The measurement error in marine survey catches: the bottom trawl case

Vidar Hjellvik
Olav Rune Godø
Institute of Marine Research
Nordnesgt. 33
P.O. Box 1870, Nordnes
N-5817 Bergen, Norway
E-mail address (for V. Hjellvik): vidarh@imr.no

Dag Tjøstheim
Department of Mathematics
University of Bergen,
Johs. Brunsgt 12
N-5008 Bergen, Norway

Abstract—We have formulated a model for analyzing the measurement error in marine survey abundance estimates by 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 in 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.

Manuscript accepted 20 February 2002.
Fish. Bull. 100:720–726 (2002).

Surveys are vital for estimating the size and composition of marine populations. In the data collection process, trawl samples or acoustics (or both) are usually 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, 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 that 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 study was to define and quantify this residual source of random variation. This was done by analyzing measurements from parallel tows of multiple vessels, which is the closest one can come to a controlled statistical experiment in this context. The importance of quantifying this type of 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 study we looked 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 or more 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.

Materials and methods

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

Table 1

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

Group	Year	Date	Vessels	Lat.	Long.	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

of Marine Research, Bergen (IMR), (Jakobsen et al.[1]), parallel trawling experiments were used to compare the efficiency of the participating vessels with gear types as given in Table 1 and Figure 1. During a parallel haul the vessels operated about 500 meters apart and used radio contact to assure proper coordination during hauls. We analyzed 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 were excluded from the 1991 data (Michalsen et al., 1996). Similarly, two hauls from 1995 were excluded—one where trawl geometry measurements indicated problems with the doors and another with highly different recorded towed distances (0.7 and 2.2 nautical miles [nmi]). Let $d_{i;j}$ denote the towed distance for haul i and vessel j; i=1, … , n; j=1,2 where n is the total number of hauls. The average recorded distance is $\bar{d} = (2n)^{-1} \sum_{i=1}^{n}\sum_{j=1}^{2} d_{i;j} = 1.33$ nautical mile (average duration in time is 27 minutes), with $0.8 \le d_{i;j} \le 1.8$ 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.

The data were subdivided into 10 groups so that the same two vessels performed all hauls within a group, 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 nmi both in the east–west and in the north–south directions.

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 $\{y_i, i=1, \dots, n\}$ is given by

$$y_i = f(x_{i1}, \dots, x_{ip}) + \varepsilon_i, \qquad i = 1, \dots, n,$$

where f = a deterministic function, which in general is unknown; and

$x_{i1}, \dots, x_{ip}$ = 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 assumed to be linear.

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

$$y_k = f(x_1, \dots, x_p) + \varepsilon_k, \qquad k = 1, \dots, N.$$

The standard error of ε_k could then be estimated directly from the observations $\{y_k\}$ as

$$\hat{\sigma}_\varepsilon = \hat{\sigma}_y = \left\{ \frac{1}{N-1} \sum_{k=1}^{N} (y_k - \bar{y})^2 \right\}^{\frac{1}{2}},$$

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,

[1] Jakobsen, T., K. Korsbrekke, S. Mehl, and O. Nakken. 1997. Norwegian combined acoustic and bottom trawl surveys for demersal fish in the Barents Sea during winter. ICES CM 1997/Y:17:1–26. Institute of Marine Research, P.O.Box 1870 Nordnes, 5817 Bergen, Norway.

