International Council for
The Exploration of the Sea

CM 2001/P:06

The measurement error of marine survey catches: the bottom trawl case.

V. Hjellvik, O. R. Godø and D. Tjøstheim

Abstract

We have formulated a model for analyzing the measurement error of marine survey abundance estimates using data from parallel surveys (trawl haul or acoustic measurement). The measurement error is defined as the component of the variability that cannot be explained by covariates such as temperature, depth, bottom type etc. The method presented is general, but we concentrate on bottom trawl catches of cod (*Gadus morhua*). Catches of cod from 10 parallel trawling experiments in the Barents Sea with a total of 130 paired hauls were used to estimate the measurement error of trawl hauls. Based on the experimental data, the measurement error is fairly constant in size on the logarithmic scale and is independent of location, time and fish density. Compared with the total variability of the winter- and autumn surveys in the Barents Sea, the measurement error is small (approximately 2-5%, on the log scale, in terms of variance of catch per towed distance). Thus, the cod catch rate is a fairly precise measure of fish density at a given site at a given time.

Keywords: cod; measurement error; parallel trawling; trawl surveys.

Vidar Hjellvik: Institute of Marine Research, P.O. Box 1870, Nordnes, N-5817 Bergen, Norway [tel: +47 55 23 86 62, fax: +47 55 23 86 87, email: vidarh@imr.no].
Olav Rune Godø: Institute of Marine Research, P.O. Box 1870, Nordnes, N-5817 Bergen, Norway [tel: +47 55 23 86 75, fax: +47 55 23 86 87, email: olavrune@imr.no].
Dag Tjøstheim: Department of Mathematics, University of Bergen, Johs. Brunsgt 12, N-5008 Bergen, Norway [tel: +47 55 58 28 85, fax: +47 55 58 96 72, email: dagt@mi.uib.no].

1 Introduction

Surveys are vital for estimating the size and composition of marine populations. employed. It is well known that the resulting estimates are subject to substantial variations, and it is important to quantify and explain, as much as is possible of the variability in terms of relevant explanatory variables or covariates. Typically these variables will depend on the sampling tool used, but for bottom trawl catches important explanatory variables can be depth and location of the haul (Polacheck and Vølstad 1993), time of the day (Korsbrekke and Nakken 1999), season, strength of the year classes involved etc.

Generally, as the number of covariates increases, and the model becomes more complex, the residual variation (or remaining uncertainty) not explained by the model decreases. But no matter how refined the model is, there will always be an unexplained random component which cannot be attributed to any observed variable. This residual variation is caused by the interactions between the fish, the measurement device and the environment (see e.g. Engås, 1994).

The purpose of this paper is to define and quantify this residual source of random variation. This is done through analyzing measurements from parallel tows by multiple vessels, which is the closest one can come to a controlled statistical experiment in this context. The importance of quantifying this type of random fluctuation lies in the fact that it is a benchmark uncertainty which is inherent in the survey process itself, and in this sense it may be termed a measurement error. If the measurement error can be assessed from field data and is consistent over time and space, we improve our understanding and quantification of other causal factors behind the uncertainty associated with survey estimates.

In this paper we look at bottom trawl catches of cod, but we would like to stress that the concepts and techniques developed can in principle be applied to acoustic survey estimates or indeed to any type of measurements collected simultaneously by two independent parallel sampling devices. It should be noted that there is a growing related literature on comparative survey analysis. We refer to the review paper by Pelletier (1998) and references therein.

2 Materials and methods

2.1 Parallel trawling experiments

During the annual combined bottom-trawl and acoustic survey of demersal fish in the Barents Sea during winter and autumn conducted by the Institute of Marine Research, Bergen (IMR), (cf. Jakobsen et al. 1997[1]), parallel trawling experiments were used to compare the efficiency of the participating vessels. During a parallel haul the vessels operate about 500 meters apart and use radio contact to assure proper coordination during hauls.

