A linear model approach to Backcalculation of fish length

Sanford Weisberg
University of Minnesota
Technical Report No. 434

June, 1984

University of Minnesota
School of Statistics
Department of Applied Statistics
St. Paul, Minnesota 55108

A linear model approach to Backcalculation of fish length

by

Sanford Weisberg

1. Introduction

Many species of fish lay down concentric annual growth rings, or annuli, in bony structures and scales. When a fish is captured, its age can be determined by counting the number of annuli; additional information can be obtained by measuring the width of the rings, which is thought to vary with more interesting physical characteristics such as length or size.

A frequent problem in fishery management is as follows. A sample of n_f fish of several ages are captured and on each fish, total length, age, and annuli widths are measured. From these data we may wish to infer the length of the fish in previous years, called <u>backcalculation</u>. These values may be used to judge the effects of management policies instituted several years before, the effects of variation in the environment, or even to predict future fish harvests. A large literature on this problem exists; see, for example, Hile(1970), Ricker(1975), Carlander(1981, 1982), Frie(1982), Bartlett et al. (1984).

The usual method of backcalculation is as follows, with notation used only in this section. Let L_k be the length of the k-th fish and let S_k be its total width of all annuli at capture; generally, annuli are measured on the scales of the fish, and S stands for scale width. The relationship between length and scale width is assumed to be of some parametric form, with a linear regression

$$E(L_k) = \tau_0 + \tau_1 S_k \qquad (1.1)$$

the most common. For a sample of 78 Bluegills, to be discussed in detail in Section 4, Figure 1 shows the relationship between L and S, lending plausibility to (1.1). Estimates of the τ's may be obtained via least squares or some other method (Duncan, 1980); determination of τ_0 by a table lookup, specific to a species, has been advocated by Carlander (1982).

Backcalculation of fish length does not use (1.1), but rather assumes each fish to have its own slope, corresponding to the model

$$E(L_{ki}) = \tau_0 + \tau_{1k}S_{ki} \tag{1.2}$$

where L_{ki} and S_{ki} are the total length and scale width of the k-th fish at age i; of course, the L_{ki} are not observed. The use of a common intercept allows the estimation of one parameter, a slope, from one complete data point, (L_k, S_k) on each fish. The usual estimators are

$$\hat{\tau}_{1k} = (L_k - \hat{\tau}_0) / S_k \tag{1.3}$$

$\hat{\tau}_0$ estimated from between fish analysis

This procedure has served remarkably well, although its shortcomings are many. First, the method is very sensitive to the estimation of the common intercept, obtained from the questionable model (1.1). Second, (1.3) can be expected to be unstable since it uses information on only a single observation. Third, estimates of variability are not clear using this method.

The purpose of this paper is to provide an alternative method for modelling fish growth based on a single sample of fish of various ages collected in a closed ecosystem such as a lake. The approach is quite different than that used in the fishery literature, although it does rely on fairly standard linear model theory. Section 2 gives details of assumptions that are made here; most of these can be modified. Section 3 gives estimation methods and standard errors. An example is worked in Section 4 with conculsions and extensions discussed in Section 5. Technical details are presented in an appendix.

2. Notation and Assumptions

For an _a_ year old fish in the sample, _a_ measurements on the covariate (scale radius _increments_) are taken. By convention, denote the year of the sample as 1, the year before as 2, and so on. Then the values of the covariate are given by x_{akj}, where a is the age of the fish ($a = 1, \ldots, m$), k is the fish number in that age class ($k = 1, \ldots, n_a$), and j is the particular covariate measurement ($j = 1, \ldots, a$). Thus, for example, x_{342} is the scale radius increment for the fourth fish of age three last year (when the fish was two years old), and so on (see Figure 2). We collect the covariate measurements for a particular fish into the vector $\mathbf{x}_{ak}$ of length a.

Similarly, we can define a vector of fish length increments y_{akj} as the yearly increases in length of the fish. These, too, can be collected into an a-vector $\mathbf{y}_{ak}$, but we observe only the total, $z_{ak} = \mathbf{1}^T\mathbf{y}_{ak}$, the length of the fish at capture. It is also convenient to define the total for the covariate in a fish, $t_{ak} = \mathbf{1}^T\mathbf{z}_{ak}$.

Assumptions. To understand the analysis proposed here, it is helpful to make all assumptions needed explicit, and to mention their possible biological implications. Most of the assumptions can be modified, but usually at the cost of increased complexity. We first assume that the fish in the sample are independent, so growth for one fish is independent of growth of another. Within a single fish, we further assume that the yearly growth increments are independent from year to year. That is, the pair (x_{akj}, y_{akj}) is independent of the pair $(x_{akj'}, y_{akj'})$ for $j \neq j'$. This is a very strong assumption, when in fact any of negative, positive or zero correlation between successive periods may be more appropriate, depending on the population under study and the environment. The specific role of this assumption is to simplify the modelling; without it, the problem is inherently multivariate, and since the y_{akj} are unobserved, the covariances between the x's and the y's are not directly estimable. Further, one can hope that this assumption could be made more plausible by allowing transformations of the x's and the y's. For descriptive purposes, we will assume that this particular assumption of independence holds for the untransformed values.

