acoustic fish stock estimates. *Can. J. Fish. Aquat. Sci.* 45: 605-616.

Massé J. (1989) Daytime detected abundance from echosurveys in the Bay of Biscay. *ICES CM 1989/B:24*.

Matheron, G. 1971. The theory of regionalised variables and its applications. Les cahiers du centre de morphologie mathématique de Fontainebleau, Fasc.5. ENSMP, Centre de Géostatistique, Fontainebleau. 212p.

Motos, L., Petitgas, P. and Truong, B. 1997. Statistical analysis of the daily egg production biomass equation. Final report to the European Commission DG-Fish, project n° 95/009.

Reid D. & Maravelias C. 2001. Relationships between herring school distribution and seabed substrate derived by Roxan. *ICES Journal of Marine Science* 58: 1161-1173.

Rivoirard, J. 1994. Introduction to disjunctive kriging and non-linear geostatistics. Clarendon, oxford.

Rivoirard J. & Wieland K. 2001. Correcting for the effect of daylight in abundance estimation of juvenile haddock (Melanogrammus aeglefinus) in the North sea: an application of kriging with external drift. *ICES Journal of Marine Science* 58: 1272-1285.

Rivoirard J., Simmonds J., Foote K., Fernandes P. & Bez N. (2000) *Geostatistics for estimating fish abundance*. Blackwell Science Ltd.

Stefanson, G. 1996. Analysis of groundfish survey abundance data: combining GLM and delta approaches. ICES Journal of Marine Science, 53: 577-596.

Thompson, S. 1992. Sampling. John Wiley and Sons, New York.

ICES / CIEM
Workshop on Survey Design and Analysis Working Document
21-25 June 2004, Aberdeen

About non-linear geostatistics and adaptative sampling

Pierre Petitgas

IFREMER, rue de l'Île d'Yeu, 44311 cedex 3 Nantes, France. tel: +33240374163; Fax: 33240374075; pierre.petitgas@ifremer.fr

Introduction

Homogeneous survey designs make scientists spend their time sampling zero values when fish are aggregated in particular areas. Multivariate methods as well as process knowledge on biological aggregation did not allow to reduced estimation variance significantly. The high variance in survey data of homogeneous designs lies in the inability of such design to measure variability at different scales and it is thought that significant progress will be made if we change survey design. Adaptative designs are then of particular interest.

Geostatistics models the spatial structure and uses that structure in the estimation procedure. The result is model-based estimates. Their advantage is to separate data analysis from survey design.

Non-linear geostatistics models the spatial correlation structure between paires of cutoffs, conditionally to one of them. The conditioning idea is helpful in the analysis of adaptative adaptative sampling because adaptative sampling is based on adding extra samples conditionally to already sampled ones.

This document illustrates a case study on surveying anchovy eggs in Biscay and provides an example for deriving estimates of mean and variance from an adaptative design. It is based on Rivoirard (1989, 1994), Petitgas (1993, 1997), Motos et al. (1997).

The case study

The adaptative design of the case study is documented in Petitgas (1997) (Fig. 1). Level 1 samples follow a systematic design, independent of process values. Then level 2 samples are added in the vincinity of level 1 samples conditionally on the values observed at those level 1 samples. In contrast to other adaptative rules (e.g., Thompson, 1992) which add samples conditionally to the value of a particular point, the rule used here was to add level 2 samples conditionally on the mean of several level 1 samples.

First, data are rectified on a regular grid: each node grid is attributed the value of the nearest sample. variographic structure is modelled from the rectified data.

Bias in the adaptative design is tested by geostatistical simulations and resampling the simulated process with different survey designs (Motos et al., 1997). The egg process is geostatistically simulated generating 100 fields. Each field is sampled using the adaptative rule. The fields are also sampled with a systematic design that has a similar nb of points than the adaptative design. Bias is estimated by the difference between the sampled mean and the process mean. The distribution of the bias is provided by the 100 repetitions. The result was that the adaptative rule generated little bias and allowed for a reduction in variance in comparison to the systematic design.