We analyse ten parallel trawl experiments performed by the IMR during the last decade (Table 1). The data from 1991 are described in Michalsen et al. (1996). Two hauls with unstable bottom contact are excluded from the 1991 data (cf. Michalsen et al. 1996). Similarly, two hauls from 1995 are excluded; one where trawl geometry measurements indicate problems with the doors and another with highly different recorded towed distances (0.7 and 2.2 nautical miles). Let d_{ij} denote the towed distance for haul i and vessel j; $i = 1, \ldots, n$; $j = 1, 2$, where n is the total number of hauls. The average recoded distance is $\overline{d} = (2n)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{2} d_{ij} = 1.33$ nautical mile (average duration in time is 27 minutes), with $0.8 \leq d_{i;j} \leq 1.8$ in 98% of the cases. The absolute values $|d_{i;1} - d_{i;2}|$ of the differences in towed distance for the two vessels in the same haul, are 0, 0.1, 0.2, 0.3, 0.4 and 0.6 in 43, 45, 31, 6, 4 and 1 cases, respectively.

group	year	date	vessels	lat	lon	n	w
1	1991	3-5 Mar.	LIZY, JH	71.3	26.2	10	2.01
2	1994	10 Feb.	LIZY, JH	71.2	36.0	5	0.81
3	1994	22-23 Feb.	LIZY, GS	71.3	26.3	8	1.98
4	1995	22-23 Feb.	JM, JH	71.3	25.4	12	1.05
5	1995	23-25 Feb.	JH, GS	71.3	25.4	23	0.94
6	1995	15-17 Aug.	MS, JH	74.3	17.3	29	0.72
7	1996	17 Feb.	JM, GS	70.4	36.5	4	0.03
8	1996	24-25 Feb.	JM, GS	71.8	23.8	10	0.19
9	1997	8-10 Feb.	JH, GS	71.3	27	17	0.72
10	1997	2-3 Aug.	MS, JH	72-73	27-30	12	0.21

Table 1 *Summary statistics for the parallel trawling experiments; time, vessel, position (lat, lon), the number of hauls n and the average per cod weight w in kg within each group. The vessels involved are Johan Hjort (JH), Anny Kræmer (LIZY), G. O. Sars (GS), Jan Mayen (JM) and Michael Sars (MS).*

The data have been subdivided into 10 groups so that all hauls within a group are done by the same two vessels, within a period of one to three days, and usually in a small geographical area. Group 10 is an exception where the trawl stations are evenly spread over about 60 nautical miles both in the east-west and in the north-south directions.

---

[1] Jakobsen, T., Korsbrekke, K., Mehl, S., and Nakken, O. 1997. Norwegian Combined Acoustic and Bottom Trawl Surveys for Demersal Fish in the Barents Sea During Winter. ICES CM 1997/Y:17: 1-26.

2.2 The statistical model

Any study of uncertainty depends on the stochastic model adopted. Two different statistical models may yield quite different uncertainty estimates. A general model for a series of survey measurements $\{y_i,\ i = 1, \ldots, n\}$ is given by

$$y_i = f(x_{i1}, \ldots, x_{ip}) + \epsilon_i, \quad i = 1, \ldots, n,$$

where f is a deterministic function, which in principle is unknown, $x_{i1}, \ldots, x_{ip}$ are the ith measurements of p explanatory variables, such as geographical location, depth or time. If all of the relevant explanatory variables were included, ϵ_i would represent the residual uncertainty. In practice, all conceivable explanatory factors will not be observed, and often f is be assumed to be linear.

A difficulty in assessing the uncertainty of fish abundance estimates is that we cannot carry out a series of controlled experiments, where the setting of each experiment is identical. In such an idealized series of experiments the explanatory variables $x_{i1}, \ldots, x_{ip}$ would be fixed, and ϵ is the only source of random variation so that for a series of N experiments,

$$y_k = f(x_1, \ldots, x_p) + \epsilon_k, \quad k = 1, \ldots, N.$$

Since f does not depend on k, the standard error of ϵ_k can then be estimated directly from the observations $\{y_k\}$ as

$$\hat{\sigma}_\epsilon = \left\{ \frac{1}{N-1} \sum_{k=1}^{N} (y_k - \overline{y})^2 \right\}^{\frac{1}{2}}$$

and the correctness of the model could be tested by a new series of experiments for a new set of fixed values for the explanatory variables.

