Northwest Atlantic Fisheries Organization

Serial No. N1055 NAFO SCR Doc. 85/81

SCIENTIFIC COUNCIL MEETING - SEPTEMBER 1985

Some Statistical Techniques for Estimating Abundance Indices from Trawl Surveys

by

Michael Pennington

Northeast Fisheries Center, Woods Hole Laboratory
Woods Hole, Massachusetts, 02543, USA

ABSTRACT

Methods are presented for estimating an index of relative abundance from trawl survey catch per tow data. The estimated variance of the index takes into account the within survey variability in catch and possible yearly changes in catchability. Applying the techniques to a series of surveys for yellowtail flounder (Limanda ferruginea) off the northeast coast of the United States yields an abundance index with a variance which is 40% lower than the variance of the original survey index for the current value and 57% lower for values not near the ends of the survey series.

SPECIAL SESSION ON BIOLOGICAL SURVEYS

INTRODUCTION

The average number of fish caught per tow during a trawl survey is often used as an index of a species's relative abundance (Grosslein, 1969; Clark, 1979). Catch per tow data are usually quite variable due to the heterogeneous distribution of many fish stocks (Byrne et al., 1981). A further source of variability for survey indices of abundance is that the catchability of a particular species with respect to the survey trawl may change from year to year (Byrne et al., 1981; Collie and Sissenwine, 1983). As a result, the observed time series of abundance indices reflects changes in the population, within survey sampling variability, and varying catchability over time.

This paper uses various statistical methods to construct from the catch per tow data an index of abundance which more closely tracks the population than does the original (average catch per tow) series. Specifically, since the distribution of catch per tow data is often highly skewed and contains a proportion of zeros, estimates of the mean catch per tow for each survey are made based on the Δ-distribution (Aitchison and Brown, 1957). Next, time series techniques are used to estimate the component of the series generated by the actual changes in the population.

The methods are applied to data for yellowtail flounder (*Limanda ferruginea*) from a series of groundfish trawl surveys conducted off the northeast coast of the United States as part of the National Marine Fisheries Service's MARMAP program. The resulting index of abundance is substantially more precise than the original index.

STATISTICAL METHODS

Sources of Variability

Let y_t denote the observed average catch per tow for the survey conducted in year t and $z_t' = E[y_t]$, the expected value of y_t. Since a species catchability may change from year to year with respect to the survey trawl, let $z = E[z'|p]$ denote the expected value of z' given a population level p. Then

$$y_t = z_t + e_t.$$

The error term, e_t, can be expressed as

$$e_t = (y_t - z_t') + (z_t' - z_t),$$

where the first error component is due to the within survey variability and the second is due to changes in catchability.

In order to construct an index of abundance, it is necessary to assume a functional relationship between z_t and p_t. A reasonable assumption made in practice (and in this paper) is that

$$z_t = ap_t.$$

If the relationship is not linear, then the unadjusted catch per tow index will be a biased measure of relative abundance.

Estimating the mean catch per tow

The distribution of marine survey data often can be described by what is called a Δ-distribution (Aitchison and Brown, 1957). That is, the data contain a proportion of zeros and the nonzero values are distributed lognormally. The minimum variance unbiased estimates of the mean and its variance for the Δ-distribution are given by (Pennington, 1983),

$$c = \begin{cases} \frac{m}{n} \exp(\bar{y}) G_m(s^2/2), & m>1, \\ \frac{x_1}{n}, & m=1, \\ 0, & m=0, \end{cases} \qquad (1)$$

and

$$\mathrm{var}(c) = \begin{cases} \frac{m}{n} \exp(2\bar{y}) \{\frac{m}{n} G_m^2(s^2/2) - (\frac{m-1}{n-1}) G_m (\frac{m-2}{m-1}) s^2)\}, & m>1, \\ (\frac{x_1}{n})^2, & m=1, \\ 0, & m=0, \end{cases} \qquad (2)$$

where n is the number of tows, m is the number of nonzero values, $\bar{y}$ and s^2 are the sample mean and variance respectively of the nonzero $\log_e$ values, x_1 is the single (untransformed) nonzero value when m=1, and

$$G_m(x) = 1 + \frac{m-1}{m} x + \sum_{j=2}^{\infty} \frac{(m-1)^{2j-1} x^j}{m^j (m+1)(m+3)\ldots(m+2j-3) j!}.$$