The pair (x_{akj}, y_{akj}) are next assumed to follow a bivariate normal distribution,

$$\begin{pmatrix} x_{akj} \\ y_{akj} \end{pmatrix} \sim N\left(\begin{pmatrix} E(x_{ajk}) \\ E(y_{akj}) \end{pmatrix}, f(a,j)\begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}\right) \qquad (2.1)$$

where f(a,j) is a known positive function what can depend on $\underline{a}$ and j but not on k. Possibly, f will be decreaing in its first arguement to recognize decreasing growth with age. The assumption of proportional covariance matrix for all ($\underline{a}$, k, j) is a strong one whose plausibility may again be improved by transformation. The assumption of normality is also in some doubt because of potential catch selectivity: fish of certain sizes are more likely to be captured. This may lead to perhaps a truncated normal model, assuming that the selectivity could be successfully modelled (see Aitkin, 1981, and Schmee and Hahn, 1979, for discussion of regression problems with truncated normals). In the following, we take $f(a,j) = 1$, for all (a,j); use of other f's will require only modest modificaton of the technique, although of course f will never be known exactly in any application.

Given the normality assumption concerning the x's and the y's, and the independence assumption of the pervious paragraph, it follows that the pair (t_{ak}, z_{ak}) will also be normally distributed,

$$\begin{pmatrix} t_{ak} \\ z_{ak} \end{pmatrix} \sim N\left(\begin{pmatrix} E(\mathbf{1}^T\mathbf{x}_{ak}) \\ E(\mathbf{1}^T\mathbf{y}_{ak}) \end{pmatrix}, \begin{pmatrix} a\sigma_x^2 & a\sigma_{xy} \\ a\sigma_{xy} & a\sigma_y^2 \end{pmatrix}\right) \tag{2.2}$$

<u>Linear model</u>. Next, we assume that the means for the covariate $\mathbf{x}_{ak}$ can be described by a linear model,

$$E(\mathbf{x}_{ak}) = \mathbf{D}_a\boldsymbol{\theta} \tag{2.3}$$

where $\mathbf{D}_a$ is an $\underline{a} \times p$ design matrix that will differ for each $\underline{a}$, p is the total number of parameters in the model, and $\boldsymbol{\theta}$ is the unknown p-vector of parameters. One useful model assumes that $E(x_{akj})$ depends on <u>additive</u> age and environment effects,

$$E(x_{akj}) = (\text{age effect})_{a+1-j} + (\text{year effect})_j \tag{2.4}$$

so that the growth in a particular year depends on the age of the fish and the year in which it attained that age. For this model, if m is the maximum age of a fish in the sample, $p = 2m - 1$, with the first m components of θ corresponding to age effects, and the last m-1 components of θ corresponding to year or environmental effects (we set the environmental effect in the current year to zero to make the model full rank). As an example, the design matrix for an $\underline{a} = 3$ year old fish with m = 6 is

$$\mathbf{D}_3 = \begin{bmatrix} 0&0&1&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0&0&1 \\ 1&0&0&0&0&0&0&0&0&1&0 \end{bmatrix} \tag{2.5}$$

Alternative models are possible, allowing for general or specialized interactions between age and environment by suitably modifying $\mathbf{D}_a$.

A similar linear model will be assumed for the unobserved $\mathbf{y}_{ak}$, namely

$$E(\mathbf{y}_{ak}) = \mathbf{D}_a\eta \tag{2.6}$$

assuming the same design matrix, $\mathbf{D}_a$, but of course η is not the same as θ. Later, we add further assumtions linking θ to η.

The analysis can be divided into two parts: an analysis based on fish totals (t_{ak} and z_{ak}) and an analysis based on the yearly measurements, the $\mathbf{x}_{ak}$ and the unobserved $\mathbf{y}_{ak}$. For the analysis based on totals, we see easily that

$$E(z_{ak}) = \mathbf{1}^T\mathbf{D}_a\eta$$
$$E(t_{ak}) = \mathbf{1}^T\mathbf{D}_a\theta$$

From the fish totals, not all parameters need be estimable. In particular, for the additive model,

$$E(z_{ak}) = \sum_{i=1}^{a}(\eta_i + \eta_{2m+1-i}) \tag{2.7}$$

$$E(t_{ak}) = \sum_{i=1}^{a}(\theta_i + \theta_{2m+1-i})$$

where we take $\eta_{2m} = \theta_{2m} = 0$. Thus, if $a > 1$,

$$E(z_{ak}) - E(z_{a-1,k}) = \eta_a + \eta_{2m-a}$$

so each age coefficient ($a>1$) is confounded with an environmental coefficient. It follows, therefore, that separation of age of environmental effects is impossible from the fish totals alone with a single sample of fish. Estimates are possible from samples from more than one year as in Section 5,or by adding assumptions.

Combining all the results so far, we have for the observed data,

$$\begin{pmatrix} \mathbf{x}_{ak} \\ z_{ak} \end{pmatrix} \sim N\left(\begin{pmatrix} D_a\theta \\ 1^T D_a^T \eta \end{pmatrix}, \begin{pmatrix} \sigma_x^2 I & \sigma_{xy} 1 \\ \sigma_{xy} 1^T & a\sigma_y^2 \end{pmatrix}\right) \tag{2.8}$$

3. Estimation

Given the distribution (2.8), it is not hard to write down the likelihood for $\mathbf{x}_{ak}$, z_{ak}. This likelihood in turn can be rewritten as a product of the marginal likelihood based on the $\mathbf{x}_{ak}$ times the conditional likelihood based on z_{ak} given $\mathbf{x}_{ak}$. Usual multivariate normal calculations give

$$\begin{aligned} &\mathbf{x}_{ak} \sim N(D_a\theta,\ \sigma_x^2 I) \\ &z_{ak}|\mathbf{x}_{ak} \sim N\left(1^T D\ \eta_a + \beta(t_{ak} - 1^T D_a\theta),\ a\sigma_{y|x}^2\right) \end{aligned} \tag{3.1}$$

where we have defined the new parameters $\beta = \sigma_{xy}/\sigma_x^2$ and $\sigma_{y|x}^2 = \sigma_y^2 - \sigma_{xy}^2/\sigma_x^2$. When mle's of these parameters are found, these equations can be inverted to give estimates of the original parameters, if needed. However, the following analysis will use the conditional distribution defined by (3.1).

