A Stochastic Implementation of an Age-Structured Production Model

Victor R. Restrepo and Christopher M. Legault
University of Miami, Rosenstiel School of Marine and Atmospheric Science, Miami, Florida

Abstract
An age-structured production model (ASPM) has been used by the International Commission for the Conservation of Atlantic Tunas for the assessments of some stocks. That model is sometimes preferable to traditional biomass-based surplus production models because it can accommodate age-structured indices of relative abundance. The traditional ASPM estimates a deterministic stock-recruitment relationship, a property that may result in inconsistencies between the estimated level of recruitment and the observed level of catches. In this work we relax the deterministic assumption by incorporating stochasticity in recruitment around the deterministic predictions as a first-order, autoregressive time-series process. We use data for western Atlantic bluefin tuna to contrast the deterministic and stochastic model fits.

Introduction
Age-structured production models (ASPM) have been used in assessments carried out by the International Commission for the Conservation of Atlantic Tunas (ICCAT) in the past, particularly for albacore tuna (Thunnus alalunga) in the South Atlantic and for bluefin tuna (Thunnus thynnus) tuna in the Western Atlantic. Conceptually, ASPMs fall somewhere between simple biomass-based production models (e.g., Schaefer 1957, Prager 1994) and the more data-demanding sequential age-structured population analyses (Megrey 1989). Simple production models estimate parameters related to carrying capacity, rate of productivity, biomass at the start of the time series, and coefficients that scale indices of abundance to the absolute magnitude of biomass. ASPMs estimate similar parameters but make...
explicit use of age-structured computations, rather than lumped-biomass ones, and directly estimate parameters of a stock-recruitment relationship. Their main advantage over simpler production models is that they can make use of age-specific indices of relative abundance.

During the 1996 assessments of Atlantic bluefin tuna (ICCAT 1997), an implementation of an ASPM (Restrepo 1997) was criticized because its estimates of recruitment in recent years could not be reconciled against available catch estimates for extant cohorts. This could be in part attributable to the fact that the ASPM implementation estimated a deterministic stock-recruitment relationship and, therefore, estimated levels of recruitment would vary smoothly over time at the deterministic predictions. The main objective of this paper is to extend the ASPM formulation in order to account for stochastic recruitment. The same approach to modeling recruitment can easily be incorporated into the more general “integrated approaches” (e.g., Fournier and Archibald 1982, Deriso et al. 1985, Methot 1990) or into tuned sequential population analyses as was done by Porch (in press).

Model Formulations and Estimation

Throughout this paper, a Beverton and Holt (1957) type of stock recruitment relationship (SRR) is assumed.

Deterministic Formulation

The deterministic model is similar to that of Punt (1994), which was based on ideas presented by Hilborn (1990), with modifications to account for multiple fisheries. Details of the fitting procedure can be found in Punt (1994) and only a brief summary is presented here, along with some concepts needed for the presentation.

The resource dynamics are modeled by a forward population projection using standard fishery equations. The projection includes a “plus” group (ages $p$ and older) and each year’s recruitment is obtained from a deterministic stock-recruitment relationship. The fishing mortality values needed to project the population forward are computed based on total yields and selectivities that are input and assumed exact. The Restrepo (1997) formulation allows for gear-specific yields and selectivities because, in multi-gear cases, it is easier to obtain annual input selectivities by gear, rather than for all gears combined. The estimation consists of finding the values of the stock-recruitment relationship parameters that result in stock size trajectories which best explain observed indices of relative abundance by minimizing the negative log-likelihood

$$-\ln(L) = \sum_{t} \left[ \frac{n_t}{2} \sum \ln(\sigma_{it}^2) + \sum \frac{1}{2\sigma_{it}^2} (I_{it} - \hat{I}_{it})^2 \right]$$

(1)
where \( t \) denotes year, \( i \) denotes each available series of relative abundance comprised of \( n_i \) observations, the last term is for the squared differences between observed \((\hat{I})\) and predicted \((\hat{I})\) indices of abundance (these could be in logarithmic units if a lognormal error is assumed), and \( \sigma_i^2 \) are abundance index variances that are either input or estimated. Additional parameters include catchability coefficients that scale the indices to absolute abundance (or biomass), and possibly a parameter related the initial conditions. In many ASPM applications, the initial conditions are not reliably estimated and the initial age structure is fixed, e.g. by assuming that the stock was in a virgin state at the start of the time series if the series extends back to the onset of fishing.