Figure 1, which is an extension of a graph in Aitchison and Brown (1957, p. 98), shows the large sample efficiency of the ordinary sample statistics as compared with their most efficient estimates for the Δ-distribution with 50% zeros. Estimates of σ^2, the variance of the nonzero $\log_e$ values, are often between 1 and 2 for trawl surveys. Thus (see Figure 1) the sample mean is a fairly efficient estimator of the mean for trawl surveys, but the sample variance is highly inefficient. Though for larger values of σ^2, which, for example, are common for egg surveys (Pennington and Berrien, 1984), the sample mean is also very inefficient.

Estimating the Index of Abundance

The estimated mean catch per tow, y_t, as an index of relative abundance has two drawbacks. First, its estimated variance when derived from the within survey variance, can be an underestimate since catchability may vary from year to year. A second and more serious deficiency is that the index for a particular year is based only on that year's survey which disregards relevant information contained in the surveys for other years.

One method to construct an abundance index based on the entire survey series is briefly as follows. More details can be found in Pennington (1985).

Suppose the population (or z_t) can be represented by the autoregressive integrated moving average process (Box and Jenkins, 1976, Chap. 4)

$$\phi(B)z_t = \theta(B)a_t.$$

where the a_t's are independently identically and normally distributed with mean zero and variance σ_a^2 [iid $N(0,\sigma_a^2)$]. If $y_t = z_t + e_t$, and the e_t's are assumed iid $N(0,\sigma_e^2)$, then y_t will follow the model

$$\phi(B)y_t = \eta(B)c_t, \qquad (3)$$

where the c_t's are iid $N(0,\sigma_e^2)$. Now if model (3) and the ratio σ_e^2/σ_c^2 are known, then the maximum likelihood estimate of z_t is given by

$$\hat{z}_t = y_t - \frac{\sigma_e^2}{\sigma_c^2}(\hat{c}_t - \pi_1 \hat{c}_{t+1} - \pi_2 \hat{c}_{t+2} -, \ldots, - \pi_{T-t}\hat{c}_T), \quad (4)$$

where T denotes the last year of the series, the $\hat{c}_t$'s are the estimated residuals generated by model (3), and the π values are calculated using the identity

$$\phi(B) = (1 - \pi_1 B - \pi_2 B^2 - \ldots)\eta(B). \quad (5)$$

The variance of $\hat{z}_t$ is given approximately by

$$\mathrm{var}(\hat{z}_t) \doteq \sigma_e^2 \, [1 - (\pi_0^2 + \pi_1^2 + \ldots + \pi_{T-t}^2) \frac{\sigma_e^2}{\sigma_c^2}], \quad (6)$$

where $\pi_0 = 1$.

The model for y_t [equation (3)] is usually obtained in practice by fitting a model to the observed series using procedures described in Box and Jenkins (1976). If catchability is constant over time, the within survey sampling variance provides an estimate of σ_e^2. But if catchability varies, another approach is necessary.

Toward this end, consider the expression

$$z_t = z_{t-1} e^{a_t}$$

or

$$(1-B)\ln z_t = a_t. \quad (7)$$

Suppose the factors causing the change in population from year t-1 to year t (such as recruitment, fishing mortality, natural mortality, and migrations) produce a_t's which are approximately iid $N(0,\sigma_a^2)$. If the measurement errors are multiplicative, then

$$\ln y_t = \ln z_t + e_t. \quad (8)$$

Assuming the e_t's are iid $N(0,\sigma_e^2)$ and independent of the a_t's, then it follows as above that y_t can be represented by the model

$$(1-B)\ln y_t = (1-\theta B)c_t. \qquad (9)$$

where the c_t's are iid $N(0,\sigma_c^2)$.