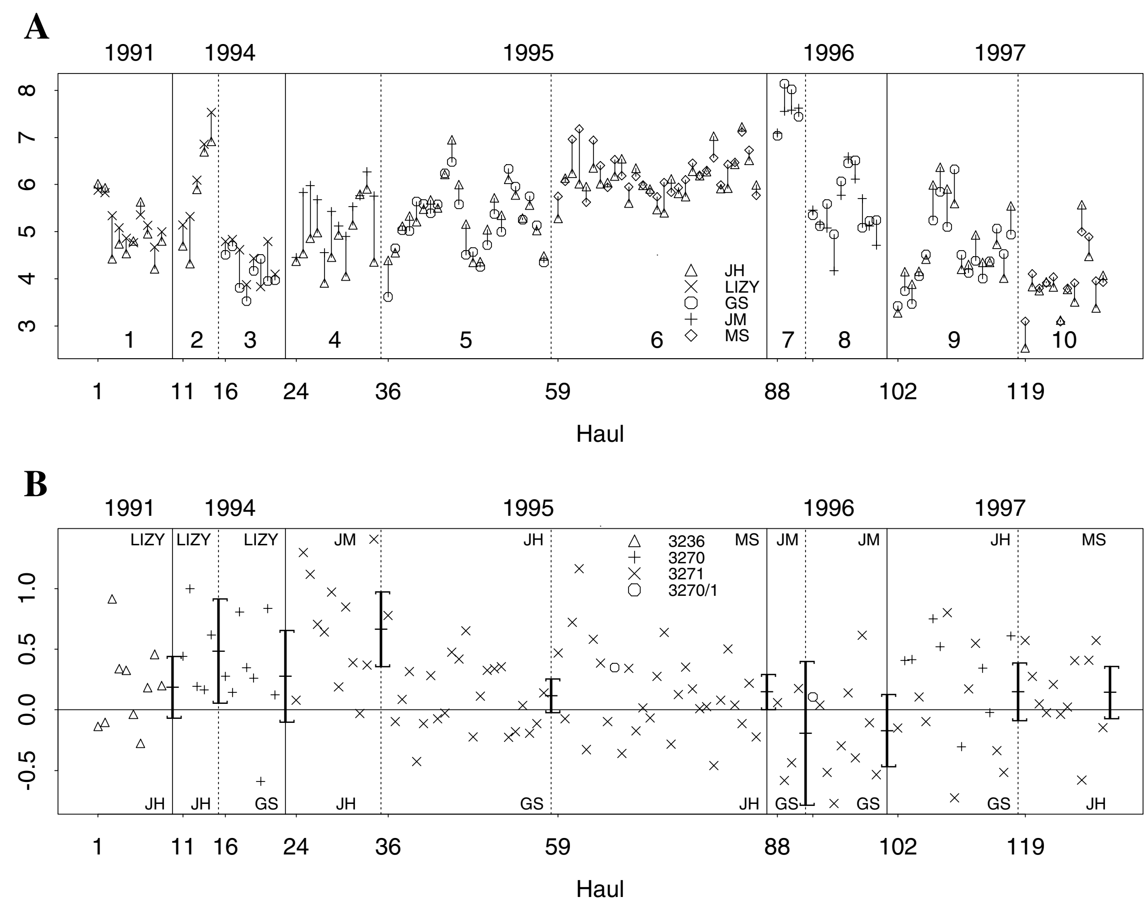

Figure 1

(**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 (see 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, 95% confidence intervals for the group means are given. The first and second vessel in each group are given at the top and bottom, respectively, for each group.

but as an approximation they will be considered identical. However, we will allow for an additive individual vessel effect α_j, j=1,2, which permits differences in equipment and efficiency for the two vessels. This leads to the following model for the observations $\{y_{i;j}, i=1, \dots, n; j=1,2\}$ where j=1, 2 corresponds to the first and second vessel in Table 1:

$$y_{i;1} = f(x_{i1}, \dots, x_{ip}) + \alpha_1 + \varepsilon_{i;1}$$
$$y_{1;2} = f(x_{i1}, \dots, x_{ip}) + \alpha_2 + \varepsilon_{i;2}. \quad (1)$$

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

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

$$z_i = y_{i;1} - y_{i;2} = \alpha_1 - \alpha_2 + \varepsilon_{i;1} - \varepsilon_{i;2}.$$