Then, a non-linear spatial structure is modelled in Petitgas (1997) using an approach developed in Rivoirard (1989, 1994). The structure is such that rich areas can be defined that are spatially independent of their surroundings, allowing then for post-stratifying the data and estimating mean and variance. Computations were done using the software EVA (Petitgas and Lafont, 1997) which has a option for heterogeneous sampling.

Conclusion

The example shows that with simulations to estimate bias and with a non-linear spatial model to estimate mean and variance, it is possible to use adaptative designs.

References:

Petitgas P. 1993. Use of disjunctive kriging to model areas of high pelagic fish density in acoustic fisheries surveys. *Aquatic Living Resources* 6: 201-209.

Petitgas, P. 1997. Use of disjunctive kriging to analyse an adaptative survey design for anchovy eggs in Biscay. Ozeanografika, 2: 121-132.

Petitgas P. & Lafont T. (1997) EVA2: estimation variance version 2 - a geostatistical software on windows for the precision of fish stock surveys. *ICES CM 1997/Y:22*.

Motos, L., Petitgas, P. and Truong, B. 1997. Statistical analysis of the daily egg production biomass equation. Final report to the European Commission DG-Fish, project n° 95/009.

Rivoirard, J. 1989. Models with orthogonal indicator residuals. pp.91-108, *In*: Geostatistics, M. Armstrong (Ed.). Kluwer Academic Publishers, Dordrecht.

Rivoirard, J. 1994. Introduction to disjunctive kriging and non-linear geostatistics. Clarendon, oxford.

Thompson, S. 1992. Sampling. John Wiley and Sons, New York.

