ICES WKSAD Report 2005

ICES Fisheries Technology Committee
ICES CM 2005/B:07
Ref. D, G

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

9–13 May 2005

Sète, France

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

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

Recommended format for purposes of citation:
ICES. 2005. Report of the Workshop on Survey Design and Data Analysis (WKSAD), 9–13 May 2005, Sète, France. ICES CM 2005/B:07. 170 pp. https://doi.org/10.17895/ices.pub.9631

Contents

ICES WKSAD 2005 Executive summary

TERMS OF REFERENCE. The Workshop on Survey Design and Analysis [WKSAD] met in Sète, France, from 9–13 May 2005 to: a) evaluate alternate analyses of surveys of a simulated fish population and several real survey datasets; b) 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; c) evaluate analyses of covariate data which could provide improved precision of abundance estimates; d) review methods for combining surveys of the same resource using different methods; e) evaluate the sensitivity of methods to estimate biological parameters in terms of analytical assumptions and measurement error.

SIMULATED SURVEYS. A simulation exercise was conducted whereby a variety of trawl survey designs and design types were applied to two simulated fields of fish density. As expected this exercise demonstrated the advantage of using more systematic designs in the presence of more autocorrelation. However, the exercise also showed how random surveys can perform better when combined with route optimisation algorithms which, in a fixed time, allow for more trawl samples to be taken than a systematic design; the latter only occurs when the autocorrelation is low.

SURVEY DECISION TREE. As a result of the simulations and subsequent discussions a decision tree was proposed with the objective of providing advice on the best survey design to implement given the objective of deriving a precise estimate of the abundance of a marine resource. Generally, the decisions are aided by knowledge of the spatial distribution of the fish: the more autocorrelation there is in the distribution, the greater the advantage of introducing some form of regular spacing to the survey design.

TOW DURATION. In many cases, distinct advantages can be gained from reducing the duration of a trawl tow. These include: an increase in survey precision; less wear on gear; less sorting time, providing more time to take other biological measurements. Such advantages may be specific to certain conditions so the possibility of reducing the tow duration should be examined by conducting experiments such as those described in this report (Section 3.2). If and when it can be demonstrated that reducing tow duration increases survey precision, then that reduced tow duration should be employed.

USE OF COVARIATES. Covariate information can be used to improve both survey design and analysis, as well as provide useful information on possible causes of inter-annual variation in mean abundance and other parameters. An example was described where survey design and wind conditions explained about half the interannual variation in survey density indices.

COMBINING SURVEYS. Where the relationship between acoustic data and trawl catch data is strong, the between-station acoustic data can be used to extrapolate fish abundance and improve the overall index of bottom trawl surveys. Independently derived indices can be combined according to a weighting scheme derived directly from the observed sampling variability in the indices: an example is given of a (herring) stock assessment model which uses this.

BIOLOGICAL SAMPLING. The effective sample size to determine biological parameters such as a length distribution can be much smaller than the number of samples taken. This has implications for the efficiency of the sampling process and should be examined more widely. Further development of coherent mapping of biological parameters would be desirable.

1 Introduction

1.1 Terms of reference

According to C.Res. 2004/2B07 the **Workshop on Survey Design and Data Analysis** [WKSAD] (Co-chairs: P.G. Fernandes, UK, and M. Pennington, Norway) met in Sète, France, from 9–13 May 2005 to:

a) evaluate alternate analyses of estimates of the abundance, associated variance, and density maps, from surveys of a simulated fish population whose abundance is known and then expand this to several actual survey datasets;

b) 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;

c) evaluate analyses of covariate data which could provide improved precision of abundance estimates;

d) review methods for combining surveys of the same resource using different methods;

e) evaluate the sensitivity of methods to estimate biological parameters in terms of analytical assumptions and measurement error.

WKSAD will make its report available by 20 June 2005 for the attention of the Fisheries Technology Committee, the Living Resources Committee and Resource Management Committee.

1.2 Participants

Jean Adams	U.S.A.	
Nicola Bez	France	
Robert Brown	UK, England	
Noel Cadigan	Canada	
Ian Doonan	Ireland	
Abdelmalek Faraj	Morroco	
Paul Fernandes	UK, Scotland	(Co-chair)
Joakim Hjelm	Sweden	
Leire Ibaibarriaga	Spain	
Johan Lövgren	Sweden	
Jean Claude Mahe	France	
Michael Pennington	Norway	(Co-chair)
Jacques Rivoirard	France	
John Simmonds	UK, Scotland	
Konstantin Sokolov	Russia	
Arnauld Souplet	France	
David Stokes	Ireland	
Verena Trenkel	France	
Paul Walline	USA	
Kai Wieland	Greenland	
Mathieu Woillez	France	

Participants' affiliations and e-mail addresses are given in Annex 1.

1.3 Structure of the report

The Terms of Reference (ToR) 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; and (e) in Section 6. Recommendations are given in Section 7 and a bibliography is given in Section 8. Eight working documents were presented to the meeting: these are listed in Annex 2 and the documents are appended. A short piece of code to carry out abundance estimates based on the delta distribution is attached as Annex 3.

Section 2 examines a number of different approaches to designing and analysing surveys. This comprises three main sub sections. Firstly, results from 19 trawl surveys of two simulated fish populations are compared; a variety of analysis methods are applied to these data. This is accompanied with a study in which 2000 simulated surveys were conducted to compare the precision of a random design with a systematic one. In the second part of this section, a few examples of designs and analyses of various different types of actual surveys are described. This includes a brief review of the MEDITS programme which has adopted an ecosystem approach by delivering a series of indicators based on surveys in the Mediterranean. In a conclusion to this section, a decision tree is proposed which provides guidance on which survey design approach might used in order to obtain the most precise estimate of the abundance of marine resources.

Section 3 examines the issue of tow duration and reviews a number of studies which have indicated that taking shorter tows can confer a number of advantages compared to longer tows. An example of how to go about determining whether this is the case for a particular survey is then given in reference to the west Greenland survey for shrimps (*Pandalus borealis*) and halibut (*Reinhardtius hippoglossoides*).

Section 4 considers the use of covariates. This section gives an overview of the discussion and examines a case study which considered the use covariates to describe possible causes of inter-annual variation in mean abundance from a survey. Section 5 deals with combining survey indices. Although the anticipated review of this subject was not submitted, two relevant contributions are described. In one case the combination of two of the most common survey methods is described – that of combining trawl and acoustic survey indices – based on a three year research project. In another case the methods used in an assessment model to combine four survey indices are described.

Section 6 examines the issue of biological sampling. The first part of this section examines the effective sample size to determine biological parameters such as length and age. A number of studies are reviewed which suggest that the number of biological samples taken could be drastically reduced without a significant loss in precision. The second part looks at the spatial mapping of fish length in acoustic surveys.

2 Comparative analyses of survey data

There are a number of different ways of analysing survey data. Many of these are conditional on the type of design applied and/or on the assumptions behind the particular analysis method. Generally, the design type and analysis method are linked. Design-based methods of analysis require few assumptions at the analytical stage, but require that the samples are located at random within the interpolated field (area, strata or block). Model-based methods allow for a more flexible allocation of samples, such as systematic or regular designs, but have more assumptions when estimating variance. Although meeting the latter assumptions is often considered a hindrance, a systematic design provides a more precise estimate in the presence of moderate to high local positive autocorrelation (see ICES, 2004, Section 5.1.5).

A number of analyses are presented below, covering a range of different survey designs. In the first case a simulation exercise is described where participants were invited to survey two two-dimensional fields of fish density of known properties (but unknown to them). This is followed by a comprehensive test of two survey designs on the same data. In the second part, a number of miscellaneous survey analyses are presented.

2.1 Simulation exercise

2.1.1 Methods

Six participants took part in a limited survey simulation exercise. This was intended to provide a greater shared understanding of analytical methods and an appreciation of the effects of deviations from certain assumptions of the methods.

Two virtual fields were generated to base the simulation exercise on two contrasted complete known realities (Figure 1). Amongst the available geostatistical simulations techniques, the Turning Band method (e.g. Lantuéjoul, 2002) was used. This allows simulating a random function whose spatial structure, defined by the variogram, is predefined. Simulations techniques allow generating as many surrogates as necessary. Here, only one simulation was performed for each of the two targeted situations. Simulations usually generate Gaussian statistical distributions. In order to get statistical distributions more like those expected for fish density (approximately log normal shape), the output of the simulations were transformed as follows:

$$z \rightarrow e^{0.2 \cdot z}$$

This transformation is known to modify the spatial structure and the initial parameters of the variograms were chosen so that after the transformation, two fields were generated with the following characteristics:

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

To account for a gradual reduction in abundance from the heart of the distribution to the borders, the outputs were finally multiplied by a bell shaped curves. Both fields were square areas of 120 by 120 n.mi., and were discretised into points representing potential trawl sampling units of 0.25 n.mi.^{2} (57 600 points). They contained an unknown proportion of structural zeros, representing areas where fish do not occur beyond a certain boundary. These latter points where generated by addressing a zero value to all the points below a given threshold.

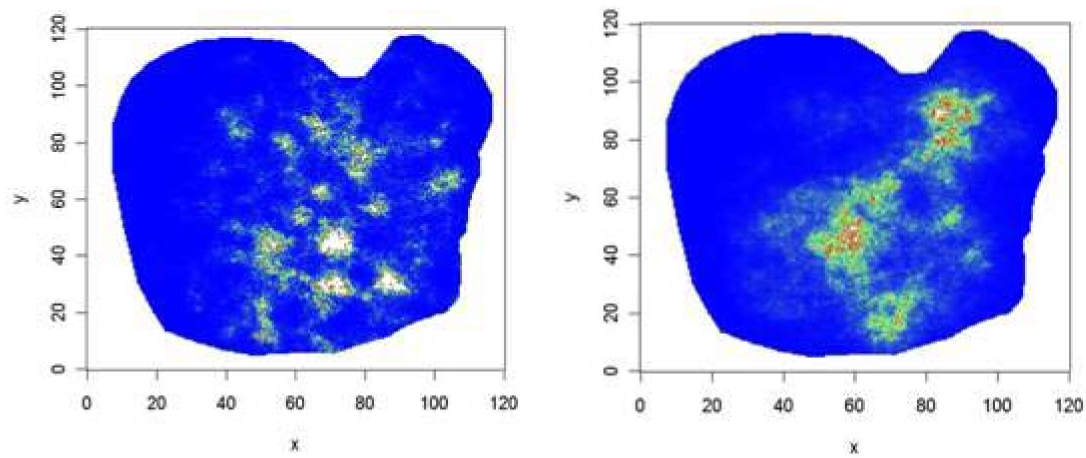


Figure 1. Two dimensional representation of the simulated fish density fields. The density scale goes from blue to green to red to white with increasing density.

Their respective characteristics of the two fields in terms of variograms and statistical distributions were as follows (see also Figure 2):

Field 1:

- Coefficient of variation = 3.3
- Mean fish density in the field of presence = 4 10^7 ind n.mi.$^{-2}$
- Total abundance = 10^7 ind
- Variogram = nugget effect (sill = 2.5 10^6 ind^2 n.mi.$^{-4}$) + spherical (sill = 8.3 10^6 ind^2 n.mi.$^{-4}$; range = 10 n.mi.); the nugget effect represents 23% of the total variance.

Field 2:

- Coefficient of variation = 1.7
- Mean fish density in the field of presence = 4 10^7 ind n.mi.$^{-2}$
- Total abundance = 10^7 ind
- Variogram = nugget effect (sill =0.23 10^6 ind^2 n.mi.$^{-4}$) + spherical (sill = 2.25 10^6 ind^2 n.mi.$^{-4}$; range =25 n.mi.); the nugget effect represents 9% of the total variance.

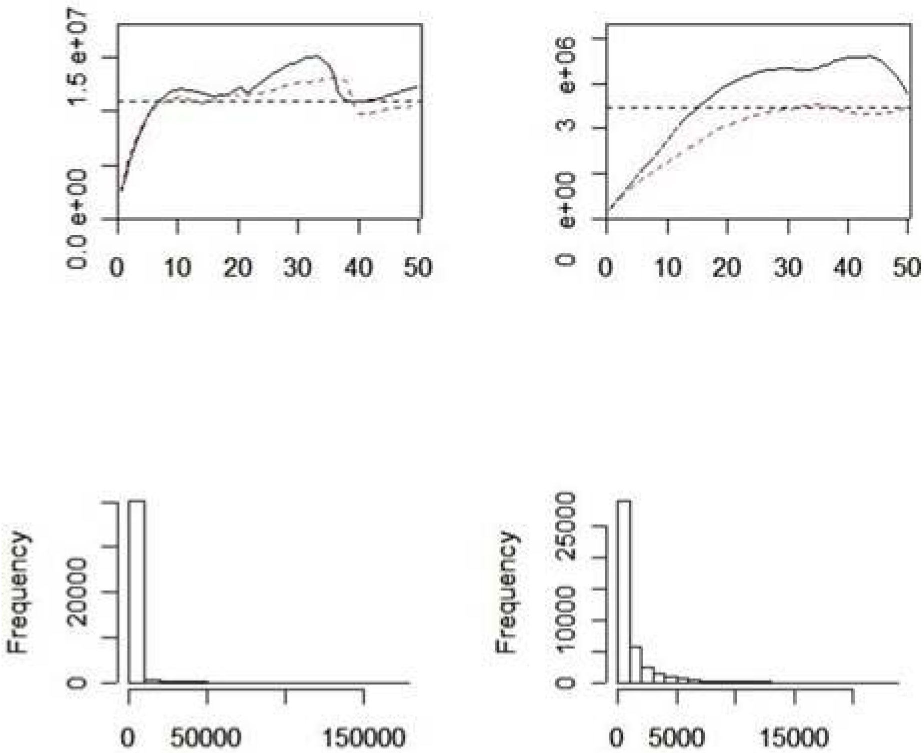


Figure 2. Variograms (upper panels) and histograms (lower panels) of the simulated Field 1 (left) and Field 2 (right). Variograms were computed in two spatial directions (0° straight lines and 90° dashed lines).

The following rules were applied:

1) The fields were generated using geostatistical techniques (Lantuéjoul, 2002) by a simulator at the Centre de Géostatistique, France.

2) The properties of the population (abundance and distribution) were unknown to all participants, until the meeting.

3) Participants were given the opportunity to locate samples in each field using a survey design of their choice. Participants could choose up to 3 designs (i.e., 3 surveys) for each field, but must have submitted their designs at the same time (i.e., no designs were submitted after an analysis of a previously submitted design).

4) The assumed sampling tool was a bottom trawl, delivering fish densities in number per square nautical mile.

5) Each survey must have been completed in 9 whole days (216 hours).

6) Each survey must have started and ended at the origin, coordinates (0, 0).

7) Travel speed during the survey was not to exceed 10 knots at any time.

8) Each 0.25-nm^2 pixel took 0.5 hours to sample. The sampling point was defined as the midpoint of any pixel(s) sampled. The cruise track was to proceed from the midpoint of each sampling point, such that there was no travel through the pixel(s) being sampled, just the relevant time penalty for each sampled pixel(s), 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 was more than one pixel to be taken for a sample, the simulator would decide which pixels were contained in the sample based on the sample midpoint location.

9) Any sample design and any sample size could have been chosen, as long as the survey was completed and the vessel was returned to port within the 9 days.

10) 1The 9 days was 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, have managed a few more samples; or a different configuration might yield fewer but longer (2 hours for 2 pixels) samples[1].

11) Submissions were to consist of:

a. Survey designs as sets of coordinates (x, y) in nm of the midpoints of sample locations (trawl stations).

b. For each sample, the trawl duration (number of pixels).

c. The total time (travel time + sampling time < 216 hours).

12) Specific outputs required:

a. Global abundance expressed as the total number of fish.

b. An estimate of the precision of the abundance estimate.

c. A map of the fish distribution.

d. The cruise track length.

e. Some interpretation of the results.

2.1.2 Results

Nineteen survey designs were submitted, including eight systematic designs, seven stratified random designs, and four other designs (Table 1). The systematic designs used either a random or a centred starting point, and were oriented along a square grid or linear transects. One systematic design, along linear transects, also allowed for two additional adaptive samples to be taken surrounding the sample along each transect with the highest catch. The stratified random samples contained either one or two samples per stratum, and each sample covered either one, two, or three pixels (corresponding to different trawl durations). Other designs included a simple random sample from the entire sample space, a random sample of points along a cruise track defined by a systematic sampling design, clusters of three samples separated by two nautical miles around randomly selected points, and a combination of a systematic design and a stratified random design with half of the samples taken from each design. In some cases, additional random samples were added to the design to use up remaining time left in the survey.

---

1 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.

Nine different combinations of estimators of the mean and variance were used to estimate the total abundance of fish over the entire sample space and the associated estimation variance (Table 1). These included the sample mean and variance; the stratified mean and variance (using different stratifications, Cochran, 1977); the cluster mean and variance (Cochran, 1977); a spline-smoothed mean from an additive model (Hastie and Tibshirani, 1990) with bootstrapped variance; a kriging-smoothed mean and variance (based on an intrinsic geostatistical variogram, Rivoirard *et al.*, 2000); a transitive geostatistical mean and variance (Rivoirard *et al.*, 2000); a geostatistical conditional mean and variance using Gaussian simulation (modified from Gimona and Fernandes, 2003); the sample mean with variance from intrinsic geostatistical variogram (Rivoirard *et al.,* 2000); and the sample mean with variance of dispersion of a point in a block (Rivoirard *et al.*, 2000).

In order to characterize and compare all of the estimates in a general sense, measures of the accuracy and precision of the estimates were defined as follows. The accuracy of each estimate was defined as the difference between the estimated ($\hat{T}$) and true (T) total abundance,

$$Accuracy = \hat{T} - T .$$

The precision of each estimate was defined as the root mean squared difference between the estimated total abundance plus or minus the standard error and the true total abundance,

$$Precision = \sqrt{\frac{(\hat{T} - \hat{s}_T - T)^2 + (\hat{T} + \hat{s}_T - T)^2}{2}} = \sqrt{(\hat{T} - T)^2 + \hat{s}_T^2} ,$$

where $\hat{s}_T$ is the standard error of the estimate $\hat{T}$. Estimates for Field 2 tended to be more accurate (accuracy closer to 0) and more precise (precision closer to zero) than estimates for Field 1 (Figure 3). The difference in precision was expected, because the coefficient of variation (CV) of Field 1 was 3.3 and the CV of Field 2 was 1.6. Estimates based on shorter surveys (total duration < 180 hours) tended to be less precise than those based on longer surveys (Figure 3).

In 13 out of 76 cases (17%), the estimated total abundance was:

$$\hat{T} - T > 1.96\hat{s}_T .$$

Thus, 95% confidence intervals (based on the assumption of approximate normality) did not contain the true total abundance in 17% of the simulations. This is a significantly higher proportion than the 5% expected if the estimates were approximately normally distributed. Six of these cases were generated from two survey designs that used systematic sampling centred along transects (both with and without adaptive sampling) applied to Field 2. This outcome highlights the importance of more "even" spatial distribution of samples, especially in the presence of high autocorrelation. Widely spaced transects may miss relatively large regions of high density, resulting in underestimation of both the total abundance and the variance of the estimate. Eliminating these two surveys from the collection of simulations, left 7 out of 60 cases (12%) with confidence intervals not containing the true total abundance. This proportion is not significantly different from the 5% expected, and could have occurred by chance alone.

Example analyses of the simulation exercise were submitted as working documents WD2 and WD3 attached in Annex 2.

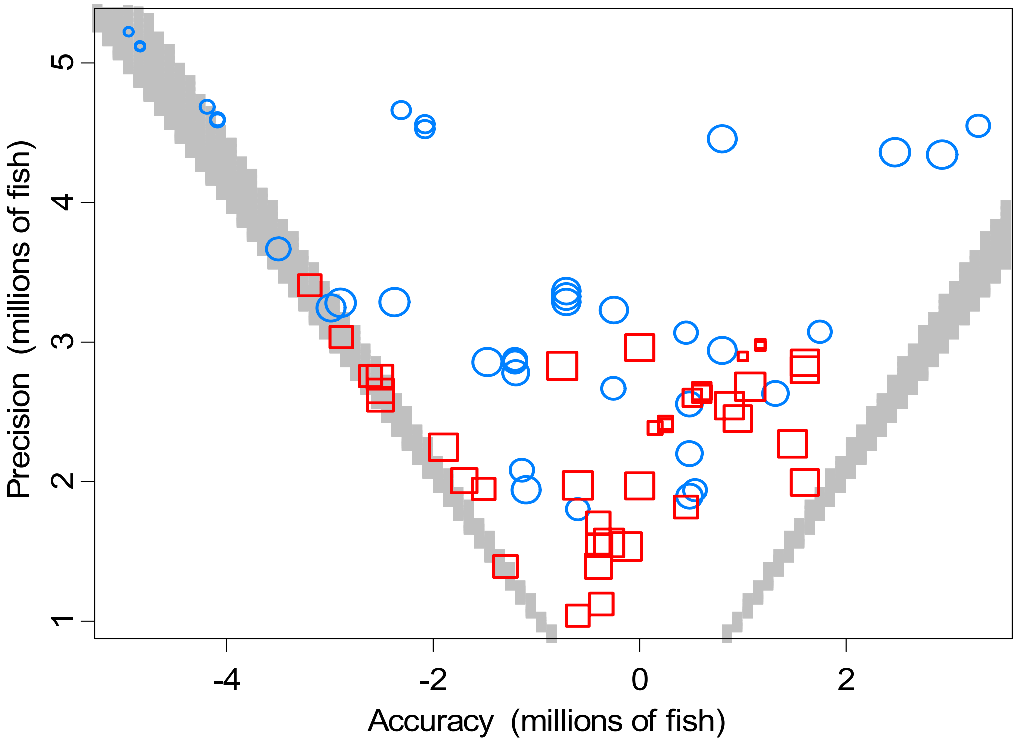


Figure 3. Accuracy and precision of estimates from the simulation exercise (see text for definitions of accuracy and precision). Symbol shape identifies the simulated field (circles for Field 1, squares for Field 2); symbol size corresponds to length of survey (hours at sea). Shading is used to highlight the region where estimates were more than 1.96 times the standard error from the true abundance (10 million fish). One extreme point (accuracy 28, precision 14, in millions) is excluded from this plot (corresponding to the first row of Table A).

Table 1. Summary of results from the simulation exercise, including survey design, time to complete survey (hours at sea), number of tows (n), estimator type, and the estimated fish abundance (total in millions of fish) with associated standard error (SE) and relative standard error (RSE = 100%*SE/Total) for two simulated fields. Each sample covered one pixel, unless otherwise specified. Estimates which were more than 1.96 times the standard error from the true abundance (10 million fish) are shaded.

				Field 1			Field 2		
Design	Hours	n	Estimator	Total	SE	RSE	Total	SE	RSE
systematic random start	211	64	geostatistical intrinsic with variogram and kriging	37.9	26.5	70	10.9	2.4	22
systematic random start	193	64	geostatistical conditional Gaussian simulation	**6.5**	**1.1**	**17**	**8.7**	**0.5**	**6**
systematic random start	~ 216	72	geostatistical transitive	12.9	3.2	25	9.9	1.5	16
systematic centered	215	78	sample mean and variance	7.6	2.3	30	9.4	1.9	20
systematic centered	202	64	geostatistical intrinsic with variogram and kriging	8.8	2.5	29	9.6	1.3	14
systematic centered	192	64	stratified mean and variance (16 strata)	8.8	2.6	30	9.6	1.7	17
			stratified mean and variance (32 strata)	8.8	2.6	29	9.6	1.5	15
			geostatistical intrinsic with variogram and kriging	8.9	1.7	20	9.6	1.1	11
			spline with bootstrapped variance	9.7	2.7	27	10.4	1.8	17
systematic centered along transects	201	96	stratified mean and variance (16 strata)	10.5	2.1	20	**7.5**	**0.8**	**11**
			stratified mean and variance (48 strata)	10.5	1.8	17	**7.5**	**0.6**	**9**
			geostatistical intrinsic with variogram and kriging	10.5	2.5	24	**7.5**	**1.1**	**15**
			spline with bootstrapped variance	11.3	2.3	20	8.3	1.1	13
systematic centered along transects + adaptive	190	68	stratified mean and variance (16 strata)	11.7	2.5	22	**7.4**	**0.9**	**12**
			stratified mean and variance (20 strata)	10.5	1.9	18	**7.1**	**0.9**	**13**
			geostatistical intrinsic with variogram and kriging	10.4	3.0	29	**6.8**	**1.2**	**17**
			spline with bootstrapped variance	13.3	3.2	24	8.5	1.2	14
stratified random (1 sample per stratum)	137	36	geostatistical intrinsic with variogram and kriging	**5.1**	**1.7**	**33**	11.0	2.7	25
			sample mean with geostatistical intrinsic variogram	**5.2**	**1.7**	**33**	11.2	2.8	25
			sample mean with variance of dispersion of a point in a block	**5.2**	**1.7**	**32**	11.2	2.7	24
stratified random (1 sample per stratum, 2 pixels)	155	36	geostatistical intrinsic with variogram and kriging	**5.8**	**2.1**	**36**	10.2	2.4	23
			sample mean with geostatistical intrinsic variogram	5.9	2.1	36	10.3	2.4	24
			sample mean with variance of dispersion of a point in a block	**5.9**	**2.1**	**36**	10.3	2.4	23
stratified random (1 sample per stratum, 3 pixels)	173	36	geostatistical intrinsic with variogram and kriging	7.7	4.1	53	10.5	2.6	24
			sample mean with geostatistical intrinsic variogram	7.9	4.1	51	10.6	2.6	24
			sample mean with variance of dispersion of a point in a block	7.9	4.0	51	10.6	2.6	24
stratified random (2 samples per stratum)	208	72	stratified mean and variance (36 strata)	9.3	3.3	35	11.6	1.2	10
			geostatistical intrinsic with variogram and kriging	9.8	3.2	33	11.0	2.3	21
			sample mean with geostatistical intrinsic variogram	9.3	3.3	35	11.6	2.4	20
			sample mean with variance of dispersion of a point in a block	9.3	3.2	35	11.6	2.3	20
stratified random (1 sample per stratum)	189	64	geostatistical conditional Gaussian simulation	9.4	1.7	18	9.4	0.9	9
stratified random (1 sample per stratum)	~ 216	64	geostatistical transitive	12.5	3.6	29	9.7	1.5	16
stratified random (1 sample per stratum, 2 pixels)	~ 216	47	geostatistical transitive	7.1	1.5	22	11.1	2.5	22
Random	211	80	geostatistical conditional Gaussian simulation	8.9	1.6	18	8.1	1.2	15
random along path	215	82	sample mean and variance	8.5	2.4	29	9.3	2.7	30
clusters	210	78	sample mean and variance	10.8	2.8	26	10.0	2.0	20
			cluster mean and variance	10.8	4.4	41	10.0	3.0	30
half stratified random, half systematic centered	~ 211	64	geostatistical intrinsic with variogram and kriging	7.0	1.3	18	11.5	1.7	15

2.1.3 Comparing random and systematic designs

The two simulated distributions used for the survey strategy evaluation (Section 2.1.1) were used to evaluate the differences between a systematic survey design and a fully random survey design. The two methods each with 1000 different sampling realisations were defined as the following:-

- **Systematic:** a regular grid of 64 points, arranged in an equally spaced 8 by 8 grid with a spacing of 1/8 of survey dimension with a 2D random starting location on a scale of 1/8 by 1/8 of dimension of the area.
- **Random:** the procedure starts with initially 64 stations, the number of stations is then increased by adding new random stations and checking for time available using the travelling salesman algorithm (Harbitz and Pennington, 2004), until the maximum number possible in the time allocated is reached. The number of stations for each of the 1000 random sampling realisations is given in Figure 4.

The results of the simulation were evaluated through examination of the distribution of the estimates of the total abundance for each method. These distributions are given separately for each simulated surface in Figure 5. For both methods and both simulated surfaces the estimates of mean abundance are unbiased at $1*10^7$.

Figure 5a shows the results from simulated surface 1 which has high variance and low spatial autocorrelation. In this case, the results indicate that the random survey, which has the higher number of observations, has the lower RSE (49%) and provides a more precise estimate than the systematic survey (RSE = 56%). Note also that the distribution is very skewed.

Figure 5b shows the results from surface 2 with the lower variance and higher spatial autocorrelation. In contrast to surface 1, the improved precision due to even allocation of sampling with the systematic survey delivers improvement in the estimate of abundance over the random survey. In this case the systematic survey RSE = 14%; even with the extra samples obtained for the random survey, the RSE (23%) is poorer. These contrasting results for the two different spatial distributions show that there is an interaction between spatial autocorrelation and sampling design. Further investigation of a wider range of surfaces with different proper-

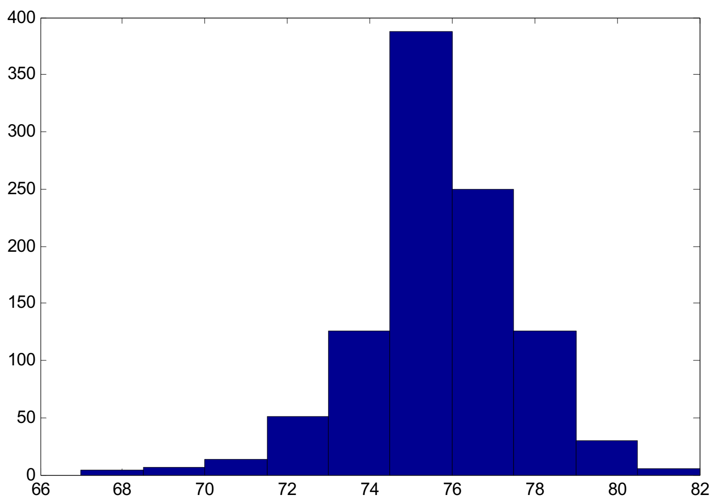


Figure 4. Number of randomly located stations in a fixed time with minimum track obtained using the travelling salesman algorithm. (9 days with a survey speed of 10knots and trawling time of 1.5 hours in a 120 $N.mi^2$ area.

ties should help to refine the parameters that influence the point at which different survey strategies are more efficient estimators of the abundance and variance.

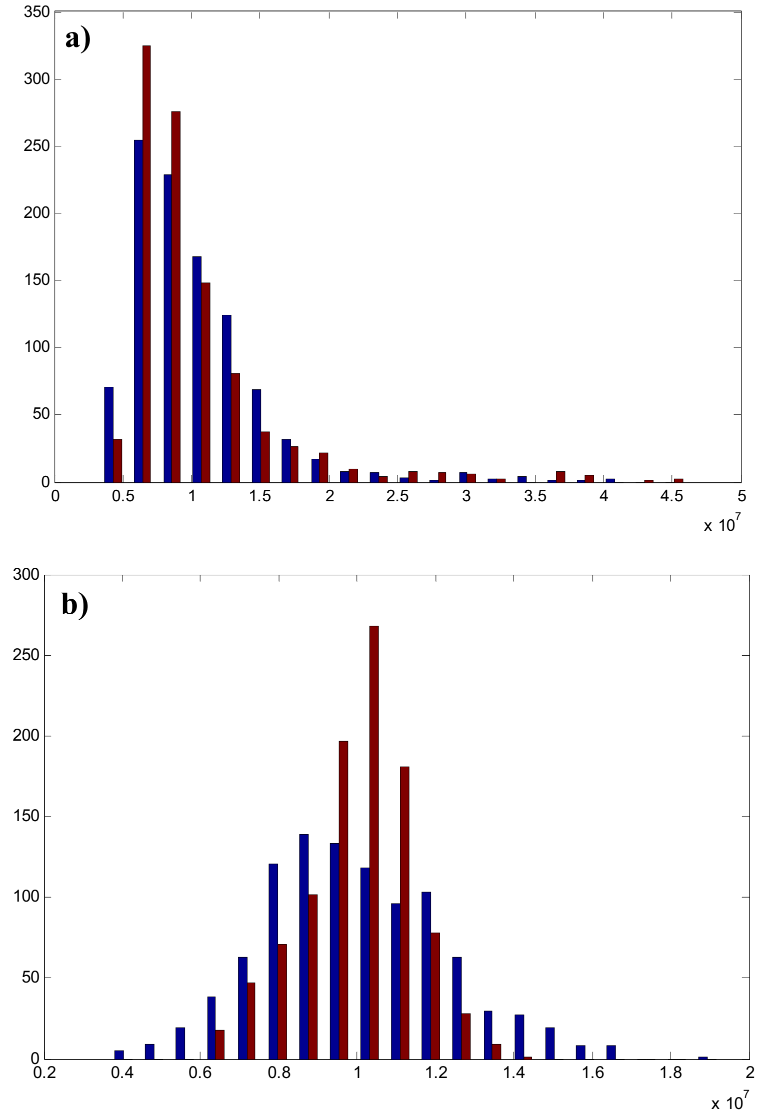


Figure 5. Frequency distribution of estimates of total abundance for systematic survey (red) and random survey (blue) for: a) high variance low correlation surface (upper panel); and b) lower variance more correlated surface (lower panel).

2.2 Miscellaneous methods

2.2.1 Estimating the precision of echo-integration trawl surveys of walleye pollock standing stocks in an area near Kodiak Island, Alaska

Acoustic data from a series of repeated echo-integration trawl surveys in an area near Kodiak, Alaska were analyzed using a variety of methods to produce estimates of variance of the mean density or total biomass of walleye Pollock (*Theragra chalcogramma*) in the survey area. The methods included: replicate surveys, a 1D transitive geostatistical method (Williamson and

Traynor, 1996), conditional Gaussian geostatistical simulation (Gimona and Fernandes, 2003), a random field linear model (Lai and Kimura, 2002), and cluster analysis (Williamson, 1982). The single survey with the highest skew and the highest single observation was analyzed using all the methods for comparison purposes.

Relative standard error (RSE) from four sets of surveys repeated three times each (Barnabas 2001: 27.6%, Barnabas 2004: 13.1%, Chiniak 2001: 7.4%, Chiniak 2004: 13.7%) are considered to be overestimates of the variance associated with a single survey because fishing effects (for Barnabas) and possible temporal changes in fish abundance over the survey period are included in the error term if the repeated surveys are treated as replicates. For the first survey in Barnabas Trough in 2001 considered here, RSEs obtained from the 1D geostatistical method (11.7%) and the conditional Gaussian geostatistical simulation (15.5%) were lower than the estimate obtained from repeated surveys. Estimates made using methods ignoring the non-random sample design (and thus invalid) had higher RSEs than geostatistics-based estimates: transect cumulates as replicates 25.9%; paired transects in strata 18.2%; assuming independent 0.5 nmi EDSUs (Equivalent Distance Sampling Units) 29.4%. Estimates of RSE using classical approaches (reversible field line mapping 72.9%, cluster analysis: 30.6%), in which autocorrelation is considered to reflect a redundancy of information, reducing the effective degrees of freedom and increasing the variance in the integral/abundance estimate, are much higher than the estimates from repeated surveys, and are not recommended for analysis of acoustic survey data (ICES, 1993).

2.2.2 The use of cluster analysis for stratification in the Celtic Sea

Given the range of habitat types in Eastern Atlantic areas, and the steep bathymetric gradient along the edge of the continental shelf, stratification in many of the eastern North Atlantic surveys is based primarily on the interpretation of ecologically meaningful strata (e.g., as determined by cluster analyses of catches).

A first analysis was carried out in the Bay of Biscay and a stratification scheme was established to be used in the sampling design of the French EVHOE (Evaluation des ressources halieutiques de l'Ouest de l'Europe) survey initially conducted in the Bay of Biscay only. When the survey area was extended to the Celtic Sea, and in the absence of any data on fish distribution, this depth stratification was extended and coupled with a geographic stratification from north to south (North, Centre, South).

After 7 years of surveys, and in order to check the adequacy of the existing stratification, data collected were used to analyse the spatial organisation of species assemblages on the continental shelf and upper-slope of the Celtic sea in the period 1997–2003 (Poulard and Mahé, 2004). The study of the multispecies spatial structures over time requires the combined analysis of different tables of species density sampled at different stations. This was done using multitable factorial analysis. The table of the total number of individuals per survey and per species (matrix with seven surveys and 52 species) was used as input in a between-class correspondence analysis (CoA) to test a survey effect in the overall species composition.

Automatic classification techniques were used to establish a cluster distribution of the sampling sites. Hierarchical ascending classification was applied to the factorial co-ordinates of sites in the space defined by the multitable analysis.

The spatial distribution and species composition of the 5 different groups identified are shown in Figure 6 as well as the actual strata borders.

Species assemblages and EVHOE stratification sampling scheme

Table 2 compares the number of hauls per species assemblage and per stratum used for the EVHOE sampling scheme. It shows that southern (cluster 1), Northeast (cluster 4) and central (cluster 5) Celtic shelf assemblages fit individually with a limited number of strata. On the contrary, the western assemblage encompasses a large depth range (120–400 m) along the shelf edge. The transition zone is more evident between 120–160 m and slightly more in the north of the study area (Cc4).

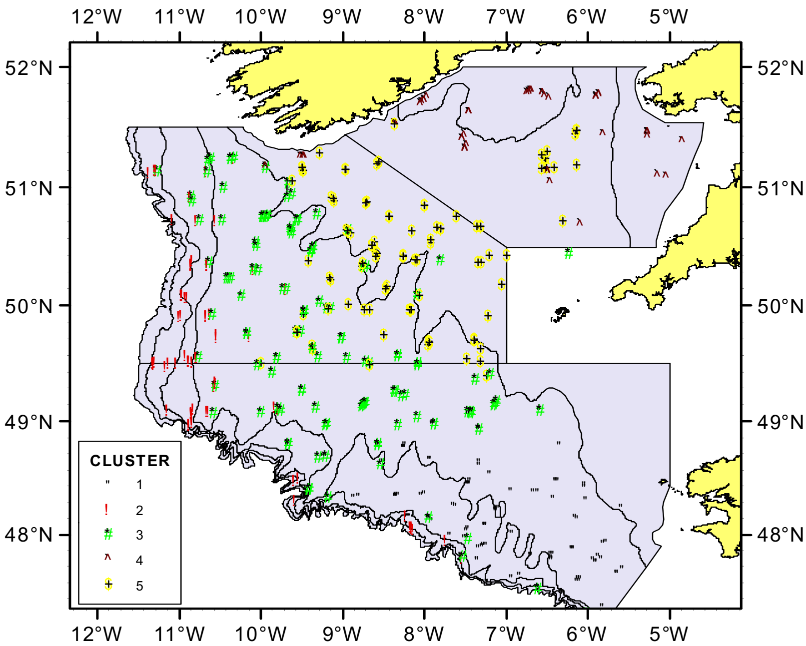


Figure 6. Distribution of the fish species assemblages in the Celtic sea based on 458 hauls sampled during autumn surveys from 1997 to 2003. Space partitions in 5 clusters were obtained by ascending hierarchical classification of the factorial scores of the hauls on the first three axes.

In all of the strata, more than 50% of the hauls belong to one cluster. For 6 out of 9 strata, more than 68% of the hauls belong to one cluster. This provides some evidence that the current definition of the strata is relevant.

Table 2. Number of hauls per fish species assemblage and stratum for the EVHOE surveys.

	Depth	Cluster					
Stratum	range (m)	1	2	3	4	5	Total
Cc3	80–120			10	4	51	65
Cc4	120–160		15	57		33	105
Cc5	160–200		11	5			16
Cc6	200–400		14	1			15
Cn2	30–80				28		28
Cn3	80–120			1	20	15	36
Cs4	120–160	63	12	43		2	120
Cs5	160–200	39	8	10			57
Cs6	200–400	3	11	2			16
Total		105	71	129	52	101	458

Implementation in the coordinated Western division IBTS surveys

At the 2005 IBTSWG (International Bottom Trawl Survey Working Group) meeting, the nations operating in this region reached general agreement on this stratification scheme, given that only minor modifications to alternative national schemes would be required. Furthermore, it is hoped that these bathymetric strata can also be extended northwards off the western coasts of Ireland and Scotland. Comparable strata will be developed for the Irish Sea, though the sedimentary environment in this area will also be incorporated in strata design, as sediment type and bathymetry are key determinants for assemblages in this region (Ellis *et al.*, 2000, 2002; Ellis and Rogers, 2004).

At a cursory level the strata constructed during the above analysis were in agreement with the rudimentary sediment maps available at the time. However, the current technology available on the relevant vessels is facilitating more routine capture of seabed discrimination data. As this applied habitat covariate data currently being acquired becomes available for all surveys, the efficacy of the above stratification will be reviewed.

As a general point, while stratification should in principal improve the precision of survey estimates, where a survey has several target species of interest this can often be confounded by differing species-specific spatial patterns of distribution. In such cases it may be beneficial to do some preliminary analysis to establish a hierarchy of which species in particular might benefit from stratification and concentrate on these for the analysis (see Smith and Gavaris, 1993).

2.2.3 Abundance estimator based on distribution assumption

A random effects model for disentangling population abundance and capture efficiency effects on bottom trawl catches was proposed (Trenkel and Skaug, in press). The spatial distribution of individual fish is assumed random, leading to a Poisson distribution for the number of individuals in the trawl path (no schooling). Capture efficiency, i.e., the proportion of individuals in the trawl path being retained by the gear, is modeled as a random variable. The proposed model extensions include the effects of mean body size on capture efficiency and of mean age on average abundance. Estimation is carried out by Maximum Likelihood. The precision of the average density (mean of Poisson distribution) is estimated from the observed Fisher information matrix using AD Model builder. The method was applied to several species from the Celtic Sea groundfish community based on small-scale repetitive hauls. The ratio between the obtained abundance estimates and the average catches ranged from about 5 to 20 for the different species. The relative standard errors of the estimated mean densities were between 4 and 17% with the exception of haddock (*Melanogrammus aeglefinus*, 160%). The estimated capture efficiencies were comparable between species and showed that generally capture efficiency increases for larger species with the exception of haddock, which had low estimated capture efficiency despite its large body size. Model identifiability was studied using simulations and an independent trawl data set from the same area.

2.2.4 Lake Ontario Alewife Abundance

The U.S. Geological Survey's Great Lakes Science Center conducts annual surveys of alewife (*Alosa pseudoharengus)* with bottom trawls in U.S. waters of Lake Ontario in cooperation with the New York State Department of Environmental Conservation. For the purposes of comparing analyses of survey data, we focused on the relative biomass of adult alewives (age two and older) in 2003. The sample space was limited to the depth range (0 to 160 m) where bottom trawl catches of the target species have been highest historically. A fixed survey design was used, consisting of sampling at up to 13 sites at each of 12 ports. Tow duration was targeted at 10 minutes.

Biomass estimates were calculated using two methods. First, we assumed that the fixed survey was, in fact, a stratified random survey, with 20-m depth zones from 0 to 160 m as strata, and the fixed sampling stations were random samples. Relative mean biomass and its variance were then estimated using standard methods (Cochran, 1977). Second, biomass estimates were calculated based on the assumption that alewife biomass could be described by a smooth spline function of fishing depth. Predictions were made across the entire sample space to estimate the overall mean, and variance was estimated using bootstrapping. Estimates from both methods were essentially identical (mean 27 kg per 10-minute tow and RSE about 25%).

Information from the 2003 survey was used to investigate the effects of optimal allocation of sampling effort. Because the time to take a single bottom trawl sample increases with bottom depth, optimal allocation has to take cost of sampling into account. In 2003, the total on-site sampling time for 98 stations was 50 hours (this does not include travel time). Using this as our fixed on-site sampling cost, we calculated the optimal allocation as 84, with most of the samples (74%) being placed in the depth strata from 80 to 120 m. Application of the optimal allocation, through resampling of the 2003 data and calculation of the design-based estimator, resulted in a reduction in the error in the estimated mean abundance (RSE = 15%).

However, because the depth distribution of alewives in Lake Ontario may change annually (O'Gorman *et al.*, 2000), a single fixed allocation of sampling effort will be not be optimal every year. Thus, it may be beneficial to incorporate some adaptive sampling in the survey design, taking more samples in those depth zones yielding large catches of alewives, and taking fewer samples in those depth zones yielding smaller catches. Use of an adaptive design would necessitate the use of a model-based estimate.

Further details of this work are provided in the working document WD5 attached in Annex 2.

2.2.5 The geostatistical transitive approach

When geostatistics is applied, it is often done in the so called intrinsic approach using variograms (Rivoirard *et al.,* 2000; Petitgas, 2001). However, the estimation of the variogram is often difficult in practice due to the characteristics of the fish data (i.e., the location of the high values in the field, the numerous low or zero densities), and due to the hypotheses associated to the use of the variogram (Matheron, 1971; Petitgas, 1993; Bez and Rivoirard, 2001). Although some authors are suggesting more robust estimators for the variogram (Cressie, 1991), the method itself might be regarded as based on too strong hypotheses. In this regard, one usually looks for estimations based on as few hypotheses as possible (principle of parsimony) as this reduces the possibilities to observe discrepancies between the characteristics of the data and the assumptions on which the estimator is based (robustness).

To estimate global estimation variance in case of regular sampling, Matheron (1971) developed the transitive approach, a model-based method which requires fewer hypotheses than the intrinsic approach. Bez (2002) provides a detailed description of the method with two examples of fisheries applications.

The transitive method is an appropriate technique for systematic sampling schemes (i.e., regular designs with random origin). It can also be applied to random stratified designs, i.e., designs with one point located at random in each block of a regular lattice. There are two fundamental reasons for that. The first one is that realistic estimates of the covariogram are only available for regular samplings where each observation gets the same area of influence. In case of an irregular sampling, a complex weighting procedure based on the surfaces of influence of each samples has to be used (Bez *et al.,* 1995). The second reason is that the estimation variance it furnishes is based on the fact that combining all the possible outcomes of the random starting point of a sampling grid and the grid nodes, amount to cover space entirely. This is no longer true for irregular samplings.

The theory makes relatively few assumptions: it assumes the randomness of either the origin of the sampling grid (Figure 7.) or the location of data points in grid cells. These assumptions are easily controllable in practice (i.e., falsifiable). Together with the low number of parameters to be estimated, this ensures robust results.

Let x represents a point in space. The fish density $z(x)$, taken as a function of space, is a regionalised variable expressed, for instance, as the number of individuals per unit surface

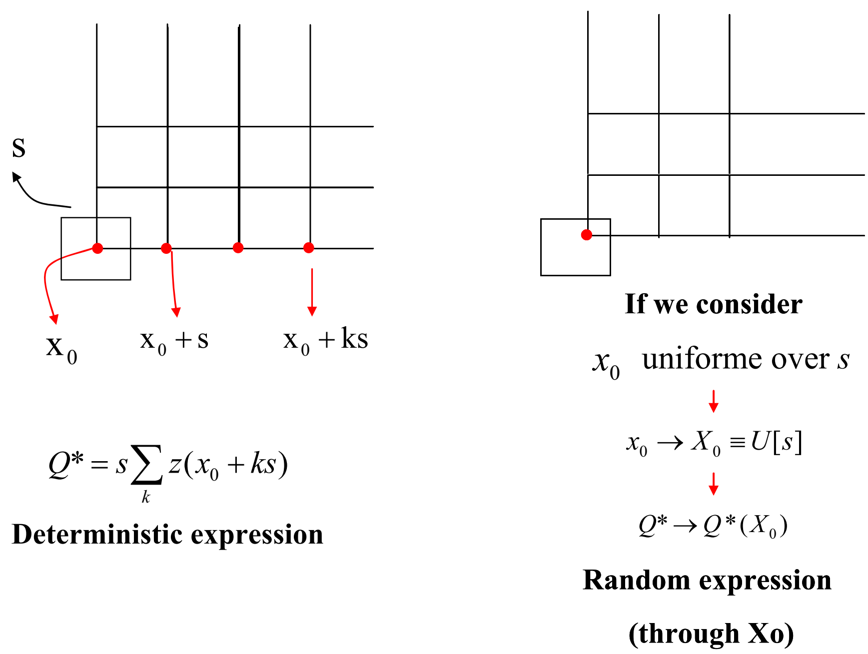


Figure 7. Notations and status of the origin of the sampling grid.

area (e.g., ind·m^{-2}). The total fish abundance is $Q = \int z(x)dx$. Assuming the location of the origin of the sampling grid is randomly located we get the estimator denoted $Q^*(X_0)$.

The estimator is unbiased due to the uniform distribution of the origin of the grid. After Matheron (1971), the estimation variance can be expressed as the difference between the discrete and the exact integral of the covariogram. When a significant nugget effect exists, it explains nearly all the RSE (Figure 8). The RSE can then be approximated by:

$$RSE \approx \sqrt{\text{block area } \times \text{nugget effect}}$$

2.2.6 Confidence intervals for trawlable abundance from random stratified bottom-trawl surveys

An approximately pivotal statistic is proposed that can be used to construct confidence intervals about average and total trawlable abundance from stratified random bottom-trawl fisheries surveys. The statistic is based on the strata area-weighted average that is commonly computed from the survey catches. The distribution of the statistic is derived from both the random selection of sites to survey and the random fish capture process at a site. This is in contrast with the commonly used "design-based" approach to statistical inference that includes only the randomness in the sites selected for trawling. The method is applied to case studies, and simulations based on these case studies are used to examine the coverage accuracy of the confidence intervals.

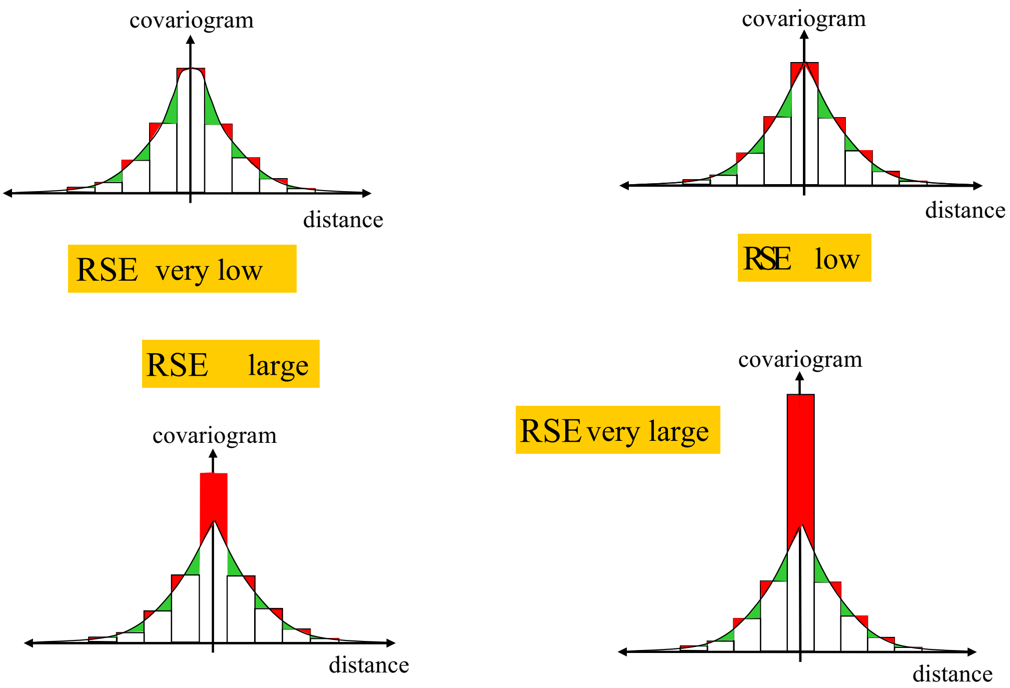


Figure 8. Fluctuation of the relative standard error with the level of heterogeneity of the fish spatial distribution. Justification of the approximation of the RSE based on the nugget effect component only when this exists.

Further details of this work is provided in the working document WD4 attached in Annex 2

2.2.7 Current thoughts in Geostatistical conditional simulation

Linear geostatistics, i.e. geostatistics based on the variogram, allows for the estimation of abundance with its estimation variance. However, in complex situations, such as when combining acoustic and biological data (mean length, proportion at age), these methods are limited. As an alternative, geostatistical conditional simulations can be used to link the uncertainty of these variables and to determine the resulting estimation variance. Simulations made for this purpose have to deal with the specific distributions and relations being considered, e.g. highly skewed distributions with many zeroes or categorical variables. Such analyses are under development to determine the error structure of Scottish herring survey estimates based on the specific multivariate model presented in Rivoirard *et al.*, 2000.

2.3 An example of an ecosystem approach: the MEDITS programme

2.3.1 Introduction

The MEDITS programme was initiated at the request of the European Commission. It started in 1993 with the collaboration of the four Mediterranean Members States (Spain, France, Italy and Greece) and the first surveys took place in June–July 1994. In 1996, the European Commission funded the participation of Slovenia, Croatia and Albania, under Italian co-ordination, to cover the whole Adriatic Sea. In 1999 and 2000, Morocco participated in the survey with the financial backing of FAO/COPEMED to cover the Alboran Sea (east of the Gibraltar Strait). In the same year Malta joined the programme with EU funding. In 2002, the EU signed a agreement on Data Collection Regulation with the Member States which obliged them to produce the basic data needed to fishery regulation (fishery statistics, length frequencies of the landings, survey at sea data). In 2004 Slovenia, Malta and the Greek part of Cyprus became Member States and, hence, full members of the programme. Cyprus will start the survey in June 2005.

2.3.2 MEDITS survey methods

Currently, the survey covers all the trawlable areas from the strait of Gibraltar to the Aegean Sea plus, from 2005 onwards, the Greek part of Cyprus, between 10 and 800 metres depth (Figure 9).

All participants use the same trawl (GOC 73) with the same rigging. The sampling scheme is strata based with several areas for each country (for example in the French Waters two areas in the Gulf of Lions, west and east of 4°E and two areas in the eastern coast of Corsica, north and south of 42°N). Each area is divided into 5 depth strata : 10–50 m, 50–100 m, 100–200 m (these 3 strata covering the continental shelf), 200–500 m and 500–800 m (on the slope). The locations of the trawling position have been fixed at random in each stratum, taking into account knowledge about the seabed (stones, wrecks, etc.). The haul duration is ½ hour on the shelf (10–200 m) and 1 hour on the slope (200–500 m).

All species caught are numbered and weighed, and for 36 of them (26 fish, 4 crustaceans, 6 cephalopods of commercial interest in at least one of the participant countries) sex and maturity stages are determined and the individuals or a subsample of them are measured by sex and maturity stage.

The MEDITS programme has co-ordinated the work of more than 50 scientists from nearly 20 institutes. To date, 40 scientific papers have been published, plus the proceedings of a symposium held in Pisa (Italy) in 1998 and a special issue of the revue Scientia Marina in 2003.

The first goal of the programme was to provide abundance indices by area together with length frequency distributions. But considering both the huge quantity of data provided by the survey and the new tools developed for example within the Fishery Information System (SIH) of IFREMER, it has been decided to enlarge the analysis by developing population and community indicators.

2.3.3 Population indicators

Given that a fishery reduces the population abundance, the population growth rate (r) fromthe population growth model:

$$N(t) = lN(t-1) = N(t-1)e^{-r}$$

is used an as indicator, such that if r

>0 growing population

=0 stationary population

<0 decreasing population

- Fishery decreases the mean length of the individuals within the populations

- When the fishing mortality increases, the total mortality Z increases as well.

Z Can be estimated by the number of individuals by length class:

$age = t_0 - 1/k \log(1 - l/L_\infty))$

$N_a(t) = N_{a-1}(t-1) \exp(-Z)$ $Z_a = -\log(N_a(t)/N_{a-1}(t-1))$

- Fishing induces an earlier maturity at smaller lengths

L_{50} = length at which 50% of the population is mature, estimated by logistic regression (GLM) including the time trend

2.3.4 Community indicators

Fishing activity decreases

- Overall biomass
- Overall abundance
- Mean individual weight in the community
- Proportion of "large" individuals in the community

«large», i.e. larger than an arbitrary threshold, empirically fixed here at 27 cm (but it is still needed to check the effects of changes of this threshold on the results).

2.3.5 Conclusions

These tools have been used for the first time at a large scale (all the areas covered by the MEDITS programme) in March 2005 and this work is still under development but promising (e.g. Table 3). Some results are, for example, that in French waters there are generally no significant trends for the various indicators. But is no trend good news? For instance, east of Corsica, which is currently poorly lightly exploited it can be considered as good news: the state of the resources will stay at a sustainable level. On the other hand, in the Gulf of Lions which is known to be fully or strongly exploited, the absence of any trend can be interpreted as there are no chance to observe a recovering of the fish stock in a near future.