The expected difference between the two vessels is then given by

$$\mathrm{E}(z_i) = \alpha_1 - \alpha_2$$

and because of the independence of $\varepsilon_{i;1}$ and $\varepsilon_{i;2}$,

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

and the standard error σ_ε can be estimated as

$$\hat{\sigma}_\varepsilon = \frac{1}{\sqrt{2}}\hat{\sigma}_z = \frac{1}{\sqrt{2}}\left\{\frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2\right\}^{\frac{1}{2}}, \quad (2)$$

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

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

$$y_{i;j} = \log(n_{i;j}/d_{i;j}),\ i = 1, \ldots, n; \qquad j = 1,2, \qquad (3)$$

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

Tests for differences between experimental 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 $\hat{\sigma}_\varepsilon$, should be the same for all 10 experimental groups. If the z_i's originate from a Gaussian distribution, the null hypothesis of equal variance can be tested by Bartlett's test (*cf.* all groups tested simultaneously, Bickel and Doksum, 1977, p. 304) and if needed, followed by a series of F-tests where the groups are tested against each other in pairs.

The possible differences in efficiency caused by different vessels or fishing gears (or both) 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 are a prerequisite for such tests.

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

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

where, $k=k(i)$, $k=1, \ldots, 10$, denotes the group that haul i, $i=1, \ldots, n$, belongs to; and $\bar{z}_k$ and s_k are the average and the estimated standard deviation of the z-values in group k, respectively. 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).

Results

The log-catches $\{y_{i;j}\}$ and the corresponding differences $\{z_i\}$ are presented in Figure 1. The catches range 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 $\{x_i = (z_i - \bar{z}_k)/s_k\}$ appears linear (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. Bartlett's test for testing equality of variances 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 source of additional information 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., could not be rejected.

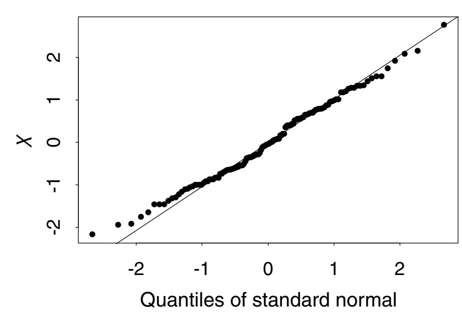

Figure 2
Normal probability plot of $\{x_i = (z_i - \bar{z}_k)/s_k\}$.

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 finding was consistent with earlier findings in calibration experiments (Pelletier, 1998). Thus, because $E(Z_i)$ cannot be considered equal for all groups (also the confidence intervals in Fig. 1B), a pooled variance could be used for estimating σ_ε^2. Alternatively, z_i in Equation 2 could be replaced by the variables $z_i' = z_i - \bar{z}_k$, which are adjusted for group means and are identical to the residuals from the ANOVA fit. The resulting estimates with the last approach are $\hat{\sigma}_\varepsilon^2 = 0.069$ and $\hat{\sigma}_\varepsilon = 0.263$. The bootstrapped standard errors of $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_\varepsilon$ are 0.0077 and 0.0147, respectively (1000 bootstrap replicates were used). Some caution should be exercised in interpreting these numbers (see e.g. Srivastava and Chan, 1989).

Compared with the total variability of the survey, the measurement error of a single haul is relatively small. For the last 5 years (1996–2000), $\mathrm{var}(y_i)$ for the nonzero catches varied 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 log scale. If antilogs of the catch rate were to be used, the additive model (Eq. 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 (Eq. 3) by, $m_{i;j}$, where $m_{i;j}$ is the weight of the catch in kilograms. Only hauls where both vessels

Table 2

Estimates of σ_ε^2 (with bootstrapped means and standard errors) for catches measured in kg (first column), total number of fish caught (second column), and catch of fish in various length-stratified groups measured in numbers (last four columns). The number of hauls where both vessels caught at least 10 cod is given in the fourth row, the groups excluded due to less than four remaining hauls are given in the fifth row, the average catch of the remaining hauls in the sixth row and *P*-values from Bartlett's test for equal variances, in the seventh row. The means and standard errors were estimated by using 1000 bootstrap replicates.

