Fol. 41 D

International Council for
the Exploration of the Sea

C.M. 1992/D:9

Measuring the Effect of Changes in Catchability on the Variance of Marine Survey Abundance Indices

Michael Pennington
National Marine Fisheries Service
Woods Hole Laboratory
Woods Hole, Mass. USA 02543

and

Olav Rune Godø
Institute of Marine Research
N5024 Bergen, Norway

Abstract

The average catch per tow by a research vessel survey is often used as an index of abundance. An estimate of the variance of such indices that is based only on the between station variability in catch may underestimate the true variance if catchability varies over time. In this paper, the survey index variance is estimated indirectly by cross calibrating time series of VPA estimates and trawl survey indices of abundance. The method is applied to several species on Georges Bank and to haddock in the Barents Sea. For these surveys, it appears that the true variance of the survey indices is approximately twice as large as the usual estimates based on the within survey variance. As an application, a time series technique, which requires an estimate of the survey index variance, is used to generate a more precise index of abundance. The results indicate that for the surveys examined the variance of the estimated abundance index is 30 - 40% smaller than the original (average catch per tow) index.

1. Introduction

A potentially significant source of variability for trawl survey indices of abundance is that the catchability of a species with respect to the survey trawl may change from year to year (Sissenwine and Bowman 1978; Byrne et al. 1981; Collie and Sissenwine 1983; Pope 1988; Shepherd 1988). Many factors may cause such fluctuations. These may include variations in environmental conditions, changes in the spatial or vertical distribution of a stock, or varying equipment and crew. If catchability does change from year to year then an estimate of the error variance that is based only on the within survey variability would be an underestimate.

It would be useful to have an estimate of the true survey variance. For tuning a VPA, the estimated variance would provide a weighing factor for the survey series when several abundance indices are available. For survey design, it is needed for determining the relationship between sample size and survey precision. Finally, an estimate of survey error is necessary for a time series method which attempts to generate a more precise abundance index.

Since the true population size is rarely if ever known, it is necessary to estimate indirectly the variance of a survey index. In this paper the actual survey variance is estimated by cross calibrating VPA abundance estimates and survey catch per tow indices. As examples, the technique is applied to surveys for several species and some applications of the results are presented.

2. Statistical Methods

For a particular species let y_t denote the survey index of abundance (e.g., the average catch per tow), z_t an estimate of stock size from a VPA and p_t the actual population size for year t. It is assumed that the expected value of the survey index and of the VPA are proportional to population size. That is $E[y_t] = kp_t$ and $E[z_t] = k'p_t$. Since for marine data errors often appear to be multiplicative, let

$$y_t = kp_t\mathrm{EXP}(\delta_t+\epsilon_w),$$

where δ_t is the error term due to yearly changes in catchability and ϵ_w is the error due to within survey variability.

For simplicity, from here on, the same notation will be used to denote the logarithm of the variables. Thus

$$y_t = k + p_t + e_t \ ,$$

where $e_t = \delta_t + \epsilon_w$. If the error components are independent, then

$$\sigma_e^2 = \sigma_\delta^2 + \sigma_w^2. \qquad (1)$$

Likewise, the VPA series can be expressed as

$$z_t = k' + p_t + \xi_t \ ,$$

where ξ_t measures the precision of the VPA. Again, it is assumed that the expected value of the VPA is not necessarily equal to, but only that it is proportional to stock size .

In order to express the error components in terms of the observed series, consider first that

$$\mathrm{Var}(y_t - z_t) = \mathrm{Var}(e_t - \xi_t)$$

$$= \sigma_e^2 + \sigma_\xi^2 - 2\rho\sigma_e\sigma_\xi ,$$

where $\rho = \mathrm{Cor}(e_t, \xi_t)$. If $\rho = 0$, then

$$\mathrm{Var}(y_t - z_t) = \sigma_e^2 + \sigma_\xi^2. \tag{2}$$

The assumption that $\rho = 0$ is the reason a survey series and a VPA series are considered rather than, for example, two independent trawl surveys. If catchability does vary from year to year, then it is unlikely that the measurement errors for the two surveys would be uncorrelated.