For model (9) [generated by equations (7) and (8)]

$$\theta = \sigma_e^2/\sigma_c^2$$

and

$$(1-\theta)^2 = \sigma_a^2/\sigma_c^2 . \qquad (10)$$

Therefore, assuming the above approximations to the population dynamics, fitting model (9) to the observed survey series provides an estimate, $\hat{\theta}$, of σ_e^2/σ_e^2. The π-weights for the model are from equation (5) given by

$$\pi_i = (1-\hat{\theta})\,\hat{\theta}^{\,i-1}, \qquad i\geq 1. \qquad (11)$$

AN APPLICATION

The Northeast Fisheries Center conducts an intensive groundfish trawl survey as part of its MARMAP program two times a year; in fall since 1963 and in spring since 1968 (Grosslein, 1969). The survey region is divided into sampling strata based on geographic boundaries and depth contours (Figure 2). For each survey, trawl stations are chosen randomly within each stratum. One of the objectives of the surveys is to provide indices of abundance for the many species of commercial value in the region.

Yellowtail flounder (<u>Limanda ferruginea</u>) is an important New England fishery resource whose population has fluctuated considerably over the survey period (Clark et al., 1984). Commercial catch statistics exist for yellowtail, but age data suitable for a VPA analysis are unavailable. Major yellowtail fisheries are off Southern New England (Strata 5, 6, 8, 9) and on Georges Bank (Strata 13-21). The two stocks are fairly distinct but with some intermixing (Clark et al., 1984).

The nonzero catch per tow survey data for yellowtail are approximately lognormally distributed within a stratum. Therefore, the estimators based on the Δ-distribution [equations (1) and (2)] were used to estimate the mean catch per tow and its variance in each stratum. The regional estimates for Southern New England and Georges Bank were then calculated in the usual manner for each survey (see e.g. Pennington and Brown, 1981).

Model (9) was fit to each series (spring 1968-1984 and fall 1963-1984 in both regions) and the model's adequacy checked (see Box and Jenkins, 1976, Chap. 8). Table 1 contains summary statistics and parameter estimates for the four series. Since the series are relatively short, the average of the areal and seasonal estimates are used as the final estimates of θ and σ_e^2 (last line in Table 1).

Abundance indices for the two regions and seasons were calculated by applying to each series equation (4) with, $\hat{\theta} = .4$, the π-weights given by equation (11), and the $\hat{c}_t$'s (for each series) generated by model (12). An estimate of σ_e^2 equal to .20 and of σ_a^2 equal to .18 were obtained from equation (10). The estimated variance of the index equals, from equation (6), .12 for the current value and declines to .09 for values not near the series' endpoints. This compares with a variance of .20 ($= \hat{\sigma}_e^2$) for the original index. Figures 3 (log scale) and 4 (linear scale) show plots of the estimated index and the observed catch per tow series for the fall surveys off Southern New England.

DISCUSSION

The major advantage of estimating an index of abundance from the entire survey series is that it can produce an index which has a variance which is considerably smaller than the variance of the observed series. But the application also demonstrates that estimates of the accuracy of an index based only on the within survey sampling variance can be misleading. For example, the 1972 survey value for yellowtail off Southern New England is considered an anomaly (Collie and Sissenwine, 1983). It does appear anomalous if comparisons are made using .11, the estimated variance based on the within survey variance, but not if the estimate of .20 (= $\hat{\sigma}_e^2$) is considered (see Figure 3).

Assessing the accuracy of an index of abundance for marine stocks is difficult since the true levels are never known with certainty. But they can be compared with other indicators of abundance. The methods were applied to the haddock stock on Georges Bank (Pennington, 1985) for which a VPA exists. It was found that model (7) adequately describes the dynamics of the VPA series, and the survey series follows model (9). The resulting index of abundance is quite similar to the VPA estimates.

Collie and Sissenwine (1983) give a method for estimating the relative abundance of a fish stock using survey data and commercial catch statistics. They observe that their method produces estimates which compare favorably with VPA estimates. Figure 5 shows plots of Collie and Sissenwine's estimate of the relative abundance of Southern New England yellowtail and the index based only on the survey data.