The closest we can come to such an idealized experiment is that of parallel trawling described above. The values of the explanatory variables, such as geographical location and depth, will vary somewhat from one vessel to another, but as an approximation they will be considered identical. We will allow for an additive individual vessel effect $\alpha_j,\ j = 1, 2$, though, permitting differences in equipment and efficiency for the two vessels. This leads to the following model for the observations $\{y_{i;j},\ i = 1, \ldots, n;\ j = 1, 2\}$, where $j = 1, 2$ correspond to the first and second vessel in Table 1:

$$\begin{aligned} y_{i;1} &= f(x_{i1}, \ldots, x_{ip}) + \alpha_1 + \epsilon_{i;1} \\ y_{i;2} &= f(x_{i1}, \ldots, x_{ip}) + \alpha_2 + \epsilon_{i;2}. \end{aligned} \tag{2.1}$$

All the factors affecting jointly tow performance are supposed to be in the function f, and therefore the residuals $\{\epsilon_{i;j},\ i = 1, \ldots, n;\ j = 1, 2\}$ are assumed to be independent zero-mean identically distributed random variables and $\sigma_\epsilon = \mathrm{sd}\,(\epsilon_{i;j})$ is the measurement error.

We can now eliminate f and the explanatory variables by taking differences, i.e.

$$z_i \overset{\mathrm{def}}{=} y_{i;1} - y_{i;2} = \alpha_1 - \alpha_2 + \epsilon_{i;1} - \epsilon_{i;2}.$$

The expected difference between the two vessels is then given by

$$E\,(z_i) = \alpha_1 - \alpha_2$$

and due to the independence of $\epsilon_{i;1}$ and $\epsilon_{i;2}$,

$$\sigma_z^2 = \mathrm{var}\,(z_i) = \mathrm{var}\,(\epsilon_{i;1} - \epsilon_{i;2}) = 2\sigma_\epsilon^2$$

and the standard error σ_ϵ can be estimated as

$$\hat{\sigma}_\epsilon = \frac{1}{\sqrt{2}}\hat{\sigma}_z = \frac{1}{\sqrt{2}}\left\{(n-1)^{-1}\sum_{i=1}^{n}(z_i - \overline{z})^2\right\}^{1/2}, \qquad (2.2)$$

whereas $\delta = \alpha_1 - \alpha_2$ is estimated by $\hat{\delta} = \overline{z}$.

It should be noted that in the actual computation of $\hat{\sigma}_\epsilon$ and $\hat{\delta}$ the data $y_{i;j}$ are log transformed, i.e.

$$y_{i;j} = \log(n_{i;j}/d_{i;j}), \quad i = 1, \ldots, n; \quad j = 1, 2, \qquad (2.3)$$

where $n_{i;j}$ and $d_{i;j}$ denote the catch in numbers and the towed distance, respectively, for vessel j at the i-th haul. Log transformed data are used to reduce the heterogeneity of the variance.

Tests for differences between experiment groups

If our hypothesis that parallel trawling experiments can be used to quantify a measurement error inherent in the cod catching process itself is correct, we expect that this error, as estimated by $\hat{\sigma}_\epsilon$, should be the same for all 10 experimental groups. If the z_is originate from a Gaussian distribution, the null hypothesis of equal variance, $\sigma_{\epsilon,k}^2$ the same, $k = 1, \ldots, 10$, for the 10 groups can be tested by Bartlett's test (all groups tested simultaneously, Bickel and Doksum 1977, p. 304) and if needed followed by a series of F-tests (the groups could be tested against each other in pairs).

The possible differences in efficiency caused by different vessels and/or fishing gears can be tested by an ANOVA test followed by a series of t-tests if the ANOVA test leads to rejection. Again, normally distributed observations is a prerequisite for such tests.

Our first task is therefore to check if the z_i-data follow a Gaussian distribution. It seems plausible to assume that observations from different groups follow the same distribution, but possibly with differences in mean and variance. Therefore, when checking for normality, we consider the standardized variables

$$x_i = \frac{z_i - \overline{z}_k}{s_k}$$

where $k = k(i)$, $k = 1, \ldots, 10$, denotes the group that haul i, $i = 1, \ldots, n$, belongs to, $\overline{z}_k$ and s_k are the average and the estimated standard deviation of the z-values in group k. Deviations from normality of $\{x_i\}$ can be checked visually by inspecting a normal plot, and formally by e.g. the Kolmogorov-Smirnov test (Bickel and Doksum 1977, ch. 9.6).