Next consider the variance of the year to year changes in the series. These changes can be written using the backward difference operator ∇ which is defined by $\nabla z_t = z_t - z_{t-1}$. The value of $\mathrm{Var}(\nabla z_t)$ is calculated, for example, rather than $\mathrm{Var}(z_t)$ because z_t often appears to be nonstationary or nearly so (i.e., it does not seem to fluctuate about a fixed mean) while ∇z_t usually seems stationary. If z_t is nonstationary, then $\mathrm{Var}(z_t)$ is not defined.

For the VPA series, if ∇p_t and $\nabla \xi_t$ are uncorrelated, then

$$\mathrm{Var}(\nabla z_t) = \mathrm{Var}(\nabla p_t + \nabla \xi_t)$$

$$= \mathrm{Var}(\nabla p_t) + 2(1 - \rho)\sigma_\xi^2,$$

where $\rho = \mathrm{Cor}(\xi_t, \xi_{t-1})$. Thus if $\rho = 0$, then

$$\mathrm{Var}(\nabla z_t) = \mathrm{Var}(\nabla p_t) + 2\sigma_\xi^2. \tag{3}$$

Similarly, if ∇p_t and ∇e_t are uncorrelated, and $\mathrm{Cor}(e_t, e_{t-1}) = 0$, then

$$\mathrm{Var}(\nabla y_t) = \mathrm{Var}(\nabla p_t) + 2\sigma_e^2. \tag{4}$$

It follows from equations (3) and (4) that

$$[\mathrm{Var}(\nabla y_t) - \mathrm{Var}(\nabla z_t)]/2 = \sigma_e^2 - \sigma_\xi^2. \tag{5}$$

In summary, given the assumptions that have been made, the error variances can be expressed in terms of the observable series and the variance caused by within survey variability, σ_w^2. The variance of the survey abundance index can be written as

$$\sigma_e^2 = \{\mathrm{Var}(y_t - z_t) + [\mathrm{Var}(\nabla y_t) - \mathrm{Var}(\nabla z_t)]/2\}/2, \tag{6}$$

and the variance of the VPA error component is given by

$$\sigma_\xi^2 = \{\mathrm{Var}(y_t - z_t) - [\mathrm{Var}(\nabla y_t) - \mathrm{Var}(\nabla z_t)]/2\}/2. \tag{7}$$

Finally, the variance component of the survey index due to varying catchability is given by

$$\sigma_\delta^2 = \sigma_e^2 - \sigma_w^2. \tag{8}$$

3. Examples and Applications

3.1 Example: estimating error components

In Table 1 are the results of substituting sample estimates into the appropriate equations for several species with both a VPA series and a series of trawl surveys. This was done using the entire series [Table 1 (a)] and to lessen any effect that the survey indices may have on the VPA estimates, the series through 1985 [Table 1 (b)]. As an indirect check of the assumption that both series are proportional to the true population, the log of the survey index was regressed on the log of the more precise VPA estimates. If this assumption is correct, then the slopes should be equal to one. None of the estimated slopes (column 2) is significantly different from one. In column 3 are estimates of the sum of the error components from equation (2) and in column 4 are estimates of the differences [equation (5)]. In the next two columns are estimates of the error terms [equations (6) and (7)]. In the last column [Table 1 (b)] are estimates of the variance of the survey index, σ_w^2, based on the within survey variability.

3.2 Application: effect of sample size on survey precision

Increasing the number of stations in a survey will only reduce the component of the survey variance due to within survey variability. For example, using equations (1) and (8) and estimates of σ_e^2 and σ_w^2 [Table 1 (b)], it appears that doubling the number of stations for the Georges Bank survey would reduce the variance of the haddock index from .20 to .16. The smallest the variance could be if just sample size is increased is approximately $\hat{\sigma}_\delta^2 = .12$.

3.3 Application: estimating an index of abundance

If the error variance of a survey is known, time series methods can be used to estimate an abundance index that has a smaller variance than the original series. Briefly, this can be done as follows (for details, see Pennington, 1985; 1986).