	kg	Total no. of fish caught	Catch of fish in length stratified groups			
	>0 cm	>0 cm	<30 cm	30–59 cm	≥60 cm	≥30 cm
$\hat{\sigma}_\varepsilon^2$	0.0744	0.0690	0.1124	0.0801	0.0855	0.0643
boot. mean of $\hat{\sigma}_\varepsilon^2$	0.0740	0.0684	0.1113	0.0788	0.0844	0.0637
boot. SE of $\hat{\sigma}_\varepsilon^2$	0.0096	0.0077	0.0140	0.0114	0.0122	0.0083
No. of hauls with at least 10 cod	130	130	113	117	103	118
groups excluded	none	none	2,3	10	2,7,10	10
$\bar{y}$	4.77	5.28	4.61	4.06	3.61	4.52
P-value, Bartlett	0.11	0.81	0.31	0.87	0.60	0.97

collected 10 specimen or more were included, and only groups with at least four such hauls. The null hypothesis of equal variance was not rejected for any of the length groups (*P*-values from Bartlett's test are given in Table 2). It is seen that $\hat{\sigma}_\varepsilon^2$ is highest for small fish, and the bootstrapped standard errors indicate that the difference is statistically significant. Indeed, let $D = \hat{\sigma}^2_{\varepsilon,small} - \hat{\sigma}^2_{\varepsilon,large}$ denote the difference in measurement error between small (<30 cm) and large (≥30 cm) fish. If $\hat{\sigma}^2_{\varepsilon,small}$ and $\hat{\sigma}^2_{\varepsilon,large}$ are independent and normally distributed (their bootstrap distributions are approximately normal) with standard errors as given in Table 2, it follows that $D \sim N(0, 0.0140^2 + 0.0083^2)$ under the null hypothesis of equal measurement errors. The corresponding one-sided *P*-value for the observed D is 0.0016.

The hauls differed in towed distance, with the two most frequently recorded values being 1 and 1.5 nmi which correspond to 20 and 30 minutes tow duration. Stratifying on tow duration (which is recorded more precisely 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 $\hat{\sigma}_\varepsilon^2$=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 $\bar{y}_i = (y_{i;1} + y_{i;2})/2$ as the independent variable. The regression equation was $|z_i'|=0.20+0.019y_i$ and the *P*-value under the null hypothesis of no relationship was 0.31. However, the residuals from the analysis were skewed with a long right-hand tail; therefore 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.

Discussion

We have estimated the measurement error (σ_ε^2) of a trawl haul by 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 to be independent of year, time of the year, and geographical position at which the haul was taken. It also seems to be independent of the catch size on a logarithmic scale. The magnitude of σ_ε^2 is small (≈2–5%) compared with 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 another investigation, Pelletier (1998) 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 Iilende (2001) examined 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 (catch in kg/h) were kindly made available to us, and an estimate of $\hat{\sigma}_\varepsilon^2$=0.19 was obtained for the measurement error. The variance of y_i for all the 365 hauls was 2.0; the measurement error for this study on the log scale was about 10% of the survey variance.