We now proceed to find the mle's without additional assumptions concerning η. Defining

$$\beta_{0a} = 1^T D_a(\eta - \beta\theta) \tag{3.2}$$

the log-likelihood, say $\ell = \ell_1 + \ell_2$, where

$$\begin{aligned} \ell_1 &= -\frac{n_x}{2}\log(2\pi) - \frac{n_x}{2}\log(\sigma_x^2) - \frac{1}{2\sigma_x^2}\sum\sum(x_{ak} - D_a\theta)^T(x_{ak} - D_a\theta) \\ \ell_2 &= -\frac{n_f}{2}\log(2\pi) - \sum\frac{a}{2}\log(a\sigma_{y|x}^2) - \frac{1}{2\sigma_{y|x}^2}\sum\frac{1}{a}\sum(z_{ak} - \beta_{0a} - \beta t_{ak})^2 \end{aligned} \tag{3.3}$$

and $n_x = \Sigma a n_a$ and $n_f = \Sigma n_a$. Determining the mle's from this model is easy since information concerning θ is available only from ℓ_1, and information concerning the β's is only in ℓ_2. The mle's for all parameters in (3.3) are given in Appendix A1. $\hat{\underset{\sim}{\theta}}$ and $\hat{\sigma}_x^2$ are determined by usual regression calculations, while $\hat{\beta}$, $\hat{\sigma}_{y|x}^2$, and

$$\hat{\beta}_0 = (\hat{\beta}_{01}, \ldots, \hat{\beta}_{0m})^T$$

are computed by weighted least squares. Usual formulas are available for estimating standard errors.

<u>Estimation of η</u>. Let D^* be an m×p matrix with i-th row $1^T D_a$. Then, by

(3.2),

$$\hat{\beta}_0 = D^*(\hat{\eta} - \hat{\beta}\hat{\theta}) \tag{3.4}$$

or

$$D^*\hat{\eta} = \hat{\beta}_0 + \hat{\beta}D^*\hat{\theta} \tag{3.5}$$

Since D^* will not have full column rank, the elements of η are not estimable, but description of the set of all mle's of η is useful. Partition $D^* = (D_1\ D_2)$ so D_1 is p×p positive definite, and $D^* = D_1(I,$ $D_1^{-1}D_2$. If we partition $\eta^T = (\eta_1^T\ \eta_2^T)$ and $\theta^T = (\theta_1^T\ \theta_2^T)$ conformably, so η_1 and θ_1 are age parameters, (3.5) becomes

$$D_1(I\ D_1^{-1}D_2)\hat{\eta} = \hat{\beta}_0 + \hat{\beta}D_1(I\ D_1^{-1}D_2)\hat{\theta}$$

$$\hat{\eta}_1 + D_1^{-1}D_2\hat{\eta}_2 = D_1^{-1}\hat{\beta}_0 + \hat{\beta}\hat{\theta}_1 + \hat{\beta}D_1^{-1}D_2\hat{\theta}_2$$

$$\hat{\eta}_1 = D_1^{-1}\hat{\beta}_0 + \hat{\beta}\hat{\theta}_1 + D_1^{-1}D_2(\beta\hat{\theta}_2 - \hat{\eta}_2)$$

Thus, the set of all possible solutions is

$$\begin{pmatrix} \hat{\eta}_1 \\ \hat{\eta}_2 \end{pmatrix} = \begin{pmatrix} D_1^{-1}\hat{\beta}_0 + \hat{\beta}\hat{\theta}_1 + D_1^{-1}D_2(\beta\hat{\theta}_2 - \eta_2) \\ \eta_2 \end{pmatrix} \tag{3.6}$$

for all η_2. If all $n_a > 0$ and model (2.3), (2.4) is used, each element η_a, a=1,...,m, can be written explicitly as

$$\hat{\eta}_1 = \hat{\beta}_{01} + \hat{\beta}\hat{\theta}_1$$
$$\hat{\eta}_2 = \hat{\beta}_{02} - \hat{\beta}_{01} + \hat{\beta}\hat{\theta}_2 + (\hat{\beta}\hat{\theta}_p - \eta_p) \tag{3.6}$$
$$\hat{\eta}_a = \hat{\beta}_{0a} - \hat{\beta}_{0,a-1} + \hat{\beta}\hat{\theta}_a + (\hat{\beta}\hat{\theta}_{m-a} - \eta_{m-a})$$
$$- (\hat{\beta}\hat{\theta}_{m-a+1} - \eta_{m-a+1}) \quad a=3,\ldots,m.$$