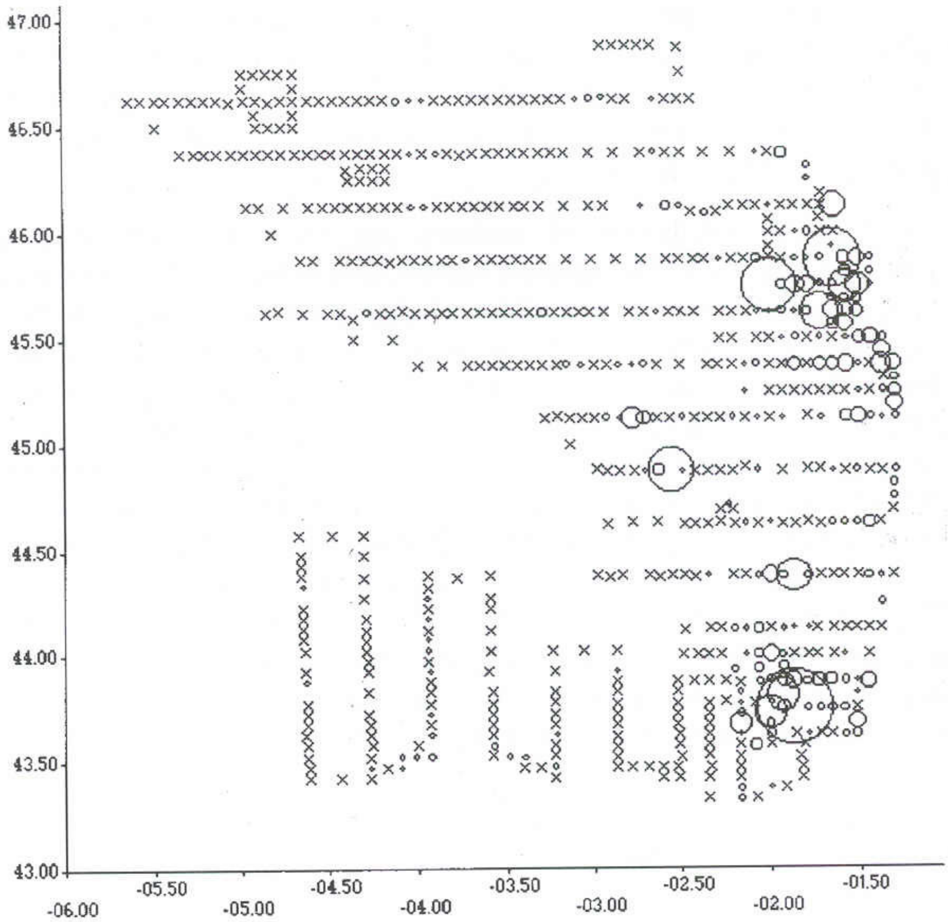


Figure 1: The 1992 adaptive egg survey with proportional representation of the day-1 egg density (count per 0.05 square meter of sea surface). Circle radius is linearly proportional to he egg density.

Simulation-test of the adaptative rule [Motos et al. (1997)]

2.5. *Geostatistical simulations of an adaptive design*

A simulation was performed. Geostatistical simulations were performed using the turning bands methods. We used the version enabling a great number of bands. The simulated variable is Gaussian. We choose a 100x101 field. The variogram range is 20. The Gaussian simulated values were exponentiated to generate a skewer distribution than a Gaussian. The mean of the logs was set to unity and the CV to 1.5. We generated 100 simulations. Each simulation was sampled with an adaptive design and a regular design. The regular design is made of parallel regularly spaced transects, the inter-transect distance beeing 20 (ie variogram range). There are 5 transects. We considered 2 sampling efforts, 100 points (mesh along transect of 5) and 170 points (mesh along transect of 3).

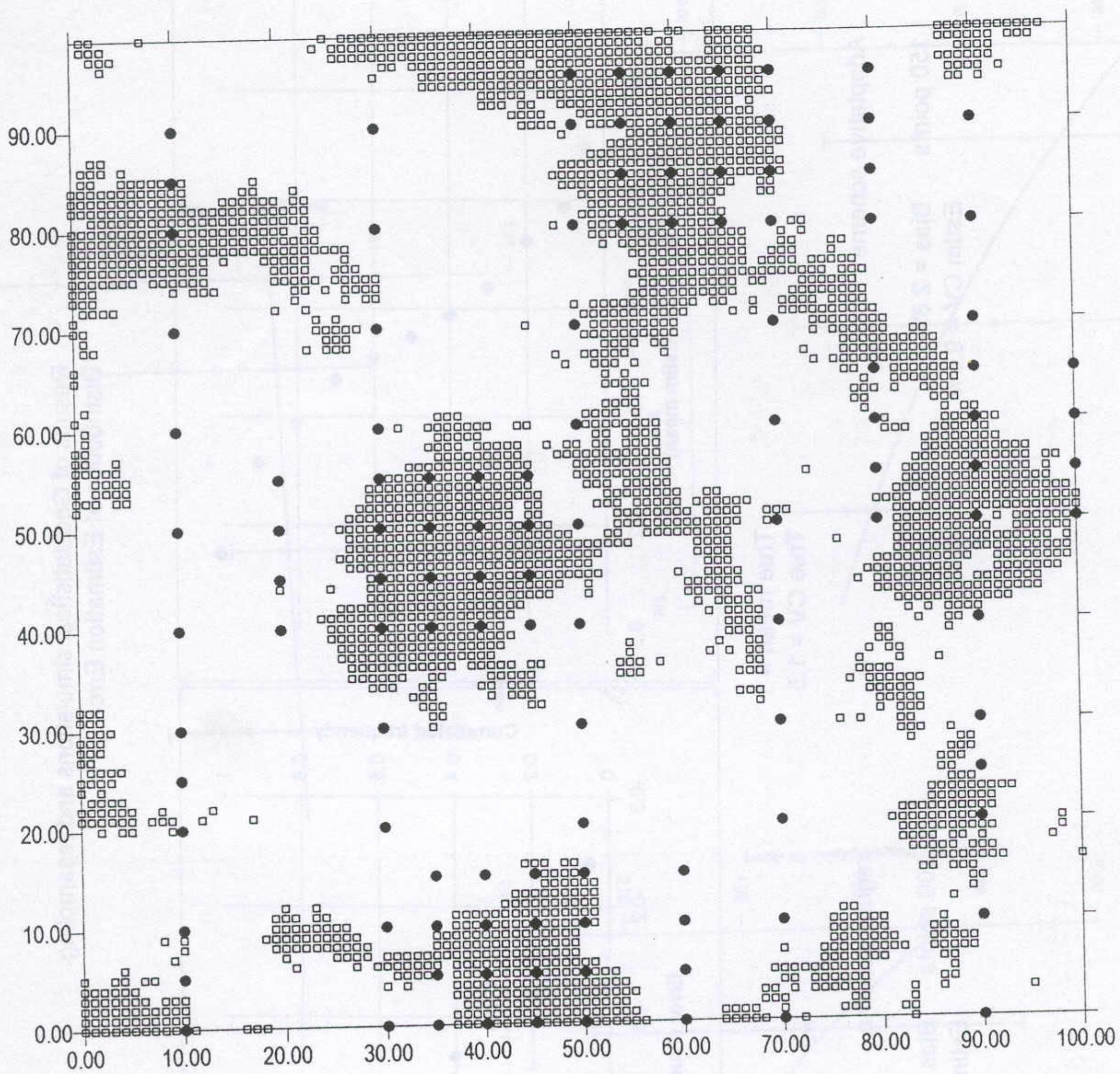
The adaptive design is based on the regular grid of transects with mesh 5 along the transects. The adaptive rule is highly inspired from the rule of AZTI in the 1992 survey. The sampling mesh along the transects is widened when low values are encountered. When a rich area is encountered as seen from a certain number of transect values (level 1), a small transect is added (level 2 transect) on both sides. If values stay rich on the level 2 transect, a transect of level 3 is added between transect of level 1 and the small transect of level 2.