3 Results

The log-catches $\{y_{i;j}\}$ and the corresponding differences $\{z_i\}$ are presented in Fig. 1. The catches ranges from approximately $e^3 \approx 20$ to $e^8 \approx 3000$, but on the log-scale the difference in catch between the vessels does not seem to depend on the size of the catches (see formal test at the end of this section). A normal plot of the standardized observations $\{x_i = (z_i - \overline{z}_k)/s_k\}$ appears linear (cf. Fig. 2), and the Kolmogorov-Smirnov test does not reject the null hypothesis of normality at a 10% level. Testing each group separately (except group 7 where the sample size is too small) yields the same result, i.e. normality is not rejected at a 10% level for any group, thus justifying the use of Bartlett, F and t-tests. Bartletts test yields a p-value of 0.81. In view of this it is not really necessary to test the groups in pairs for equality in variance using an F-test, but as a check we have carried out the tests obtaining the lowest p-value of 0.078 for groups 4 and 5. Thus, based on these data, the hypothesis of a uniform measurement error independent of geographical location, time, depth etc., can not be rejected.

To investigate possible differences in efficiency for the participating vessels we did an ANOVA test. We found a p-value less than 10^{-7} indicating significant differences. This is consistent with earlier findings in calibration experiments (cf. Pelletier 1998). From a statistical point of view, the next natural step would be to carry out a multiple comparison to locate and order differences, but since our main objective is not to quantify differences between vessels, we satisfy ourselves with an informal analysis consisting of a series of t-tests. Using a 5% significance level, we found significant differences in efficiency between the participating vessels in group 2, 4 and 6 with $p-$values 0.035, 0.001 and 0.038, respectively. Thus, since $E(z_i)$ cannot be considered equal for all groups, a pooled variance could be used for estimating σ_ϵ^2. Alternatively, z_i in (2.2) could be replaced by the variables $z_i' = z_i - \overline{z}_k$ which are adjusted for group means and identical to the residuals from the ANOVA fit. The resulting estimates using the last approach are $\hat{\sigma}_\epsilon^2 = 0.069$ and $\hat{\sigma}_\epsilon = 0.263$. The bootstrapped standard errors of $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_\epsilon$ are 0.0077 and 0.0147,

