J. Cons. int. Explor. Mer, 47: 284–294. 1991

Multiplicative modelling of catch-at-age data, and its application to catch forecasts

J. G. Shepherd and M. D. Nicholson

Shepherd, J. G., and Nicholson, M. D. 1991. Multiplicative modelling of catch-at-age data, and its application to catch forecasts. – J. Cons. int. Explor. Mer, 47: 284–294.

A natural interpretation of fish catch-at-age data leads to an approximate multiplicative model with three factors: year, age, and year class. These factors are however, interrelated (year class = year − age) and estimates of the parameters of such a model are not unique. Specific solutions may be obtained by imposing biologically meaningful constraints on the parameters: for example, by specifying the trend in the year effect. The choice of error distribution and its parameters is also important in fitting such models. Examination of the sampling procedures used in data collection suggests a simple approximate formulation for the error variance of the log-transformed data. The model may be fitted by standard least-squares methods, or by a simpler calculation based on log-catch ratios. The model may be applied to any coherent set of catch-at-age data, representing the total international catch, that of a single fleet, or a research survey. The method is independent of VPA, but conceptually closely related to separable VPA. It may be used to estimate the steady-state age composition (i.e. a corrected catch curve), which is a required starting point for some assessment procedures, and also to estimate relative year-class strength for all year classes represented in the data, even those only present as older ages in early years. The fitted parameters may also be used as the basis of a simple forecast of catch-at-age for the data set to which it has been fitted.

J. G. Shepherd and M. D. Nicholson: Ministry of Agriculture, Fisheries and Food, Directorate of Fisheries Research, Fisheries Laboratory, Lowestoft, Suffolk NR33 0HT, England.

1. Introduction

Shepherd and Nicholson (1986) pointed out that much of the variance of catch-at-age data can usually be explained by a simple multiplicative (log-linear) model, with three factors representing ages, years, and year classes. This merely formalizes the conventional wisdom of fisheries science: that catch-at-age is primarily determined by year-class strength, the overall level of fishing effort in each year, and the combined effect of selection and survival as a function of age. The use of a formal statistical model is useful, however, as it focuses attention on the necessity for appropriate fitting techniques, preferably taking account of the likely error structure of the data.

In this paper the development of such models is discussed with emphasis on the practical aspects rather than the statistical ones. The application to commercial catch-at-age data is considered, and an alternative very simple method of parameter estimation (based on log-catch ratios) is proposed. The significance of the parameters of the model fitted is discussed, and the relationship with VPA (especially separable VPA) is explored. Finally, the use of the parameter estimates as the basis for a simple forecast of catch-at-age is described.

2. Basis of the model

2.1. Derivation of the multiplicative model

For many fish stocks subject to routine assessment, data are available for catch numbers by age group for several years for both commercial fisheries and research surveys. The standard theory of fishing (e.g. Gulland, 1983) interprets these as the product of an instantaneous fishing mortality rate F_{ya} and the average size of the population $\bar{P}_{ya}$ at each age (a) during the year y.

A major determinant of the population size for most marine fish is the year-class strength (i.e. cohort size), which is highly variable. This is denoted here by R_k for the (k)th year class, where y, a, and k are of course related linearly.

Fish are subject to a natural instantaneous mortality rate M_a (conventionally assumed constant with time) as well as the total fishing mortality F_{ya}. The total instantaneous mortality rate is then

$$Z_{ya} = M_a + F_{ya} \tag{1}$$

The average population size in year y is given – to a good approximation – by

$$\bar{P}_{ya} = x_{ya} R_k \tag{2}$$

where x_{ya} is the fraction surviving to age a in the (y)th year and is given by

$$x_{ya} = \exp[-\mathrm{cum}(Z_{ya})] \tag{3}$$

where $\mathrm{cum}(Z_{ya})$ is the cumulative mortality from recruitment to the mid-point of the final fishing season for the cohort in question.

Pope and Shepherd (1982) have pointed out that F_{ya} may often be expressed approximately as a product of year and age effects,

$$F_{ya} = \bar{F}_y S_a \tag{4}$$

where $\bar{F}_y$ is a measure of overall fishing mortality and S_a is the selection of fish of age a. Thus,

$$C_{ya} = \bar{F}_y R_k S_a x_{ya}$$

This has the form of a multiplicative model involving year, age, and year-class effects, and the interaction of age and year effects. Models of this form, retaining the interaction term, have been studied by Pope and Stokes (1989). Here we concentrate on the simplest possible form, by noting that if fishing mortality were constant, the interaction term would reduce to a simple age effect. Assuming that this is a sufficient approximation, and therefore writing

$$S'_a = S_a x_{ya}$$

for the overall age effect, expressing the effects of both selection and survival, we obtain

$$C_{ya} = \bar{F}_y R_k S'_a \tag{5}$$

Equation (5) is exact only if the fishing mortality is separable and the overall fishing mortality is constant over the period considered. If either of these conditions is violated, it becomes an approximation only. Nevertheless, the results of Pope (1979, 1983) imply that the errors involved should not be of major significance in normal circumstances – most of the variance of the data may be explained when the three main effects are fitted.

The development of the simple model of Equation (5) has been expressed in terms of total fishing mortality, but it is easy to see that it also applies for the partial fishing mortality of any individual fleet or survey, although it is again only exact if the *total* fishing mortality and exploitation pattern remain constant. The model may therefore be fitted to partial data sets when complete data are lacking. This is a considerable practical advantage over VPA-like procedures, which require complete total international catch data.

Whether total or partial data are fitted, the results will be imprecise (biased) if either the total fishing mortality or the exploitation pattern vary with time. This would lead to a non-zero year/age interaction (x_{ya} becomes a function of year and age, not age alone), which could be large enough to be of practical significance. Whether or not the resulting biasses of the parameters estimated are serious in comparison with other possible biasses and the inevitable variance of the results will depend on the nature of the data. This could probably best be studied by simulation tests, and is an interesting subject for further work. More complete models, allowing for year/age and other interactions, have also been applied by Pope and Stokes (1989). In the present case any such unfitted interaction terms will also add to the residual error variance.

