SOME PROBLEMS IN ESTIMATING POPULATION SIZES FROM CATCH-AT-AGE DATA

Roy Mendelssohn[1]

ABSTRACT

A new method for estimating population sizes from catch-at-age data is given. The method treats the observed population sizes as missing data and uses a combination of the Kalman filter and the EM algorithm to derive maximum likelihood estimates of the parameters and minimum mean square error estimates of the population sizes. The algorithm does not assume that the observation errors and the errors in the population dynamics are uncorrelated with equal variances, which is a common assumption of existing techniques. A new parameterization for both recruitment and fishing mortality is given, based on smoothness priors. Recruitment (or fishing mortality) is estimated as a nonparametric function of time by calculating an "optimal" tradeoff between goodness-of-fit and smoothness of the function. The algorithm allows for multiple sources of observations (fishing, surveys, etc.) and allows for missing data in the observations, which can arise if the different sources of the observations occur on different time scales. An example suggests that the new algorithm may better capture variation that is important when using the population estimates to study the role of the environment (or other exogenous variables) on the population dynamics.

I can address the motivation of this paper by considering a slightly modified version of a model proposed by Collie and Sissenwine (1983). Assume that the underlying population dynamics satisfy

$$N(a + 1, t + 1) = [N(a,t) - C(a,t)] \, \tilde{m} + w(a,t)$$

$$a = 1, A \tag{1}$$

where $N(a,t)$ is the number of fish age a at time t, $C(a,t)$ is the catch of age a fish at time t, $\tilde{m} = exp(-m)$ is the mortality rate and the vector $w(t) = (w(1,t), \ldots, w(A,t))^\tau$ is a sequence of independent, identically distributed normal random vectors with mean 0 and covariance matrix Q, and for any vector a, the notation a^τ denotes the transpose of the vector. We assume that the initial population vector $N(0)$ is gaussian with a mean of μ and covariance Σ.

The population itself is not observed. Instead we observe that

$$n(a,t) = q(t)N(a,t) + v(a,t) \tag{2}$$

where q is an unknown parameter and the vector $v(t) = (v(1,t), \ldots, v(A,t))^\tau$ is a sequence of independent, identically distributed normal random vectors with mean 0 and covariance matrix R. It is assumed that $E(v(t) \, w(t)^\tau) = 0$; that is, the observation error and the underlying randomness in the population are uncorrelated. There is some interest in the value of the estimates of the $q(t)$ (or if mortality is to be estimated, in the estimate of $\tilde{m}$) but the major interest lies in estimating the unobserved population sizes $N(a,t)$. The estimates of the $N(a,t)$ should reflect not only the trend in the population, much as a regression might, but also the period-to-period variation of the population, such as might be related to environmental changes. This will be the emphasis throughout the paper.

The model described in Equations (1) and (2) differs from that of Collie and Sissenwine (1983) in that I do not assume that the mortality rate is known; here I allow the underlying population dynamics to be random, and the observation errors in Equation (2) to be additive rather than multiplicative. For known $\tilde{m}$, Collie and Sissenwine (1983) suggested minimizing

$$\sum_{t=1}^{T} \sum_{a=1}^{A} (v(a,t)^2 + w(a,t)^2) \tag{3}$$

over the parameters (in our notation) $\theta = (q, N(t))$

where $N(t)$ is the vector of unobserved population sizes. The concern here is not whether the multiplicative or additive form of the model is more correct, but rather what elements of the model are data, what are parameters, and how to calculate appropriate estimates of each.

If there were all-seeing observers, both the population process $N(t)$ and the observation process $n(t)$ would be available as data. Equations (1) and (2) imply a sequence of conditional probability distributions $f_N(N(t + 1) \mid N(t), \theta_1)$ and $f_n(n(t) \mid N(t), \theta_2)$, where θ_1, θ_2 are parameters of the distributions to be estimated. Assuming additive gaussian errors, then θ_1 would include the mean vector $FN(t)$ and a covariance matrix that can be calculated recursively (see below). The mortality rate $\bar{m}$ serves as a constraint on the form of the estimates of the mean vector, much as in the theory of regression. Similarly, the parameters for the observation process are a mean vector $H(t)N(t)$ and a covariance matrix, where $H(t)$ in this case depends on the parameters $q(t)$ that constrain the estimates of the mean vector. With an all-seeing observer, both $n(t)$ and $N(t)$ are realized values of random vectors and hence, are the data to be used to estimate the unknown parameters of the distributions, θ_1 and θ_2.

Thus, the unobserved population sizes are most appropriately treated as missing data. The estimation scheme proposed by Collie and Sissenwine (1983) treats the $N(t)$ as parameters to be estimated. Little and Rubin (1983, 1987: sec. 5.4) showed that treating missing data as parameters in likelihood equations does not produce maximum likelihood estimates of the parameters unless the proportion of missing data approaches zero as the sample size increases. This is because much of the asymptotic theory of maximum likelihood estimation depends on the number of observations becoming large, relative to the number of parameters. Little and Rubin (1983, 1987) showed that for a regression-like situation, the bias due to treating data as parameters can be quite large.

The alternate approach discussed by Little and Rubin (1983, 1987) is to integrate out the missing data from the complete data likelihood and maximize this function over the parameters as usually defined in estimation theory. This is the approach taken by Shumway and Stoffer (1982a), who used the EM (expectation-maximization) algorithm of Dempster et al. (1977) and Kalman filtering to derive maximum likelihood estimates for the parameters of the model and minimum mean square error estimates of the missing data.

I can explain some of the problems with estimating Equations (1) and (2), using the likelihood of Equation (3). Under the gaussian assumptions of the model, the complete data log-likelihood is given by (Shumway and Stoffer 1982a)

$$-\frac{1}{2} \log | \Sigma | \; - \frac{1}{2} (N(0) - \mu(0))^{\tau} \Sigma^{-1}$$

$$(N(0) - \mu(0)$$

$$-\frac{T}{2} \log | Q | \; - \frac{1}{2} \sum_{t=1}^{T} (N(t) - FN(t - 1))^{\tau}$$

$$Q^{-1} (N(t) - FN(t - 1))$$

$$-\frac{T}{2} \log | R | \; - \frac{1}{2} \sum_{t=1}^{T} (n(t) - H(t)N(t))^{\tau} R^{-1} (n(t)$$

$$- H(t)N(t)) \tag{4}$$

where $F = \bar{m}I$ and $H(t) = q(t)I$. Similarly, the complete data log-likelihood in Equation (3) by substitution is

$$-\sum_{t=1}^{T} || N(t) - FN(t - 1) ||^2 + || n(t)$$

$$- H(t)N(t) ||^2. \tag{5}$$

Collie and Sissenwine (1983) noted that their estimation scheme assumes that the process and observation errors have the same variance. However, from Equations (4) and (5) it can be seen that they make the stronger and unlikely assumption that both the errors in the population dynamics and the errors in the observation process are uncorrelated. Further, we can see from Equation (4) that when the $N(t)$ are treated as parameters, the estimates of the $N(t)$ depend on the observed data $n(t)$ for $t = 1,T$. Following Shumway and Stoffer (1982a), the expected log-likelihood conditioned on the observed data comprises three parts: a term due to estimating the expected value of the initial population size,