Table 3. Indicator summary : green/red signals for Corsica and Gulf of Lions

Indicator	East Corsica, 1994-2004	Gulf of Lions, 1994-2004
Population abundance	2/43 ↘ 6/43 ↗	0/48 ↘ 6/48 ↗
Average length in population	1/30 ↘ 1/30 ↗	4/35 ↘ 4/35 ↗
Total mortality	1/4 ↘ 0/4 ↗	0/11 ↘ 0/11 ↗
Length at maturity	10/18 ↘ 2/18 ↗	14/25 ↘ 6/25 ↗
Total abundance	→	→
Total biomass	↗	→
Average weight in community	→	→
Average length in community	→	→

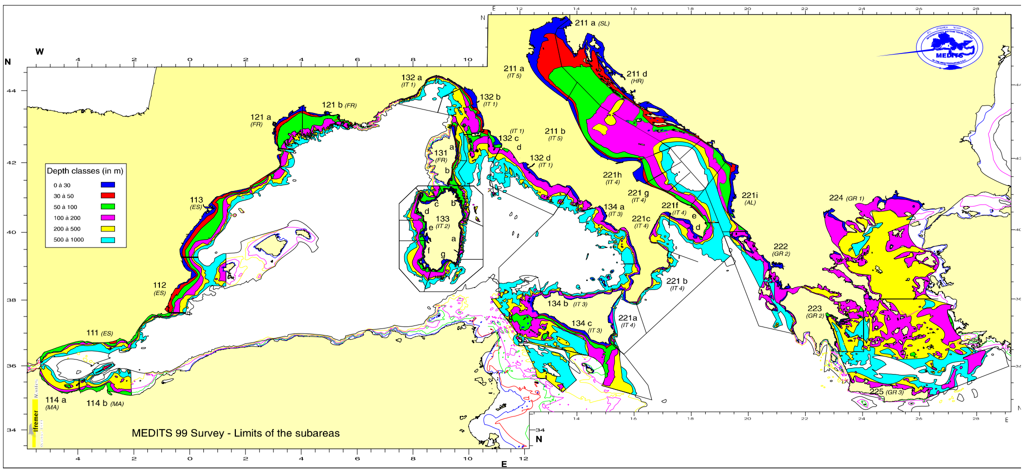


Figure 9. Map of the Mediterranean Sea showing the MEDITS survey area coverage (with colour-coded depth strata).

2.4 Estimating the variance of an abundance estimate

When estimating the abundance of a population, or equivalently the mean fish density over a domain when this domain is fixed, we are also interested in the precision of this estimate, to know how close the estimate is to the "true" value. A basic way to do this is through the variance of the difference between the true value and the estimator, that is, the variance of the estimation error, or estimation variance:

Var(estimator - true)

This can be computed by considering the random elements in the process used: e.g. the random locations of sample points (or the random origin of a grid), or the fish density considered as a random function or process.

When the fish density is not considered as a random function (or if it is considered random but with no spatial structure), the true value is not a random quantity but a fixed one. It follows that the estimation variance (variance of the estimation error) coincides with the estimator variance (variance of the estimator):

Var(estimator - true) = var(estimator)

This applies notably when only the location of sample points is randomized (or only the origin of the grid is randomized, in particular in the transitive geostatistical approach).

In the usual intrinsic geostatistical approach, however, the fish density is considered as a random function, so that its mean value over the domain (representing the true value to be estimated) is also a random quantity. In this case we have (supposing the different terms can be defined in the geostatistical model):

Var(estimator - true) = var(estimator) + var(true) - 2 cov(estimator, true)

so that the variance of the estimator would be in general different from the desired estimation variance. (For instance, the same arithmetic mean of, say, regularly spaced sample points will correspond to different estimation variances when changing the frontiers of the estimated domain).

In particular, what is known as the kriging variance corresponds to the estimation variance of the kriging estimator, not to the variance of the estimator itself. Moreover, when kriging the mean density over a domain, its kriging variance is not directly related to the kriging variances of the points within the domain, nor to the variance of kriged points.

2.5 Conclusions: a general synthesis of ideas on survey design

It is clear from the descriptions above that analyses of survey data are possible using a variety of model based and design based techniques, almost regardless of the survey design (Figure 10) This is likely to continue to be the case and reflects peoples particular expertise, experience and in some cases, philosophical preferences. The issue of survey design, however, may be more open to achieving some consensus.

A decision tree has been constructed (Figure 11) with the objective of determining which design might be appropriate given the objective of estimating the abundance of a single marine resource from a survey in a fixed time. It leads to one of the following generic survey designs: random, stratified random, stratified random (blocked), and systematic. There may be further subdivisions of these (e.g. star, zig-zag or parallel transect systematic designs) which are not addressed here.

The first choice is made on the basis of any existing knowledge about the spatial distribution

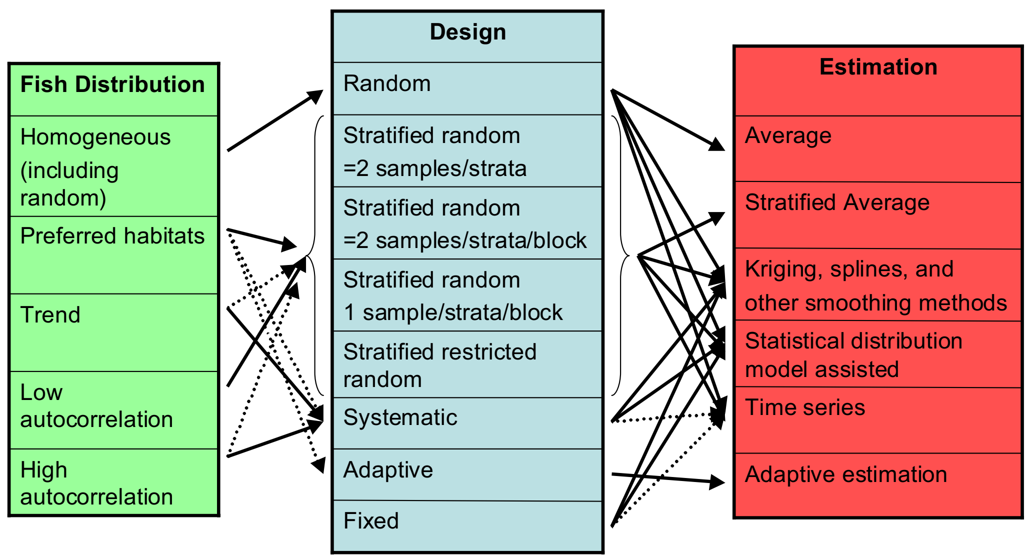


Figure 10. A table indicating the various routes available for survey design and analysis.

of the target species. If nothing is known about the spatial distribution, then a stratified random design should be employed with the area (domain) divided up into small areas (blocks) and two or more samples located in each block. If the fish distribution is known to be random, or random within a stratum, then a random or stratified random survey, with samples chosen at random within strata, will provide as precise an abundance estimate.

The choices are primarily influenced by the spatial distribution of the fish. A distinction is made between a stationary process, where the expected mean and variance of fish density is consistent throughout the area (strictly speaking this is second order stationarity), and a trend, where these may change systematically. A process may be stationary over the whole domain or within strata, hence the second question.

The two extreme cases are likely to be correct: (1) random field: random survey; (2) high autocorrelation: systematic survey. With no autocorrelation, a (stratified) random survey has higher precision due to more samples being available in a fixed time from savings made due to passage optimisation (or "travelling salesman") algorithms. With low autocorrelation, a stratified random survey with 2 samples per block is recommended because this is an expected compromise between some gain in the precision due to regular placement of samples (from some autocorrelation), and some gain in samples numbers from application of the travelling salesman and a simple design based estimate of variance. This is therefore a compromise design between fully random and systematic. The recommendations are based on the perception that as spatial structure increases, there will be matching advantages in increasing survey design structure rather than just a direct switch from one to another. This should be investigated with simulation, considering changes in nugget, range (and a trend could possibly be included as a long range component) and survey design in a FIXED time.

It is implicitly assumed that effort (either within strata or not) is allocated in proportion to area and variance. As with most fish populations a mean variance relationship is expected and therefore the effort can be allocated in proportion to the mean. This proportional allocation, however, may not be continuous and logistically is likely to be increasing factors of e.g. 2 (i.e. 5 transects per unit area in low abundance areas; 10 transects per unit area in medium abundance areas; and 20 transects per unit area in high abundance areas).

The choice is also influenced by the type of sampling device used, specifically, if samples are taken continuously (as in an acoustic, or visual survey) or in a discrete manner (using a trawl, ichthyoplankton or dredge). When collecting data continuously, there are no advantages to random designs (other than perhaps the convenience of calculating the variance). This is because there is no increase in the number of samples among any of the designs for a continuous sampling tool. When sampling continuously all the time is used, no matter what the area, and, therefore, whether a random track or a systematic track is carried out, the amount of sampling in a fixed amount of time is the same. As the random design confers no advantage in the amount of sampling done, then there is always the possibility of taking advantage of (perhaps unknown) small range spatial autocorrelation with a systematic design. In contrast, for discrete sampling designs, there are increases in the number of samples (for the same fixed time) for more random designs when passage optimisation (or "travelling salesman") algorithms are employed: this will deliver improved precision for more fields wit low or no autocorrelation.

The outcomes are colour coded: black text is the survey design recommended; blue the estimator to determine the abundance; and red the estimator to determine the precision of the abundance estimate. Note that the geostatistical estimation variance requires a model variogram, the parameters of which are influenced greatly by the quality of the experimental variogram.

The final question asks how much autocorrelation is present: this is currently an issue of debate, but based on the variogram, high autocorrelation may refer to a model variogram with no or a small nugget effect (<50% of the semivariance) and a long range (>3 times the mean intersample distance). Low autocorrelation would have a high nugget (>50% of the semivariance) and short range (<2 times the intersample distance). No autocorrelation would either have a pure nugget effect (100 % of the semivariance) or very poor structure in the variogram (i.e. autocorrelation is unknown).

The tree does not consider adaptive surveys. These are advantageous where the target is static and distributed in patches which are large relative to the inter sample distance (i.e. the probability of taking the wrong adaptive decision is very low). They can be considered therefore as a subset of outcome 1b.

Measurement bias is not included in any analytical method (unless it is measured independently). Measurement error may then be considered in a continuum between a process that occurs in short time scales (in which case it IS included); or one that occurs over longer time scales (in which case its inclusion depends on the interaction between rate of coverage, strata size and design/analysis method). Large strata with random designs include more long time scale measurement error as oppose to small strata with systematic designs which model long time scale measurement error as autocorrelation or trend. For example, trawl surveys taking place over the course of two weeks may have lunar cycle effects as trend that would not be included in the measured error. A short time scale error, could be for example, induced by current: this will affect the volume of water sampled (through variations in speed and gear geometry), but over the course of a survey one would expect these to be a random process and therefore contribute to the overall error. The tree does not explicitly deal with estimation of individual components of the variance. More importantly, the decisions which the tree address, are based on optimising the design for spatial sampling not for consideration of total survey error.

Where possible, absolute measures should be evaluated rather than indices. Absolute measures serve more purposes: they can be used as indices; they may be useful purely as absolute estimates (e.g. capelin in the Barents Sea); and, most significantly of all, given the needs of an ecosystem approach, they are essential to compare the abundance of one species with another. It is recognised, however, that absolute abundance requires knowledge of whole gear selectivity which is currently lacking for many survey sampling tools such as trawls. The transition to absolute abundance estimates may therefore be slow.

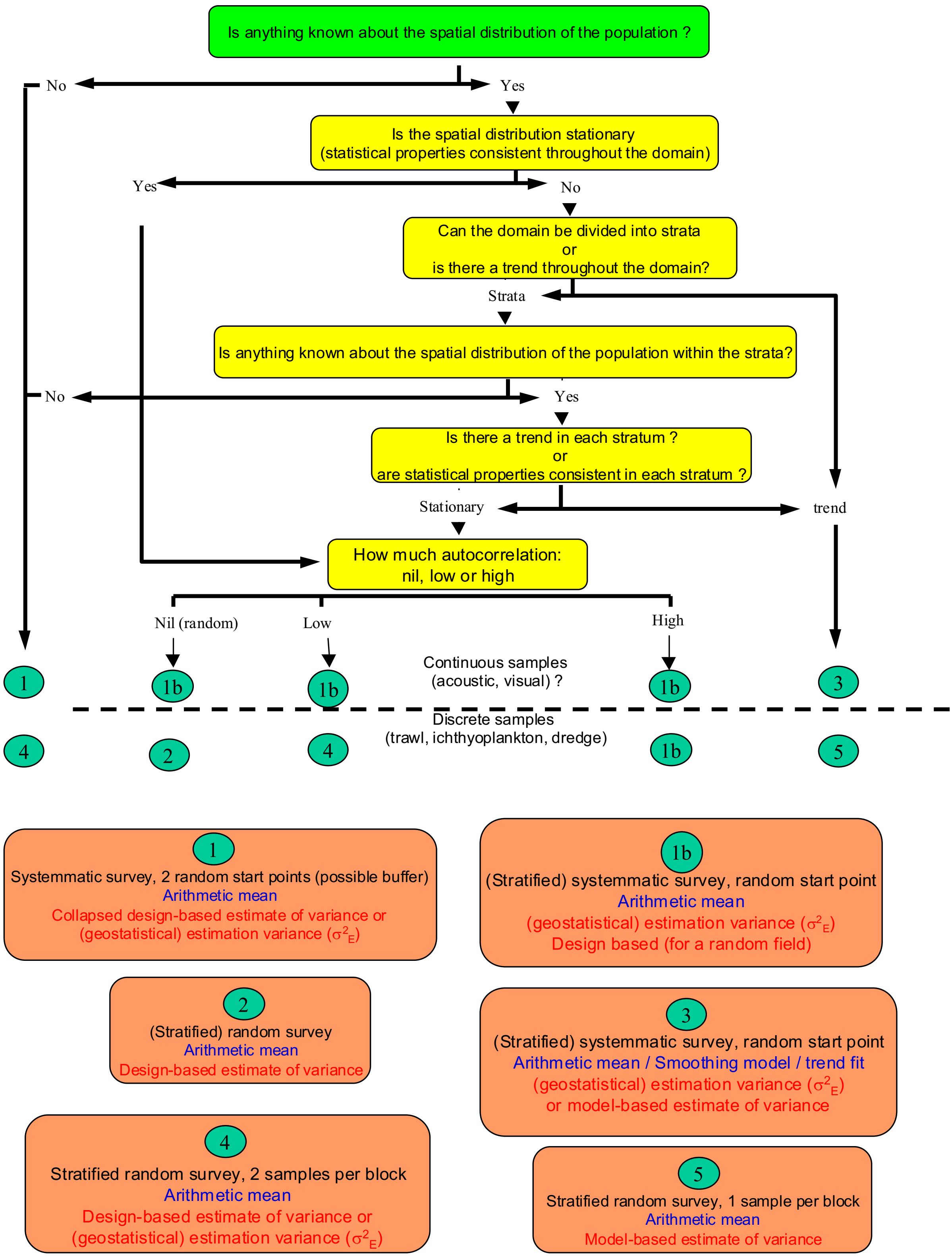


Figure 11. A decision tree to determine the type of design which would be optimal for the most precise estimate of the abundance of a single fish species. The term "stationary" in relation to the spatial distribution refers to a constant (expected) mean and variance throughout the area. Stratification is used in a rather loose manner and could indicate any subdivision of the area into strata based on for example, depth or substrate. There is a distinction between the latter, and smaller regular strata which are referred to as "blocks": a block might be, for example, an ICES statistical rectangle. See text for further details.

3 Survey tow duration

3.1 A Review of survey tow duration

The following is a brief description of the effect trawl duration has on survey precision. For more details see Pennington and Vølstad (1991, 1994). Tables 4 through 6 are examples of the relation between sampling unit size and the associated coefficient of variation.

Table 4. Estimated coefficient of variation ($cv = s/\bar{x}$) with approximate standard errors (in parentheses) for various sampling unit sizes for some sea scallop (*Placapecten magellanicus*) populations on Georges Bank.

Area	Country conducting survey	Average catch/100m^2	Average $\hat{cv}$	Sampling unit size (m^2)	Number of samples
Northeast part, 1982–1984	U.S.A.	5.79	1.41 (.15)	3,954	235
	Canada		1.34 (.08)	3,013	589
South Channel, 1983	U.S.A.	14.42	1.64 (.52)	3,954	32
	U.S.A.		1.55 (.47)	1,318	32
All areas, 1975, 1977, 1978	U.S.A.	3.42	1.63 (.16)	4,943	343

Table 5. Estimated coefficient of variation for haddock from a tow duration experiment on Georges Bank in January 1965. Each estimate is based on 16 tows. The trawl swept 1,145 m^2 of bottom per minute.

Length of tow (min)	15	30	60	120
Avg. catch per tow	190	394	627	1,341
$\hat{cv}$	1.16	1.90	1.56	1.53

Table 6. Estimated coefficient of variation for haddock from two tow duration experiments in the Barents Sea. Each estimate in the first experiment is based on 20 tows, and in the second experiment on 8 tows. The trawl swept 1,574 m^2 of bottom per minute.

Length of tow (min)	5	15	30	60
Avg. catch per tow[a]		184	319	438
$\hat{cv}$		.94	1.17	.80
Avg. catch per tow[b]	33		126	
$\hat{cv}$	.72		.68	

[a]October 1988.
[b]January 1989.

For these examples, the CV does not appear to decrease with increasing sampling unit size. The sampling distribution converges to the Poison for sufficiently small unit sizes and because the CV is approximately constant for larger units the relation between the mean and the variance for varying tow durations (unit sizes) is approximately:

$$\sigma_t^2 = \mu_t + b\mu_t^2, \qquad (1)$$

where μ_t and σ_t^2 are the mean and variance, respectively, of catch-per-tow for tow duration t and b is a constant. Therefore the CV_t as a function of t is given by:

$$CV_t = (1/\mu_t + b)^{1/2}. \quad (2)$$

Figure 12 is a plot of the CV_t versus tow duration for ocean pout (*Macrozoarces americanus*), which had a rather low catch rates during experimental towing on Georges Bank.

The total time, *C*, to conduct either a random or systematic survey is given approximately by the cost function:

$$C = (c_1 + t)n + c_2\sqrt{n}, \quad (3)$$

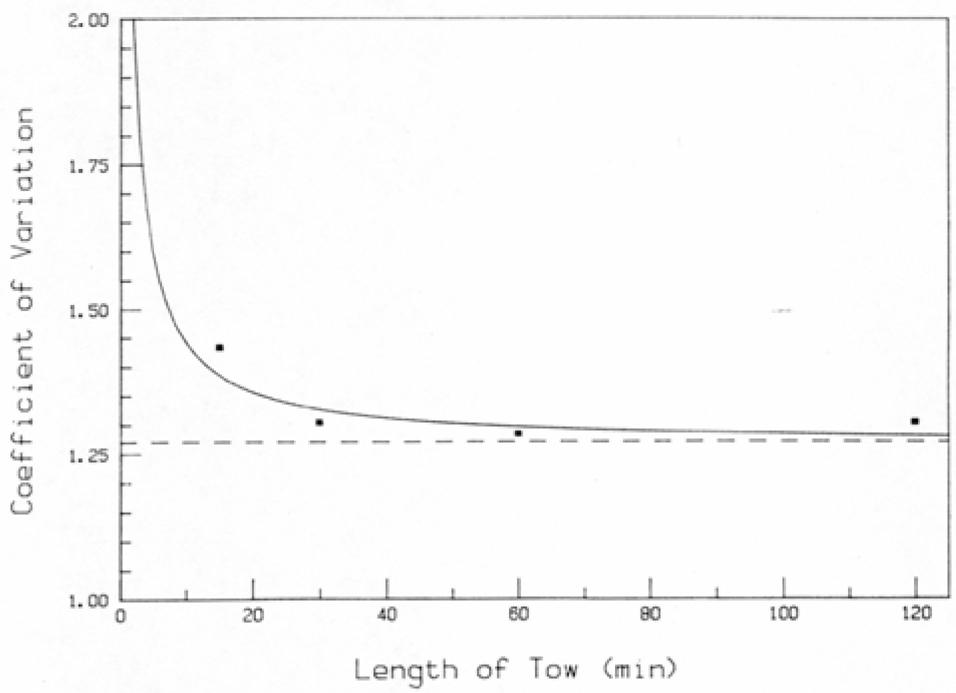


Figure 12. The coefficient of variation as a function of tow length for ocean pout. The graph is given by equation (2), and the points are estimates each of which is based on 16 tows. The average numbers of fish per tow were 2.3, 7.5, 13.1 and 24.9 for the 15-, 30-, 60- and 120-min tow respectively.

where c1 is the average time needed to set and haul the trawl, t is tow duration, n is the sample size and c2 is a constant that depends on the cruising speed, v, and the area, A, of the survey.

For random surveys, $c_2 \approx \frac{0.8}{v}\sqrt{A}$ and $c_2 \approx \frac{\sqrt{A}}{v}$ for systematic surveys.

Assuming that the catch rate is proportional to tow duration, $\mu_t = m_0 t$, then

$$ECV_t = \left\{ \frac{\frac{1}{m_0 t} + b}{n} \right\}^{1/2}. \quad (4)$$

The tow duration, t_0, when *C* is fixed that minimizes (4) is the solution of the equation

$$\frac{c_1 + t}{t(1 + m_0 bt)} + \frac{1}{\left[1 + 4(c_1 + t)C / c_2^2\right]^{1/2}} = 1 \qquad (5)$$

The sample size is

$$n_t = \left\{ \frac{\left(c_2^2 + 4(c_1 + t\ )C\right)^{1/2} - c_2}{2(c_1 + t\ )} \right\}^2 \qquad (6)$$

and

$$t\ = \left\{ \frac{c_1 + \dfrac{c_2}{2\sqrt{n_t}}}{m_0 b} \right\}^{1/2} \qquad (7)$$

Equation (5) can be solved numerically or iteratively using (6) and (7), which also defines the minimum.

As examples, the ECV_t as a function of tow duration for the ocean pout estimates is shown in Figure 13 and estimates of the optimum tow duration, $\hat{t}_o$, for estimating the mean catch-per-tow of haddock on Georges bank. If the tow duration for the Georges Bank survey was reduced from 30 to 10 minutes, then the number of stations could be increased by about 30% and the precision improved accordingly (Table 4).

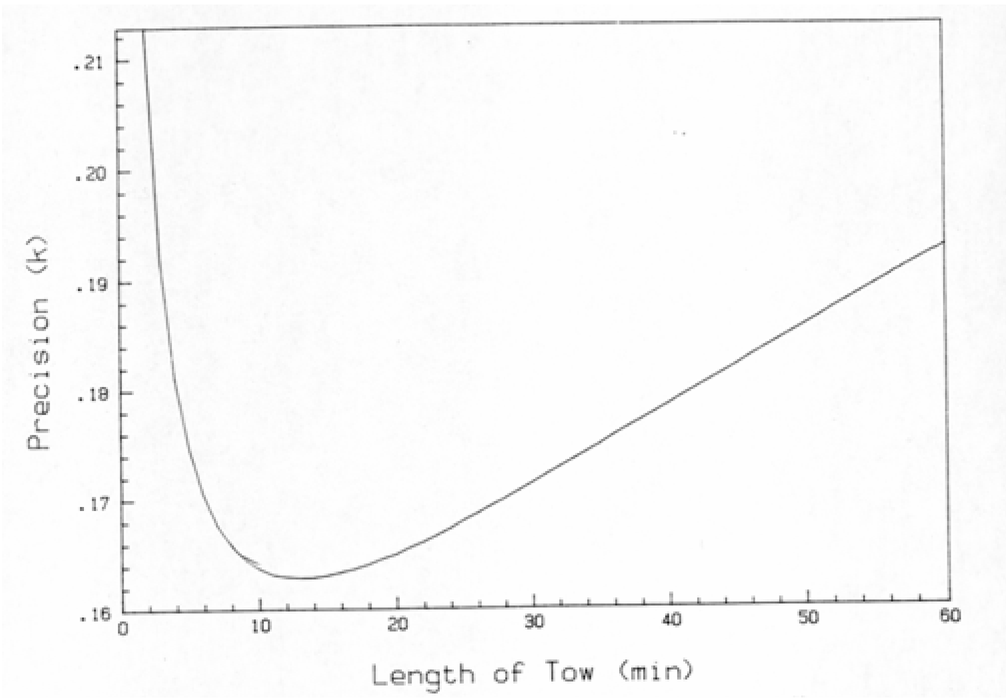


Figure 13. The precision $k = cv / \sqrt{n_t}$ versus tow length for a survey of ocean pout with fixed total cost.

Table 7. Parameter estimates for determining the effect of reducing unit size for the Georges Bank surveys. In column 10 are estimates of the resulting reduction in $(cv_m)^2 / n_t$ for density, R_1, and in the last column that for $Var(\bar{x}_r)$, R_2

Year	$\hat{\sigma}_x$	$\hat{\rho}$	$\bar{m}$	s_m	$\widehat{m_0 b}$	$\hat{t}_0$	$\sqrt{\{\hat{b}/[(\hat{b}+1)\hat{\rho}]\}}$	n_{10}	R_1	R_2
1963	16.1	.68	97.0	172.7	10.4	2.3	1.1	94	.78	.78
1964	9.7	.41	115.2	186.6	10.1	2.4	1.3	94	.78	.74
1965	7.4	.40	62.2	97.0	5.0	3.3	1.3	99	.78	.76
1966	13.6	.58	20.3	34.2	1.9	5.4	1.1	95	.80	.81
1967	10.6	.68	11.5	25.8	1.9	5.4	1.1	101	.80	.81
1968	10.1	.36	5.2	13.3	1.1	7.1	1.6	104	.82	.90
1969	17.9	.83	1.9	3.6	.2	16.8	1.0	109	1.0	.98
1970	14.1	.56	5.6	21.3	2.7	4.5	1.3	105	.79	.80
1971	25.5	.79	3.3	7.4	.5	10.5	1.0	109	.86	.86
1972	19.6	.77	7.5	4.1	1.2	6.7	1.0	110	.81	.82
1973	15.2	.55	9.5	28.7	2.9	4.3	1.3	109	.79	.79
1974	16.1	.76	2.9	6.9	.5	10.5	1.1	110	.86	.88
1975	17.5	.90	23.3	54.9	4.3	3.6	1.0	109	.75	.78
1976	7.3	.64	47.8	194.2	26.3	1.4	1.2	101	.77	.78
1977	6.3	.48	41.9	216.2	37.2	1.2	1.5	148	.76	.81
1978	18.5	.93	24.9	76.2	7.7	2.5	1.0	235	.75	.75
1979	7.0	.62	71.4	637.5	189.7	.5	1.3	229	.75	.75
1980	16.0	.89	38.5	126.0	13.7	1.9	1.0	134	.77	.77
1981	11.7	.54	11.3	23.0	1.5	6.1	1.2	106	.80	.83
1982	18.5	.71	4.8	11.9	.9	7.9	1.1	102	.83	.84
1983	18.6	.78	9.5	19.8	1.3	6.5	1.0	105	.81	.83
1984	9.9	.65	7.2	28.4	3.7	3.9	1.2	104	.79	.79
1985	13.4	.85	14.8	38.2	3.3	4.1	1.0	100	.78	.79
1986	10.2	.73	8.6	32.3	4.0	3.7	1.1	102	.79	.79
1987	18.1	.90	5.4	19.4	2.3	4.9	1.0	100	.80	.79
1988	12.5	.80	7.7	19.9	1.7	5.7	1.0	100	.80	.80

If tow duration is reduced, then the number of fish will be reduced. For example, if tow duration is reduced from 30 to 10 minutes for the Georges Bank survey, then the number of stations would increase from 77 to about 100, but towing time would decease from 2300 to 1000 minutes. Therefore, the expected total catch using 10-minute tows would be 57% less than when 30-minute tows are employed.

To determine the effect of this reduction in sample size, consider the estimate of the mean length of fish

$$\hat{\mu}_r = \frac{\sum\sum x_{ij}}{\sum m_i}, \qquad (8)$$

where $\hat{\mu}_r$ is a ratio estimator and m_i is the number of fish caught at station i. Based on some assumptions, the expected variance of (8) is given by (Pennington and Vølstad, 1994)

$$V(\hat{\mu}_r) \approx \frac{\sigma_x^2\left\{1 + \left(\overline{M} - 1 + \sigma_m^2 / \overline{M}\right)\rho\right\}}{\overline{M}n}, \qquad (9)$$

where σ_x^2 is the variance of the population's length distribution, σ_m^2 denotes the expected tow to catch variance, $\overline{M}$ is the expected catch-per-tow, and ρ is the intra-haul correlation coefficient for length. Then the tow duration that minimizes (9) subject to the constraint (3) is the iterative solution of (6) and

$$t = \left\{ \frac{c_1 + \dfrac{c_2}{2\sqrt{n_t}}}{m_0(1+b)\rho} \right\}^{1/2} . \qquad (10)$$

Comparing (7) and (10), it can be seen that if $\{b/[(1+b)\rho]\}^{1/2}$ is near 1 than the optimum tow durations for the two estimators will be nearly the same. If ρ =0, then the optimum tow duration for estimating mean length would be a single, long tow and as $\rho \to 1$ the optimum decreases. In Table 4 are estimates of the effect on the precision of estimates of mean haddock length of reducing the tow duration to 10 minutes.

Estimates of population characteristics, such as length distributions, appear not to be a function of tow duration. For example, Goddard (1997) concluded that estimated length distributions based on 15-minute tows were not significantly different than those based on 30-minute tows (Figure 14).

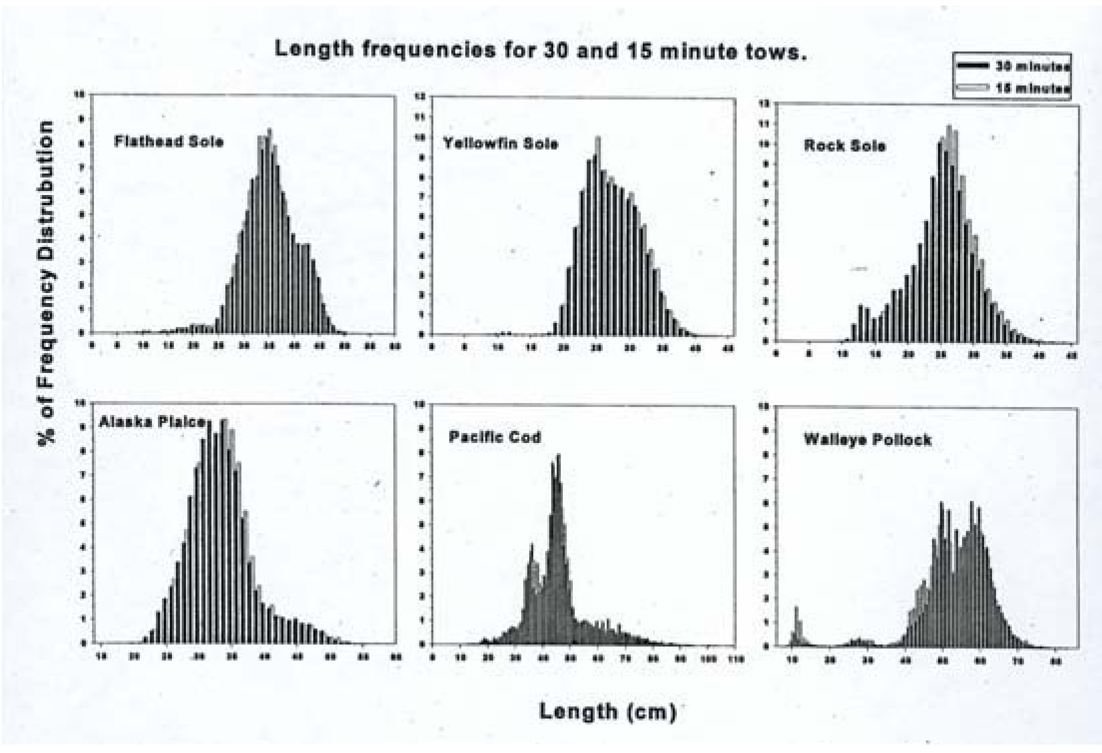


Figure 14. Estimated length distributions from a tow duration experiment in the Bering Sea (from Goddard, 1997).

In addition to increasing survey precision additional benefits from reducing tow duration include:

- Since total towing time will be reduced, there will be less gear and equipment wear and less fuel will be consumed.
- There will be fewer large catches that will have to be subsampled.
- Shorter tows can be made at more locations in survey area.
- The problem of gear saturation will be reduced
- The resultant smaller catches will require less sorting time, which will provide more time for taking other biological measurements. This is particularly relevant in the context of an ecosystems approach where it may be preferable to make measurements of parameters relating to a wide variety of species rather than just the few commercially exploited ones that have traditionally been the focus of the survey.

3.2 Methods for determining the effect of reduced tow duration: an example from western Greenland

A detailed description on the effect of tow duration on catch rates and length distributions of Northern shrimp and Greenland halibut in the West Greenland Bottom Trawl Survey is given in working document WD6 in Annex 2. The main results of that study are summarized below.

The West Greenland Bottom Trawl Survey for shrimp and fish follows a stratified random design and has been established since 1988. Standard towing time was initially 60 min. There were concerns that the continuity of the survey time series would be severely impacted in particular when making drastic changes in tow duration too quickly, i.e., from one year to the next (Kingsley *et al.*, 2002). Hence, tow duration was reduced stepwise over the years allocating the shorter duration randomly to the survey stations. 30-min and 15-min tows were introduced in 1991 and 1999, respectively, 60-min tows were replaced by 45-min tows in 2000 and since 2001 solely 30- and 15-min tows were used 2001 (Wieland *et al*., 2004). However, abundance of northern shrimp and Greenland halibut in particular has increased considerably off West Greenland in the most recent years (Wieland *et al.*, 2004, Storr-Paulsen and Jørgensen, 2004). As a consequence, 30-min tows often result in large catches, which are difficult to handle and for which time-consuming subsampling procedures have to be applied. It would be desirable to shorten tow duration to 15 min on all survey stations in order to reduce the need for subsampling as well as to gain the opportunity for an increase of the total number of stations from which an improvement of the precision of the survey estimates is expected. Hence, catches of northern shrimp and Greenland halibut from 15- and 30-min tows have been analysed to examine whether a reduction of tow duration to 15 min on all stations would influence the catch per swept area, its precision and the size distribution of the two species.

Tows conducted during the routine part of the survey in the years 1999 to 2004 were grouped into two intervals of 15-min and 30-min with a tolerance of 10% of the reported towing time, and tows of other duration were discarded. Strata for which at least two hauls in each group of tow duration have been available in a given year were selected for further analysis. This resulted in an initial data set of 185 15-min and 217 30-min tows from 18 strata and 6 years. At these sampling locations, which were distributed over a large part of the survey area, only few zero catches of northern shrimp occurred and 43 pairs of stratum and year combinations were used for analysis. For Greenland halibut, however, a considerable number of zero catches were recorded in the southern part of the study area and limiting the analysis cases for which at least two non-zero catches were available for each tow duration reduced the data set to 160 15-min and 197 30-min tows and 37 pairs of stratum and year combinations.

Catch data by haul were analysed using a Generalized Linear Model (GLM) approach (McCullagh and Nelder, 1989):

$$\log_e(CPUE_{ijk}+1) = Stratum_j + Year_k + Depth_i + TowDuration_i + \varepsilon_{ijk}$$

and

$$\log_e(CPUE_{ijk}+1) = \log_e(MeanCPUE_{jk}) + TowDuration_i + \varepsilon_{ijk}$$

where *CPUE* is the catch of each tow divided by its swept area and ε is the error term. The errors were assumed to be independent and identical distributed, i.e., with the mean $\mu = 0$ and the variance σ^2. Stratum and year were considered as factors. Depth was included as a continuous variable to account for additional within-stratum variability of depth for the different tow locations.

Biomass densities (in kg/m^2), which include all length intervals, were used for both species as CPUE. Numerical densities (in numbers/km^2) were used for three sexual stages of northern shrimp. Northern shrimp is a protandric hermaphrodite, in which males, primiparous and multiparous females correspond to groups of increasing mean carapace length (Wieland *et al.*, 2004). For Greenland halibut, numerical densities (in numbers/km2) were categorized into three length groups: < 20 cm, 20–40 cm and > 40 cm.

Within stratum and year means and within stratum and year standard deviations of biomass densities were computed, and a pair wise comparison of the standard deviation over mean ratios for the two tow durations (paired t-test) was used to evaluate the effect of tow duration on the precision of the density estimates.
Fish and shrimp are usually caught in clusters resulting in a non-independence of length measurements within a haul (Pennington and Vølstad, 1994, see also section on cluster sampling). To account for this, the effect of tow duration on mean length was studied following the approach by Godø *et al.* (1990). Population mean lengths from clustered observations can be estimated as

$$\mu = \sum c_i x_i / C$$

where c is the number of individuals, x is its mean length in the ith haul and C is the total number in the n hauls (Cochran, 1977). However, as the number of hauls has been small, jackknife estimates of mean length and its standard error were calculated according to Cochran (1977) where

$$\mu_{(i)} = \sum_{i \neq j} c_i x_i / (C - c_j)$$

is the weighted mean length deleting the nth haul and

$$\mu_{(.)} = \sum \mu_{(i)} / n$$

is the estimate of the population mean length, and

$$se = \sqrt{((n-1)/n)\sum(\mu_{(i)} - \mu_{(.)})^2}$$

is the corresponding standard error.
The effect of tow duration on mean length was then examined using the model

$$L_{mean,ijkl} = \mu_{(.)\,jk} + TowDuration_i + \varepsilon_{ijk}$$

where L_{mean} is the mean length in each tow and $\mu_{(.),jk}$ is the jackknife estimate of mean length in stratum *j* and year *k*. Relative standard errors ($RSE_{jk} = se_{jk} / \mu_{(.)\,jk}$) were computed to evaluate whether mean length was adequately estimated by the given number of observations or whether certain strata and year combinations were more inhomogeneous than others and should be excluded from the analysis. Two threshold levels of the relative standard error of the jackknife estimates of mean length were defined arbitrarily as $RSE_{jk} < 0.075$ and $RSE_{jk} < 0.050$.

To examine the effect of tow duration on maximum observed length, the largest observed length in the tows belonging to the different combinations of stratum and year were used for a pair wise comparison of the two classes of tow duration (paired t-test).

Overall average catch rates of 15-min tows were higher than for 30-min tows for both, northern shrimp and Greenland halibut in most years and for all years combined. However, no significant effect of tow duration on the catch rates were detected, irrespectively whether total biomass density or numerical density of the different size categories for the two species were considered. Normal quantile plots of the residuals for northern shrimp and Greenland halibut do not indicate that the models for the catch rates were inappropriate. Tow duration remained also non-significant in models in which the stratum and year were replaced by mean densities for the respective stratum and year combinations. It is therefore concluded that 15-min tows are as efficient as 30-min tows to measure the density of all sexual stages of northern shrimp and different size categories of Greenland halibut, and that the higher level of catches rates in 15-min tows were due to the significant effects of other factors, e.g. sampling location.

Average ratios of within stratum and year standard deviation of biomass density and within stratum and year mean biomass density were 1.146 for northern shrimp and 1.072 for Greenland halibut. For both species, the average ratios for 15-min and 30-min tows were rather similar (Northern shrimp: 1.148 and 1.143, Greenland halibut: 1.116 and 1.027) and no significant difference between the two tow durations was found (paired t-test; Northern shrimp: $t = 0.048$, d.f. = 42, $p = 0.962$; Greenland halibut: $t = 1.000$, d.f. = 36, $p = 0.324$). The result of the statistical analysis did not change when the minimum number of observation in each stratum, year and tow duration combination were increased from two to four (Northern shrimp: $t = -0.035$, d.f. =14, $p = 0.851$; Greenland halibut: $t = 0.103$, d.f. = 14, $p = 0.920$). Hence, there is no indication that 15-min tows give less precise results than 30-min tows.

Mean carapace length of northern shrimp ranged from 15.8 to 23.6 mm in the different strata and years. Mean total length of Greenland halibut varied between 14 and 39 cm. Large relative standard errors for the jackknife estimates of mean length were observed in several cases for both species, but in particular for Greenland halibut. This indicates a considerable within-stratum variability of mean size and that the number of hauls in some strata was too small for a reliable estimation of the overall mean population length. In addition to the inclusion of all possible hauls, i.e., all non-zero catches, the analysis was therefore also done for reduced data sets, in which the most inhomogeneous strata were removed. Here, thresholds of the relative standard error for the jackknife estimates of mean length for a given stratum and year combinations were applied. In all cases, however, the analysis of variance did neither for northern shrimp nor for Greenland halibut reveal a significant effect of tow duration on the mean length. For northern shrimp, the normal quantile plots of the residuals do not indicate that the models were inappropriate. This was also the case for Greenland halibut although the normal quantile plots were less satisfactory, which might be related to a much lower number of observations. However, the length frequency distributions of both species were apparently not affected by tow durations of 15 and 30-min.

Maximum observed length of northern shrimp and Greenland halibut in the different strata and years were highly variable for both, the 15-min and 30-min tows. No significant effect of tow duration was found (paired t-test; Northern shrimp: $t = -0.682$, d.f. = 42, $p = 0.499$; Greenland halibut: $t = -0.020$, d.f. = 36, $p = 0.984$). This suggests that also extreme values, i.e. the largest individuals, can be sampled adequately by 15-min tows.

Previous studies of the effect of tow duration have shown that the mean sizes of several flatfish and gadoids were not affected by tow durations from 60 to 5-min (Godø *et al.*, 1990; Walsh, 1991), and similar results were reported for three crab species comparing 30 and 15-min tows (Somerton *et al.*, 2002). This indicates that the effect of tow duration, if there is any, is the same for all sizes, which is consistent with our findings for northern shrimp and Greenland halibut.

Carothers and Chittenden (1985) found for two species of penaeid shrimp a significant relation between catch and tow durations of 5 to 30-min, but reported also that tow duration accounted for only a small proportion of the total variation in catch. Godø *et al.* (1990) and Walsh (1991) observed that catch per unit effort (CPUE) of flatfish and gadoid species increased significantly with decreasing tow duration only at tow durations below 15-min and remarked that higher catch rates of short tows were difficult to explain. Somerton *et al.* (2002) measured significant higher CPUE values in 15-min tows than in 30-min tows for two out of the three crab species studied, but it was not possible to identify the definite causal mechanism for this result.

The present study was not designed to detect mechanism that are independent of tow duration such as catch-by-surprise due to herding or errors in the measurements of the length of the tow path (Godø *et al.*, 1990) or escapement below the footrope (Walsh, 1992; Somerton *et al.*, 2002), which would effect the results from short tows relatively more than those from long tows. However, the present analysis of the mixture of 30 and 15-min tows randomly allocated to the sampling locations within a stratum in the West Greenland Bottom Survey show that differences in catch rates of northern shrimp and Greenland halibut were due to stratum and year effects rather than caused by tow duration. This implies that a bias introduced by using only 15-min instead of 30-min tows due to an 'end-effect' is rather small and that other sources of variation, such as the within stratum differences of depth at the various sampling locations, are much more important. The fact that the existence of en end-effect could not be demonstrated here does not necessarily mean that such an effect does not exist and could result in a systematic bias. Kingsley *et al.* (2002) estimated that the amount of northern shrimp caught outside the nominal tow period equals to 2.78 min additional towing time, which corresponds to about 9 % of a 30-min tow but to 18 % of a 15-min tow. This so-called 'end-effect' has been estimated with a high uncertainty (relative standard error: 42 %) and the relevance of such an effect was not confirmed by a later study (Kingsley, 2001). Moreover, the results of the present study indicate that the magnitude of the 'end-effect', if existing at all, is rather small and variable. Hence, the risk of introducing a bias to the time series of biomass estimates from the survey due to a reduction of tow duration to 15-min appears to be negligible.

Furthermore, no indication was found that 15-min tows give less precise estimates of biomass and numerical density for northern shrimp and Greenland halibut than 30-min tows, and there was also no significant difference between the two tow durations concerning their efficiency to sample extreme values such as maximum length. These conclusions, however, were derived from paired t-tests, and the lack of difference found here should be taken with some caution as its power depends very much on the number of observations (Sokal and Rohlf, 1995).

In summary, we conclude that the actual mixture of 15 and 30-min tows in the West Greenland Bottom Trawl for shrimp and fish can be replaced by 15-min tows on all stations without interrupting the time series of survey estimates. The implementation of a standard towing time of 15-min would be advisable because it reduces the frequency of large catches, which are time consuming to handle, and because the gain in survey time related to the shorter average tow duration could be used for an increase of the total number of stations in order to improve the overall performance of the survey.

3.3 Estimating trawl capture before and after official haul duration

3.3.1 Introduction

Recent investigations have shown that CPUEs for fish and crustaceans can be relatively higher for 15-mins hauls compared to 30-mins hauls (Godø *et al*., 1990; Walsh, 1991; Somerton *et al.*, 2002). These observations are used as arguments to systematically shorten haul durations for scientific trawl surveys, because if hauls are shorter, the number of hauls can be increased, resulting in an increased precision of abundance index estimates (Folmer and Pennington, 2000; Kingsley *et al.*, 2002; Pennington *et al.*, 2002). However, the reasons for proportionally increased catches in shorter hauls are not obvious and several hypotheses can be put forward: i) fish escapement is lower at the beginning of the haul as individuals are surprised by the arriving haul, ii) fish are caught before and after the official haul duration, iii) the net saturates for longer hauls which increases escapement. First, evidence for the "surprise" effect comes from visual observations at the trawl opening. Albert *et al.* (2003) observed that proportionally more Greenland halibut entered the trawl in the first couple of minutes on the sea floor compared to the later parts of the haul. Second, the time elapsed after arrival at the sea floor but before trawl geometry is stabilised, which is often taken as the nominal starting time, and while hauling, has been estimated to be non-negligible for shrimps in Greenland waters (Kingsley, 2001). Third, net saturation will probably occur in certain circumstances but should not generally be a problem. Thus, the two most plausible hypotheses, not excluding that other factors will also contribute, are a surprise effect and fishing before and after the nominal haul duration. Both effects are independent of haul duration and thus might be proportionally more important for shorter hauls. These negative effects might counter balance the expected benefits from shorter hauls. However, the second problem might be overcome by technology, i.e., using trawls that are open while lowered and closed before hauling.

3.3.2 Case study

In 2003, a study was carried out to estimate the importance of fish catches before and after the nominal haul duration. For this, 6 30-min hauls were carried out. In addition, for each full haul, 3 "zero-duration" hauls were carried. In zero-duration hauls, the trawl is retrieved as soon as the nominal haul duration would start in an ordinary haul. These zero-duration hauls were located approximately at the beginning, middle and end of the total length covered by the full haul. All catches were identified, weighed and counted. For 9 species enough individuals were caught to allow comparison between full and zero-duration catches.

An analysis of variance on the log-transformed catch numbers (ln(catch+1)) showed that the location of the zero hauls along the length of the nominal haul did not matter, i.e., the three zero-duration hauls can be considered as replicates. This was true for all species.

Table 8. Catch ratio (numbers) zero-duration / full hauls

SPECIES	MEAN	STD
Arnoglossus laterna	0.25	0.37
Callionymus lyra	0.16	0.2
Dicologlossa cuneata	0.07	0.09
Merluccius merluccius	0.34	0.64
Merlangius merlangus	0.19	0.19
Sepia officinalis	0.09	0.16
Solea solea	0.05	0.06
Trachurus trachurus	0.18	0.35
Trisopterus luscus	0.22	0.5

zero-duration hauls can be considered as replicates. This was true for all species.

Comparative plots indicated positive correlations between zero-duration and full haul catches (Figure 15). The catch ratio was strongly species dependent (Table 8). For highly mobile species such as hake (*Merluccius merluccius,* MERLMCC in Figure 15), the proportion caught in "zero-duration" hauls was rather high while for benthic species such as sole (*Solea solea*, SOLEVUL in Figure 15) it was below 10%. In conclusion, for less mobile species the catch due to fishing before and after the nominal haul duration might be proportional to the duration, while for highly mobile species it can be hypothesised that the surprise effect is more important. This later observation sheds strong doubts on swept area based abundance indices for highly mobile species.

3.4 Conclusions on tow duration

There is evidence that a reduced tow duration can increase the precision of a survey by allowing time to collect more samples at the expense of collecting longer ones. This may to be specific to certain conditions such as the species, areas, (whole) gear, and time (of day or year). Survey planners should be encouraged to examine the possibility of reducing the tow duration by conducting the types of experiments described in Section 3.2. If and when it can be demonstrated that reducing tow duration has the expected effect increasing precision, then that reduced tow duration should be employed and the extra time allocated to obtaining more samples.

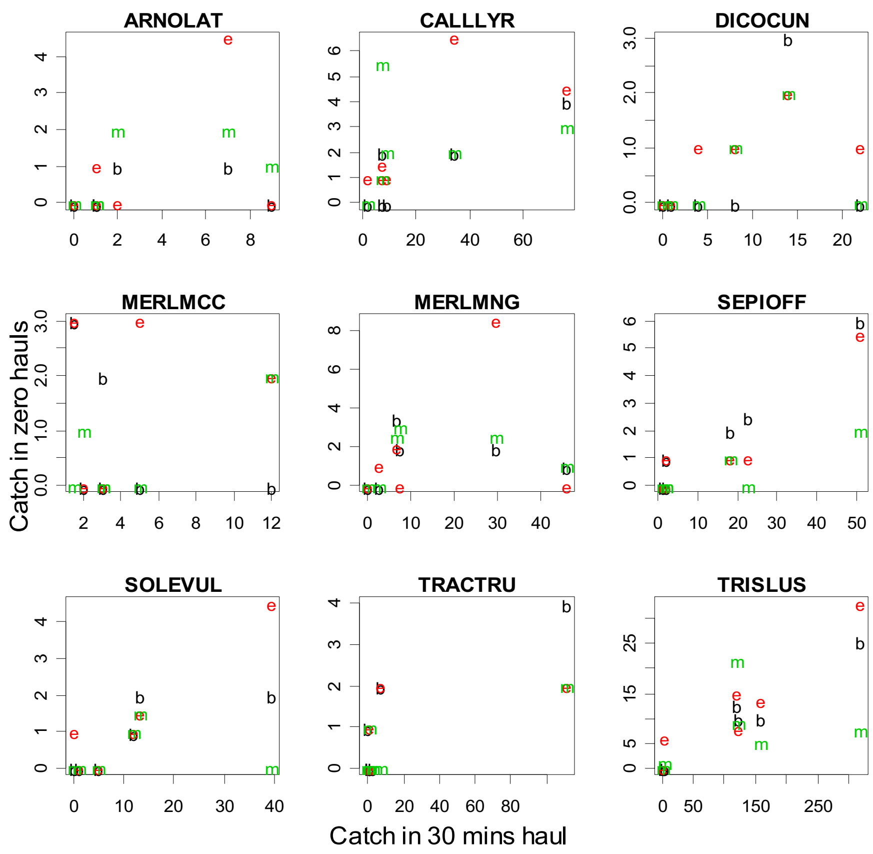


Figure 15. Comparison of catches in 30-mins and "zero-duration" hauls for nine species. Location of zero-duration hauls: b=beginning, m= middle, e=end of full haul.

4 Analysis of covariates

Covariate information can be used to improve both survey design and analysis. For example, habitat classification could be used to improve survey design through better stratification, if species distribution is related to the habitat classification. Habitat classification could also be used to improve survey analyses through post-stratification. It should be noted that the potential for improving multi-species survey design using covariate information is limited because it is unlikely that the same or similar covariate relationships exist for all species of interest. However, improved analyses may be feasible. The degree of improvement in survey precision and accuracy is dependent on the strength of the covariate relationship and the amount and accuracy of the covariate information for the surveyed area. Covariate information can also provide useful information on possible causes of inter-annual variation in mean abundance and other parameters.

Some covariates may not be fixed within the time-frame of a survey. If information on such covariates could be obtained during the survey then there is potential for using an adaptive sampling design to improve abundance information. For example, one could decide to adaptively increase the number of samples in an area if the temperatures were found to be favorable for the species being surveyed. There is a potential for considerable improvements in survey performance if parameters that influence the target population's 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 objectives or high numbers of target species (e.g., multi-species surveys) are not likely to be strong candidates for greater refinement in survey design. Improving survey precision for one species may reduce survey precision for other species. The ability to address multiple objectives including those that may be conceived in the future is likely enhanced by relatively simple survey designs. At 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.

It is also worth considering the use of covariate data in terms of reducing the variation in catch data through improved standardization of the sampling unit. This is particularly the case for demersal trawl surveys where stability and catchability of the sampling unit, or trawl, changes constantly in relation to bottom type, environmental conditions as well as fish behaviour. The premise for many research surveys is standardised catchability of the gear, and where parameters affecting catchability cannot be standardised such as depth, weather, time of day and more recently bottom contact and speed of water through the net, these parameters are often recorded in great detail. Although many of these variables are reviewed in real time by the senior scientist on the bridge, and kept within 'working' tolerances, if they were combined in a multivariate analysis they might still convey further benefit to post survey analysis. There are at least two ways this could happen.

First, some tows may appear as outliers where several parameters, while within an agreed tolerance, may combine to produce an extreme sampling situation. These tows could be removed or weighted prior to calculation of survey estimates. Secondly, a time series of these parameters may provide some relative index of catchability. For instance, where a survey coincides with a period of unusually poor weather, it is likely that catches will be reduced overall and precision is likely to appear to improve. However, ground contact, sea state and other parameters are likely to worsen, adversely affect catchability and those intending to use the data should be aware of this. The assessment of the efficiency of the sampling unit, other than purely number of minutes on bottom, could be provided along with estimates of survey precision. This could act as a check that a change in precision is real and not simply a particularly bad survey year or, worse, a positive or negative bias over time.

4.1 Evaluating the impact of survey design and environmental variables on survey abundance estimates

For many species, survey density estimates vary significantly between years, often more than what is expected based on biological theory. The question is whether estimates of variance for such density estimates are too small or whether a systematic, but varying, bias could be identified. If the problem is caused by biased estimates, some covariates might exist to explain the bias. For example, fish availability or more generally, survey catchability could vary between years. The effect of such a survey catchability would be expected to affect several species simultaneously. A study was carried out to investigate the relationship between survey indices (density estimates, mean weight and coefficient of variations of density estimates) and covariables describing survey design and conditions for the Bay of Biscay groundfish community (1987–2003). All survey indices were normalised across years by species to remove species effects (53 species which occurred in at least 5% of hauls on average). Survey design conditions were described by the starting date, the mean haul location and the number of coastal hauls. Environmental conditions during the survey period were represented by various indicators describing wind conditions (average speed and direction, standard deviation of half-daily wind speed and wind persistence).

In order to explore annual catchability effects due to survey conditions, the relationship between individual indicators and survey condition variables were investigated using generalised additive models (GAM). All explanatory variables except wind direction (3 level factor) were modelled by non-linear relationships using regression splines (mgcv, , minimizing generalised cross validation, package in R). Explanatory variables were selected based on significance tests and visual inspection of the form of the fitted relationship. Only variables for which the relationship was significantly different from zero over at least part of the parameter range were retained. Species were divided into three groups, depending on their global trend over the study period: group 1 increasing, group 2 stable and group 3 decreasing in density.

The normalised density indices of the three species groups differed as expected by their time trend patterns (Figure 16a). Group 1 increased in steps, group 2 fluctuated strongly with no trend and group 3 decreased generally but in waves. All three groups displayed an abrupt increase in the early 90s. Group 1 and 2 also showed a similar decrease at the end of the series. The best model explaining these interannual variations contained the number of coastal hauls (NHC; $p=0.0005$) and the wind speed variability (MDW.std; $p <0.0001$) as linear covariates and average wind speed (MDW; $p=0.001$) and survey starting date (date; $p<0.0001$) as smooth functions. Normalised abundance slightly increased up to about an average wind speed of 17 and then levelled off (Figure 16a). Earlier survey starting dates seemed to lead to somewhat higher average abundances (Figure 16b). Overall this model explained 19% of deviance. This has to be compared to the model including only year effects, which explained about 36%. Thus survey and wind conditions can explain about half the global interannual variation in species abundance. The remaining variability is due to other factors and species specific causes such as non-synchronous recruitment variations. When adding time trends by species group to the best model, these time trends where much smoother than the raw data trends (Figure 16 b). The change was most visible for species group 2. The originally strongly fluctuating trend became a smooth increasing trend, when covariates were included. Thus it is possible that due to the slight changes in survey design and resulting wind conditions, the global increasing trend of this species group was hidden in the raw abundance indices. In contrast, the originally decreasing trend of group 3, seemed to be less so once the survey conditions were taking into account.