The results of fitting the model (Equation (5)) are estimates of the parameters R_k representing relative year-class strength, S'_a representing the combined effect of selection and survival, and $\bar{F}_y$ representing relative fishing mortality, either partial or total, depending on the data set in question. Note that S'_a is in fact an estimate of the steady-state age composition of the catch – a 'catch-curve' corrected for fluctuations of year-class strength and (to a first approximation) for modest changes of fishing mortality. This may be further analysed (e.g. by cohort analysis) to give some estimate of average total fishing mortality. All this information may be obtained, to what seems in practice to be a useful degree of approximation, from a single partial data set, including that from a survey series, provided *total* fishing mortality and exploitation pattern do not vary too much.

Much of fish stock assessment revolves around estimating and using these parameters, often employing complicated and circuitous procedures. The assessment and subsequent forecasting procedures should be clearer and simpler if the model (5) is fitted directly to the data. This also permits the estimation of R_k for early year classes not fully represented in the data, which is sometimes of interest.

It is well known, however, that not all of the relevant parameters can be determined from catch data alone (Pope and Shepherd, 1982). In the formulation of Equation (5), the linear model obtained by taking logarithms of catches-at-age is in fact indeterminate because of the relationship between year, year class, and age. This indeterminacy may be removed, however, by imposing biologically meaningful constraints on the parameters, as discussed below.

The model can be linearized by taking logarithms

$$\ln(C_{ya}) = \mu + f_y + r_k + s_a \tag{6}$$

where the lower case f_y, r_k, s_a denote the logarithms of their uppercase equivalents, apart from some additive constants collected in a constant term (μ). These factors (f, r, and s) may be defined subject to the usual arbitrary but convenient normalization

$$f_1 = r_1 = s_1 = 0 \tag{7}$$

and are here subscripted with $a = 1$ for the youngest age, $y = 1$ for the first year, and $k = 1$ for the earliest year class, so that

$$k = y - a + m \tag{8}$$

where m is the number of age groups, excluding the plus group (if any), which is not treated by this model.

2.2. Indeterminacy

Equation (6) may be fitted to the data by any of several least squares procedures, including the use of standard statistical packages such as GLIM (Baker and Nelder, 1978) or by a simple method described below. The problem is structurally aliased, however, so the solution is not unique.

The form of this indeterminacy may easily be demonstrated. Suppose a solution for μ, f_y, r_k, and s_a has been found. Now $k = y - a + m$, and therefore $(y - a - k + m)$ is zero, and any multiple of it may be added to the fitted value without altering the goodness-of-fit. Adding (say) δ times this null expression to both sides of (6) then leads to

$$\ln(C_{ya}) = \mu + (m-1)\delta + f_y + (y-1)\delta + r_k - (k-1)\delta + s_a - (a-1)\delta$$

There are thus an infinite number of equally good solutions corresponding to the parameter values

$$\mu + (m-1)\delta$$

$$f_y + (y-1)\delta$$

$$r_k - (k-1)\delta$$

$$s_a - (a-1)\delta$$

These differ from the original solution to the extent of an arbitrary additive linear trend $(0, \delta, 2\delta, 3\delta \ldots)$ on the effects for each factor.

This indeterminacy may be removed by applying suitable additional constraints on the parameters. In general, appropriate constraints may be found by examining the estimable functions of the parameters (Shepherd and Nicholson, 1986). The commonsense approach, however, is that any constraint on the overall trend of the effects for any factor will serve: one must in effect apply a constraint which will fix the value of δ.

Various possibilities exist. In practice, we have found that the most generally useful form of the constraint on the year effect (f_y) is to specify that its trend is equal to some value, g, determined perhaps from the analysis of effort data, or assumed to be zero.

The easiest way to do this is simply to require that

$$f_n = f_1 + (n-1)\,g$$

where n is the number of years. This is particularly easy to implement when fitting using log-catch ratios as discussed below. Standard statistical packages do not usually allow for the convenient specification of additional constraints on the parameters when fitting categorical ("analysis of variance") models such as this. They can usually be tricked into doing what is required, however, by supplying an additional imaginary data point corresponding to an infeasible combination of y, k, and a, e.g. $y = 1$, $k = 1$, $a = 1$. This has the effect of fixing the grand mean, and thus the value of g: the required value for the spurious data point corresponding to a given trend can be found by trial and error.

In special circumstances it might also be of interest to examine solutions for a specified (e.g. zero) trend in the year class (recruitment) effect, but this would be of less general applicability, since recruitment trends are not generally known *a priori*, nor can they often be assumed to be zero with any confidence.

3. Identification of variance structure

In this section we propose an inessential but useful embellishment of the model, being an approximate allowance for the error structure of the data. This is a problem of much wider occurrence, usually ignored, often because it is tricky to handle. In a directly fitted model such as that discussed here, there is less difficulty in doing something sensible. We would however suggest that the method discussed below could find wider application in the analysis of catch-at-age data.

The processes leading to errors in catch-at-age data are quite complicated, and the error structure is difficult to specify precisely. In the first place there are process errors, due to real but unpredictable variations in the fishing process itself or the behaviour of the fish being caught. In addition, there are sampling errors, although in practice these cannot normally be separated from the process errors. These arise because only a sample of the catch is usually measured, and a further subsample is taken for age determination (actually this is often not even a true subsample). It remains true, of course, that the larger numbers of fish observed are more precisely determined, but because of the multi-stage nature of the sampling, the extensive use of subsampling (and therefore raising factors), and the conversion of length to age, it is very difficult to set down a precise formulation. In addition, the samples taken for measurement tend in practice to be of more or less fixed maximum size (in weight or volume, i.e. a basket or a bucket of fish), so that the improvement in precision as abundance increases in fact reaches some sort of limit.