The Beverton and Holt SRR is usually described by the equation

\[
R_{i+1} = \frac{\alpha S_i}{\beta + S_i},
\]

where \( R \) is the number of recruits and \( S \) is the product of numbers, maturity and fecundity, summed over all ages. For simplicity, we refer to \( S \) as “spawning biomass,” which is often used as a proxy for reproductive output. During estimation, the ASPM uses a different parameterization, following Francis (1992). It consists of defining a “steepness” parameter, \( \tau \), which represents the fraction of the virgin recruitment \((R_0)\) that is expected when \( S \) has been reduced to 20% of its virgin level: \( R = \tau R_0 \) when \( S = \gamma/5 \), where \( \gamma \) is the virgin spawning biomass. The SRR is thus defined in terms of steepness and virgin biomass, two parameters for which initial values are somewhat easier to guess than \( \alpha \) and \( \beta \). (For a Beverton-Holt relationship, virgin biomass should generally be of similar magnitude to the largest observed yields, while steepness should fall somewhere between 0.2 and 1.0). \( R_0 \) is computed as the ratio of virgin spawning biomass to spawning biomass per recruit in the absence of fishing, \((S/R)_{F=0}\),

\[
R_0 = \frac{\gamma}{(S/R)_{F=0}},
\]

and \( \alpha \) and \( \beta \) are given by

\[
\alpha = \frac{4\tau R_0}{5\tau - 1}, \quad \text{and}
\]

\[
\beta = \frac{\gamma(1 - \tau)}{5\tau - 1}
\]

The computation of statistics such as maximum sustainable yield \((MSY)\) and related benchmarks (e.g., \( S_{MSY} \), \( F_{MSY} \)) follows the procedure summarized
by Shepherd (1982). Conditional on a given \( F \) (including an overall selectivity pattern), equilibrium spawning biomass, recruitment and yield are computed as (for the Beverton and Holt SRR),

\[
S_F = \alpha (S/R)_F - \beta, \tag{6a}
\]

\[
R_F = \frac{S_F}{(S/R)_F}, \quad \text{and} \quad \tag{6b}
\]

\[
Y_F = R_F (Y/R)_F, \tag{6c}
\]

where \((S/R)_F\) and \((Y/R)_F\) are the spawning biomass per recruit and yield per recruit values resulting from exploitation at \( F \). \( F_{MSY} \) would be computed by searching for the \( F \) that maximizes equation (6c). Note that, if the selectivity pattern changes over time, then the computed \( MSY \)-related values will also change as a result of changes in the per-recruit computations.

A set of useful benchmarks for management is based on the so-called “spawning potential ratio,” \( SPR \), defined as the spawning biomass per recruit obtained under a given \( F \), divided by that under \( F = 0 \) (Goodyear 1993). An important benchmark is the \( SPR \) corresponding to the slope of the SRR at the origin, i.e., at the point when the stock is expected to “crash.” From equations (3) to (5) it follows that this \( SPR_{crash} \) is a function of steepness:

\[
SPR_{crash} = \frac{(S/R)_{crash}}{(S/R)_{F=0}} = \frac{\beta / \alpha}{\gamma / R_0} = \frac{1 - \tau}{4 \tau}. \tag{7}
\]

Deterministically, any fishing mortality that results in an \( SPR \) lower than \( SPR_{crash} \) is not sustainable.