4. Linking parameters

Thus far the models for the lengths and covariate have been left (nearly) separate, allowing only estimation of η in a subspace. We can obtain unique estimates by assuming that $\eta_2 = \gamma\theta_2$, so that environmental effects are proportional for covariate and length, but make no such assumption for age effects. Examination of (3.5) will show immediately that if we take $\hat{\gamma}=\hat{\beta}$, we get a solution in the class (3.5), and hence a maximum likelihood estimate, although, γ is not estimable from a single sample of fish. Under this model, an estimate of $\hat{\eta}$, say $\hat{\eta}_E$ is

$$\hat{\eta}_E = \begin{pmatrix} \hat{\eta}_1 \\ \\ \hat{\eta}_2 \end{pmatrix} = \begin{pmatrix} D_1^{-1}\hat{\beta}_0 + \hat{\beta}\hat{\theta}_1 \\ \\ \hat{\beta}\hat{\theta}_2 \end{pmatrix}. \tag{4.1}$$

The variance of $\hat{\eta}_E$ can be computed as follows:

$$\mathrm{Var}(\hat{n}_E) = E\{\mathrm{Var}(\hat{n}_E|\mathbf{x})\} + \mathrm{Var}\{E(\hat{n}_E|\mathbf{x})\}$$

$$= E\left\{\mathrm{Var}\left.\begin{pmatrix} D_1^{-1}\hat{\beta}_0+\hat{\beta}\hat{\theta}_1 \\ \hat{\beta}\hat{\theta}_2 \end{pmatrix}\right| \mathbf{x}\right\} + \mathrm{Var}(D_1^{-1}\beta_0+\beta\hat{\theta})$$

$$= E\begin{bmatrix} D_1^{-1}\mathrm{Var}(\hat{\beta}_0)D_1^{-T} & 0 \\ 0 & 0 \end{bmatrix} + \mathrm{Var}(\hat{\beta})\hat{\theta}\hat{\theta}^T + \beta^2\mathrm{Var}(\hat{\theta})$$

$$+ \begin{bmatrix} D_1^{-1}\mathrm{Cov}(\hat{\beta}_0,\hat{\beta})\theta_1^T+\theta_1^T\mathrm{Cov}(\hat{\beta}_0,\hat{\beta})D_1^{-T} & D_1^{-1}\mathrm{Cov}(\hat{\beta},\hat{\beta}_0)\hat{\theta}_2^T \\ \hat{\theta}_2\mathrm{Cov}(\hat{\beta}_0,\hat{\beta})D_1^{-T} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} D_1^{-1}\mathrm{Var}(\hat{\beta}_0)D_1^{-1}+D_1^{-1}\mathrm{Cov}(\hat{\beta}_0,\hat{\beta})\hat{\theta}_1^T+\theta_1\mathrm{Cov}(\beta,\beta)\bar{D}_1^T & D_1^{-1}\mathrm{Cov}(\hat{\beta}_0,\hat{\beta})\hat{\theta}_2^T \\ \theta_2^T\mathrm{Cov}(\hat{\beta},\hat{\beta}_0)D_1^{-T} & 0 \end{bmatrix}$$

$$+ \mathrm{Var}(\theta)[\hat{\beta}^2+\mathrm{Var}(\hat{\beta})] + \mathrm{Var}(\hat{\beta})\theta\theta^T \tag{4.2}$$

An estimate of $\mathrm{Var}(\hat{n}_E)$ is obtained by substituting estimates for parameters. If the environmental effects are small relative to the age effects, an approximate estimate of $\mathrm{Var}(\hat{n}_{E,a})$ $a=1,\ldots,m$, ignoring relatively small terms, is

$$\begin{aligned} \mathrm{Var}(\hat{n}_{E1}) &= \sigma^2_{y|x}\left(\frac{a}{n_1}\right) + \hat{\beta}^2\mathrm{Var}(\hat{\theta}_1) \\ \mathrm{Var}(\hat{n}_{Ea}) &= \sigma^2_{y|x}\left(\frac{a}{n_a} + \frac{a-1}{n_{a-1}}\right) + \hat{\beta}^2\mathrm{Var}(\hat{\theta}_a) \quad a=2,\ldots,m \end{aligned} \tag{4.3}$$

All of the estimates derived so far with the exception of (4.2), are easily computed using standard regression software.

<u>More structure on η</u>. It is closer to the usual methods used in the fisheries literature to assume a linear model linking η to θ,

$$\eta = \alpha_0 + \alpha_1\theta \tag{4.4}$$

but in general we can take

$$\eta = g(\theta,\alpha) \tag{4.5}$$

for some monotonic function g. Given such a model, estimation will be more complex since an iterative procedure will be usually required to compute maximum likelihood estimates. Using (4.4), the log-likelihood is $\ell_1 + \ell_3$, where

$$\ell_3 = -\frac{n_f}{2}\log(2\pi) - \frac{1}{2}\sum n_a \log(a\sigma^2_{y|x})$$
$$-\frac{1}{2\sigma^2_{y|x}} \sum_{a=1}^{m} \frac{1}{a} \sum_{k=1}^{n_a} (z_{ak} - \alpha_0(1^T D_a 1) - \alpha_1(1^T D_a \hat{\theta}) - \beta t_{ak})^2$$

The two parts of the log-likelihood can no longer be maximized separately. However, this approach will be less attractive in practice because the information from the t_{ak} and z_{ak} has little influence in the estimation of θ, but only on α_0 and α_1. For this reason, we omit the computing details for this alternative method.