Finally, it should be noted that the purpose of the modeling stage of the estimation procedure is not necessarily to develop a realistic model for the population, but to describe the important stochastic properties of the series. As the observed series becomes longer, more precise estimates can be made. For shorter series, given the large variability inherent in marine trawl surveys, a preliminary estimate of between .3 and .4 for the smoothing parameter θ appears to be an appropriate initial value to use for estimating an abundance index until more information becomes available.

LITERATURE CITED

Aitchison, J., and J.A.C. Brown. 1957. The Lognormal Distribution. Cambridge University Pres, London.

Box, G.E.P., and G.M. Jenkins. 1976. Time Series Analysis: Forecasting and Control. Revised edition, Holden-Day, San Francisco.

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.

Clark, S.H. 1979. Application of bottom trawl survey data to fish stock assessment. Fisheries 4: 9-15.

Clark, S.H., M.M. McBride, and B. Wells. 1984. Yellowtail flounder assessment update. Woods Hole Laboratory Reference Document No. 84-39. 29 p.

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.

Grosslein, M.D. 1969. Groundfish survey program of the Bureau of Commercial Fisheries, Woods Hole. Comm. Fish. Rev. 31: 22-35.

Pennington, M. 1983. Efficient estimators of abundance, for fish and plankton surveys. Biometrics 39: 281-286.

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

Pennington, M., and P. Berrien. 1984. Measuring the precision of estimates of total egg production based on plankton surveys. J. Plank. Res. 6: 869-879.

Pennington, M., and B.E. Brown. 1981. Abundance estimators based on statified random trawl surveys. Can. Spec. Publ. Aquat. Sci. 58: 2-7.

Table 1. Summary statistics and parameter estimates for the yellowtail survey series. The first three sample autocorrelations (r_1, r_2 and r_3) are for the first differenced logged series.

Survey		Number of years	r_1	r_2	r_3	$\hat{\theta}$	SE($\hat{\theta}$)	$\hat{\sigma}^2_c$
Southern New England	Spring	17	-.23	.12	-.18	.21	.28	.57
	Fall	22	-.26	.07	-.31	.40	.22	.71
Georges Bank	Spring	17	-.32	0.0	-.09	.61	.23	.36
	Fall	22	-.30	-.06	.18	.36	.23	.33
Average			-.28	.03	-.10	.40	.12*	.50

*Assuming the estimates of θ are independent.

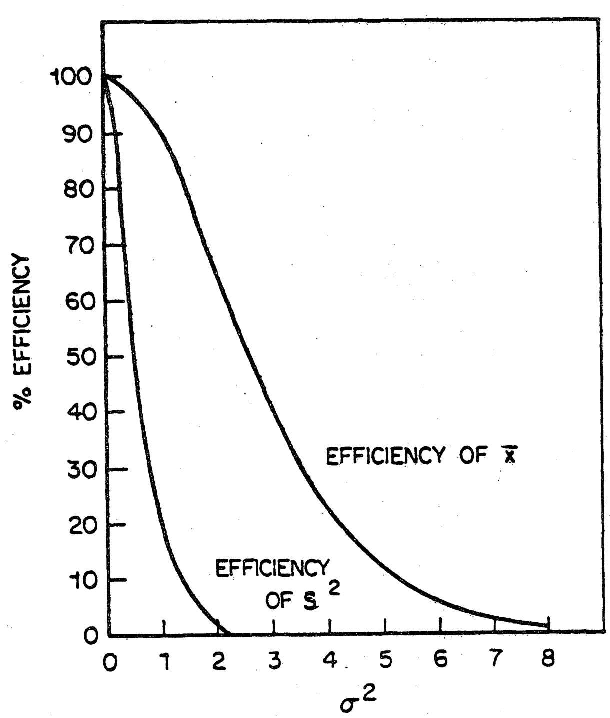

Figure 1. The efficiency of $\bar{x}$ and s^2 (the sample mean and variance, respectively) for the Δ-distribution with 50% zeros.