$$-\frac{1}{2}\log|\Sigma| - \frac{1}{2}\,tr\,\{\Sigma^{-1}\,[P(0|T) + (N(0) - \mu)$$

$$\times\,(N(0) - \mu)^{\tau}]\}, \tag{6}$$

a term due to the unobserved dynamics,

$$-\frac{1}{2}\sum_{t=1}^{T}\log|Q(t)|$$

$$-\frac{1}{2}\sum_{t=1}^{T}tr\,\{Q(t)^{-1}\,[(N(t|T) - FN(t-1|T))$$

$$\times\,(N(t|T) - FN(t|T)^{T}$$

$$+ P(t|T) + FP(t-1|T)\,F^{\tau} - P(t-1|T)$$

$$\times\,F^{\tau} - FP(t, t-1|T)]\}, \tag{7}$$

and a term due to the observation process,

$$-\frac{1}{2}\sum_{t=1}^{T}\log|R(t)|$$

$$-\frac{1}{2}\sum_{t=1}^{T}tr\{R(t)^{-1}[(n(t) - H(t)N(t|T))$$

$$\times\,(n(t) - H(t)N(t|T))^{\tau}$$

$$+ H(t)P(t|T)H(t)]\} \tag{8}$$

where $N(t|s)$ denotes

$$N(t|s) = E[N(t) \mid n(1), \ldots, n(s)], \tag{9}$$

$P(t|s)$ denotes

$$P(t|s) = E[(N(t) - N(t|s))$$

$$\times\,(N(t) - N(t|s))^{\tau} \mid n(1),$$

$$\ldots, n(s)], \tag{10}$$

and $P(t, t-1|s)$ denotes

$$P(t, t-1|s) = E[(N(t) - N(t|s))$$

$$\times\,(N(t-1) - N(t-1|s))^{\tau}$$

$$\times\,\mid n(1), \ldots, n(s)]. \tag{11}$$

As shown in Equations (6) through (8), the proper estimates of the $N(t)$ and the related covariance matrices should be conditional expectations based on all of the data rather than on only the data up to time t. Assuming that all quantities are calculated properly, estimates that include only the data up to time $t - 1$ are termed "predicted" estimates, estimates that include only the data up to time t are termed "filtered" estimates, while estimates given all the data are termed "smoothed" estimates. I shall show below that the appropriate formulas for predicted, filtered and smoothed estimates differ significantly. Thus using Equation (5) as the likelihood and treating the $N(t)$ as parameters does not produce proper estimates of the $N(t)$.

In the rest of this paper, I review state-space models and methods for estimating both the parameters and the unobserved components of the model. A very readable background for what follows is chapter 3 in Shumway (1988). The estimation scheme described does not require that the components of R, Q, and Σ have equal variance and are uncorrelated. Explicit estimates of these matrices are given. Then I show that a variety of age-based models proposed in the literature can be formulated as a state-space model, but that the formulations as presented make the same error of treating the unobserved components as parameters, and assume zero covariance in the errors. Auxiliary information as in Deriso et al. (1985) can be put into this format. And I show that the state-space formulation can include multiple observations of the population, but where some of the observations are missing. This can arise when the population is observed from fishing and from a variety of surveys, but some of the surveys are not done every year. This is essentially the problem discussed in Methot[2]. I give true maximum likelihood estimates for this model, allowing the different observation processes to have different error structures and estimate the relative weight that should be given each. This is a significant advance over the procedure in Methot (fn. 2).

A related paper is the analysis of Brillinger et al. (1980) who use a modified Kalman filter and maximum likelihood estimation to estimate the average birth and death rates and population structure of Nicholson's blow-fly data when only total numbers

of births and deaths at any time are available. In their model, knowing the births is equivalent to knowing the recruitment at each time period and the deaths are observed directly without error. Also, they appear to use the filtered estimates of the population structure, while the smoothed estimates are the minimum mean squared estimates given all the data. P. Sullivan[3] in his Ph.D. dissertation independently developed a length-based fishery model using Kalman filtering and maximum likelihood estimation.

I reiterate that the models considered here assume additive errors, while much of the existing literature prefers multiplicative errors, particularly for the observation equation. The basis of this preference appears to be that models with multiplicative errors have given a better "fit" to the data for other estimation schemes. In these estimation schemes the errors are assumed to have equal variances and are assumed to be uncorrelated. The better fit found for multiplicative errors in these algorithms may be due to these assumptions on the error variances. In effect, the errors are being scaled by the observed data which suggests that the assumption of equal variances is incorrect. In the present formulation I assume additive errors, but I can allow for errors with unequal variances, for errors that depend on the size of either the observed catch $n(a, t)$ or on the unobserved population $N(a, t)$, and for missing observations, and I can obtain simple to compute standard errors of the estimated underlying population sizes. The algorithm is also simple to program. If multiplicative errors are assumed, it is more difficult to calculate exact maximum likelihood estimates. Approximate likelihood methods (such as the extended Kalman filter) have known undesirable properties. When the full richness of the assumptions allowed in additive error models is used, it is an open question if multiplicative errors are to be preferred.

STATE SPACE MODELS

The State-Space Model can be written in the form (Jazwinski 1970; Anderson and Moore 1979; Ljung and Soderstrom 1983):

$$x(t) = F(t)x(t - 1) + B(t)u(t - 1) + Gw(t) \quad (12)$$

$$y(t) = H(t)x(t) + v(t) \quad (13)$$

where $x(t) = (x_1(t), \ldots, x_p(t))^\tau$ is the p-dimensional unobserved state of the system; $u(t) = (u_1(t), \ldots, u_p(t))^\tau$ is a p-dimensional vector of deterministic inputs; $y(t) = (y_1(t), \ldots, y_q(t))^\tau$ is the observed data of the system; $w(t) = (w_1(t), \ldots, w_p(t))^\tau$ is a sequence of zero mean normal vectors with common covariance matrix Q; $v(t) = (v_1(t), \ldots, v_q(t))^\tau$ is a sequence of zero mean normal vectors with common covariance matrix R; and F, B, G, and H are appropriately dimensioned matrices that may depend on an unknown parameter vector θ. Note that q, the dimension of the observation vector, can be larger than p, the dimension of the state vector. Thus several different observation processes of the underlying dynamics are allowed.