It is common practice when analysing catch-at-age data to ignore all this, and either to make unweighted fits to the untransformed data (implicitly treating the data as though it had constant normal additive errors) or to log-transformed data (implicitly assuming that it has constant multiplicative errors – a constant coefficient variation). Clearly, neither assumption is likely to be realistic. It has often been remarked that unweighted fits to untransformed data of this type are dominated by the large numbers, whilst unweighted fits to log transformed data are dominated by the small numbers. Something in between these extremes is really required. Attempts have been made to allow for other error structures of catch data (see, for example, Fournier and Archibald, 1982; Deriso, Quinn, and Neal, 1985) but even these do not fully allow for the gory complexity of the real situation. We propose a rather rough and ready approach, which ignores the shape of the error distribution (thus implicitly treating it as lognormal) but attempts to allow for the varying precision of the data as abundance changes. According to Gilchrist (1984), this should capture the most important features of the problem.

The argument is given in detail by Shepherd and Nicholson (1986), but the idea is quite simple. If fish of a certain age (or size) are scarce, all those present in a sample are likely to be counted, measured, and (very likely) aged. If, on the other hand, they are abundant, they will be subsampled, and only one box, basket, or bucketful will be measured. In the first case the variance of the number caught should be roughly proportional to the mean number caught (not equal to it, because there is almost always some raising factor involved in producing the final catch-at-age tables). The coefficient of variation is therefore inversely proportional to the square root of the mean. In the second case the coefficient of variation remains fixed at whatever level is appropriate for the subsample size, since further increases of abundance do not lead to more fish being counted and measured, but to a smaller proportion of the catch being sampled, and a larger raising factor being applied to the observations (and multiplication preserves coefficients of variation).

Thus, as abundance increases, we should expect the coefficient of variation of the numbers estimated to fall at first, but beyond some threshold abundance to stabilize at a constant level. This behaviour may be allowed for by weighting the residuals in the fitting procedure appropriately. Since we use a logarithmic transformation to convert a multiplicative model into a linear one and a constant coefficient of variation corresponds (approximately) to a constant logarithmic variance, the appropriate form of weighting is (using inverse variance weights)

$$
\begin{aligned}
w &= C \; (C < N) \\
w &= n \; (C \geq N)
\end{aligned}
\qquad (9)
$$

where N is the appropriate threshold abundance (in final raised units) above which subsampling will have been implemented.

Clearly, such a weighting scheme accords with commonsense, and interpolates nicely between the two extreme regimes outline above – it could be described as "asymptotically sensible". Note that the scaling of the weights is completely arbitrary and of no importance – there is no need to agonize over the correct constant of proportionality. Also, zero catches get zero weight, removing a perennial problem with logarithmically transformed data.

The remaining problem is the choice of an appropriate value of n. This is tricky. The true value relates to something like the probable number of average sized fish in a box, basket, or bucket, within a factor of a few, but is thereafter usually veiled in a mist of raising factors and combinations of data sets. In practice, we suggest that it should be sufficient to guess a value which separates abundant age groups from scarce ones – something like the median of the data as presented, as an order-of-magnitude estimate. The abundance estimates usually span several orders of magnitude at least, so this should be sufficient.

Finally, we remark that this weighting is no more than an embellishment, and by no means a necessary part of the model. We feel that it is a useful practical procedure, but it may be omitted if it offends. However, one should then choose carefully between omitting it by letting N tend to zero (thereby accepting a constant coefficient of variation), or letting it tend to infinity (thereby accepting a variance proportional to the mean)!

4. Fitting using log-catch ratios

The model may be fitted using a general purpose statistical package, or standard subroutines for least squares solutions to overdetermined linear equations. In addition, it may be fitted using row and column summations of log-catch ratios, in the same way that ANOVA models may be fitted using row and column means, provided that the catch-at-age matrix is complete. A similar procedure is used by Separable VPA (Pope and Shepherd, 1982) which is closely related.

Consider the log-catch ratio

$$d(y,a) = \ln(C_{ya}) - \ln(C_{y+1,\,a+1}) \approx$$
$$\approx f_y - f_{y+1} + s_a - s_{a+1} \tag{10}$$

since the terms involving the mean and the year-class effect cancel out. Then

$$\mathrm{sum}_y\{d(y,a)\} = f_1 - f_n + (n-1)(s_a - s_{a+1}) \tag{11}$$

where n is the number of years in the data set. And

$$\mathrm{sum}_a\{d(y,a)\} = (m-1)(f_y - f_{y+1}) + s_1 - s_m \tag{12}$$

However, we require $f_n - f_1 = (n-1)g$ to control the indeterminacy and thus

$$s_{a+1} = s_a - \mathrm{sum}_y\{d(y,a)\}/(n-1) - g \tag{13}$$

and

$$f_{y+1} = f_y - [\mathrm{sum}_a\{d(y,a)\} - s_1 + s_m]/(m-1) \tag{14}$$

Given that $s_1 = 0$, because of the choice of normalization, and g is specified, Equation (13) may be applied repeatedly to determine all the other values of s_a including s_m, and then Equation (14) may similarly be used repeatedly to determine all the values of f_y, starting from $f_1 = 0$.