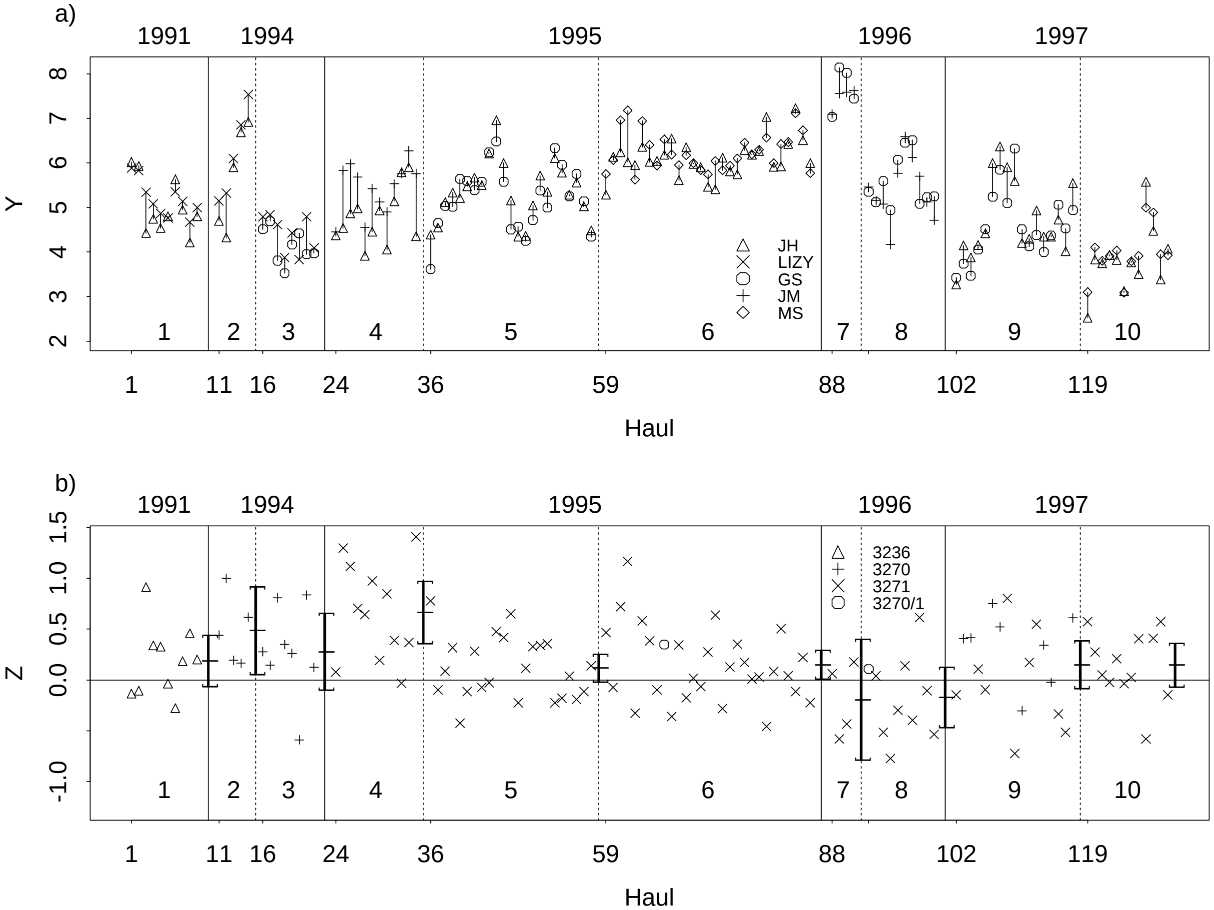

Figure 1 *a) Catches* $y_{i;j}$ *from the parallel trawling experiments in Table 1. The groups are separated by vertical lines and the group numbers corresponding to Table 1 are given at the bottom of the figure. The symbols represent vessels (cf. Table 1), and the cumulative number of hauls are given at the horizontal axis. b) Corresponding differences* $z_i = y_{i;1} - y_{i;2}$. *The symbols indicate gear type; 3236: Campelen 1800 shrimp trawl with 35 mm mesh size, 40 m sweeps and rockhopper gear; 3270: same as 3236, but 20 mm mesh size; 3271: same as 3270 but with strapping. 3270/1 indicate that one vessel uses 3270, the other 3271. To the right for each group 95% confidence intervals for the group means are given.*

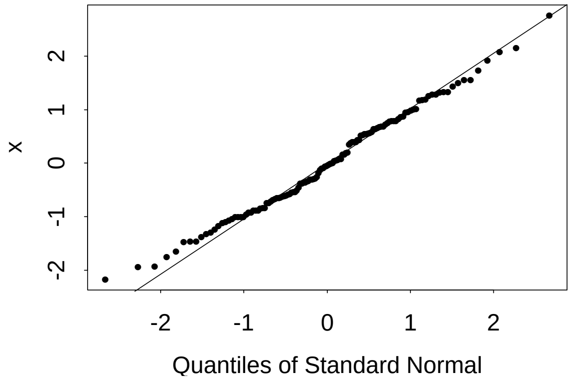

Figure 2 *Normal plot of* $\{x_i = (z_i - \overline{z}_k)/s_k\}$.

	kg	number				
	> 0 cm	> 0 cm	< 30 cm	30-59 cm	≥ 60 cm	≥ 30 cm
$\hat{\sigma}_\epsilon^2$	0.074	0.069	0.112	0.080	0.086	0.064
sd. error of $\hat{\sigma}_\epsilon^2$	0.0096	0.0077	0.0140	0.0114	0.0122	0.0083
n	130	130	113	117	103	118
groups excluded	none	none	2,3	10	2,7,10	10
$\overline{y}$	4.77	5.28	4.61	4.06	3.61	4.52

Table 2 *Estimates of σ_ϵ^2 with bootstrapped standard errors for catches measured in kg, total number of fish caught, and length stratified numbers. The number of hauls where both vessels caught at least 10 cod is given in the third row, the groups excluded due to less than 4 remaining hauls are given in the fourth row and the average catch of the remaining hauls in the fifth row. The standard errors are estimated using 1000 bootstrap replicates.*

respectively (1000 bootstrap replicates were used). Some caution should be excersized in interpreting these numbers (see e.g. Srivastava and Chan 1989).