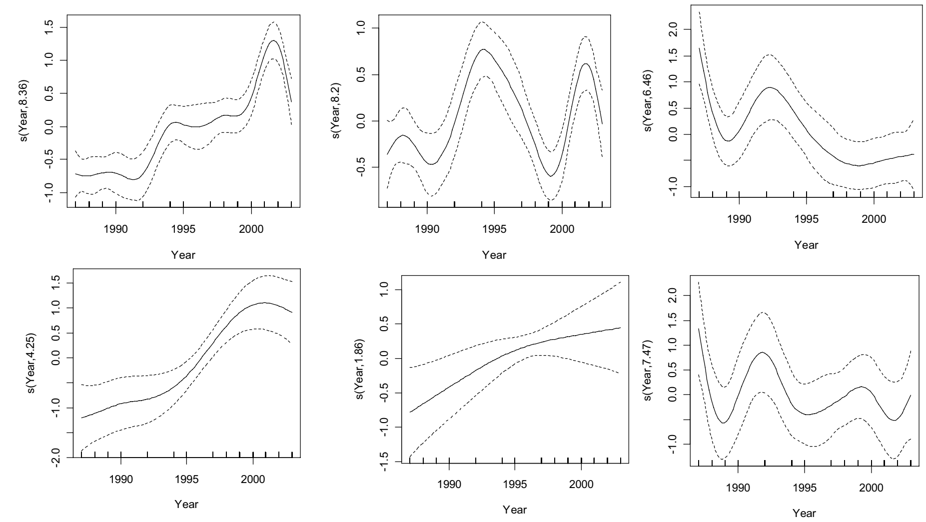


Figure 16. Smooth relationships between abundance estimates and survey year by species groups. a) Relationships for null model without covariates (upper row); b) Relationships when adding year to the best fitting model (lower row). Inset marks along the x-axis indicate position of data points.

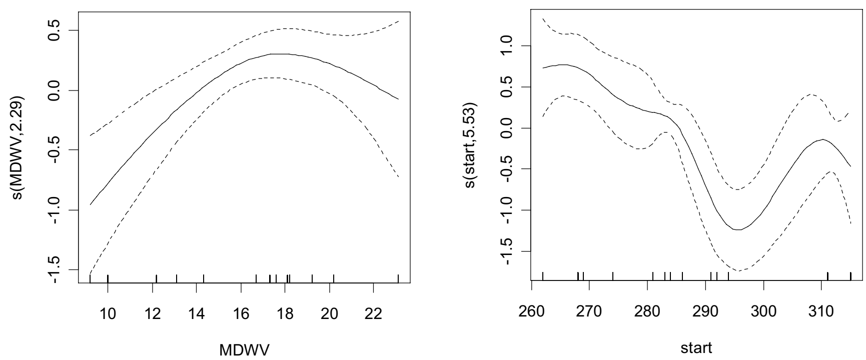


Figure 17. Smooth relationships between abundance and average wind speed (MDWV, left panel) and survey starting date (right panel) for best fitting model. Inset marks along the x-axis indicate position of data points.

Average species weight was significantly explained by an increasing function of survey starting date but none of the other covariates. CVs of density estimates depended only on wind direction. In conclusion, survey design and wind conditions can explain about half the interannual variation common to all species in survey density indices. For certain species (group 2), this effect might even hide a true underlying increasing trend.

5 Methods of combining surveys

5.1 Combining acoustic and bottom trawl data: lessons from the CATEFA project

The CATEFA project set out to determine links between simultaneously collected trawl and acoustic data for stock assessment purposes. Nineteen bottom trawl surveys with coincident acoustic measurements comprising of five different survey series were selected:

1) The ICES co-ordinated International Bottom Trawl Surveys (IBTS) in the North Sea. They follow a random design stratified by ICES rectangle (Figure 18b).

Trawls and acoustic data are only taken in daylight hours. The surveys used were those carried out by CEFAS (2000, 2001 and 2002), FRS (1999, 2000 and 2002) and IFREMER (2002 and 2003). Each survey comprises between 60 and 80 hauls.

2) The Northern Irish Bottom Trawl Surveys (NIBTS) in the Irish Sea. These surveys are mostly small (20 or 30 hauls). They follow a random sampling design stratified by depth and substrate (Figure 18c). Depth varied between 25 and 150 m. Five surveys carried out by DARDNI were available: autumn 1997, spring 2000, spring and autumn 2001 and spring 2002.
3) The combined acoustic and bottom trawl surveys for cod and haddock in the Barents Sea – carried out by IMR Bergen. Sampling follows a regular grid with a haul every 20 n.mi. (Figure 18d) The number of hauls varied between 200 and 300. Available surveys were 1997, 1998, 1999, 2000, 2001 and 2002.

The challenge in correlating acoustic backscatter to bottom trawl catch probably lies in their different catchabilities or efficiencies. For instance, when echograms are scrutinized carefully some puzzling conclusions emerge: it is often possible to observe big catches in the trawl but very little or nothing whatsoever on the associated echotrace and one must recognise the limitations of both sampling approaches in the context of estimating demersal fish assemblages.

The swept area of bottom trawls varies as a function of depth, gear type, the amount of warp paid out and the doors' angle of attack. Similarly, beam ‘footprint’ of an echosounder varies geometrically according to beam angle and depth. Figure 19 compares the increase in the acoustic footprint with depth using a transducer with a 7 degree beam angle and the door spread of the trawl gear setup used in CATEFA. This is complemented by a scaled representation of the average situation of respectively the North Sea surveys, the Irish Sea surveys and the Barents Sea surveys (Figure 20) where the trawl to vessel distance is also taken into consideration.

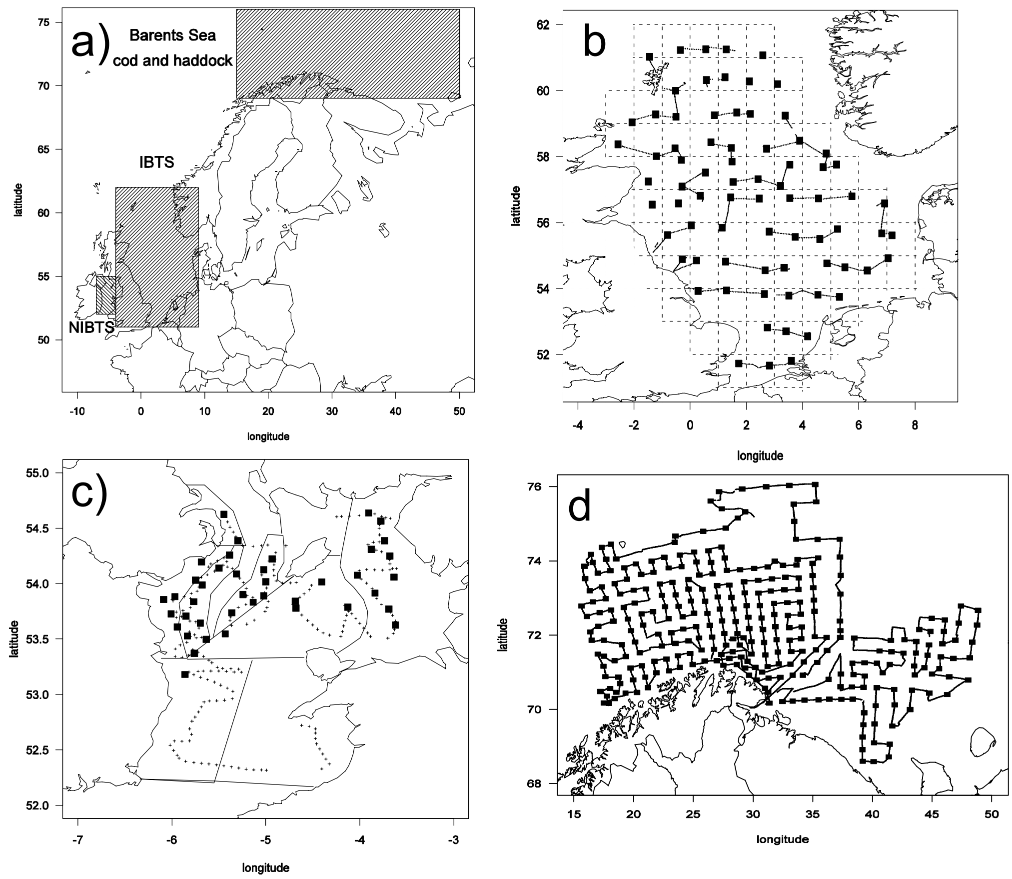


Figure 18. Study areas (a) and sampling schemes for survey series (b) IBTS (c) NIBTS, and (d) the combined surveys in the Barents Sea. Solid squares represent stations. Small dots represent between stations recordings. They appear as lines when the density of between stations observations is large.

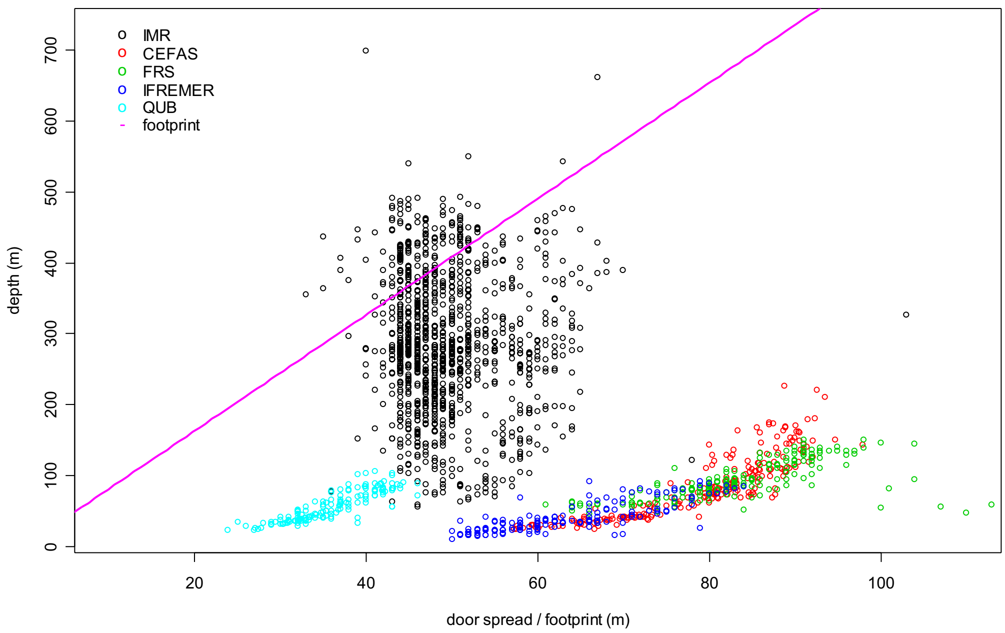


Figure 19. Comparison between acoustic footprint for a 7 degrees beam angle and door spread reported for the different partners.

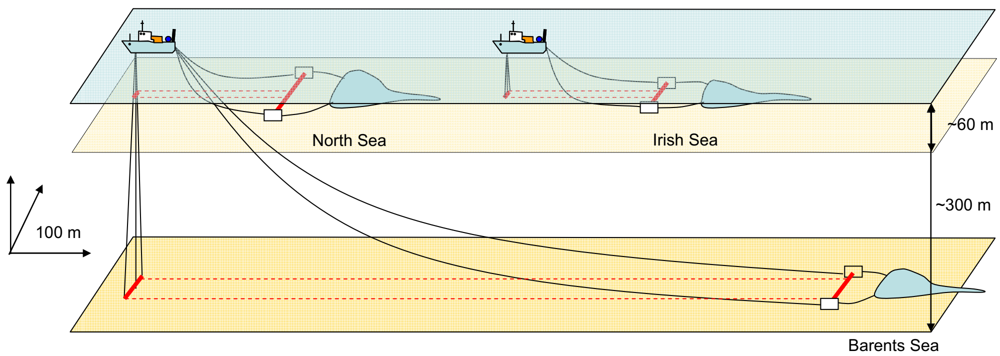


Figure 20. Relative size and location of acoustic beam and trawl track. Case of North Sea and Irish Sea as opposed to Barents Sea

The acoustic beam generally will sample a smaller area of the seabed when compared to the bottom gear, except for samples in the Barents Sea, taken in deeper waters. This is likely to introduce discrepancies between acoustic observations and net catch.

Considering the sources of perturbation and differences presented above, the echosounder is much less obtrusive. However, assessing demersal species using acoustic methods depends strongly on their vertical distribution or aggregation behaviour. The better correlation found with some demersal species in the Barents Sea as opposed to that found in the North Sea provides clear evidence of this, as in the first case the species aggregate in schools off the seabed, whilst in the latter such aggregations do not occur.

On the other hand, the acoustic method allows for a better appraisal when large and dense aggregations occur, as the fishing gear will tend to only capture a proportion of these and possibly miss-represent the true density. In such cases it might be the case that the acoustic index could offer some correction value.

Four different methodologies have been developed with applications based on the CATEFA database. However, all the methodologies have not been tested on the complete set of 19 surveys neither on the various surveys types (Barents Sea, North Sea, Irish Sea). This renders an exhaustive comparison of the behaviour of the models' performances biased. In general *direct* relationships between trawl and acoustic data were weak, especially in the southern North Sea, while patterns or trends revealed *indirectly* were often very similarly shaped. By the term 'indirect' we mean that when trawl and acoustic data were plotted, or modelled, independently against another covariate such as depth, the pattern of dependence was very similar. For the North Sea surveys, when statistical procedures were used to select auxiliary variables with the strongest explanatory power (GAM, Fuzzy Model), the acoustic variable was generally not retained. At the extreme, a fuzzy model was suggested with only longitude as the explanatory variable, hence with poor estimation power.

Other approaches were based by construction on the use of NASC data. Artificial Neural Network did get an input node for NASC data, double sampling postulated some (linear in its simple version and non linear in its generalised version) point to point relationship between trawl and acoustic data, and geostatistics looked for spatial correlations between the two variables.

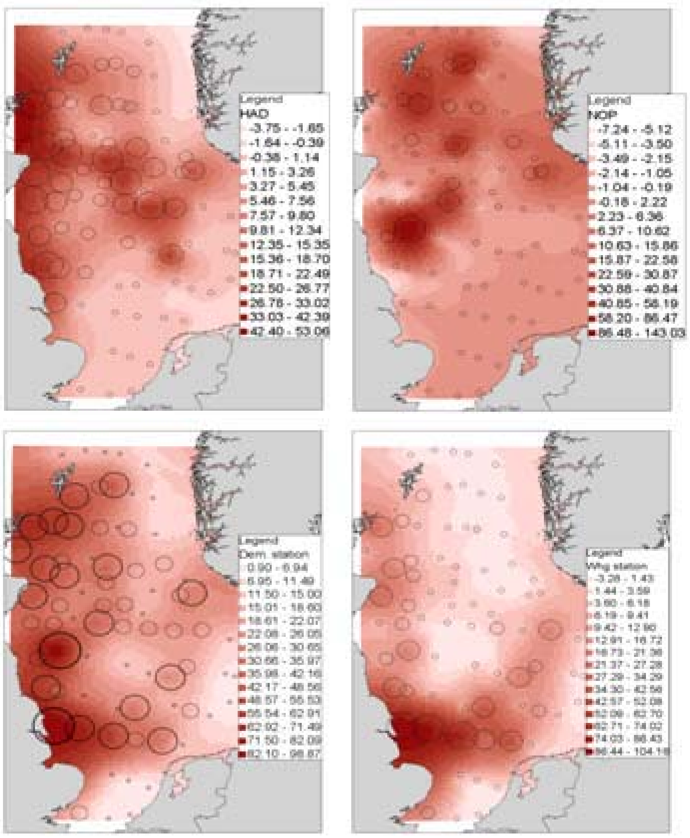


Figure 21. Survey maps produce by ANN. Trawl data are denoted by black circles overlaid onto a contour map of the interpolated between station trawl data (haddock, Norway pout, demersal, whiting).

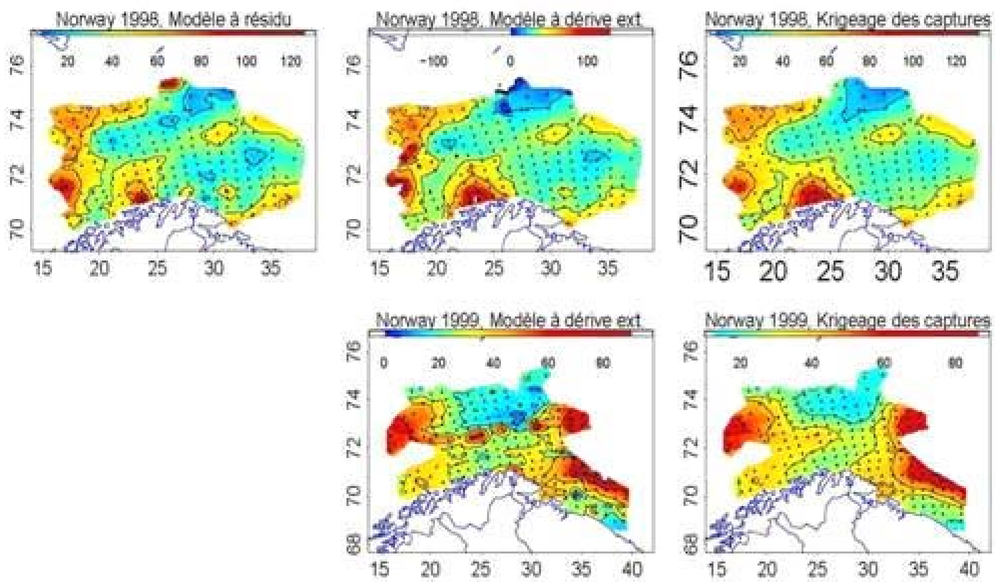


Figure 22. Estimation of demersal equivalent NASC. Barents Sea surveys. Left: simplified co kriging. Middle: kriging wih external drift. Right: simple kriging of trawl data. The colour scales are identical for each year but different from year to year.

All methods did end up with the same qualitative conclusions:

- In general, noise and signal are difficult to disentangle in the relationships between trawl and acoustic data. This is less true as the number of trawling stations (i.e. statistical conditions) increases.
- The behaviour of the pair of variables (trawl and acoustic) is well appraised in the Barents Sea surveys, reasonably well appraised in the Irish Sea surveys and poorly appraised in the North Sea surveys (with some punctual exceptions).
- Combination leads to improved indices in the Barents Sea and Irish Sea case studies (same order of magnitude for the index but larger confidence). Combination is not operational in the North Sea (with some punctual exceptions), at least under the present CATEFA framework (layered NASC values over large ESDUs).

Initially, the acoustic data were treated as a single overall integrated value. No attempt was made to partition by species or other grouping. Given the weakness of the relationships, it was decided to investigate partitioning methods to see if these would improve the relationships. Two approaches were adopted, both derived from standard practice in other acoustic surveys. In the first approach, the acoustic data were partitioned according to the proportion of a given species or grouping in the catch. This was able to improve the relationships to some extent, but not to a level where they could prove useful. The second approach was to directly assign echo traces to species, based on the trawl catch and experience – the "scrutiny" approach. With the exception of pelagic species, this proved unsuccessful, probably as a result of the difficulty of certain identification of echo traces to species.

Given that most of the demersal fish targeted by the trawl survey are found close to the seabed, a further alternate approach was investigated. This involved a careful re-analysis of the acoustic signal to include data as close to the seabed as possible. Again, this was capable of improving the relationships to some degree, and explaining some of the mis-matches between trawl and acoustic data. However, the improvement was insufficient to make all the relationships useable.

In situations where there are useable relationships between on station acoustic data and the trawl catch, the next step was to examine the relationships between acoustic data on and between stations. The null hypothesis in this case was that there would be no significant differences seen between the acoustic data in the two situations. Investigations of this showed that generally the acoustic observations were similar on and between stations, both in terms of NASC and of depth profiles. This finding makes it possible to use relationships at the trawl station to infer fish abundance and distribution away from the trawl stations

Having derived useable relationships between trawl and on station acoustic data, and established that on and between station acoustic data were likely to be observations of the same phenomenon, it was then be possible to derive combined indices. The advantage of the combination would be that it allows many more observations (real trawl data and *virtual* trawl data derived from the acoustic data between station and the observed relationships). This has been attempted in a number of different ways. The results showed broadly similar biomass levels to the traditional index, but with improved estimation variances and sometimes a different pattern of years of high abundance.

Fuzzy logic models used longitude and depth as covariates as well as the acoustic data. The acoustic data actually had the least explanatory power. The resulting combined indices tended to have a much lower average level than the catch based indices, although with a reduced variance. One possible explanation for the differences was the strong influence of a small number of catches with many fish.

Similar results were found using the Artificial Neural Net (ANN) approach (Figure 21). Reasonable models were found for the Barents Sea data, but for the North Sea data, the ANNs performed less well. In both cases however, the models tended to underestimate the fish abundance compared to the trawl data. Again, this could be partially attributed to a failure to capture the rare, high amplitude observations.

In the case of the Barents Sea cod and haddock, the relationships between trawl and acoustic data were much more robust than in the North Sea, and a combined estimate more supportable. The on and between station acoustic data were therefore combined using spatial models (Geostatistics, Generalised Double Sampling). The resultant indices showed reasonable correlation with the trawl based indices, but had sometimes a higher estimation variance.

All the different approaches highlight one of the intrinsic difficulties in evaluating output indices. The trawl survey indices are often taken as the “truth”, and the failure of the new index to match this can be taken as “failure”. However, here are also reasonable grounds for being unsure about the accuracy of the pattern shown by the trawl data alone, e.g. age structures and trawl geometry. Further evaluation would be possible by comparing the two indices performance in an assessment model, but given the uncertainties associated with the combined index this was not considered valuable. Other features that can be distilled include the tendency of any modelling approach to “smooth” data leading to a failure to capture the high amplitude observations, and the general weak explanatory power of the acoustic data.

Finally, in considering the results of this project it is important to remember the basic rationale for the approach. Bottom trawl surveys have been part of the stock assessment process for many years, and provide valuable input data, particularly for demersal fish. However, it is widely recognised that the results have quite a high associated variance, and are possibly biased. The surveys are carried out for this purpose anyway, so the collection of acoustic data during them represented a cheap and potentially useful additional data source. The question was, therefore, can we use these data to improve the quality of the indices derived from these surveys.

All the analytical tools deployed to study the relationships were able to describe some relationships between the acoustics and the trawl data. However, with the notable exception of the Barents Sea data, these relationships were weak and had little explanatory power. The original premise that the trawl and the acoustics were sampling the same phenomenon was therefore correct, but there were clearly many other factors that influence the results of the two approaches and these tended to confound any relationships. The analytical tools were also all capable of generating combined indices, but these often deviated considerably from the trawl derived indices for the same reasons. So the answer to the question asked, at least in the case of the North & Irish Sea surveys is; no we cannot currently use the acoustic data to improve the quality of indices derived from bottom trawl surveys.

An important exception to this conclusion was for the Barents Sea surveys. Here, a combination of a larger data set (more survey hauls), and fish distributions that were more amenable to acoustic measurement does promise a potential improvement in the index.

Future developments in multi-frequency or multi-beam acoustic methods may improve the quality and discriminatory power of the acoustic data. Ongoing work on the quantification of whole gear selectivity may also increase our knowledge of what is actually sampled by a trawl net differences. Once these technologies mature, it may be possible to revisit this approach. However, for the time being, and at least in the North & Irish Seas it is unlikely that a useable combined trawl and acoustic index can be determined.

Further information is available on the project web site at http://www.cg.ensmp.fr/~bez/catefa/.

5.2 Combining survey indices: lessons from assessment models

An example of a combination of multiple surveys is provided by the use of four different data sets within an assessment of North Sea herring (*Clupea harengus*, Simmonds, 2003). Three different internationally coordinated research vessel surveys provide data for the assessment of North Sea herring. Herring larvae surveys started in 1972 and routinely the data since 1973 have been used to provide an index of spawning stock biomass for North Sea herring. North Sea 1st quarter IBTS surveys started in 1971 but only by 1983 was the fishing gear and operating procedures sufficiently standardised to deliver a consistent index for ages 1 to 5+. On the same IBTS survey an ichthyoplankton net (MIK, Methot Isaacs Kidd) has been used at night since 1979 and provides an index of 0 group herring. The North Sea herring acoustic survey was started in 1979 and by July 1984 was an internationally co-ordinated survey conducted annually. Though this survey continued to expand area and coverage, by 1989 it was running consistently and providing indices at age 2 to 9+. Since 1995 the area has been extended and indices of age 1 have been provided. These three surveys giving four datasets have been used to tune a catch at age assessment model of North Sea herring. The survey data has been extensively examined to determine the best use of each component. Initially each age group in each survey was given equal weight but it became obvious that different sources of data were of different quality. In 1999 an ICES study group evaluated weighting of the indices (ICES, 2001). The precision of the acoustic survey was estimated, using data at ICES statistical rectangle level, using bootstrap resampling methods modified by geostatistical estimates of the spatial autocorrelation. Similar techniques were applied to the larvae, MIK and trawl surveys but at individual station level. The comparison of survey performance was also included the bootstrap estimates of abundance at age to give 1000 simulated assessments of North Sea herring using the assessment method (Integrated Catch at Age, ICA), the outcome of these comparisons is described in Simmonds (2003). Using the results of these analyses the appropriate weighting of all the various indices of herring abundance at age within the assessment was investigated. Several methods were tested:

1) Equal weighting of all age groups in all indices (previous assessment method).
2) Adaptive weighting with weights estimated within the ICA model.
3) Inverse variance weights with a single weight for each age group based in mean inverse variance.

4) Inverse variance / Adaptive weights (fixed as the mean of 2 and 3 above).

In conclusion the inverse variance weighting method was selected which provided the most precise method for estimating the stock among the weighting methods tested. The more precise assessments were checked for retrospective pattern and an assessment was proposed which provided the most precise stock estimates with the best retrospective pattern. This assessment has been reviewed and accepted by ICES Advisory Committee on Fisheries Management and used in subsequent years up to and including this year (2005).

There are two important aspects to the use of survey data in this manner:

- The weighting selected was derived directly from the observed sampling variability in the indices. These selected weights did not conform to the weights that would be selected by adaptive within model fits. Table 9 compares the weights used with weights for adaptive ICA fit or XSA (eXtended Survivors Analysis) fit. The this use of prior weights gives a less variable assessment over eight different terminal years than did the use of adaptive weights estimated within the assessment. The major difference between the weights can be seen in Table 9 which shows that the prior selected weights give higher importance to younger ages which are known estimated more precisely. The adaptive weighting gives more weight to older ages as it appears to be easier to fit older ages in such an assessment model, however, results suggest this does not produce the best results.

- The method uses a single value weighting factor for each age for each index, not a value that depends on the precision of each annual estimate. The reason for this is that the estimate of the variance in a single year is found to be correlated with the abundance estimate in that year. This because low estimates of abundance occur when by chance a set of relatively low sample values are obtained and as these values are also used to calculate the variance this values is also low. If yearly estimates were used with inverse variance weighting the lower abundance would apparently indicate a more precise estimate. It was considered that this might bias the assessment. The mean variance over a number of years using a consistent survey would better reflect the underlying precision of the estimate of the age class. If the survey design was changed there would however, be a need to estimate the variance again for each period of the survey.

Table 9. Comparison between the weighting factors currently used in the combination of indices in the assessment of North Sea herring. The weights are derived from an analysis of sampling variance of the survey indices. For comparison alternative weights derived from minimum variance estimation within the assessment model are shown. The weights used have been shown to provide more repeatable assessments than when the weights are estimated within the model. Section A shows the weights used directly in the models, section B shows the same weights standardised to 5 for the acoustic survey for easier comparison.

	Survey Inverse Variance Wts			ICA adaptive weights			XSA with weak shrinkage					XSA with high shrinkage				
Age	Acoustic survey 2-9+	IBTS: 1-5+ wr	MIK 0-wr	Acoustic survey 2-9+	IBTS: 1-5+ wr	MIK 0-wr	Acoustic survey 2-9+	IBTS: 1-5+ wr	MIK 0-wr	P shrinkage mean	F shrinkage mean	Acoustic survey 2-9+	IBTS: 1-5+ wr	MIK 0-wr	P shrinkage mean	F shrinkage mean
0			2.05			0.30			0.69	0.26	0.05			0.39	0.15	0.46
1	0.74	0.67		0.73	0.53		0.40	0.40	0.13	0.05	0.01	0.36	0.36	0.11	0.04	0.14
2	0.75	0.24		0.81	0.14		0.57	0.34	0.09		0.01	0.51	0.31	0.07		0.11
3	0.64	0.06		0.83	0.19		0.64	0.30	0.06		0.01	0.58	0.27	0.05		0.10
4	0.27	0.03		0.68	0.19		0.67	0.28	0.04		0.01	0.62	0.26	0.04		0.09
5	0.14	0.03		0.85	0.13		0.72	0.24	0.03		0.01	0.63	0.21	0.02		0.15
6	0.13			0.78			0.80	0.18	0.02		0.01	0.71	0.15	0.02		0.12
7	0.12			0.38			0.86	0.12	0.01		0.01	0.75	0.11	0.01		0.13
8	0.07			0.65			0.91	0.07	0.01		0.01	0.78	0.07	0.00		0.15
Total	2.86	1.03	2.05	5.69	1.18	0.30	5.56	1.93	1.07	0.32	0.12	4.92	1.72	0.71	0.20	1.45
B) Weighting for each assessment method normalised to give a standard weight of five to the acoustic survey in each assessment, in order to facilitate comparison																
0			3.58			0.26			0.62	0.24	0.04			0.39	0.15	0.47
1	1.29	1.17		0.64	0.46		0.36	0.36	0.12	0.05	0.01	0.36	0.36	0.11	0.04	0.14
2	1.31	0.42		0.71	0.12		0.51	0.30	0.08		0.01	0.52	0.31	0.08		0.11
3	1.12	0.10		0.72	0.17		0.57	0.27	0.05		0.01	0.59	0.27	0.05		0.10
4	0.47	0.05		0.59	0.17		0.60	0.25	0.04		0.01	0.62	0.26	0.04		0.09
5	0.24	0.05		0.74	0.11		0.65	0.22	0.02		0.01	0.64	0.21	0.02		0.15
6	0.23			0.68			0.72	0.16	0.02		0.01	0.72	0.15	0.02		0.12
7	0.21			0.34			0.77	0.11	0.01		0.01	0.76	0.12	0.01		0.13
8	0.12			0.57			0.82	0.07	0.00		0.01	0.79	0.07	0.00		0.15
Total	5.00	1.80	3.58	5.00	1.04	0.26	5.00	1.73	0.96	0.28	0.11	5.00	1.75	0.72	0.20	1.47

6 Estimating biological parameters

6.1 Estimating population characteristics based on cluster samples

Fish that are caught together at a station form a cluster. From each cluster, fish for aging, measuring, *etc.* are selected, i.e., data on population characteristics are often generated by two-stage cluster sampling. When the sample consists of a total of m fish from n clusters; the individual animals are not a random sample from the entire population. This is because animals caught together tend to be more similar than animals in the entire population (i.e., there is positive intra-cluster correlation). The practical implication of positive intra-cluster correlation is that a sample of animals caught in clusters will generally contain much less information on the population structure than an equal number of fish sampled at random, i.e., the effective sample size is much smaller than the number of animals sampled (Pennington *et al.,* 2002; Aanes and Pennington, 2003).

Given a random sample of n clusters and a random subsample of m_i fish from a total of M_i individuals in cluster i, then the design-based estimator

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{n} M_i \tilde{x}_i}{\sum_{i=1}^{n} M_i} \qquad (1)$$

is an approximately unbiased and a consistent estimator of; 1) the mean age or length of the population if $\tilde{x}_i$ is the average age or length of the sample of m_i fish from cluster i or; 2) the proportion at age or length in the population if $\tilde{x}_i$ is the estimated proportion of fish of a specific age or length class in cluster i (Skinner *et al.,* 1989; Lehtonen and Pahkinen, 2004). This is a weighted average of the $\tilde{x}$'s, where the cluster sizes are the weights. Since both the numerator and denominator are random variables this is a ratio type estimator (Cochran, 1977), and an exact variance formula does not exist. The variance may be approximated using a Taylor expansion of (1) or by resampling techniques, such as nonparametric bootstrapping (*e.g.,* Efron, 1983).

An alternative to the design-based estimator, which in some situations may have a smaller variance than estimator (1), is the unweighted average of the $\tilde{x}$'s

$$\hat{\mu}_2 = \frac{\sum_{i=1}^{n} \tilde{x}_i}{n}. \qquad (2)$$

In general, the unweighted estimator, $\hat{\mu}_2$, may be biased and this bias may not decrease with increasing sample size, but if $\tilde{x}_i$ and M_i are uncorrelated, then $\hat{\mu}_2$ may be an acceptable estimator (Cochran, 1977). If M_i and $\tilde{x}_i$ are correlated, then the expected bias of estimator (2) is;

$$Bias(\hat{\mu}_2) = -\frac{Cov(M_i, \tilde{x}_i)}{\overline{M}},$$

where $\overline{M}$ is the mean cluster size. One reason that estimator (2) is sometimes used is that the sizes of the clusters, M_i, are unknown or not recorded, and, hence, the resulting estimate may contain an unknowable bias.

To evaluate the precision of the two estimators, consider the standard random effects model

$$x_{ij} = \mu + A_i + \varepsilon_{ij},$$

where A_i is the cluster effect and it is assumed that $E(A_i \mid M_i) = 0$, $Var(A_i \mid M_i) = \sigma_A^2$, $E(\varepsilon_{ij} \mid M_i) = 0$, $Var(\varepsilon_{ij} \mid M_i) = \sigma_\varepsilon^2$ and $m_i \propto M_i$. Then (Aanes and Pennington, 2003)

$$Var(\hat{\mu}_1 \mid \mathbf{m}) = \frac{\sigma_\varepsilon^2}{n\overline{m}} + \frac{\sigma_A^2}{n}\left(1 + \frac{s_m^2}{\overline{m}^2}\right) \tag{3}$$

and

$$Var(\hat{\mu}_2 \mid \mathbf{m}) = \frac{\sigma_\varepsilon^2}{n^2}\sum_{i=1}^{n}\frac{1}{m_i} + \frac{\sigma_A^2}{n}, \tag{4}$$

where *m* is the vector of m_i's.

If the intra-cluster correlation, ρ, is 0, then $\sigma_A^2 = 0$, and it follows from the Cauchy-Schwarz inequality that $\sigma_\varepsilon^2/(n\overline{m}) \le \sigma_\varepsilon^2/n^2\sum_{i=1}^{n} 1/m_i$ for all *m* with equality only if the sample sizes, m_i, are equal. Therefore if $\rho = 0$, then for any *m* the variance of estimator (1), which is the ordinary average of all the values, will be less than or equal to the variance of the unweighted estimator (2).

If $\rho > 0$, then which is smaller, $Var(\hat{\mu}_1)$ or $Var(\hat{\mu}_2)$, will depend on the sizes of the various components in equations (3) and (4). In particular, it should be noted that the first term in equation (4) can be considerably larger than the first term in equation (3) when the samples consist of many small values and a few very large values, which is often the case for scientific trawl surveys of fish stocks.

Because of positive intra-cluster correlation, the effective sample is more informative than the total number of fish sampled (Lehtonen and Pahkinen, 2004). The effective sample size m_{eff} is defined as the number of individuals that would need to be sampled at random so that the estimate generated by simple random sampling would have had the same precision as the estimate obtained based on a more complex sampling scheme (Kish, 1965; Skinner *et al.*, 1989). To estimate the effective sample size for estimating, for example, the age distribution, first estimate, based on the sampling scheme, the mean age, $\hat{\mu}_a$, its variance and the variance of the age distribution of the target population, $\hat{\sigma}_a^2$. Then the estimated effective sample size, $\hat{m}_{eff}$, is defined by

$$\frac{\hat{\sigma}_a^2}{\hat{m}_{eff}} = \mathrm{var}(\hat{\mu}_a). \tag{4}$$

As an example, in Table 10 are summary statistics for estimating the mean age of the Norwegian commercial catch of Northeast Arctic cod (*Gadus morhua*) in 2000 (Aanes and Pennington, 2003). A number of cod were aged from individual fishing trips, and thus the fish caught during a trip form a cluster. For these data there was positive intracluster correlation (Table

10), but the size of the clusters did not appear to be correlated with average age in a cluster. Therefore, the estimates of mean age generated by estimators (1) and (2) were similar. For this data set, estimator (2) appeared to be more precise, on average, than (1), but the effective sample size for both estimators was rather small compared with the number of fish aged. For example during the first quarter a total of 6000 cod were aged and the estimated effective sample sizes for estimators (1) and (2) were 59 and 223, respectively.

Table 10. Summary statistics for estimating the mean age of the catch of Northeast Arctic cod in 2000 using estimators $\hat{\mu}_1$ and $\hat{\mu}_2$; where *n* is the number of fishing trips sampled; *m* the total number of fish aged from the *n* trips; $\hat{\mu}_i$ and $se(\hat{\mu}_i)$ are the estimated mean age and its standard error for *i=1,2*, respectively; $\hat{m}_{i,eff}$ is the effective sample size for *i=1,2*; and $\hat{\rho}$ is the estimated intracluster correlation coefficient. The approximate 95% confidence intervals are in parenthesis. The estimated standard errors and confidence intervals are based on 500 bootstrap replicates (from Aanes and Pennington, 2003).

QUARTER	n	m	$\hat{\mu}_1$	$se(\hat{\mu}_1)$	$\hat{\mu}_2$	$se(\hat{\mu}_2)$	$\hat{m}_{1,eff}$	$\hat{m}_{2,eff}$	$\hat{\rho}$
1	70	6000	6.75 (6.35,7.23)	0.23	7.25 (6.99,7.45)	0.11	59 (37,212)	223 (158,330)	0.26 (0.20,0.33)
2	26	2277	5.33 (5.20,5.46)	0.07	5.33 (5.21,5.46)	0.06	211 (93,393)	213 (135,333)	0.10 (0.06,0.15)
3	13	1077	5.23 (4.98,5.60)	0.17	5.27 (5.08,5.55)	0.12	32 (18,193)	56 (29,215)	0.20 (0.06,0.33)
4	17	1342	5.05 (4.89,5.18)	0.07	5.06 (4.83,5.29)	0.12	182 (81,428)	57 (29,206)	0.23 (0.07,0.38)

For survey data, the effective sample size is often approximately equal to the number of hauls (Pennington and Vølstad, 1994). In Table 11 are estimates of the effective sample size for estimating the mean length of Northeast Arctic cod from data collect by the Institute of Marine Research's winter survey in the Barents Sea.

Table 11. Summary statistics for assessing the precision of the estimated length distributions of Northeast Arctic cod based on the winter bottom trawl surveys in the Barents Sea. The estimated effective sample size is denoted by $\hat{m}_{eff}$, *n* is the number of stations at which cod were caught, *M* is the total number of cod caught, *m* is the number measured, $\hat{R}$ is the estimate of mean length and var($\hat{R}$) is its variance (from Pennington *et al.,* 2002).

YEAR	N	*M*	*M*	$\hat{R}$ (CM)	VAR($\hat{R}$)	$\hat{m}_{eff}$	$\hat{m}_{eff}$ / N	$(\hat{m}_{eff} / m) \times 100\%$
95	296	175006	47286	20.0	0.7	313	1.1	0.7
96	314	209114	44021	18.0	0.3	511	1.6	1.1
97	177	71418	25689	19.0	2.1	119	0.7	0.7
98	197	60746	32536	22.1	0.7	394	2.0	1.2
99	223	50192	21760	25.0	1.9	107	0.5	0.5
	Avg.	113295	34258			289	1.2	0.8%

It should be noted that if the effective sample size is small, then this implies that the estimate of the entire age or length distribution is rather imprecise (Pennington and Vølstad, 1994; Pennington *et al.,* 2002). For example, the effective sample size for estimating the mean length of haddock in 1967 was 10. To demonstrate that a low effective sample size implies that the estimate of the entire length distribution is rather imprecise, the 59 stations at which haddock were caught were randomly split into two groups and the length distribution was estimated based on each group (Figure 23). As can be seen, the estimated length distributions appear to be markedly different, even though they are not statistically different. Likewise, during the 1998 shrimp survey off West Greenland 7 stations were sampled in a small stratum and at each station approximate 4 kg of shrimp were subsampled. A total of 5341 shrimp were measured and the effective sample is approximately 24 (Folmer and Pennington, 2000). As for haddock, the stations were randomly split into two groups. The estimated length distribution using all 7 stations (based on 5341 shrimp) is in Figure 24(a), in Figure 24(b) is the length distribution based on four stations (2696 shrimp) and in Figure 24(c) is the distribution for the other three stations (2645 shrimp). Finally, the effect of reducing the number of Northeast Arctic cod measured during the 1999 winter survey from 21,760 to 2,597 is shown in Figure 25. In particular, it should be noted that even for the larger sample ($\hat{m}_{eff} = 107$) the confidence intervals are rather wide even though a large number of fish were measured.

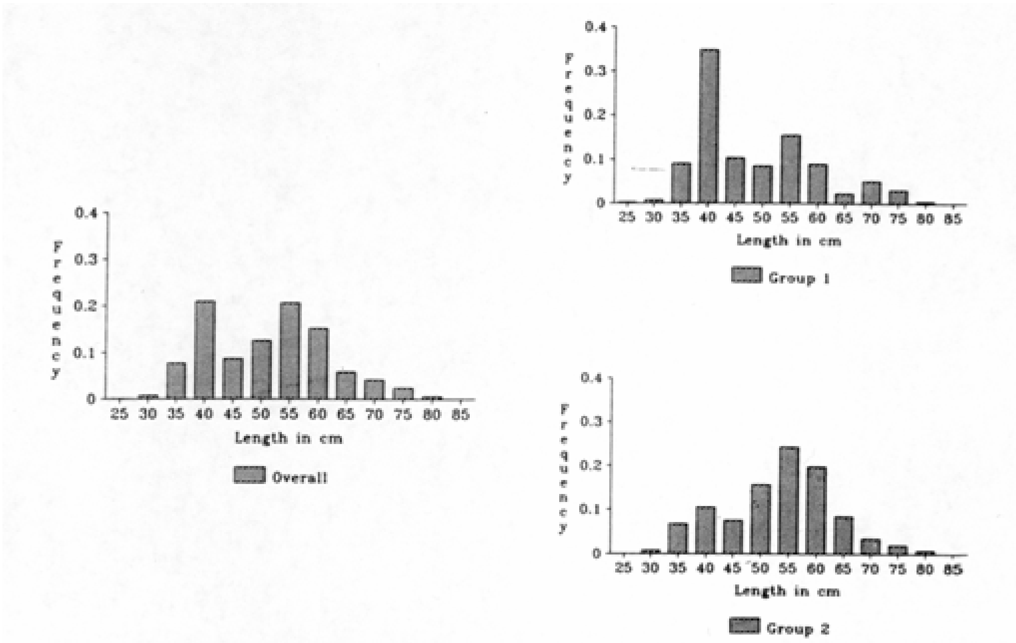


Figure 23. The overall length-frequency distribution of haddock for 1967 is based on 893 fish from 59 stations. The stations were randomly split into two groups. The distributions for groups 1 and 2 are based on 384 and 509 fish, respectively.

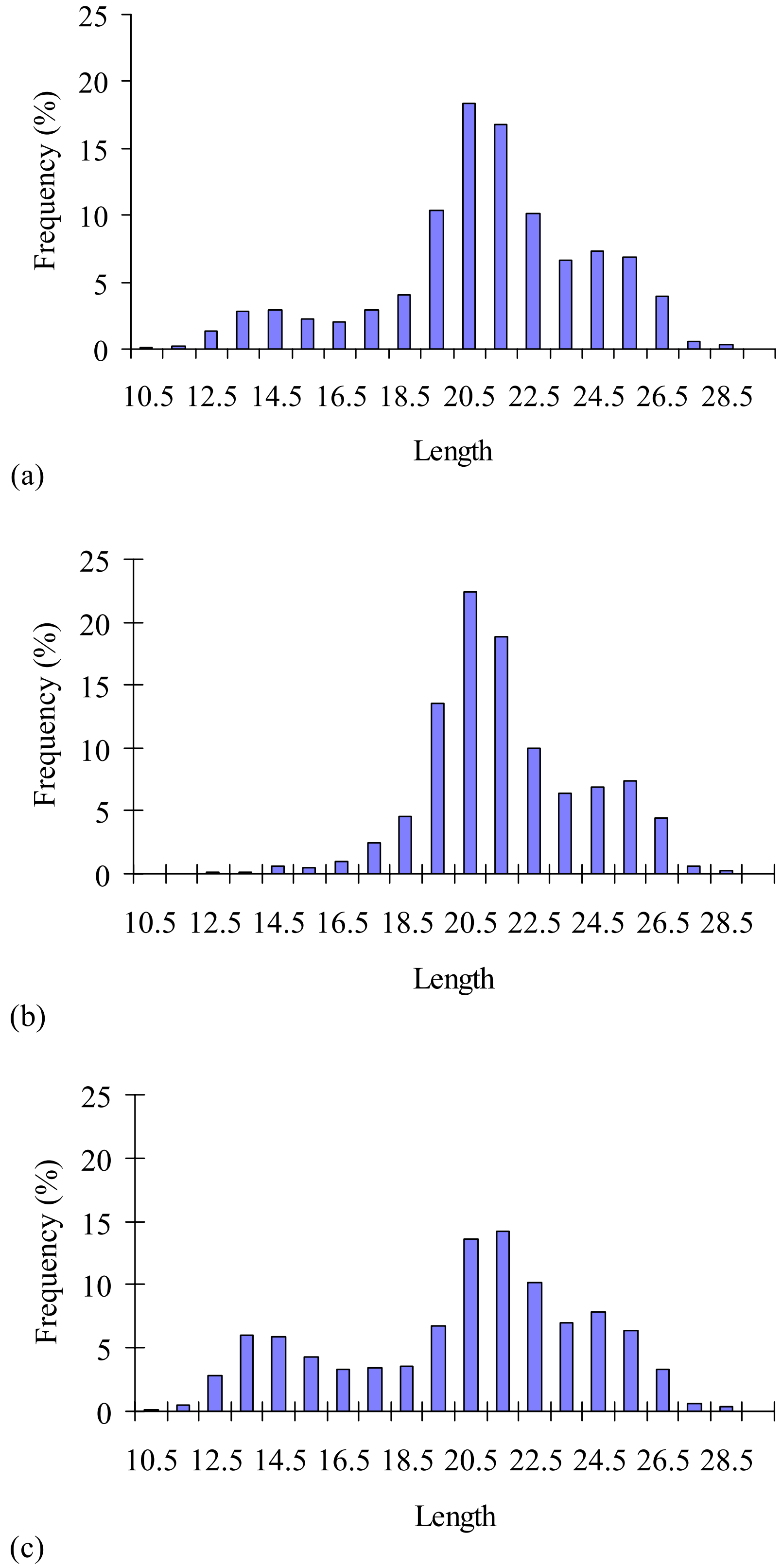


Figure 24. (a) The overall length-frequency distribution based on a total of 5341 measured shrimp from seven stations in a small stratum. (b) The length-frequency distribution based on four of the stations chosen at random (2696 shrimp) and (c) the distribution based on the other three stations (2645 shrimp) from Folmer and Pennington (2000).

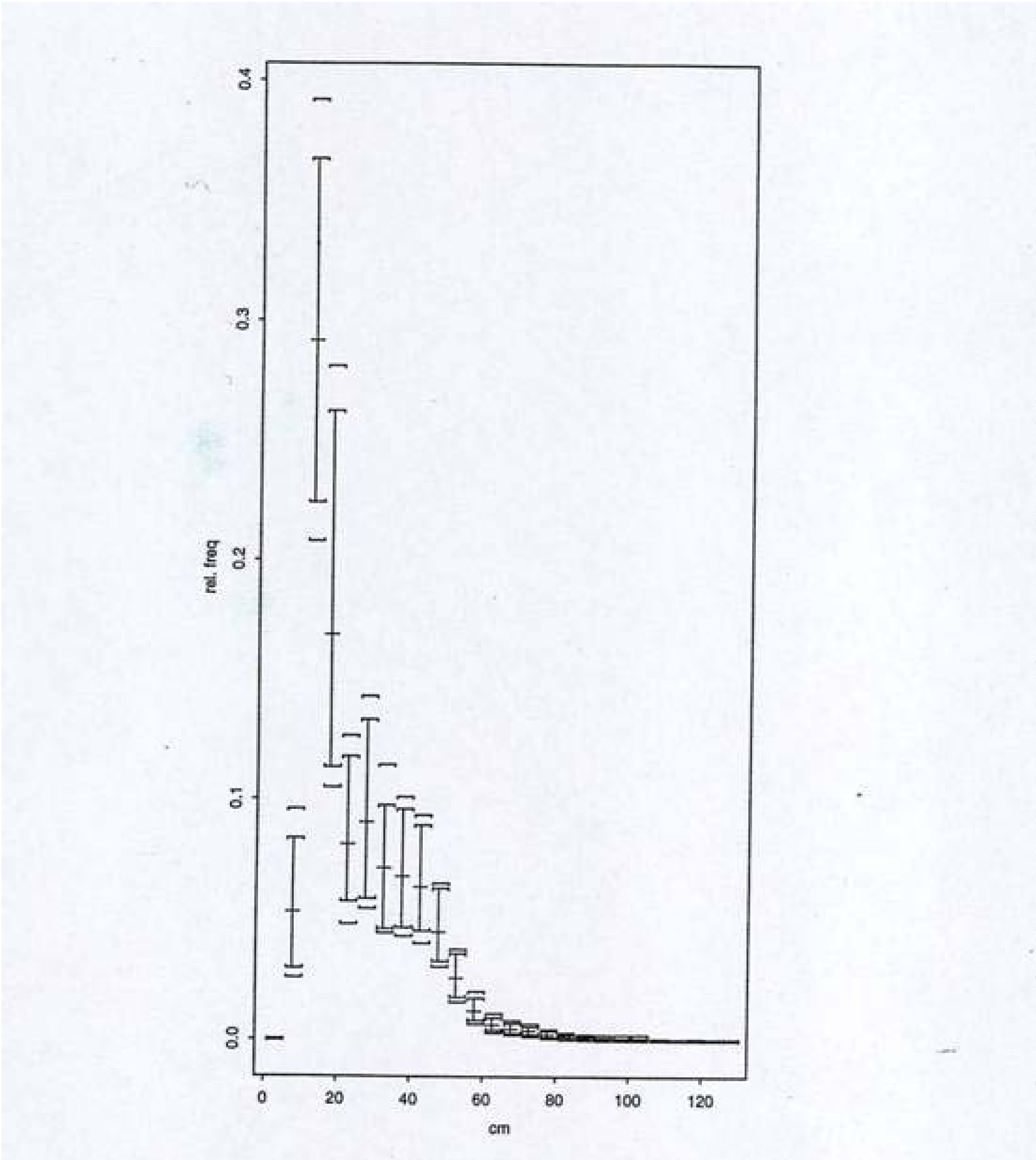


Figure 25. Bootstrapped estimates of the 95% confidence intervals for the relative length frequency distribution of cod in the Barents Sea in winter 1999. The inner brackets denote the confidence intervals if the estimates are based on all the cod measured (m = 21,760 fish) and the outer brackets, if 10 fish were measured per subsample (m = 2,597 fish).

6.2 Interpolating biological data from acoustic surveys

For pelagic species, abundance is usually estimated through echosounder data and the biological composition from directed opportunistic trawl hauls. Trawl data collected in this way provides a non-random discrete estimate which can be linked to fish traces detected by the echosounder.

The data collected often shows spatial trends in fish size and other biological characteristics or metrics. There are several reasons why the size distribution of a fish stock might not be the same over the large area covered in the typical acoustic survey. When the surveyed area is bordered by a coastline, for example, there may be differences in the age structure between inshore fish and those offshore in deeper water. Or the fish may associate with others near the same size and age, resulting in clusters of fish in particular age groups. If the fishing samples indicate consistent differences between one region and another, different size, age, sex and maturity distributions may need to be determined for different regions within the overall area.

The observed size distributions will vary due to sampling error as well as real changes in the population structure between the trawl stations. Thus, there is a need to match trends in biological structure while allowing for some measurement error. If the measurement error is small enough to be negligible then a nearest neighbour estimate will suffice to map the data. If the measurement error completely dominates the observed variability and there is no apparent spatial structure in the distribution a global mean will be the appropriate method. Only in the intermediate situation is there a need for a more sophisticated method, accommodating some sampling error and some important spatial structure. There is currently no well established method for mapping all the variables obtained by sampling biological data together. Mapping of different parameters by different methods might lead to inconsistencies at the unsampled locations, so a comprehensive procedure involving all the variables is preferred. Currently one procedure used (Simmonds and MacLennan, 2005) is to select relatively homogeneous regions. This might be evident from inspection of the numerical data if there are clear differences between clusters of trawl stations. Alternatively, a more objective approach to the problem is to apply the Kolmogorov-Smirnov (KS) test (Campbell, 1974) on the length distributions. This method is sensitive both to the position of modes and to the spread of size distributions obtained. The result of the test is a number P_{KS} in the range 0 to 1. $P_{KS} = 0$ when the distributions are identical, and $P_{KS} = 1$ when there is no overlap at all. If the distributions come from the same population, subject only to sampling error, P_{KS} might be expected to be small, e.g., less than 0.1.

It is not suggested that the KS test should be applied as an automatic procedure or that significance using the measure number N be used directly as this has little meaning, expressing only the number of fish taken from the sample of the cluster not the effective number in the sample (see section 6.1). A more thoughtful approach is necessary. We might begin by assigning the trawl stations to groups within which P_{KS} for each pair is less than 0.2, and see whether this suggests a sensible division of the surveyed area into homogeneous regions. The calculations are then repeated for similarity thresholds $P_{KS} = 0.15$ and 0.25. If the grouping of the trawl stations is sensitive to the threshold value of P_{KS}, this indicates that the differences in size distribution are small enough to be ignored, and only one region (the surveyed area) needs to be considered. Otherwise, the boundaries between the homogeneous regions are determined by the condition that any part of the surveyed area is assigned to the region corresponding to that of the closest trawl station.

Figure 26 shows an example of the P_{KS} values obtained by comparing samples of herring collected at 40 trawl stations, displayed as a dendrogram. In Figure 26, the distributions are grouped spatially according to the value of P_{KS}. There is a clear statistical separation into five groups, with a suggestion of three more which are spatially disparate. This suggests that the surveyed area should be divided into eight regions. The boundaries of these regions in relation to the trawl stations are shown in Figure 27. The mean length of the herring caught at each location is also shown in this figure as an indication of the spatial variability. Comparison of the fish size in regions I, III, V and VI shows why the area must be partitioned in this case. The evident local variability suggests that some averaging is better than the nearest-neighbour method, since the latter fits a stepwise surface at each trawl location and implies no error in the fish-size distribution.

This methodology may be extended by combining Euclidean distance, P_E, with P_{KS} through $(P'_E \cdot P_{KS})$^0.5. Here, P_E is not used directly but normalised to P'_E with the same mean as P_{KS} to make the two terms comparable. A new dendrogram and the spatial boundaries are shown in Figures 28 and 29. This method makes the selection of a single threshold for separating areas much easier as it combines spatial and statistical aspects in a single metric and tends not to mix similar size groups from non-contiguous areas.

Further development of coherent mapping of the full suite of biological parameters would be desirable. Considering each point in space as a linear weighted combination of biological data from hauls in the same region seems reasonable, as stated above one extreme is a nearest neighbour method, (the weighting factors are all 1 and 0), the other extreme is the global mean (the weighting factors are all 1). The regional method described here delivers a stepwise approximation to a surface with hauls in groups given equal weight. It would be even better to fit a continuous surface. The benefits in using the same spatial weighting for all biological data would be considerable, it would ensure that raising the other biological data through length stratified sampling of biological data would be explicitly consistent through space. Currently there are no explicit methods for this kind of multivariate mapping except possibly co-kriging. Suggestions for obtaining a set of weighting factors suggest further investigation including: kriging weights through kriging mean length or quartiles; 2D kernel smoothing; and General Additive Models (GAMs). It is thought that most of these techniques will deliver suitable results and the overall difference might be expected to be small. Further examination and testing of these methods is to be encouraged.