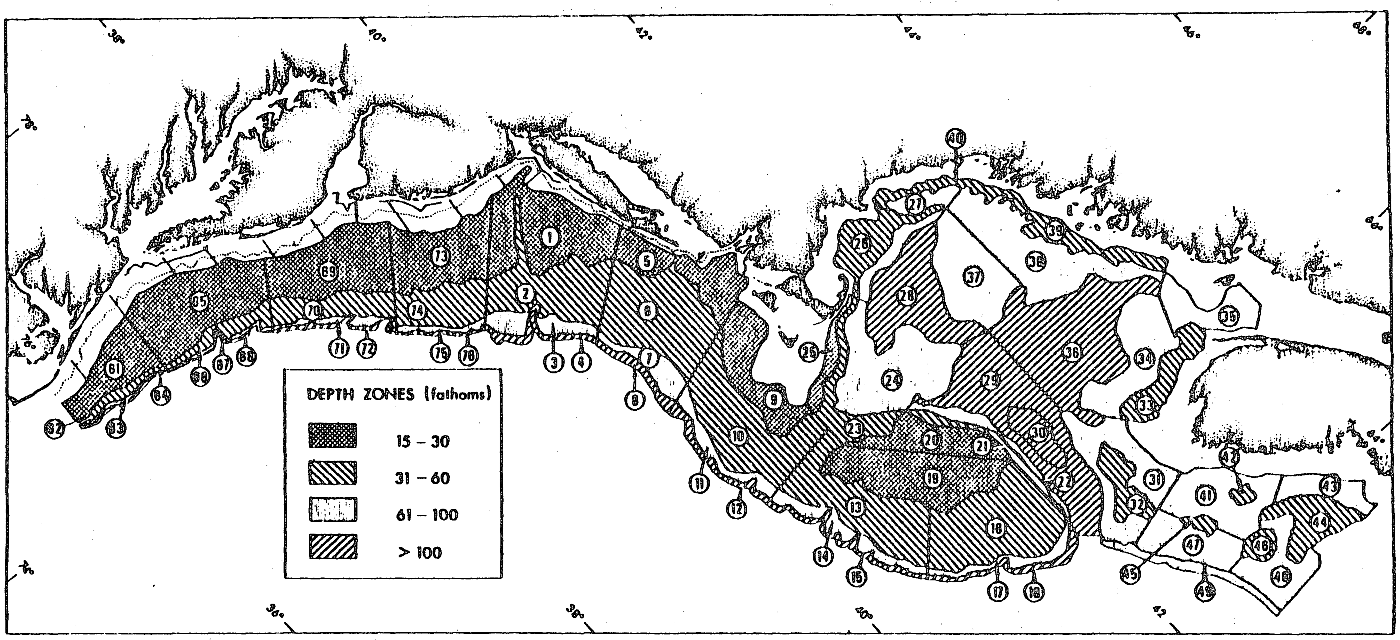

Figure 2. The National Marine Fisheries Service's MARMAP survey strata.

