International Council for the Exploration of the Sea

C.M.1977/F:33
Demersal Fish (N) Committee
Ref: Pelagic Fish (N) Committee

AN ITERATIVE DERIVATION OF FISHING AND NATURAL MORTALITY FROM CATCH AND EFFORT DATA GIVING MEASUREMENTS OF GOODNESS OF FIT

by

D. F. Gray

Marine Fish Division
Bedford Institute of Oceanography
Dartmouth, N.S., Canada

ABSTRACT

This paper presents a method for estimating natural and fishing mortalities from catch-at-age data and catch-per-unit-effort data. This method is an iterative, two-phase approach that starts with approximations to the mortalities and corrects them at each iteration until the values become constant. Convergence of the method is examined empirically. Trouble using it on existing data is also discussed.

INTRODUCTION

Virtual population methods (Fry 1949; Murphy 1964; Jones 1964; Gulland 1965; Pope 1972) were developed to estimate fishing mortalities and numbers of fish at age from catch at age data only. In particular, they avoided the use of effort data which are difficult to obtain in a way that is consistent enough in time to be of use when studying a time series of catch data. In their papers presenting least squares methods, Pope (MS 1974) and Doubleday (1976), show that the original methods are equivalent to statistically fitting at least one parameter for each piece of data and do not allow any estimation of the error in fitting. To reduce this problem, Pope and Doubleday assume that the fishing mortality of age a fish in year n, ${}_aF_n$, is made up of an effort or time effect f_n and an age effect (selectivity, partial recruitment and availability) ${}_as$ so that ${}_aF_n = {}_as \cdot f_n$, ${}_as$ is constant over all years and f_n is constant over all ages. Even with these more restrictive assumptions the number of observations per parameter is usually quite low and the principal component analysis of the example in Doubleday (1976) shows that the method is quite unstable to changes in stock vs changes in mortalities.

Both the virtual population and least squares methods referenced above assume a given constant natural mortality M. This parameter must be estimated from other data or, as often happens, guessed at by using an accepted figure from some other similar stock. As is known, an error in M can lead to large errors in the number at age estimated from cohort analysis and the errors grow as the analysis is run. (See, for example, the detailed discussion in Ulltang (MS 1976) or briefly in Tables 1-3). If we attempt to fit M as well as ${}_a s$ and f_n using just catch data, the instabilities noted above are increased and it is difficult to get reliable estimates even using manufactured data containing little error. Hence some other data are needed, for example, catch per unit effort.

This paper presents a method for estimating natural and fishing mortalities from catch-at-age data and catch-per-unit-effort data. This method is an iterative, two-phase approach that starts with approximations to the mortalities and corrects them at each iteration until the values become constant. Convergence of the method is examined empirically. Trouble using it on existing data is also discussed.

ASSUMPTIONS AND NOTATION

The analysis of this paper is based on the standard catch equation of Beverton and Holt (1957). The model presented assumes natural mortality is constant over all ages and years though this can be generalized somewhat (see Conclusions).

The model assumes that catch per unit effort (CPUE) at age is a multiple of numbers at age. Thus, it is assumed that these numbers are adjusted for selectivity or availability. This assumption, crucial to the model, is discussed in the section Attempted Application of the Method. This model also assumes that numbers have been sampled to keep a constant coefficient of variation in the data.

Subscripted prefixes refer to age; subscripted suffixes refer to year.

M - natural mortality, constant over age and time

${}_a s$ - availability and selectivity part of fishing mortality

f_n - year effect part of fishing mortality

${}_a F_n = {}_a s \cdot f_n$ - fishing mortality at age a in year n

${}_a N_n$ - number of age a fish in year n

${}_a C_n$ - number of age a fish caught in year n

${}_aCPUE_n$	- catch per unit effort of age a fish in year n
exp	- exponential function
ℓn	- natural logarithm function

THE MODEL

In the above notation, the catch equation becomes:

$$ {}_aC_n = \frac{{}_aF_n \cdot {}_aN_n}{{}_aF_n + M} \cdot (1 - \exp(-{}_aF_n - M)) \qquad (1)$$