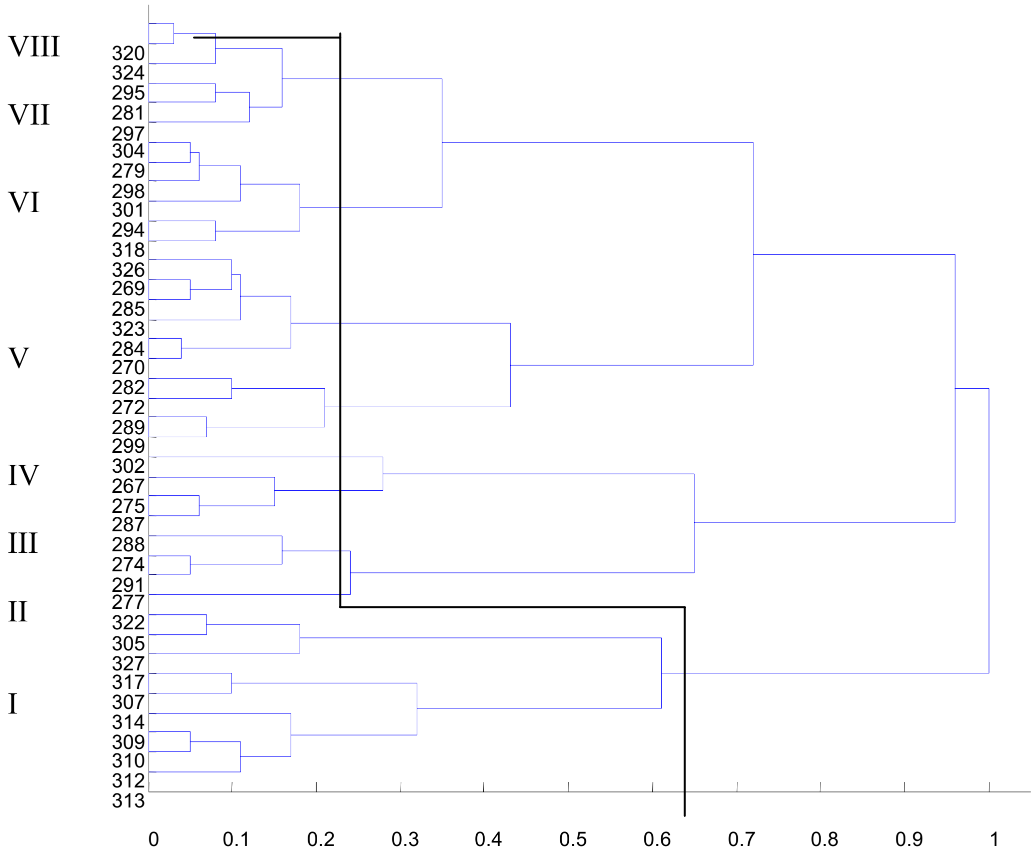


Figure 26. Dendrogram of P_{KS} for 39 trawl hauls located as shown in Figure 27

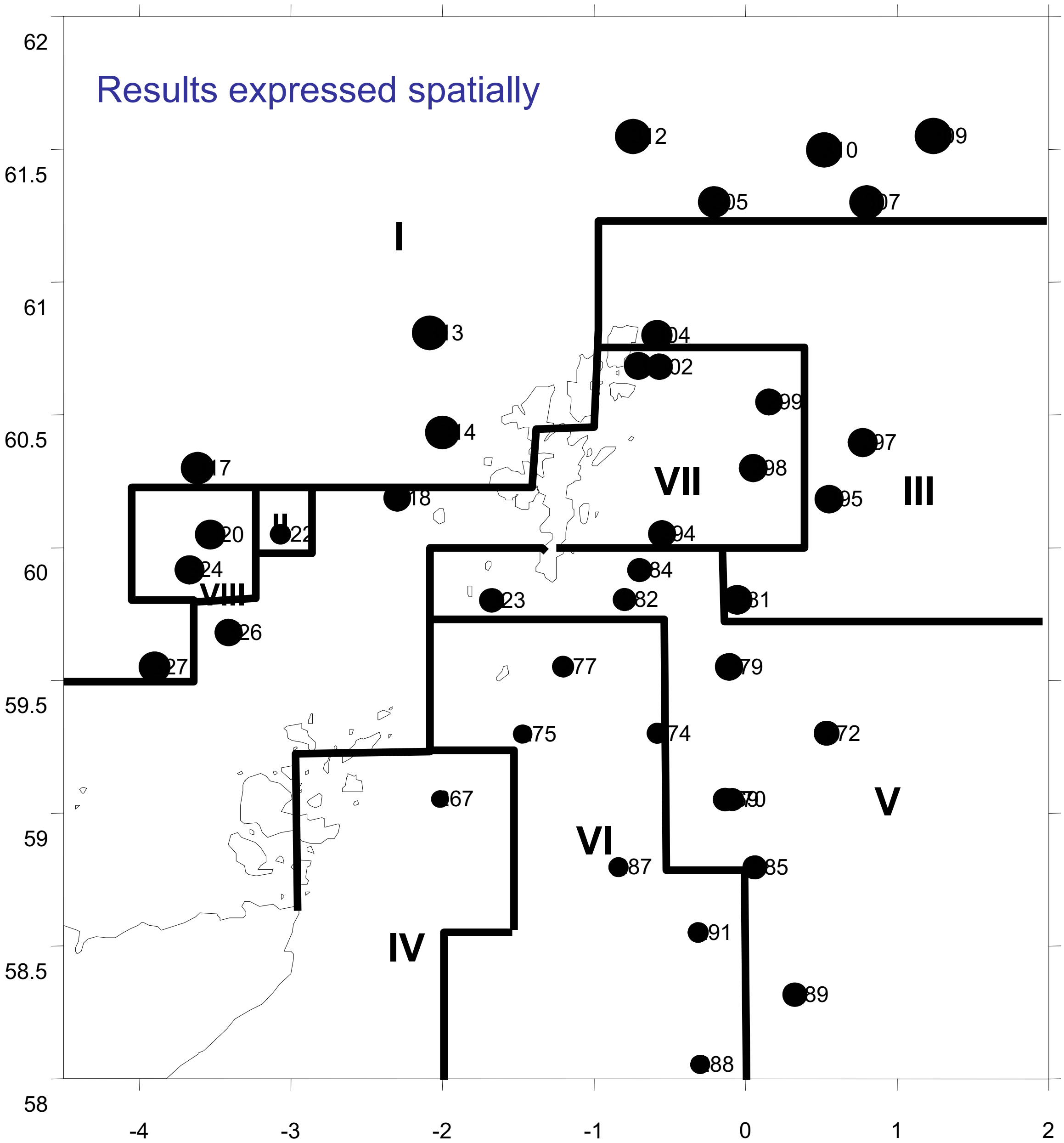


Figure 27. Location of trawl hauls, mean length (dot size) and region boundaries from the threshold line in Figure 26

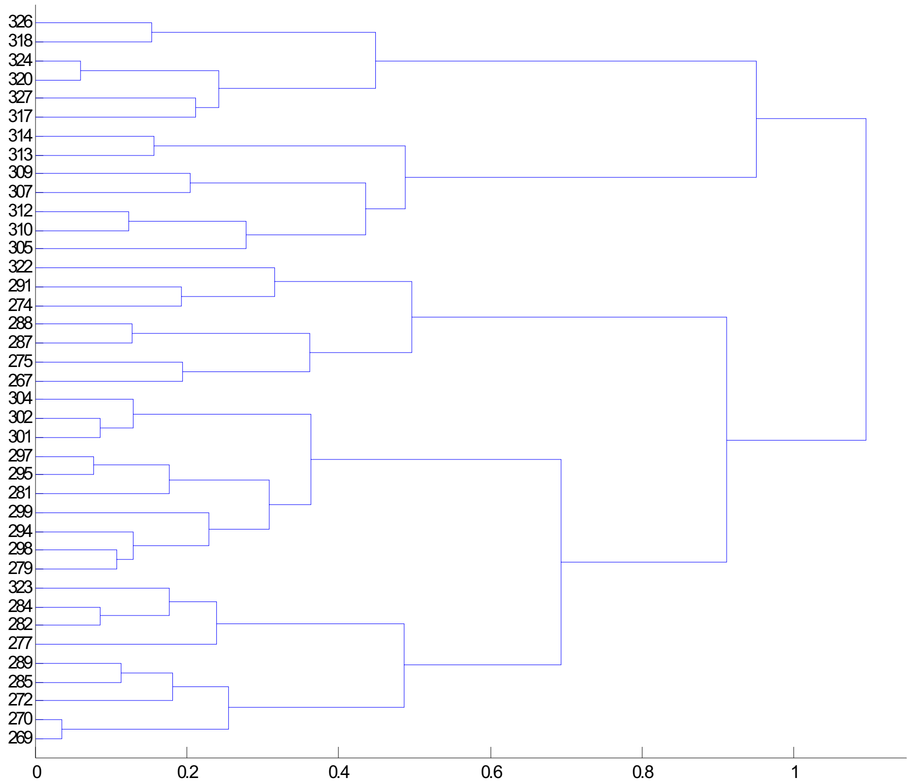


Figure 28. Dendrogram of (P_{KS} P'_E)^0.5 for 39 trawl hauls located as shown in Figure 29.

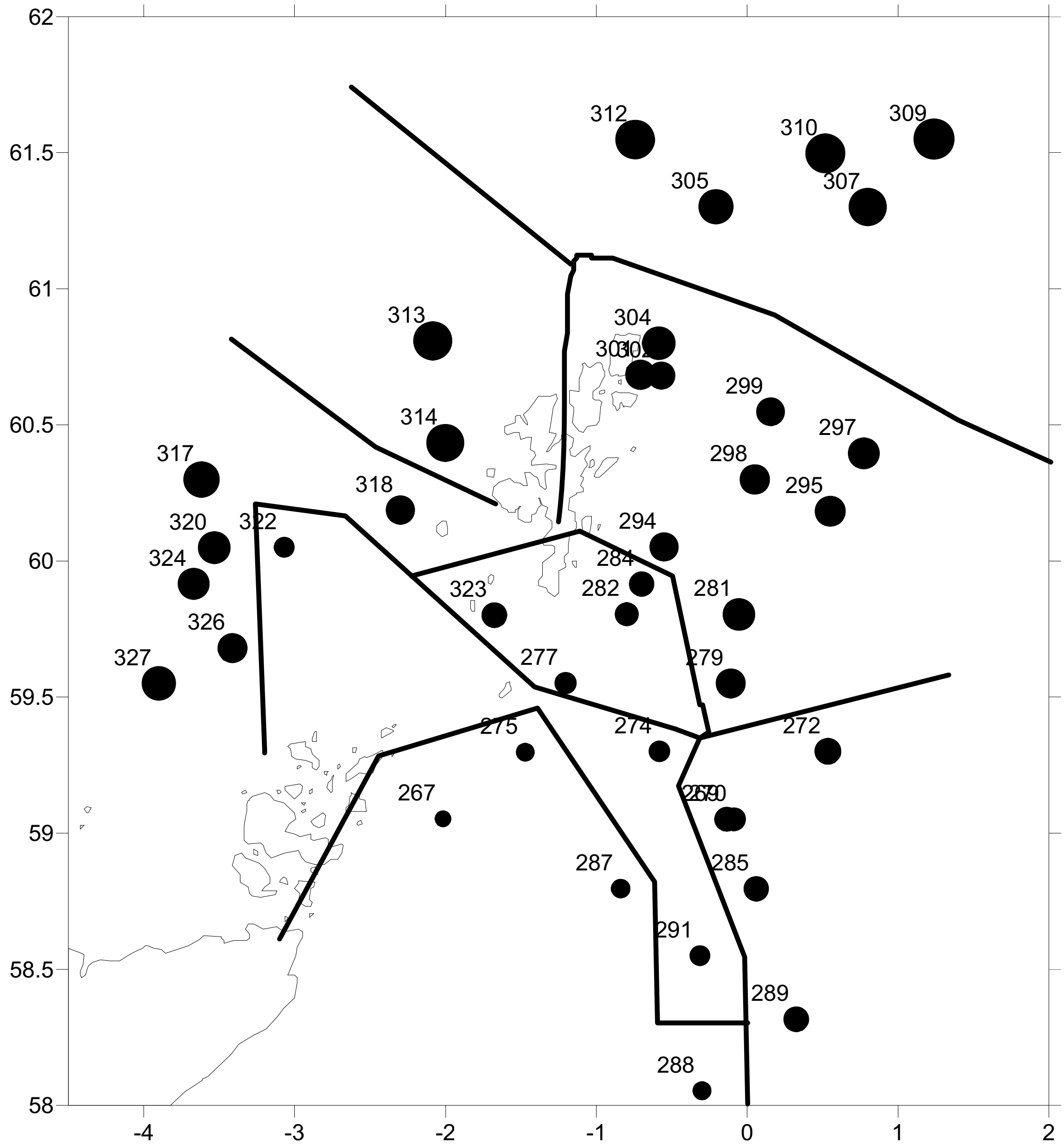


Figure 29. Location of trawl hauls, mean length (dot size) and region boundaries derived from the threshold line in Figure 28.

7 Recommendations

1) **The spatial distribution of the fish** should be considered when designing and analysing surveys. A decision tree has been provided to assist in the choice of methods available. Survey planners should be fully aware of the assumptions allied to any model-based estimation technique.

2) **The survey specific effect of tow duration**, should be investigated in individual surveys. Shorter tows should be implemented if found to provide an improvement in the precision of the survey.

3) **Covariates should be used, if available**, where they provide an improvement in the precision of the survey. Be aware that the covariates must have a good relationship with the response and be available over the entire sample space (not just the sampled area).

4) **Inverse variance weighting should be considered to combine survey data**. When combining indices of the same resource, the inverse variance of the individual indices is a useful weighting scheme.

5) **The effective sample size to determine biological parameters** should be investigated. The *effective* sample size of fish selected for ageing, measuring, etc. can be much smaller than the actual number of animals sampled, it is, therefore, important to account for this when reporting information on biological parameters. In cases where this can demonstrated to be smaller than current sample sizes more effort can be incorporated into sampling other species (including non-fish species) for consideration of an ecosystems approach (e.g. to compile community-based indicators).

6) Quantiles of individual distributions can be used to map biological data rather than interpolating a summary statistic (e.g. mean length).

7) **Further meetings of ICES WKSAD**: The group does not recommend meeting until such time as certain analyses have been carried out which demonstrate progress and can form the basis of further discussion. The following areas require further investigation and participants are encouraged to pursue appropriate studies in:

- Simulations to determine the levels of autocorrelation required for optimal survey design strategies.
- The effect of reduced tow duration (and subsequent increased sample size) on the precision of the survey.
- The effective sample size of biological (trawl) samples.
- Methods for incorporating covariates which improve the estimation of fish abundance.
- Methods to interpolate statistical distributions, for the purposes of, for example, improving the interpolation of acoustic survey data.
- Methods of determining the total precision in surveys

8 References

Aanes, S. and 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.

Albert, O.T., Harbitz, A., Høines, Å.S. 2003. Greenland halibut observed by video in front of survey trawl: behaviour, escapement, and spatial patterns. Journal of Sea Research, 50: 117–127.

Bez, N., Rivoirard, J., and Poulard, J.C. 1995. Approche transitive et densités de poissons. Compte-rendu des journées de Géostatistique, 15–16 juin 1995, Fontainebleau, France. Cahiers de Géostatistique, 5: 161–177.

Bez, N. and Rivoirard, J. 2001. Transitive geostatistics to characterize spatial aggregations with diffuse limits: an application on mackerel ichtyoplankton. Fisheries Research, 50: 41–58.

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

Campbell, R.C. 1974. Statistics for Biologists, 2nd ed. Cambridge University Press, 383 pp.

Carothers, P.E., and Chittenden, M.E. 1985. Relationships between trawl catch and tow duration for penaeid shrimp. Transactions of the American Fisheries Society, 114(6): 851–856.

Cochran, W.G. 1977. Sampling techniques. 3rd edition. Wiley, New York, 428 pp.

Cressie, N.A.C. 1991. Statistics for spatial data. Wiley, New York.

Efron, B. 1983. The Jackknife, the Bootstrap and Other Resampling Plans, 2nd ed. Society for Industrial and Applied Mathematics, Philadelphia.

Ellis, J.R., Armstrong, M.J, Rogers, S.I., and Service, M. 2002. The distribution, structure and diversity of fish assemblages in the Irish Sea. In Marine biodiversity in Ireland and adjacent waters, pp 93–114. Ed. by J.D.Nunn. Ulster Museum, Belfast.

Ellis, J.R. and Rogers, S.I. 2004. Distribution and structure of faunal assemblages and their associated physical conditions on the Atlantic continental shelf of the British Isles. ICES CM 2004/P:03. 25pp.

Ellis, J.R., Rogers, S.I. and Freeman, S.M. 2000. Demersal assemblages in the Irish Sea, St George's Channel and Bristol Channel. Estuarine and Coastal Shelf Science, 51(3): 299–315.

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.

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

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

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.

Hastie, T.J. and Tibshirani, R.J. 1990. Generalized Additive Models. Chapman and Hall, London.

Harbitz, A. and Pennington, M. 2004. Comparison of shortest sailing distance through random and regular sampling points. ICES Journal of Marine Science, 61: 140–147.

ICES 1993. Report of the workshop on the Applicability of Spatial Statistical Techniques to Acoustic Survey Data. ICES Cooperative Research Report, 195. 87 pp.

ICES. 2001. Report of the study group on evaluation of current assessment procedures for north sea herring. ICES CM 2001/ACFM:22.

ICES. 2004. Report of the Workshop on Survey Design and Data Analysis (WKSAD), 21–25 June 2004, Aberdeen, UK. ICES CM 2004/B:07. 261 pp.

Kingsley, M. C. S. 2001. Studies in 2001 on the end effect of the Skjervøy 3000 trawl in the West Greenland shrimp survey. NAFO SCR Doc. 01/177, 7 pp.

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. Fisheries Research, 58: 171–183.

Kish, L. 1965. Survey Sampling. Wiley, New York.

Lai, H.-L. and Kimura, D.K. 2002. Analyzing survey experiments having spatial variability with an application to a sea scallop fishing experiment. Fisheries Research, 56: 239–259.

Lantuéjoul, C. 2002. Geostatistical simulation. Models and Algorithms. Springer, Berlin, 256 pp.

Lehtonen, R. and E. Pahkinen, 2004. Practical Methods for Design and Analysis of Complex Surveys, 2nd ed. Wiley, New York.

Matheron, G. 1971. The theory of regionalized variables and its applications. Les Cahiers du Centre de Morphologie Mathématique 5.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. 2nd edition. Chapman and Hall, New York, 225 pp.

O'Gorman, R., J.H. Elrod, R.W. Owens, C.P. Schneider, T.H. Eckert, and B.F. Lantry. 2000. Shifts in depth distributions of alewives, rainbow smelt, and age-2 lake trout in southern Lake Ontario following establishment of dreissenids. Transactions of the American Fisheries Society, 129: 1096–1106.

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

Pennington, M., and Vølstad, 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., Burmeister, L.-M., and Hjellvik, V. 2002. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fishery Bulletin, 100: 74–80.

Petitgas, P. 1993. Geostatistics for fish stock assessments: a review and an acoustic application. ICES Journal of Marine Science, 50: 285–298.

Petitgas, P. 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. Fish and Fisheries, 2: 231–249.

Poulard J.-C. and J.-C. Mahé, 2004. Structure and spatial distribution of fish assemblages in the Celtic sea. WD for IBTSWG, Lisbon 2004, 14p.

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

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. and MacLennan, D.M. 2005. Fisheries acoustics 2nd edition. Chapman and Hall, London.

Skinner, C.J., Holt, D. and Smith, T.M.F. (eds.). 1989. Analysis of Complex Surveys. Wiley, New York.

Sokal, R.R. and Rohlf, F.J. 1995. Biometry – The principles and practice of statistics in biological research. 3rd edition. W.H. Freeman, New York, 887 pp.

Smith, S. J. and Gavaris, S. 1993. Improving the precision of abundance estimates of Eastern Shelf Atlantic cod from bottom trawl surveys. North American Journal of Fisheries Management, 13:35–47.

Somerton, D. A., Otto, R. S., and Syrjala, S. E. 2002. Can changes in tow duration on bottom trawl surveys lead to changes in CPUE and mean size? Fisheries Research, 55: 63–70.

Storr-Paulsen, M. and Jørgensen, O. 2004. Biomass and abundance of demersal fish stocks off West Greenland estimated from the Greenland shrimp survey, 1988–2003. NAFO SCR Doc. 04/18, 28 pp.

Trenkel, V.M. and Skaug, H. In press. Disentangling the effects of trawl efficiency and population abundance on catch data using random effects models. ICES Journal of Marine Science.

Walsh, S.J., 1991. Effect of tow duration on gear selectivity. NAFO SCR Doc. 91/84, 9 pp.

Walsh, S.J., 1992. Size-dependent selection at the footgear of a groundfish survey trawl. North American Journal of Fisheries Management, 12(3): 625–633.

Wieland, K., Kanneworff, P. and Bergström, B., 2004. Results of the Greenland Bottom Trawl Survey for Northern shrimp (Pandalus borealis) off West Greenland (NAFO Subarea 1 and Divison 0A), 1988–2004. NAFO SCR Doc. 04/72, 31 pp.

Williamson, N.J.. 1982. Cluster sampling estimation of the variance of abundance estimates derived from quantitative echo sounder surveys. Canadian Journal of Fisheries and Aquatic Sciences, 39(1): 229–231.

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.

Annex 1: List of participants

NAME	ADDRESS	PHONE/FAX	EMAIL
Jean Adams	USGS, Marquette, U.S.A.	+1906226 1212 +1906226 3632	jvadams@usgs.gov
Nicola Bez	IFREMER, Sète, France	+33(0)499573204	bez@cg.ensmp.fr
Robert Brown	CEFAS, Lowestoft, England	+44(0)1502 524417	r.g.brown@cefas.co.uk
Noel Cadigan	DFO, Newfoundland, Canada	+17097725028	Cadigann@dfo-mpo.gc.ca
Ian Doonan	MI, Galway, Ireland	+35391730440 +35391730410	ian.doonan@marine.ie
Abdelmalek Faraj	INRH, Casablanca, Morroco	+212(0)61079909 +212(0)22266967	faraj@inrh.org.ma
Paul Fernandes	FRS, Aberdeen, Scotland	+441224295403 +441224295511	fernandespg@marlab.ac.uk
Joakim Hjelm	IMR, Lysekil, Sweden	+4652318750 +4652313977	joakim.hjelm@fiskeriverket.se
Leire Ibaibarriaga	AZTI, Pasaia, Spain	+34943004800 +43943004801	libaibarriaga@pas.azti.es
Johan Lövgren	IMR, Lysekil, Sweden	+4652318750 +4652313977	johan.lovgren@fiskeriverket.se
Jean Claude Mahe	IFREMER, Nantes, France	+33297873515 +33297873536	Jean.Claude.Mahe@ifremer.fr
Michael Pennington	IMR, Bergen, Norway		michael.pennington@imr.no
Jacques Rivoirard	CDG, Fontainebleau, France	+33164694764	rivoi@cg.ensmp.fr
John Simmonds	FRS, Aberdeen, Scotland	+441224295366 +441224295511	simmondsej@marlab.ac.uk
Konstantin Sokolov	PINRO, Murmansk, Russia	+7(8152)472532 +7(8152)473331	sokol_km@pinro.ru
Arnauld Souplet	IFREMER, Sète, France		Arnauld.Souplet@ifremer.fr
David Stokes	MI, Galway, Ireland	+35391730400 +35391730470	david.stokes@marine.ie
Verena Trenkel	IFREMER, Nantes, France	+33240374053	Verena.Trenkel@ifremer.fr
Paul Walline	NMFS, Seattle, USA	+12065264681	Paul.Walline@noaa.gov
Kai Wieland	GINR, Nuuk, Greenland	+299361248 +299361212	wieland@natur.gl
Mathieu Woillez	CDG, Fontainebleau, France	+33164694776	mathieu.woillez@cg.ensmp.fr

Annex 2: Working Documents

WD1. Rivoirard, J. and Woillez, M. Abundance estimation for stratified random surveys using intrinsic geostatistics.

WD2. Adams, J. The gaming exercise: grid, transect and adaptive designs.

WD3. Brown, R. Results of the simulation study for ICES WKSAD.

WD4. Cadigan, N. Confidence intervals for trawlable abundance from random stratified bottom-trawl surveys.

WD5. Adams, J. and O'Gorman, R. Lake Ontario alewife abundance.

WD6. Effect of tow duration on catch rates and mean length of Northern shrimp (*Pandalus borealis*) and Greenland halibut (*Reinhardtius hippoglossoides*) in the West Greenland Bottom Trawl Survey, 1999-2004.

WD7. Oeberst, R. Optimum duration /distance of tows during surveys.

WD8. Oeberst, R. Species composition in scattered layer based on control hauls

Annex 3: Delta distribution code

At the meeting, two types of estimation methods were described in relation to the simulation study. The intrinsic geostatistical methods (Rivoirard et al 2000) and the delta distribution (pennington 1996). Programmes to utilise both methods were distributed at the meeting. Hoever, an additional piece of code to run the delta method in R (with a set cut off of zero) was provided by Jean Adams and is listed here. The author accepts no responsibility for the use or misuse of this code and readers are encouraged to send any updates & modifications to jvadams@usgs.gov

```
# DeltaFunctionR.prg - R function to estimate mean and var(mean) using the delta distribution
# Pennington. 1996. Estimating the mean and variance from highly skewed marine data. Fishery Bulletin 94:984-505.
# technically nrep is supposed to be infinity, but there doesn't seem to be any practical benefit to going beyond 50

delta <- function(x, nrep=50) {
        # estimate mean(x) and var(mean(x)) using the delta distribution
        # Pennington. 1996. Estimating the mean and variance from highly skewed marine data. Fishery Bulletin 94:984-505.
        # technically nrep is supposed to be infinity, but there doesn't seem to be any practical benefit to going beyond 50
        # need also the g() function to run this
  n <- length(x)  # sample size
  m <- sum(x>0)   # number of nonzeroes
  y <- log(x[x>0])
  ybar <- mean(y) # mean of ln of nonzeroes
  s2y <- var(y)   # variance of ln of nonzeroes
  xbar <- mean(x) # mean of all values
  s2x <- var(x)   # variance of all values
  est <- m/n * exp(ybar) * g(m, s2y/2, nrep)  # mean of delta distribution
  d <- m/n * exp(2*ybar) * (g(m, 2*s2y, nrep) - (m-1)/(n-1) * g(m, (m-2)*s2y/(m-1), nrep)) # variance of delta dist'n
  varest <- m/n * exp(2*ybar) * (m/n * g(m, s2y/2, nrep)^2 - (m-1)/(n-1) * g(m, (m-2)*s2y/(m-1), nrep))  # variance of est
  cbind(est, d, varest, xbar, s2x, n, m, ybar, s2y)
  }

g <- function(m, x, nrep=50) {
  tot <- 1 + (m-1)*x/m
  for(j in 2:nrep) {
    i <- seq(1, 2*j-3, 2)
    part3over2 <- paste("(x/1)*", paste("(x/(", 2:j, "*(m + ", i, ")))", sep="", collapse="*"), sep="")
    part3o2 <- eval(parse(text=part3over2))
    section <- ( (m-1)^(2*j-1)/(m^j) ) * part3o2
    tot <- tot + section
    }
  tot
  }

# example
x <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 8, 3, 1, 1)
delta(x)
```

Annex 4: Working Document 1

Simulation exercise

Simulation exercise

Abundance estimation for stratified random survey using intrinsic geostatistics

Context

A stratified random survey is used. Because of the a priori ignorance of any directional effect (anisotropy) or conditioning factor (e.g. isodepth), strata are taken as square blocks dividing the 120 x 120 domain. A number of N x N blocks is thus chosen, with N even to have go-and-return pairs of “transects”. A number of 6 x 6 = 36 blocks of size 20 x 20 is chosen. This will allow taking one random sample of size 2 or 3 within each block, or two samples with size 1, within the duration limit. For a better comparison between sizes, the 36 locations for size 2 and 3 are the same and correspond to one of the two locations per block for size 1.

An estimate of the mean fish density over the domain can be obtained by the arithmetic mean of the data. The abundance is obtained by multiplying by the total area. Geostatistics allows to compute (an estimate of) the estimation variance, to describe the spatial structure, and to make map by kriging interpolation.

Brief about geostatistics used here

Support

A basic geostatistical concept is the **support** of the variable, i.e. in 2D the generic area (in shape and orientation) on which the fish density is considered: this can be the total domain considered (say V), the 20 x 20 nm block size (say v), or the size of the sample, that is, the trawled area for the haul. In the following, what is considered as a point (say x) will be in fact the size of samples (or the size 1 if several sizes are considered together).

Let z(x) (or Z(x)) be the fish density at location x. The abundance is the sum of the densities over the domain:

$$Q = \int_V z(x)\,dx$$

directly linked by Q = |V| Z(V), to the mean density over the domain:

$$Z(V) = \frac{1}{V}\int_V Z(x)\,dx$$

Variogram

In intrinsic geostatistics (working within such a domain V with fixed boundaries, as opposed to transitive geostatistics), the spatial structure is described by the variogram, which measures the mean half variability between two points, as a function of their distance.

The experimental, or sample, variogram is classically computed from sample points as:

$$\gamma^*(h) = 0.5\frac{1}{N(h)}\sum_{x_i - x_j \sim h}[z(x_i) - z(x_j)]^2 \quad :$$

where $N(h)$ is the number of pairs of points (x_i, x_j), separated by the vector distance h, in the summation.

Then this is fitted by a variogram model $\gamma(h)$, representing the expectation of 0.5 $\mathrm{E}[Z(x+h) - Z(x)]^2$. By definition, the variogram is 0 for distance 0. The model often is the sum of different components:

- a discontinuity from distance 0, the nugget effect (this corresponds to the variance of the spatially uncorrelated component of the variable, including the variance of a random error, if any);
- and one or more continuous components ("spherical", exponential, linear…).

When "regularizing" a variable from a given support to a multiple support, the histogram is regularized (less extreme values, lower variance within a domain), and the variogram is regularized (showing more spatial continuity).

In fact each component of the variogram is regularized. In particular, the nugget effect is reduced, for instance by a ratio of 3 when going from a support of size 1 to a support of size 3.

Estimation variances

The variogram allows the computation of the **estimation variance** (the **variance of the error**) when estimating the mean density:

$$Z(V) = \frac{1}{V}\int_V Z(x)\,dx$$

over a domain V, by the average:

$$Z(V)^* = \frac{1}{N}\sum_i Z(x_i)$$

on the set *I* of samples x_i.
The estimation variance is:

$$\sigma_{\mathrm{E}}^2 = 2\bar{\gamma}(I,V) - \bar{\gamma}(V,V) - \bar{\gamma}(I,I)$$

where for instance $\bar{\gamma}(I,V)$ is the mean of $\gamma(x-y)$ when x describes *I* and y describes V independently.

In particular a quantity like $\bar{\gamma}(v,v)$ represents the **variance of point values within the support v**. It also represents the estimation variance of a support v by one random point in v, while the estimation variance by n independent random points in v is $\bar{\gamma}(v,v)$/n.

So, in the case of a stratified random sampling, with N strata having the same support and n points within each strata, the estimation variance of the domain is $\bar{\gamma}(v,v)$/(n N).

Note: if n > 1, the term $\bar{\gamma}(v,v)$ can be replaced by the mean of the variances of point values within each strata; this variance, and so the estimation variance of the domain, can be computed directly without geostatistics (J. Simmonds).

Ordinary Kriging

This allows to estimate the value over a domain, a block or at a target point (for mapping), by a linear combination $\sum \lambda_i Z(x_i)$, with $\sum \lambda_i$ = 1, of sample values $Z(x_i)$. The weights are chosen so as to minimize the estimation variance (which can be computed by the variogram).

Annex 5: Working Document 2

Gaming Exercise *by* Jean V. Adams

Working Document prepared for the ICES Workshop on Survey Design and Analysis II
9-13 May 2005, Sète, France

Gaming Exercise

Jean V. Adams
U.S. Geological Survey Great Lakes Science Center
Marquette Biological Station, 1924 Industrial Parkway, Marquette, MI 49855, USA, jvadams@usgs.gov

1. *Sampling designs*

With no prior information on the fish population to be surveyed, the surveys were designed to have broad coverage of the area within the given time frame (10 nm/hr, 1.5 hr/sample, 216 hours total). In all cases, locations of initial samples were determined systematically.

1.1 *Grid*

The Grid design had eight transects with eight samples each, covering the sample space in an evenly spaced sampling grid (Figure 1). The cruise track covered 956 nautical miles, and yielded 64 samples, for a total time of 191 hours (Table 1). The resulting catches are shown in Figure 2.

1.2 *Transect*

The Transect design had 4 transects with 24 samples each, concentrating closely-spaced sampling along one axis of the sample space, at the expense of sampling effort along the other axis (Figure 1). The cruise track covered 565 nautical miles, and yielded 96 samples, for a total time of 201 hours (Table 1). The resulting catches are shown in Figure 2.

1.3 *Adaptive*

The Adaptive design was a modification of the Transect design, with fewer initial samples along each transect, such that there was enough time left to take two additional adaptive samples. The Adaptive design had 4 transects with 15 initial samples each (Figure 1). Upon completion of each transect, two additional adaptive samples were taken on either side of the site that yielded the largest catch on the transect. Adaptive samples were located at the midpoint between previously sampled sites (Figure 1). The cruise track covered 816 nautical miles (on average), and yielded 60 samples, for a total time of 190 hours (on average, Table 1). The resulting catches are shown in Figure 2.

2. *Estimators*

2.1 *Design-based*

2.1.1 *Strata*

I arbitrarily defined 16 strata, each stratum a square with 30 nautical miles on a side. For each survey design, I assumed that initial systematic samples and additional adaptive samples were, in fact, generated from a stratified random survey. Global abundance (number of fish) and its variance were estimated for each field using standard methods (Table 2, Cochran 1977).

2.1.2 *Collapsed strata*

Additionally I re-analyzed the data using location-defined (collapsed) strata. A regularly-spaced square grid was expanded until each square grid contained at least two samples. Again, all samples were assumed to be generated from a stratified random survey, and abundance was estimated using standard methods (Table 2). Note that in the case of the Grid design, the collapsed strata were identical to the originally defined 16 strata.

2.2 Model-based

2.2.1 Spline

Total fish abundance was estimated based on the assumption that the relation between fish numbers and the x and y coordinates could be described by a smoothed surface. The relation was fit to the data using an additive model,

$$total = \begin{cases} s(x)+s(y) & s(x)+s(y) \geq 0 \\ 0 & s(x)+s(y) < 0 \end{cases},$$

where *s*() is a nonparametric smoothing spline function with four degrees of freedom. Predictions were made across the entire sample space (each 0.25 nm^2 pixel). Total abundance was calculated as the sum of these predictions, and variance of the estimator was estimated using bootstrap resampling (Table 2).

2.2.2 Geostatistical

For each survey design and field, I fit an exponential variogram to the catch data, with the minimum number of pairs set to 20 and the maximum distance set to 80 nm. Ordinary kriging was used to predict catch across the entire sample space (each 0.25 nm^2 pixel). Total abundance was calculated as the sum of these predictions (Table 2). Variance of the estimator was estimated using simulations applied to a broader grid (6 nm H 6 nm squares) and bias-corrected boostrap confidence limits were estimated (Table 2).

3. Discussion

Use of the Transect design in combination with the Collapsed strata estimator generated abundance estimates with the lowest residual standard errors (RSEs) for both fields (Table 2, Figure 3). This may be a result of spatial autocorrelation in the population, in which case sampling along a transect yields more similar catches than randomly distributed (or evenly spaced) samples.

The Transect design always yielded estimates with lower RSEs than the Adaptive design. This may largely be a function of sample size, because the application of adaptive sampling to transects resulted in 29% fewer samples, given the time constraints of the gaming exercise. Or it may be the result of the adaptive sampling tending to add samples with higher catch, which would increase the variability of the estimate.

For those surveys that used the Grid design, the Geostatistical estimator yielded estimates with the lowest RSE. This may be a result of spatial autocorrelation in the population, in which case accounting for autocorrelation in the samples using geostatistics (for which the grid design is best suited) should reduce the variability in the estimated abundance.

Predictions from the Collapsed strata estimator and the Geostatistical estimator are compared visually for both fields with the Grid and Transect designs in Figures 4 and 5.

4. References

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

Table 1. Design of three surveys, including number of initial and adaptive samples, total distance of cruise track, and total survey time.

Design	No. Samples Initial	Adaptive	Total	Distance (nm)	Time (hr)
Grid	64	0	64	956	192
Transect	96	0	96	565	201
Adaptive	60	8	68	876 [1]	190 [2]

[1] Distance is not fixed due to the adaptive samples. Reported is the mean distance covered in applying this survey to Fields 1 and 2 (879 and 873 nm).

[2] Time is not fixed due to the adaptive samples. Reported is the mean time taken in applying this survey to Fields 1 and 2 (190 and 189 hrs).

Table 2. Estimates of global fish abundance (in millions) in two simulated fish population fields, using three survey designs, and four estimators. Variance of the estimator is reported as standard error (SE), relative standard error (RSE), and 95% confidence intervals.

Design	Estimator	No. Strata	Field 1 Total	SE	RSE	95% CI	Field 2 Total	SE	RSE	95% CI
Grid	Strata	16	8.8	2.6	30%	(3.6, 14.0)	9.6	1.7	17%	(6.3, 12.9)
	Collapsed	32	8.8	2.6	29%	(3.6, 14.0)	9.6	1.5	15%	(6.6, 12.6)
	Spline	0	9.7	2.7	27%	(5.3, 15.1)	10.4	1.8	17%	(7.3, 13.8)
	Geostatistical	0	8.9	1.7	20%	(5.3, 12.3)	9.6	1.1	11%	(7.5, 11.7)
Transect	Strata	16	10.5	2.1	20%	(6.2, 14.7)	7.5	0.8	11%	(5.8, 9.2)
	Collapsed	48	10.5	1.8	17%	(6.8, 14.1)	7.5	0.6	9%	(6.2, 8.8)
	Spline	0	11.3	2.3	20%	(7.2, 16.0)	8.3	1.1	13%	(6.2, 10.4)
	Geostatistical	0	10.5	2.5	24%	(5.4, 15.5)	7.5	1.1	15%	(5.2, 9.5)
Adaptive	Strata	16	11.7	2.5	22%	(6.7, 16.8)	7.4	0.9	12%	(5.6, 9.2)
	Collapsed	20	10.5	1.9	18%	(6.8, 14.2)	7.1	0.9	13%	(5.3, 9.0)
	Spline	0	13.3	3.1	24%	(8.5, 20.6)	8.5	1.2	14%	(6.1, 10.7)
	Geostatistical	0	10.4	3.0	29%	(4.3, 16.3)	6.8	1.2	17%	(4.4, 9.0)

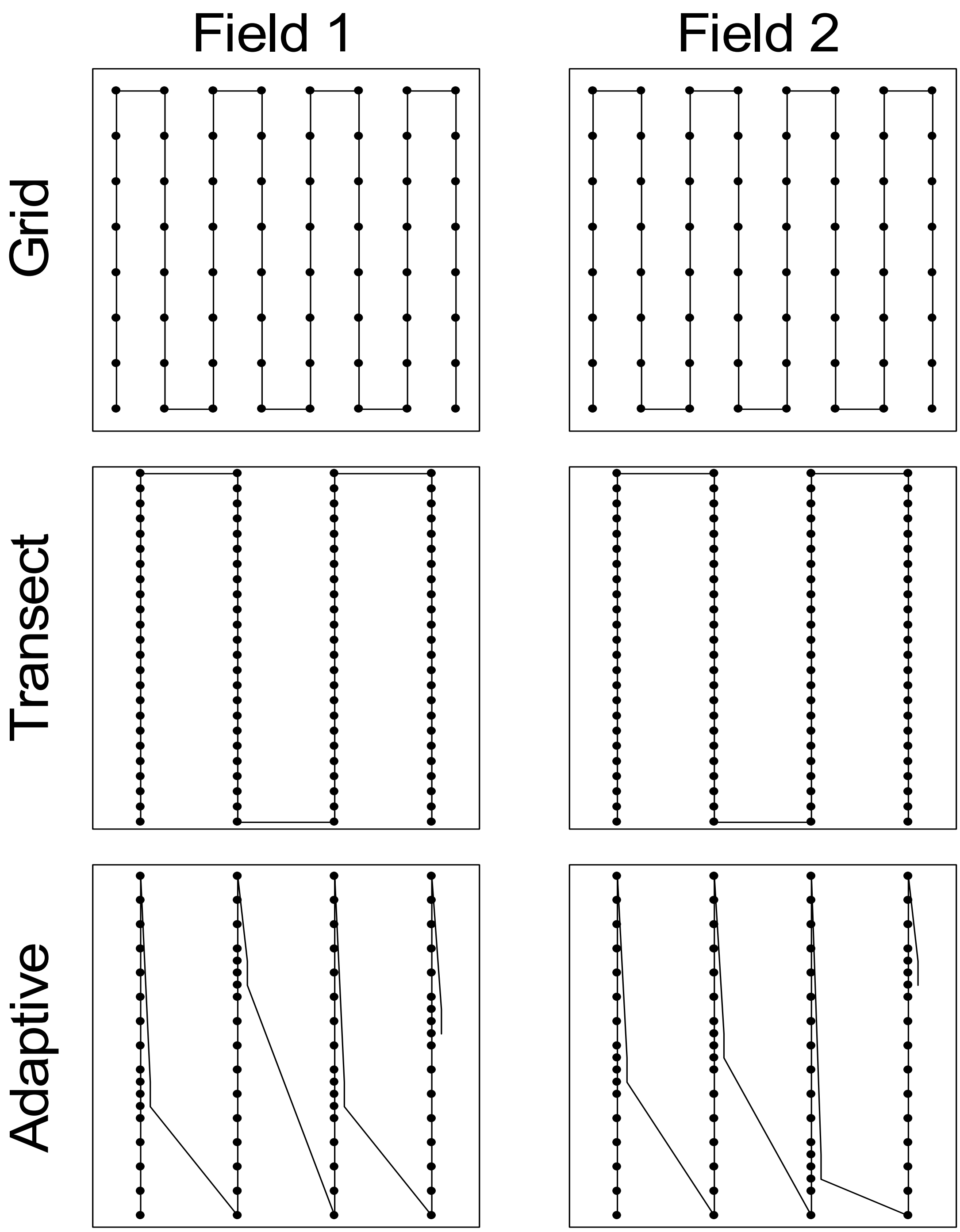


Figure 1. Cruise tracks of the three survey designs applied to the two fish population fields. To distinguish overlapping cruise tracks in the Adaptive design, the path was drawn offset to the right for the adaptive samples. For clarity, the first and last legs of the cruise, from and to the origin (0, 0), are not shown. However, the distances of these legs were included in the calculations of survey distances and times.

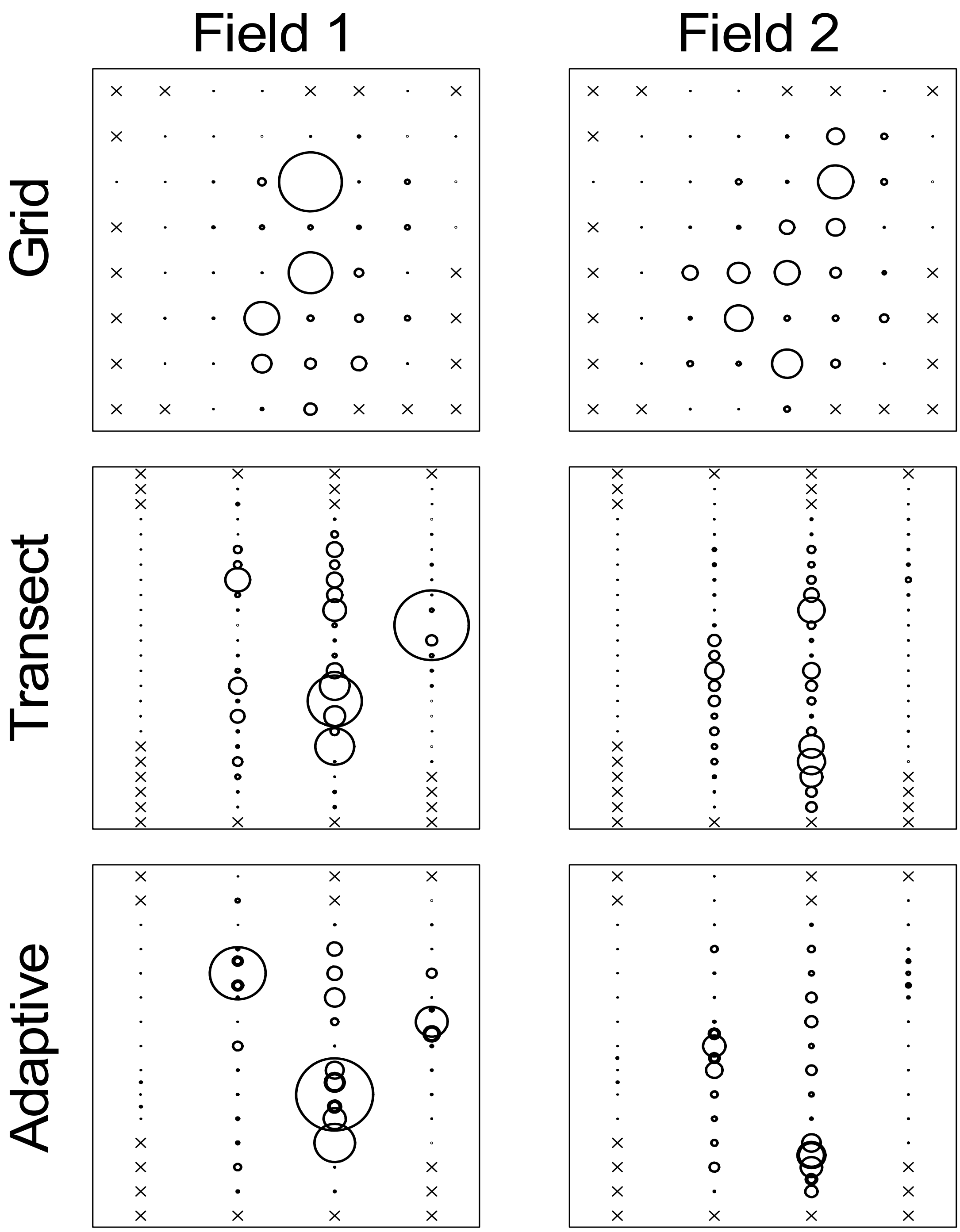


Figure 2. Catch from two simulated fish population fields using three different survey designs. Circle size represents relative size of catch, ranging from 0.04 to 44,010 fish per square nautical mile, Xs represent samples with zero catch. Circles representing adaptive samples were drawn using bold lines.

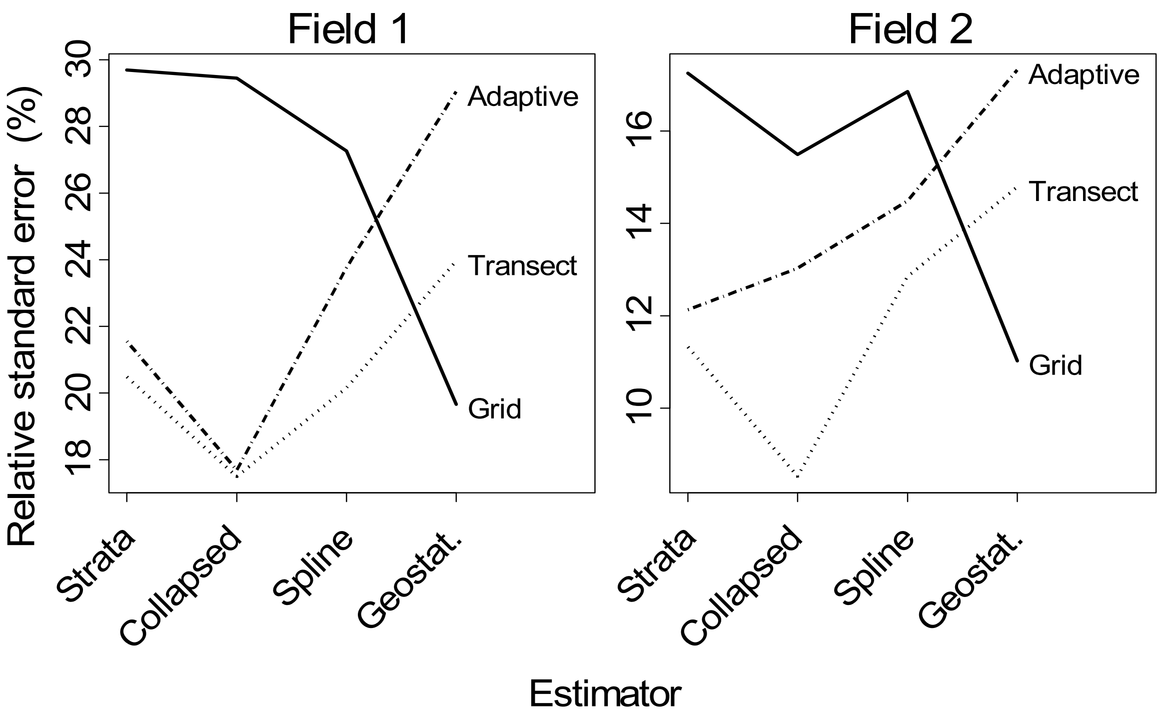


Figure 3. Relative standard error of total abundance estimates from two simulated fish populations fields, sampled with three survey designs (Grid, Transect, Adaptive), and analyzed with four estimators (Strata, Collapsed, Spline, and Geostatistical).

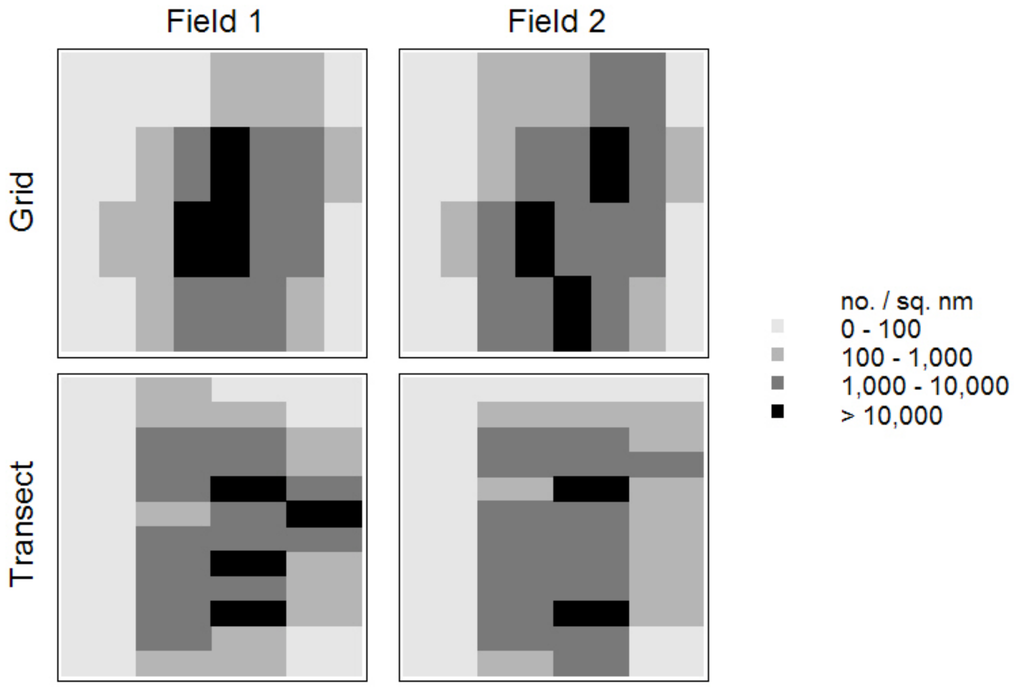


Figure 4. Map of fish abundance predicted over the entire sample space, based on the collapsed strata estimator for two simulated fish population fields, sampled with two survey designs (Grid and Transect).

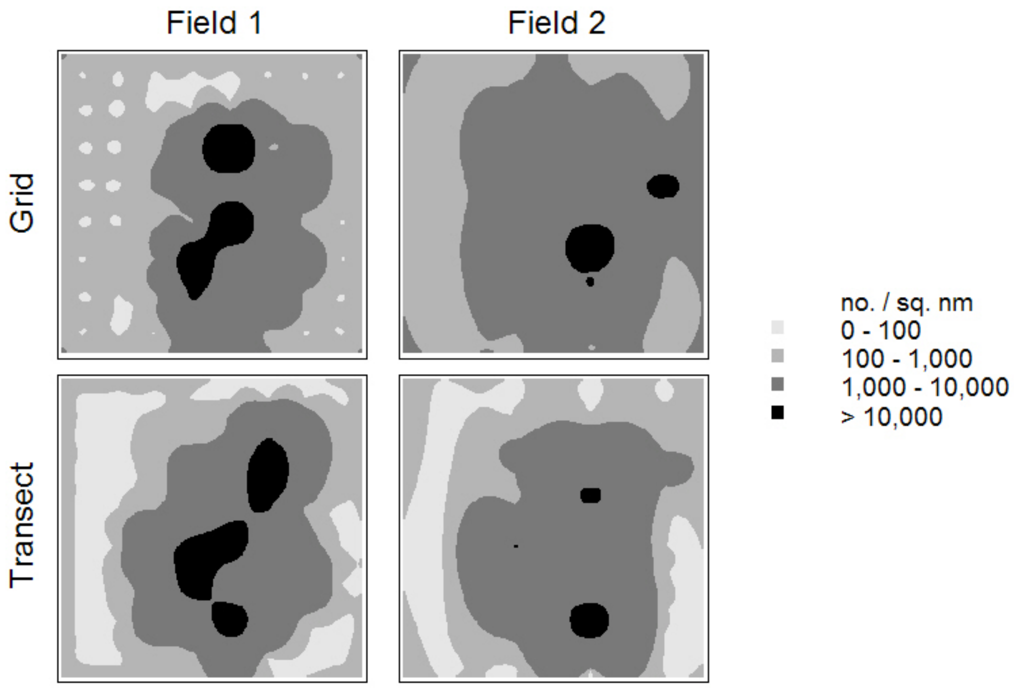


Figure 5. Map of fish abundance predicted over the entire sample space, based on the geostatistics estimator for two simulated fish population fields, sampled with two survey designs (Grid and Transect).

Annex 6: Working Document 3

Results of simulation study for ICES WKSAD *by* Robert Brown

Results of simulation study for ICES WKSAD

Robert Brown
April 2005

Introduction

I was asked to undertake this simulation analysis by the staff that attended the workshop last year. However, the background information and resources for this were limited and as a result the analysis conducted is quite basic. It's now been agreed that I'll attend the workshop and I hope that I'll be able to devote more resources to this work and make a larger contribution in the future.

Method

Three major methods of sampling are implemented for comparison. The first survey used was a simple systematic survey along a grid chosen as the most efficient within the constraints. The first study also has a few additional points chosen at random along the cruise path to maximise the use of the available time. The second survey used the same grid with the sample points along this path selected at random. The third sample used clusters of three samples, separated by 2 n.m. around sample points chosen at random along the same grid. In all cases it is assumed that the sampling is 100% efficient and that the catch for the 0.25 n.mi.^{2} represents the actual abundance in that area.

Global abundance

This section presents the requested results for global abundance and relative standard error (RSE) for the different surveys and different fields. These were calculated using the survey command in Stata 8 to account for clustering in the third survey.