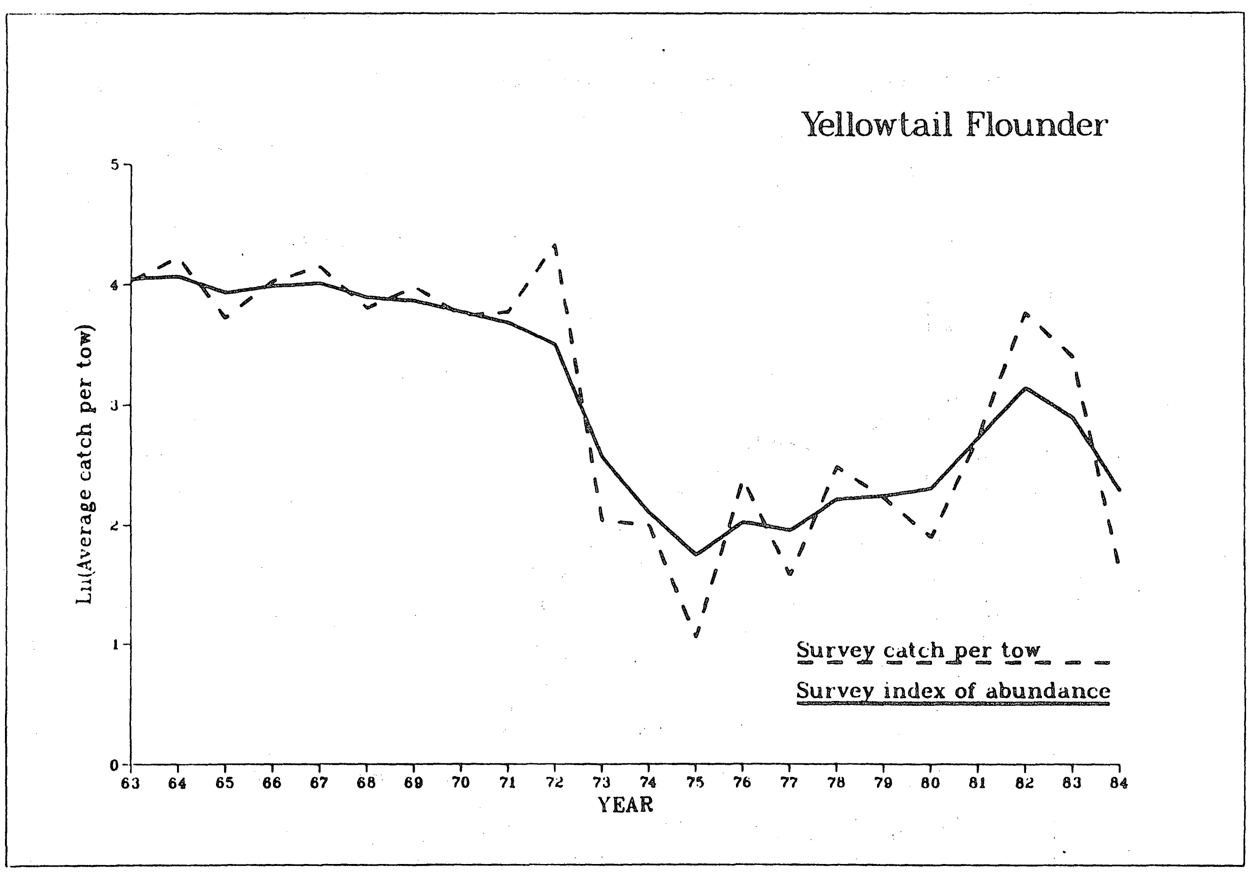

Figure 3. Logged average catch per tow and the estimated index of abundance for Southern New England yellowtail flounder.