The simulation is calibrated. The threshold used to define low or rich was 1. This was so because the values greater than unity represent 30% of the field. 30% of the surveyed area is made of the two major spawning grounds. The inter-transect distance was 20, e.g. the variogram range. The re-sampling rule is defined as follows: along the transect the mesh is 5. When at a point x, the value is lower than 1, the next sample is performed at a distance of 10. Along the transect, a moving average is computed on 3 points. If the average is greater than the threshold, a transect of level 2 is added on both sides at a distance of 10. Then, on the transect of level 2, moving averages of 3 points are computed. If they are greater than the threshold, a transect of level 3 is added on both sides at a distance of 5.

Fig. 16 shows one simulated image and its adaptive sampling. Fig. 17 shows the cumulative distributions of the error for the adaptive scheme and a regular scheme with a sampling mesh of 5 along the transects. Weighting the sample values by their area of influence produces the mean estimate. We compare results in the following text table:

	Average number of sampling points	Bias	Estim CV
Regular 1	100	-1%	13.8%
Regular 2	170	-0.8%	11.8%
Adaptive	150	2.8%	9.5%

In comparison to the regular schemes, the adaptive survey has a little bias (underestimation) but is more precise. No general rule can be given, a different adaptive sampling rule may have different results. The present rule adds more stations in large patches of values greater than the threshold and this is why we think it behaves well. Smaller patches are crossed by the regular part of the transects. This justifies to stratify the data for the estimation purpose as we have done. Note that we are in the case where inter-transect distance equals the variogram range.