This simple procedure takes no account of any weighting of the residuals in the fitting process, but has the advantage that the computation is simple enough to be carried out using a spreadsheet program if desired. The lack of weighting has only a small effect on the calculation of f and s, because these are determined mainly by ratios of catches for adjacent ages and years, which tend to have similar weights. This is not so when one comes to determine the values of r_k which depend on a wide range of ages and years. The value of r_k which minimizes the weighted sum of squares of log residuals between estimated and observed catches, i.e.

$$SSQ = \Sigma\Sigma W_{ya}\{\ln C_{ya} - \mu - r_k - f_y - s_a\}^2 \tag{15}$$

is easily shown to be

$$r_k = \Sigma W_{ya}(\ln C_{ya} - \mu - f_y - s_a)/\Sigma W_{ya} \tag{16}$$

where the summation extends over all age groups in each cohort, and the weights are those given by Equation (9). The constant μ is chosen so that $r_1 = 0$, in accordance with the normalization condition. This is easily done by calculating the r_k using Equation (16) with μ set to zero, and then setting $\mu = r_1$, and subtracting μ from all the values of r_k. Finally, the sum of squares of residuals (15) may be calculated and used to give an estimate of the lack of fit of the model to the data

$$CV = \sqrt{(SSQ/\Sigma\Sigma W_{ya})} \tag{17}$$

which would be an estimate of the coefficient of variation of the data if all the lack of fit was due to measurement error and none to model and process error (it is therefore an upper limit for the measurement (sampling) error). The standard errors of the parameters can also be estimated together with their covariance matrix, but the interpretation of these is not straightforward and will be dealt with elsewhere (Nicholson, to appear).

5. Catch forecast

The estimated parameters of the model may be used to forecast future catches, using the model equation for year y = n + 1, n + 2, etc.

$$C_{n+1,a} = \exp\{\mu + f_{n+1} + s_a + r_k\} \tag{18}$$

Table 1. North Sea cod: total international catch numbers: 1976–1985.

Year / Age	1	2	3	4	5	6	7	8	9	10
1	5 145	58 279	26 368	35 319	59 344	20 416	61 191	23 395	62 720	8 086
2	90 263	45 947	156 479	86 133	98 856	177 309	56 340	118 144	60 215	109 944
3	16 485	22 823	13 358	39 843	29 578	26 739	50 002	16 932	27 801	15 909
4	6 721	4 300	8 386	3 584	9 988	7 352	6 639	9 869	3 493	6 745
5	1 661	2 099	2 850	3 188	1 595	3 829	3 002	2 584	3 126	1 179
6	2 746	757	980	713	1 164	757	1 769	1 235	956	1 104
7	836	1 029	383	371	411	571	333	575	413	319
8	120	335	376	131	191	135	204	142	233	158
9	59	238	141	145	71	65	68	83	57	70
10	57	23	33	39	54	37	23	22	43	12
11	22	9	15	2	18	17	10	16	13	18
12	16	43	22	13	6	1	5	2	4	2
13	1	35	2	0	0	3	0	0	0	0

To complete these estimates, values of f_y and r_k must be supplied for any future years and year classes, as for any catch forecast – and, similarly, one or more of the most recent estimates of year-class strength (those based on only a few poorly sampled age groups) may need to be replaced by independent estimates, based for example on survey indices, or by a recent average value.

Table 2. Analysis by SRMCM2 of North Sea cod: total international catch numbers: 1976–1985. Delta year effect = 0.0; CV = 0.205; Constant = 10.22.

	Estimated effects		
Level	Year	Age	Year class
1	0.00	0.00	0.00
2	0.38	1.25	1.91
3	0.41	−0.02	1.01
4	0.05	−1.32	0.33
5	0.13	−2.19	−0.50
6	0.03	−3.10	−0.04
7	0.00	−3.89	0.24
8		−4.85	0.58
9		−5.43	−0.69
10		−6.56	−0.41
11		−7.84	−0.62
12		−8.03	−0.21
13		−9.53	−1.07
14			0.25
15			−0.14
16			−0.01
17			0.62
18			−0.43
19			0.80

Mean year-class effect = 0.09

Assumptions for catch forecast

Year effects		
Year	8	9
Effect	0.00	0.00
Year-class effects		
Year class	20	21
Effect	0.10	0.10

	Fitted catch numbers							Forecast	
Age	1	2	3	4	5	6	7	8	9
1	9 358.5	51 393.9	36 010.4	28 709.8	57 957.5	18 422.8	61 191.0	30 291.6	30 291.6
2	77 809.3	47 762.7	184 675.1	88 009.4	108 512.9	182 462.3	62 435.2	213 561.7	105 721.5
3	14 512.4	31 949.0	13 808.5	36 314.5	26 763.7	27 486.3	49 752.2	17 532.0	59 970.4
4	4 857.6	5 755.0	8 921.2	2 621.9	10 665.9	6 547.3	7 238.3	13 493.6	4 754.5
5	1 541.6	2 982.9	2 488.1	2 623.1	1 192.0	4 040.3	2 669.6	3 039.5	5 666.7
6	2 196.2	907.8	1 236.8	701.3	1 144.0	432.6	1 580.1	1 075.0	1 224.1
7	711.1	1 464.1	425.7	394.3	345.8	470.0	191.0	720.0	489.7
8	204.2	396.3	574.5	113.0	162.3	118.4	173.5	72.2	274.1
9	71.6	167.0	227.9	224.6	68.1	81.5	63.9	96.7	40.0
10	53.1	33.4	55.1	51.0	78.2	19.2	24.9	20.0	30.7
11	28.7	20.9	8.8	9.9	14.6	18.8	4.4	6.2	4.8
12	59.3	34.9	17.6	4.7	8.7	10.6	14.9	3.5	4.9
13	1.0	18.8	7.3	1.9	0.4	1.0	1.5	2.6	0.0

When fishing mortality is high, or recruitment is highly variable, this will be a crucial part of the calculation, as it is for any catch forecast (see, for example, Shepherd (1991)). If an average is used, the most recent values should of course be excluded, because they are based on a few points and have high standard errors. The values of f_y required must as usual be based on *assumptions* about the level of fishing mortality. These again are relative values, but since such assumed values are usually based on recent values increased or decreased appropriately, this usually causes no difficulty.