5. Example

Table 1 gives data on 78 Bluegills captured in Lake Mary, Wright County, Minnesota, in June, 1981. The data on each fish consists of age at capture, length at capture (t_{ak}), and scale radius increments for each of the a years, most recent increase first (x_{ak}). The method of measurement is described by Frie (1982). The fish in the sample range in age from 1 to 6 years old at capture, and the n_a range from 1 to 41; see summary statistics in Table 2. In the data, the x_{ak1} actually consist of the increment in scale radius between the last two annuli plus any additional growth in the current year. The annuli are typically laid down in the Spring, so some additional growth will be apparent on most fish. This will increase the value of the year effect for the current year, by confounding it with the increment that would be included in the growth for the following year. We return to this additional complexity in the next section.

We begin the analysis by considering the linear model (2.3), (2.4) for the covariate. Residual and influence analysis (Cook and Weisberg, 1982, Chapters 2 and 3), not reported here in any detail, do not suggest the need to transform the response, nor are any outliers nor overly influential observations evident. The fit of the linear model to the covariate is summarized in Table 3. We see that the age effects, $\hat{\theta}_1, \ldots, \hat{\theta}_6$, decrease with age, and, except for $\hat{\theta}_6$, all have fairly small standard errors. The year/environment effects, $\hat{\theta}_7$ to $\hat{\theta}_{11}$, are comparatively smaller, as one might expect, and are all negative; by the parameterization, the environmental effects are based on comparisons to the current year.

The analysis of variance summary shown in Table 3 indicates a test for lack of fit that can be computed using the sum of squares $\sum\sum\sum(x_{akj} - \bar{x}_{ak+})^2$ to compute a "pure error" estimate of σ_x^2 (Weisberg, 1980). We see from this test that the linear model used is not adequate, which may suggest the need for a more complicated model. We will not persue this need for model enhancement here, since the model does explain most of the variation (R^2 = .580, with a maximum value possible of R^2 = .642), and the lack of fit is likely only to make small adjustments to the analysis.

Table 3 summarizes the analysis based on fish totals. Referring to Figure 1, we now see that this graph should be modelled by 6 parallel lines, with intercepts given by the $\hat{\beta}_{0a}$, and common slope $\hat{\beta}$.

Table 4 gives $\hat{\eta}$ computed given $\hat{\eta}_2 = \hat{\beta}\hat{\theta}_2$, along with the standard errors computed from (3.8); approximation (3.9) gives very similar results. Growth is seen to decrease with age. The fitted total lengths, computed from the $\hat{\eta}$, for each age cohort at each age are given also in Table 4, along with standard errors. Of the 21 averages in this table, 8 differ from the "usual" values (Frie, 1982) obtained by the method outlined in Section 1 by at least 2 standard errors.

6. Conclusions

The method proposed here for backcalculation is quite different from the standard in the fisheries literature. The assumptions of a common intercept and separate growth rates for each fish have been replaced by a linear model for age and environmental effects.

Although the η's are not estimable from a single sample of fish without further assumptions, the θ's are estimable, and they could be used to test for environmental or management changes without first obtaining backcalculated lengths. Such tests would make use of the available information without requiring further assumptions.

Also, η <u>will</u> be estimable from the fish totals if samples of fish are drawn from the same lake in repeated seasons. Consequently, it is recommended that such repeated samples be taken, if only to verify the applicability of the method proposed here.

The effect of current season growth has been confounded with the environmental effect of the current year. Alternatively, this could be modelled by addition of one or more parameters to the model; one parameter will suffice if current year growth can be assumed not age dependent. It is unclear whether or not this added complexity will be required. As a practical matter, samples collected in the period of about December to May can avoid this problem entirely.

All of the assumptions used in the analysis are severe, and may not be appropriate in practice. Nearly all assumptions can be modified, at the cost of increased complexity. The true test of this methodology will come from its application in other fish populations.

Appendix

A1. Estimates for model (2.3),(2.4). Let

$$\mathbf{D} = \sum_{a=1}^{m} n_a D_a^T D_a \qquad \text{(A.1)}$$

$$\mathbf{Dx} = \sum_{a=1}^{m} n_a D_a^T \bar{x}_{a+}$$

where $\bar{\mathbf{x}}_{a+} = \sum x_{ak}/n_a$ is an a-vector of averages. Then, for the covariate parameters,

$$\hat{\theta} = \mathbf{D}^{-1}(\mathbf{Dx})$$

$$\hat{\sigma}_x^2 = \sum_a \sum_k (\mathbf{x}_{ak} - \mathbf{D}_a\hat{\theta})^T(\mathbf{x}_{ak} - \mathbf{D}_a\hat{\theta})/n = RSS_x/n \ . \qquad \text{(A.2)}$$

The intermediate quantities (A.1) are easily computed using usual regression software by creating a long vector **X** of length $n_x = \sum n_a$ by stringing the $\mathbf{x}_{ak}$ one below the next, and by defining an $n_x \times p$ matrix $\tilde{\mathbf{D}}$ by appending the $\mathbf{D}_a$ in the same order as was used for **X**. Then the regression of **X** on $\tilde{\mathbf{D}}$ will give $\hat{\theta}$ and $\hat{\sigma}_x^2$. For $\mathbf{D}_a$ given by (2.4), this method is not efficient, and direct calculation of (A.1) is preferable. The Cholesky factorization of **D** to compute (A.2) is easily obtained using LINPACK using routine SCHUD (Dongarra et al, 1977).