Using the same notation as in Equations (9) and (10), the predicted, filtered, and smoothed estimates of the state vector and the covariance matrices can be calculated recursively as follows: for prediction and filtering,

$$x(t|t - 1) = F(t)x(t - 1|t - 1) + Bu(t - 1) \quad (14)$$

$$P(t|t - 1) = F(t)P(t - 1|t - 1)F(t)^\tau + GQ(t)G^\tau \quad (15)$$

$$K(t) = P(t|t - 1)H(t)^\tau \times (H(t)P(t|t - 1)H(t)^\tau + R(t))^{-1} \quad (16)$$

$$x(t|t) = x(t|t - 1) + K(t) \times (y(t) - H(t)x(t|t - 1)) \quad (17)$$

$$P(t|t) = P(t|t - 1) - K(t)H(t)P(t|t - 1) \quad (18)$$

where $x(0|0) = \mu$ and $P(0|0) = \Sigma$, and for smoothing,

$$J(t - 1) = P(t - 1|t - 1)H(t)^\tau(P(t|t - 1))^{-1} \quad (19)$$

$$x(t - 1|T) = x(t - 1|t - 1) + J(t - 1) \times (x(t|T) - x(t|t - 1)) \quad (20)$$

$$P(t - 1|T) = P(t - 1|t - 1) + J(t - 1) \times (P(t|T) - P(t|t - 1))J(t - 1)^\tau. \quad (21)$$

The predicted state variable $x(t|t - 1)$ differs from the quantity often used as the predicted value in the fisheries literature in that it is based on the last period's filtered estimate rather than on the last

period's predicted estimate. (For example, Deriso et al. (1985) suggested using backward VPA or cohort analysis to obtain predicted values.) The filtered estimate $x(t|t)$ is a weighted average of the predicted value of $x(t|t - 1)$ and the observed error in estimating $y(t)$, where the weighting term $K(t)$ (the Kalman gain matrix) is regression-like. Similarly, the covariance of the estimate, as measured by $P(t|t - 1)$, increases due to the prediction, but decreases by the amount $K(t)H(t)P(t|t - 1)$ after the observation has been made. The smoothed estimate $x(t|T)$, which is the correct estimate of the underlying population (it satisfies the conditional expectation), is found by a backward recursion on the filtered estimates, where the filtered estimate is adjusted by a regression on the error between the smoothed and predicted estimates of the following period. Thus the smoothed estimates correct the predicted estimates both by the error in predicting the observed data as well as by the error in predicting the underlying population when using the filtered estimates of the underlying population. The square roots of the diagonal terms of the various P matrices produced by the Kalman recursions are the standard errors of the predicted, filtered, and smoothed estimates of the population.

Equations (12) and (13) are the basic form of the state-space model. It is a simple extension to the model to allow any of the matrices F, B, G, or H to be nonlinear functions of the past values of the $y(t)$ (see, for example, Shiryayev (1984), section VI.7), to allow the error vectors $w(t)$ and $v(t)$ to depend on past values of the $y(t)$ (Shiryayev 1984), or to allow the $v(t)$ to depend on the underlying state vector $x(t)$ (Zehnwirth 1988).

In a typical fisheries problem, the matrix $H(t)$ represents fishing. If some age-specific measure of effort $E(a,t)$ is known, then $H(t)$ is a diagonal matrix with $E(a,t)$ on the diagonals. Or it may be assumed that the exploitation rate is of the form $s(a)E(t)$, where $E(t)$ is known and the $s(a)$ values are to be estimated. Then for given values of $s(a)$, the matrix $H(t)$ has $s(a)E(t)$ on its diagonals. The matrix $F(t)$ is formed in a similar manner to represent the population dynamics.

In some parameterizations, it is assumed that a known vector is subtracted from the state vector either before or after the effect of F on the population. For example, the known vector might be the catch from the previous time period. The extension to the Kalman filter in this case is straightforward, an example of which can be found in Jazwinski (1970). Essentially, all predicted estimates of the

state are corrected by the constant amount. The covariance and gain calculations are unaffected by the known vector.

The Kalman filter, Equations (14) through (21), assumes that the matrices F, B, G, H, R, Q, and Σ, and the vector μ are known. For fisheries problems, the matrices F and H usually depend on a set of parameters to be estimated (e.g., $F = mI$), and R, Q, and μ are to be estimated. Let θ be a vector containing the parameters that F and H depend on, and let $\Theta = (\theta, R, Q, \mu)$ be the total parameters of the model. Shumway and Stoffer (1982a) showed that conditional on Θ, the complete data likelihood is given by Equation (4). They apply a result of Dempster et al. (1977), which shows that maximum likelihood estimates of the parameters can be obtained by finding the conditional expectation (the E-step) of the complete data likelihood with respect to the missing "data" (in this case the missing data are the sufficient statistics of the normal distribution) and alternately estimating the expected value of the missing data, and then maximing the likelihood (the M-step) using the completed data. Shumway and Stoffer (1982a) showed that the expected conditional log-likelihood is given by Equations (6) through (8). All of the terms in this likelihood, for a given value of Θ, can be found by the Kalman filter. Moreover, given these values, the maximization problem is a deterministic one.

If we assume that the matrix F is independent of time and unrestricted, then Shumway and Stoffer (1982a) showed that the maximization step is accomplished by setting

$$F = S_t(1)S_{t-1}(0)^{-1} \tag{22}$$

$$Q = (S_t(0) - S_t(1)S_{t-1}^{-1}(0)S_t(1)^{\tau})/T \tag{23}$$

$$R = T^{-1} \sum_{t-1}^{T} [(y(t) - H(t)x(t|T)$$

$$\times (y(t) - H(t)x(t|T))^{\tau}$$

$$+ H(t)P(t|T)H(t)^{\tau}] \tag{24}$$

$$\mu = x(0|T). \tag{25}$$

$$\text{where} \quad S_t(j) = \sum_{t-1}^{T} (P(t,t - j|T)$$

$$+ x(t|T)x(t - j|T)^{\tau}). \tag{26}$$

Shumway and Stoffer (1982a) gave a recursive formula for calculating $P(t, t - 1|T)$ while performing the backward smoothing recursion.

If we assume that F is constrained to be of the form $F = \bar{m}D$ where $\bar{m}$ is a constant and D is a known matrix, then Shumway and Stoffer (1982b) showed that

$$Q = (C - \bar{m}BD^\tau - \bar{m}DB^\tau + \bar{m}^2DAD^\tau) \quad (27)$$

$$\bar{m} = \frac{tr(Q^{-1}BD^\tau)}{tr(Q^{-1}DAD^\tau)} \, . \quad (28)$$

Equations (27) and (28) can be solved by taking an initial guess for $\bar{m}$, then iteratively solving for Q and $\bar{m}$ until the values converge.

