Northwest Atlantic Fisheries Organization

Serial No. N1991 NAFO SCR Doc. 91/99

SCIENTIFIC COUNCIL MEETING - SEPTEMBER 1991

Modelling the Effect of Environment on Growth of Cod: Fitting to Growth Increment Data *Versus* Fitting to Size-at-age Data

by

Russell B. Millar

Science Branch, Dept. of Fisheries and Oceans, P. O. Box 5667
St. John's, Newfoundland, Canada A1C 5X1

COD SYMPOSIUM, SEPTEMBER 1991

Abstract

If size-at-age data are collected from annual surveys over a number of consecutive years then variability in growth can be investigated for association with environmental influences by formulating annual growth increments as a function of the environmental conditions prevailing at the time. Because growth increments are not observed directly, but are calculated as the difference of size-at-age measurements, successive growth increments are statistically correlated. This has resulted in a division in the method of analysis of such data because some studies ignore the correlation while others accomodate it. Here the performance of three different methods of analysis are compared.

Introduction

Recent studies (e.g. Akenhead et al. 1982; Nakken and Raknes 1987; Millar and Myers 1990; P-Gandaras and Zamarro 1990; Anon 1991) have been directed at relating environmental variables (e.g., bottom temperature; population biomass; prey abundance) to cod growth using data from annual surveys. The objective of these studies is to investigate the influence of environmental conditions on the growth increments observed between successive surveys.

Due to the availability of length-at-age data these studies used length as the measure of fish size and the same will be done here since the arguments do not depend on how size is quantified. In the analysis of these data a division of methodology has occurred due to the fact that it is length-at-age data that is observed, but incremental growth that is directly modelled as environmentally dependent. Akenhead et al. (1982) and Anon (1991) difference the length-at-age data to obtain growth increment data, to which an environmentally dependent growth increment model is fitted. Millar and Myers (1990) and Anon (1991) sum the (environmentally dependent) growth increments from these models to produce an environmentally dependent length-at-age model and fit this to the observed length-at-age data. These two approaches are presented below as methods 1 and 3 respectively.

Method 1 below uses ordinary least squares to fit the expected (environmentally dependent) growth increments to the growth increment data. It is shown that the growth increment data are statistically correlated and method 2 (generalized least squares) is presented as a method that incorporates this correlation into the fitting procedure. Method 3 fits expected length-at-age, obtained from summing the expected growth increments, to the observed length-at-age data using ordinary least squares.

This study examines the rationale behind these three different methods of fitting the same growth model. Simulation is used to observe the behaviour of the methods and to compare their relative performance. In the presentation below t is used interchangeably as an index for time or age. For rigour, one should interpret t as an age index for each yearclass, and so the "time interval" $(t, t+1)$ is in fact the year in which the yearclass aged from t to $t+1$.

Methods

Method 1: *Fitting to growth increment data using ordinary least squares*

The expected growth increment over time interval $(t, t+1)$ will be denoted by D_t and the environmental variables suspected of influencing D_t will be denoted by x_t, possibly a vector. The expected growth increment is modelled as

$$D_t = f(x_t; \beta) \qquad (1)$$

where f is some pre-specified function and β is a vector of parameters to be estimated. For example, if x_t is the average temperature anomaly over the time period $(t, t+1)$ then one might wish to begin the analysis with a least squares fit of the simple linear model

$$D_t = \beta_0 + x_t\beta_1$$

to the observed growth increments. Inspection of the residuals from this model will indicate whether other terms in the model are required, such as a quadratic effect of temperature.

Equation (1) assumes that growth does not depend on current size or age. There is some evidence to suggest that this may be a reasonable assumption for cod over a restricted range of ages. For example, the Northeast Arctic cod data of Nakken and Raknes (1987) does not show any apparent size effects on growth increments (they present data for ages 1-7). P-Gandaras and Zamarro (1990) observe linear growth in Flemish Cap cod, up to age 11 in some yearclasses. Moreover, if a size effect is suspected then age can be used as a proxy for size (Anon 1991) and its statistical significance tested by including it as a factor in the linear model.