Compared to the total variability of a survey, the measurement error of a single haul is relatively small. The last 5 years (1996-2000) $\text{var}\,(y_i)$ for the non-zero catches varies between 1.38 and 2.06 for the winter survey and between 2.53 and 3.92 for the autumn survey. Thus, σ_ϵ^2 is about 2-5 % of the total variation. This is the percentage of the variation that we cannot expect to be able to explain by explanatory variables. One should carefully note that these numbers are on the og scale. If antilogs of the catch rate were to be used, the additive model (2.1) would have to be replaced by a multiplicative model, and the relative magnitude of variances would be changed.

The results for length stratified data are shown in Table 2, as well as the results obtained by measuring the catches by weight instead of by numbers, i.e. by replacing $n_{i;j}$ in (2.3) by $m_{i;j}$, where $m_{i;j}$ is the weight of the catch in kilograms. Only hauls where both vessels got 10 specimen or more are included, and only groups with at least 4 such hauls. It should be noted that for fish less than 31 cm, group 1 has significantly higher variance than group 5 (p-value 0.025). For all other pairs, the null hypothesis of equal variance is accepted, again confirming the hypothesis of a uniform measurement error. It is seen that $\hat{\sigma}_\epsilon^2$ is highest for small fish, and this difference is statistically significant: Comparing the bootstrap replicates $\sigma_\epsilon^{2\star}$ of $\hat{\sigma}_\epsilon^2$ for fish smaller than 30 cm to those for fish larger than 30 cm by grouping them in pairs, we get the largest value of $\sigma_\epsilon^{2\star}$ for small fish in all of the 1000 bootstrap replicated pairs. This means that the bootstrapped p-value for the observed data under the null hypothesis of equal measurement error for the two length groups is less than 0.001 if the alternative hypothesis is that the measurement error is largest for small fish.

The hauls differ in towed distance, with modes at 1 and 1.5 nautical miles of the towed distance distribution, or alternatively 20 and 30 minutes tow duration. Stratifying on tow duration (which is recorded more exactly than towed distance), with hauls of less than 25 minutes duration (33 %) in group A and the remaining hauls (67 %) in group B, and estimating the measurement error for each group separately, we get $\sigma_\epsilon^2 = 0.0707$ and 0.0656 for group A and B, respectively. Thus, there is no significant difference due to tow duration. For fish less than 30 cm, the corresponding estimates are 0.115 and 0.107 for group A and B, respectively.

No significant relationship was found between the magnitude of the catches and their differences. A regression analysis was performed with the absolute value of the mean-adjusted differences in catch rate, $|z_i'|$, as the dependent variable and the average catch $\overline{y}_i = (y_{i;1} + y_{i;2})/2$ as the independent variable. The regression equation was $|z_i'| = 0.20 + 0.019\overline{y}_i$ and the p-value under the null hypothesis of no relationship was 0.31. However, the residuals from the analysis are skewed with a long right hand tail, so a bootstrap test was also done, resulting in an empirical p-value of 0.137, and again the null hypothesis of no relationship was not rejected at a 5% level.

4 Discussion

We have estimated the measurement error σ_ϵ^2 of a trawl haul using data from parallel trawling experiments including 130 parallel hauls from 10 groups of experiments. No significant differences in σ_ϵ^2 among the groups were found. Thus, σ_ϵ^2 seems independent of year, time of the year and geographical position at which the haul is taken. It also seems to be independent of the catch size on a logarithmic scale. The magnitude of σ_ϵ^2 is small (approximately 2-5%) compared to the total variability in the survey trawl catches. The results are preliminary in that they are based on a limited set of hauls, and more extensive experiments would be of interest to check their consistency. In other experiments (cf. Pelletier 1998) one has examined the vessel effect between two research vessels. These data could possibly be used to test the general pattern revealed by analyses of data in the current study. Strømme and Illende[2] are currently examining a total of 365 paired hauls from intercalibration experiments off Namibia in 1998 and 1999 between the research vessel Dr. Fridtjof Nansen and commercial trawlers. These data of Namibian hake have kindly been made available to us, and an estimate of $\hat{\sigma}_\epsilon^2 = 0.19$ for the measurement error was obtained. The variance of y_i for all the 365 hauls was 2.0, so in this case the measurement error on the log scale was about 10% of the survey variance.