The remaining parameters, β, β_0, $\sigma_{y|x}^2$ require weighted least squares calculations based only on fish totals. Again, the calculations are easily carried out using standard software, with data on each fish consisting of

(1) its total length, z_{ak}, (2) its total scale radius, t_{ak}, (3) its age and (4) a set of m dummy variables, the a-th of which has value 1 for an a year old fish, all the others being zero. The regression of z_{ak} on t_{ak}, and the dummy variables, through the origin and weighed by 1/a, will give all estimates of parameters. Explicit formulae are as follows. Let

$$STT = \sum \frac{1}{a} \sum (t_{ak} - \bar{t}_{a+})^2$$

$$STZ = \sum \frac{1}{a} \sum (t_{ak} - \bar{t}_{a+})(z_{ak} - \bar{z}_{a+})$$

$$SZZ = \sum \frac{1}{a} \sum (z_{ak} - \bar{z}_{a+})^2$$

Then,

$$\hat{\beta} = STZ/STT$$

$$\hat{\beta}_{0a} = \bar{z}_{a+} - \hat{\beta}\,\bar{t}_{a+}$$

$$\hat{\sigma}^2_{y|x} = (SZZ - \hat{\beta}^2/STT)/n_f$$

(Usual software will use a divisor of $(n_f - a)$ in place of n_f for $\hat{\sigma}^2_{y|x}$. Either form is acceptable.)

Bibliography

Aitkin, M. (1981). "A note on the regression analysis of censored data", Technometrics 23, 161-163.

Bartlett, J. R., Randerson, P. F., Williams, R., and Ellis, D. M. (1984). "The use of analysis of covariance in the back-calculation of fish growth. J. Fish. Biol., 24, 201-13.

Carlander, K.D. (1981). "Caution on the use of the regrèssion method of back-calculating lengths from scale measurements" Fisheries 6, 2-4.

Carlander, K.D. (1982). "Standard intercepts for calculating lengths from scale measurements for some centrarchid and percid fish" Transactions of the American Fisheries Society 111, 332-336.

Cook, R.D. and Weisberg, S. (1982). Residuals and Influence in Regression. London and New York: Chapman-Hall.

Dongarra, J., Bunch, J.R., Moler, C.B. and Stewart, S.W. (1979). The LINPACK Users Guide. Philadelphia:SIAM.

Duncan, K.W. (1980). "On the back-calculation of fish lengths; modifications and extensions to the Fraser-Lee equation" Fisheries Society of the British Isles, 725-730.

Frie, Richard (1982). "Measurement of fish scales and backcalculation of body lengths using a digitizing pad and microcomputer: Fisheries 7, 5-8.

Hale, Ralph (1970). "Body-scale relation and calculation of growth in fishes" Transactions, American Fisheriesociety, 468-474.

Kennedy, W. and Gentle, J. (1980). Statistical Computing. New York: Dekker.

Ricker, W.E. (1975). "Computation and interpretation biological statistics of fish populations" Ottawa: Dept. of the Environment Fisheries and Marine Service Bulletin #191.

Schmee, J. and Hahn, J. (1979). "A simple method for regression analysis with censored data" Technometrics 21, 417-432.

Weisberg, S. (1980a). Applied Linear Regression. New York: Wiley.

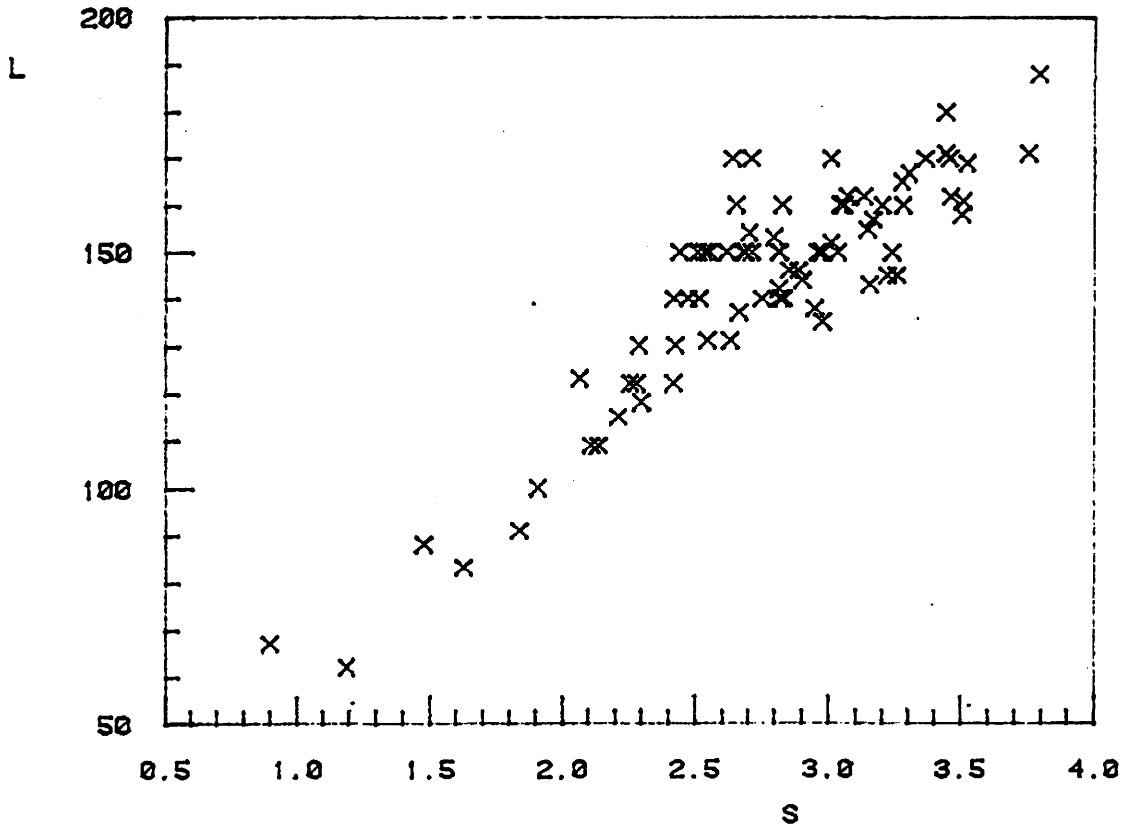
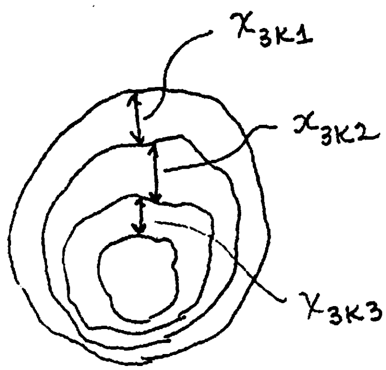