Suppose p_t follows the autoregressive integrated moving average process (see, e.g., Box and Jenkins 1976)

$$\Phi(B)p_t = \Theta(B)a_t ,$$

where the a_t's are iid $N(0, \sigma_a^2)$. If $y_t = k + p_t + e_t$ and the e_t's are iid $N(0, \sigma_e^2)$ and independent of the a_t's, then it follows that y_t can be represented by the process

$$\Phi(B)y_t = \eta(B)c_t \ , \qquad (9)$$

where the c_t's are iid $N(0, \sigma_c^2)$. Now if model (9) and σ_e^2 are known then the maximum likelihood estimate of the current value of $k + p_t$ is given by

$$y_t - (\sigma_e^2/\sigma_c^2)\hat{c}_t \ , \qquad (10)$$

where the $\hat{c}_t$'s are the estimated residuals from model (9). The variance of the index is approximately equal to

$$\sigma_e^2[1 - (\sigma_e^2/\sigma_c^2)],$$

as compared with σ_e^2, the error variance of the original survey series.

Equation (10) is convenient for computing the index, but its structure can be more easily seen if it is written in the form

$$(1-\theta)y_t + \theta\hat{y}_{t-1}(1), \qquad (11)$$

where $\theta = \sigma_e^2/\sigma_c^2$ and $\hat{y}_{t-1}(1)$ denotes the one-step-ahead forecast of y_t at time $t - 1$. That is the index is the weighted average of the current survey value and the predicted population level based on the previous values of the series. From (11) it also can be seen that the index is an unbiased estimator of $k + p_t$.

For the four species considered in section 3.1, the autoregressive model

$$y_t' = \phi y'_{t-1} + c_t \ , \qquad (12)$$

where $y_t' = y_t - \mu$, fits all the series adequately. The results are in Table 2. In the second column are the estimates of ϕ. The ratio, σ_e^2/σ_c^2, was estimated using the estimate of σ_e^2 [from Table 1 (b)] and the estimate of the residual variance for model (12). In column 4 are estimates of $Var(z_t - \hat{y}_t)$ for the series through 1985 (except for Barents Sea haddock, which is based on the entire series), where $\hat{y}_t$ denotes the index of abundance from equation (10). For comparison, in the last column are estimates of $Var(z_t - y_t)$, a measure of the performance of the original index.

Another approach to estimate an index of abundance, which does not require VPA estimates, is to assume that the population follows the model

$$p_t = p_{t-1} + a_t \ , \qquad (13)$$

where the a_t's are iid $N(0, \sigma_a^2)$. Then if y_t is related to p_t as above and the e_t's are independent of the a_t's, it follows that

$$(1 - B)y_t = (1 - \theta B)c_t \ , \qquad (14)$$

and $\theta = \sigma_e^2/\sigma_c^2$. Thus $\hat{\theta}\hat{\sigma}_c^2$ is an estimator of σ_e^2 (again, for further details see Pennington 1985; 1986).

Therefore if the assumptions hold σ_e^2 can be estimated using only the survey series. It is also not necessary to assume that the series has a mean. In contrast, forecasts based on model (12) assume the mean of the entire series has some relevance to future stock sizes (for more on the advantages of using a nonstationary model, see Box and Jenkins 1976, p. 192).

In Table 3 are the estimates based on model (14). Two estimates of θ were used to estimate an abundance index; $\hat{\theta}_1$ from fitting model (14), and $\hat{\theta}_2$ based on the model and the estimates of σ_e^2 [Table 1 (b)]. The resulting estimation variances based on the internal and external estimates of σ_e^2 (last two columns of Table 2) are similar and are comparable to those from model (12) [Table 2]. In figure 1 are plots of the original survey series, the VPA series and survey abundance indices based on model (14). It appears that the estimated survey index tracks the VPA fairly well except toward the end of the series; the period when the VPA estimates are the most imprecise. It may be noted that the estimated survey indices (Figure 1) are based only on information available during each year. A more precise historical index may be generated by using data from both previous and subsequent years (Pennington 1985).

4. Discussion

Since yearly changes in catchability appear to contribute significantly to the variance of the survey abundance indices examined, increasing the number of stations sampled would only marginally improve their precision. A more efficient approach may be one that determines the causes of catchability changes and adjusts the indices accordingly. For example, the area swept by a trawl varies with depth (Godø and Engås 1989). Thus if the spatial distribution of a stock changes, say to deeper water, its catchability may be affected. Given sufficient understanding of survey trawl performance, catch at a station could, perhaps, be adjusted for depth. Another possibility is to use acoustics to track changes in a stock's vertical and spatial distribution and quantify the effect these changes have on catchability.