The growth increment data for each yearclass are calculated as $d_t = l_{t+1} - l_t$ where l_t is the observed length at age t. Note that successive growth increments $d_t = l_{t+1} - l_t$ and $d_{t+1} = l_{t+2} - l_{t+1}$ are not statistically independent because they both depend on l_{t+1}. Therefore the use of (ordinary) least squares is not strictly appropriate. The appropriate recourse for dealing with correlated observations in a least squares model is to use generalized least squares (Seber 1977), as described below.

Method 2: *Fitting to growth increment data using generalized least squares*

In generalized least squares a linear transformation is applied to the data to eliminate the correlations, as outlined below. For full details the reader is referred to Seber (1977) and Seber and Wild (1989).

If M is the age of the oldest fish observed and the youngest age observed is 1 then, in matrix notation,

$$d = Al$$

where $d^T = (d_1, d_2, ..., d_{M-1})$, $l^T = (l_1, l_2, ..., l_M)$ and A is the $M-1 \times M$ matrix given by

$$A = \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ . & . & . & \ldots & . & . \\ . & . & . & \ldots & . & . \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix}.$$

Since the length-at-age observations are statistically independent the covariance matrix $\sigma^2\Sigma$ of l is a diagonal matrix and the covariance matrix of d is $\Psi = \sigma^2 A\Sigma A^T$ (Seber 1977). Since Ψ is a positive definite symmetric matrix it can be represented by its Cholesky decomposition (Seber and Wild 1989, p. 680), that is, Ψ can be written in the form KK^T where K is a non-singular lower triangular matrix. Then the linearly transformed vector $K^{-1}d$ has covariance matrix $\sigma^2 K^{-1}\Psi(K^T)^{-1} = \sigma^2 I_{M-1}$, where I_{M-1} is the $M-1 \times M-1$ identity matrix. That is, β is estimated by using ordinary least squares to fit the model

$$K^{-1}d = K^{-1}f(x;\beta) + e$$

where $f(x;\beta) \equiv (f(x_1;\beta), f(x_2;\beta), ..., f(x_{M-1};\beta)) = (E[d_1], E[d_2], ..., E[d_{M-1}])$ and $e \equiv (e_1, e_2, ..., e_{M-1})$ are independent and identically distributed as Normal$(0, \sigma^2)$ random variables.

Method 3: *Fitting to length-at-age data using ordinary least squares*

L_t will denote expected length at age t. If expected length at age 1, L_1, is modelled as a function of environmental conditions through to the end of the first year (denoted by x_0) then the expected length at age t is

$$L_t = L_1 + \sum_{i=1}^{t-1} D_i \quad .$$

That is, expected length at age t is simply given by summing expected length at age 1 with the expected growth increments up to time t. Whatever model of D_t is used by methods 1 or 2 above can be used here.

Using this approach D_t can be formulated as a function of current size. For example, Millar and Myers (1990) used the discretized version of the von Bertalanffy growth curve,

$$D_t = L_{t+1} - L_t = (L_{\max} - L_t)(1 - e^{-k}) \tag{2}$$

and modelled $L_{\max}$ or k to be functions of bottom temperature and cod biomass. That is, $k \equiv k(x_t;\beta)$ or $L_{\max} \equiv L_{\max}(x_t;\beta)$ where $x_t = (\text{temp}_t, \text{biom}_t)$ is the annual temperature anomaly and cod biomass in the time interval $(t, t+1)$. The values L_t can be fitted to the observed length-at-age data l_t by ordinary least squares. Since the observed length-at-age data are independent this is a statistically correct procedure.

This approach of modelling environmentally dependent growth increments by fitting to length-at-age data has been used by Cloern and Nichols (1978) and Campana and Hurley (1989). Cloern and Nichols (1978) modified the continuous form of the von Bertalanffy growth equation to account for seasonal variation in the growth of a bivalve mollusc and flathead sole. Campana and Hurley (1989) applied modified exponential and logistic growth models to growth of cod and haddock larvae.