**Stochastic Formulation**

A stochastic ASPM requires that a recruitment value be estimated for every year. In this work, we have chosen to constrain the (log) recruitment deviations from the equilibrium SRR to follow a first-order autoregressive (AR[1]) process. The population projection equations are as in the deterministic model, except that recruitment is estimated as

\[
N_{a,t} = R_0 e^{\nu_t}, \tag{8}
\]

where the notation \( N_{a,t} \) denotes population numbers for age \( a \) in year \( t \). Thus, recruitment is estimated as deviations from a virgin level. Instead of estimating \( \gamma \) and \( \tau \) directly as parameters, the model estimates \( \gamma \) and all the \( \nu_t \). \( R_0 \) is computed from equation (3). Besides the initial population
size, these are all the parameters that are needed to project the population forward. The AR[1] process is incorporated by assuming that the recruitment estimates thus obtained vary around the expected SRR as

\[ N_{t+1} = R_{t+1}e^{\varepsilon_{t+1}} = \frac{\alpha S_t}{\beta + S_t}e^{\varepsilon_{t+1}} \quad (9) \]

with \( \varepsilon_{t+1} = \rho \varepsilon_t + \eta_{t+1} \), where \(|\rho| < 1\), and the \( \eta \) have zero expectation and variance equal to \( \sigma^2_\eta \). In equation (9) we distinguish between recruitment values estimated as parameters in the search \( (N_{t+1}) \) and those predicted from the estimated stock-recruitment relationship \( (R_t) \).

The negative log-likelihood for these recruitment “residuals” is (Seber and Wild 1989):

\[
-\ln(L) = \frac{n}{2} \ln(\sigma^2_\eta) - \frac{1}{2}\ln(1 - \rho^2) + \frac{1}{2\sigma^2_\eta} \left(1 - \rho^2\right)e^2_t + \sum_{t=2}^{n_t} \left(\varepsilon_t - \rho \varepsilon_{t-1}\right)^2 \quad (10)
\]

where \( n_t \) is the number of years in the analysis. It follows from equation (9) that the residuals are computed as

\[
\varepsilon_{t+1} = \ln\left(N_{t+1}\right) - \ln\left(\frac{\alpha S_t}{\beta + S_t}\right) \quad (11)
\]

Computation of the first residual would depend on the initial conditions. For example, in a virgin state, it would be \( \varepsilon_1 = \ln(N_1) - \ln(R_0) \). Note that \( \alpha \) and \( \beta \) in equations (9) and (11) could be computed from knowledge of virgin biomass and steepness (see equations 4 and 5). However, only the former is being estimated directly as a parameter. To include steepness as an additional parameter to be directly estimated by the search would confound the information contained in \( R_0 \) and \( \gamma \). Our approach is to replace \( \alpha \) and \( \beta \) in the SRR of equation (11) by a function of those parameters being estimated in the search, and steepness. From equations (4) and (5) it follows that deterministic recruitment can also be predicted by

\[
R_{t+1} = \frac{4R_0 S_t \tau}{\tau(5S_t - \gamma) - S_t + \gamma}, \quad \text{such that} \quad (12)
\]

\[
\varepsilon_{t+1} = \ln\left(N_{t+1}\right) - \ln\left(\frac{4R_0 S_t \tau}{\tau(5S_t - \gamma) - S_t + \gamma}\right). \quad (13)
\]

We use these relationships in order to solve for \( \tau \), noting that, for a given \( \rho \) and \( \sigma^2_\eta \), equation (10) will be at a minimum when
is also at a minimum. Thus, in every iteration in the search, a subprocedure is invoked to find the $\tau$ that minimizes (14). Having thus calculated the steepness (and, consequently, $\alpha$ and $\beta$), the log-likelihood of equation (10) is added to the overall objective function.

It remains to be mentioned what to do about the parameters $\rho$ and $\sigma_\eta^2$. In theory, there is a potential for these to also be estimated. In practice, however, it is unlikely that the data will contain so much information as to determine the relative contribution from recruitment variability with respect to the variability in the index values. In our limited experience with this model, it appears that these values should be controlled by the analyst in much the same way as contributions to the likelihood from different data sources are weighted externally in other assessment methods (e.g., Deriso et al. 1985). Lower $\sigma_\eta^2$ values will result in lower stochasticity in recruitment, while higher $\sigma_\eta^2$ values will allow recruitment to fluctuate more widely. A value of $\rho = 0$ would assume no autocorrelation between successive recruitment deviations. Empirical studies such as those of Beddington and Cooke (1983) and Myers et al. (1990) may yield information about likely ranges of values for $\rho$ and $\sigma_\eta^2$ for species groups.