Finally, we can make explicit the effect of assuming equal variances and no covariances for both $w(t)$ and $v(t)$. For given estimates of Q and R, since both are square, symmetric matrices, they can be factored as

$$Q = U\,DU^\tau$$

$$R = LL^\tau$$

where U is an upper triangular matrix, D is a diagonal matrix, and L is the lower triangular square root of R. To obtain an underlying population dynamic that has an uncorrelated error vector $\tilde{w}(t)$ and uncorrelated observations with variances of 1, we make the following transformations:

$$\tilde{G} = GU$$

$$\tilde{y}(t) = L^{-1}y(t)$$

$$\tilde{H}(t) = L^{-1}H(t)$$

$$\tilde{v}(t) = L^{-1}v(t)$$

and replace G, w, y, H, and v in Equations (14) through (21) with the transformed values. Then $\tilde{v}(t)$ has covariance matrix I, and $\tilde{w}(t)$ has covariance matrix D. The assumption that both the error in the dynamics and in the observations are equal, further constrains the values of D to be identical. This is a very strong assumption.

AN EXAMPLE

As an example of these methods, I use the data

for Pacific mackerel published in Parrish and MacCall (1978). I emphasize that I am only using these data for illustrative purposes and do not claim to be making a careful, thorough reexamination of the problem. Though $\bar{m}$ can be estimated using Equation (28), I assume that the value of $\bar{m}$ is known a priori. If I were to use a different value of $\bar{m}$, it would be difficult to judge to what extent the new estimates differ solely due to the different mortality rate, rather than due to the estimation scheme. I assume, as in the reference, that the mortality rate m is equal to 0.5, so that the F matrix in my notation is a matrix with a value of 0.6065 in position $(i, i - 1)$, $i = 2, \ldots, 7$, corresponding to the underlying dynamics for age groups 1 through 6.

Recruitment in Pacific mackerel is highly variable. I want to obtain estimates of recruitment that accurately reflect this variability while still being consistent with the observed data. Also, I do not want to a priori assume a functional relationship between recruitment and population size. To this end, I assume that the recruitment time series, after taking differences of a given order, is a random variable, i.e.,

$$\nabla^k r(t) = w(t) \quad (29)$$

where $w(t)$ is a normal random variable with a mean of zero and with an unknown variance σ^2 and ∇^k denotes kth order finite differencing. Akaike (1979) originally showed that this formulation is the discrete equivalent of fitting a spline to the data (in this case as a function of time), where the estimate of the variance σ^2 expresses the tradeoff between the degree of smoothness in the fitted curve with fidelity to the observed data. In this approach, k and the variance of $w(t)$ are treated as hyperparameters of the model. A fitting criterion such as AIC is then used to determine the best value of k given the data. Following Kitagawa and Gersch (1984), I could use smoothness priors to more generally decompose recruitment as

$$r(t) = T(t) + S(t) + \xi(t) + w(t) \quad (30)$$

where $T(t)$ is a trend term (as in Equation (29)), $S(t)$ is a seasonal term, and $\xi(t)$ is an irregular stationary term. A decomposition such as Equation (30) would be useful, for example, in modeling the monthly anchoveta recruitment considered in Mendelssohn and Mendo (1987). However, for convenience in this paper, I restrict recruitment to be of the form in Equation (29).

The relationship between this "smoothness priors" approach, "smoothing splines", and other penalized likelihood methods is discussed further for a variety of contexts in Brotherton and Gersch (1981), Kitagawa and Gersch (1984, 1985, 1988), Ansley and Kohn (1986), and Kohn and Ansley (1987, 1988). Wahba (1977) and O'Sullivan (1986) discussed the relationship between generalized cross-validation, penalized likelihood functions, and determining the tradeoff between smoothness and fit.

For $k = 0$, Equation (29) models recruitment as a random variable around a fixed but unknown mean value. For $k = 1$, Equation (29) models recruitment as a random walk with unknown mean level and drift (variance). Higher values of k have similar interpretations. Values of k higher than two or three rarely need to be considered, since these include the discrete equivalent of cubic splines. Cubic splines can approximate most functionals (in this case of time) to a reasonable degree of accuracy.

For this example, I assume $k = 1$, so that

$$r(t) = r(t - 1) + w(t) \qquad (31)$$

which is a random walk with unknown variance. (A more complete analysis of this data would probably also include an irregular stationary term as in Equation (30) and determine the "best" order of differencing using a given criterion.) Equation (31) can be incorporated into the state space model by letting the $(1, 1)$ element of the matrix F be equal to

1. The matrices $H(t)$ are diagonal matrices whose values are calculated from table 13 in Parrish and MacCall (1978). Because I am assuming that the estimates of F are known, then the value of Q for the M step is maximized as

$$T^{-1}[S_t(0) - S_t(1)F^\tau - FS_t(1) + FS_{t-1}(0)F^\tau].$$

$$(32)$$

As in Parrish and MacCall (1978), I treated age groups 4 through 6 as fully selected by the fishery, and will refer to these age groups as "adults". Similarly, I refer to the number of age-1 fish at the start of the season as the number of recruits. I assume that F and $H(t)$ are known, so the estimation problem is reduced to determining the means of the initial population sizes and the values of the two covariance matrices Q and R.

The resulting maximum likelihood estimates of Q and R (Tables 1, 2) show that the variances of the error terms differ by up to two orders of magnitude, hardly meeting the usual assumption of equal variances. Moreover, the covariances (expressed as correlations in the tables) are quite high, so that using $\frac{1}{\sigma}$ as a weighting factor will not be adequate. The predicted, filtered, and smoothed estimates of the adults (Fig. 1) are very similar, reflecting that the errors have been "filtered out" over time by the population dynamics.

TABLE 1.—Estimated values of the matrix Q presented as a variance-correlation matrix. The diagonal terms are the variances, and the off-diagonal terms are the cross-correlations.

	Age 0	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6
Age 0	8.72E+09	-0.9075	0.2429	0.0592	0.1785	-0.2096	0.0249
Age 1		4.59E+09	-0.1750	-0.0574	-0.1658	0.1803	-0.0432
Age 2			1.07E+09	0.0396	0.2416	0.1109	0.0675
Age 3				5.28E+08	0.7030	0.5964	0.3140
Age 4					1.90E+08	0.6717	0.3944
Age 5						3.08E+07	0.4854
Age 6							2.18E+06

TABLE 2.—Estimated values of the matrix R presented as a variance-correlation matrix.