The forecasts obtained are of catch-at-age, as for a standard age-based catch forecast, and may be converted to catch weights by forming sums of products with appropriate weight-at-age estimates. It is interesting to note that the forecast is obtained without VPA-like retrospective estimation of fishing mortality or absolute population size, and is based entirely on fitted (and therefore somewhat averaged) parameters and not at all on raw data, and should thus be relatively insensitive to sampling errors in recent catch data.

6. Example: Applications to real data

An example of the application of the model to a survey data set was given by Shepherd and Nicholson (1986), and real-life applications to both survey and commercial data have been given by Cook (1988) and Borges (1990). Cook's application to a very heterogeneous data set with missing data is particularly interesting.

An application to a standard total international commercial data set is given here. The data (for North Sea cod from 1976 to 1985, taken from Anon. (1986)) are given in Table 1, and the results, assuming no trend of year effect in Table 2, and assuming an increase of 0.1 per year (referred to as "delta year effect") in the logarithmic year effect in Table 3. The results for the year, year class, and age effects for both assumptions are given in Figures 1 to 3: these make it clear that the only difference between the two interpretations is an

Table 3. Analysis by SRMCM2 of North Sea cod: total international catch numbers: 1976–1985. Delta year effect = 0.1; CV = 0.205; Constant = 11.42.

	Estimated effects		
Level	Year	Age	Year class
1	0.00	0.00	0.00
2	0.48	1.15	1.81
3	0.61	−0.22	0.81
4	0.35	−1.62	0.03
5	0.53	−2.59	−0.90
6	0.53	−3.60	−0.54
7	0.60	−4.49	−0.36
8		−5.55	−0.12
9		−6.23	−1.49
10		−7.46	−1.31
11		−8.84	−1.62
12		−9.13	−1.31
13		−10.73	−2.27
14			−1.05
15			−1.54
16			−1.51
17			−0.98
18			−2.13
19			−1.00

Mean year-class effect = −0.81

Assumptions for catch forecast

Year effects		
Year	8	9
Effect	0.60	0.60
Year-class effects		
Year-class	20	21
Effect	−0.80	−0.80

	Fitted catch numbers							Forecast	
Age	1	2	3	4	5	6	7	8	9
1	9 358.5	51 393.9	36 010.5	28 709.8	57 957.6	18 422.8	61 191.0	74 506.7	74 506.7
2	77 809.3	47 762.8	184 675.1	88 009.4	108 512.9	182 462.5	62 435.2	193 238.6	235 288.6
3	14 512.4	31 949.0	13 808.5	36 314.5	26 763.7	27 486.3	49 752.2	15 863.5	49 099.5
4	4 857.6	5 755.0	8 921.2	2 621.9	10 665.9	6 547.3	7 238.3	12 209.4	3 892.5
5	1 541.6	2 982.9	2 488.1	2 623.1	1 192.0	4 040.3	2 669.6	2 750.2	4 639.4
6	2 196.2	907.8	1 236.8	701.3	1 144.0	432.6	1 580.1	972.6	1 002.0
7	711.1	1 464.1	425.7	394.3	345.8	470.0	191.0	651.4	400.7
8	204.2	396.3	574.5	113.0	162.3	118.4	173.5	65.3	224.2
9	71.6	167.0	227.9	224.6	68.1	81.5	63.9	87.4	32.6
10	53.1	33.4	55.1	51.0	78.2	19.2	24.9	18.0	24.9
11	28.7	20.9	8.8	9.9	14.6	18.8	4.4	5.5	3.8
12	59.3	34.9	17.6	4.7	8.7	10.6	14.9	3.1	3.9
13	1.0	18.8	7.3	1.9	0.4	1.0	1.5	2.2	−0.2

increasing additive offset as the ages, years, and year classes progress: the residual CV is identical at 0.205. The relative overall fishing mortality and year-class strength, as subsequently estimated by VPA (Anon., 1988), are also plotted in Figures 1 and 2 for comparison. It is clear that the interpretation is very close to that obtained by more conventional methods. The year effects are small (relative to the residual error of 0.205): the short-term fluctuations are closely paralleled after year 2 (especially for delta year effect (DYE) = 0.1), but the overall trend is more closely matched for DYE = 0.0. The recruitment pattern is very closely reproduced, with the overall trend again more closely matching that for DYE = 0.0. The corrected catch curves (Fig. 3) are remarkable for their smoothness and systematic pattern, which are in no way enforced by the method. The slope of the right-hand limb of the curve for DYE = 0.0 corresponds to a total mortality of 1.0 over the age range 2 to 11, with a suggestion of a higher value of about 1.3 for ages 2 and 3. This is close to the conventional interpretation (Anon., 1988).

The data have in fact been fitted for seven years only (1976 to 1982), in order that the forecast catch-at-age may be compared with the out-turn for subsequent years. The fitted and forecast values of catch-at-age are given in Tables 2 and 3, for a simple forecast assuming that the year effect remains at its most recent level (i.e. a *status quo* assumption) and that year-class strength is equal to the mean (including the most recent ones even though these have high standard errors).