Numerical study

One criticism of fitting growth curves (method 3) is that in practice the residuals from each yearclass may display autocorrelation. For example, if an environmental measurement x_t is spurious at time t then it will affect the entire growth curve of that yearclass. Autocorrelated residuals could also arise from model misspecification, such as the existence of highly influential environmental variables that are not included in the model. Method 2 is susceptible to the same phenomenon because of the linear transformation applied to the vector of expected growth increments - the entire transformation is affected if any element of the vector is spurious. Thus, although methods 2 and 3 are statistically correct, it is worth exploring how they compare to method 1 in detecting and quantifying the effect of environmental variables on growth.

Simulations were conducted to compare the performance of method 1, growth increment ordinary least squares fit; method 2, growth increment generalized least squares fit; and method 3, growth curve ordinary least squares fit. The scope of the simulation study was chosen to reflect the type of data summarized by Nakken and Raknes (1987). For seven yearclasses of Northeast Arctic cod they presented a summary of length-at-age data for ages 1-7 and bottom temperature data considered to be representative of that experienced by the cod.

To best emulate the difficulties arising in the application of these environmentally modified fits to cod growth, several sources of variability and error were introduced. For simplicity, growth increments were modelled as a function of only one environmental variable, temperature anomaly. For each of seven yearclasses the simulated length-at-age data were generated in the following way:

1: The temperature anomolies temp_t in the time intervals $(t, t+1), t = 0, 1, \ldots, 6$ are distributed as independent standard normal random variables.

2: The expected length increment (cm) over time interval $(t, t+1)$ is given by

$$D_t = L_{t+1} - L_t = 10 + \beta \text{temp}_t \quad t = 1, 2, \ldots, 6 \tag{3}$$

and the expected length at age 1 by

$$L_1 = 15 + \beta \text{temp}_0 \quad . \tag{4}$$

Note that in the event of no temperature variation the expected lengths of fish from ages 1 to 7 are 15, 25, 35, 45, 55, 65 and 75 cm respectively.

3: The observed length at age t is

$$l_t = L_t + \epsilon_t \quad t = 1, ..., 7$$

where each ϵ_t is an independent Normal$(0, \sigma_L^2)$ random variable.

4: The observed temperature anomaly over time $(t, t+1)$ is

$$x_t = \text{temp}_t + \varepsilon_t \quad t = 0, ..., 6 \tag{5}$$

where each ε_t is an independent Normal$(0, \sigma_T^2)$ random variable.

Simulations were performed for all combinations of β = 0.0, 0.5, 1.0 and 2.0; σ_L^2=0.25 and 1.0; and σ_T^2= 0.0, 0.25 and 1.0. Note that $\sigma_T^2 = 0$ corresponds to no measurement error on the explanatory variable, while $\sigma_T^2 = 1$ puts as much measurement error on temperature anomaly as there is actual true variability (since temperature anomaly is distributed as a standard normal) and so will result in some "spurious" x_t. This large amount of measurement error can be considered as incorporating the fact that the annual temperature anomaly experienced by the fish is estimated using some convenient temperature index, often calculated from temperature data collected at a fixed location. The underlying assumption is that the temperature anomaly so measured is positively correlated with the temperature anomaly experienced by the fish.

Results

The simulation results are given in Tables 1-3. When there is no measurement error on temperature anomaly ($\sigma_T^2 = 0$) the root mean squared error (rmse) of the estimators does not depend on the value of β (Tables 1 and 2). This is expected, since in this case all three $\hat{\beta}$ are unbiased and their variance depends only on the design matrix and the magnitude of the error in observing the length-at-age data. The growth curve estimator $\hat{\beta}_3$ has lower rmse than the growth increment generalized least squares estimator $\hat{\beta}_2$. This is because $\hat{\beta}_3$ takes advantage of the information provided by L_1, which $\hat{\beta}_2$ does not use. The effect of ignoring autocorrelation results in $\hat{\beta}_1$ having the highest rmse of the three estimators.