Figure 1. Total length, L, versus total scale radius, S, for 78 Bluegills.
(Source: R. Frie)

Figure 2

scale increments



length increments

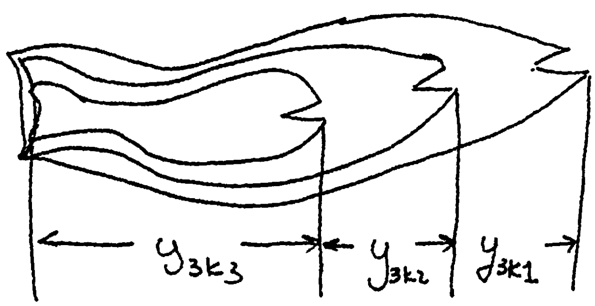


Table 1. Age ($\underline{a}$), length (t_{ak}) and scale radius increments (x_{ak}) for 78 bluegills (Source: R. Frie)

	a	t_{ak}	x_{ak}			
1	1	67.	.89565			
2	1	62.	1.18563			
3	2	109.	1.16264	.94200		
4	2	83.	1.06000	.56418		
5	2	91.	.91950	.91584		
6	2	88.	.58768	.88550		
7	2	123.	1.23648	.82861		
8	2	100.	.62759	1.27505		
9	2	109.	1.15402	.97780		
10	3	137.	.64573	.80386	1.21270	
11	3	131.	1.22368	.77625	.62998	
12	3	122.	1.23220	.60218	.44559	
13	3	122.	1.20180	.63200	.42302	
14	3	118.	1.03872	.73841	.51950	
15	3	115.	1.09591	.55270	.56159	
16	3	131.	.90666	.81443	.82484	
17	3	143.	1.06900	.93330	1.14889	
18	3	142.	.96284	.90991	.93541	
19	3	122.	1.41237	.55664	.44770	
20	3	140.	.57339	1.02448	1.14557	
21	3	150.	1.06616	.61789	.84482	
22	3	140.	.91923	.78998	.75961	
23	3	150.	.90172	.84111	.94100	
24	3	150.	.91540	.77027	.92725	
25	3	140.	1.07718	.79677	.93864	
26	3	150.	.89100	.91584	1.22241	
27	3	130.	.80005	.76616	.85675	
28	3	130.	.91527	.67605	.69509	
29	4	138.	.38935	.59632	.77020	1.18864
30	4	135.	.66205	.26780	.82495	1.21923
31	4	146.	.36134	.47730	.87500	1.17045
32	4	146.	.37827	.41691	.80216	1.25136
33	4	145.	.72325	.41911	.80341	1.27227
34	4	145.	.66968	.59573	.67545	1.31232
35	4	144.	.62346	.48820	.74170	1.04645
36	4	140.	.42809	.63773	.73914	1.01341
37	4	150.	.70900	.28327	.89039	1.08270
38	4	152.	.48170	.39666	.93084	1.19766
39	4	157.	.28197	.76841	.79657	1.31752
40	4	155.	.43382	.74841	.72757	1.23211
41	4	153.	.52572	.53184	.74770	.98641
42	4	154.	.73754	.19395	.66743	1.09932
43	4	158.	.71050	.41214	.91484	1.46277
44	4	162.	.42018	.38061	1.04202	1.22186
45	4	161.	.54447	.51591	.95230	1.49352

Table 1 (continued)