Survey 1

Survey	Abundance	RSE	n	Track length
1a	30.5 million	29.8%	78	
1b	37.6 million	20.0%	78	

Survey 2

Survey	Abundance	RSE	n	Track length
2a	34.1 million	28.7%	82	
2b	37.0 million	29.5%	82	

Survey 3a

Strata	Clusters	Abundance	RSE	n	Track length
1 strata	yes	43.2 million	40.6%	78	
1 strata	no	43.2 million	26.2%	78	

Survey 3b

Strata	Clusters	Abundance	RSE	n	Track length
1 strata	yes	40.0 million	29.6%	78	
1 strata	no	40.0 million	19.7%	78	

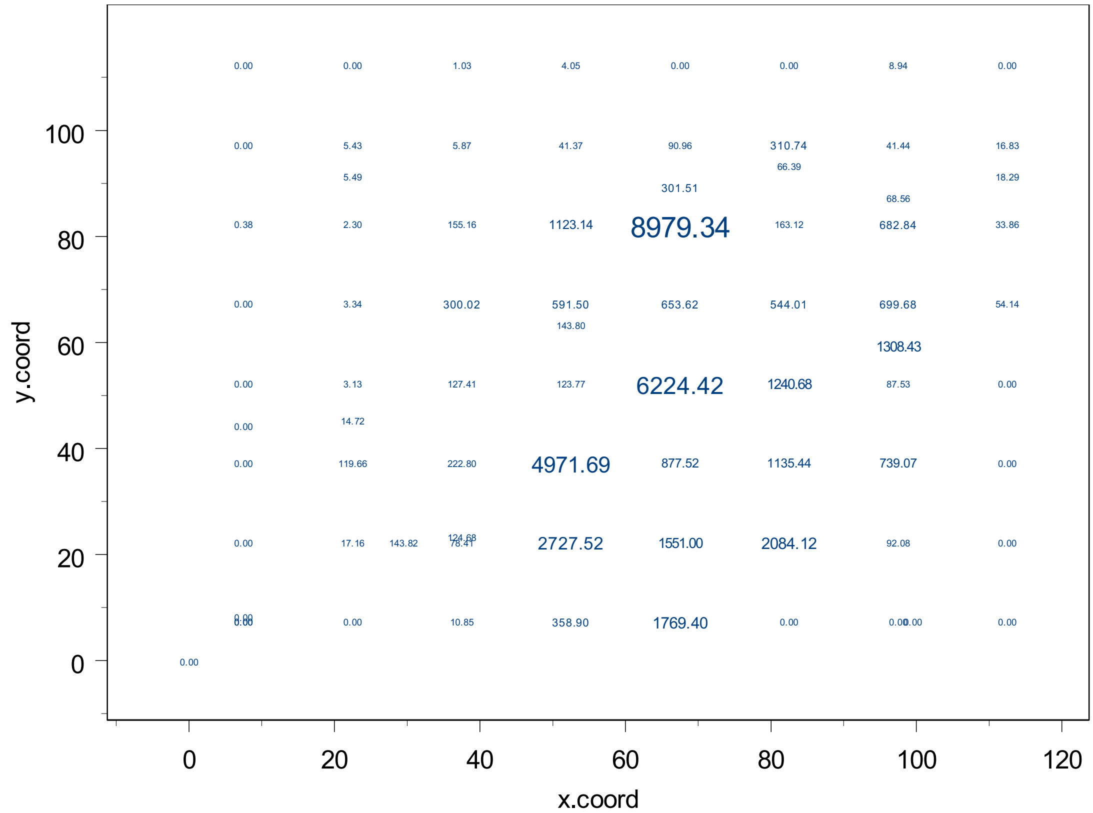
Interpretation of the results will depend to a large extent on the details of the generating data, which are not available at present. However, it can be noted that clustering reduces the precision of the estimates.

Maps of fish distribution

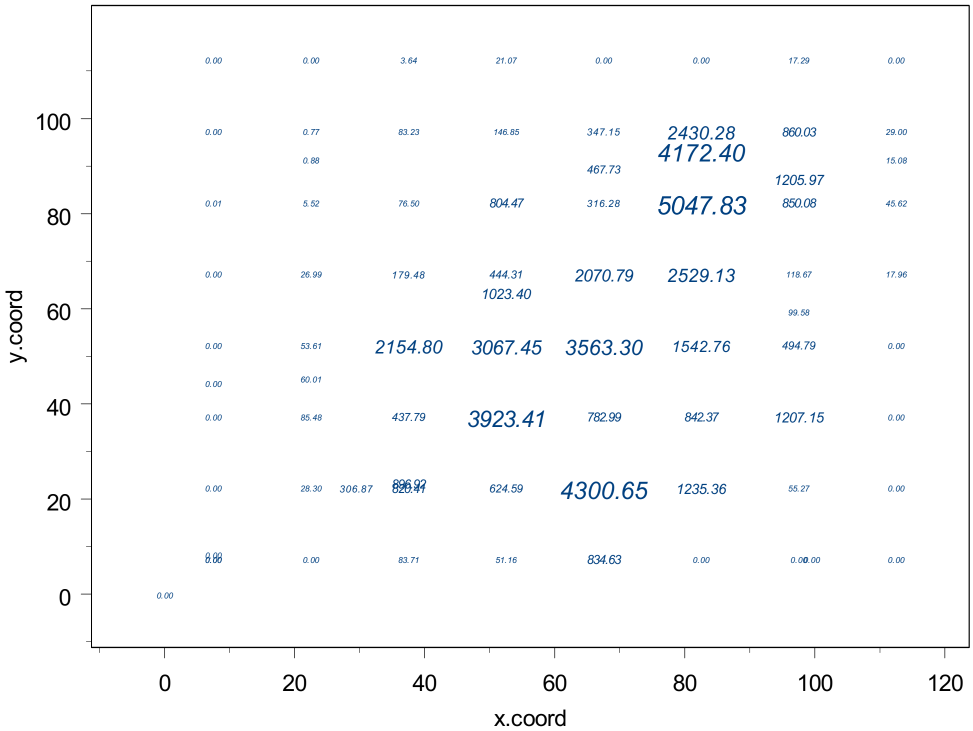
The records are plotted as bubble plots using S-plus. The figures are the results of the sample at that point. The size of the figures is related to the value recorded. Contour maps are developed using the contour plotting function in S-plus, using loess smoothing set subjectively by comparison with the bubble plot. The values of the contours are in the units of the sampling i.e. number per 0.25 n.mi^2.

a. Bubble plots

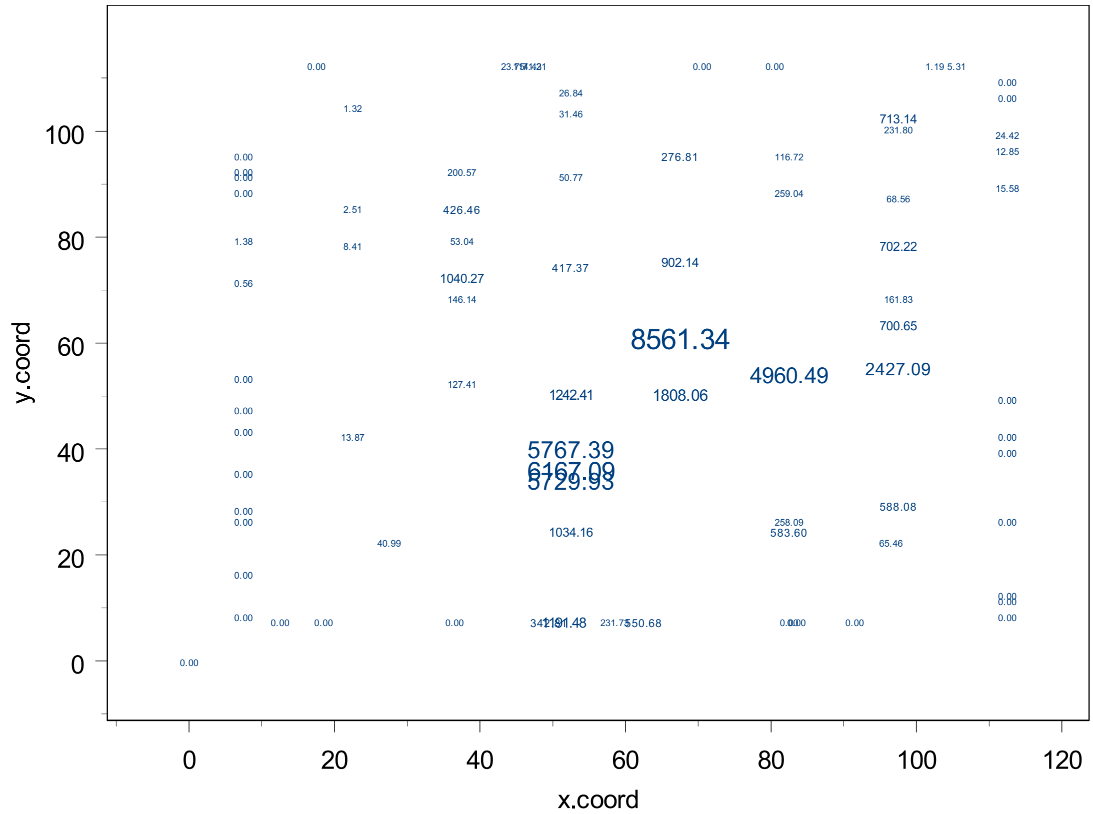
Survey 1a



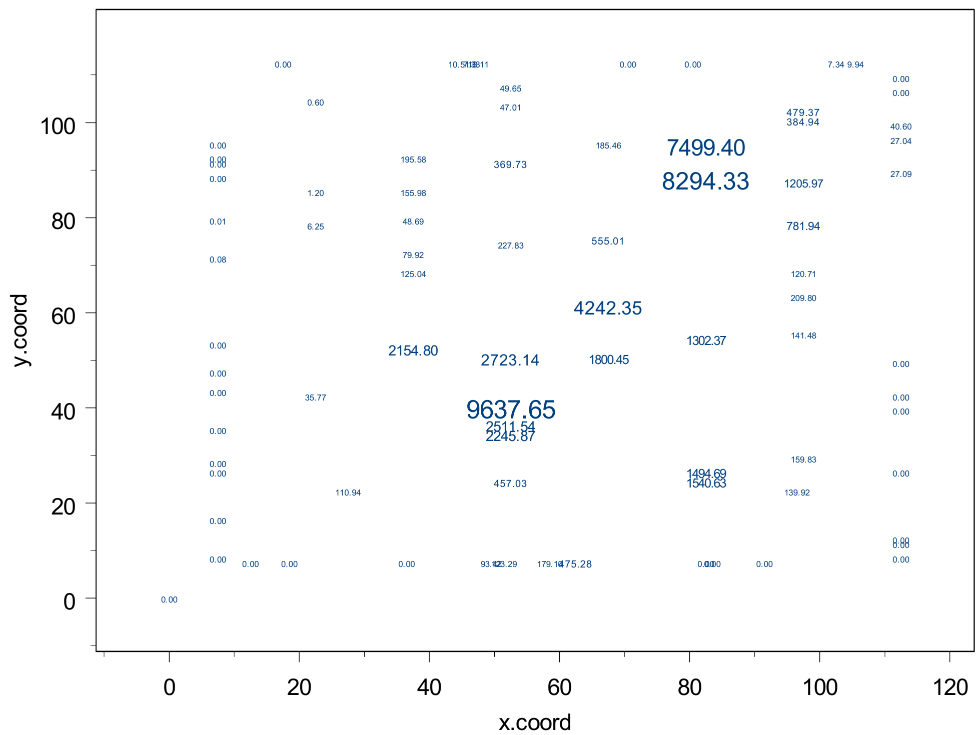
Survey 1b



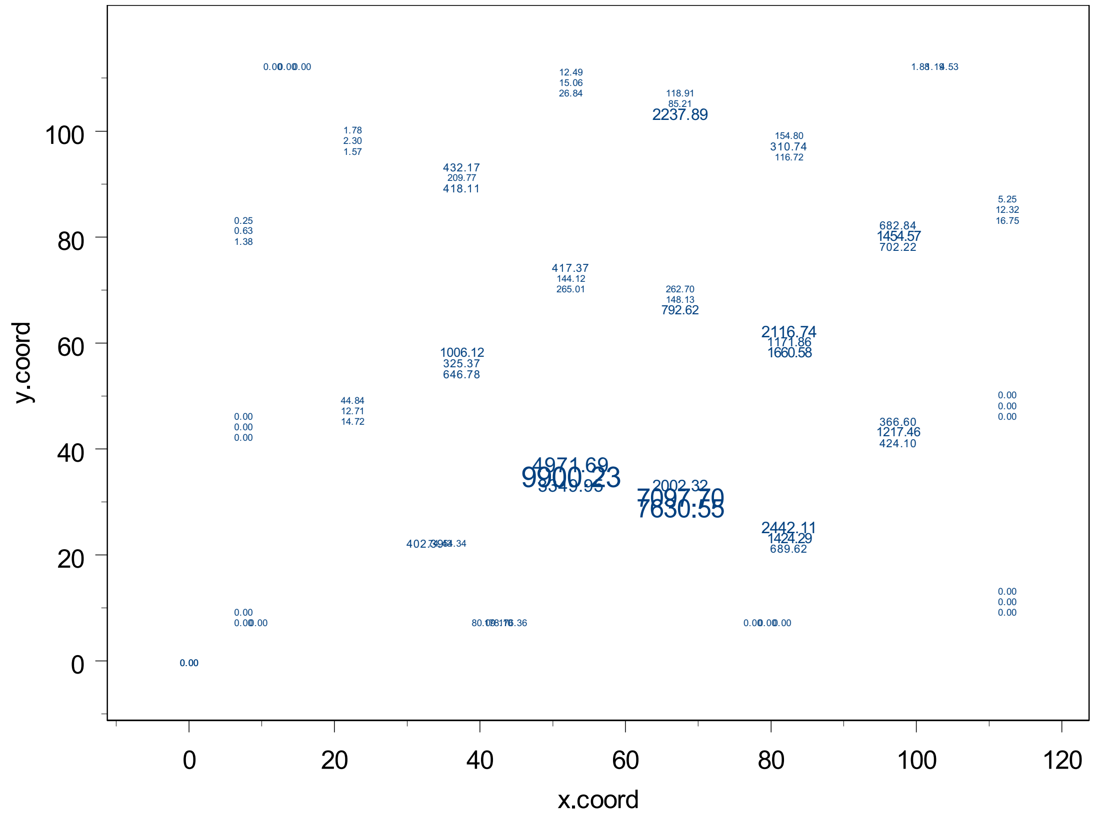
Survey 2a



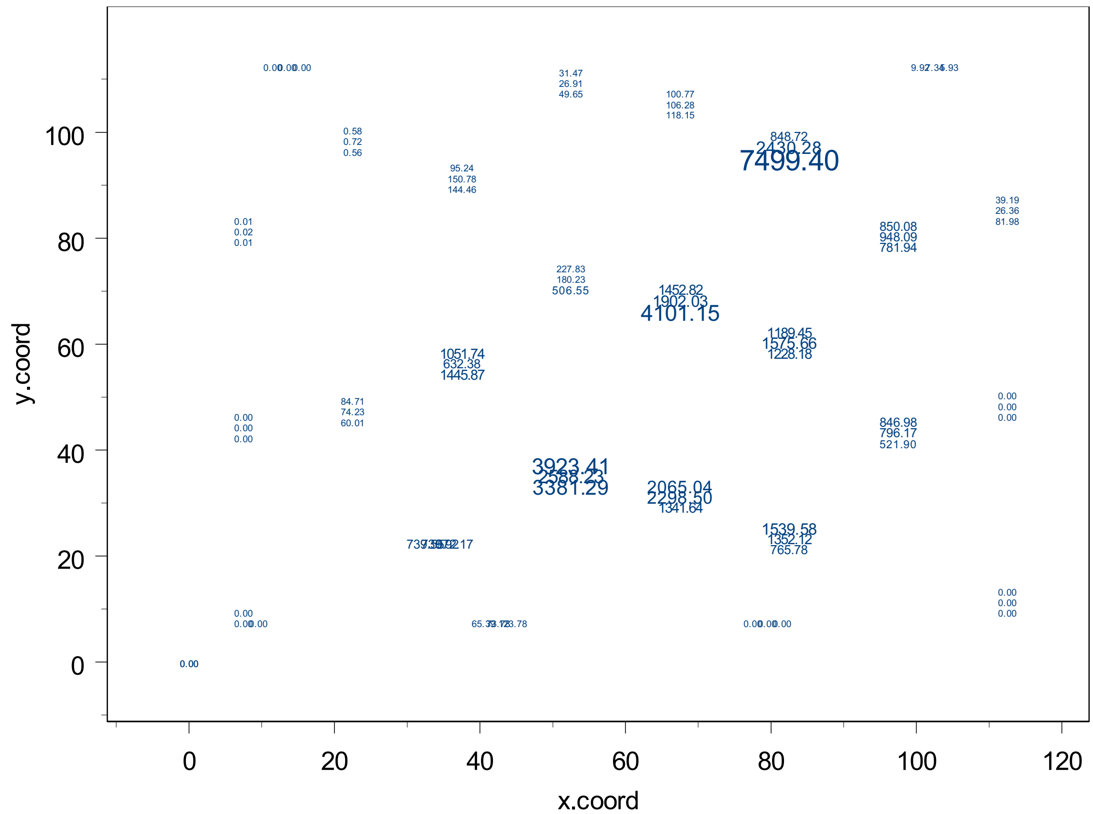
Survey 2b



Survey 3a

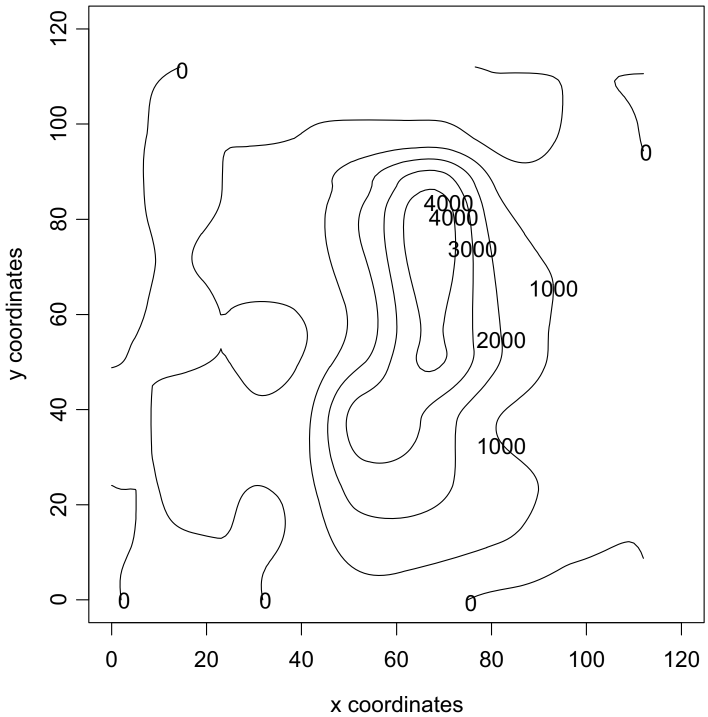


Survey 3b

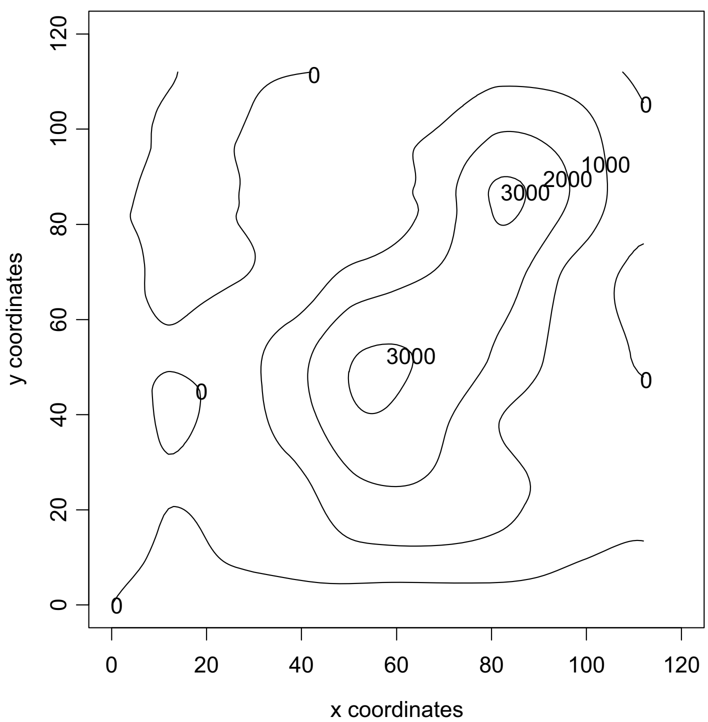


b. Contour maps

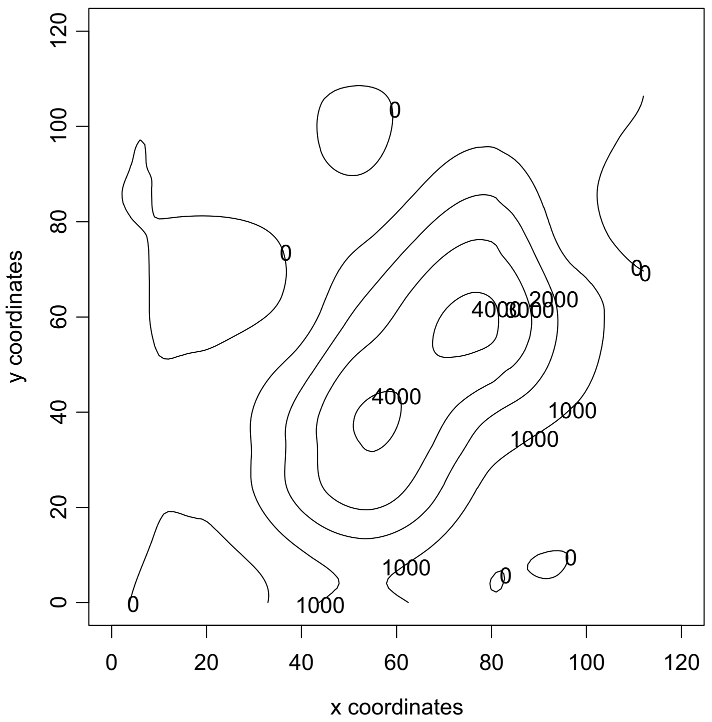
Survey 1a, using loess smoothing with span = 2



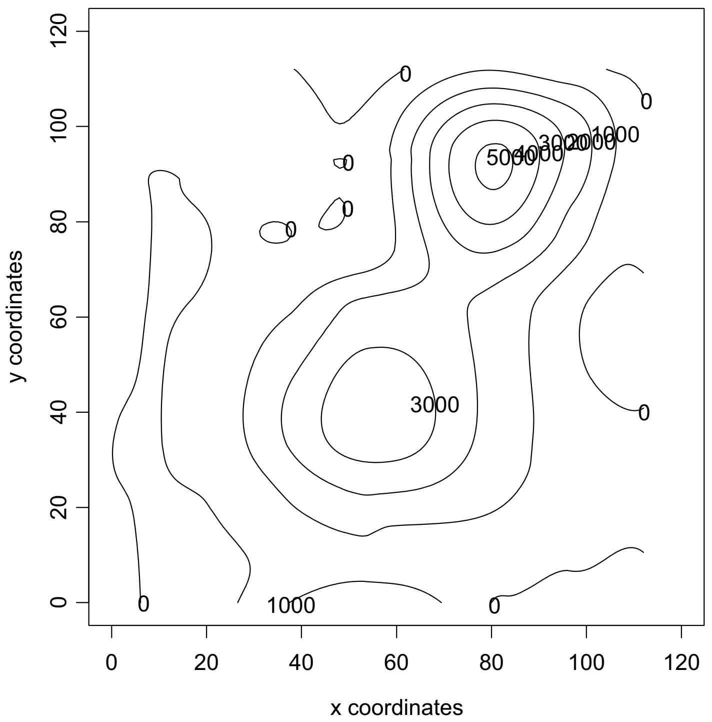
Survey 1b, using loess smoothing with span = 0.3



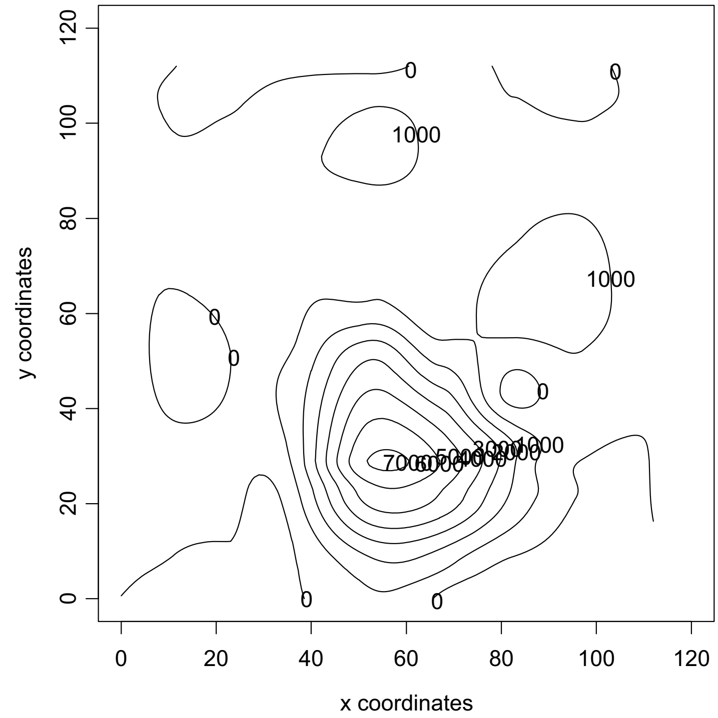
Survey 2a using loess smoothing with span = 0.35



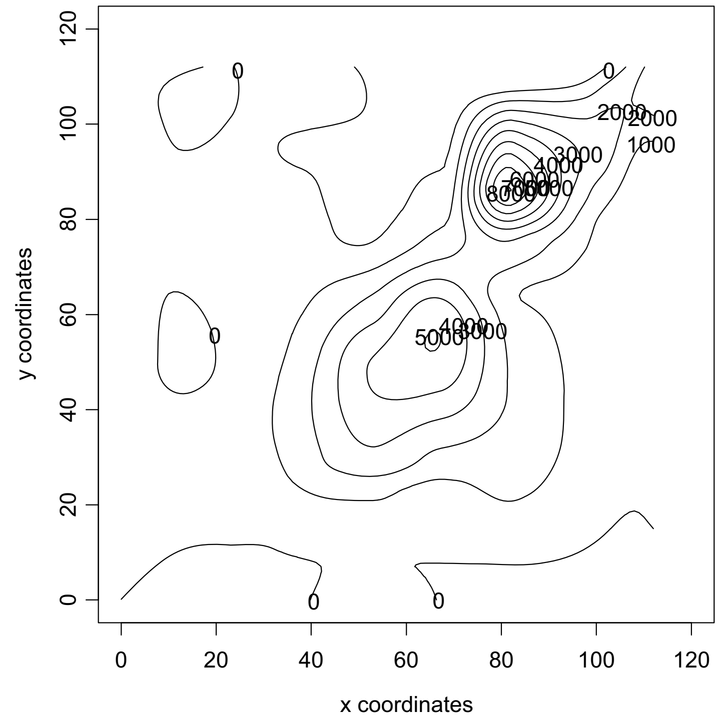
Survey 2b using loess smoothing with span = 0.3



Survey 3a using loess smoothing with span = 0.25



Survey 3b using loess smoothing with span = 0.25



Interpretation

The contour plots are heavily influenced by occasional extreme values resulting from the large small scale variation. This is particularly noticeable in plot 1a and might be expected considering the low autocorrelation and high nugget. In consideration of this it can also be noted there is far more agreement about the distribution for field 2. In this circumstance, for field 1 in particular, the clustered records may provide value in assessing distribution. In all cases relatively large span values have been used to reduce small scale variation which is viewed as having little validity with this level of sampling.

Annex 7: Working Document 4

Confidence intervals for trawlable abundance from random stratified bottom-trawl surveys *by* Noel Cardigan

Confidence intervals for trawlable abundance from random stratified bottom-trawl surveys

Noel Cadigan

Fisheries and Oceans Canada

St. John's, NL Canada

A1C 5X1

Abstract

An approximately pivotal statistic is proposed that can be used to construct confidence intervals about average and total trawlable abundance from stratified random bottom-trawl fisheries surveys. The statistic is based on the strata area-weighted average that is commonly computed from the survey catches. The distribution of the statistic is derived from both the random selection of sites to survey and the random fish capture process at a site. This is in contrast with the commonly used "design-based" approach to statistical inference that includes only the randomness in the sites selected for trawling. The method is applied to case studies, and simulations based on these case studies are used to examine the coverage accuracy of the confidence intervals.

1 Introduction

Bottom trawl research surveys provide important information for the assessment and management of many fish stocks worldwide. In a bottom trawl survey, data are collected by towing a standard research trawl at a constant speed and usually for a fixed duration of time. A variety of other abundance and biological information for many species is also collected. We focus on the total number or weight of a species. Essentially the unit of measurement is a three-dimensional quadrat sample of relative stock abundance. Measurements are relative because not all fish in the path of a trawl are caught, and a portion of the local abundance will be above the trawl and also not caught. A review of bottom trawl surveys is given in Gunderson (1993).

In the Northwest Atlantic Ocean, fisheries surveys are usually based on stratified random sampling. For example, the annual multi-species bottom-trawl surveys conducted by the Canadian Department of Fisheries and Oceans (DFO) use this sampling design. Doubleday (1981) presented a comprehensive manual on the survey methodology. Strata boundaries are primarily based on depth, which is important in delineating distributional patterns of some species. Strata are also divided to be spatially contiguous. Stratification is useful when the catch variability within a strata is less than the variability between strata. An example of the stratification for Northwest Atlantic Fisheries Organization (NAFO) Subdivision 3Ps is given in Figure 1. Most NAFO management areas are shown in Figure 2.

Each time a station is trawled many variables are measured, such as the time and location of a tow, light conditions, wind and sea states, depth of a tow, water temperature, and the number and weight of every species caught. Samples of the catch are taken to measure individual fish lengths, whole and gutted weights, liver and gonad weights, stomach contents, sexual and maturity status, and presence of parasites. Otoliths (ear bones) are also taken for age determinations in a lab. We consider only to the numbers caught per tow or the catch weight; however, clearly the surveys provide much more information than this.

The objective of most fisheries surveys is to provide inferences about the populations that are surveyed. Statistical inferences are usually based on the so-called "design-based" approach (e.g. see Cochran, 1977); however, a variety of other approaches involving models have also been used (e.g. Smith, 1990; Sullivan, 1991; Pelletier and Parma, 1994; Malinen and Peltonen, 1996; Warren, 1997). In the survey design-based approach the catch is treated as a fixed quantity and inferences are based only on the random selection of sites to survey. Design-based inferences are focused on what the catch might be if the entire population (i.e. all sites or quadrats) were sampled. For fisheries surveys this is insufficient. Even if it were possible to sample all sites, the survey would still result in imprecise information because the survey trawl provides an uncertain and proportional measurement of stock size at the tow site. In a good survey the predominate factor affecting the catch should be the local stock size available to the trawl, which we refer to as local trawlable abundance. However, other factors may affect the catch so that it is at best a random measurement of local trawlable abundance. Gunderson (1993) and Godø (1994) gave comprehensive reviews of factors affecting trawl catches.

Let N_i be the local stock size at tow site i. The trawl usually catches only a fraction q of N_i, and $\mu_i = qN_i$ is local trawlable abundance. If we denote the survey catch at site i as R_i then we expect in many idealized replications of the trawl sampling process that the average catch should equal μ_i. We denote this stochastic model expectation as

$$E_\xi(R_i) = \mu_i. \tag{1}$$

If the total number of sites in the survey area is M then the average trawlable abundance for the survey area is $\mu = M^{-1}\sum_{i=1}^{M}\mu_i$ and the total trawlable abundance is $\mu_M = M\mu$.

In this paper we consider improved confidence intervals (CI's) for μ and μ_M compared to the standard ones based on the sampling variance of $\hat{\mu}$ and the "t-statistic"

$$T_s = \frac{\hat{\mu} - \mu}{\widehat{Var}^{1/2}(\hat{\mu})} \tag{2}$$

The denominator in (2) is the square root of an estimate of $Var(\hat{\mu})$. The improved CI's will be described in the next section. Our inferences are based on two sources of variability. The first is local variability in trawl catches, $Var_\xi(R_i)$, and the second is the random selection of survey sites. We restrict our attention to the commonly used area weighted-average estimator,

$$\hat{\mu} = \sum_{h=1}^{H} W_h \bar{R}_h, \tag{3}$$

where $W_h = M_h/M$ is the weight or fractional size of stratum h. This estimator is commonly computed and is unbiased if the trawl catches are also unbiased for local trawlable abundance; that is, if (1) is true. It may be possible to find more efficient estimators of μ; however, this is beyond the scope of this paper.

Of course the fraction of fish caught in most bottom trawls is length dependent. Changes in trawlable abundance may reflect changes in stock abundance as well as changes in size distribution. For simplicity we only consider total trawlable abundance; however, our CI's can also be applied to size disaggregated data.

2 Methods

Let $R_1, ..., R_M$ be random variable (rv's) for the trawl catches that could have been obtained at the M tow stations in the survey region. Observed catches were obtained from a random stratified survey design in which stratum h had M_h tow stations, $h = 1, ..., H$. We assume a sample of m tow stations was selected, with m_h stations chosen at random from each strata, where $m = \sum_h m_h$. We denote the sample of tow stations as s. The survey sample of observed catches is $\{r_i : i \in s\}$. Note that we use the lower case notation for observations of the rv R. We know from standard sampling theory that $E_D(\hat{\mu}) = M^{-1} \sum_{i=1}^{M} R_i$ and

$$Var_D(\hat{\mu}) = \sum_h W_h^2 \left(1 - \frac{m_h}{M_h}\right) \frac{S_{R,h}^2}{m_h},$$

where $S_{R,h}^2 = (M_h - 1)^{-1} \sum_{i=1}^{M_h} (R_i - \bar{R}_h)^2$ and $\bar{R}_h$ is the average catch at all tow sites in stratum h. Note that we use the suffix D when we refer to expectations with respect to the randomness in the survey design, and we use the suffix ξ when we refer to expectations with respect to the randomness in the trawl sampling process. These two types of expectations are described in more detail in Section 6.4 of Särndal et al. (1992). We restrict our attention to $\hat{\mu}$. Estimates and CI's for μ_M may be obtained from multiplying by M the corresponding result for μ.

The Negative Binomial (NB) distribution is often suggested to be appropriate for modelling the variability in survey trawl catches (see Gunderson, 1993). The NB variance is

$$Var_\xi(R_i) = \sigma_i^2 = \mu_i + \mu_i^2/k.$$

There are a variety of generating mechanisms that can produce this type of over-dispersion (e.g. see Cameron and Trivedi, 1998). We use this distribution to develop a variance expression for $\hat{\mu}$ that is also useful for constructing CI's.

As stated previously, as long as (1) holds then the total expectation of $\hat{\mu}$ is

$$E(\hat{\mu}) = E_\xi\{E_D(\hat{\mu})\} = E_\xi\left(\bar{R}\right) = \mu.$$

The total expectation is with respect to both the sample selection and trawl catch processes. The variance is more difficult to derive; however, we can use 16.4.5 and 16.4.7 in Särndal et al. (1992) to show that $Var(\hat{\mu}) = V_{11} + V_2$, where V_{11} is the measurement variance component,

$$V_{11} = \sum_h \frac{W_h^2}{M_h m_h} \sum_{i=1}^{M_h} \sigma_i^2,$$

and V_2 is the sampling variance component,

$$V_2 = \sum_h W_h^2 \left(1 - \frac{m_h}{M_h}\right) \frac{S_{\mu,h}^2}{m_h},$$

where $S_{\mu,h}^2$ is defined similar to $S_{R,h}^2$.

The standard design-based estimator of $Var_D(\hat{\mu})$ is

$$v_{st}(\hat{\mu}) = \sum_h W_h^2 \left(1 - \frac{m_h}{M_h}\right) \frac{s_h^2}{m_h},$$

where $s_h^2 = (m_h - 1)^{-1} \sum_{i=1}^{m_h} (r_i - \bar{r}_h)^2$. Särndal et al. (1992) show that $v_{st}(\hat{\mu})$ underestimates $Var(\hat{\mu})$ when there is measurement error. To estimate $Var(\hat{\mu})$ we need to estimate V_2 and V_{11}, and the latter variance term requires estimates of σ_i^2 and the NB k parameter. We cannot estimate k without some replicate sampling at tow sites or additional assumptions about the distribution of μ_i within strata.

2.1 Constant within-strata means model

If there is little within-strata variation in the factors that affect trawlable abundance, or if these factors vary randomly within strata, then we could regard catches in a strata as replicate samples from a NB distribution with a common mean. The number of strata in surveys of Northwest Atlantic fish stocks tend to be relatively large. Fish also tend to be spatially autocorrelated along contours of equal depths, so it is reasonable to assume that stock densities are homogeneously distributed within strata. In this case $Var(\hat{\mu}) = V_{11}$. An approximately unbiased estimator of $Var(\hat{\mu})$ is

$$v_{csm}(\hat{\mu}) = \sum_h \frac{W_h^2}{M_h m_h} \sum_{i=1}^{M_h} \hat{\sigma}_i^2, \tag{4}$$

where

$$\hat{\sigma}_{i \in h}^2 = \frac{m_h \hat{k}}{m_h \hat{k} + 1} \left(\hat{\mu}_h + \frac{\hat{\mu}_h^2}{\hat{k}}\right),$$

and both $\hat{\mu}_h = \bar{r}_h$ and $\hat{k}$ are NB maximum likelihood estimators when the within-strata means are constant. We refer to this as the *csm* model. If $\hat{k}$ was known (i.e. not estimated) then $\hat{\sigma}_i^2$ would be exactly unbiased for σ_i^2.

We do not directly use $v_{csm}(\hat{\mu})$ in CI's for μ. Rather, we hypothesize that the major mode of variation in $\hat{\mu}$ is equal to $m^{-1}(\mu + \mu^2/k_p)$ and we use this to construct a CI. In many of the surveys we examine the sampling allocation is approximately proportional to the strata area in which case $M_h/m_h \approx M/m$ and

$$\begin{aligned} Var(\hat{\mu}) &\approx m^{-1}\sum_h W_h\left(\mu_h + \frac{\mu_h^2}{k}\right) \\ &= m^{-1}\left(\mu + k^{-1}\mu^2\right) + m^{-1}k^{-1}\sum_h W_h(\mu_h - \mu)^2. \end{aligned}$$

If $\sum_h W_h(\mu_h - \mu)^2$ is approximately equal to $\beta\mu^2$, then $Var(\hat{\mu}) \approx m^{-1}(\mu + \mu^2/k_p)$, where $k_p = k/(1+\beta)$. Empirical evidence (see **Results** section) suggests that between-strata variation in μ_h's is often approximately proportional to μ^2. At the least we expect that it should increase with μ.

For CI's we propose the approximately pivotal statistic

$$T_p = \frac{\hat{\mu} - \mu}{m^{-1/2}(\mu + \mu^2/k_p)^{1/2}}. \tag{5}$$

We define our estimator of k_p below. The p subscript indicates that this is a parameter that is pooled over strata, and $k_p < k$. A pivotal statistic has a known distribution that does not depend on unknown parameters. We assume that T_p is normally distributed. If k_p is not too small then a $(1 - 2\alpha)100\%$ CI for μ is

$$\frac{a}{2} \pm \sqrt{\frac{a^2}{4} + b}, \tag{6}$$

where

$$a = \frac{2\hat{\mu} + Z_\alpha^2/m}{1 - Z_\alpha^2/mk_p}, \text{ and } b = \frac{-\hat{\mu}^2}{1 - Z_\alpha^2/mk_p}.$$

Z_α is the lower a quantile from a standard normal distribution. If k_p is too small so that the denominator term is negative (i.e. $k_p < Z_\alpha^2/m$) then the upper confidence limit is infinite and the lower limit is the larger term in (6).

We use a moment-type estimator for k_p,

$$k_p^{-1} = \frac{mv_{csm}(\hat{\mu}) - \hat{\mu}}{\hat{\mu}^2}.$$

The is obtained by solving $\hat{\mu} + \hat{\mu}^2/k_p = mv_{csm}(\hat{\mu})$ for k_p. Note that although we used approximate proportional allocation to motivate the variance term in (5), we do not assume proportional allocation when computing $v_{csm}(\hat{\mu})$. Advantages of the T_p statistic are considered in the **Discussion** section.

2.2 Variable within-strata means model

The approach may still work even if the μ_i within a strata are not constant. If their variation is proportional to the square of the stratum mean, $\mu_h = M_h^{-1} \sum_{i=1}^{M_h} \mu_i$, and if the between-strata variation in μ_h's is also proportional to μ_h^2, then one can show with proportional allocation that $Var(\hat{\mu}) \approx m^{-1}(\mu + \mu^2/k_t)$. The k_t term will be smaller than k_p in the previous section; however, we can estimate it in the same way as before, and use the same procedure for CI's. However, we have not yet studied the reliability of the method in this situation.

3 Results

We computed CI's for population average survey catch using (2) and (5). We did this for annual time series of surveys for five stocks, which are listed in Table 1. Included in this table are the number of survey years, average catch, and average number of survey sets.

Table 1. Example species.

Species	Division	Years	Catch	Sets
Cod	2J+3KL	22	71	376
Cod	3Ps	22	36	137
American Plaice	3LNO (Fall)	15	120	346
American Plaice	3LNO (Spring)	20	124	341
Striped Wolffish	3Ps	22	0.6	154

The CI's are presented in Figure 3. Three points are apparent. The first is that the NB CI's are often considerably asymmetric, as expected. The second point is that the NB CI's often have a shorter width than the standard ones, and this was not expected. We consider this point further in **Section 3.1**. The third point which we anticipated is that the standard CI's can cover infeasible negative values (see 1995 3Ps cod; panel 2), whereas the NB intervals cover only feasible values. Occasionally the standard intervals are shorter than the NB intervals (e.g.

1992 and 1994 for 3Ps cod). However, the standard intervals were almost always shorter for 3Ps wolffish. In the **Discussion** section we consider why the NB method may not work well for this stock.

We used annual estimates of the NB k parameter when computing $v_{csm}(\hat{\mu})$. This was because likelihood ratio tests indicated significant inter-annual variation in k.

It is important to check the validity of the NB model assumptions. These seem necessary to accommodate for randomness in the trawl sampling; however, the specific NB variation or the csm model may not be appropriate. We examine deviance residuals obtained by fitting the NB csm with k estimated annually. The results are presented in Figure 4. The residuals for the two plaice stocks (panels 3 and 4) do not indicate any violations in the model assumptions. The residuals for the two cod stocks (panels 1 and 2) are somewhat skewed, with a heavier left-tail than the normal distribution. This suggests extra-variation in catches with small μ's than the NB model can accommodate. The wolffish survey data (panel 5) produced residuals with heavier tails than NB variation could accommodate.

Another assumption we made was that the between-strata variation in μ_h's was proportional to μ^2. In Figure 5 we plot the square root of estimates of $V(\mu_h) = \sum_h W_h(\mu_h - \mu)^2$ versus μ. We use a difference estimator for the between-strata variation in strata means,

$$\hat{V}(\mu_h) = k \sum_h W_h \hat{\sigma}_h^2 - k\hat{\mu} - \hat{\mu}^2 + v_{csm}(\hat{\mu}).$$

These results confirm that $V(\mu_h)$ increases with μ in an approximately linear manner. The large discrepancy for 3Ps cod (panel 2) is for 1995. The survey for that year had an anomalously large catch.

3.1 Simulations

We used the estimated strata means to generated simulated data from the NB distribution, with k fixed at the annual estimates. The strata sample sizes were also fixed as the actual values in our survey data-sets. This is basically a parametric bootstrap procedure. The advantage of using the estimated strata means is that we can examine the efficacy of the NB CI method to errors in the assumptions about the between-strata variability in means; that is, the strata means we simulated from have variability the same as in Figure 5. However, our simulations do not address CI coverage errors that may occur if the within-strata distribution of survey catches are not independent and identically distributed (iid) NB rv's.

The simulated exceedance probabilities are compared to their nominal values in Figure 6. The exceedance probability is $1 - \Pr(L \leq \mu \leq U)$, where L and U are the lower and upper CI endpoints. We examined CI's for three type 1 errors, $\alpha = 0.2,\ 0.1$, and 0.05. Overall the NB CI's have better coverage properties than the standard CI's. This is particularly true for the two plaice stocks (panels 3 and 4). The NB intervals also tended to work well for the two cod stocks (panels 1 and 2). In some years the coverage probabilities were considerably different than their nominal values, but to a much lesser extent than the standard CI's. For 2J+3KL cod the problem years were 1983, 1991, and 1994. For 3Ps cod the problem years were 1985, 1988, 1991, 1992, and 1995. These years do not always seem to be associated to anomalous years in Figure 5.

The lower exceedance probabilities are shown in Figure 7. The standard limit tended to be too conservative. This is not surprising from a method that can cover infeasible negative values. The NB limit was reasonably accurate for the two plaice stocks and 2J+3KL cod (panels 1, 3, and 4), with a slight tendency to be conservative. The lower NB CI was also reasonably accurate for 3Ps cod (panel 2), although for this survey the interval was slightly optimistic (i.e. not low enough). The interval was substantially optimistic in 1995, the year with one very large catch. The NB lower limit for $\alpha = 0.05$ was in error approximately the same amount as the standard lower limit for 3Ps wolffish; however, the NB lower limit was optimistic rather than conservative (like the standard lower CI), and this is not desirable. The NB limits were closer to their nominal values, although still optimistic, for $\alpha = 0.1$ and 0.2 compared to the standard limits.

The upper exceedance probabilities are shown in Figure 8. The NB upper CI's performed uniformly better than the standard upper limits. The upper coverage errors were still occasionally in substantial error, especially for the 3LNO American plaice fall survey (panel 3) and 3Ps wolffish in 1999-2004 (panel 5).

A somewhat surprising result was that the NB CI's tended to be shorter than the standard

CI's, although both confidence intervals were based on the area-weighted average catch. This was surprising because the NB CI's were based on both survey design and trawling variability whereas the standard CI's were based only on the survey design. As mentioned above, $Var_D(\hat{\mu})$ underestimates the total variance of $\hat{\mu}$ when there is trawling variability. Hence, one might expect the NB CI's to be wider. In fact, the situation is even worse because we showed in Figure 6 that the coverage probabilities of the standard intervals are too low, and that these intervals need to be wider for their coverage probabilities to be closer to the nominal values. This is consistent with the fact (see Särndal et al., 1992) that $v_{st}(\hat{\mu})$ underestimates $Var(\hat{\mu})$ when there is measurement error. Under-coverage tends to be less of a problem for the NB intervals.

The main reason why the NB CI's are shorter is that $Var(T_p)$ is usually less than $Var(T_s)$. The average simulated variance over all years is shown in Table 2. The bias tends to be smaller as well, and this also improves the lower and upper limit coverage accuracy. In Table 2 we also present the simulated bias and variance of another possible confidence interval statistic,

$$T_d = \frac{\hat{\mu} - \mu}{v_{csm}(\hat{\mu})}. \tag{7}$$

This gives some indication of the effect of using the total variance estimate to construct a confidence interval. We speculate based on these results that confidence intervals based on (7) would be more similar, but slightly shorter, than the standard CI's based on (2).

Table 2. Simulated bias and variance of the T_s, T_p, and T_d approximate pivots. The minimum absolute bias and variance values are in bold.

		Bias			Variance		
Species	Division	T_s	T_p	T_d	T_s	T_p	T_d
Cod	2J+3KL	-0.165	**-0.066**	-0.170	1.278	**1.069**	1.163
Cod	3Ps	-0.512	**-0.101**	-0.466	5.932	**1.262**	3.224
American Plaice	3LNO (Fall)	-0.184	**-0.084**	-0.190	1.391	**1.072**	1.172
American Plaice	3LNO (Spring)	-0.113	**-0.040**	-0.114	1.145	**1.017**	1.059
Striped Wolffish	3Ps	-0.406	**-0.101**	-0.394	2.333	2.455	**1.909**

4 Discussion

Our main conclusion is that standard confidence intervals (CI's) from design-based sampling theory using (2) can have coverage properties that are considerably different from their nominal type 1 error rates, and that the new CI's based on (5) that we developed have better coverage properties. The new intervals are also guaranteed not to cover negative values which is clearly desirable; however, they may cover very large and possibly infinite values which is undesirable. This reflects the substantial uncertainty that can exist in bottom trawl surveys for quantifying how large a stock might actually be. From a conservation point of view it is more important to have reliable and informative lower confidence intervals. Our new CI method provides a substantial improvement in this regard.

Our approach involved analytically evaluating the total variance of $\hat{\mu}$, including variability from the survey-design and variability from the trawl capture process. We speculated that the variation should be approximately Negative Binomial (NB) in form, and our case studies supported this speculation. It was not necessary to estimate the μ components in our approximation of $Var(\hat{\mu})$ to construct a CI for μ. We estimated only the k_p variance parameter which could be estimated with greater precision than $Var(\hat{\mu})$ itself. This is why the T_p statistic had less variability than the T_s statistic, and why the NB CI's were shorter and more accurate than the standard CI's. This improvement is the result of the judicious use of model-based assumptions about the within-strata variability of survey catches, and the between-strata variability in stratum means. In statistical inference it is always desirable to produce shorter intervals with a fixed significance level. In fisheries science it is desirable to produce as precise statements about stock status as possible. Inflated uncertainties because of inefficient analyses should be discouraged.

The NB CI's worked best with the 3LNO American plaice spring survey, for which our assumptions were more appropriate. The method also performed better than the standard CI's in the five case-studies we considered. Further research is required to understand why the NB CI method performed more poorly in some years than others, and for some stocks than others. It may not simply be coincidental that the NB CI's worked worse for the two stocks with the lowest average catch per tow, which were 3Ps cod and wolffish (see Table 2). The simulated bias in the T_p statistic was largest for these stocks. It is possible that this is related to non-normality of the T_p CI statistic for small μ's. For count data we usually anticipate normality when either the sample sizes are large or the means are large. Further improvements may be possible if a better approximation to the distribution of T_p can be formulated, particularly when μ's are small.

If our assumptions are not correct about the within-strata variability of survey catches and the between-strata variability in stratum means then the strata area-weighted average estimator is still unbiased because of the random allocation of sets within strata. This is an important advantage of the stratified random-sampling survey design. Fixed station or other purposive or non-probability sampling designs can not have this advantage without additional assumptions about the population distribution of catches. Such assumptions can be categorized as model-based.

If NB variance for within-strata variation in trawl catches is not appropriate then $v_{csm}(\hat{\mu})$ will provide a biased estimate of $Var(\hat{\mu})$. An alternative estimator that is unbiased for most types of within-strata variability is

$$v_{rb}(\hat{\mu}) = \sum_h W_h^2 \frac{s_h^2}{m_h}.$$

The rb subscript indicates that this estimator is distributionally robust. We could have used this estimator to compute k_p, but the T_p statistic would not be appropriate to use if the NB variation assumption was not appropriate. In this case we could use $v_{rb}(\hat{\mu})$ directly in (7) to replace $v_{csm}(\hat{\mu})$. This will produce more robust but wider confidence intervals.

5 References

Cameron, A. C., and Trivedi, P. K. (1998). *Regression analysis of count data.* Cambridge: Cambridge University Press.

Cochran, W. G. (1977). *Sampling Techniques.* 3rd Ed. New York: John Wiley.

Doubleday, W. G. (ed.) (1981). Manual on groundfish surveys in the Northwest Atlantic. *NAFO Sci. Coun. Stud.*, **2**, 7-55.

Godø, A. R. (1994). Factors affecting the reliability of groundfish abundance estimates from bottom trawl surveys, p. 166-199 *in* Fernö, A., and Olsen, S. (eds.) *Marine fish behavior in capture and abundance estimation.* Oxford: Fishing News Books.

Gunderson, D. R. (1993). *Surveys of fisheries resources.* New York: John Wiley.

Malinen, T., and Peltonen, H. (1996). Optimal sampling and traditional versus model-based data analysis in acoustic fish stock assessment in Lake Vesijärvi. *Fish. Res.* 295-308.

Pelletier, D., and Parma, A. M. (1994). Spatial distribution of Pacific halibut (*Hippoglossus stenolepis*): An application of geostatistics to longline survey data. *Can. J. Fish. Aquat. Sci.,* **51**, 1506-1518.

Särndal, C., Swensson, B, and Wretman, J. (1992). *Model assisted survey sampling.* New York: Springer-Verlag.

Smith, S. J. (1990). Use of statistical models for the estimation of abundance from groundfish survey data. *Can. J. Fish. Aquat. Sci.,* **47**, 894-903.

Sullivan, P. J. (1991). Stock abundance estimation using depth-dependent trends and spatially correlated variation. *Can. J. Fish. Aquat. Sci.,* **48**, 1691-1703.

Warren, W. G. (1997). Changes in the within-survey spatio-temporal structure of the northern cod (*Gadus morhua*) population, 1985-1992. *Can. J. Fish. Aquat. Sci.,* **54**, 139-148.

6 Figures

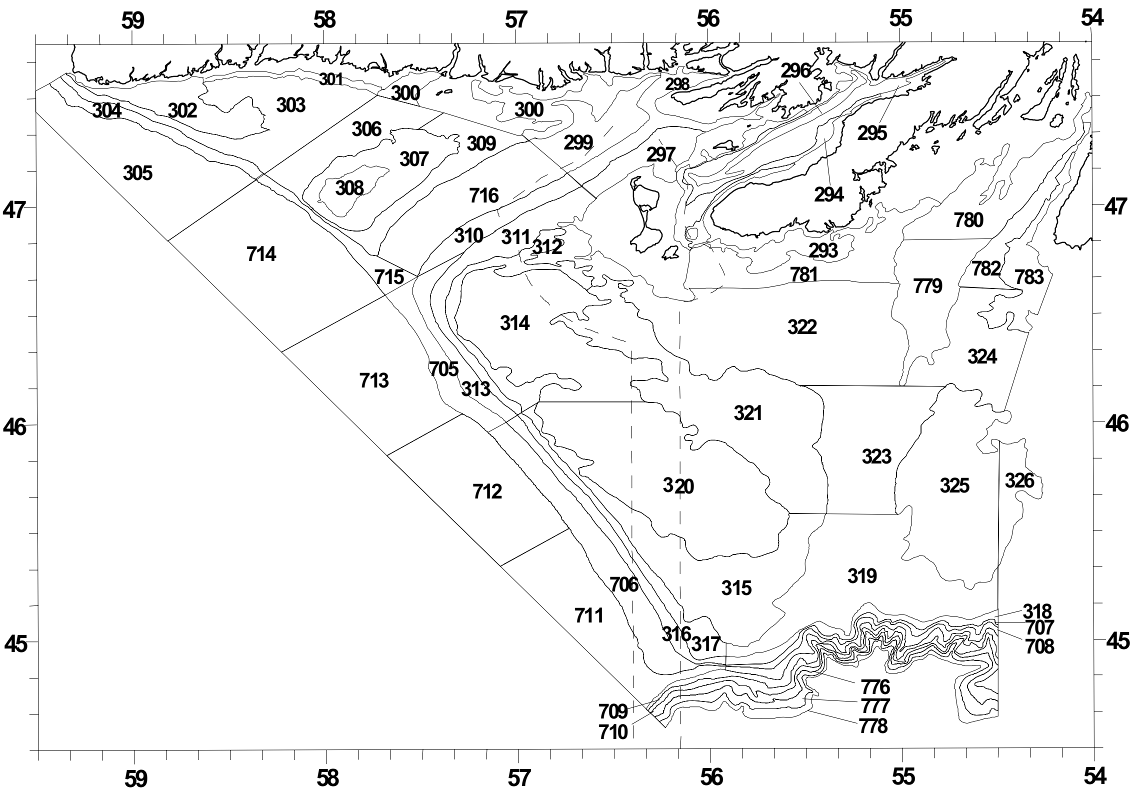


Figure 1: Stratum boundaries in NAFO Subdivision 3Ps. The dashed lines delineate the French economic zone.