With a moderate amount of measurement error on temperature ($\sigma_T^2 = 0.25$) the performance of the estimators becomes dependent on β. For β equal to 0 or 0.5, $\hat{\beta}_3$ continues to have smaller rmse than $\hat{\beta}_2$, which in turn has smaller rmse than $\hat{\beta}_1$. For β equal to 1.0 there is little difference in performance but for $\beta = 2.0$ the ranking of the estimators reverses, with $\hat{\beta}_1$ having smallest rmse and $\hat{\beta}_3$ the largest rmse in both Tables 1 and 2.

With a larger amount of measurement error ($\sigma_T^2 = 1.0$) there is little difference in relative performance at any of the non-zero values of β. For $\beta = 2.0$ estimator $\hat{\beta}_1$ has the smallest variance, but the bias of all three estimators is so large as to make the difference negligible in terms of rmse.

It is interesting to observe that in all cases the three estimators displayed the same degree of bias. Moreover, the degree of bias appears to be determined by the signal to noise ratio in the temperature data. The true temperature anomaly is distributed Normal(0,1) and so when its measurement error is distributed as Normal(0,0.25) the total variability in the temperature observations is distributed as Normal(0,1.25). The proportion of total observed variability explained by true temperature anomaly is thus $1.0/1.25 = 0.8$. The observed mean of the estimators is this same proportion of β (Tables 1 and 2). The same holds true when measurement error on temperature anomaly is Normal(0,1) - the total variability is Normal(0,2) and the ratio is 0.5.

To study the ability of the three estimators to detect an environmental influence on growth the null hypothesis $\beta = 0$ was tested at each simulation (Table 3) at the 5% level of significance. Under the null hypothesis the observed significance level of all three estimators was not significantly different from 0.05 for all combinations of σ_L^2 and σ_T^2. For $\beta = 0.5$ method 1 had great difficulty in detecting the effect of temperature anomaly when σ_L^2 was 1.0. For example, it reported a significant effect in only 34% of the simulations when σ_L^2 and σ_T^2 were both equal to 1.0. In contrast, methods 2 and 3 detected the effect in 62% and 85% of the simulations respectively. For stronger effects, $\beta = 1$ or 2, there was not much difference in power of the tests.

Discussion

Three different ways of fitting the incremental growth models have been demonstrated. Any model used by method 1 can be used by methods 2 and 3, but the converse is not true because only method 3 can include current size as a variable affecting expected growth increment.

Fitting growth curves (method 3) requires determination of L_1, the expected length at age 1. Due to variation in yearclass timing and prevailing conditions during the pelagic stage it may be desirable to model L_1 (equation 4) differently from D_t (equation 3) in practice. If reasonable explanatory variables for L_1 can not be found then one is restricted to using a growth increment fit - methods 1 or 2.

The problem of correlated growth increment data is avoided by methods 2 and 3, but in so doing they incur the feature that the fit to an entire yearclass can be affected by a single spurious covariate data point. However, the simulations showed that this does not greatly affect their ability to estimate and detect the significant environmental effects and that model 1 is as much affected by these spurious data. When the model was well specified (small σ_T^2), methods 2 and 3 generally did considerably better than method 1, and when σ_T^2 was large there was little difference between the three methods.

On the practical side, the three methods were easily implemented in the above simulations because they were all fitted by linear least squares. Method 3 becomes computationally complex when nonlinear growth curves are fitted because finding the least squares solution then typically requires an iterative procedure.

When the survey data include older fish the growth increment fits may not be able to take advantage of all the data because as a yearclass ages and becomes scarcer the length-at-age data will become patchy and will include gaps where no fish of that yearclass were observed. This is frequently the case in the Atlantic Canada cod data used by Millar and Myers (1990). Without successive length-at-age measurements annual growth increments can not be calculated.