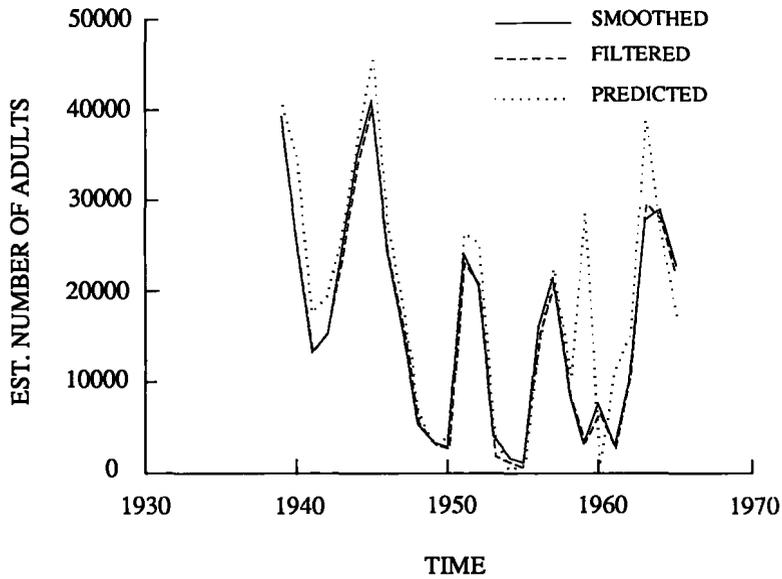|	Age 0	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6
Age 0	7.30E+07	-0.0497	0.8505	0.7367	0.6594	0.5120	-0.4019
Age 1		2.62E+07	0.1594	0.0896	-0.0544	-0.3411	0.4325
Age 2			4.95E+06	0.6764	0.6412	0.3830	-0.2418
Age 3				3.43E+06	0.7286	0.4908	-0.2260
Age 4					7.11E+05	0.5708	-0.2030
Age 5						2.05E+05	-0.1331
Age 6							5.40E+04

FIGURE 1.—Estimated number of adults, 1939–65, using values from Parrish and MacCall.

The predicted, filtered, and smoothed estimates of the recruits (Fig. 2), unlike those of the adults, are not similar. The filtered and smooth estimates are often indistinguishable, but there are some years (such as 1941 and 1943) where there are significant differences. The predicted values are very smooth, tending to emphasize the trend in the recruitment. In an analysis of recruitment estimates produced by standard cohort analysis of the anchoveta off Peru, Mendelssohn and Mendo (1987) found the estimates
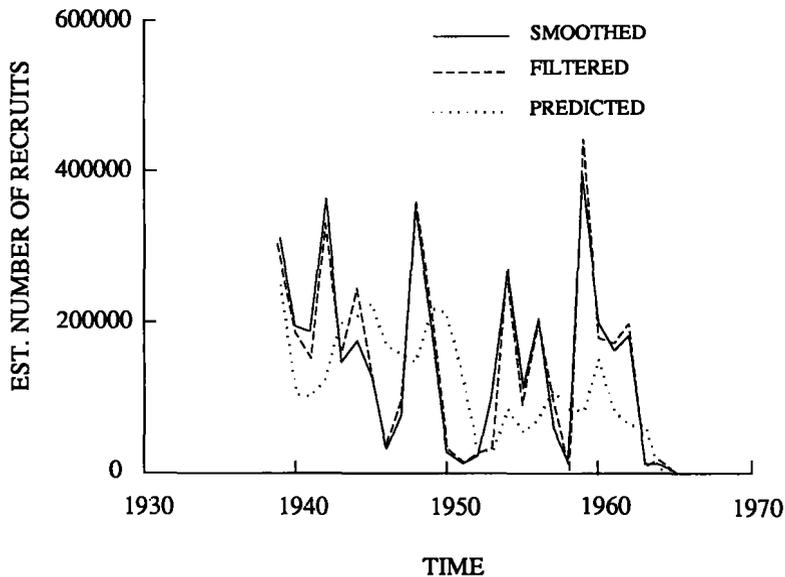
FIGURE 2.—Estimated number of recruits, 1939–65, using values from Parrish and MacCall.

to be far too smooth. The results of this present example may explain their observation. If these values were used in a subsequent analysis, say to determine the role of the environment on recruitment, a totally false picture of this relationship could emerge. I emphasize that the predicted estimates are calculated from the previous season's filtered values, whereas it is often true in the fisheries literature that the predicted values are estimated from the previous predicted values, rather than from the filtered values. Direct comparison with the estimate in Parrish and MacCall (1978) are difficult because restricting recruitment to be of the form (Equation (29)) with k fixed, rather than the more general form (Equation (30)) with k variable, may not be appropriate for the Pacific mackerel data. But overall, their estimates tend to resemble the smoother trend of my predicted estimates.

RELATIONSHIP TO OTHER LITERATURE AND A NEW PARAMETERIZATION

If the models are restricted to additive errors, then most of the simpler difference equation models proposed in Collie and Sissenwine (1983), Deriso et al. (1985), Fournier and Archibald (1982), Fournier and Doonan (1987), and among others can be formulated as I did. Some of these models assume recruitment is a nonlinear function of the underlying population, which cannot be handled in this model without some modifications (suggested below). However, all of these authors treat the underlying population sizes as parameters of the data rather than as missing data. As discussed earlier, it is questionable whether this will produce proper estimates of the underlying populations. Biases from treating missing data as parameters in a regression setting are explicitly discussed in Little and Rubin (1983, 1987).

A very broad class of possible models that can be selected to model catch-at-age data are given by Schnute (1985). He correctly identified the values that are parameters of the difference equations he discusses, and these are sufficient for estimating the likelihood (if evaluated properly). However, if we assume observation error, then it can be shown (see, for example, Shumway 1988) that the innovations are determined by predictors calculated from the previous period's filtered, rather than predicted values. Moreover, the minimum mean squared error estimates of the underlying populations, as discussed earlier, are the smoothed estimates. It appears that Schnute (1985) used the predicted or,

at best, the filtered estimates of the underlying populations.

A popular parameterization that appears to have been first suggested by Doubleday (1976) is to assume that the observation matrix $H(t)$ is of the form $H(t) = \{s(a)f(t)\}$ where $s(a)$ is an age-dependent selectivity factor and $f(t)$ is a time-dependent exploitation rate. These values can be found by using a minimization routine during the M-step of the algorithm. However, it should be noted that the estimate of $f(t)$ for each t will depend on R and that the estimate of R will depend on both $s(\cdot)$ and $f(\cdot)$, so that either R, s, and f should be solved for together, or else they should be successively solved for using Equation (7) while holding the other parameter values fixed.

Alternatively, Equation (7) can be differentiated. Then for given values of $f(t)$, the optimum value of the vector $s(a)$ at each iteration are the diagonals of the matrix S given by

$$S = AB^{-1} \qquad (33)$$

where $A = \sum_{t=1}^{T} y(t)x(t|T)^{\tau}f(t)$, and $B = \sum_{t=1}^{T} f^2(t)$

$(P(t|T) + x(t|T)x(t|T)^{\tau}$. However, this is an unconstrained estimate and does not guarantee that $s(a)$ is between (0,1). It can be shown that the optimal solution is to set $s(a)$ at zero if $s(a)$ is negative or to 1 if $s(a)$ is greater than 1.