Since sampling is done to control the relative variance, it is more appropriate to fit the logarithm of ${}_aC_n$ than ${}_aC_n$ itself. However, the right-hand side of equation (1) is awkward to handle under a log transformation. From studying the coefficients in the Taylor series expansions it can be seen that

$$\frac{1 - \exp(-{}_aF_n - M)}{{}_aF_n + M}$$ is approximated closely by

$$\frac{-{}_aF_n - M}{2.25}$$. In fact the error is less than 2.0% if

$0 < {}_aF_n + M < 1.6$. However, this approximation must be used with more care than the cohort analysis approximation because the error rises exponentially outside this range (the cohort approximation error rises linearly). Some particular points are:

${}_aF_n + M$	Error %
.5	1.8
1.0	1.4
1.5	.9
2.0	4.9
3.0	16.8
4.0	31.1

With this approximation, the catch equation becomes:

$$ {}_aC_n = {}_aF_n \cdot {}_aN_n \cdot \exp \frac{-{}_aF_n - M}{2.25} \qquad (2)$$

and on taking natural logarithms we get:

$$\ell n\ {}_aC_n = \ell n\ {}_aF_n + \ell n\ {}_aN_n - \frac{1}{2.25}({}_aF_n + M) \qquad (3)$$

or, rearranging and using ${}_aF_n = {}_as \cdot f_n$

$$\ell n\ ({}_aC_n/{}_aN_n) = \ell n\ {}_as + \ell n\ f_n - \frac{{}_as \cdot f_n}{2.25} - \frac{M}{2.25} \qquad (4)$$

Before this equation is fitted to the data several other manipulations are necessary. It is assumed that numbers at age are not known but that catch per unit effort is and that:

$${}_aN_n = k \cdot {}_aCPUE_n$$

All the ${}_as$ and f_n are not independent so one must be set beforehand. If one ${}_as$ is set to 1, $\ell n_a s$ is zero for that age (let this age be $\bar{a}$). The equation to be fitted now becomes:

$$\ell n\ ({}_aC_n/{}_aCPUE_n) - \ell nk = \ell n_a s + \ell nf_n - \frac{{}_as \cdot f_n}{2.25} - \frac{M}{2.25} \qquad (5)$$

$$a \neq \bar{a}$$

To fit equation (5) we introduce dummy variables for age and year. The parameters we want to determine are then functions of the coefficients of these variables. We let:

$$I_{a^1} = \begin{cases} 1 \text{ if } a = a^1 \\ 0 \text{ else} \end{cases}$$

$$J_{n^1} = \begin{cases} 1 \text{ if } n = n^1 \\ 0 \text{ else} \end{cases}$$

For a regression problem the independent variable matrix must be non-singular and this is not true if dummy variables are used for all years. Hence, we cannot fit the equation quite as it stands. However, if we measure all f_n's as a proportion of $f_{\bar{n}}$ for a particular year $\bar{n}$, we can get around this problem. Let:

$$f_n = {}_nd_{\bar{n}} \cdot f_{\bar{n}}$$

Now ${}_{\bar{n}}d_{\bar{n}} = 1$, so $\ell n\ {}_{\bar{n}}d_{\bar{n}} = 0$ and the dummy variable for year $\bar{n}$ can be avoided. This problem does not arise for the dummy variables for ages since the dummy variable for age $\bar{a}$ never appears.

We are now left with fitting the following equation:

$$\ell n\ {}_aC_u/{}_aCPUE_n = \sum_{a \neq \bar{a}} \ell n\ ({}_as) \cdot I_a + \sum_{n \neq \bar{n}} \ell n\ ({}_nd_{\bar{n}}) \cdot J_n - \frac{{}_as \cdot f_n}{2.25} \qquad (6)$$