Fitting of growth curves (method 3) enables some interesting mechanistic questions to be addressed. For example, - over a prolonged period of unfavourable environmental conditions will maximum length be attained at the same age (implying a smaller maximum length), or will fish eventually reach the same maximum length but at an older age? In the context of the von Bertalanffy growth curve the above question is asking whether it is $L_{\max}$ or k (equation 2) that is a function of the environment. Another question: "is loss of growth in an unfavourable year eventually recovered (when environmental conditions change) or permanently lost?", is also addressed by Millar and Myers (1990).

To conclude, since the simulations can not hope to cover the range of all possible parameter values or ways in which the model can be mis-specified or corrupted it may be unwise to universally recommend one of the three methods over the others. However, method 3 has the advantages that it enables all of the length-at-age data to be utilized and provides a valuable management tool - an environmentally sensitive growth curve.

References

Akenhead, S. A., J. Carscadden, H. Lear, G. R. Lilly, and R.Wells. 1982. Cod-capelin interactions off northeast Newfoundland and Labrador. *In* M. C. Mercer (ed)., Multispecies approaches to fisheries management advice. Can. Spec. Publ. Fish. Aquat. Sci. 59: 141-148.

Anon. 1991. Report of the multispecies assessment working group. ICES C. M. 1991/Assess: 7.

Campana, S. E., and P. C. F. Hurley. 1989. An age- and temperature-mediated growth model for cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*) larvae in the Gulf of Maine. Can. J. Fish. Aquat. Sci. 46: 603-613.

Cloern, J. E., and F. H. Nichols. 1978. A von Bertalanffy growth model with a seasonally varying coefficient. J. Fish. Res. Board Can. 35: 1479-1482.

P-Gandaras, G., and J. Zamarro. 1990. Changes in the cohort growth rate of Flemish Cap cod. NAFO, SCR Doc. 90/40.

Millar, R. B., and R. A. Myers. 1990. Modelling environmentally induced change in growth for Atlantic Canada cod stocks. ICES C. M. 1990/G:24.

Nakken, O. and A. Raknes. 1987. The distribution and growth of Northeast Arctic cod in relation to bottom temperatures in the Barents Sea, 1978-1984. Fish. Res. 5: 243-252.

Seber, G. A. F. 1977. Linear regression analysis. Wiley, New York.

Seber, G. A. F., and C. J. Wild. 1989. Nonlinear regression. Wiley, New York.

Table 1. Mean, standard error and root mean squared error of the ordinary least squares fit to increments estimator ($\hat{\beta}_1$), generalized least squares fit to increments estimator ($\hat{\beta}_2$) and the ordinary least squares fit to length-at-age data estimator ($\hat{\beta}_3$). The results are calculated from 500 simulations. Random error on observed length has variance $\sigma_L^2 = 0.25$.

	σ_L^2=0.25								
	σ_T^2=0			σ_T^2=0.25			σ_T^2=1.0		
	mean	sd	rmse	mean	sd	rmse	mean	sd	rmse
$\beta = 0.0$									
$\hat{\beta}_1$	-0.01	0.12	0.12	0.00	0.10	0.10	0.00	0.08	0.08
$\hat{\beta}_2$	-0.01	0.08	0.08	0.00	0.07	0.07	0.00	0.05	0.05
$\hat{\beta}_3$	0.00	0.04	0.04	0.00	0.04	0.04	0.00	0.03	0.03
$\beta = 0.5$									
$\hat{\beta}_1$	0.49	0.11	0.11	0.40	0.11	0.15	0.26	0.09	0.26
$\hat{\beta}_2$	0.50	0.07	0.07	0.40	0.09	0.13	0.26	0.09	0.25
$\hat{\beta}_3$	0.50	0.04	0.04	0.40	0.08	0.12	0.26	0.09	0.26
$\beta = 1.0$									
$\hat{\beta}_1$	1.00	0.11	0.11	0.79	0.13	0.24	0.50	0.11	0.51
$\hat{\beta}_2$	1.00	0.07	0.07	0.79	0.13	0.24	0.50	0.14	0.52
$\hat{\beta}_3$	1.00	0.04	0.04	0.79	0.16	0.27	0.50	0.17	0.53
$\beta = 2.0$									
$\hat{\beta}_1$	2.01	0.11	0.11	1.60	0.16	0.43	1.00	0.17	1.02
$\hat{\beta}_2$	2.00	0.07	0.08	1.60	0.22	0.45	0.99	0.27	1.04
$\hat{\beta}_3$	2.00	0.04	0.04	1.61	0.28	0.48	0.99	0.35	1.07