It is easily seen that the fitted values are, as expected, *identical* for the two assumptions. The forecasts differ, however, particularly for the younger age groups. The results are plotted in Figure 4 for 1983, together with the observed catch-at-age for that year (Anon., 1988). For ages greater than 3 the agreement is excellent, and extends through to age 7 and beyond, as may be seen in the logarithmic plot (Fig. 4a). For catch forecast purposes, the absolute numbers as displayed in Figure 4(b) are more significant, however. The agreement for ages 3 to 5 is excellent. That for age 2 is less satisfactory, because the estimates of year-class strength are determined solely by one data point (for the youngest age group). The estimates for the newly recruiting age group 1 are poor, since the estimates are based on averages

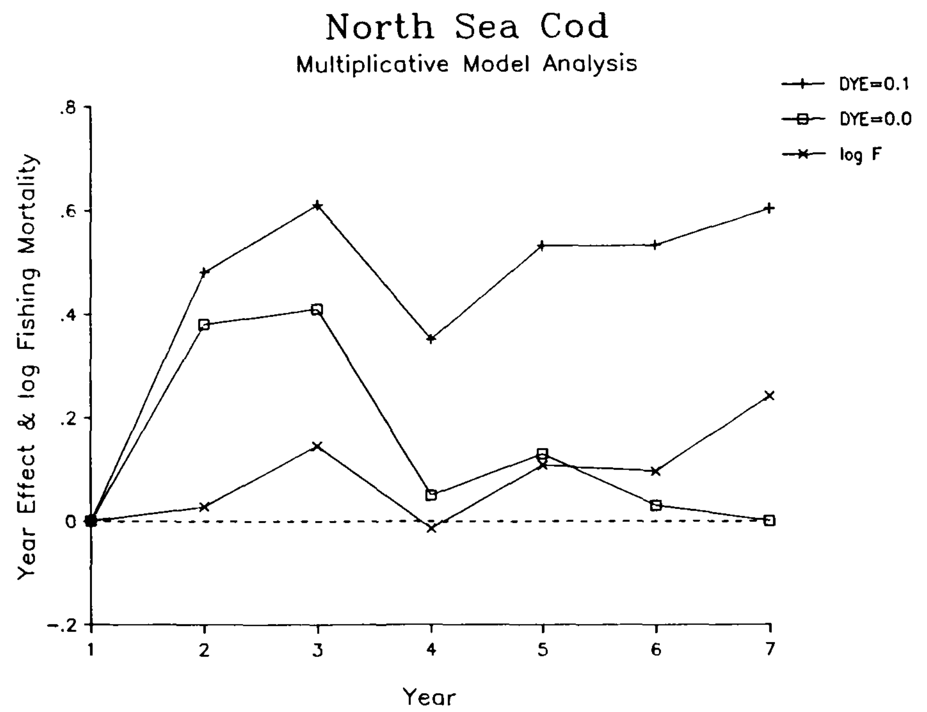

Figure 1. Estimated year effects for two attempts at fitting total international catch data for North Sea cod, with the logarithm of mean standardized fishing mortality from VPA, for delta year effect (DYE) equal to zero and 0.1 per year.

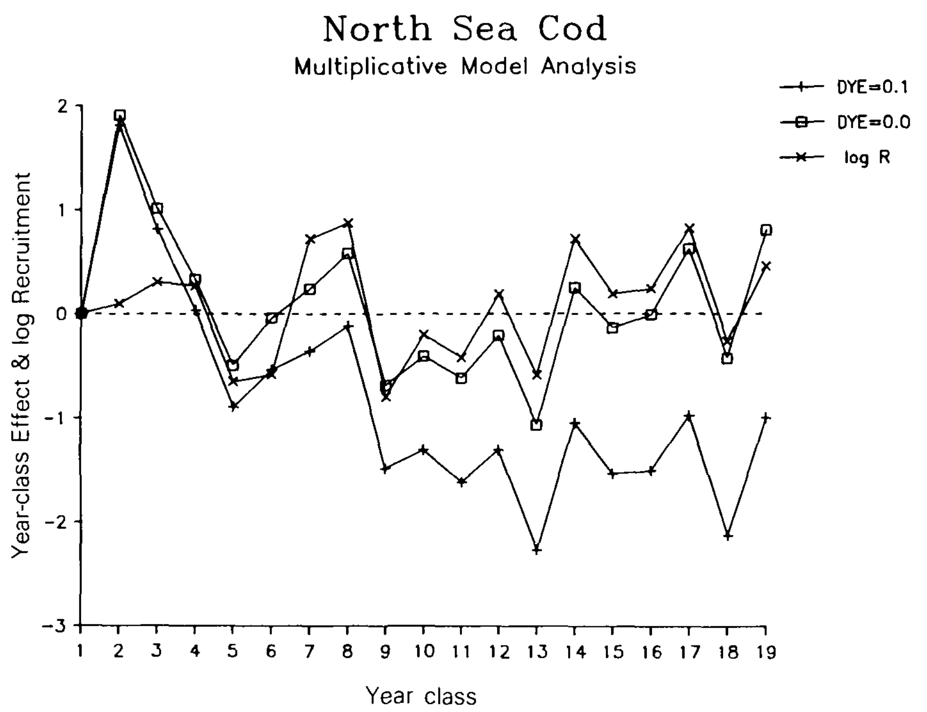

Figure 2. Estimated year-class effects and logarithm of standardized recruitment for the same analyses as for Figure 1.