$$+ \left(\ell nf_{\bar{n}} + \ell nk - \frac{M}{2.25}\right) + \varepsilon$$

where, by assumption ε is $n(0, \sigma_1^2 + \sigma_2^2)$ where σ_1^2 is the variance of the errors in $\ln {}_aC_n$ and σ_2^2 is the variance of the errors in $\ln {}_a\text{CPUE}_n$. Since I_a and J_n are exact, there is no problem of errors in the independent variables. So far we have not considered the term $-\frac{{}_as \cdot f_n}{2.25}$. If we add the dummy variables to this term too, the problem becomes nonlinear and the parameters we are attempting to fit are included in more than one term. We avoid this by shifting this term to the left and putting in initially guessed values of ${}_as$ and f_n. Now we have a linear regression in which the constant term estimates $\left(\ln f_{\bar{n}} + \ln k - \frac{M}{2.25}\right)$ and the other parameters estimate the $\ln({}_as)$ and the $\ln({}_nd_{\bar{n}})$. Now we need to be able to split up the constant term to obtain $f_{\bar{n}}$ (and hence the f_n's) and M.

We now turn to the standard equation:

$$ {}_{a+1}N_{n+1} = {}_aN_n \cdot \exp(-{}_as \cdot f_n - M) \tag{7}$$

With the same manipulations as above, this becomes:

$$ \ln ({}_{a+1}\text{CPUE}_{n+1}/{}_a\text{CPUE}_n) = -({}_as \cdot {}_nd_{\bar{n}} \cdot f_{\bar{n}}) - M + \varepsilon^1 \tag{8}$$

${}_as$ and ${}_nd_{\bar{n}}$ are estimated from the fitting of equation (6). Hence equation (8) is a linear equation that can be used to estimate $f_{\bar{n}}$ and M. The error in the left hand side is, by assumption, $n(0, 2\sigma_2^2)$ but since ${}_as$ and ${}_nd_{\bar{n}}$ are only estimates there is also error in the independent variable. Hence (8) should be fitted using a geometric mean regression (see Ricker (1973)).

So the complete method is as follows: Guess initial values of ${}_as$ and f_n. Solve equation (6) with a standard linear regression routine. Use the estimated values for ${}_as$ and ${}_nd_{\bar{n}}$ to fit equation (8) and derive values for M and $f_{\bar{n}}$ (hence all the f_n). Now put the new values of ${}_as$ and f_n in (6) and repeat until the method converges.

Once finished, we have values for M, ${}_as$ and f_n and a value of k can easily be derived. There are now two ways to estimate ${}_aN_n$:

$$ {}_aN_n = k \cdot {}_a\text{CPUE}_n $$

or

$$ {}_aN_n = \frac{{}_aC_n \cdot ({}_aF_n + M)}{{}_aF_n \cdot (1-\exp(-{}_aF_n - M))} $$

Error bounds for the coefficients in the regressions yield approximate error bounds for the ${}_aF_n$ and M. The goodness of fit of the final parameters can be studied by considering the residuals in the regression equations and the two estimates of the ${}_aN_n$.

DISCUSSION OF CONVERGENCE AND STABILITY

To test convergence and stability of the model, a number of runs were made on generated data. Data were generated from the standard catch equation with an assumed M and ${}_as$'s and with f_n picked randomly between .1 and 1.2 and recruitment picked randomly between 10 thousand and 10 million. This variation, especially since the fluctuations may be erratic, is probably much more extreme than found in most applications. Numbers were produced for 11 ages and 13 years and catches for the first 10 ages and 12 years. Each value was then multiplied by the exponential of a number picked from a normal distribution with mean zero and a selected standard deviation. This reflected the assumption that the numbers have a constant coefficient of variation. Twenty examples were run at standard deviation values of .1, .2, .4, and .6. In all examples k = 1 was used.

For each example, the method needed about six iterations before the third decimal of each of the ${}_as$, f_n, and M became constant. Since the matrix of independent variables in the first phase never changes, the matrix inverses and products involving just these values can be calculated for all iterations. The second regression has only two parameters so is equally quick. Hence computationally the method is quite efficient.

In other methods tried, especially if just catch data were used, it was difficult to obtain a stable value of M and small amounts of error often led to unrealistic negative values. This does not happen in the model presented and estimated M values are usually quite reasonable. Figure 1 shows the mean and one standard deviation for the estimated M values at a number of levels of error. For smaller error values (standard deviations up to .2) ${}_as$ and f_n are estimated very closely. Some of these values start to diverge from the input values at higher error levels, but this usually was more marked when f_n changed radically from year to year.