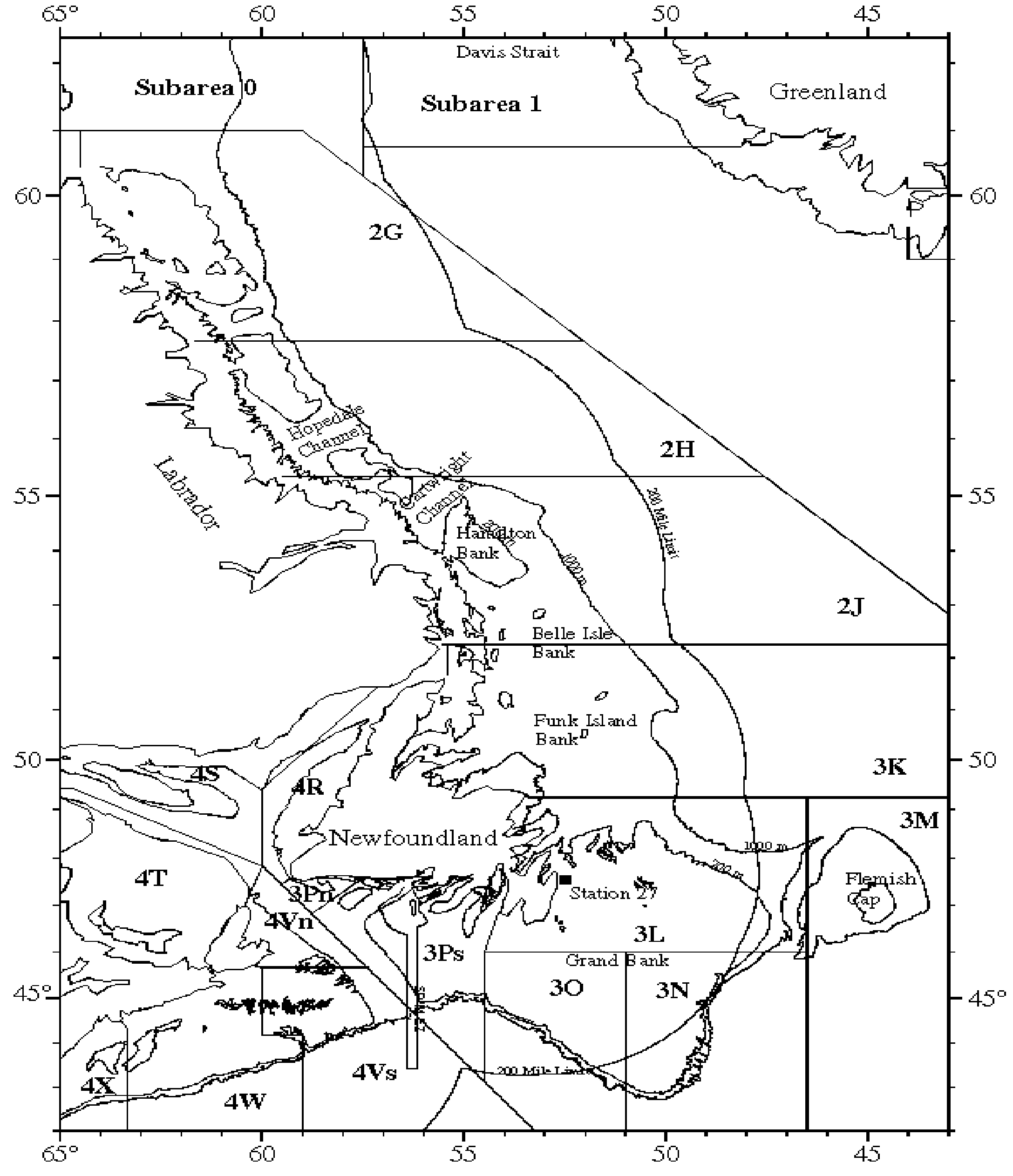


Figure 2: Northwest Atlantic Fisheries Organization (NAFO) divisions.

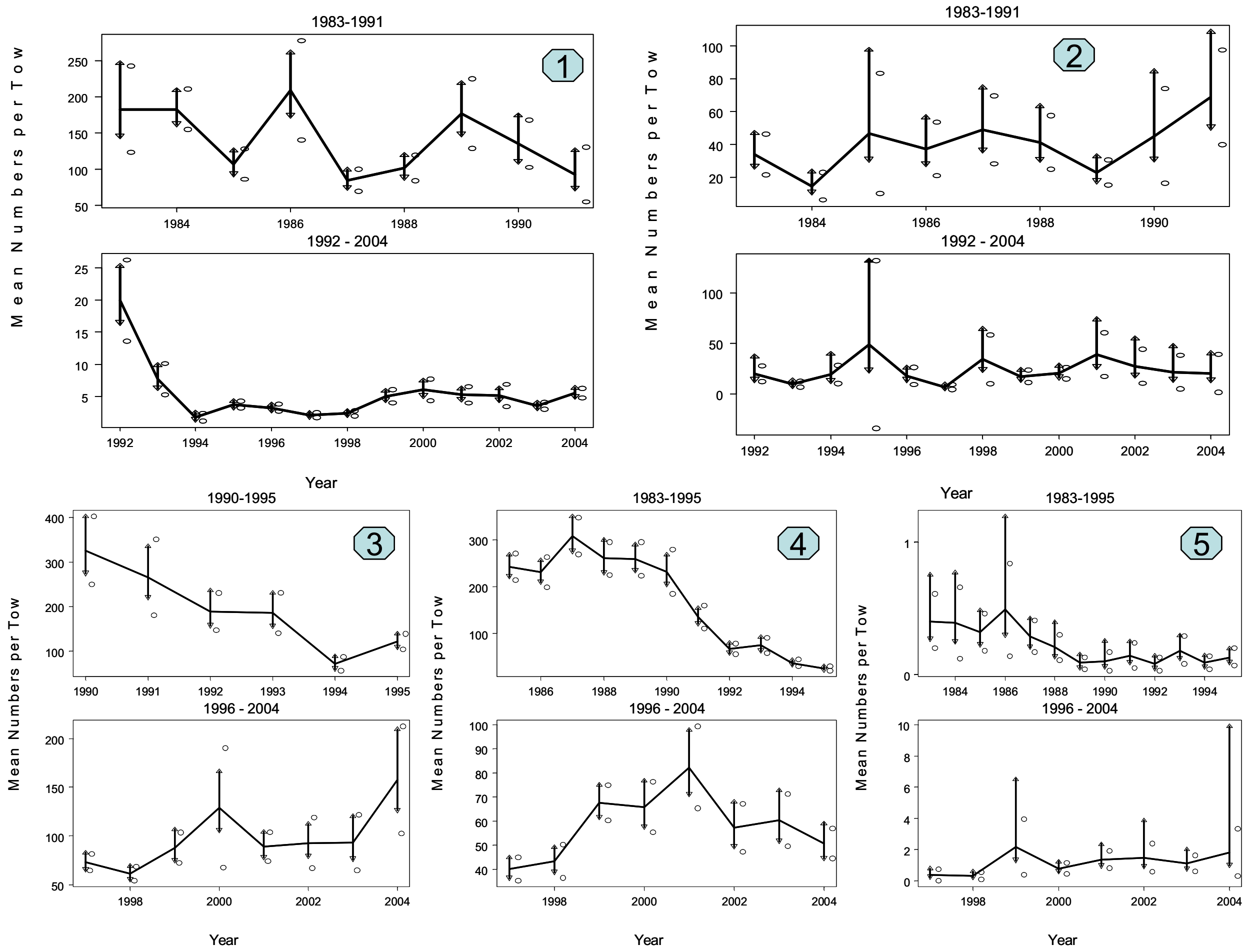


Figure 3: Negative Binomial (segments) and standard (circles) confidence intervals for population average trawlable abundance. Stocks: 1 - 2J+3KL cod; 2 - 3Ps cod; 3 - 3LNO Fall American plaice; 4 - 3LNO Spring American plaice; 5 - 3Ps striped wolffish. A solid arrow (1995 3Ps cod) indicates an infinite limit.

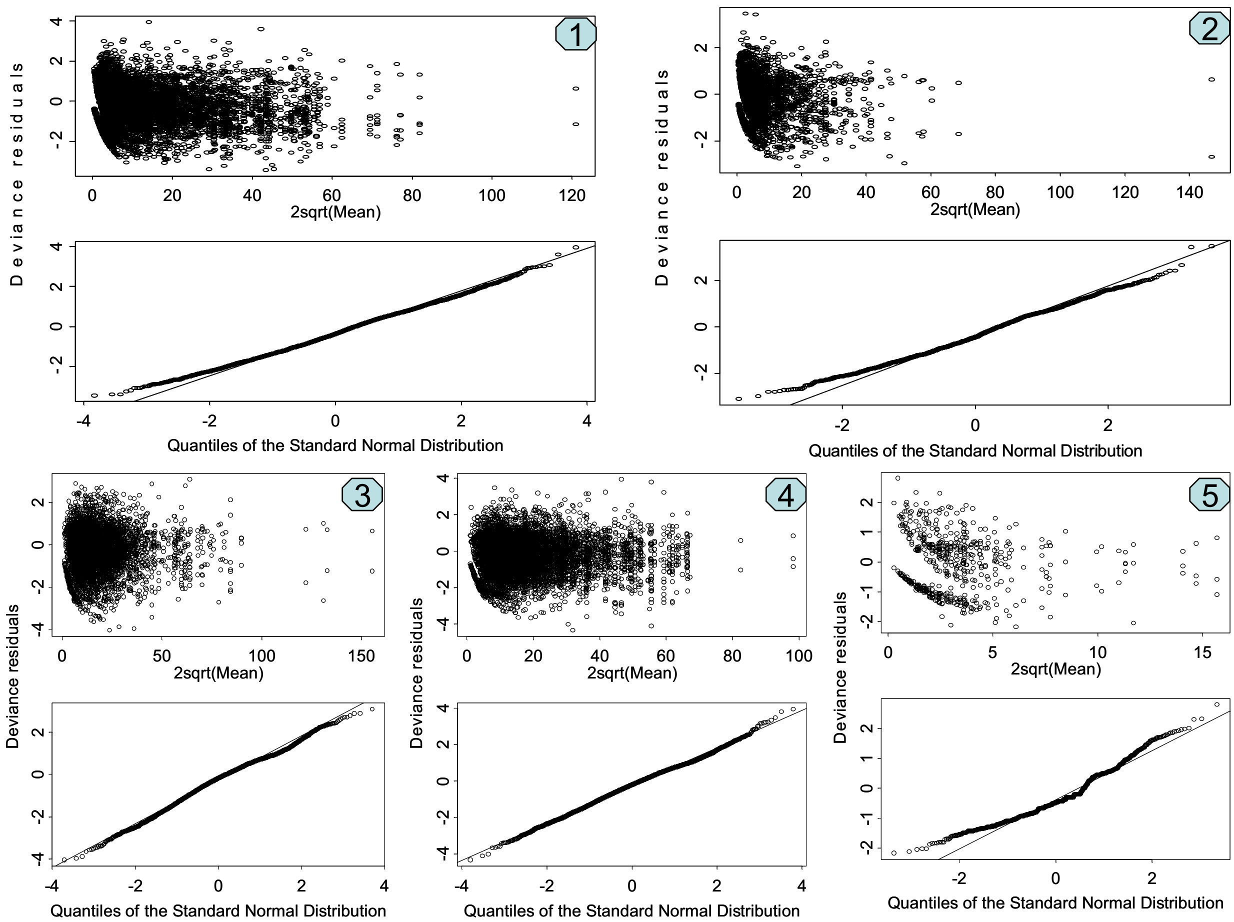


Figure 4: Residual plots. Stocks: 1 - 2J+3KL cod; 2 - 3Ps cod; 3 - 3LNO Fall American plaice; 4 - 3LNO Spring American plaice; 5 - 3Ps striped wolffish.

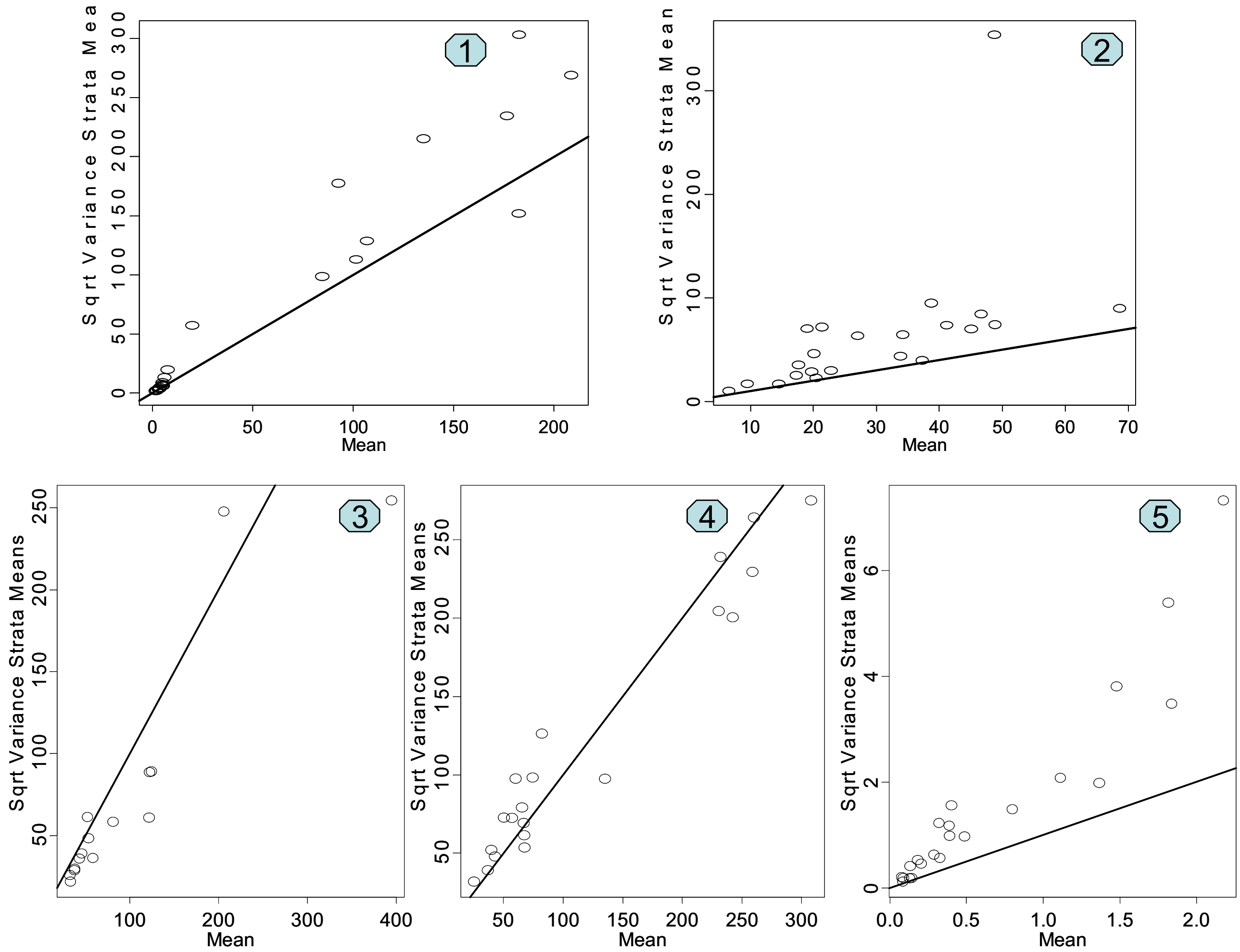


Figure 5: Estimates of the square root of the between strata variation in trawlable abundance, versus $\hat{\mu}$. The solid line has a slope of one. Stocks: 1 - 2J+3KL cod; 2 - 3Ps cod; 3 - 3LNO Fall American plaice; 4 - 3LNO Spring American plaice; 5 - 3Ps striped wolffish.

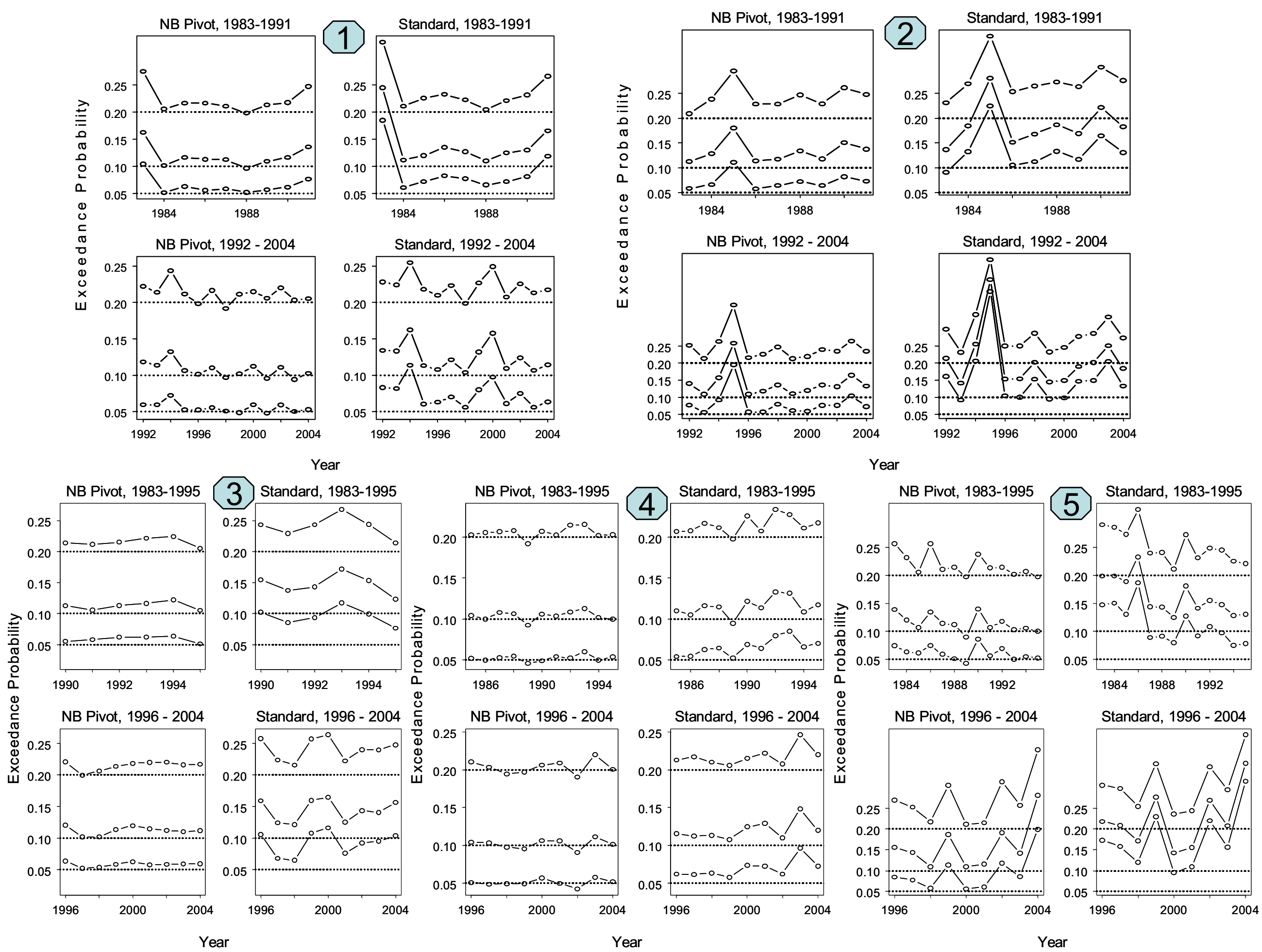


Figure 6: Simulated (solid lines) and nominal (dotted lines) exceedance probabilities. Stocks: 1 - 2J+3KL cod; 2 - 3Ps cod; 3 - 3LNO Fall American plaice; 4 - 3LNO Spring American plaice; 5 - 3Ps striped wolffish.

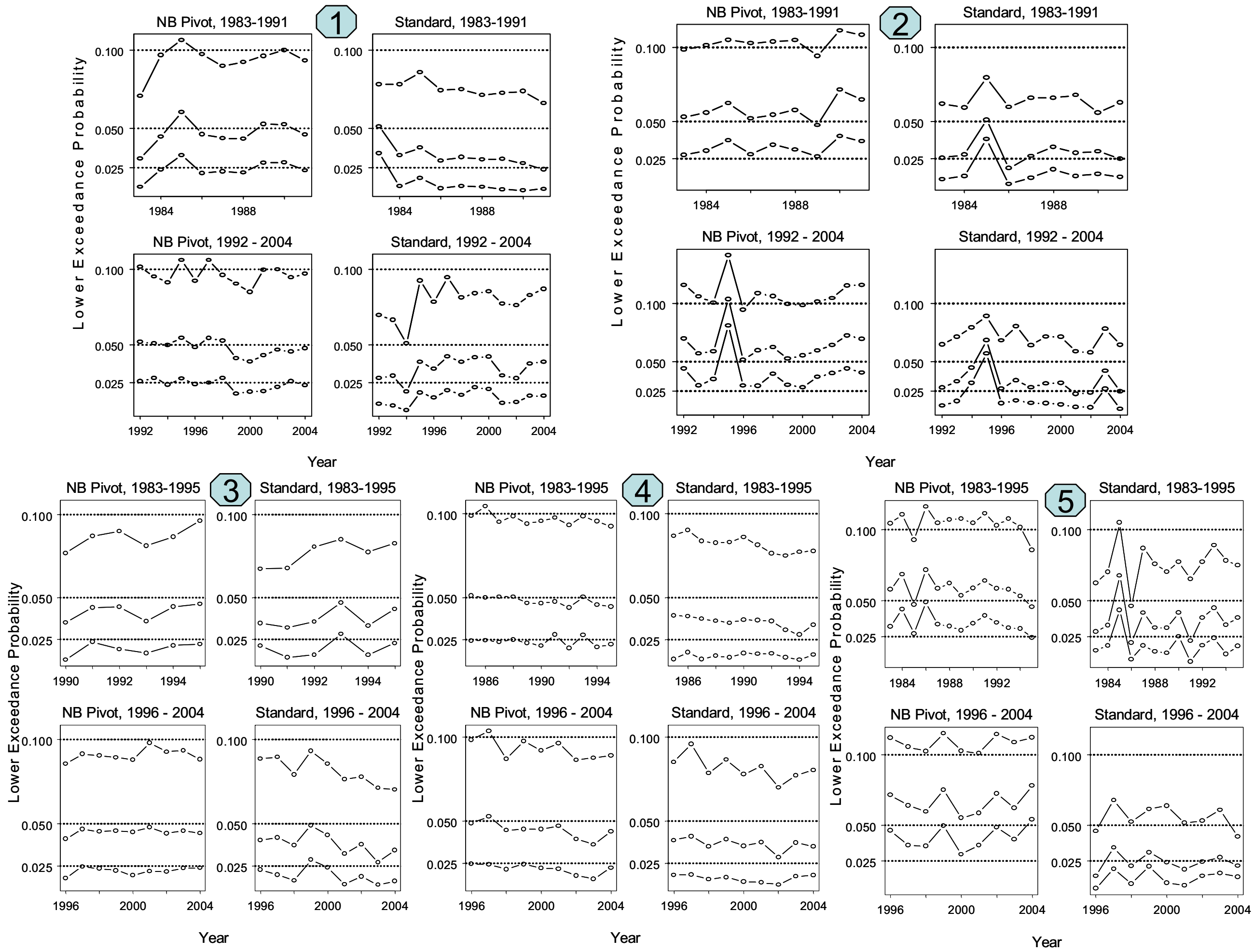


Figure 7: Simulated (solid lines) and nominal (dotted lines) lower exceedance probabilities. Stocks: 1 - 2J+3KL cod; 2 - 3Ps cod; 3 - 3LNO Fall American plaice; 4 - 3LNO Spring American plaice; 5 - 3Ps striped wolffish.

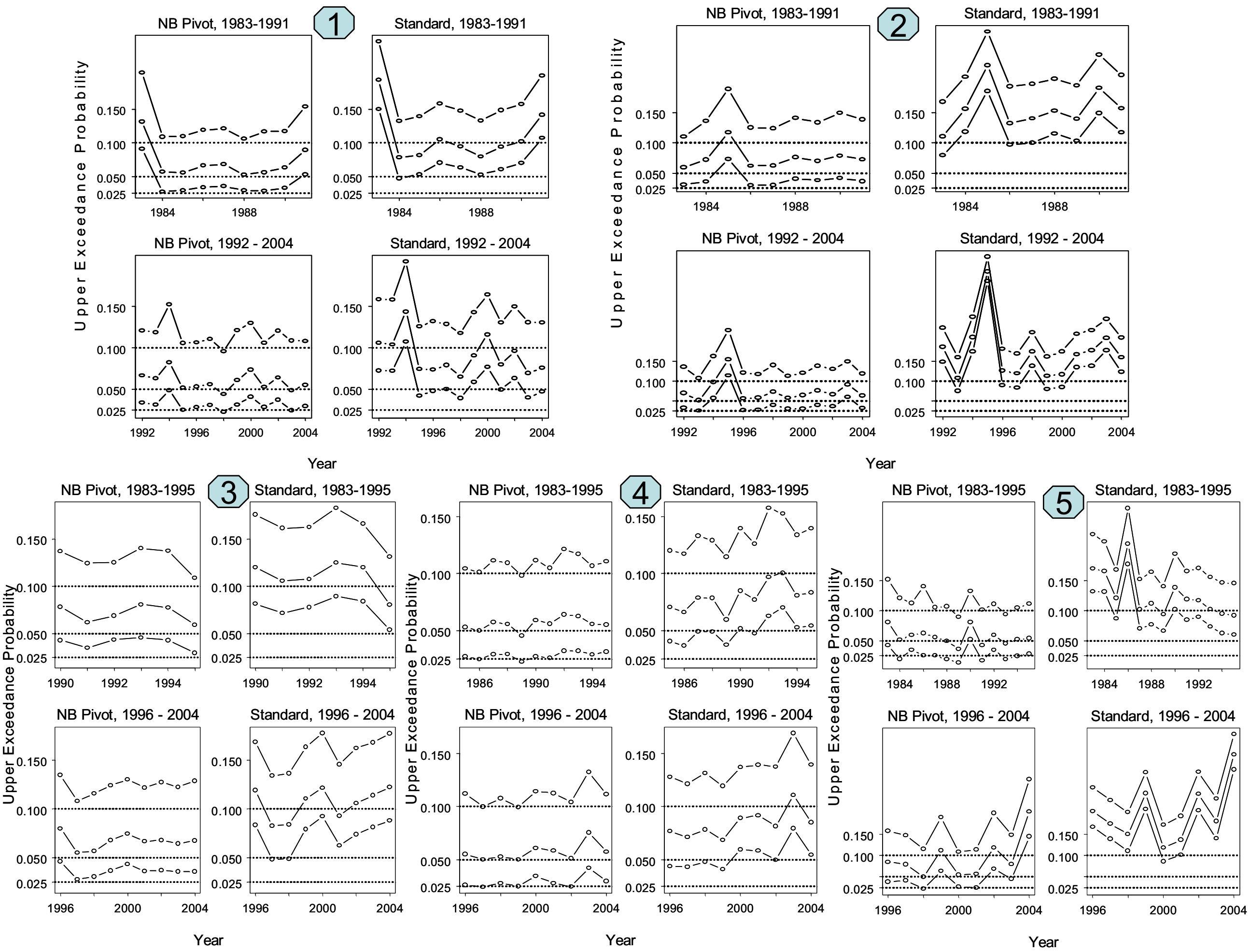


Figure 8: Simulated (solid lines) and nominal (dotted lines) upper exceedance probabilities. Stocks: 1 - 2J+3KL cod; 2 - 3Ps cod; 3 - 3LNO Fall American plaice; 4 - 3LNO Spring American plaice; 5 - 3Ps striped wolffish.

Annex 8: Working Document 5

Lake Ontario alewife abundance *by* Jean V. Adams and Robert O'Gorman

Working Document prepared for the ICES Workshop on Survey Design and Analysis II
9-13 May 2005, Sète, France

Lake Ontario Alewife Abundance, 2003

Jean V. Adams[1] and Robert O'Gorman[2]
U.S. Geological Survey Great Lakes Science Center

[1] Marquette Biological Station, 1924 Industrial Parkway, Marquette, MI 49855, USA, jvadams@usgs.gov
[2] Lake Ontario Biological Station, 17 Lake Street, Oswego, NY 13126, USA, rogorman@usgs.gov

1. Objectives of the survey

The U.S. Geological Survey's Great Lakes Science Center (GLSC) conducts annual surveys of alewife *Alosa pseudoharengus* with bottom trawls in U.S. waters of Lake Ontario in cooperation with the New York State Department of Environmental Conservation. Alewife biomass estimates are critical in setting stocking levels of trout and salmon that maintain quality fisheries and minimize the risk of creating a predator-prey imbalance (Jones et al. 1993). Alewives are considered an important species in the prey-fish community as laid out in the Lake Ontario fish community objectives (Stewart et al. 1999). For the purposes of comparing analyses of survey data, we focused on the relative biomass of adult alewives in 2003.

2. Population to be sampled

The target population was defined to be adult alewife (age two and older) in U.S. waters of Lake Ontario. In practice, the sampled population was defined somewhat differently. The adult life stage was defined using a length cutoff, which was verified annually by aging a subsample of juveniles. All alewives greater than 109 mm total length were considered adults. Although the lake's maximum depth is 244 m, the sample space was limited to the depth range (0 to 160 m) where bottom trawl catches of the target species have been highest historically. The sample space was also limited to waters over trawlable bottom substrate. Sampling was conducted in late April through early May, coinciding with maximum availability of alewife to the bottom trawl (based on 1972 catch records).

3. Data to be collected

All alewives in small catches were grouped according to life stage (yearlings and adults) and each life stage was weighed in bulk. When the total catch was large (greater than about 20 kg), we sorted, counted, and weighed a random subsample of 10 to 12 kg; the rest of the catch was weighted in aggregate, and composition was estimated by direct proportion. Trawl duration (tow time) was measured. Location of trawl tow was recorded (latitude and longitude).

4. Degree of precision required

No specific requirements for precision were targeted. The precision of the survey was limited by time and cost constraints (day time sampling during a three-week cruise).

5. Methods of measurement

The survey was conducted by the GLSC's R/V Kaho, equipped with a 20.4-m headrope, 3-in-1 bottom trawl. At each of 12 ports, sampling was conducted at up to 13 fixed sites. Tow duration was targeted at 10 minutes.

6. The frame

The sampling frame was limited to a relatively short list of trawlable locations (98 fixed stations) in the U.S. waters of Lake Ontario, ranging in depth from 8 to 150 m.

7. Selection of the sample

The entire collection of trawlable locations (98 fixed stations) were selected for sampling.

8. The pretest

Because this survey has been conducted annually since 1978, no additional pretest sampling was required.

When the 3-in-1 trawl was introduced in 1997, 100 paired trawl tows were conducted for calibration with the "standard" bottom trawl, the 12-m headrope, Yankee trawl. The change in nets was necessary to avoid gear saturation with exotic mussels (*Dreissena* spp.), which invaded Lake Ontario in the early 1990s. The 3-in-1 trawl does not fish as close to the bottom as the Yankee trawl, catching fewer mussels as well as fewer small demersal fish.

9. Organization of the field work

Individuals participating in the survey were trained in the methods used in the survey, including fishing the trawl, processing the samples, and recording the data.

10. Summary and analysis of the data

Weight of adult alewife caught (biomass) was corrected for the greater fishing power of the 3- in-1 trawl, relative to the historic standard Yankee trawl, and for tow duration (standardized to a 10-minute trawl tow). A map of the observed biomass is shown in Figure 1.

10.1 Design-based approach

Biomass estimates were calculated based on the assumption that the fixed survey was, in fact, a stratified random survey, with 20-m depth zones from 0 to 160 m as strata, and the fixed sampling stations as random samples. Relative mean biomass and its variance were estimated using standard methods (Table 1, Cochran 1977). We also calculated bias-corrected confidence intervals (Table 1). The sample means for each stratum are shown in Figure 2. Predictions were made across a grid (at one minute intervals of latitude and longitude) within the sample space (Figure 3).

Efficiency of the design was characterized by estimating the design effect and the effective sample size through simulation (Table 1). Because the original samples taken were not proportionally allocated to the depth strata, resampling was conducted using probabilities proportional to the relative weighting of the depth zones (according to surface area of the lake). Sampling with replacement was repeated 1,000 times for sample sizes ranging from 50 to 200. The mean biomass was calculated for each sample, and the variance of the 1,000 means was calculated for each sample size (Figure 4). The design effect of the design-based approach was calculated as the ratio of the observed variance to the simulated variance from simple random sampling (estimated from the graph, Figure 4) using the same sample size (n = 98),

$$deff_{design} = \frac{Var_{obs}}{Var_{srs}} = \frac{6.6^2}{8.7^2} = 0.58\,.$$

The effective samples size, n_c, was estimated from the graph (Figure 4) to be 164.

10.2 Model-based approach

Biomass estimates were calculated based on the assumption that the relation between alewife biomass and fishing depth could be described by a smooth line. The relation was fit to the data using an additive model,

$$biomass = \begin{cases} s(depth) & s(depth) \geq 0 \\ 0 & s(depth) < 0 \end{cases},$$

where *biomass* is alewife biomass (in kg per 10-minute tow), *depth* is fishing depth (in m), and *s*() is a nonparametric smoothing spline function with five degrees of freedom (Figure 2). Predictions were made across a grid (at one minute intervals of latitude and longitude) within the sample space (Figure 3). Mean biomass was calculated as the mean of these predictions, and precision was estimated using bootstrap resampling (Table 1).

Efficiency of the design was characterized by estimating the design effect and the effective sample size through simulation using the same methods as described for the design-based model (Table 1). The design effect of the model-based approach was calculated as the ratio of the observed variance to the simulated variance from simple random sampling (estimated from the graph, Figure 4) using the same sample size,

$$deff_{model} = \frac{Var_{obs}}{Var_{srs}} = \frac{6.4^2}{7.3^2} = 0.76\,.$$

The effective samples size, n_c, was estimated from the graph (Figure 4) to be 130.

11. Information gained for future surveys

Information from the 2003 survey was used to investigate the effects of optimal allocation of sampling effort. Because the time to take a single bottom trawl sample increases with bottom depth, optimal allocation has to take cost of sampling into account. For the R/V Kaho on Lake Ontario, the time it takes to conduct a 10-minute trawl tow is a linear function of depth,

$$T = 15 + 0.2D,$$

where T is time in minutes and D is bottom depth in meters (Robert O'Gorman, GLSC, unpublished data). This relation can be used to estimate the time to take a single sample in each 20-m depth zone,

$$c_h = (15 + 0.2m)/60,$$

where c_h is the cost of taking a sample in stratum h (in hours) and m_h is the midpoint of the depth stratum (in m).

Other factors may affect sampling time, but we ignore them here, for simplicity. These include increased cost in travel time to sample deeper stations, and increased cost in catch processing time to sample stations with higher catches.

In 2003, the total on-site sampling time for 98 stations was 50 hours (this does not include travel time). Using this as our fixed on-site sampling cost, we calculated the optimal allocation as

$$n = \frac{50\sum(N_h s_h / \sqrt{c_h})}{\sum(N_h s_h \sqrt{c_h})} \text{ and}$$

$$n_h = \frac{n N_h s_h / \sqrt{c_h}}{\sum(N_h s_h / \sqrt{c_h})},$$

where n is the total number of samples taken, N_h is the surface area of the lake in stratum h (in ha), s_h is the sample standard deviation among samples in stratum estimated from 2003 data, and n_h is the number of samples taken in stratum h (Cochran 1977). The optimal sample size was n = 84, and the corresponding allocation is shown in Figure 5.

We estimated the effect of optimal allocation on the 2003 abundance estimate using resampling. Because the optimal allocation suggested only one sample taken in waters less than 60 m, we shifted five samples from the 80-120 m depth zones to the 0-60 m depth zones to ensure at least two samples per stratum. This allocation reduced the error in the estimated mean abundance by 38% (from SE = 6.5 and RSE = 24% to SE = 4.0 and RSE = 15%) for the same cost using the design-based estimator.

However, because the depth distribution of alewives in Lake Ontario changes every year (mean depth of capture ranged from 35 to 85 m during 1978-1998, O'Gorman et al. 2000), a single fixed allocation of sampling effort will be not be optimal every year. Thus, it may be beneficial to incorporate some adaptive sampling in the survey design, taking more samples in those depth zones yielding large catches of alewives, and/or taking fewer samples in those depth zones yielding smaller catches. Use of an adaptive design would necessitate the use of a model-based estimate.

12. Assumptions

Several assumptions were made in design and analysis of the Lake Ontario alewife survey. Alewives greater than 109 mm were assumed to be adults. This was verified by aging, and adjustments were made when necessary. Waters deeper than 160 m were assumed to have no alewives. This can be verified by sampling deeper waters. The relative abundance of alewives was assumed unaffected by bottom substrate (trawlable vs. nontrawlable). This would require verification using gear other than a bottom trawl, e.g., hydroacoustics or gill nets. Fixed stations were assumed to be randomly selected. The effects of this assumption could also be investigated using alternative survey gear. The proportion of the alewife population, near bottom and available to the sampling gear was assumed to vary little between years. Acoustics could be used to examine this assumption.

13. References

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

Jones, M.L., J.F. Koonce, and R. O'Gorman. 1993. Sustainability of hatchery-dependent salmonine fisheries in Lake Ontario: conflicts between predator demand and prey supply. Transactions of American Fisheries Society 122:1002-1018.

O'Gorman, R., J.H. Elrod, R.W. Owens, C.P. Schneider, T.H. Eckert, and B.F. Lantry. 2000. Shifts in depth distributions of alewives, rainbow smelt, and age-2 lake trout in southern Lake Ontario following establishment of dreissenids. Transactions of the American Fisheries Society 129:1096–1106.

Stewart, T.J., R.E. Lange, S.D. Orsatti, C.P. Schneider, A. Mathers, M.E. Daniels. 1999. Fish-community objectives for Lake Ontario. Great Lakes Fish. Comm. Spec. Pub. 99-l. 56 p. *http://www.glfc.org/pubs/SpecialPubs/Sp99_1.pdf*

Table 1. Estimates of mean adult alewife biomass (kg/10-min) using different estimation methods. Precision is reported as standard error (SE), relative standard error (RSE), and 95% confidence intervals. Bias-corrected bootstrap estimates of precision, design effect, *deff*, and effective sample size, n_c, are also reported.

		Precision			Bootstrap			Efficiency	
Estimation	Mean	SE	RSE	95% CI	SE	RSE	95% CI	*deff*	n_c
Design-based	27	6.6	25%	(14, 40)	6.5	24%	(17, 45)	0.58	164
Model-based	27	-	-	- -	6.4	24%	(17, 43)	0.76	130

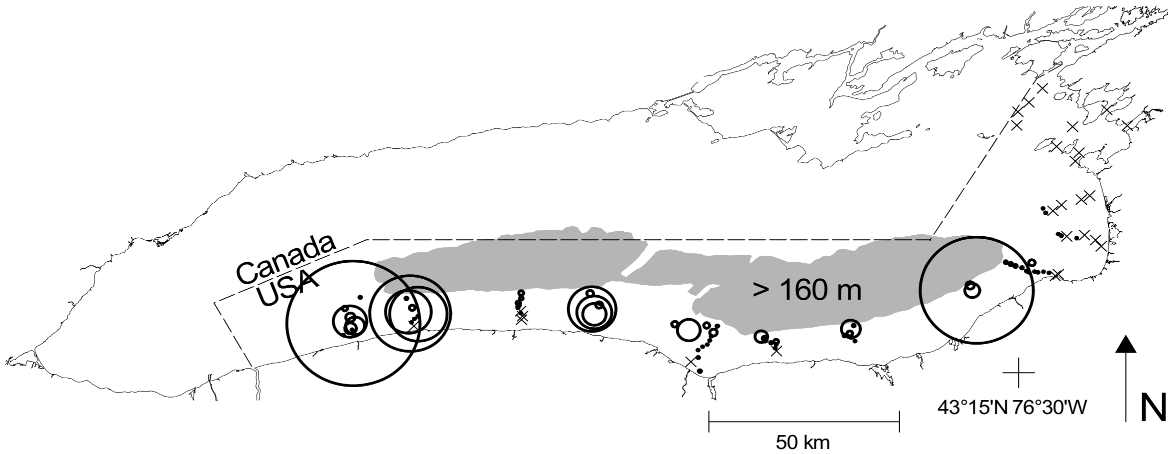


Figure 1. Adult alewife biomass in U.S. waters of Lake Ontario, 2003. Circle size represents relative size of trawl catch, ranging from 0.01 to 650 kg per 10-minute tow, Xs represent trawl samples with zero catch. Sample space includes area from shore to international boundary (dashed line) excluding waters greater than 160 m (shaded area).

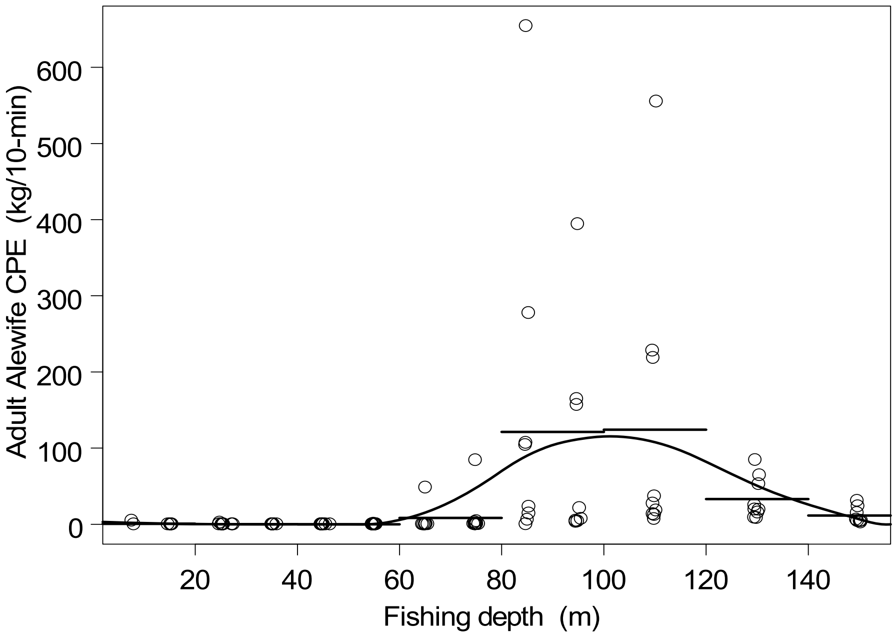


Figure 2. Relation between adult alewife biomass and bottom trawl fishing depth in U.S. waters of Lake Ontario, 2003. Points represent observed values, lines indicate predicted values (line segments for the design-based estimator and curve for the model-based estimator).

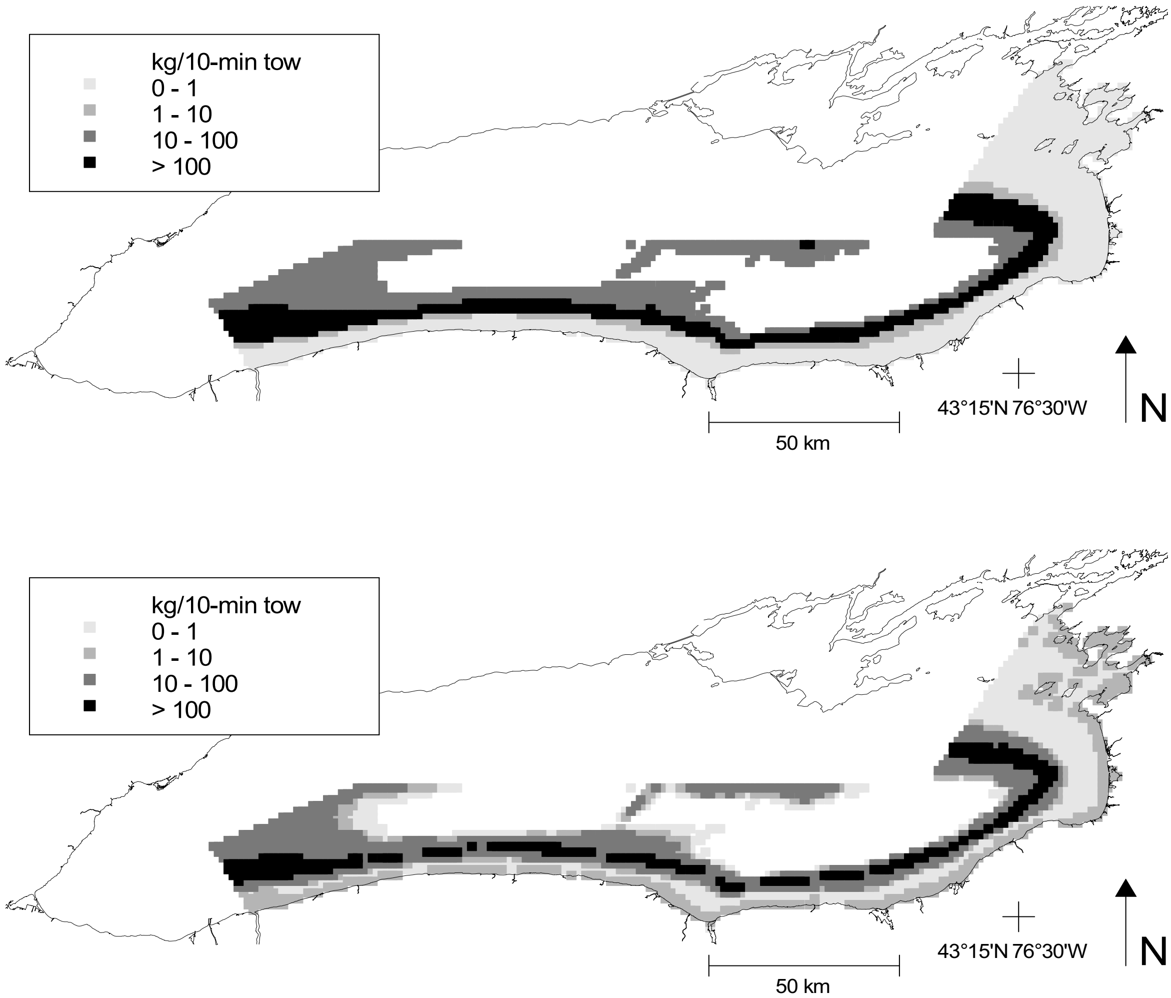


Figure 3. Map of adult alewife biomass in U.S. waters of Lake Ontario, 2003 based on (a) design-based and (b) model-based estimators.

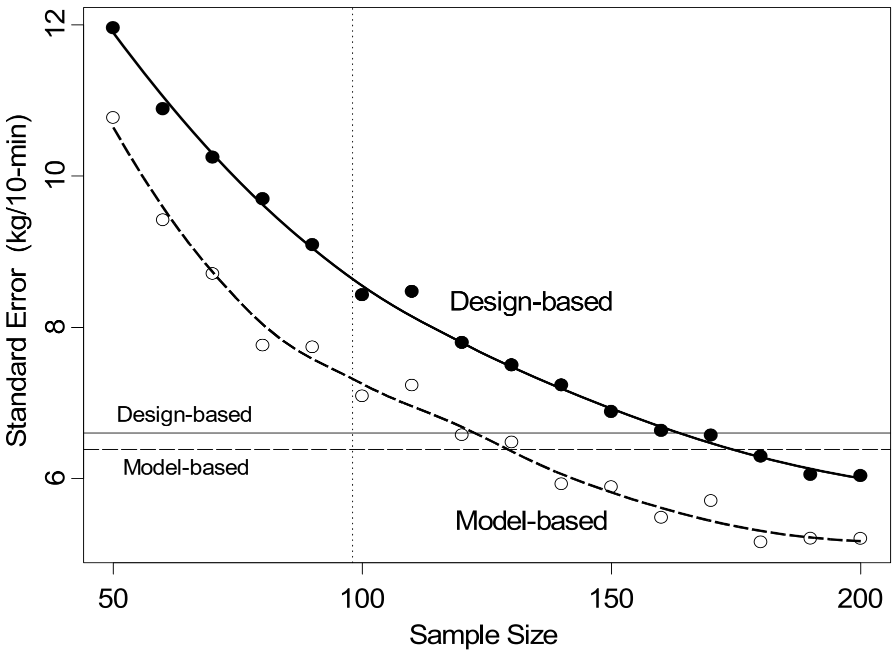


Figure 4. Relation between sample size and standard error for simulated, simple random sampling using design-based (filled circles) and model-based (open circles) estimators of adult alewife biomass in 2003 (with loess smoothed lines). The horizontal lines indicate the standard errors from stratified random sampling, and the dotted vertical line indicates the actual sample size (n = 98).

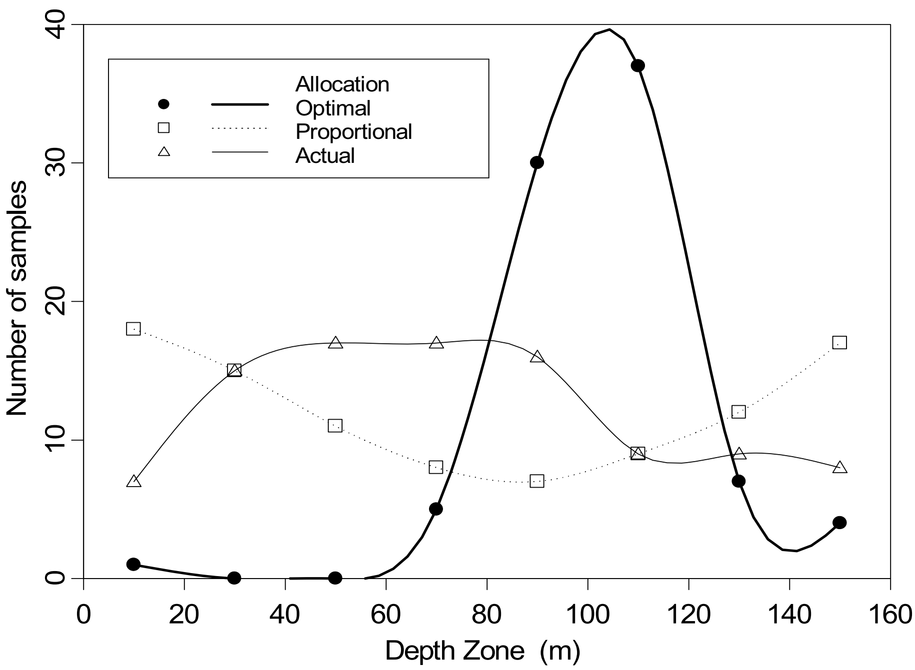


Figure 5. Alternative allocations of samples to 20-m depth strata, for a fixed cost (50 hours of on-site sampling).

Annex 9: Working Document 6

Effect of tow duration on catch rates and mean length of Northern shrimp (*Pandalus borealis*) and Greenland halibut (*Reinhardtius hippoglossoides*)in the West Greenland Bottom Trawl Survey, 1999-2004 *by* Kai Wieland and Marie Storr-Paulsen

Effect of tow duration on catch rates and mean length of Northern shrimp (*Pandalus borealis*) and Greenland halibut (*Reinhardtius hippoglossoides*) in the West Greenland Bottom Trawl Survey, 1999-2004

Kai Wieland and Marie Storr-Paulsen

Greenland Institute of Natural Resources
PO Box 570, 3900 Nuuk, Greenland

Abstract

Standard towing time in the annual West Greenland Bottom Trawl Survey for shrimp and fish has initially been 60 min. Shorter tow durations have been gradually introduced over time and a mixture of 30 and 15 min tows have been used in the recent years. Catches of northern shrimp and Greenland halibut from 15 and 30 min tows have been analysed to examine whether a reduction of tow duration to 15 min would influence the catch per swept area, its precision and the size distribution of the two species. For both species, neither total biomass density nor numerical densities of different size groups differed significantly between 15 and 30 min tows. No indication was found that 15 min tows give less precise results than 30 min tows. Tow duration had also no significant effect on mean size and maximum length of both species. The results suggest that the mixture of 15 and 30 min tows can be replaced by 15 min tows on all stations without impact on the continuity of the time series of survey estimates and the gain in survey time should be used to increase the total number of stations.

Introduction

Marine resources are often distributed with strong short-range spatial autocorrelation, which implies that a small sample at a station provides almost as much information as a large one (Pennington and Vølstad 1991, Gunderson 1993). Consequently, Pennington and Vølstad (1991, 1994) suggested that reducing tow duration and increasing appropriately the number of stations would result in an increase in the precision of the survey estimates of e.g. abundance and biomass but also for other properties such as mean length. However, decreasing tow duration may alter the species and length selectivity of a trawl, and effects of tow duration on catch rates have been reported for several cases (Carothers and Chittenden 1985, Godø et al. 1990, Walsh, 1991, Somerton et al. 2002).

The West Greenland Bottom Trawl Survey for shrimp and fish has been conducted annually by the Greenland Institute of Natural Resources since 1988. The initial standard tow duration was 60 min. The results from Pennington and Vølstad (1991, 1994) indicated that the use of shorter tows would improve the efficiency of the survey. This information, however, was based on data for other marine species than northern shrimp, which is the main target species of the West Greenland Bottom Survey. Furthermore, there were concerns that the continuity of the survey time series would be severely impacted in particular when making drastic changes in tow duration too quickly, i.e. from one year to the next (Kingsley et al. 2002). Hence, tow duration were reduced stepwise over the years introducing 30 min and 15 min tows in 1991 and 1999, respectively, replacing 60 min tows with 45 min tows in 2000 and using solely 30 and 15 min tows in 2001 (Kanneworff and Wieland 2001). Kingsley et al. (2002) studied the effect of the mixture of tow durations on the precision of the survey estimates of northern shrimp biomass density for the years 1999 and 2000. Their results in indicate, that using shorter tows has little effect on the precision of the estimate of total survey biomass. However, an underestimation of the nominal swept area due imprecise measure of tow duration and catching before and after the time for which tow duration is recorded would effect short tows relatively more than long tows (Godø et al. 1990). Kingsley et al. (2002) estimated this 'end-error' to be equivalent to a towing time of 2.78 min and concluded that biomass estimates from survey based on short tows would be biased upwards.

Abundance of northern shrimp and Greenland halibut in particular has increased considerably off West Greenland in the most recent years (Wieland et al. 2004, Storr-Paulsen and Jørgensen 2004). As a consequence, 30 min tows often result in large catches, which are difficult to handle and for which time consuming subsampling procedures have to be applied. It would be desirable to shorten tow duration to 15 min on all survey stations in order to reduce the need for subsampling as well as to gain the opportunity for an increase of the total number of stations from which an improvement of the precision of the survey estimates is expected.

The present study provides new information whether a reduction in tow duration from 30 to 15 min is advisable and can be done without risking an interruption of the continuity of the survey time series for northern shrimp and Greenland halibut. No special experiment has been carried out for this purpose, but the use of the mixture of 30 and 15 min tows randomly allocated to the sampling locations within the strata in the West Greenland Bottom Survey since 1999 allowed an analysis of the effect of tow duration on catch rates and length composition for the two target species.

Material and Methods

Fishing gear and survey stratification

The fishing gear used in the West Greenland Bottom Trawl Survey for shrimp and fish is a 3000/20-mesh *Skjervøy* bottom trawl equipped with a heavy bobbin ground rope. The bobbins have a diameter of 21" (53 cm) and there are no rubber discs between them. The mesh size in the cod-end was reduced from 44 mm to 20 mm (stretched) in 1993, and the fine mesh cod-end has been used in all years thereafter. Average towing speed has been about 2.5 knots in all years. From 1988 to 2003 the trawl doors were of the type *Greenland Perfect*, measuring 9.25 m² and weighing 2420 kg. They were replaced in 2004 by *Injector International* 7.5 m² trawl doors with a weight of 2800 kg. The change of the trawl doors, however, had no significant effect on the vertical opening of the trawl and the door spread (Wieland et al. 2004). Nominal swept area is calculated from the straight-line track length between start and end positions multiplied by the mean wingspread for each tow, where mean wingspread is derived from measurements of the door spread in intervals of about 5 min and the trawl geometry. Headrope height is monitored using a *Furuno* sensor, and the length of the tow path is measured from the moment the headline distance from the bottom is stable until the moment the haulback begins.

The survey covers West Greenland waters at depths from < 150 m to 600 m and extends from 59°30'N, the southern tip of Greenland, to 72°30'N. It has been conducted annually since 1988 and follows a stratified random design. The survey strata correspond to geographical areas and four main depth zones: 150-200 m, 200-300 m, 300-400 m and 400-600 m (Wieland and Kanneworff 2004). Stations conducted at depths < 150 m were not considered in the present analysis because neither northern shrimp nor Greenland halibut are usually found at such shallow depths.

From 1988 to 1999, the number of stations was allocated to strata in proportion to stratum area. Since 2000 allocation has been weighted towards strata with historically high densities of northern shrimp and where high variances are observed. An exponential smoothing technique for the weighting was applied to give higher influence of the more recent observations (Kingsley et al. 1999). This was done in order to improve the precision of the biomass estimates for northern shrimp, as this has been the main objective of the survey.

Station positions have been selected randomly and in 1999 a method was introduced which combines the use of a minimum between-station-distance rule (a buffer zone) with a random allocation scheme (Kingsley et al. 2004).

Tow duration has been 60 min in the years 1988 to 1990 and a mixture of 60 min and 30 min in the years 1991 to 1998. 15 min tows were introduced the first time in 1999 (≈ 20 % of the stations). In the years 2001 to 2003 tow durations of 30 and 15 min tows in the proportion 2:1 were used. Thereafter, an equal proportion of 30 and 15 min tows, which were randomly distributed over all strata with depths ≥ 150 m, has been used.