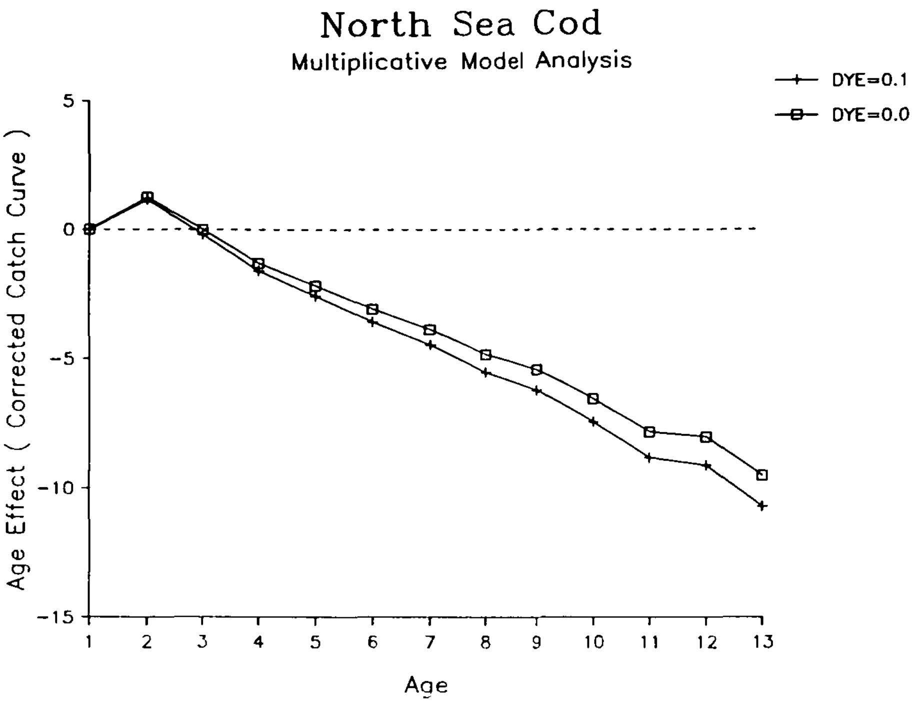

Figure 3. Estimated age effects for the same analyses as for Figure 1.

only. Clearly, it would in practice be necessary to estimate the size of recruiting year classes from a careful analysis of pre-recruit data, as usual.

7. Discussion

The distinctive features of the method are that (unlike VPA) it allows for the possibilities of sampling error in catch-at-age data, and may be applied to data for individual fleets as well as to total international data. It also permits estimation of (relative) year-class strength for any cohort represented in the data, and the steady-state age composition (i.e. an average catch curve corrected for the effects of varying recruitment and fishing mortality). This may be used to estimate total mortalities directly, and is the necessary starting point for the method of Jones (1961), which permits the usual long-term yield and biomass calculations to be carried out very simply. In addition, however, since year-class effect is also estimated, the parameters may be used (with an appropriate assumption about future year-class strengths and fishing mortality – e.g. that any trends detected will continue) to construct short-term catch forecasts as well. In this context the method is essentially a generalization of the leapfrog and ANOVA methods of Pope (1983), but uses the data more efficiently and should be more stable, since recent catch data are not taken to be error-free (except for the last year class represented).

It is interesting to note that the method has only one degree of indeterminacy, compared with separable VPA, which has two. This is because the multiplicative model only estimates its factors relatively, whilst separable VPA uses the essential additive feature of VPA (that a sum of catches is a dimensional estimate of population) to arrive at absolute estimates of population and fishing mortality. If a pseudocohort analysis of the age effect is made, to obtain absolute estimates of fishing mortality, then the extra level of indeterminacy re-appears in the form of the terminal F for the pseudocohort.

The use of a multiplicative model may be regarded as intermediate between a simple catch-curve analysis and a VPA. The former assumes that year-class strength, fishing mortality, and exploitation pattern are all constant (and therefore well estimated by their averages). Such an analysis is therefore very parsimonious – perhaps dangerously so, particularly in respect of varying year-class strength. This particular assumption is relaxed by the multiplicative model, and the first-order effects of varying partial fishing mortality are also allowed for by the year effect (although the second-order effect due to varying total F expressed by the year/age interaction is still ignored). In a full VPA, all these parameters (year-class strength, exploitation pattern, fishing mortality) are allowed to vary arbitrarily, at the expense of severe indeterminacy, associated with no statistical degrees of freedom.