For fixed values of $s(a)$ and R, Equation (7) is maximized when $f(t)$ takes the value

$$f(t) = \frac{tr(R^{-1}y(t)x(t|T)^{\tau}S^{\tau})}{tr(R^{-1}S(P(t|T) + x(t|T)x(t|T)^{\tau})S^{\tau})} \qquad (34)$$

where the matrix S is as above. This is the unconstrained solution. The constrained solution again is to force the estimate to lie within the closed interval $(0, 1)$ as with the estimate of the $s(a)$. Since the estimates of R, $s(a)$, and $f(t)$ are interrelated, I have found it to be a workable procedure to first estimate Q as given above and then for a given number of iterations, iteratively solve for $f(t)$ then $s(a)$. When these values stabilize, estimate R using the formula given above. While this procedure does not necessarily maximize Equation (7), it is sufficient for the generalized EM algorithm the new values increase the function given in Equation (7).

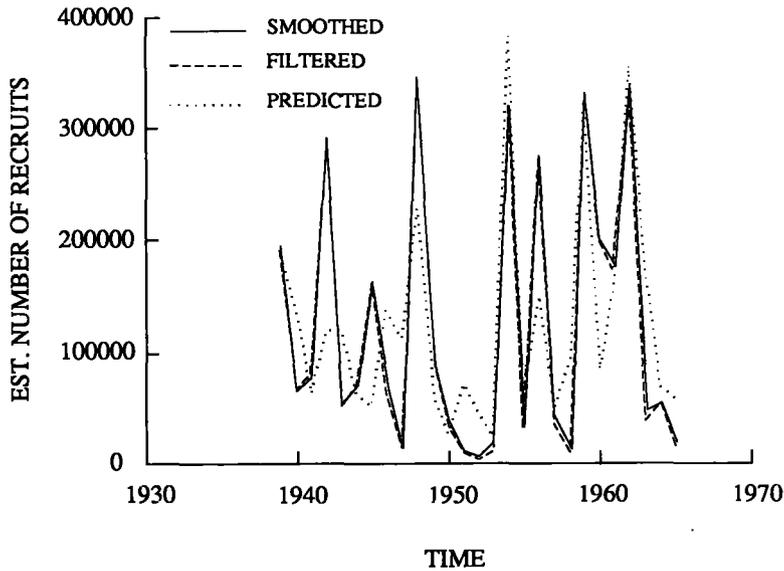As with the original estimates, the smoothed and filtered estimates of recruitment (Fig. 3) are close

625

FIGURE 3.—Estimated number of recruits 1939–65, assuming $q(a, t) = f(t)s(a)$.

to each other, while the predicted values are too smooth. The estimated values of $f(t)$ (Fig. 4) generally decline with time. All three estimates are significantly different from the previous estimates. Years where relative highs and lows occur differ, showing the sensitivity of this class of models to the assumed form of the H matrix. Since the estimates of R, $s(a)$ and $f(t)$ are interrelated, I suspect that the parameter estimates are highly correlated and hence unstable. While I have not calculated the parameter covariance matrix, it should be checked for any serious analysis using these techniques.
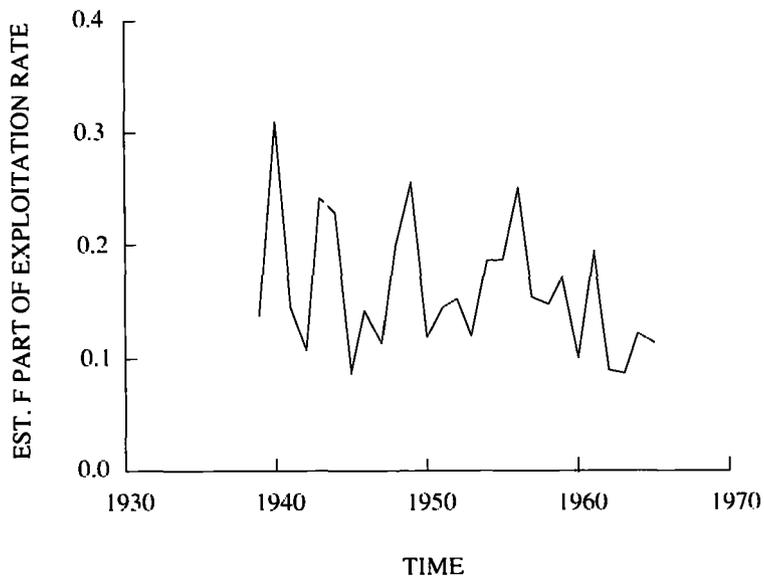

FIGURE 4.—Estimated values of $f(t)$ when $q(a, t) = f(t)s(a)$.

While the parameterization $H(t) = \{f(t)s(a)\}$ greatly reduces the number of parameters, I still face the same problem as when I treated the unobserved population as parameters: each new period adds another parameter to the model. The model still appears to be overparameterized, and the asymptotic theory for maximum likelihood estimation may be invalid.

As with the recruitment estimates, the number of effective parameters can be reduced by adding a smoothness prior on $f(t)$. However, $f(t)$ is a proportion and not likely to be a normal variable. The $f(t)$ are also constrained to lie in the interval $(0, 1)$, so a transformation to an unconstrained variable would also be desirable. If $f(t)$ is a binomial random variable then arcsin $(\sqrt{f(t)})$ is approximately normal with nearly equal variance. This suggests the transformation of variables

$$f(t) = \sin^2(e(t)) \qquad (35)$$

where $e(t)$ is an unconstrained normal variable. Then the smoothness prior becomes

$$\nabla^k e(t) = w(t) \qquad (36)$$

(a smoothness prior that includes a seasonal component or irregular stationary part, as in Equation (30), can also be used), and for any age class, the observation equation becomes

$$y(a, t) = s(a)\sin^2(e(t))x(a, t). \qquad (37)$$

The underlying population dynamics must also be expanded to include the smoothness prior constraint (Equation (36)).

Unfortunately, the observation equation is no longer linear in the state vector. The smoothness prior is a prior distribution on the $f(t)$, and a full Bayesian analysis can be done to obtain the overall distribution. The variance of $w(t)$ is then treated as a hyperparameter in the analysis.

A simpler approach is to evaluate the filter equations approximately by using any one of a number of nonlinear filters (see Anderson and Moore 1979). One that is easy to implement, given the nonlinearities in this problem, is the extended Kalman filter (EKF), which at each time period just linearizes all the nonlinear terms around the value of the predicted state vector. The EKF, however, can have divergence problems and is not guaranteed to find the true penalized likelihood estimates.