Data selection

Tows conducted during the routine part of the survey in the years 1999 to 2004 were grouped into two intervals of 15 and 30 min with a tolerance of 10 % of the reported towing time, and tows of other duration were discarded. Strata for which at least two hauls in each group of tow duration have been available in a given year were selected or further analysis. This resulted in an initial data set of 185 15 min and 217 30 min tows from 18 strata and 6 years (Tab. 1). At these sampling locations, which were distributed over a large part of the survey area (Fig. 1), only few zero catches of northern shrimp occurred and 43 pairs of stratum and year combinations were used for analysis. For Greenland halibut, however, a considerable number of zero catches were recorded in the southern part of the study and limiting the analysis cases for which at least two non-zero catches were available for each tow duration reduced the data set to 160 15 min and 197 30 min tows and 37 pairs of stratum and year combinations.

Effect of tow duration on catch rate

Catch data by haul were analysed using a Generalized Linear Model (GLM) approach (McCullagh and Nelder 1989):

$$\log_e(CPUE_{ijk} + 0.001) = Stratum_j + Year_k + Depth_i + TowDuration_i + \varepsilon_{ijk}$$

and

$$\log_e(CPUE_{ijk} + 0.001) = \log_e(MeanCPUE_{jk}) + TowDuration_i + \varepsilon_{ijk}$$

where $CPUE$ is the catch of each tow divided by its swept area and ε is the error term. The errors were assumed to be independent and identical distributed, i.e. with the mean $\mu = 0$ the variance σ^2. Stratum and year were considered as factors. Depth was included as a continuous variable to account for additional within-stratum variability of depth for the different tow locations.

Biomass densities (in kg/m^2), which include all length intervals, were used for both species as CPUE. Numerical densities (in numbers/m^2) were used for three sexual stages of northern shrimp. Northern shrimp is a protandric hermaphrodite, in which males, primiparous and multiparous females correspond to groups of increasing mean carapace length (Wieland et al. 2004). For Greenland halibut, numerical densities (in numbers/km^2) were categorized into three groups: < 20 cm, 20 – 40 cm and > 40 cm.

Effect of tow duration on precision

Within stratum and year mean biomass densities and within stratum and year standard deviations were computed and a pair wise comparison of the standard deviation over mean ratios for the two tow durations (paired t-test) was used to evaluate the effect of tow duration on the precision of the density estimates.

Effect of tow duration on mean length

Fish and shrimp are usually caught in clusters resulting in a non-independence of length measurements within a haul (Pennington and Vølstad 1994). To account for this, the effect of tow duration on mean length was studied following the approach by Godø et al. (1990). Population mean lengths from clustered observations can be estimated as

$$\mu = \sum c_i x_i / C$$

where c is the number of individuals, x is its mean length in the i^{th} haul and C is the total number in the n hauls (Cochran 1977). However, as the number of hauls has been small (Tab. 1), jackknife estimates of mean length and its standard error were calculated according to Cochran (1977) where

$$\mu_{(i)} = \sum_{i \neq j} c_i x_i / (C - c_j)$$

is the weighted mean length deleting the n^{th} haul and

$$\mu_{(.)} = \sum \mu_{(i)} / n$$

is the estimate of the population mean length, and

$$se = \sqrt{((n-1)/n) \sum (\mu_{(i)} - \mu_{(.)})^2}$$

is the corresponding standard error.

The effect of tow duration on mean length was examined using the GLM model

$$L_{mean,ijkl} = \mu_{(.)jk} + TowDuration_i + \varepsilon_{ijk}$$

where L_{mean} is the mean length in each tow and $\mu_{(.),jk}$ is the jackknife estimate of mean length in stratum j and year k. Relative standard errors ($RSE_{jk} = se_{jk} / \mu_{(.)jk}$) were computed to evaluate whether mean length was adequately estimated by the given number of observations or whether certain strata and year combinations were more inhomogeneous than others and should better be excluded from the analysis. Two threshold levels of the relative standard error of the jackknife estimates of mean length were defined arbitrarily as $RSE_{jk} < 0.075$ and $RSE_{jk} < 0.050$.

Effect of tow duration on maximum observed length

Maximum values of the largest observed length in the tows belonging to the different combinations of stratum and year were used for a pair wise comparison of the two classes of tow duration (paired t-test).

Results

Overall average catch rates of 15 min tows were higher than for 30 min tows for both, northern shrimp and Greenland halibut in most years and for all years together (Tab. 2). However, no significant effect of tow duration on the catch rates were detected, irrespectively whether total biomass density or numerical density of the different size categories for the two species were considered (Tab. 3). Normal quantile plots of the residuals for northern shrimp (Fig. 2) and Greenland halibut (Fig. 3) do not indicate that the models for the catch rates were inappropriate. Tow duration remained also non-significant in models in which the stratum and year were replaced by mean densities for the respective stratum and year combinations (Tab. 4). It is therefore concluded that 15 min tows are as efficient as 30 min tows to measure the density of all sexual stages of northern shrimp and different size categories of Greenland halibut, and that the higher level of catches rates in 15 min tows were due to the significant effects of other factors, e.g. sampling location.

Average ratios of within stratum and year biomass density standard deviation and within stratum and year mean biomass density were 1.146 for northern shrimp and 1.072 for Greenland halibut (Fig. 4). For both species, the average ratios for 15 min and 30 min tows were rather similar (Northern shrimp: 1.148 and 1.143, Greenland halibut: 1.116 and 1.027) and no significant difference between the two tow durations was found (paired t-test; Northern shrimp: $t = 0.048$, d.f. = 42, $p = 0.962$; Greenland halibut: $t = 1.000$, d.f. = 36, $p = 0.324$). The result of the statistical analysis did not change when the minimum number of observation in each stratum, year and tow duration combination were increased from two to four (Northern shrimp: $t = -0.035$, d.f. =14, $p = 0.851$; Greenland halibut: $t = -0.103$, d.f. = 14, $p = 0.920$). Hence, there is no indication that 15 min tows give less precise results than 30 min tows.

Mean carapace length of northern shrimp ranged from 15.8 to 23.6 mm in the different strata and years. Mean total length of Greenland halibut varied between 14 and 39 cm. Large relative standard errors for the jackknife estimates of mean length were observed in several cases for both species, but in particular for Greenland halibut (Fig. 5). This indicates a considerable within-stratum variability of mean size and that the number of hauls in some strata was too small a reliable estimation of the overall mean population length. In addition to the inclusion of all possible hauls, i.e. all non-zero catches, the analysis was therefore also done for reduced data sets, in which the most inhomogeneous strata were removed. Here, thresholds of the relative standard error for the jackknife estimates of mean length for a given stratum and year combinations were applied. In all cases, however, the analysis of variance did neither for northern shrimp nor for Greenland halibut reveal a significant effect of tow duration on the mean length (Tab. 5). For northern shrimp, the normal quantile plots of the residuals (Fig. 6) do not indicate that the models were inappropriate. This was also the case for Greenland halibut although the normal quantile plots were less satisfactory (Fig. 7), which might be related to a much lower number of observations. However, the length frequency distributions of both species were apparently not affected by tow durations of 15 and 30 min.

Maximum observed length of northern shrimp and Greenland halibut in the different strata and years were highly variable for both, the 15 min and 30 min tows (Fig. 8). No significant effect of tow duration was found (paired t-test; Northern shrimp: $t = -0.682$, d.f. = 42, $p = 0.499$; Greenland halibut: $t = -0.020$, d.f. = 36, $p = 0.984$). This suggests, that also extreme values, i.e. the largest individuals, can be sampled adequately by 15 min tows.

Discussion

Previous studies of the effect of tow duration have shown that the mean sizes of several flatfish and gadoids were not affected by tow durations from 60 to 5 min (Godø et al. 1990, Walsh 1991), and similar results were reported for three crab species comparing 30 and 15 min tows (Somerton et al. 2002). This indicates that the effect of tow duration, if there is any, is the same for all sizes, which is consistent with our findings for northern shrimp and Greenland halibut.

Carothers and Chittenden (1985) found for two species of penaeid shrimp a significant relation between catch and tow durations of 5 to 30 min, but reported also that tow duration accounted for only a small proportion of the total variation in catch. Godø et al. (1990) and Walsh (1991) observed that catch per unit effort (CPUE) of flatfish and gadoid species increased significantly with decreasing tow duration only at tow durations below 15 min and remarked that higher catch rates of short tows were difficult to explain. Somerton et al. (2002) measured significant higher CPUE values in 15 min tows than in 30 min tows for two out of the three crab species studied, but it was not possible to identify the definite causal mechanism for this result.

The present study was not designed to detect mechanism that are independent of tow duration such as catch-by-surprise due to herding or errors in the measurements of the length of the tow path (Godø et al. 1990) or escapement below the footrope (Walsh 1992, Somerton et al. 2002), which would effect the results from short tows relatively more than those from long tows. However, the present analysis of the mixture of 30 and 15 min tows randomly allocated to the sampling locations within a stratum in the West Greenland Bottom Survey show that differences in catch rates of northern shrimp and Greenland halibut were due to stratum and year effects rather than caused by tow duration. This implies that a bias introduced by using 15 min instead of 30 min tows due to an ‘end-error’ is rather small and that other sources of variation, such as the within stratum differences of depth at the various sampling locations, are much more important. The fact that the existence of en end-error could not be demonstrated here does not necessarily mean that such an effect does not exist and could result in a systematic bias. Kingsley et al. (2002) estimated that the amount of northern shrimp caught outside the nominal tow period equals to 2.78 min additional towing time, which corresponds to about 9 % of a 30 min tow but to 18 % of a 15 min tow. This ‘end-error’ has been estimated with a high uncertainty (relative standard error: 42 %) and the relevance of such an effect was not confirmed by a later study (Kingsley 2001). Moreover, the results of the present study indicate that the magnitude of the 'end-error' is so small that the risk of introducing a serious bias to the time series of biomass estimates from the survey due to a reduction of tow duration to 15 min appears to be negligible.

Furthermore, no indication was found that 15 min tows give less precise estimates of biomass and numerical density for northern shrimp and Greenland halibut than 30 min tows, and there was also no significant difference between the two tow durations concerning their efficiency to sample extreme values such as maximum length. These conclusions, however, were derived from paired t-tests, and the lack of difference found here should be taken with some caution as its power depends very much on the number of observations (Sokal and Rohlf 1995).

In summary, we conclude that the actual mixture of 15 and 30 min tows in the West Greenland Bottom Trawl for shrimp and fish can be replaced by 15 min tows on all stations without interrupting the time series of survey estimates. The implementation of a standard towing time of 15 min would further be advisable because as it reduces the frequency of large catches, which are time consuming to handle, and because the gain in survey time related to the shorter average tow duration could be used for an increase of the total number of stations in order to improve the overall performance of the survey.

References

Carothers, P.E., Chittenden, M.E., 1985. Relationships between trawl catch and tow duration for penaeid shrimp. Trans. Am. Fish. Soc. 114 (6): 851-856.

Cochran, W.G., 1977. Sampling techniques. 3rd edition. Wiley, New York, 428 pp.

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

Gunderson, D.R., 1993. Surveys of fisheries resources. Wiley, New York, 248 pp.

Kanneworff, P., Wieland, K., 2001. Stratified-random trawl survey for northern shrimp (*Pandalus borealis*) in NAFO Subareas 0+1 in 2001. NAFO SCR Doc. 01/175, 23 pp.

Kingsley, M. C. S., 2001. Studies in 2001 on the end effect of the Skjervøy 3000 trawl in the West Greenland shrimp survey. NAFO SCR Doc. 01/177, 4 pp.

Kingsley, M. C. S., Kanneworff, P., Carlsson, D.M., 1999. Modifications to the design of the trawl survey for *Pandalus borealis* in West Greenland waters: Effects on bias and precision. NAFO SCR Doc. 99/105, 14 pp.

Kingsley, M.C.S., Carlsson, D.M., Kanneworff, P., Pennington, M., 2002. Spatial structure of the resource of *Pandalus borealis* and some implications for trawl survey. Fish. Res. 58: 171-183.

Kingsley, M. C. S., Kanneworff, P., 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 J. Mar. Sci. 61: 12-24.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. 2nd edition. Chapman and Hall, New York, 225 pp.

Pennington, M., 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.

Sokal, R.R., Rohlf, F.J., 1995. Biometry – The principles and practice of statistics in biological research. 3rd edition. W.H. Freeman, New York, 887 pp.

Somerton, D.A., Otto, R.S., Syrjala, S.E., 2002. Can changes in tow duration on bottom trawl surveys lead to changes in CPUE and mean size ? Fish. Res. 55: 63-70.

Storr-Paulsen, M., Jørgensen, O., 2004. Biomass and abundance of demersal fish stocks off West Greenland estimated from the Greenland shrimp survey, 1988-2003. NAFO SCR Doc. 04/18, 28 pp.

Walsh, S.J., 1991. Effect of tow duration on gear selectivity. NAFO SCR Doc. 91/84, 9 pp.

Walsh, S.J., 1992. Size-dependent selection at the footgear of a groundfish survey trawl. N. Am. J. Fish. Manage. 12(3): 625-633.

Wieland, K., Kanneworff, P., 2004. Revision of depth contours and stratification of the West Greenland Bottom Trawl Survey for Northern shrimp. Technical report No. 56, Greenland Institute of Natural Resources, 30 pp. (http://www.natur.gl/publikationer/tekniske rapporter).

Wieland, K., Kanneworff, P., Bergström, B., 2004. Results of the Greenland Bottom Trawl Survey for Northern shrimp (*Pandalus borealis*) off West Greenland (NAFO Subarea 1 and Divison 0A), 1988-2004. NAFO SCR Doc. 04/72, 31 pp.

Tab. 1: Number of stations by stratum, year and tow duration (last number in stratum name denotes depth intervals, i.e. –1: 150-200m, -2: 200-300m, -3: 300-400m, -4: 400-600m.

		Number of tows		Number of zero catches	
Stratum	Year	15 min	30 min	Northern shrimp	Greenland halibut
I1-3	2001	4	3	0	0
I1-3	2002	3	5	0	0
I1-3	2003	3	2	0	0
I1-4	2001	3	3	0	0
I2-2	2004	3	2	0	1
I2-4	2003	3	2	0	0
W1-2	1999	3	5	0	1
W1-2	2004	4	2	0	0
W1-3	1999	6	6	0	0
W1-3	2001	4	10	0	0
W1-3	2002	4	8	0	0
W1-3	2003	4	5	0	0
W1-3	2004	6	3	0	0
W2-2	2002	3	3	0	0
W2-3	2000	3	4	0	0
W2-3	2001	5	5	0	0
W2-3	2003	3	3	0	0
W2-4	2001	3	3	0	0
W3-2	1999	3	3	0	1
W3-2	2000	3	4	1	1
W3-2	2001	6	13	0	2
W3-2	2002	12	15	0	7
W3-2	2003	6	13	0	1
W3-2	2004	8	8	0	1
W4-2	1999	3	2	0	1
W4-2	2004	3	3	0	0
W4-3	2000	3	6	0	1
W4-3	2002	4	5	0	1
W5-1	2003	3	3	1	5
W5-2	2000	4	7	0	3
W5-2	2001	6	9	0	7
W5-2	2002	8	8	0	5
W5-2	2003	4	6	0	4
W5-2	2004	7	6	2	6
W6-2	2000	3	3	0	3
W6-2	2001	3	5	0	3
W7-2	2001	4	4	1	5
W7-2	2002	7	2	0	7
W7-2	2003	4	2	2	6
W7-2	2004	4	6	4	9
W8-3	2004	4	2	0	1
W8-4	2000	3	4	0	0
W8-4	2001	3	4	0	1

Tab. 2: Mean catch rates of northern shrimp and Greenland halibut by year and tow duration.

a) Northern shrimp:

Year	Tow duration	Number of hauls	Total biomass (kg/km^2)	Males (n/m^2)	Primiparous females (n/m^2)	Multiparous females (n/m^2)
1999	15 min	15	1174.318	0.261	0.018	0.007
	30 min	16	2195.119	0.381	0.050	0.021
2000	15 min	19	7573.347	1.027	0.134	0.193
	30 min	28	10014.280	2.038	0.165	0.143
2001	15 min	41	6677.144	0.963	0.168	0.147
	30 min	59	8676.349	0.923	0.240	0.254
2002	15 min	41	6804.320	1.420	0.129	0.085
	30 min	46	5897.337	0.874	0.150	0.093
2003	15 min	30	10863.820	1.198	0.268	0.404
	30 min	36	10070.520	1.293	0.326	0.163
2004	15 min	39	16865.170	1.605	0.682	0.399
	30 min	32	5939.267	0.743	0.743	0.209
1999 - 2004	15 min	185	9224.376	1.194	0.271	0.223
	30 min	217	7609.672	1.048	0.207	0.155

b) Greenland halibut:

Year	Tow duration	Number of hauls	Total biomass (kg/km^2)	< 20 cm (n/km^2)	20 - 40 cm (n/km^2)	> 40 cm (n/km^2)
1999	15 min	15	59.267	2174.141	75.730	8.694
	30 min	16	128.048	4759.454	194.354	9.699
2000	15 min	16	132.368	2002.274	226.969	32.004
	30 min	25	66.112	823.384	102.736	24.883
2001	15 min	37	370.712	5635.299	1052.802	76.266
	30 min	55	295.996	8110.324	857.108	30.529
2002	15 min	34	183.694	1828.765	909.842	16.557
	30 min	44	133.586	2233.558	433.926	11.225
2003	15 min	23	506.640	8790.022	1561.821	68.944
	30 min	31	279.144	1452.149	1192.492	72.090
2004	15 min	35	84.265	2021.223	151.418	14.770
	30 min	26	56.163	857.546	174.757	2.463
1999 - 2004	15 min	160	234.818	3801.537	724.233	38.312
	30 min	197	182.603	3609.879	577.204	26.740

Tab. 3: Results of the analysis of variance for the effect of stratum, year, depth and tow duration on catch rates of northern shrimp and Greenland halibut ([1]: records of length frequencies by sexual group missing for three tows, [2]: records of length frequencies missing for two tows; number in brackets: p-values for the effect of tow duration when non-significant terms, i.e. Year or Year and Depth, were removed).

	Response	N_{Haul}	Effect	d.f.	F	p	
Northern shrimp	Total biomass (kg/km^2)	402	Stratum	17	9.873	< 0.001	
			Year	5	2.294	0.045	
			Depth	1	9.210	0.003	
			Tow duration	1	0.137	0.712	
	Males (n/m^2)	399 [1]	Stratum	17	8.028	< 0.001	
			Year	5	1.581	0.164	
			Depth	1	3.015	0.083	
			Tow duration	1	0.018	0.893	(0.961)
	Primiparous females (n/m^2)	399 [1]	Stratum	17	5.460	< 0.001	
			Year	5	3.056	0.010	
			Depth	1	13.126	< 0.001	
			Tow duration	1	0.172	0.679	
	Multiparous females (n/m^2)	399 [1]	Stratum	17	5.489	< 0.001	
			Year	5	3.900	0.002	
			Depth	1	11.246	0.001	
			Tow duration	1	0.011	0.918	
Greenland halibut	Total biomass (kg/km^2)	357	Stratum	15	15.723	< 0.001	
			Year	5	0.713	0.614	
			Depth	1	4.465	0.035	
			Tow duration	1	0.587	0.444	(0.411)
	< 20 cm (n/km^2)	355 [2]	Stratum	15	14.776	< 0.001	
			Year	5	0.165	0.975	
			Depth	1	0.007	0.933	
			Tow duration	1	1.199	0.274	(0.260)
	20 - 40 cm (n/km^2)	355 [2]	Stratum	15	19.389	< 0.001	
			Year	5	1.652	0.140	
			Depth	1	27.408	< 0.001	
			Tow duration	1	2.532	0.115	(0.157)
	> 40 cm (n/km^2)	355 [2]	Stratum	15	5.489	< 0.001	
			Year	5	3.900	0.002	
			Depth	1	11.246	0.001	
			Tow duration	1	0.011	0.918	

Tab. 4: Results of the analysis of variance for the effect of stratum mean density ($\log_e$-transformed) and tow duration on catch rates of northern shrimp ([1]: records of length frequencies by sexual group missing for three tows, [2]: records of length frequencies missing for two tows and cases with zero mean density excluded).

	Response	N_{Haul}	Effect	d.f.	F	p
Northern shrimp	Total biomass	402	Mean density	1	38.159	< 0.001
	(kg/km^2)		Tow duration	1	1.825	0.177
	Males	399 [1]	Mean density	1	58.773	< 0.001
	(n/m^2)		Tow duration	1	0.360	0.549
	Primiparous females	399 [1]	Mean density	1	34.062	< 0.001
	(n/m^2)		Tow duration	1	1.487	0.223
	Multiparous females	399 [1]	Mean density	1	30.407	< 0.001
	(n/m^2)		Tow duration	1	0.759	0.384
Greenland halibut	Total biomass	357	Mean density	1	147.283	< 0.001
	(kg/km^2)		Tow duration	1	0.911	0.340
	< 20 cm	349 [2]	Mean density	1	167.638	< 0.001
	(n/km^2)		Tow duration	1	0.989	0.321
	20 - 40 cm	355 [2]	Mean density	1	223.403	< 0.001
	(n/km^2)		Tow duration	1	1.844	0.175
	> 40 cm	302 [2]	Mean density	1	140.183	< 0.001
	(n/km^2)		Tow duration	1	0.296	0.587

Tab. 5: Results of the analysis of variance for the effect of tow duration on the mean length of northern shrimp and Greenland halibut (a): all possible hauls, b) and c): reduced data sets using threshold levels of 0.075 and 0.050 for the relative standard error of the mean length in a given stratum and year).

	N_{Haul}	Effect	d.f.	F	p
Northern shrimp	a): 387	Mean length in stratum/year	1	118.726	< 0.001
		Tow duration	1	0.723	0.396
	b): 329	Mean length in stratum/year	1	110.516	< 0.001
		Tow duration	1	0.616	0.433
	c): 241	Mean length in stratum/year	1	76.8335	< 0.001
		Tow duration	1	1.208	0.272
Greenland halibut	a): 308	Mean length in stratum/year	1	218.404	< 0.001
		Tow duration	1	0.062	0.804
	b): 201	Mean length in stratum/year	1	99.932	< 0.001
		Tow duration	1	0.450	0.503
	c): 144	Mean length in stratum/year	1	17.255	< 0.001
		Tow duration	1	0.590	0.443

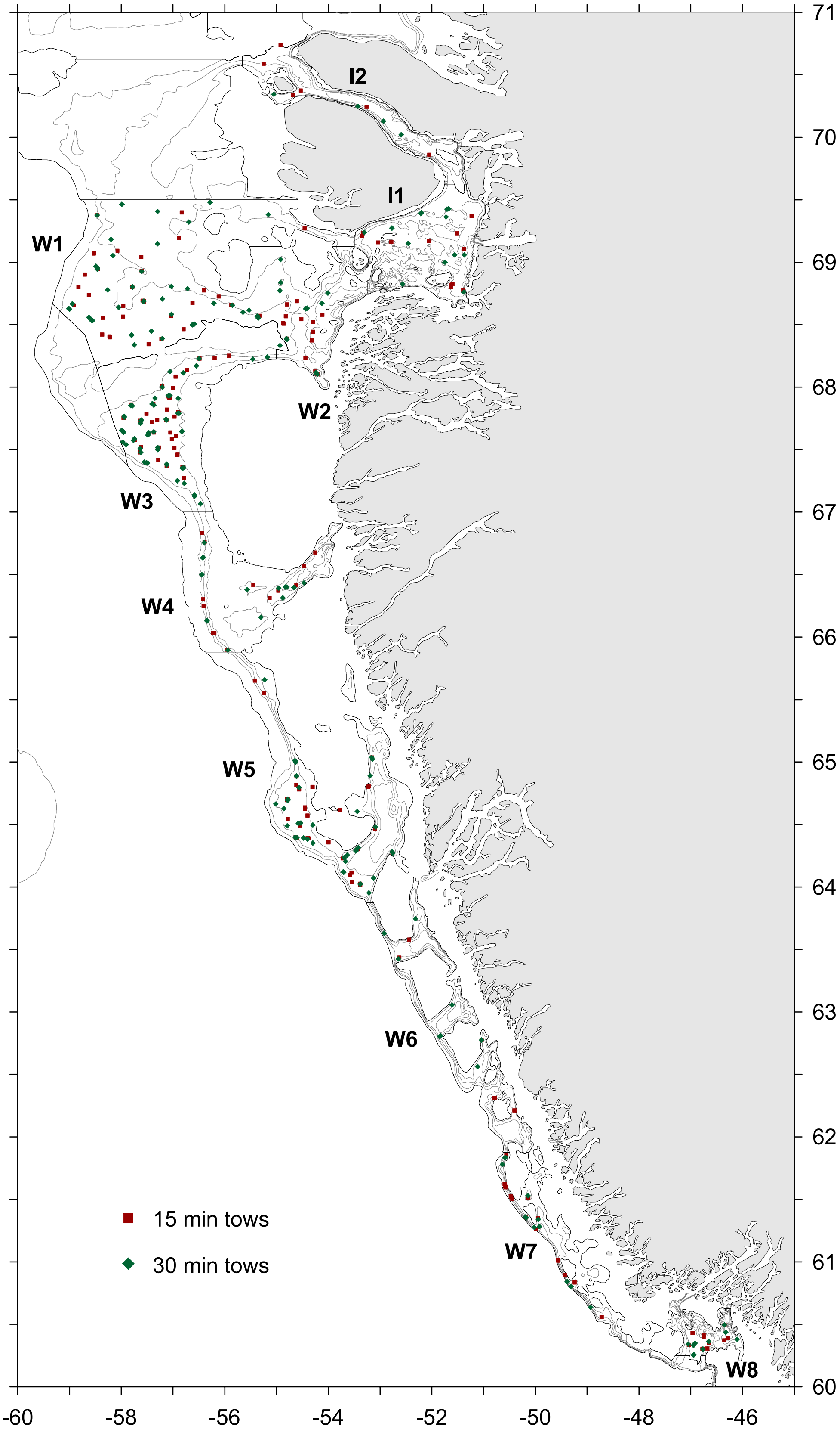


Fig. 1: Location of 15 and 30 min tows in 1999-2004 and stratification of the West Greenland Bottom Trawl Survey (Labels on map indicate areas, each area divided into four depth strata: 150-200m, 200-300m, 300-400m and 400-600m).

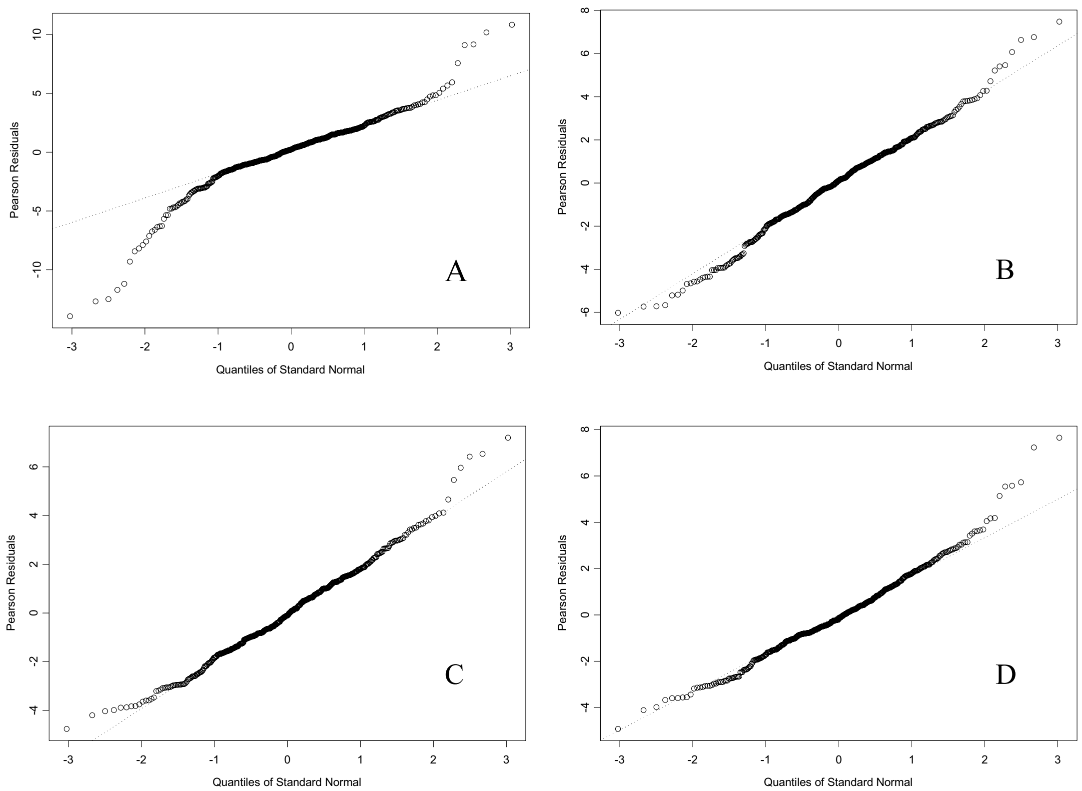


Fig. 2: Normal quantile plots for the GLM analysis of the effect of stratum, year, depth and tow duration on catch rates of northern shrimp (A: total biomass, B: males, C: primiparous females, D: multiparous females).

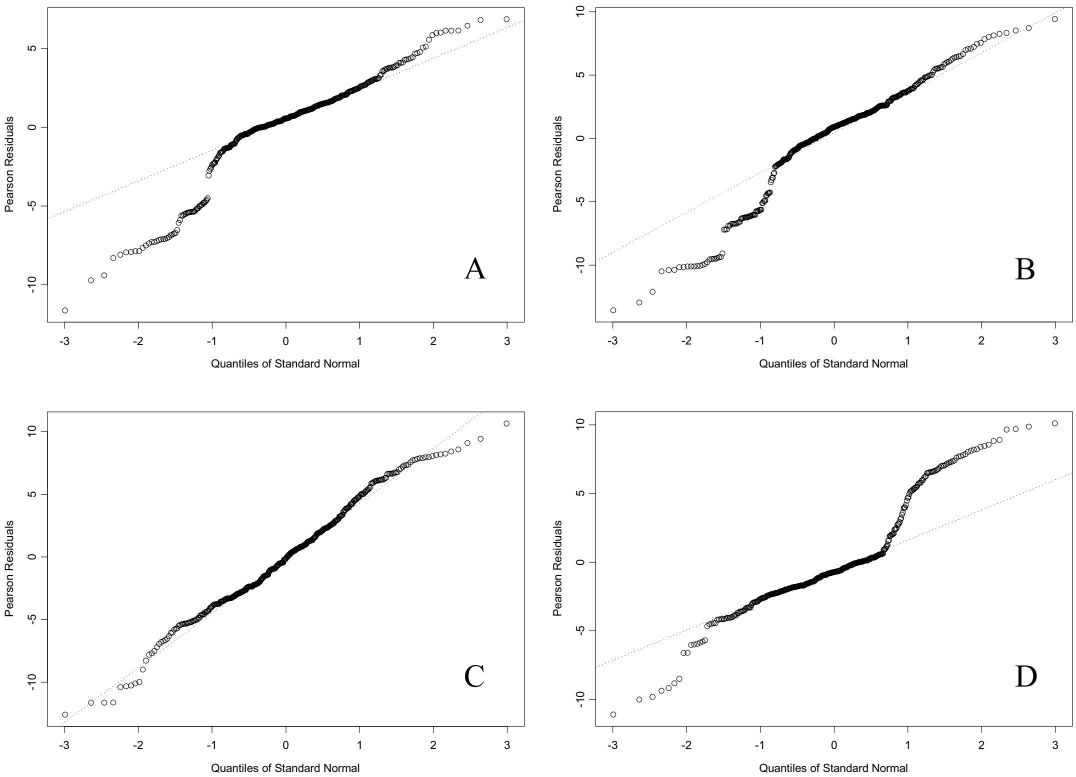
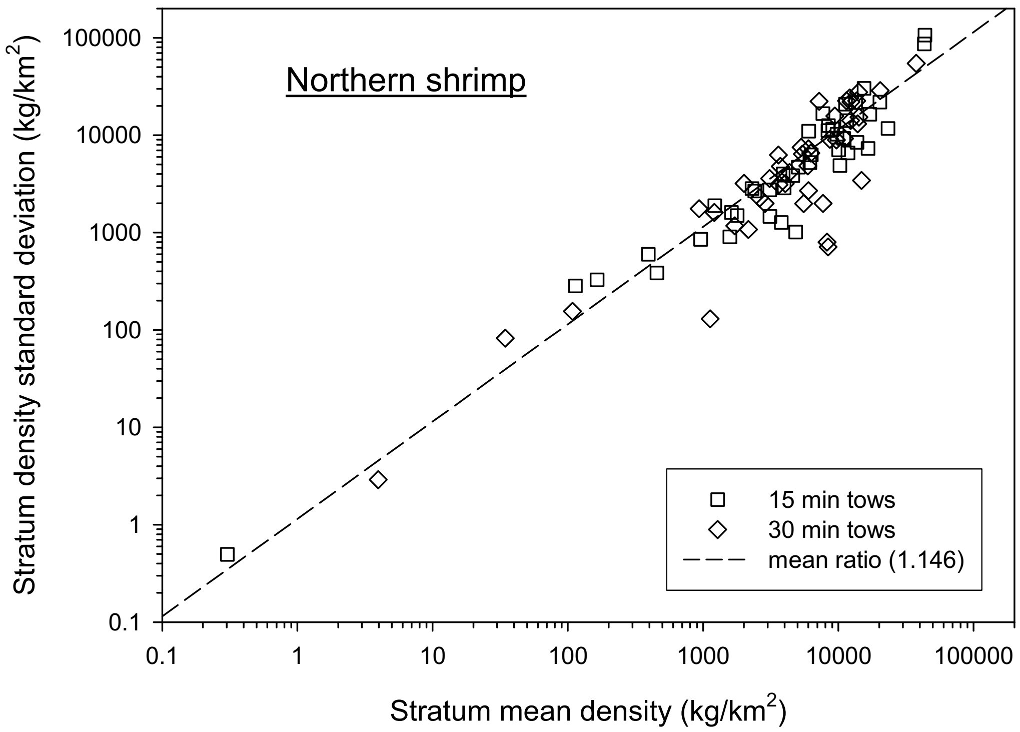


Fig. 3: Normal quantile plots for the GLM analysis of the effect of stratum, year, depth and tow duration on catch rates of Greenland halibut (A: total biomass, B: < 20 cm, C: 20 – 40 cm, D: > 40 cm).



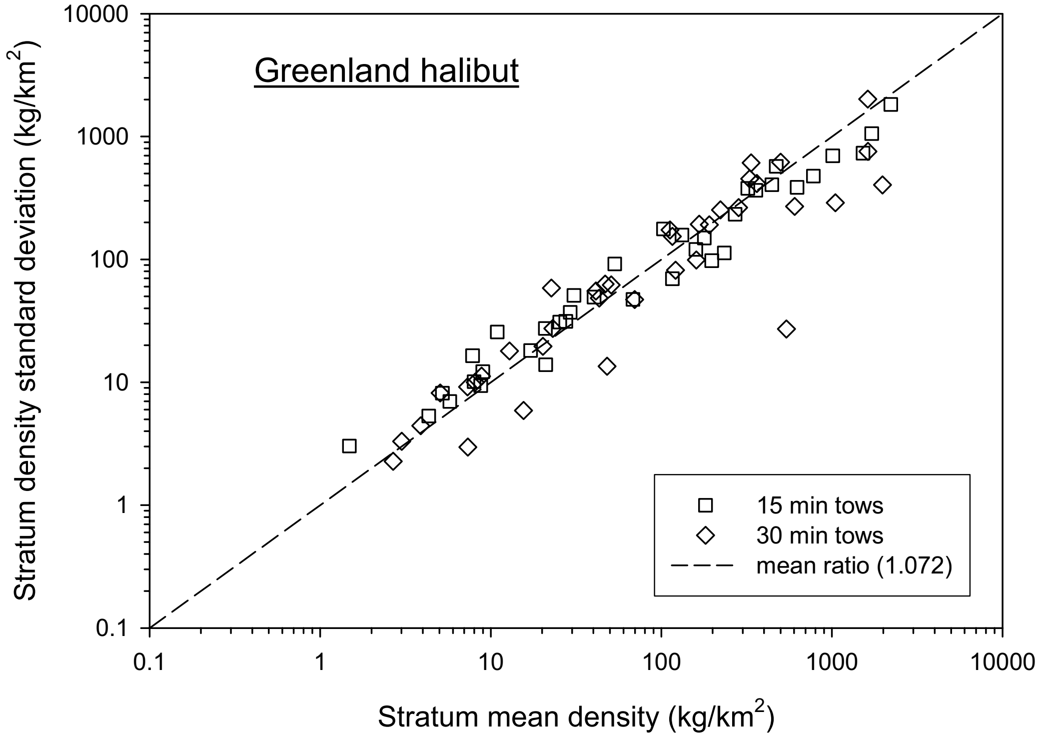


Fig. 4: Within stratum and year standard deviation of biomass density vs. within stratum and year mean biomass density of 15 and 30 min tows for northern shrimp and Greenland halibut.

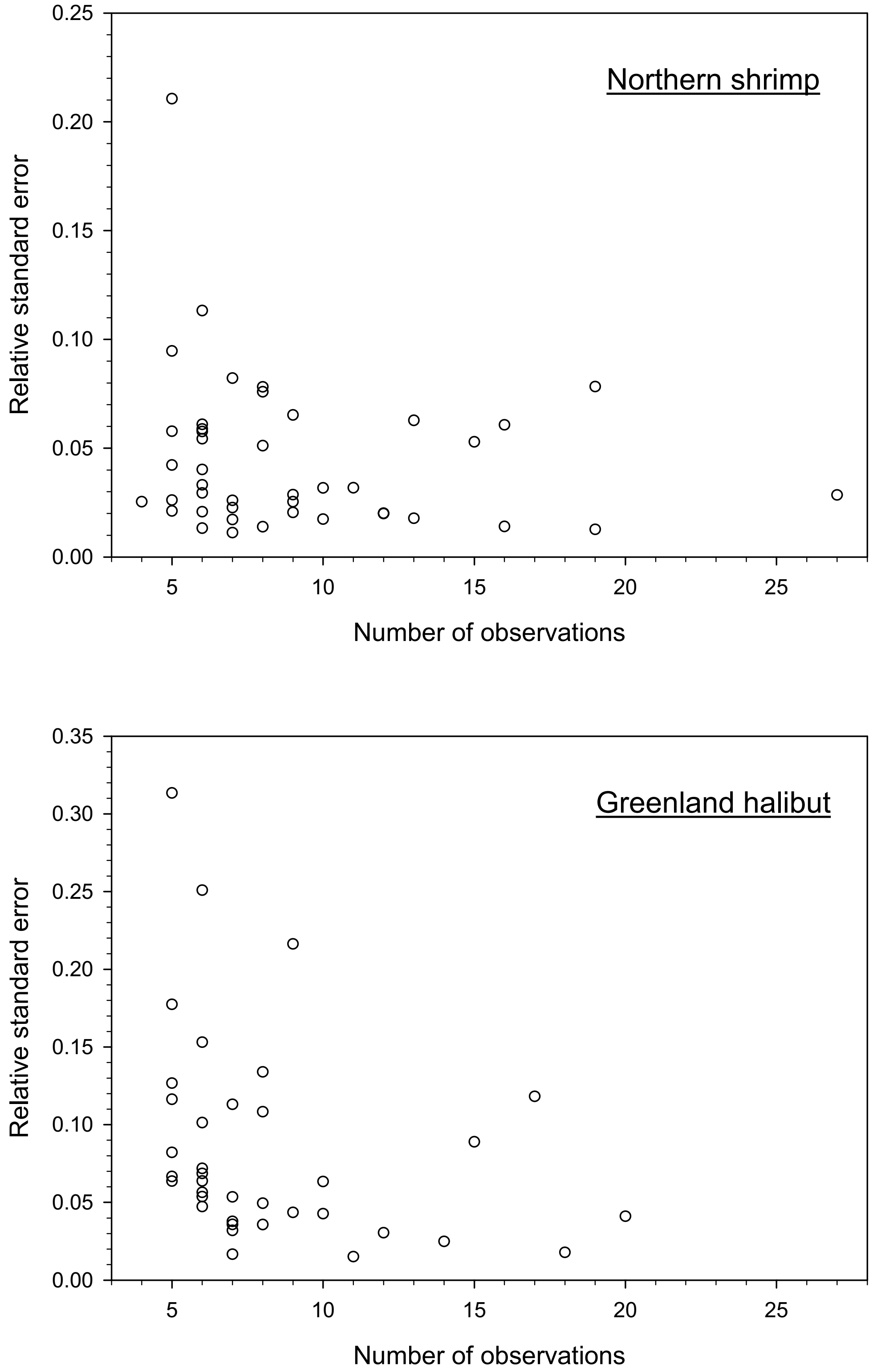


Fig. 5: Relative standard error (RSE = standard error / mean) for the jackknife estimates of mean length of northern shrimp and Greenland halibut vs. number of observations within a stratum and year (15 and 30 min tows combined).

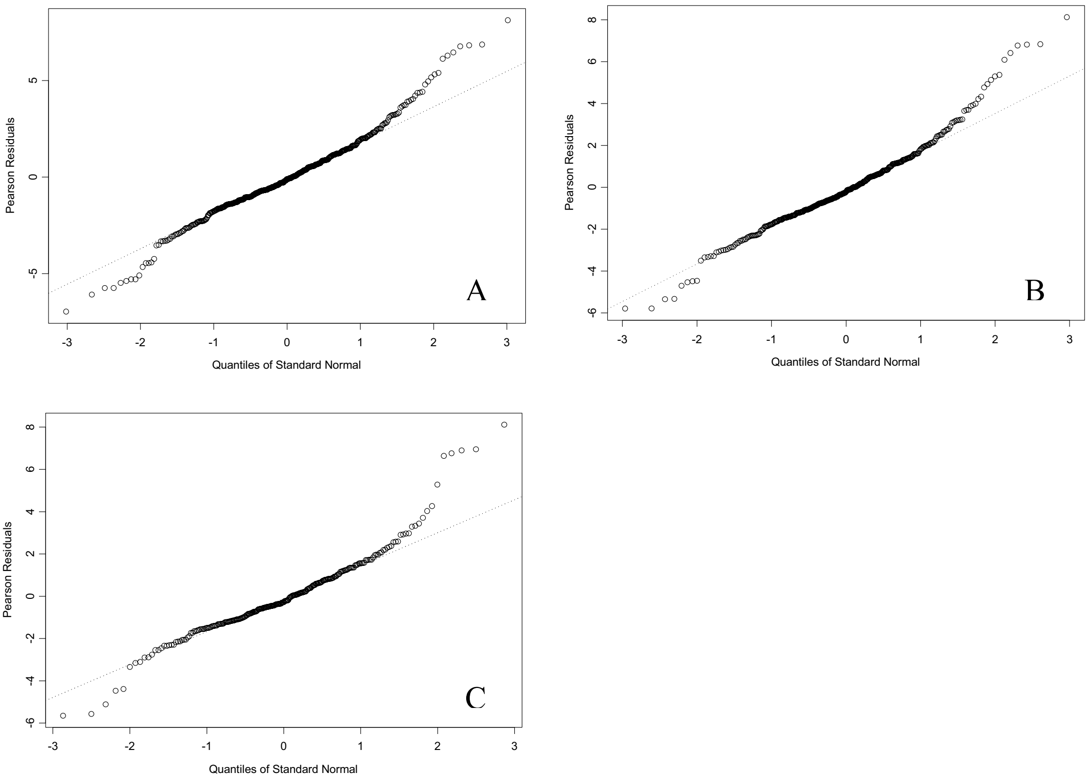


Fig. 6. Normal quantile plots for the GLM analysis of the effect tow duration on mean length of northern shrimp (A: all hauls, B and C: reduced data sets).

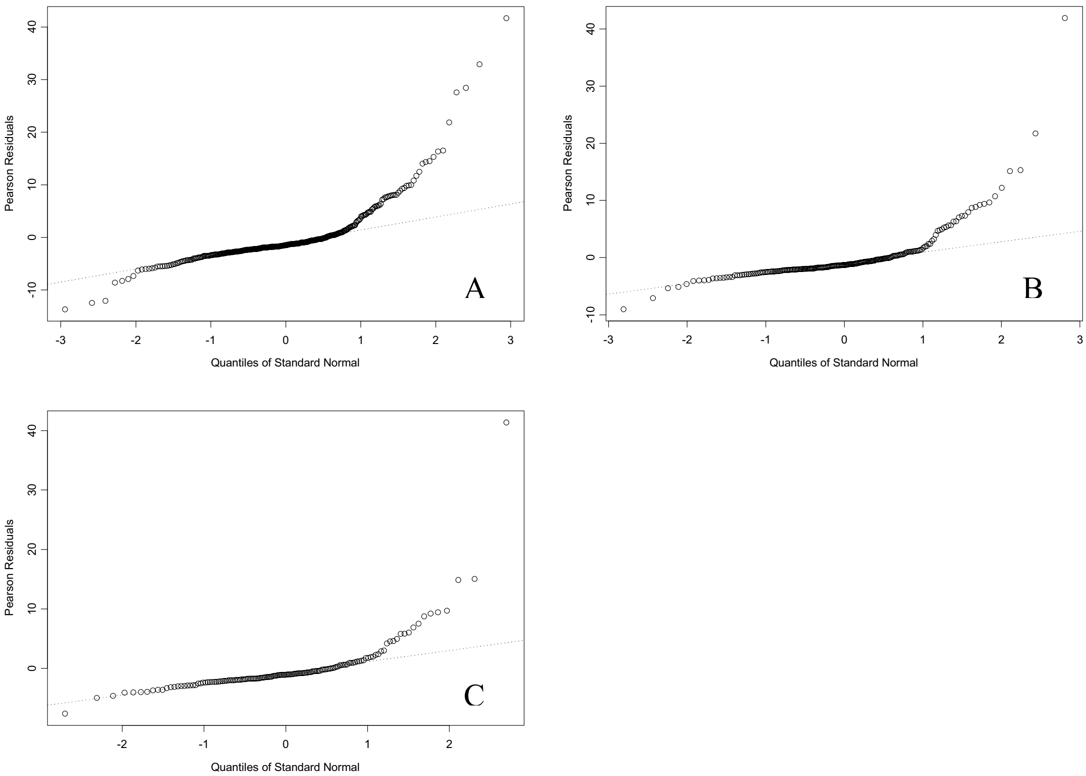


Fig. 7. Normal quantile plots for the GLM analysis of the effect tow duration on mean length of Greenland halibut (A: all hauls, B and C: reduced data sets).

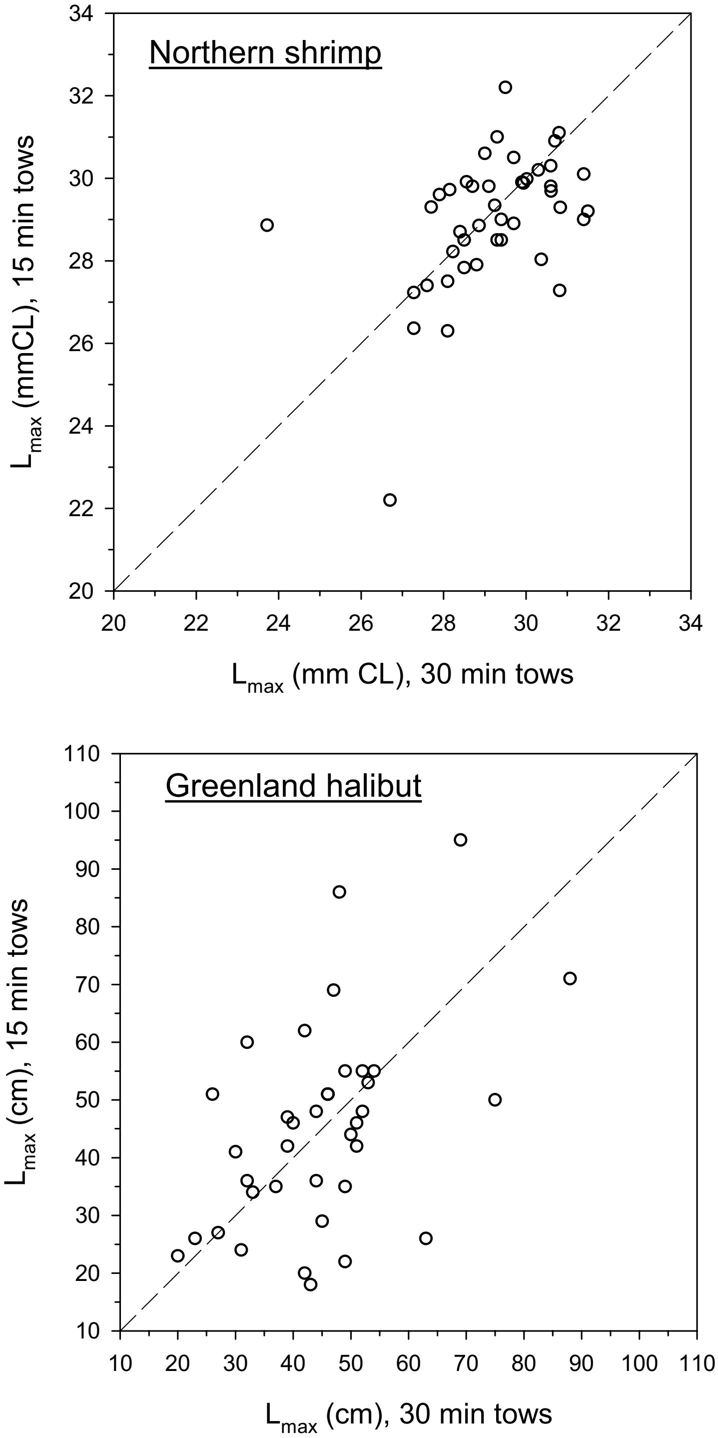


Fig. 8. Maximum length of northern shrimp and Greenland halibut (for strata and years with at least two non-zero catches in each of the two classes of tow duration).

Annex 10: Working Document 7

Optimum duration/distance of tows during surveys *by* R. Oeberst

Working Paper **Workshop on Survey Analyses and Design**
Sete, France 09. – 13.05.2004

OPTIMUM DURATION/DISTANCE OF TOWS DURING SURVEYS

by

R. Oeberst
Bundesforschungsanstalt für Fischerei Hamburg
Institut für Ostseefischerei Rostock
An der Jägerbäk 2
D - 18069 Rostock, Germany

Introduction

Tow durations from 30 min to 2 h are used during different trawl surveys. The lowest value of 30 minutes is used during the Baltic Internationals Trawl surveys (ICES 2002). Furthermore, the total period of the surveys is limited and can not be expanded depending on the observed situation during the survey. The defined standard duration of the haul is important because changes influence the number of stations which can be realized during a survey when the total survey period is limited. On the other hand a change of the standard haul duration influences the variability of the catch per unit effort (CPUE).

Oeberst (1985) studied the variance of the CPUE values of tow which were realized for estimating the species composition during acoustic surveys and suggested that the variability of CPUE increases with duration of the tow based on the autocorrelation between the subsequent parts of the tow. Godø et al. (1990) compared the length distribution of catches with tow duration from 5 min to 2 h and did not found significant changes in mean length. Pennington and Vøstad (1991) showed that reducing of tow duration results in increasing appropriately the number of stations and in an increase in the precision of survey estimates when the level of survey resources is fixed. Pennington and Vøstad (1994) have shown that the reduced tow duration in combination with an increase of the number of hauls at more locations can results in more precise estimates of population parameters due to intra-cluster correlation when the required assumptions are fulfilled.

The study analyses the effects of increasing tow duration/distance, the periods which are necessary for veering and heaving, the number of hauls which can be realised and the autocorrelation of fish density of subsequent parts of the tow concerning the mean and the variance of the catch per unit effort.

Mathematical background

It is assumed that x presents the CPUE values of the target species of the smallest possible haul duration (or distance), d, the standard duration. The mean and variance of x are E[x] and V[x]. Furthermore, it is also assumed that the duration of the haul can be chosen by n times of subsequent parts of d.

The CPUE value, y_{nj}, of the n subsequent parts of the standard duration of d can be given by

$$y_{nj} = \sum_{i=1}^{n} x_{ij} \tag{1}$$

with i used as the index of the sequence of parts of d during the haul and j as index of the haul realized during the survey. M_n denotes the number of hauls which can be realized during the total survey period. The mean of $y_{n,j}$ can be given by

$$E[y_{n,j}] = nE[x]. \tag{2}$$

Equation 2 shows that the CPUE value increases with increasing distance of the tow.
The variance of $y_{n,j}$ can be given by

$$V[y_{n,j}] = \sum_{i=1}^{n} V[x_{i,j}] + 2\sum_{k=1}^{n-1}(n-k)Cov(x_{1,j}, x_{1+k,j}) \tag{3}$$

with $Cov(x_{1,j},x_{1+k,j})$ as notation of the covariance between the parts $x_{1,j}$ and $x_{1+k,j}$ of the haul. Since the variance of each part of the haul, $x_{i,j}$, is equal to the variance of x follows

$$V[y_{n,j}] = nV[x] + 2\sum_{k=1}^{n-1}(n-k)Cov(x_{1,j}, x_{1+k,j}) \tag{4}$$

Taking into account that

$$Cov(x_{1j}, x_{1+k,j}) = r(x_{1,j}, x_{1+k,j})\sqrt{V[x_{1,j}]V[_{1+k,j}]} \tag{5}$$

with r used as notation of the autocorrelation coefficient further follows

$$V[y_{n,j}] = nV[x] + 2\sum_{k=1}^{n-1}(n-k)r(x_{1,j}, x_{1+k,j})V[x] = V[x]\{n + 2\sum_{k=1}^{n-1}(n-k)r(x_{1,j}, x_{1+k,j})\}. \tag{6}$$

That means that the variance of the hauls with a duration of n + d depends on the autocorrelation between the subsequent parts $x_{i,j}$ of haul j.
Assuming that $r(x_{1,j},x_{1+k,j})$ is very close to 1 follows that

$$V[y_{n,j}] \approx n^2 V[x] \tag{7}$$

and for $r(x_{1,j},x_{1+k,j})$ is equal zero

$$V[y_{n,j}] = n\ V[x]. \tag{8}$$

That means that $E[y_{n,j}]$ is larger than E[x] and that $V[y_{n,j}]$ is larger than V[x] for n >1 and $r(x_{1,j},x_{1+k,j}) \geq 0$ and that an increase of the haul duration normally results in an increase of mean and variance of the CPUE values of the new haul duration. When $r(x_{1,j},x_{1+k,j})$ is -1 for n = 1 follows that $V[y_{n,j}]=V[x]$.
For comparing the effects of the different haul duration and the autocorrelation between the densities of subsequent parts the quotient between the halve of the confidence interval and the mean of CPUE values is used. This step of analyses can be easily replaced by the requirement that the quotient of the halve of the confidence interval and the mean CPUE value is less or equal to a constant value of k for all realised tow durations.
Assuming that x is normally distributed the confidence interval can be given by

$$[E[y_{n,j}] \pm t^2(M_n, 1-\frac{\alpha}{2})\sqrt{\frac{V[y_{n,j}]}{M_n}} \tag{9}$$

And for n = 1 the quotient is

$$\frac{t(M_1 - 1, 1 - \frac{\alpha}{2})}{\sqrt{M_1}} \frac{\sqrt{V[x]}}{E[x]}. \tag{10}$$

For n = 2 follows

$$\frac{t(M_2-1,1-\frac{\alpha}{2})}{\sqrt{M_2}}\frac{\sqrt{V[y_{2,j}]}}{E[y_{2,j}]}=\frac{t(M_2-1,1-\frac{\alpha}{2})}{\sqrt{M_2}}\frac{\sqrt{V[x]}}{E[x]}\frac{\sqrt{2+2r(x_{1,j},x_{2,j})}}{2}. \tag{11}$$

Using Equation 10 and 11 the autocorrelation between the two subsequent parts of the haul, r_1, can be estimated where the quotients between halve of the confidence interval and the means are equal.

$$\frac{t(M_1-1,1-\frac{\alpha}{2})}{\sqrt{M_1}}\frac{\sqrt{V[x]}}{E[x]}=\frac{t(M_2-1,1-\frac{\alpha}{2})}{\sqrt{M_2}}\frac{\sqrt{V[x]}}{E[x]}\frac{\sqrt{2+2r(x_{1,j},x_{2,j})}}{2}. \tag{12}$$

Then r_1 can be estimated by

$$r_1=\frac{1}{2}\{[\frac{t(M_1-1,1-\frac{\alpha}{2})}{t(M_2-1,1-\frac{\alpha}{2})}\frac{n\sqrt{M_2}}{\sqrt{M_1}}]^2-2\}. \tag{13}$$

The autocorrelation r_1 is only dependent on the number of realized hauls and independent of the distribution parameters mean and variance of x.
When the observed autocorrelation between the subsequent parts is larger than r_1 it is not useful to increase the haul duration due to increasing of the quotient between halve of the confidence interval and the mean. On the other hand, autocorrelation coefficient of less than r_0 suggests that the duration of the haul should be expanded.
For n = 3 the autocorrelation between CPUE values of the first and third part of the haul can be estimated based on r_1 by

$$r_2=\frac{1}{2}\{[\frac{t(M_1-1,1-\frac{\alpha}{2})}{t(M_3-1,1-\frac{\alpha}{2})}\frac{n\sqrt{M_3}}{\sqrt{M_1}}]^2-2\}-2r_1. \tag{14}$$

Equation 14 can also be used for estimating r_2 dependent on observed data of the autocorrelation between subsequent parts of the haul and vice versa.
It is common practice that the total period of the survey is limited, notated by T_S. For estimating the total number of hauls, M_n, which can be realized during the survey it is necessary to take into account the periods which are necessary for veering the gear to the bottom, T_1, and for heaving the gear, T_2, as well as the duration of the haul, T. Dependent on the periods which are needed for sailing from haul to haul the total period for realizing the hauls, T_0, can be estimated. Because the period for sailing from haul to haul is very variable and in some cases the total time of for sailing between the selected positions is not necessary high correlated with the total number of planed hauls (all strata must be covered) for the studies only T0 was considered.
The number of standard hauls can be estimated by

$$M_1 \le T_0/(T_1+T_2+T). \tag{15}$$

When the duration of the hauls is expanded to n times of the standard haul follows

$$M_n \le T_0/(T_1+T_2+nT). \tag{16}$$

Because the necessary periods for veering and heaving of the trawl are dependent on the water depth the relations between M_1 and M_n differ from area to area. Using data which based on

the Baltic International Trawl Survey examples are given for the relation between the different periods of haul duration and r_1.
Dependent on the mean CPUE value per standard haul duration, d, the number of hauls, and the duration of the hauls the total number of captured individuals of the target species varied. This must be taken into account because the haul duration must be long enough to describe the length distribution of the stock with required accuracy and the get enough samples for estimating the length-age key, the length-weight relationship etc. Therefore, it may be possible that the optimum haul duration is to short to describe the all biological parameters of the stock.