Actually σ_ε^2 may be an overestimation of the measurement error because all the explanatory variables are not exactly the same for the two sets of measurements in Equation 1. For example the geographical location is not the same and the fish densities may differ from one vessel to the other because of the distance between them. However, because 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 judgments and inaccuracies in the GPS should be included in the measurement error because 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 judgments or to differences in actual towed distances. In our calculations, we used the recorded values from both vessels for $d_{i;j}$. If there is no real difference in the towed distances within a comparison, $\hat{\sigma}^2_\varepsilon$ 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 $\hat{\sigma}^2_\varepsilon$ to increase by setting $d_{i;1}=d_{i;2}$, because an extra error then is added. Actually, by using the values from vessel 2 only and by setting $d_{i;1}=d_{i;2}$ the resulting estimate of σ^2_ε is $\hat{\sigma}^2_\varepsilon=0.061$, which is a reduction by about 11%. Even though this value 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 the "Results" section, the null hypothesis of equal efficiency for all the participating vessels was rejected. By joining data from the groups where the same pair of vessels participates, the ability to detect differences in efficiency between the vessels increases through an increased sample size and a smaller number of simultaneous tests (with N tests and a nominal level α, the null hypothesis of equal efficiency is rejected for P-values smaller than the Bonferroni corrected level α/N). For the N=6 tests, we obtained P-values 0.0005 for group 4, 0.009 for groups 1 and 2, 0.011 for groups 6 and 10, 0.034 for groups 5 and 9, 0.107 for groups 7 and 8, and 0.123 for group 3. With a level α=0.05 we have $\alpha/6$=0.0083, and for group 4 the difference is clearly statistically significant. The higher efficiency of *Jan Mayen* (JM) in this group was probably due to her heavier trawl doors. At a 10% level, the vessel *Anny Kræmer* (LIZY) was significantly more efficient than *Johan Hjort* (JH) in groups 1 and 2, and *Michael Sars* (MS) was significantly more efficient than JH in groups 6 and 10. The differences between *G.O. Sars* (GS) and JH (groups 5 and 9), GS and JM (groups 7 and 8) and GS and LIZY (group 3) were not significant. However, excluding group 4, and ignoring statistical significance, the vessels can, in fact, be ranged consistently after increasing efficiency as

$$JM \xrightarrow{(7,8)} GS \xrightarrow{(5,9,3)} JH \xrightarrow{(1,2)/(6,10)} LIZY\ /\ MS. \quad (4)$$

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.

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 $\hat{\sigma}_\varepsilon$ could be expected with increasing tow length. However, for the set of tows considered, we found no significant difference in $\hat{\sigma}_\varepsilon$ 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 ≤31cm and ≥64cm, 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 were 3.03, 1.47, and 2.98 for small, large, and unstratified fish, respectively. All numbers are for nonzero 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 a short time in a limited area have demonstrated the difficulties in using the information as a relative estimate of density without an understanding of the nature and causes of the variability (Godø, 1994). Unexpected annual changes in survey indices may also be a problem for a 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 for the bottom trawl survey in the Barents Sea, catch rates and composition from the applied survey trawl are repeatable up to a relatively small and constant 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 with the total observed variability. For a particular survey it appears that most of the survey variance is caused by station-to-station differences in catches 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 distributional patterns in relation to the physical and biological environment.

Acknowledgments

We are grateful to Atle Totland for help with handling the data and to Michael Pennington for comments on the paper. We are also indebted to the scientific editor and three anonymous referees for a number of useful comments that improved the paper. The work was financially supported by the Norwegian Research Council (127198/120).

Literature cited

Bickel, P. J., and K. A. Doksum.
1977. Mathematical statistics: basic ideas and selected topics. Holden-Day, Oakland, CA, 493 p.

Doubleday, W. G., and D. Rivard.
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 (A. Fernø and S. Olsen, eds.), p. 45–68. Fishing New Books, Oxford.

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 (A. Fernø and S. Olsen, eds.), p. 166–199. Fishing New Books, Oxford.

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

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

Godø, O. R., M. Pennington, and J. H. Vølstad.
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 O. Nakken.
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., O.R. Godø, and A. Fernø.
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 O. R. Godø.
1995. Measuring the effect of changes in catchability on the variation of marine survey abundance indices. Fish. Res. 23:301–310.

Polacheck, T., and J. H. Vølstad.
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 Y. M. Chan.
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.

Strømme, T., and T. Iilende.
2001. Precision in systematic trawl surveys as assessed from replicate sampling by parallel trawling off Namibia. S. Afr. J. Mar. Sci. 23:385–396.