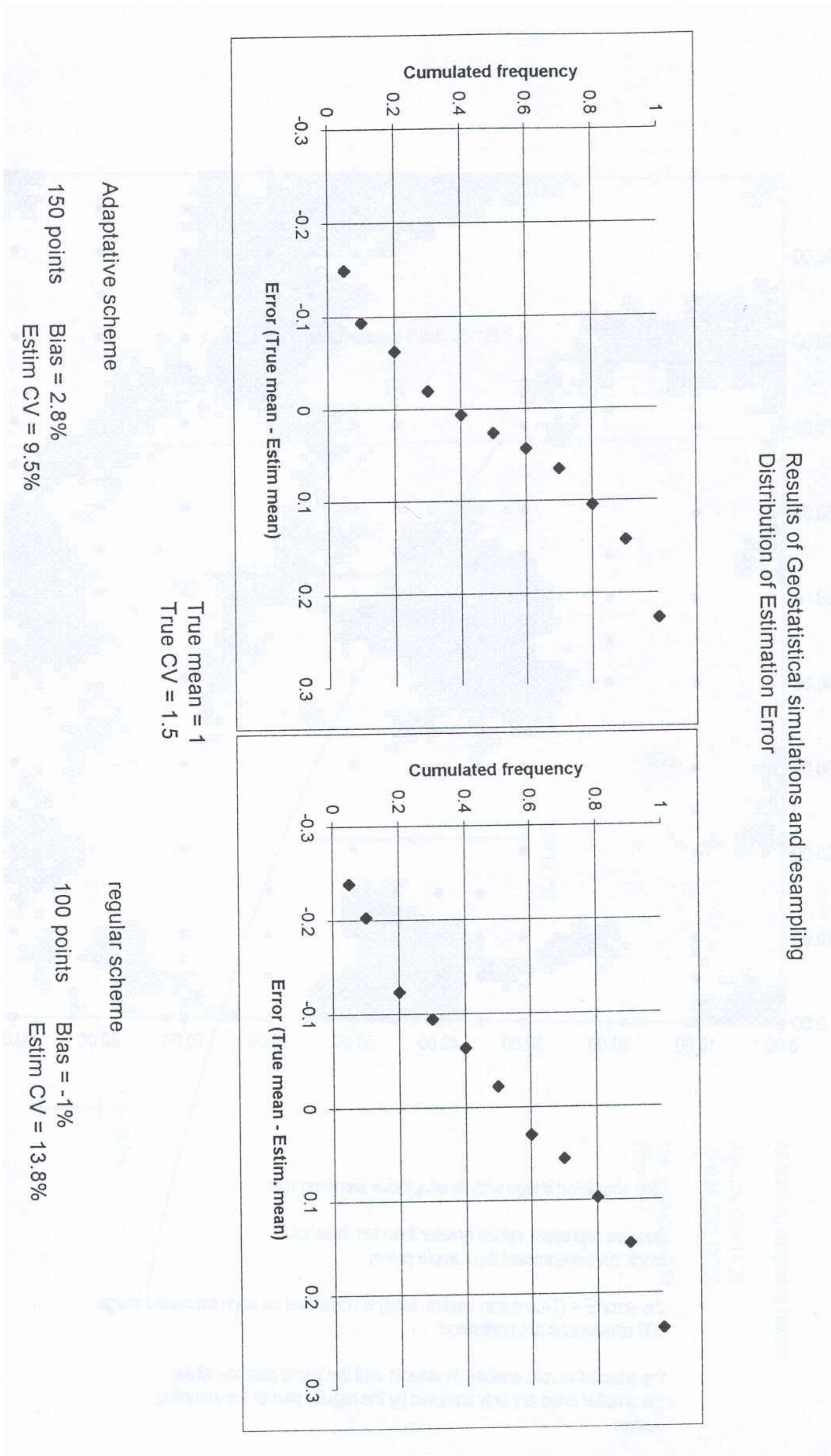


One simulated image with its adaptative sampling rule

squares represent values greater than the threshold 1
black circlesrepresent the sample points

the error E = (True mean - estim. Mea) is computed on each simulated image
100 simulations are performed

the adaptative rule enables to smaple well the larger patches while
the smaller ones are only sampled by the regular part of the sampling
design


Results of Geostatistical simulations and resampling
Distribution of Estimation Error
Cumulated frequency
Error (True mean - Estim mean)
Adaptative scheme
150 points
Bias = 2.8%
Estim CV = 9.5%
True mean = 1
True CV = 1.5
regular scheme
100 points
Bias = -1%
Estim CV = 13.8%