It would appear that, if all assumptions are met, this method will yield efficiently, fairly unbiased values for natural and fishing mortalities.

ATTEMPTED APPLICATION OF THE METHOD

A number of attempts were made to apply this method to existing data. All attempts ended with negative estimated M's though estimated numbers at age were usually reasonable. Obviously not all assumptions used in the model hold in practical examples. There may be a problem with sampling not controlling relative variance but it is more likely that the problem is with using effort data. In the attempts made, both research cruise catch per unit effort and commercial catch per unit effort were tried. In both cases, it is necessary to adjust for selectivity and/or availability. It is possible that the adjustments were incorrect but the biases were consistently in one direction whatever adjustments were attempted.

A brief study was conducted to determine in what way the CPUE figures were inadequate. For generated data, the correlation between the numbers at age with and without error was determined. This correlation may be exaggerated by the trend in numbers with age (lots of young fish, few old fish) which swamps other variation. To remove this problem, the mean for each age was subtracted from data of that age and the correlation coefficient re-derived. The error was introduced as discussed above and 20 runs were made with a standard deviation of .4 and 20 with a standard deviation of .6. These examples were compared with what was thought to be a set of good effort data and the VPA numbers from the same stock. Results were as follows:

Correlation of Numbers at Age vs Estimated Numbers

	Unaltered	Age trends removed
Standard deviation of error .4	mean .914 S.D. .022	mean .817 S.D. .074
Standard deviation of error .6	mean .785 S.D. .027	mean .619 S.D. .062
Effort data vs VPA numbers for a 'good' example	.800	.262

In this case, the effort data do not show the detailed variation adequately and the correlation in the unaltered data shows only the age trend.

The model presented does not necessarily require good effort data. What are needed are some data that give an adequate indication of changes in numbers at age from year to year. Assumptions that lead to the equation

$$ {}_aN_n = k \cdot {}_aCPUE_n \qquad (9) $$

include that of a random distribution of the stock over its range. Clustering of stock will cause bias in the estimates and some method to derive numbers at age under such conditions is needed.

Problems comparing gears and controlling for efficiency changes make it difficult to derive adequate commercial effort data. Research cruise CPUE data would also be more independent of the commercial catch data and thus the derived estimates should be more reliable. However, ways must be derived for correcting research CPUE for fish clustering. What may be adequate is an index that takes in numbers of clusters (derived from search time), size of clusters (how do we derive this?), and distributions of fish in a cluster by age (derived from a trawl through the cluster).

CONCLUSIONS

Catch data alone are inadequate for the derivation of the number of parameters needed for the assessment of a fish stock. In order to get an unbiased estimate of natural and fishing mortalities, at least some other type of independent data is needed. The model in this paper is designed to use catch and effort data. However, in practice, difficulties were encountered because the effort data did not adequately reflect the changes in year-class size. This might be because of the way in which catch-per-unit effort was derived but more likely is because the assumptions behind equation (9) do not hold. For this model to work, it is necessary that an index of numbers at age be derived that takes into account the distribution of fish in clusters. Any other model that will give adequate values for all needed parameters, will also need some good data independent of catch at age. It is possible that models can be designed that use data that are easier to collect than the index discussed above. These detailed data are needed to separate the effects of F and M, otherwise F and M can drift in opposite directions and their exact values cannot be determined precisely. The method suggested will generalize somewhat to more complicated expressions for natural mortality such as $M = M_1 + M_2 \cdot a$ or $M = M_1 + M_2 \cdot n$. However, if it is desired to derive values of natural mortality at age (${}_aM$), much more detailed data will be needed to separate the effects of natural mortality from selectivity and availability. The model in this paper is presented mainly for discussion. Any ideas concerning ways to correct the biasing problem encountered would be appreciated.

REFERENCES

Beverton, R. J. H., and S. J. Holt. 1957. On the dynamics of exploited fish populations. Fish. Invest., London, (2) 19.

Doubleday, W. G. 1976. A least squares approach to analysing catch at age data. Res. Bull. Int. Comm. Northwest Atl. Fish. No. 12, p. 69-81.

Fry, F. E. J. 1949. Statistics of a lake trout fishery. Biometrics 5: 27-67.