Results of Simulations for the Baltic Sea

The water depth in the Arkona Sea, part of the Baltic Sea which is covered by Germany during the international co-ordinated trawl surveys vary from 20 to 50 m. Based on practical experience the mean periods for veering and heaving vary between 15 and 20 minutes. The required velocity is 3 knots and the tow duration is 30 minutes (ICES 2002, Addendum). Furthermore, about 45 tows are planned for each survey in spring and autumn. Altogether 2700 minutes are used for the tows when mean periods $T_1 = T_2 = 15$ minutes is used and 3250 minutes are necessary when $T_1 = T_2 = 20$ minutes.
Using a minimum possible tow duration of T = 5 minutes and $T_0 = 2700$ minutes the effects of the expansion of tow duration was studied for different $T_1 = T_2$. The number of tows, M_n, for n = 1, 2 and 3 and the autocorrelation coefficients r_1 and r_2 (Equ. 13, 14) were estimated. Furthermore, the expected mean total number of capture individuals was estimated assuming the 5 individuals are capture per minute in mean. For n = 1 the quotient $t(M_1\text{-}1,1\text{-}\alpha/2)/\sqrt{M_1}$ was calculated. First kind of error of $\alpha = 0.05$ was used for all estimates. The number of tows decreased dependent on increasing $T_1 = T_2$ (Tab.1) and the number of capture individuals also decreased due to the decreasing of total time which can be used for the catch. With increasing n the number of tows decreased for all $T_1 = T_2$, however, the number of captured individuals increased. This aspect is important when a level of accuracy is required for estimating the different parameters of the stock (length distribution, age-length-key, …). The autocorrelation coefficient r_1 increased dependent on increasing $T_1 = T_2$. The same development was found for r_2 (Tab. 1).
The results suggest that the required level of autocorrelation between the subsequent standard parts of the tow for expanding the tow duration is dependent on the depth of the water. In shallower waters where 10 minutes are necessary for veering and heaving an expansion of the tow distance is only useful when the autocorrelation is very low. In deeper waters where 30 minutes are necessary for veering and heaving of the trawl the level of autocorrelation must be lower than 0.84 to propose a tow duration of 10 minutes. With increasing depth/periods for veering and heaving the coefficient of autocorrelation r_1 increases. That means that low autocorrelations of subsequent parts suggest an increase of tow duration. The simulated data suggest that different tow durations are proposed for the different depth layers when the autocorrelation $r_1$are observed between 0.7 and 0.8. Because the used of different tow durations during the same survey results in problem during the combination of the data of the different depth layers it seems to be useful to use the most appropriate to duration for all hauls.
On the other hand the required level of autocorrelation $r_1$decreased when the time for the tows, T, increased and the periods for veering and heaving were constant with $T_1 = T_2 =15$ minutes and the total period for tows were also $T_0 = 2700$ minutes (**Tab. 2**).

Discussion

The studies have shown that increasing tow duration/distance, the periods which are necessary for veering and heaving, the number of hauls which can be realised and the autocorrelation of fish density of subsequent parts of the tow significantly influence the mean and the variance of the catch per unit effort. Increasing tow duration always results in increasing mean CPUE values and increasing variance of the CPUE values (Equation 2, 7, 8). Comparable results were described by Pennington and Vøstad (1991) based on the comparisons of tows with tow durations from 5 minutes to 2 hours. The increase of the variance is dependent of the autocorrelation of the fish density of subsequent parts. Therefore, the effects of increasing tow duration can only be studied based on the quotient between the halve of the confidence interval of CPUE values and the mean CPUE values.
The studies have shown that the optimum haul duration can vary from area to area and dependent on the water depth as well as the autocorrelation between the subsequent parts of the haul. Studies based on acoustic estimates in the Baltic Sea have shown that the autocorrelation between subsequent parts of 1 nm can vary from 0.61 to -0.02 and for distances of 4 nm from 0.32 to -0.03 (Oeberst 1985)..
Furthermore, it can be assessed that short haul durations are more suitable in shallow water and that surveys in deeper waters should used longer haul periods. However, short tow duration results in decreasing total number of captured individuals due to the smaller total realised period of tows. That means that a very short tow duration in combination of very small density of the target species can results in to low number of captured individuals for estimating the biological parameters of the stock. Pennington and Vøstad (1994) have shown that short tow duration which results in increasing number of tows at more locations can results in more precise estimates of population parameters due to intra-cluster correlation.

Conclusion

Based on this study it can be concluded that common usable tow duration for all areas and surveys can not be defined. Dependent on the estimates of the different influencing parameters the optimum tow duration must be adapted.

References

Godø, O.R., Pennington, M., Vølstad, J.H. 1990. Effect of tow duration on length composition of trawl catches. Fisheries Research, 9 : 165-179
ICES. 2002. Report of the Baltic International Fish Survey Working group (WGBIFS). ICES CM 2002/G:05, Ref. H. 202 pp.
Oeberst, R. 1985. Zu einigen Aspekten der Planung von hydroakustischen Bestandsbestimmungen. Fischerei - Forschung Rostock 23 (4) : 77 – 88.
Oeberst, R. 1986. Some aspects of planning of acoustic stock estimations. ICES, C. M. 1986/B:21, Fish Capture Committee, 10pp.
Pennington, M., Vølstad, J. H. 1991. Optimum size of sampling unit for estimating density of marine populations. Biometrics 47, 717-723
Pennigton, M., 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.

Tables

Table 1: Level of required autocorrelation, r_1 and r_2 of subsequent parts dependent on the period of veering, T_1, and heaving, T_2, of the trawl where total time for the tows is 2700 minutes. It is assumed that 5 individuals are capture per minute in mean.

n = 1							n = 2			n = 3		
T_1 [min]	T_2 [min]	T [min]	T_0 [min]	M_1	$t/\sqrt{M}$	Catch [numb.]	M_2	r_1	Catch [numb.]	M_3	r_2	Catch [numb.]
10	10	5	2700	108	0.19	2700	90	0.66	4500	70	0.36	5775
15	15	5	2700	77	0.23	1925	67	0.73	3350	60	0.51	4500
20	20	5	2700	60	0.26	1500	54	0.79	2700	49	0.56	3675
25	25	5	2700	49	0.29	1225	45	0.83	2250	41	0.58	3075
30	30	5	2700	41	0.32	1025	38	0.84	1900	36	0.73	2700

Table 2: Level of required autocorrelation, r_0, of subsequent parts dependent on the total period for the tows, T_0, when the periods for veering, heaving and the tow duration are constant

n = 1							n = 2			n = 3		
T_1 [min]	T_2 [min]	T [min]	T_0 [min]	M_1	$t/\sqrt{M}$	Catch [numb.]	M_2	r_1	Catch [numb.]	M_3	r_2	Catch [numb.]
15	15	5	2700	77	0.23	1925	67	0.73	3350	60	0.51	4500
15	15	10	2700	67	0.25	3350	54	0.60	5400	45	0.27	6750
15	15	15	2700	60	0.26	4500	45	0.48	6750	36	0.17	8100
15	15	20	2700	54	0.27	5400	38	0.38	7600	30	0.15	9000
15	15	25	2700	49	0,29	6125	33	0.31	8250	25	0.054	9375
15	15	30	2700	45	0.30	6750	30	0.29	9000	22	-0.02	9900

Annex 11: Working Document 8

Species composition in scattered layer based on control hauls *by* R. Oeberst

Working Paper **Workshop on Survey Analyses and Design**
Sete, France 09. – 13.05.2004

SPECIES COMPOSITION IN SCATTERED LAYER BASED ON CONTROL HAULS

by

R. Oeberst
Bundesforschungsanstalt für Fischerei Hamburg
Institut für Ostseefischerei Rostock
An der Jägerbäk 2
D - 18069 Rostock, Germany

Introduction

Acoustic surveys are used for estimating stock indices of pelagic and demersal species in many areas of the world (Parrish 2004, Mac Lennan and Simmonds 1992, ICES 2004 a, b). In cases where single individuals can be detected or where schools can be assigned to species based on special characteristics the acoustic signals can be directly used for estimating the fish densities.
However, when the species are distributed in mixed scattered layers or the acoustic signal characteristic is not clearly interpretable control hauls are necessary for estimating the relative species composition. The results of the hauls are then used to assess the stock indices by combining the acoustic signals with the relative species composition.
Discussions during WG BIFS (ICES 2004a) and WKSDA (2004b) have shown that general algorithm for combining the results of control hauls is not presented until now for estimating unbiased stock indices based on acoustic surveys of scattered layers.
Based on the acoustic surveys in the Baltic Sea in April/May and October detailed description of the problem is given. The area under investigation is covered by parallel transects. In some areas (ICES subdivision 22, 23) transects follow the depth structure because the area which can be used by the research cutter is restricted by depth or shipping routs. Rectangles of 30'N x 1°E are used as strata. Furthermore, studies have shown that the distribution function of areas cross section can be described by lognormally distribution in the most cases.
During the acoustic surveys in the Baltic Sea the detected species can be divided into two types, the **main species** which dominate the total fish density and for which indices are estimated. On the other hand "**noise species**" occur which are characterized by low densities in relation to the main target species and for which unbiased estimations of stock indices are not possible based on the acoustic measurements. However, the area scattering cross section of the "noise species" significantly influences the measures of total area cross section.
During the acoustic survey in April/May the total fish density is dominated by the main target species sprat. Besides this species cod and other target "noise species" occur in the scattering layer. However, the densities of these species are significantly lower than the density of sprat (Oeberst 1985, 1986, 1987, ICES 2004a). During the acoustic survey in October two main target species exist, herring and sprat. Because the densities of cod, flatfish and other species are low and stock indices are not estimated these species can also be notated as "noise species".
It is required for the acoustic surveys in the Baltic Sea that at least two hauls are carried out in each rectangle for identifying the species composition. The hauls are realized when the total area cross section is high enough to get representative samples that means that the probability is closed to zero that hauls are realized in areas with low total fish density. Furthermore, the probability is the same that control haul is realized when the total density is larger than

required level. Because the catch per halve hour is not proportional to the fish density or the total area cross section only the relative species compositions can be used. Furthermore, it is assumed that the relative species composition in the haul and in the area is the same. The data of the different surveys have shown that the proportions of the different species can significantly differ within short distances.
The proportions of species of all hauls which are realized in a strata are combined by arithmetic mean independent of the total fish density. Because it could not be shown that the arithmetic mean produce unbiased estimates of the mean proportion of species weighted mean of the proportion of species with CPUE value of control hauls were discussed. However, studies have shown that this method does also not produce unbiased estimates.
This study bases on the hypothesis that the use of arithmetic mean for combining the proportion of the target species produces biased estimates. New method which can be used independent of possible relations between the target species is described.

Material and Methods

Following notations and basic equations were used.

i	index of species
j	index of acoustic ESDU
h	index of control haul
$S_a(i,j)$	area scattering cross section of species i at the j th acoustic ESDU
$\sigma(i)$	mean cross section of species i
TS(i)	target-strength of species i
F(i,j)	fish density of species i at the j th acoustic ESDU
G(i,h)	CPUE of species i at the h th control haul
P(i,j)	Proportion of species i at j th acoustic ESDU
E[F(i,j)]	Mean value of the fish density
V[F(i,j)]	Variance of the fish density
$Cov(S_a(i,j),S_a(k,j))$	Covariance of area cross section of species i and k

The relation between the fish density and the area cross section is given by

$$F(i,j) = S_a(i,j) \cdot c(i) = \frac{S_a(i,j)}{< \sigma(i) >} \tag{1}$$

The total area cross section of j th acoustic ESDU is the sum of the area cross section of all species which are recorded by the transducer signal (M2).

$$S_a(j) = \sum_i S_a(i,j) \tag{2}$$

The mean and the variance of $S_a(j)$ are given by

$$E[S_a(j)] = E\ [\sum_i S_a(i,j)] = \sum_j \sigma_i E[F(i,j)] \tag{3}$$

and

$$V[S_a(j)] = V[\sum_i S_a(i,j)] = \sum_j\ V[S_a(i,j)] + 2\sum_{i<m}\ Cov[S_a(i,j), S_a(m,j)] \tag{4}$$

$$V[S_a(j)] = \sum_j \sigma_i^2 V[F(i,j)] + 2\sum_{i<m} \sigma_i \sigma_m Cov[F(i,j),F(m,j)] \tag{5}$$

The acoustic measurements provide estimates of $E[S_a(j)]$, $V[S_a(j)]$ and the distribution function of $S_a(j)$. When the different species can not be identified by the acoustic equipment additional data are necessary to estimate the stock indices. The aim of the control hauls is to provide data for estimating $E[S_a(i,j)]$, $V[S_a(i,j)]$, $Cov[S_a(i,j),S_a(k,j)]$ and the mean cross section of the different species, σ(i). The trawl is positioned in the depth of scattered layer (**Fig. 1**). Dependent on the size of the scattered layer it is possible that only a part or the total scattered layer is covered. It is required for the control hauls that the species compositions in the control hauls and in the scattered layer are comparable (Equ**. 6**). However, it is **not required** that the fish density of the total scattered layer, F(h), is proportional to the CPUE value, G(h), of the control haul.

$$p(i,h) = \frac{F(i,h)}{\sum_i F(i,h)} = \frac{G(i,h)}{\sum_i G(i,h)} \tag{6}$$

The positions of the control hauls are randomly selected with following restrictions:

- The water must be deep enough to realize a haul without any problems for the cutter,
- The distance between neighbouring hauls should not be to small and
- The fish density must be high enough to get a representative sample.

That means that control hauls are normally not carried out in areas with fish density lower than a required level ($F(j) > F_o$) or where the area scattering cross section is lower than a required level ($S_a(j) > S_{ao}$) **(Fig. 2)**

The analyses are carried out using reduced model. It is assumed that only two species with different mean cross section occur in the area and that the variability of the mean cross section of both species is very small. The distribution pattern of both target species can be quite different and it is possible the fish densities of both target species are correlated or the densities are independent. In the cases where the fish densities of both species are quite different the species with low density represents "noise species".

The demand concerning the low variance of the cross section of species can be fulfilled by subdividing species in different length or age groups, e.g. herring age group 0 and age group 1+ during the acoustic surveys in the Baltic Sea in October. The restriction concerning the number of species is used for better illustration of the results.

Analyses were carried out using the different distribution pattern of fish density. The results can be easy adapted to the area scattering cross section due to the assumption of very low variance of mean cross section of the target species. Two dimensional distribution functions for the different situations and simulated data were used for the study. For each case 3000 simulations of fish densities and a given number of control hauls were realized. Models of normal and lognormal distribution were used for simulating the fish densities. Two type of relations between the densities of both species were analysed

– the densities of both species are proportional to the total density and
– the density distributions of both species are independent.

Furthermore, the effect of different required levels of total density to realize a control haul were studied.

Results

Case 1: Fish density of species 1 is constant proportion of total fish density

It is assumed that the density of species 1, F(1,j), is proportional to the total fish density, F(j),

$$F(1,j) = (p + \varepsilon) * F(j) \tag{7}$$

with normally distributed $\varepsilon \in NV(0,s_p^2)$. For species 2 follows

$$F(2,j) = (1 - p - \varepsilon) * F(j). \tag{8}$$

The mean proportion of species 1 can be given by

$$E[\frac{F(1,j)}{F(j)}] = E[\frac{(p+\varepsilon)F(j)}{F(j)}] = E[p+\varepsilon] = E[p] = p \tag{9}$$

and the mean proportion of species 2 is

$$E[\frac{F(2,j)}{F(j)}] = 1 - p \tag{10}$$

The correlation coefficient between F(1,j) and F(j) is given by

$$r_{xy} = p\frac{\sqrt{V[F(1,j)]}}{\sqrt{V[F(j)]}} \tag{11}$$

Based on simulated data Figure 3 illustrates the situation. Besides the assumed lognormal distributed total fish density the mean proportion of species 1 (65%) is presented and the different probability of areas to be used for a control hauls is shown related to the total fish density as sigmoid curve.
Furthermore, follows that the correlation

$$R[F(1,j),F(2,j)] = R[(p_1+\varepsilon_j)F(j),(1-p_1-\varepsilon_j)F(j)] \tag{12}$$

$$= \frac{Cov[(p_1+\varepsilon_j)F(j),(1-p_1-\varepsilon_j)F(j)]}{\sqrt{V[(p_1-\varepsilon_j)F(j)]V[(1-p_1-\varepsilon_j)F(j)]}}$$

is 1 when s_j is zero. With increasing s_j the correlation coefficient between the densities of both species decreases. Based on 3000 simulated data sets the effect of increasing s_j was studied. It was assumed that total fish density is normally distributed with a mean of 241 and a standard deviation of 241. The proportion of species 1 was chosen with p = 0.7. σ_j increased from 0.01 to 0.5. Following correlation coefficients were estimated R[F(j),F(1,j)], R[F(j),F(2,j)], R[F(1,j),F(2,j)], R[p_1,F(j)], R[p_1,F(1,j)], R[p_1,F(2,j)]. The simulation have shown that R[F(j),F(1,j)], R[F(j),F(2,j)], R[F(1,j),F(2,j)] significantly decrease with increasing σ_j. Furthermore, R[F(j),F(2,j)], R[F(1,j),F(2,j)] increase (Tab. 1). Only R[p_1,F(j)] did not significantly changed and did not significantly differed from zero. Additional simulation which assumed that total fish density is lognormally distributed using the same values of mean and standard deviation (Tab. 2). These data have shown that the type of distribution function does not significantly influence the effect of changing σ_j in relation to the correlation coefficients. Simulation based on different means and standard deviations have shown that the general effect of changing σ_j is the same.

Conclusions

Because the proportions of both species are independent of the total density follows that the restrictions concerning the realization of control hauls does not significantly influence the estimation of the proportion of both species in the strata when the arithmetic mean of the control hauls is used.

Case 2: Densities of both species are not correlated

It is assumed that the densities of both species are independent. Two extreme situations are possible. The parameters of the density function of both species are comparable or density of one species dominates the total fish density and the other species represents a "noise species".
The probability distribution of the proportion of species 1 in relation to the total density was estimated using different types of distribution functions, **H_1 for species 1 and H_2** for species 2, and the different values of mean and standard deviation.
The mean and standard deviation of total density can be given by
E[F(j)] = E[F(1,j)] + E[F(2,j)]
V[F(j)] = V[F(1,j)] + V[F(2,j)].
E[F(j)] and V[F(j)] were used to calculate the probability distribution of F(j). The probability that density of species 1 is within the interval $[a_1*F(j),a_2*F(j)]$ for given F(j) is dependent on the distribution function of species 2 and can be calculated by

$$P[a_1*F(j) \le Spec1 < a_2*F(j), (1-a_2)*F(j) \le Spec2 < (1-a_1)*F(j)] \quad (13)$$

$$= \mathbf{H}_1(a_2*F(j))*\mathbf{H_2}((1-a_1)*F(j)) - \mathbf{H_1}(a_1*F(j))*\mathbf{H_2}((1-a_1)*F(j)) - \mathbf{H_1}(a_2*F(j))*\mathbf{H_2}((1-a_2)*F(j)) + \mathbf{H_1}(a_1*F(j))*\mathbf{H_2}((1-a_2)*F(j))$$

where a_1 and a_2 with $a_1 < a_2$ describes the interval of the proportion of total density.
In the first step it was assumed that the densities of both species are normally distributed. The studies have shown that the correlation between the total density and the proportion of species 1 is zero when means and standard deviations of both species are equal. Figure 4 shows the probability distribution of the proportion of species 1 for different total densities for E[F(1,j)]=E[F(2,j)]=6 and V[F(1,j)]=V(F[2,j)]=1.
However, when the variances of the densities of both species are different and the means are equal the correlation between the total density and the proportion of species 1 is significantly different from zero. Figure 5 and 6 present the probability distribution for cases where the variance of the density of species 1 increased.
Increase of the mean fish density of species 1 in relation to the fish density of species 2 where the standard deviations are constant due not results in such non linear changes of the probability distribution. The studies have shown that the distributions of the proportion of both species are correlated with total density when the variances of the density are different. That means that the expected proportion of species1 increases with increasing total density when the variance of species 1 is larger then the variance of species 2. Taking into account that the control hauls are not realized in areas where the total fish density is to low (e.g. F_o=10) follows that the proportion of species 1 based on data of control haul in combination with the method of arithmetic mean is overestimated.
Simulated data with the same parameter for means and standard deviations show comparable results and for the case E[F(1,j)]=E[F(2,j)]=6 and V[F(1,j)]=4^2, V(F[2,j)]=1. The XY plots of the relation between the proportion of species 1, P, and total fish density, Total, are presented in Figure 7**.** The relation between total density and proportion of species 1 can by described by the regression model P = a + b * ln(Total) when V[F(1,j)]=4^2, V(F[2,j)]=1..

When it is assumed that the densities of both species are lognormally distributed the change of the relation between total density and the proportion of species 1 is also influenced by the density parameters of both species. Following cases were studied:

$E[F(1,j)]=10$, $V[F(1,j)]=8^2$ $E[F(2,j)]=10$ and, $V(F[2,j)]=8^2$
$E[F(1,j)]=10$, $V[F(1,j)]=8^2$ $E[F(2,j)]=10$ and, $V(F[2,j)]=4^2$
$E[F(1,j)]=10$, $V[F(1,j)]=8^2$ $E[F(2,j)]=5$ and, $V(F[2,j)]=4^2$
$E[F(1,j)]=10$, $V[F(1,j)]=16^2$ $E[F(2,j)]=5$ and, $V(F[2,j)]=4^2$
$E[F(1,j)]=40$, $V[F(1,j)]=8^2$ $E[F(2,j)]=10$ and, $V(F[2,j)]=8^2$

The probability distributions have shown that the proportion of species 1 is correlated with total density in cases the means and/or the variances of both species are different (Fig. 8 to Fig. 11). Analyses based on simulations produced comparable relations. Taking into account that the control hauls are not realized in areas where the total fish density is to low follows that the proportion of species 1 based on data of control haul in combination with the method of arithmetic mean is overestimated.

Using simulated data where the density of both species are lognormally distributed with following parameters $E[F(1,j)]=10$, $V[F(1,j)]=16^2$, $E[F(2,j)]=5$ and, $V(F[2,j)]=4^2$ the mean proportion of species 1 is estimated when control hauls are realized in areas where the total density is larger than a given level. The range of total density is 1.3 – 234.6, the mean was 17.79 and the standard deviation was 17.25. The level of total density where hauls were realized increase from 0 to 20. The mean proportion of species 1 was 55.9% when restrictions concerning the realization of hauls did not exist (Tab. 3). The mean proportion of species 1 increased dependent on the required level. Required level concerning total density of 20 results in mean proportion of species 1 of 79.3%. This biased estimation of the proportion of species 1 produces overestimation of the fish density of species 1.

The situation of case 2 is described by Figure 12 in more general form. Lognormally distributed total fish density and the relation between the proportion of species 1 and total fish density are simulated. Furthermore the probability that area with given total fish density can be used for control haul is shown.

Conclusion

The presented studies have shown that the proportion of species is dependent on the total density when

- the density of both species are not correlated and
- the means and / or the standard deviations of both species are different.

The correlation between the total density and the proportion of species 1 increases with increasing difference of the distribution parameters. The restriction concerning the realization of hauls in areas where the total density is larger than a required level results in an overestimation of the proportion of species 1. Since the estimated proportion of species 1 based on the control hauls is not related to the total density follows that the arithmetic mean of the proportions of all hauls realized in strata can produce biased estimates.

The studies have shown that it is necessary to analyse the relations between the densities of the species because the methods which must be used for combining the control hauls is dependent on the relation between the species. Furthermore, it must be taken into account that the CPUE values of the control hauls are not correlated with the total fish density of the scattered layer.

Therefore, following procedure is necessary to use the results of the control hauls.

1. Estimation of the absolute area cross section or fish density of species in areas of the control hauls

The area cross section or the absolute fish density by species in the area of control hauls can be estimated by estimation of the mean cross section of the haul σ_h. by

$$\sigma_h = \sum_i \sigma_h(i) * F(i,h) / \sum_i F(i,h) \tag{14}$$

The total fish density in the area of the control haul can be estimated by equation (1) using the mean cross section of the haul σ_h and the absolute fish densities of the species can be estimated by

$$F(i,j) = \frac{F(i,h)}{\sum_i F(i,h)} F(j) . \tag{15}$$

The area cross section of the species in the area of the control hauls can be estimated by equation (1).

2. <u>Analyses of the relation between the fish density or area cross section of the species</u>

Using the estimates of the absolute fish density or area cross section of the species in the areas of the control hauls the relations between the densities of the species or between the area cross sections of the species can be studied by estimating the correlations between the proportion of the species and the total fish density (R[p,F(j)]). The experiences from the acoustic surveys in the Baltic Sea have shown that "noise species" exist.
All species which are observed by the acoustic measurements and the control hauls can be subdivided in two groups, the "noise species" for which stock indices are not estimated and the "target species". When more than one "noise species" occur the acoustic signals of all "noise species" can be combined and handled as one target. That means that the mean $S_a(i,j)$ value of all noise species are estimated for each control haul.

3. <u>Estimation of the means and standard deviations of the area cross section or fish density by species</u>

When the analyses of the relations between the proportion of the species and the total fish density show that species 2 is a "noise species" the mean and the variance of the area cross section of the "noise species" are estimated based on the data of control hauls.

$$E[S_a(2,j)] = \frac{\sum_{h=1}^{H} S_a(2,j)}{H} \text{ and} \tag{16}$$

$$V[S_a(2,j)] = \frac{1}{H-1}\{\sum_{h=1}^{H} S^2{}_a(2,j) - H * E^2[S_a(2,j)]\} \tag{17}$$

where H denotes the number of control hauls. For the mean and variance of the area cross section of species 1 follows

$$E[S_a(1,j)] = E[S_a(j)] - E[S_a(2,j)] \tag{18}$$

and

$$V[S_a(1,j)] = V[S_a(j)] - V[S_a(2,j)] \tag{19}$$

because the fish densities of both species are independent.

When the densities of both species are correlated the proportion of species 1, p, and the area cross section of the species can be estimated.

$$E[S_a(1, j)] = p * E[S_a(j)] \text{ and} \quad (20)$$

$$E[S_a(2, j)] = (1 - p) * E[S_a(j)] \quad (21)$$

The total mean fish density can be estimated by Equation 1. Using bootstrap methods the variability of the mean area cross section and the total mean fish density can be studied.
In cases where more than two species are detected by the acoustic measurements and the control hauls the above described procedure can also be used. Based on the absolute area cross sections of the species in areas of the control hauls the relations between the densities of the species can be studied. Then mean total area cross section of the "noise species" is subtracted form the mean total area cross section. In a next step the mean proportions of the mean area cross section of the reminding species are estimated. Using bootstrap methods the variability of the estimated means can be analysed.

References

ICES. 2004a. Report of the Baltic International Fish Survey Working group (WGBIFS). ICES CM 2004/G:08, Ref. D, H. 162 pp.
ICES. 2004b. Report of the Workshop on Survey Design and Data Analysis (WKSDA), ICES CM 2004/B:07, Ref. D, G. 361 pp.
Oeberst, R. 1985. Zu einigen Aspekten der Planung von hydroakustischen Bestandsbestimmungen. Fischerei - Forschung Rostock 23 (4) : 77 – 88.
Oeberst, R. 1986. Some aspects of planning of acoustic stock estimations. ICES, C. M. 1986/B:21, Fish Capture Committee, 10pp.
Oeberst, R. 1987. Ein Ansatz zur Verbesserung von hydroakustischen Biomasseschätzungen bei gleichzeitigem Auftreten von mehreren Zielklassen. Fischerei - Forschung Rostock 25 (3) : 84 – 89.
Parrish, J.K. 2004. Invited lecture on Acoustic in the New Century: Behaviour, Ecology and Ecosystem Science. ICES Annual Science Conference, Vigo, Spain in 2004.
MacLennan, D. N. and Simmonds, E. J.; Fisheries acoustic, Chapman & Hall, London, 1992

Tables

Table 1: Correlations between total fish density and the fish densities of species 1 and 2 dependent on s_j assuming that total fish density is normally distributed with mean of 241 and standard deviation of 311 and constant proportion of species 1 of p=0.7 (3000 simulated data sets)

σ_j	R[F(j),F(1,j)]	R[F(j),F(2,j)]	R[F(1,j),F(2,j)]	R[p,F(j)]	R[p,F(1,j)]	R[p,F(2,j)]
0.01	0.9980	0.9980	0.9976	0.01	0.02	-0.03
0.25	0.9986	0.9926	0.9848	-0.01	0.03	-0.10
0.05	0.9944	0.9702	0.9391	0.02	0.10	-0.16
0.1	0.9789	0.9016	0.7943	-0.02	0.13	-0.35
0.25	0.8982	0.6769	0.2844	0.02	0.34	-0.53
0.5	0.7920	0.5683	-0.0522	0.02	0.47	-0.59

Table 2: Correlations between total fish density and the fish densities of species 1 and 2 dependent on s_j assuming that total fish density is lognormally distributed with mean of 241 and standard deviation of 311 and constant proportion of species 1 of p=0.7

σ_j	R[F(j),F(1,j)]	R[F(j),F(2,j)]	R[F(1,j),F(2,j)]	R[p,F(j)]	R[p,F(1,j)]	R[p,F(2,j)]
0.01	0.9999	0.9993	0.9985	0.01	0.02	-0.2
0.25	0.9990	0.9946	0.9890	0.01	0.04	-0.6
0.05	0.9965	0.9813	0.9619	0.01	0.06	-0.11
0.1	0.9873	0.9423	0.8772	-0.01	0.09	-0.22
0.25	0.9376	0.7670	0.4959	-0.0	0.22	-0.40
0.5	0.8376	0.6054	0.0722	0.01	0.35	-0.50

Table 3: Estimated mean proportion of species 1 when control hauls are realized where the total density is larger then the required level (densities of both species are lognormally distributed with E[F(1,j)]=10, V[F(1,j)]=16^2, E[F(2,j)]=5 and, V(F[2,j)]=4^2 , **Fig. 11,** distribution parameter of Total density; range: 1.3 – 234.6, mean: 17.79, standard deviation: 17.25)

Required level of total density	Number of data which fulfil the requirement	Mean proportion of species 1
0	1000	55.9
5	863	57.9
10	508	63.8
15	300	69.8
20	175	79.3

Figures

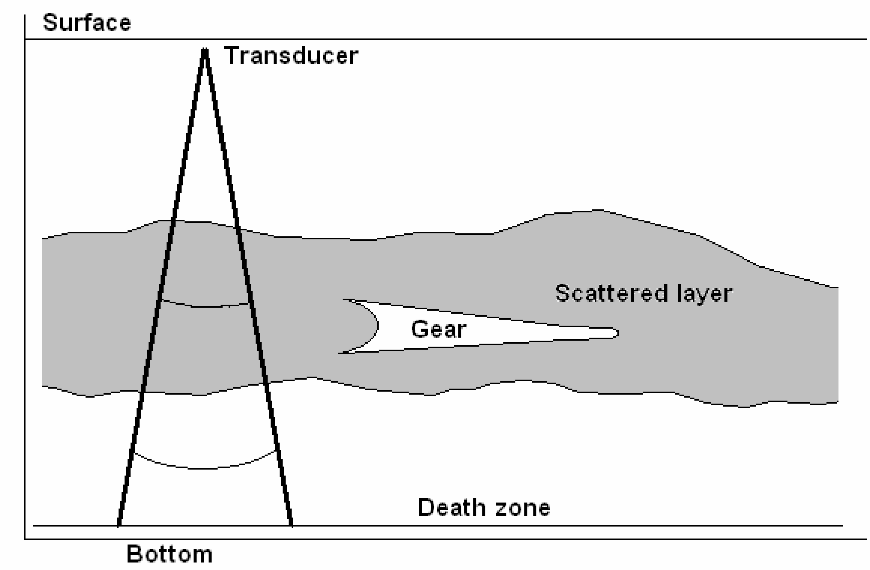


Fig. 1 (Figure1.BMP): Description of the survey situation (Using a transducer the total area scattering cross section is measured, the species composition is estimated by control hauls)

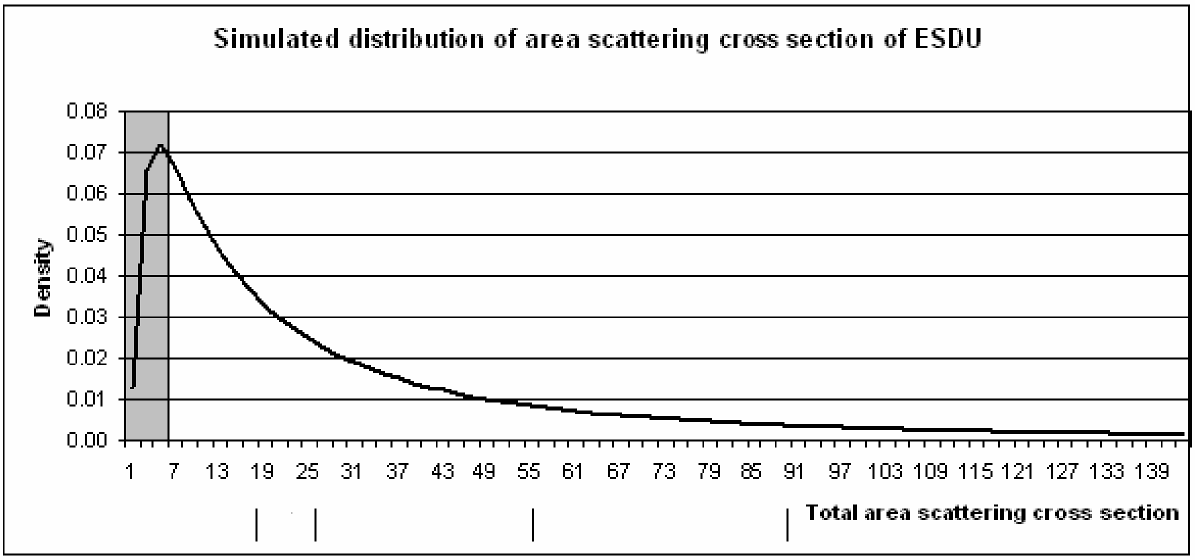


Fig. 2 (Figures_Sim01.BMP): Position of control hauls related to the distribution of total area scattering cross section of strata

Total area scattering cross section where control hauls are not realized based on the low density

Position of control hauls

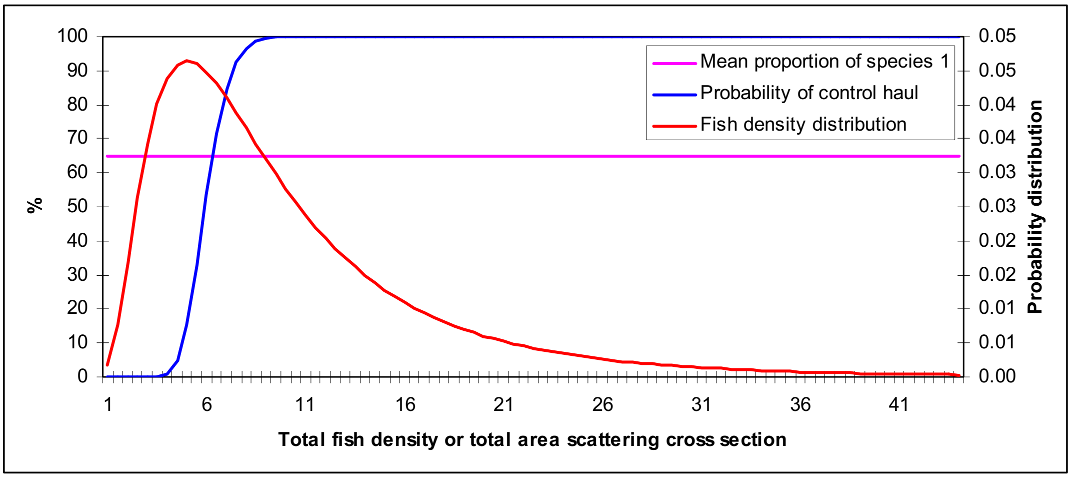


Figure 3: Simulated lognormally total fish density and constant proportion of species 1 of 65% independent of total density. Probability of areas to be used for control hauls in relation to total fish density

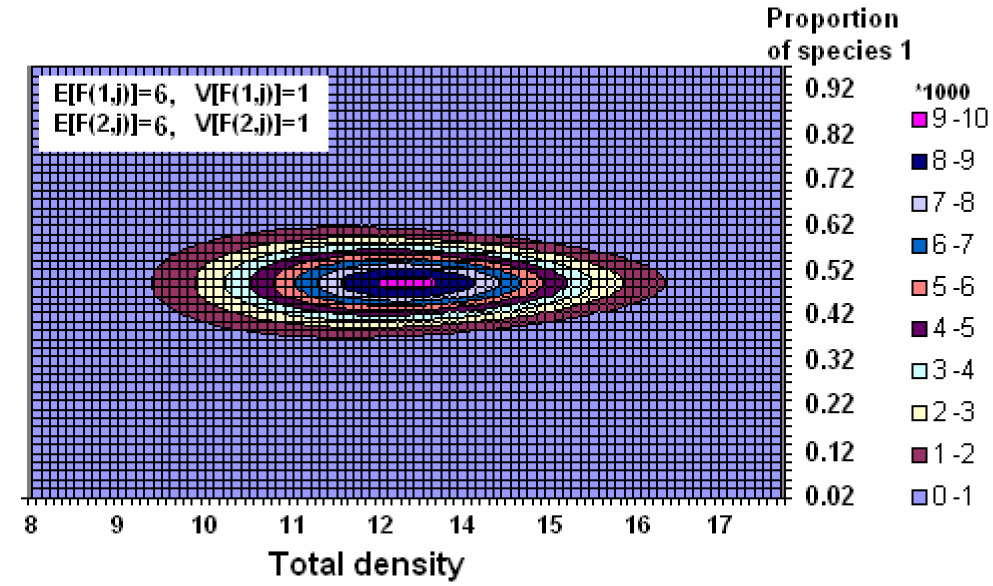


Figure 4: Probability distribution of the proportion of species 1 related to the total density when the density of both species are normally distributed and the means and standard deviations are equal

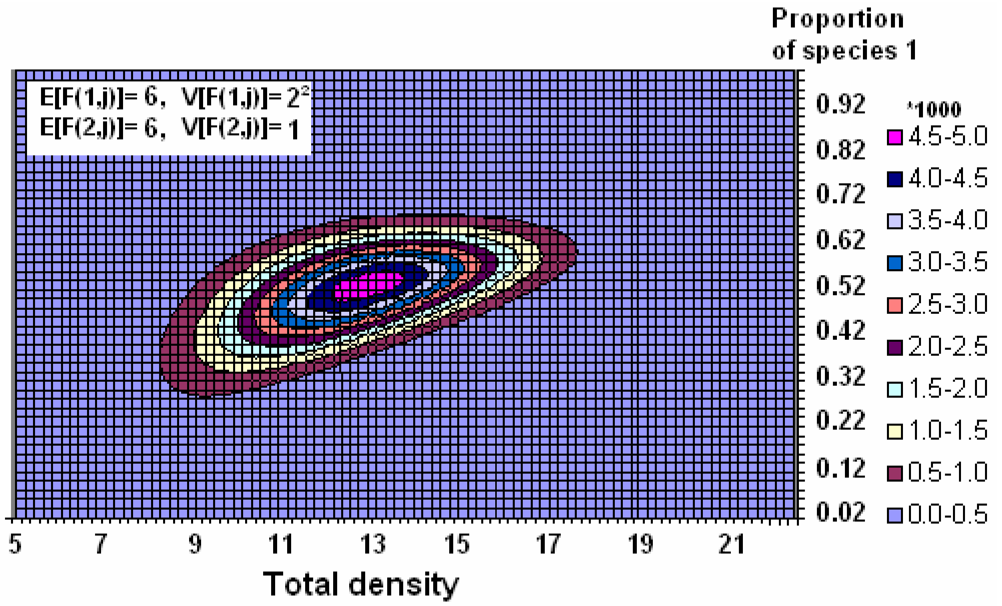


Figure 5: Probability distribution of the proportion of species 1 related to the total density when the density of both species are normally distributed and the means and standard deviation of both species are different

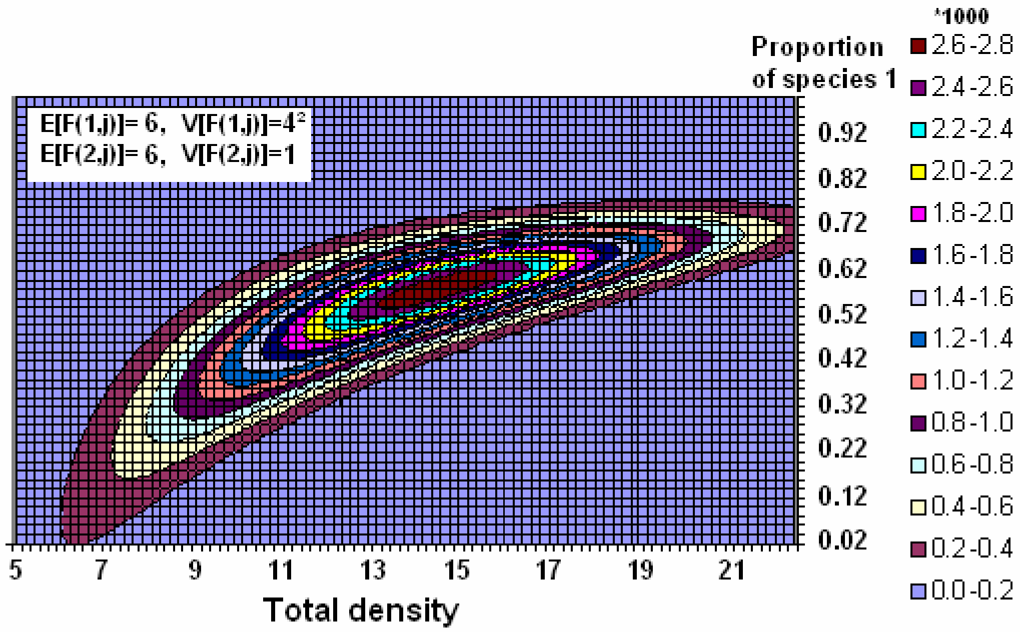
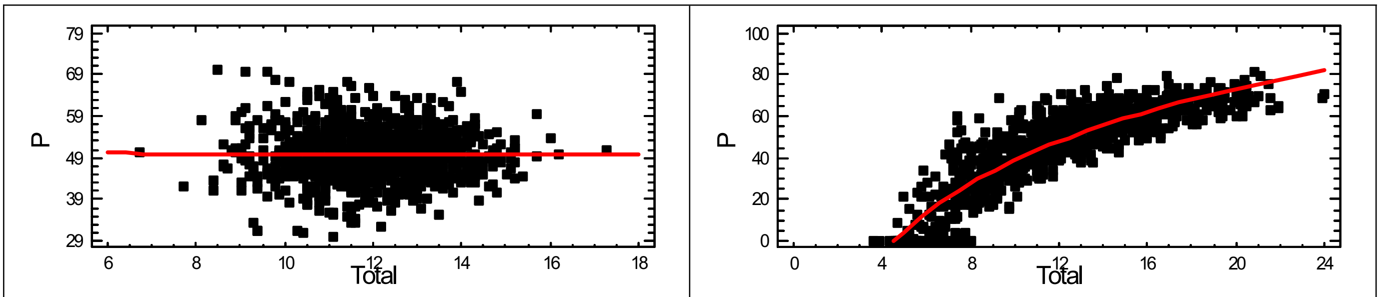


Figure 6: Probability distribution of the proportion of species 1 related to the total density when the density of both species are normally distributed and the means and standard deviation of both species are different



$E[F(1,j)]=E[F(2,j)]=6$, $V[F(1,j)]=V[F(2,j)]=1$ | $E[F(1,j)]=E[F(2,j)]=6$, $V[F(1,j)]=4^2$, $V[F(2,j)]=1$

Figure 7: Distribution of the proportion of species 1, P, in relation to the total density, Total, when total density is normally distributed

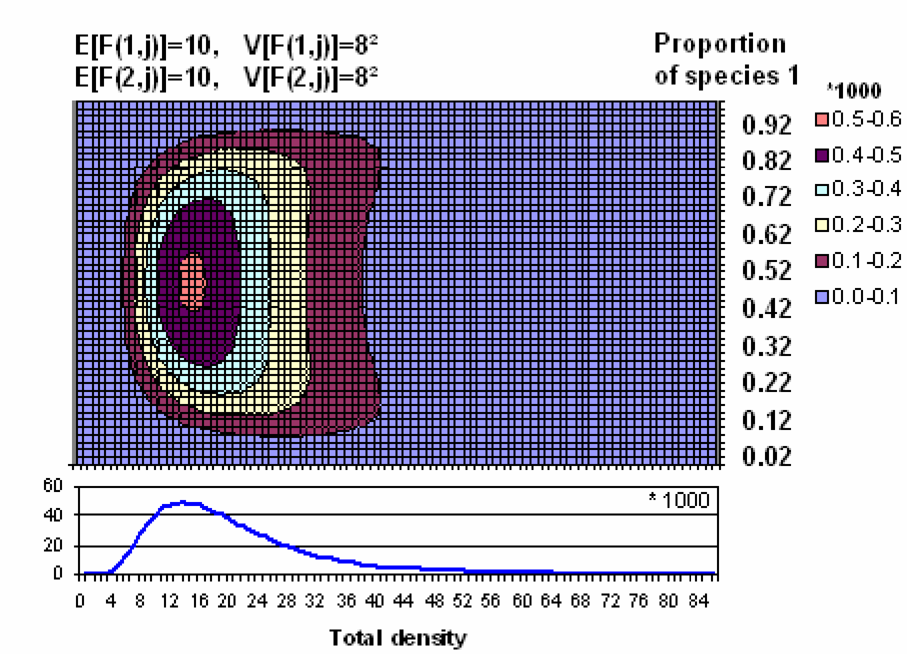


Figure 8: Probability distribution of the proportion of species 1 related to the total density when the density of both species are lognormally distributed and the means and standard deviations are equal

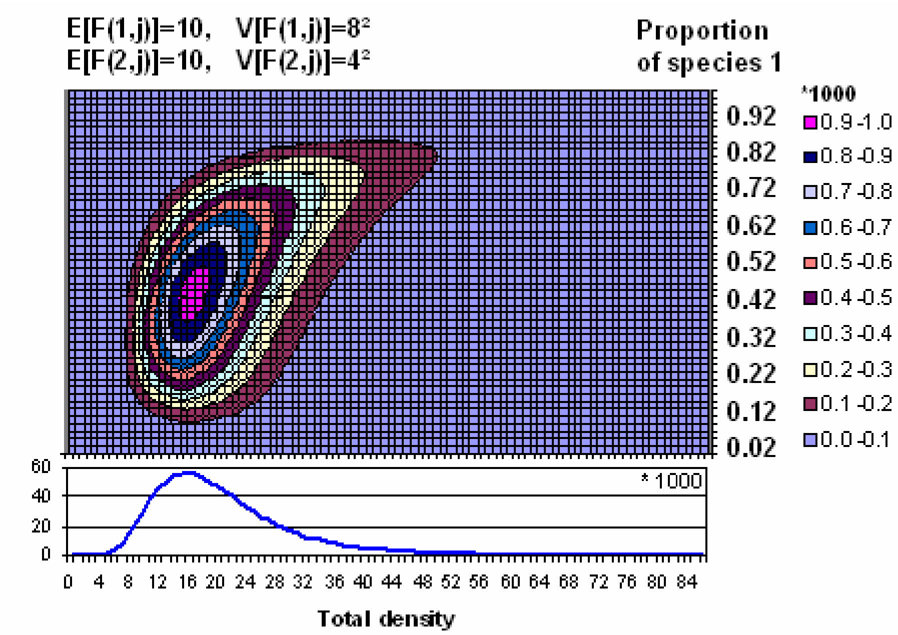


Figure 9: Probability distribution of the proportion of species 1 related to the total density when the density of both species are lognormally distributed and the means and standard deviation of both species are different

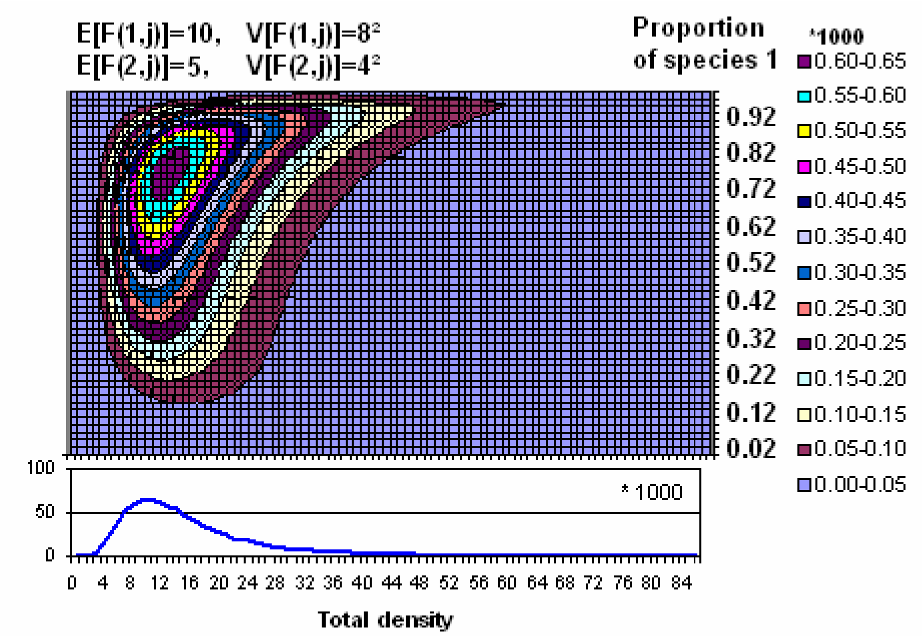


Figure 10: Probability distribution of the proportion of species 1 related to the total density when the density of both species are lognormally distributed and the means and standard deviation of both species are different

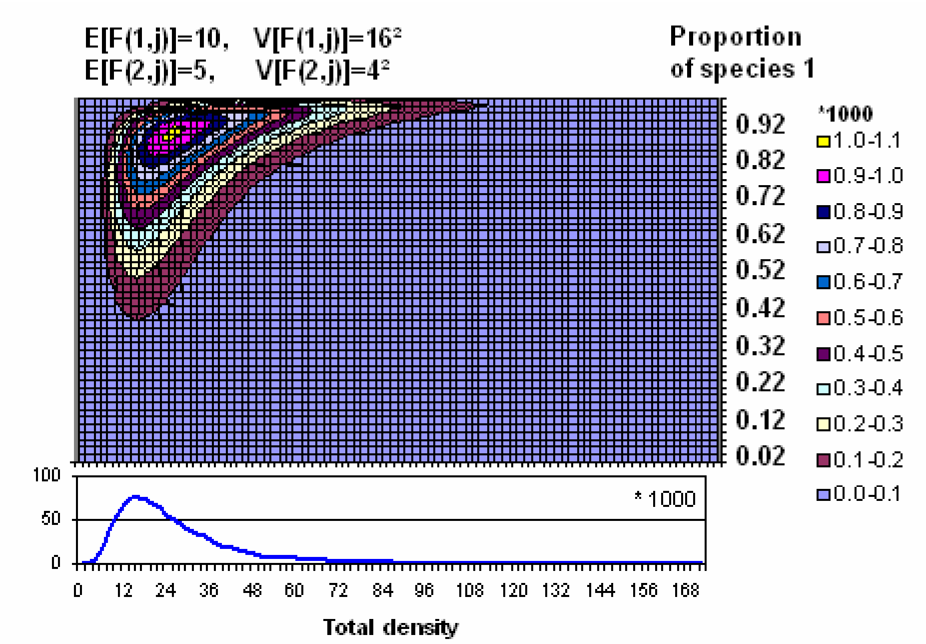


Figure 11: Probability distribution of the proportion of species 1 related to the total density when the density of both species are lognormally distributed and the means and standard deviation of both species are different

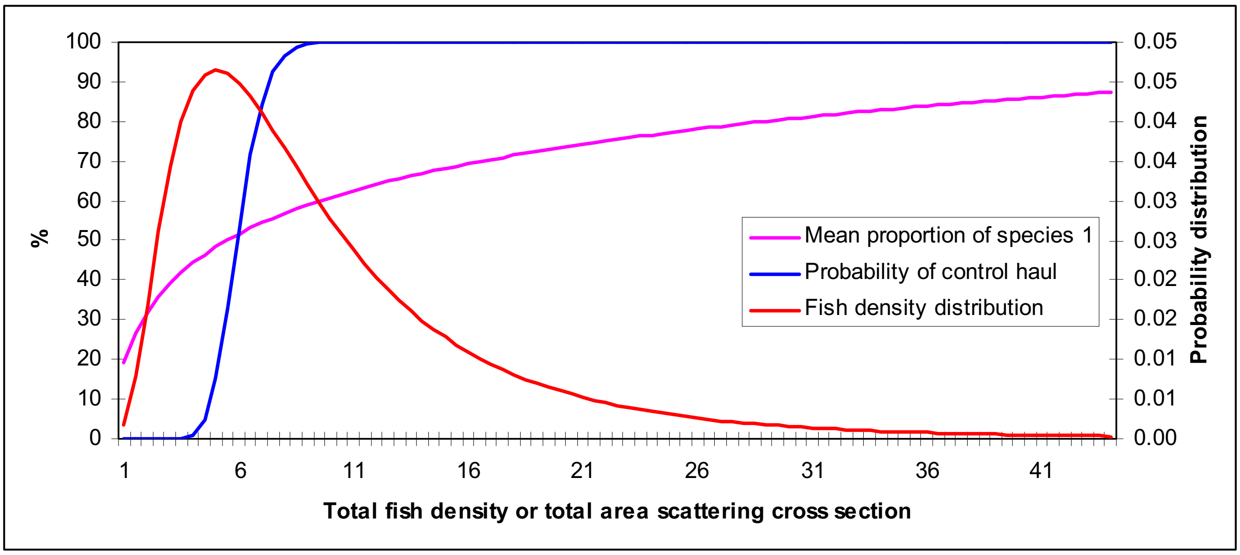


Figure 12: Simulated lognormally total fish density and correlated proportion of species 1 and the probability of areas to be used for control hauls in relation to total fish density