The estimate σ_ϵ^2 may actually be overestimating the measurement error since all the

---

[2] Strømme, T. and Illende, T. 2001. Precision in systematic trawl surveys as assessed from replicate sampling by parallel trawling. Preprint, Institute of Marine Research, Bergen, Norway.

explanatory variables are not exactly the same for the two set of measurements in (2.1). For example the geographical location is not the same and the fish densities may differ from one vessel to the other due to the distance between them. However, since the towed distance is typically 5-10 times the distance between the vessels, we believe this factor to be of minor importance.

Another problem is the determination of the towed distance. The uncertainty connected with subjective judgements and inaccuracies in the GPS should be included in the measurement error, since these factors are also present at a standard survey haul. However, it is not obvious to what extent the differences in the recorded towed distances are due to differences in subjective judgements or to differences in actual towed distances. In the calculations we have used the recorded values from both vessels for $d_{i;j}$. If there are no real difference in the towed distances within a comparison, $\hat{\sigma}_\epsilon^2$ is expected to decrease by setting $d_{i;1} = d_{i;2}$ for all hauls, thus eliminating one factor of uncertainty. At the other extreme, if the subjective judgments are perfect, and the recorded differences in towed distance are due to real differences, one would expect $\hat{\sigma}_\epsilon^2$ to increase by setting $d_{i;1} = d_{i;2}$, since an extra error then is added. Actually, using the values from vessel 2 only, and setting $d_{i;1} = d_{i;2}$, the resulting estimate of σ_ϵ^2 is $\hat{\sigma}_\epsilon^2 = 0.061$, i.e. a reduction by 11 %. Even though this is statistically insignificant, it indicates that uncertainty connected to the measurement of towed distance constitutes a part of the measurement error (see also Godø et al. 1990).

In three groups (2, 4 and 6) there were significant differences in efficiency between the vessels involved. In group 4, where the p-value was as low as 0.001, the higher efficiency of Jan Mayen was probably due to her use of heavier trawl doors. Excluding this group, the vessels can be ranged consistently after increasing efficiency (ignoring for the moment statistical significance) as

$$\text{JM} \xrightarrow{(7,8)} \text{GS} \xrightarrow{(5,9,3)} \text{JH} \xrightarrow{(1,2)/(6,10)} \text{MS/LIZY}. \tag{4.1}$$

The numbers in parentheses refer to experiment groups, and for each group one vessel to the left, and one to the right of the corresponding arrow, are involved, the one to the right being always the most efficient one. Joining data from groups where the same pair of vessels participates, thus increasing the ability to detect differences through an increased sample size, and testing for equal efficiency, we get the p-values 0.107 for groups 7 and 8, 0.034 for groups 5 and 9, 0.009 for groups 1 and 2, and 0.011 for groups 6 and 10. The consistency in the results and the low p-values indicate that there are in fact differences in efficiency between the vessels, as presented in (4.1).

There also seems to be a significant difference in the measurement error for small and large fish, it being higher for small fish. One explanation may be that the interaction between small fish and the trawl gear is more variable (Godø and Walsh 1992), another possibility is that small fish operate more in patches than do large fish. If the last assumption is correct, a reduction in $\hat{\sigma}_\epsilon^2$ could be expected with increasing tow length, since the variance of the ratio between patches encountered by the two vessels will decrease as tow distance increases. However for the set of tows considered, we found no significant difference in $\hat{\sigma}_\epsilon^2$ due to tow distance.

Consistent with the length dependency of the measurement error is the length dependency of the total variability of the surveys. The average var (y_i) for the winter surveys 1996-2000 for fish $\leq$ 31 cm and $\geq$ 64 cm, was 2.49 and 0.98, respectively, whereas for the unstratified data it was 1.66. The corresponding numbers for the autumn surveys 1996-2000 are 3.03, 1.47 and 2.98 for small, large and unstratified fish, respectively. All numbers are for non-zero catches.