Gulland, J. A. 1965. Estimation of mortality rates. Annex to Arctic Fisheries Working Group Report. ICES, C.M.1965, Doc. No. 3 (mimeo).

Jones, R. 1964. Estimating population size from commercial statistics when fishing mortality varies with age. Rapp. Cons. Explor. Mer. 155(38): 210-214.

Murphy, G. I. 1964. A solution of the catch equation. J. Fish. Res. Board Can. 22: 191-201.

Pope, J. G. 1972. An investigation of the accuracy of virtual population analysis using cohort analysis. Res. Bull. Int. Comm. Northwest Atl. Fish. No. 12: 65-74.

1974 MS. A possible alternative method to virtual population analysis for the calculation of fishing mortality from catch at age data. Annual Meeting Int. Comm. Northwest Atl. Fish. 1974, Res. Doc. No. 20, Serial No. 3166 (mimeo).

Ricker, W. E. 1973. Linear regressions in fishery research. J. Fish. Res. Board Can. 30: 409-434.

Ulltang, Ø. 1976 MS. Sources of errors in and limitations of virtual population analysis (cohort analysis). Coun. Meeting ICES 1976 (H:40), p. 1-27 (mimeo).

Table 1. Generated Numbers using Catch Equation. M=.3, selectivity given, fishing mortality and recruitment random.

Age	Year 1	2	3	4	5	6	7	9	9	10
1	9050000	5170000	9870000	2670000	9480000	5010000	2780000	5300000	9410000	7620000
2	6674368	6251146	3679853	6886066	1938818	6953078	3530488	1844948	3697685	6901736
3	2370484	4610215	4449380	2567347	5000310	1422021	4899751	2343009	1287177	2712056
4	1234831	1327230	2910354	2592869	1755709	3559073	862499	2337740	1365385	916174
5	405393	560422	743114	1416621	1669902	1212729	1857999	295845	1137900	943118
6	604906	159950	289655	320811	876582	1130619	572853	511453	127720	770423
7	99690	222533	79429	117765	194582	587591	508020	141264	207942	85613
8	97659	36674	110507	32294	71428	130432	264022	125277	57434	139388
9	46648	35927	18212	44929	19587	47880	58607	65107	50934	38499