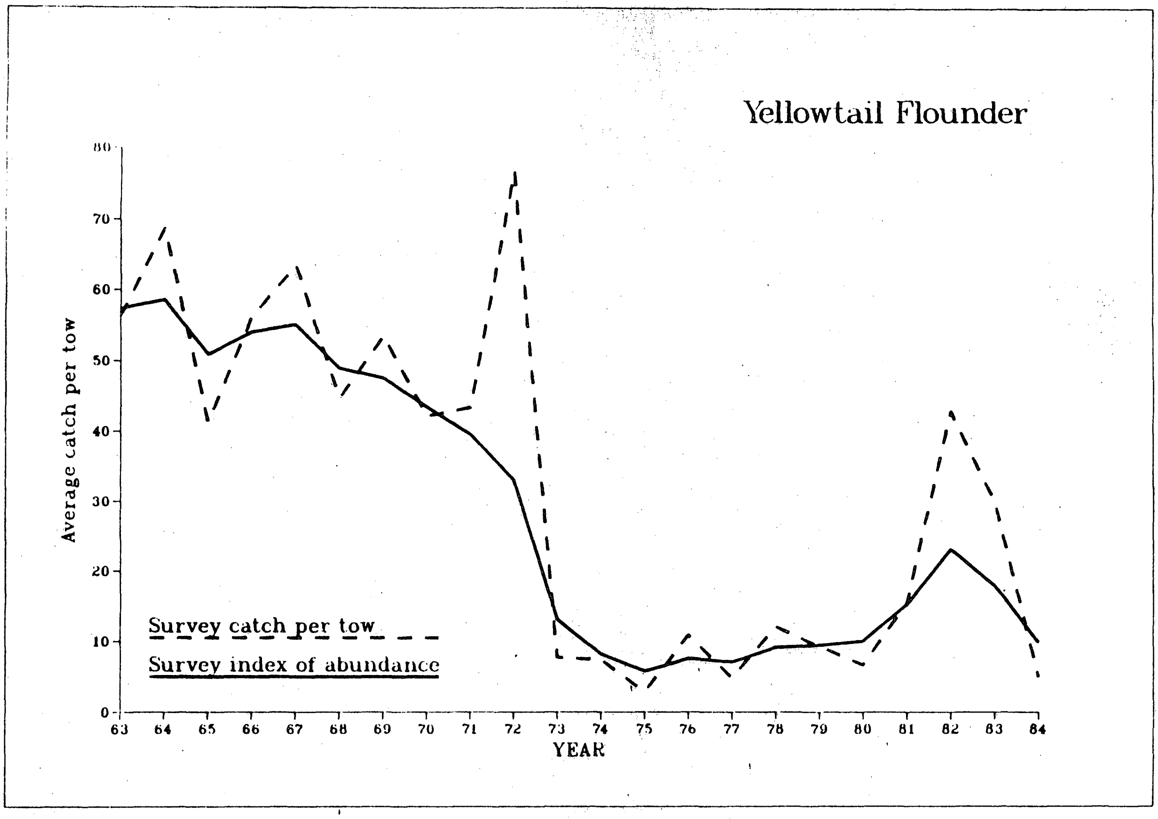

Figure 4. Average catch per tow and the estimated index of abundance for Southern New England yellowtail.

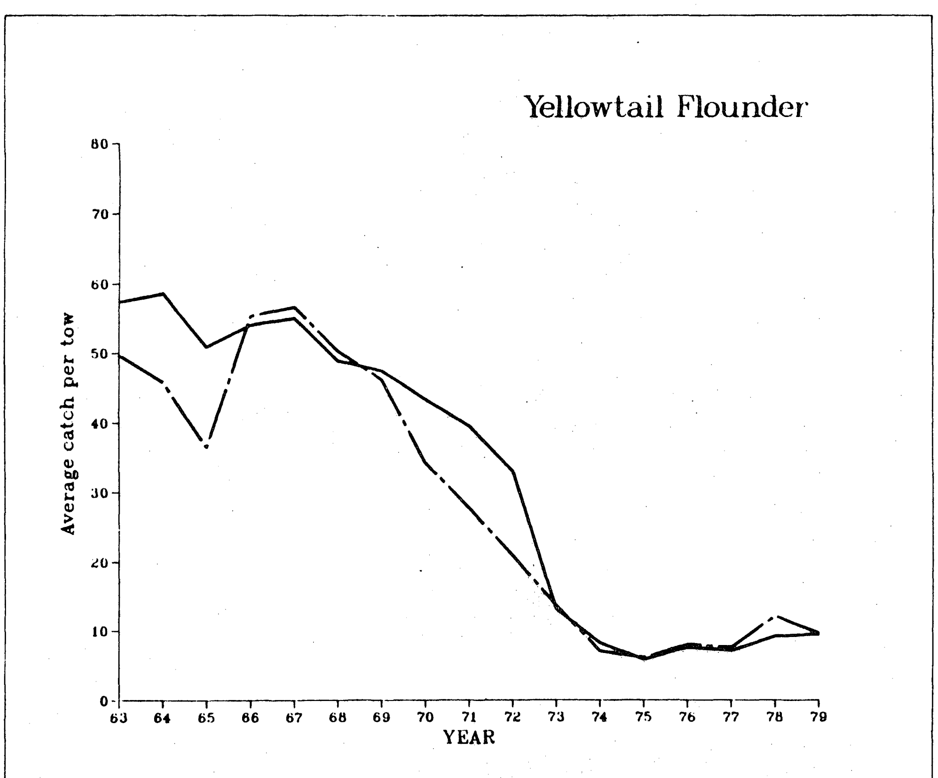

Figure 5. Survey index of abundance (solid line) and Collie and Sissenwine's index (broken line) for Southern New England yellowtail.