References

Box, G.E.P., and G.M. Jenkins. 1976. Time Series Analysis: Forecasting and Control. Rev. ed. Holden-Day, San Francisco, 575 p.

Byrne, C.J., T.R. Azarovitz, and M.P. Sissenwine. 1981. Factors affecting variability of research trawl surveys. Can. Spec. Publ. Aquat. Sci. 58:238-273.

Collie, J.S., and M.P. Sissenwine. 1983. Estimating population size from relative abundance data measured with error. Can. J. Fish. Aquat. Sci. 40:1871-1879.

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

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

Pennington, M. 1986. Some statistical techniques for estimating abundance indices from trawl surveys. Fishery Bulletin 84:519-525.

Pope, J.G. 1988. Collecting fisheries assessment data. Fish Population Dynamics (Second Edition) pp. 63-82. Ed. J.A. Gulland. John Wiley & Sons Ltd. Avon, G.B.

Shepherd, J.G. 1988 Fish stock assessments and their data requirements. Fish Population Dynamics (Second Edition) pp. 35-62. Ed. J.A. Gulland. John Wiley & Sons Ltd. Avon, G.B.

Sissenwine, M.P. and E.W. Bowman. 1978. An analysis of some factors affecting the catchability of fish by bottom trawls. ICNAF Bulletin 13:81-87

Table 1.

Summary statistics for estimating the variance components of the survey abundance indices and VPA estimates for Georges Bank haddock (age 2+) and yellowtail (1+), Southern New England yellowtail (1+) and Barents Sea haddock (3+). A dash represents a negative component estimate.

(a) Estimates based on entire series.

Survey	slope	$\hat{\sigma}_e^2 + \hat{\sigma}_\xi^2$	$\hat{\sigma}_e^2 - \hat{\sigma}_\xi^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_\xi^2$
haddock Georges Bank	.98	.18	.19	.19	-
yellowtail Georges Bank	1.09	.39	.33	.36	.03
yellowtail So. New Eng.	.84	.39	.20	.29	.10
haddock Barents Sea	1.04	.56	.48	.52	.04

(b) Estimates based on series through 1985.

Survey	slope	$\hat{\sigma}_e^2 + \hat{\sigma}_\xi^2$	$\hat{\sigma}_e^2 - \hat{\sigma}_\xi^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_\xi^2$	$\hat{\sigma}_w^2$
haddock Georges Bank	1.03	.19	.22	.20	-	.08
yellowtail Georges Bank	1.12	.20	.20	.20	.00	.11
yellowtail So. New Eng.	1.16	.29	.26	.28	.02	.12

Table 2.