Table 2. Same as Table 1, but with $\sigma_L^2 = 1.0$.

	σ_L^2=1.0								
	σ_T^2=0			σ_T^2=0.25			σ_T^2=1.0		
	mean	sd	rmse	mean	sd	rmse	mean	sd	rmse
$\beta = 0.0$									
$\hat{\beta}_1$	0.02	0.22	0.22	-0.01	0.20	0.20	0.00	0.17	0.17
$\hat{\beta}_2$	0.02	0.15	0.15	0.00	0.13	0.13	0.00	0.10	0.10
$\hat{\beta}_3$	0.01	0.08	0.08	-0.01	0.08	0.08	0.00	0.06	0.06
$\beta = 0.5$									
$\hat{\beta}_1$	0.50	0.23	0.23	0.40	0.21	0.23	0.26	0.17	0.30
$\hat{\beta}_2$	0.49	0.16	0.16	0.40	0.15	0.18	0.26	0.12	0.27
$\hat{\beta}_3$	0.49	0.09	0.09	0.40	0.10	0.14	0.25	0.11	0.27
$\beta = 1.0$									
$\hat{\beta}_1$	1.01	0.22	0.22	0.80	0.21	0.29	0.51	0.17	0.52
$\hat{\beta}_2$	1.01	0.15	0.15	0.79	0.17	0.26	0.50	0.16	0.52
$\hat{\beta}_3$	1.00	0.09	0.09	0.80	0.16	0.26	0.50	0.18	0.53
$\beta = 2.0$									
$\hat{\beta}_1$	1.99	0.22	0.22	1.60	0.23	0.46	1.01	0.23	1.02
$\hat{\beta}_2$	2.00	0.14	0.14	1.60	0.25	0.47	0.99	0.29	1.05
$\hat{\beta}_3$	2.00	0.08	0.08	1.59	0.30	0.51	1.00	0.36	1.06

Table 3. Proportion of simulations in which the null hypothesis ($\beta = 0$) was rejected at the 5% level. The results are calculated from 500 simulations.

	σ_L^2=0.25			σ_L^2=1.0		
	σ_T^2			σ_T^2		
	0.0	0.25	1.0	0.0	0.25	1.0
$\beta = 0.0$						
$\hat{\beta}_1$	0.05	0.05	0.06	0.04	0.05	0.07
$\hat{\beta}_2$	0.05	0.06	0.05	0.05	0.03	0.04
$\hat{\beta}_3$	0.05	0.05	0.06	0.04	0.05	0.04
$\beta = 0.5$						
$\hat{\beta}_1$	0.98	0.95	0.81	0.57	0.51	0.34
$\hat{\beta}_2$	1.00	0.99	0.92	0.88	0.79	0.62
$\hat{\beta}_3$	1.00	1.00	0.95	0.99	0.97	0.85
$\beta = 1.0$						
$\hat{\beta}_1$	1.00	1.00	0.98	0.98	0.94	0.81
$\hat{\beta}_2$	1.00	1.00	0.97	1.00	0.99	0.91
$\hat{\beta}_3$	1.00	1.00	0.94	1.00	1.00	0.94
$\beta = 2.0$						
$\hat{\beta}_1$	1.00	1.00	1.00	1.00	1.00	0.98
$\hat{\beta}_2$	1.00	1.00	0.99	1.00	1.00	0.98
$\hat{\beta}_3$	1.00	1.00	0.97	1.00	1.00	0.95