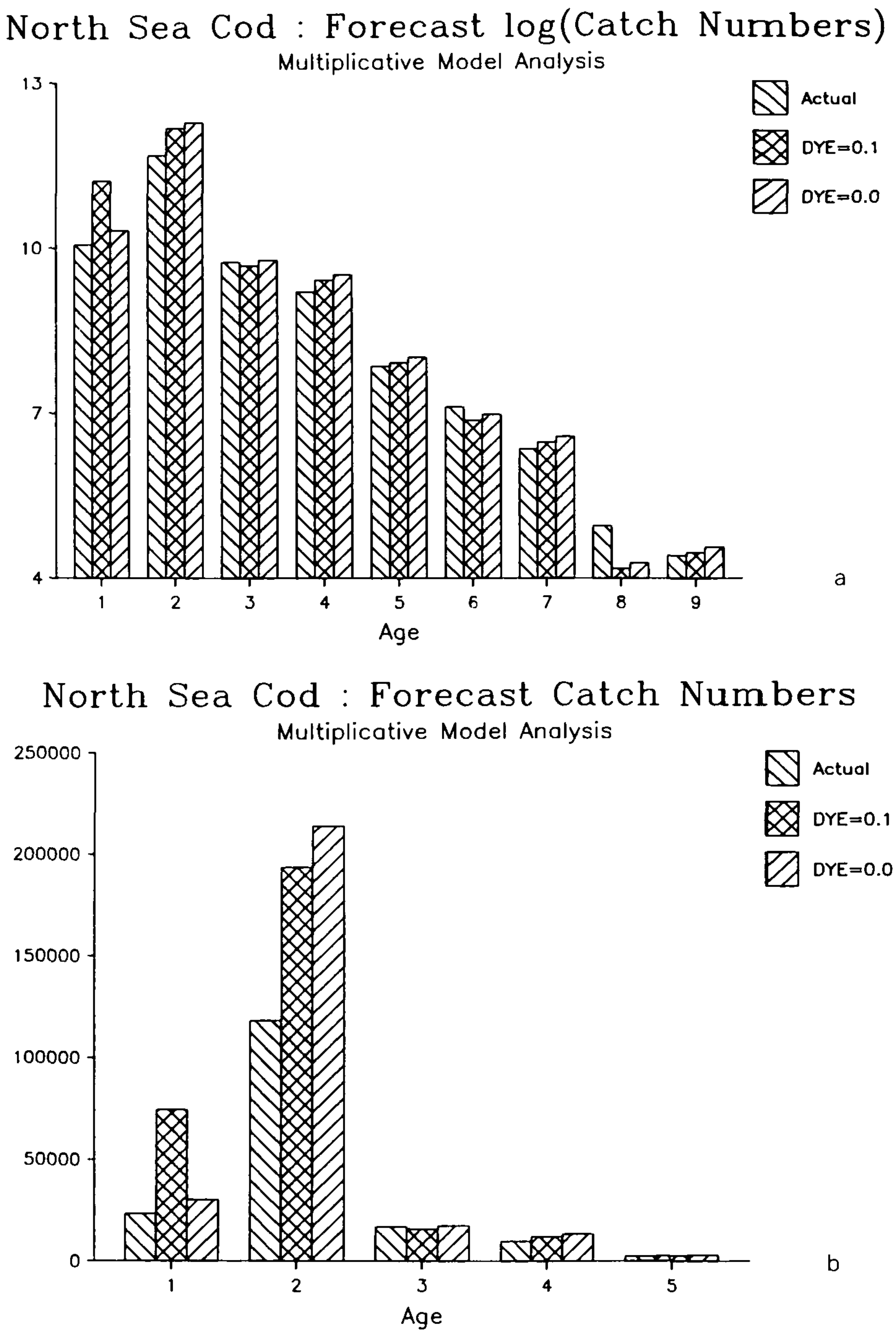

Figure 4. (a) Forecast of catch numbers for North Sea cod using the multiplicative model, compared with the actual catches, on a logarithmic scale. (b) As (a), but on a linear scale, to emphasize the age groups of greatest practical importance.

There is therefore a clear trade-off between the generality and plausibility of the model, and the indeterminacy and lack of stability of the estimates. The multiplicative model, like separable VPA to which it is closely related, is an attempt to find a middle way, with a reasonably general (and plausible) model, and only modest indeterminacy. The choice between the multiplicative model and separable VPA then depends on whether or not total catch data are available, and whether or not absolute estimates of population size and fishing mortality are required. Separable VPA can provide the latter (given total catch data), but only at the expense of one extra indeterminate parameter, and inability to use standard statistical software. The most appropriate choice depends on the circumstances.

The other main point is that any of the parsimonious models (the catch-curve corresponds to fitting the main age effect only) leave an estimate of the residual error, and some degrees of freedom, so that some estimate of the quality of the fit, and the precision of the parameters estimated, can be made. These features are absent with pure VPA, which trades them for a plausible but excessively general model.

The precision of the method used for catch forecasting has been tested by simulation studies (Sun and Shepherd, 1991) and found to be intermediate between that of traditional analytical methods and very simple ones (Shepherd, 1991), as might be expected, and it may therefore be useful when only a very short time series (say, three to five years) of data is available, so that more traditional methods cannot realistically be applied.

In addition, however, the method may be used rather generally for the analysis of incomplete data, survey data, and even weight-at-age data, in a rather straightforward and informative way, and may therefore also be applicable to more diverse data sets.

Acknowledgements

The authors would like to thank the editor and an unnamed referee for their constructive criticism of earlier drafts of this paper.

References

Anon. 1986. Report of the North Sea Roundfish Working Group. ICES CM 1986/Assess: 16.

Anon. 1988. Report of the North Sea Roundfish Working Group. ICES CM 1988/Assess: 21.

Baker, R. J., and Nelder, J. A. 1978. The GLIM system, Release 3, Generalized Linear Interactive Modelling Manual. Numerical Algorithms Group, Oxford.

Borges, M. de F. 1990. Multiplicative catch-at-age analysis of scad (*Trachurus trachurus*) from western Iberian waters. Fisheries Res.

Cook, R. M. 1988. Sections 16.5 and 16.6, pp. 23–24 in Report of the North Sea Roundfish Working Group. ICES CM 1988/Assess: 21.

Deriso, R. B., Quinn, T. J. (II), and Neal, P. R. 1985. Catch-at-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci., 42: 815–824.

Fournier, D., and Archibald, C. P. 1982. A general theory for analysing catch-at-age data. Can. J. Fish. Aquat. Sci., 39: 1195–1207.

Gilchrist, W. 1984. Statistical modelling. J. Wiley, Chichester/New York. 339 pp.

Gulland, J. A. 1983. Fish stock assessment: a manual of basic methods. Wiley-Interscience, Chichester/New York. 223 pp.

Jones, R. 1961. The assessment of long-term effects of changes in gear selectivity and fishing effort. Marine Research, No. 2 (DAFS, Edinburgh). 19 pp.

Pope, J. G. 1979. Population dynamics and management: current status and future trends. Investigacion Pesquera, 43: 199–221.

Pope, J. G. 1983. Analogies to the *status quo* TACs: their nature and variance. Can. Spec. Publ. Fish. Aquat. Sci., 66: 99–113.

Pope, J. G., and Shepherd, J. G. 1982. A simple method for the consistent interpretation of catch-at-age data. J. Cons. int. Explor. Mer, 40: 176–184.

Pope, J. G., and Stokes, T. K. 1989. The use of multiplicative models for separable VPA, integrated analysis and the general VPA tuning problem. Am. Fish. Soc. Symp., 6: 92–101.

Shepherd, J. G. 1990. Simple methods for short-term forecasting of catch and biomass. J. Cons. int. Explor. Mer (submitted).

Shepherd, J. G., and Nicholson, M. D. 1986. Use and abuse of multiplicative models in the analysis of fish catch-at-age data. The Statistician, 35: 221–228.

Sun, M., and Shepherd, J. G. 1991. Simulation testing of the performance methods for fish stock assessment and short-term catch forecasts. To appear.