When using the EKF, it works to make a forward and backward pass of the filter given the current estimates of $f(t)$ and $x(t|T)$, and then to estimate $s(a)$ and R as before. I tested the algorithm on the mackerel data with $k = 1$ in the constraint (Equation (36)). The resulting estimates (Fig. 5) are similar to the previous estimates, but the estimated values of $f(t)$ (Fig. 6) are less variable with a stronger trend than before. It is clear from Figure 5 that the
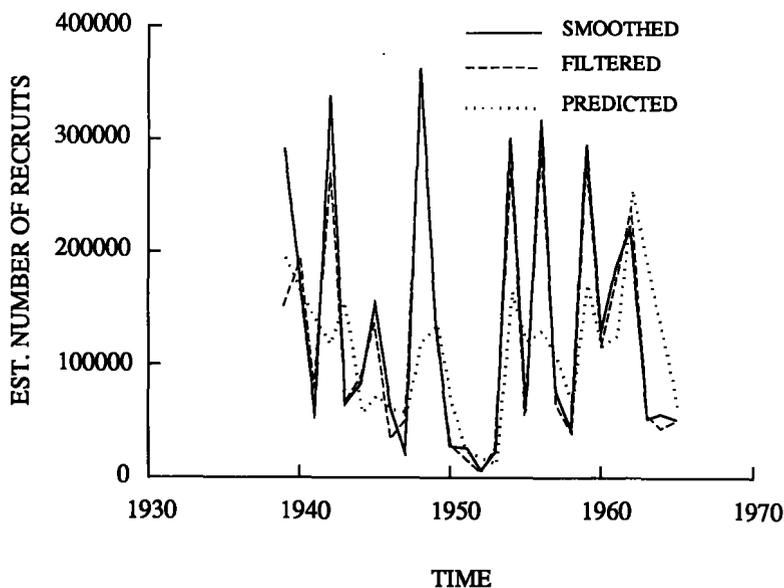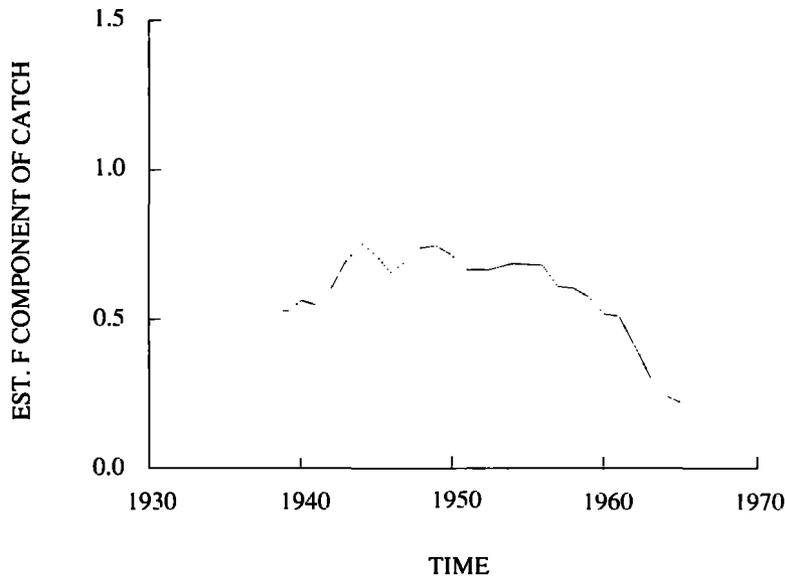

FIGURE 5.—Estimated number of recruits using a first order spline for estimating $f(t)$.

FIGURE 6.—Spline estimates of $f(t)$.

smoothness prior estimate has shifted some of the variation in the observed data to variations in the underlying population dynamics rather than variations in $f(t)$. Further research needs to be done to see which of these estimates is the most "valid".

The EM algorithm in general can be sensitive to the initial values given the parameters (see Wu 1983), and I have found that the fixed value of Σ can also affect the estimates. Initial combinations of μ, $f(t)$, and $s(a)$ that are totally inconsistent with the observed catch data can cause the algorithm to find a local maximum. This can be avoided by experimenting with several, very different starting values and determining if they converge to the same estimates.

If recruitment is thought to be a linear function of the previous population size, then there is no problem including this in the Kalman filter. If recruitment is a nonlinear function of the previous population, then the EKF can again be used to approximately determine the conditional expectations needed for the EM algorithm.

If information is available from a variety of sources, say from fishing and from surveys, as in Methot (fn. 2), then each of the vectors and matrices can be partitioned to represent this situation. For example, let $y_s(t)$ be the observed catches from a survey and $y_f(t)$ the observed catches from fishing. Let $y(t)^r = (y_s(t), y_f(t))^r$, and partition the H matrix similarly. Then the diagonal blocks of H will con-

tain the observation dynamics for the survey and for fishing, while the off-diagonal blocks will be zero. Given these modifications, all the algorithms described previously in this paper can be used to derive estimates for this situation.

Research or other surveys of the fishery usually occur less frequently than does commercial fishing, causing part of the vector $y(t)$ to be missing at given times periods. Shumway and Stoffer (1982a) and Shumway (1988) gave a straightforward modification of the Kalman filter for this case.

DISCUSSION

I have introduced a new method for estimating population sizes from catch-at-age data that includes, if additive errors can be assumed, many of the previous difference equation models. I show that it is incorrect to treat the unobserved population sizes as parameters to be estimated rather than as missing data. I also show that the minimum mean square estimates of the population sizes are the smoothed estimates rather than the predicted estimates suggested in many papers. The model assumes neither equal variances in the errors in the population dynamics nor in the observation errors and does not require that the errors be uncorrelated. For Pacific mackerel, the smoothed estimates are shown to be much more variable than the predicted estimates.

I also suggest a new parameterization for the age-specific exploitation rate $q(a, t)$. If it is assumed that $q(a, t) = s(a) f(t)$, then the model is over determined. I put a smoothness prior on $f(t)$ in order to obtain a tradeoff between the degree of smoothness in $f(t)$ as a function of time versus fidelity to the data. The degree of differencing can be treated as a hyperparameter of the model to determine the optimal amount of differencing given the data.

An advantage of the approach of this paper is that the calculations are straightforward and simple to program, and explicit formulas are given for the optimal parameter values at each iteration of the EM algorithm. Additional optimization software is not required to perform the calculations. Moreover, properties of the Kalman filter and the EM algorithm are well known. There is a large literature giving variants of the filter to calculate sensitivity to model misspecification, and recursive formulas for the derivatives with respect to a given parameter of the model also exist in the literature.

It is also simple to include environmental variables into the formulation, either as additional state variables or as fixed effects in the observation equation or both (see Sallas and Harville [1988] on how to estimate the fixed effect parameters within the context of Kalman filtering). Thus the influence of the environment can be modeled directly, rather than resorting to the conventional practice of obtaining population estimates first and correlating these estimates with the environmental variables second.

A disadvantage of my approach is that there is no guarantee that any of the estimates of the underlying population sizes will be positive. The population sizes are treated as normal random variables, and it is quite possible for the additive corrections in the filtered or smoothed estimates to make small population sizes negative if the observation error is large. P. Sullivan (fn. 3) has found that for a length-based model the Kalman filtering approach works best when there are pulses in the recruitment, that is, when the population is not in equilibrium. The likelihood surface is such that without recruitment pulses it is difficult to estimate the parameters of the growth-curve. Most fisheries are not in equilibrium, however. As the models in this paper do not contain a growth-curve, it is unclear if a similar finding will be valid.