Trawl catches have been considered highly variable (see e.g. Gulland 1964, Doubleday and Rivard 1981) and as a result the reliability of trawl survey estimates have been questioned. Abrupt changes in catch size and composition over short time in a limited area have demonstrated the difficulties in using the information as a relative estimate of density without understanding the nature and the causes of the variability (Godø 1994). Unexpected annual changes in survey indices may also be a problem for reliable evaluation of fish stocks and can be attributed to a variable bias (changes in catchability) among years (Pennington and Godø 1995). Our analysis demonstrates that catch rates and composition from the applied survey trawl are repeatable up to a constant and relatively small measurement error and are hence expected to give a reliable picture of the relative fish density at a given site and time. Further, the measurement error of this sampling gear is small compared to the total observed variability. For a particular survey it appears that most of the survey variance is caused by station-to-station differences in cathces rather than local conditions at a station. This may be taken as an indication that shorter and more frequent tows may be more efficient for monitoring this cod stock. Moreover, when controlling trawl geometry (Godø and Engås 1989) and towed distance (Godø et al. 1990) it should be possible to establish explanatory factors to be included in the survey assessment procedure. To the degree that one is able to establish models to determine fish densities at any station, the comparability of density measures throughout the distribution area will improve. The consequences will thus not only be more reliable survey estimates, but we also expect a better understanding of distribution patterns in relation to the physical and biological environment.

Akcnowledgements

We are grateful to Atle Totland for help in the data handling and to Michael Pennington for comments on the paper. Also, we are indebted to the editor and three anonymous referees for a number of useful comments, which have improved the paper. The work was financially supported by the Norwegian Research Council (050101).

References

Bickel, P.J., and Doksum, K.A. 1977. Mathematical Statistics: Basic Ideas and Selected Topics. Holden-Day, Oakland, California.

Doubleday, W.G., and Rivard, D. 1981. Bottom trawl surveys. Can. Spec. Publ. Fish. Aquat. Sci. **58** 1-273.

Engås, A. 1994. The effects of trawl performance and fish behaviour on the catching efficiency of demersal trawls. *In* Marine Fish Behaviour Related to Capture and Abundance Estimation. *Edited by* A. Fernø and S. Olsen. Fishing New Books, Oxford. pp. 45-68.

Godø, O.R. 1994. Factors affecting reliability of abundance estimates of ground fish from bottom trawl surveys. *In* Marine Fish Behaviour Related to Capture and Abundance Estimation. *Edited by* A. Fernø and S. Olsen. Fishing New Books, Oxford. pp. 166-199.

Godø, O.R., and Engås, A. 1989. Swept area variation with depth and its influence on abundance indices from trawl surveys. J. Northw. Atl. Fish. Sci. **9** 133-139.

Godø, O.R., and Walsh, S. 1992. Escapement of fish during trawl sampling — implication of resource assessment. Fish. Res. **13** 281-292.

Godø, O.R., Pennington, M., and Vølstad, J.H. 1990. Effect of tow duration on length composition of trawl catches. Fish. Res. **9** 165-179.

Gulland, J.A. 1964. Catch per unit effort as a measure of abundance. Rapp. P.-v. Reun. Cons. int. Explor. Mer **155** 8-14.

Korsbrekke, K., and Nakken, O. 1999. Length and species dependent diurnal variation in catch rates in the Norwegian Barents Sea bottom trawl surveys. ICES J. Mar. Sci. **56** 284-291.

Michalsen, K., Godø, O.R., and Fernø, A. 1996. Diel variation in the catchability of gadoids and its influence on the reliability of abundance indices. ICES J. Mar. Sci. **53** 389-395.

Pelletier, D. 1998. Intercalibration of research survey vessels in fisheries: a review and an application. Can. J. Fish. Aquat. Sci. **55** 2672-2690.

Pennington, M., and Godø, O.R. 1995. Measuring the Effect of Changes in Catchability on the Variation of Marine Survey Abundance Indices. Fish. Res. **23** 301-310.

Polacheck, T., and Vølstad, J.H. 1993. Analysis of spatial variability of Georges Bank haddock (*Melanogrammus aeglefinus*) from trawl survey data using a linear regression model with spatial interaction. ICES J. Mar. Sci. **50** 1-8.

Srivastava, M.S., and Chan, Y.M. 1989. A comparison of bootstrap method and edgeworth expansion in approximating the distribution of sample variance — one sample and two sample cases. Communication in Statistics — Simulation, **18** 339-361.