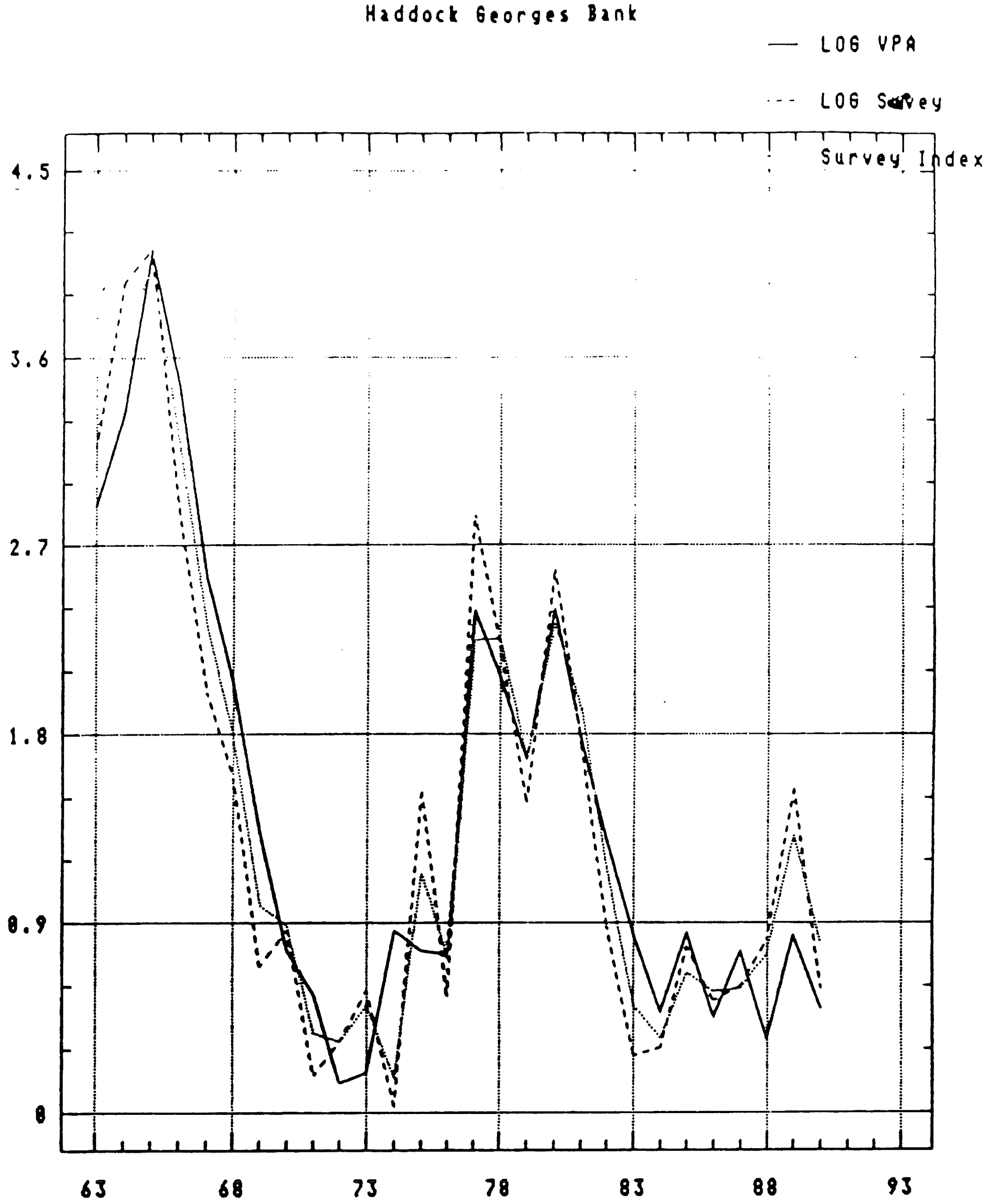
Parameter estimates for the survey abundance index based on the autoregressive model of the survey series. Except for Barents Sea haddock, a relatively short series, the comparison of the estimated index and the VPA series (col. 4) is for the series through 1985 as is that for the average catch per tow index (last column).

Survey	$\hat{\phi}$	$\hat{\sigma}_e^2/\hat{\sigma}_c^2$	$var(z_t - \hat{y}_t)$	$var(z_t - y_t)$
haddock Georges Bank	.76	.30	.08	.19
yellowtail Georges Bank	.82	.42	.05	.20
yellowtail So. New Eng.	.77	.41	.14	.29
haddock Barents Sea	.60	.34	.28	.56

Table 3.

Parameter estimates for the survey abundance indices generated by the integrated moving average model. The estimate, $\hat{\theta}_1$, is based only on the survey series and $\hat{\theta}_2$ is the estimate derived from both series.

survey	$\hat{\theta}_1$	$\hat{\theta}_2$	$var_1(z_t - \hat{y}_t)$	$var_2(z_t - \hat{y}_t)$
haddock Georges Bank	.23	.28	.08	.07
yellowtail Georges Bank	.44	.42	.05	.06
yellowtail So. New Eng.	.31	.43	.15	.16
haddock Barents Sea	.18	.34	.30	.22



(a)

Figure 1. VPA estimates of relative abundance (solid line), average catch per tow index (dashed line) and estimated index of abundance (dotted line) for Georges Bank haddock (a) and yellowtail (b), Southern New England yellowtail (c) and Barents Sea haddock (d).