Some of my results suggest that the estimates are sensitive to the form of the model chosen for the population dynamics. This is not surprising, because, unlike most missing data problems, the missing part of the data is never observed directly, but only

through the presumed form of the dynamics. For example, when modeling catch (or catch per unit effort) against an environmental variable, catch data often are not available for all periods. But there are at least some periods when both variables are observed, which can be used to estimate the relationship between the two sets of variables. This relationship is used to produce the smoothed estimates of the missing data. For estimating population sizes from catch-at-age data, the a priori estimate of the form of the observation equation replaces this empirically derived estimate.

In many of the references cited, multiplicative errors are preferred in the observation equation because variances appear to change with the size of the population. My experience is that relaxing the assumption of equal, uncorrelated errors appears to at least partially take into account the observed differences. If the model estimates are not satisfactory, assuming additive, gaussian errors, then the regular EM algorithm can be used to properly estimate the smoothed estimates of the underlying population. However, the EM algorithm requires the complete data likelihood as well as the expectation of the log-likelihood with respect to $(y(1), \ldots, y(T))$. In multiplicative models, assumptions about the error structure can lead to very complicated multivariate distributions for the complete data due to the Jacobian of the transformation. The conditional expectation of the log-likelihood may have to be evaluated by numerically integrating a nontrivial multiple integral. Certainly, as a first pass, the simpler techniques of this paper would appear to have a lot to offer as an alternative.

LITERATURE CITED

AKAIKE, H.
  1979. Likelihood and the Bayes procedure. *In* J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (editors), Bayesian statistics, p. 143–166. Univ. Press, Valencia.
ANDERSON, B. D. O., AND J. B. MOORE.
  1979. Optimal filtering. Prentice Hall, Englewood Cliffs, NJ, 357 p.
ANSLEY, C. F., AND R. KOHN.
  1986. On the equivalence of two stochastic approaches to spline smoothing. *In* J. Gani and M. B. Priestly (editors), Time series and allied processes, p. 391–405. J. Appl. Prob. 23A.
BRILLINGER, D. R., J. GUCKENHEIMER, P. GUTTORP, AND G. OSTER.
  1979. Empirical modeling of population time series data: The case of age and density dependent vital rates. *In* Some mathematical questions in biology, p. 65–90. Lectures on Mathematics in the Life Sciences 13, American Mathematical Society, Providence.

BROTHERTON, T., AND W. GERSCH.
    1981. A data analytic approach to the smoothing problem and some of its variations. Proc. 20th IEEE Conf. Decis. Control, p. 1061–1069.
COLLIE, J. S., AND M. P. SISSENWINE.
    1983. Estimating population size from relative abundance data measured with error. Can. J. Fish. Aquat. Sci. 40: 1871–1879.
DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN.
    1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. 39(Ser. B):1–38.
DERISO, R. B., T. J. QUINN II, AND P. R. NEAL.
    1985. Catch-age analysis with auxillary information. Can. J. Fish. Aquat. Sci. 42:815–824.
DOUBLEDAY, W. G.
    1976. A least squares approach to analyzing catch at age data. Res. Bull. Int. Comm. Northwest Atl. Fish. 12:69–81.
FOURNIER, D. A., AND C. ARCHIBALD.
    1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39:1195–1207.
FOURNIER, D. A., AND I. J. DOONAN.
    1987. A length-based stock assessment method utilizing a generalized delay-difference model. Can. J. Fish. Aquat. Sci. 44:422–437.
JAZWINSKI, A. H.
    1970. Stochastic processes and filtering theory. Acad. Press, N.Y., 376 p.
KITAGAWA, G., AND W. GERSCH.
    1984. A smoothness priors - state space modeling of time series with trend and seasonality. J. Am. Stat. Assoc. 79: 378–389.
    1985. A smoothness priors time varying AR coefficient modeling of nonstationary time series. IEEE Trans. Autom. Control AC-30:48–56.
    1988. Smoothness priors in time series. In J. Spall (editor), Bayesian analysis of time series and dynamic models, p. 431–476. Marcel Dekker, N.Y.
KOHN, R., AND C. F. ANSLEY.
    1987. A new algorithm for spline smoothing and interpolation based on smoothing a stochastic process. SIAM J. Sci. Stat. Comput. 8:33–48.
    1988. The equivalence between Bayesian smoothness priors and optimal smoothing for function estimation. In J. Spall (editor), Bayesian analysis of time series and dynamic models, p. 393–430. Marcel Dekker, N.Y.
LITTLE, R. J. A., AND D. B. RUBIN.
    1983. On jointly estimating parameters and missing data by maximizing the complete data likelihood. Am. Stat. 37: 218–220.

    1987. Statistical analysis with missing data. Wiley, N.Y., 278 p.
LJUNG, L., AND T. SODERSTROM.
    1983. Theory and practice of recursive identification. M.I.T. Press, Cambridge, MA, 529 p.
MENDELSSOHN, R., AND J. MENDO.
    1987. Exploratory analysis of anchoveta recruitment off Peru and related environmental series. In D. Pauly and I. Tsukayama (editors), The Peruvian anchoveta and its upwelling ecosystem: three decades of change, p. 294–306. ICLARM, Manila.
O'SULLIVAN, F.
    1986. A statistical perspective on ill-posed problems. Stat. Sci. 4:502–527.
PARRISH, R. H., AND A. MacCALL.
    1978. Climatic variation and exploitation in the Pacific mackerel fishery. Calif. Dep. Fish Game, Fish Bull. 167, 110 p.
SALLAS, W. M., AND D. A. HARVILLE.
    1988. Noninformative priors and restricted maximum likelihood estimation in the Kalman filter. In J. Spall (editor), Bayesian analysis of time series and dynamic models, p. 477–508. Marcel Dekker, N.Y.
SCHNUTE, J.
    1985. A general theory for analysis of catch and effort data. Can. J. Fish. Aquat. Sci. 42:414–429.
SHIRYAYEV, A. N.
    1984. Probability. Springer-Verlag, N.Y., 577 p.
SHUMWAY, R. H.
    1988. Applied statistical time series analysis. Prentice-Hall, Englewood Cliffs, NJ, 379 p.
SHUMWAY, R. H., AND D. S. STOFFER.
    1982a. An approach to time series smoothing forecasting using the EM algorithm. J. Time Ser. Anal. 3:253–264.
    1982b. An algorithm for parameter estimation and smoothing in space-time models with missing data. U.S. Dep. Commer., NOAA, Final Rep. P.O. 82-ABA-1190.
WAHBA, G.
    1977. A survey of some smoothing problems and the method of generalized cross-validation for solving them. In R. Krishnaiah (editor), Applications of statistics, p. 507–524. North Holland Publishing Co., Amsterdam.
WU, C. F.
    1983. On the convergence properties of the EM algorithm. Ann. Stat. 11:95–103.
ZEHNWIRTH, B.
    1988. A generalization of the Kalman filter for models with state-dependent observation variance. J. Am. Stat. Assoc. 83:164–167.