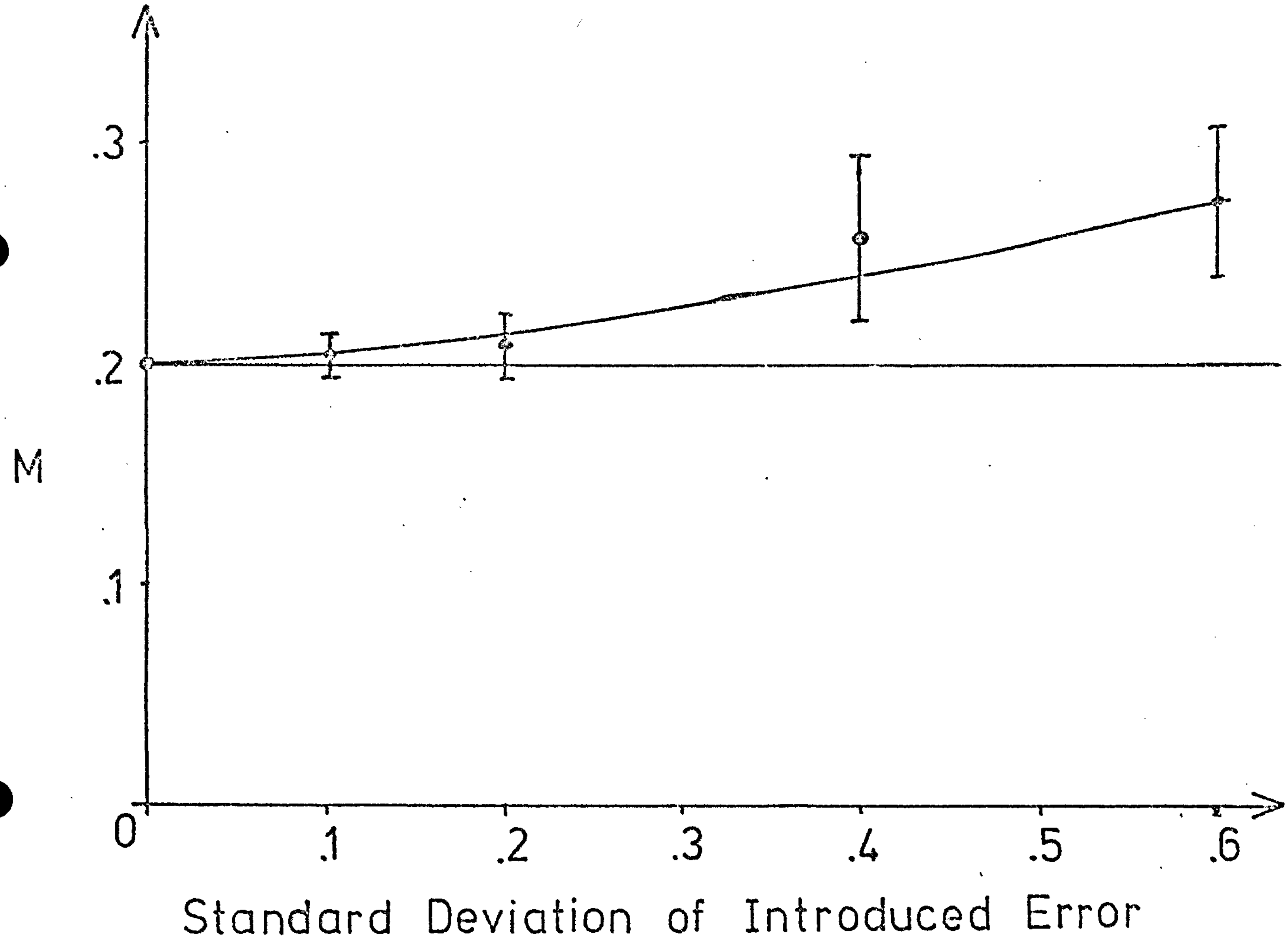
Table 2. Numbers from Cohort Aanlysis; used correct starting F's but M=.2.

Age \ Year	1	2	3	4	5	6	7	8	9	10
1	5864443	3305172	6588455	1833773	6628078	3556279	2075095	4207403	8131468	7269497
2	4685735	4322277	2547418	4944175	1460017	5352855	2720390	1471950	3203091	6584271
3	1847953	3483002	3346981	1917876	3941304	1180276	4117129	1938993	1121015	2593701
4	1026969	1057264	2316650	1993680	1415614	3073445	763818	1992429	1194362	878316
5	349494	462313	610463	1108289	1366562	1066056	1687527	256388	997961	905575
6	532466	135696	244874	254158	724147	1006240	525423	446575	112117	740329
7	89016	193757	69466	96798	162401	527509	471745	123345	182539	82269
8	89723	32967	100712	28441	62480	118451	248902	114123	50418	133943
9	44721	34359	17446	42898	18686	45828	56361	62367	48591	36995

Table 3. Percent error in cohort numbers.

Age	Year 1	2	3	4	5	6	7	8	9	10
1	35.2	36.1	33.2	31.3	30.1	29.0	25.4	20.6	13.6	4.6
2	29.8	30.9	30.8	28.2	24.7	23.0	22.9	20.2	13.4	4.5
3	22.0	24.5	24.8	25.3	21.2	17.0	16.0	17.2	12.9	4.4
4	16.8	20.3	20.4	23.1	19.4	13.6	11.4	14.8	12.5	4.1
5	13.8	17.5	17.9	21.8	18.2	12.1	9.2	13.3	12.3	4.0
6	12.0	15.2	15.5	20.8	17.4	11.0	8.3	12.7	12.2	3.9
7	10.7	12.9	12.5	17.8	16.5	10.2	7.1	12.7	12.2	3.9
8	8.1	10.1	8.9	11.9	12.5	9.2	5.7	8.9	12.2	3.9
9	4.1	4.4	4.2	4.5	4.6	4.3	3.8	4.2	4.6	3.9

Figure 1. Estimated M for generated data vs standard deviation of added error.



Mean of 20 runs.

One standard deviation.

Input value of M was .2.