46	4	162.	.77352	.30125	.96830	1.41498		
47	4	165.	.61102	.66730	.57143	1.42177		
48	4	171.	1.05304	.67555	.60841	1.10418		
49	4	162.	.64407	.79795	.72916	.95748		
50	4	169.	.56120	.77352	1.06009	1.12368		
51	4	167.	.57327	.48855	1.10895	1.12945		
52	4	150.	.51825	.62064	.87686	.95895		
53	4	170.	.28828	.68095	1.17000	1.31520		
54	4	140.	.36819	.39325	.82475	.92282		
55	4	140.	.32046	.40675	.79786	.89093		
56	4	150.	.56641	.45500	.80491	.84707		
57	4	150.	.57289	.51645	.74732	.97748		
58	4	150.	.49007	.32500	.76373	.85902		
59	4	160.	.87170	.43873	.54623	.96639		
60	4	150.	.27243	.27082	.90473	1.10016		
61	4	150.	.40316	.42720	.83595	.83927		
62	4	150.	.36814	.68082	.84714	.91445		
63	4	150.	.58650	.47259	.82252	1.07107		
64	4	140.	.44209	.36918	.97120	1.04675		
65	4	160.	.53086	.48843	1.12639	1.12991		
66	4	170.	.57250	.58280	.65759	1.19314		
67	4	160.	.34082	.64220	1.00495	1.06318		
68	4	160.	.33966	.63830	.60957	1.06316		
69	4	170.	.41291	.76780	.98305	1.19639		
70	5	171.	.44241	.27184	.47780	1.07432	1.48798	
71	5	188.	.58652	.64682	.50661	1.29943	.75109	
72	5	170.	.46634	.43807	.60143	.49505	.70648	
73	5	150.	.55993	.27095	.40916	.74484	1.25132	
74	5	150.	.47480	.58307	.67884	.44127	.52886	
75	5	160.	.63980	.34925	.48882	.52775	1.18945	
76	5	160.	.48816	.20430	.23168	.99486	1.11868	
77	5	180.	.39173	.35952	.47880	.93145	1.27793	
78	6	170.	.25909	.12164	.20761	.50466	.44593	1.09670

Table 2. Summary statistics

		Unadjusted Analysis based on fish totals				
		Total Scale Radius		Total Length		
Age	n_f	Average	SD	Average	SD	Corr.
1	2	1.040640	.205047	64.500000	3.535534	-1.0000
2	7	1.876699	.252161	100.428571	14.164005	.8487
3	19	2.570889	.263066	134.894737	11.474100	.7434
4	41	2.997337	.310879	153.829268	9.901774	.6107
5	8	3.233426	.417119	166.125000	13.653231	.6378
6	1	2.635630	0	170.000000	0	0

Scale Radius increments

Age at capture			Age at measurement 1	2	3	4	5	6
1	2	AVE	1.0406					
		SD	.2050					
2	7	AVE	.9127	.9640				
		SD	.2104	.2632				
3	19	AVE	.8148	.7641	.9920			
		SD	.2618	.1321	.2027			
4	41	AVE	1.1238	.8352	.5125	.5259		
		SD	.1695	.1504	.1581	.1710		
5	8	AVE	1.0390	.8136	.4841	.3905	.5062	
		SD	.3352	.3111	.1318	.1562	.0822	
6	1	AVE	1.0967	.4459	.5047	.2076	.1216	.2591

Table 3. Linear model for scale radius increments, $\hat{\theta}$.

	Estimate	Std. Error	t value
Age effects			
1	1.209185	.047947	25.22
2	1.082760	.040986	26.42
3	.869593	.034680	25.08
4	.555488	.028110	19.76
5	.499255	.064615	7.73
6	.259090	.193456	1.34
Year effects			
6	-.112485	.19309	-.56
5	-.222058	.079085	-2.81
4	-.120395	.052083	-2.31
3	-.305450	.043780	-6.98
2	-.321963	.036842	-8.74

Analysis of Variance on Scale Radius Increments

Source	DF	SS	MS	R^2
Regression	11	14.657407	1.332492	.5805
Lack of Fit	10	1.553525	.155352	
Pure Error	262	9.037781	.034495	
Residual	272	10.591306	.037425	
Total	282	25.248714		

Regression of Length on Total Scale Radius

Coefficient	Value	SE
$\hat{\beta}_{01}$	38.895578	4.594937
$\hat{\beta}_{02}$	54.253350	6.521553
$\hat{\beta}_{03}$	71.639298	8.511063
$\hat{\beta}_{04}$	80.081313	9.807153
$\hat{\beta}_{06}$	86.568178	11.040891
$\hat{\beta}_{06}$	105.151653	13.779959
$\hat{\beta}_{1}$	24.604496	3.239437

Number of Fish= 78
Var(scale rad)= .0374
SE (scale rad)= .1935
Var(length) = 19.4984
SE (length) = 4.4157

Table 4. Age and year effects $\hat{\eta}$ for length, assuming that $\eta_2 = \beta\theta_2$.

	Estimate	Std. Error	t value
Age effects			
1	68.646957	3.385727	20.28
2	41.998535	4.122313	10.19
3	38.781850	3.116606	12.44
4	22.109526	2.375369	9.31
5	18.770778	4.169761	4.50
6	24.958254	12.646402	1.97
Year effects			
6	-2.767629	4.959618	-.56
5	-5.463613	2.090307	-2.61
4	-2.962251	1.350087	-2.19
3	-7.515437	1.469537	-5.11
2	-7.921746	1.386995	-5.71

Length at age by age cohort

		Age					
Cohort		1	2	3	4	5	6
1	AVE	68.65					
	SD	3.39					
2	AVE	60.73	102.72				
	SD	3.32	2.85				
3	AVE	61.13	95.21	133.99			
	SD	3.27	2.69	2.19			
4	AVE	65.68	100.17	131.03	153.14		
	SD	3.20	2.53	2.12	1.86		
5	AVE	63.18	102.22	133.49	147.67	166.44	
	SD	3.51	3.01	2.73	2.82	4.47	
6	AVE	65.88	102.41	138.23	152.83	163.68	188.64
	SD	5.73	5.65	5.64	5.72	6.87	1