Estimating the initial conditions for the stochastic model can be problematic, as with the deterministic model. Estimating the age structure in year $t = 1$ would not generally be an option unless there were age-specific relative abundance data for the start of the series. Thus, using a long time series of data extending to the onset of fishing, and assuming an initial equilibrium state at $S = \gamma$, remains a useful option. In this paper we calculate a stable age structure resulting from a pre-series recruitment that is fixed. We fix $v_0$ and set the starting population sizes as

$$N_{2,1} = R_e e^{v_0} e^{-M_i},$$  \hspace{1cm} (15a)

$$N_{a,1} = N_{a-1,1} e^{-M_{a-1}} \text{ for ages } a = 3 \text{ to } p-1, \text{ and}$$  \hspace{1cm} (15b)

$$N_{p,1} = \frac{N_{p-1,1} e^{-M_{p-1}}}{1 - e^{-M_p}} \text{ for the plus group, } p.$$  \hspace{1cm} (15c)

This alternative allows the initial age structure to be either higher or lower than that corresponding to an equilibrium virgin state. The parameter $v_0$ could be estimated in the search procedure as well. If it is, it may be desirable to place a penalty on how much it can alter the initial biomass away from $\gamma$. This could be accomplished with the term
\[-\ln(L_{\gamma}) = \frac{\ln(\sigma_v^2)}{2} + \frac{[\ln(S_\gamma) - \ln(\gamma)]^2}{2\sigma_v^2}\] (16)

where $\sigma_v^2$ is fixed by the analyst.

Estimation of the stochastic model parameters for any given data set then requires several choices associated with how much recruitment can fluctuate around its deterministic predictions and about the initial conditions. In addition to choices about variances ($\sigma_\eta^2$, $\sigma_v^2$ and possibly $\sigma_{it}^2$), the log-likelihood components could be given different emphases ($\lambda_i$) to obtain model estimates by minimizing:

\[-\ln(L_{\gamma}) = -\ln(L_\gamma) - \lambda_2 \ln(L_2) - \lambda_3 \ln(L_3).\] (17)

Projections using the stochastic model results incorporate the auto-regressive process in recruitment by first generating normally distributed random deviates, $\eta_t$, with mean zero and variance $\sigma_\eta^2$, and then setting the autocorrelated residuals as

$$\varepsilon_{t+1} = \rho \varepsilon_t + \eta_{t+1}$$ (18)

for use in equation (9). The value of the first residual to “seed” the projections is available from the model fit, i.e. from equation (11).

**An Example: Western Atlantic Bluefin Tuna**

We used the stochastic ASPM model with the same data set that was used for western Atlantic bluefin tuna during the 1996 ICCAT assessment (ICCAT 1997). The input data consists of yields for four fisheries starting in 1950 (Fig. 1), and of nine indices of abundance assumed to be lognormally distributed, with associated coefficients of variation (available in ICCAT 1997). Fishery selectivities were fixed by the assessment working group for different time periods when regulations substantially affected fishing operations and were based on average patterns in the results from an age-structured VPA (Gavaris 1988, Powers and Restrepo 1992). The biological parameters used in the ASPM runs are given in Table 1.

We carried out a series of estimations making different assumptions about some of the inputs that are fixed (Table 2), or estimating different numbers of parameters, in order to examine the sensitivity of the results to these choices. Run 1 is a deterministic one, conducted for comparison to the stochastic model. Runs 2 to 5 examine the effect of estimating different numbers of recruitment deviations from $R_0$ (equation 8). In these, the $\nu_i$ were fixed at values obtained from Run 1 for years when they were not being estimated, thus forcing the initial pattern of recruitment to be smooth, like in the deterministic fits. Runs 6 and 7 allowed the initial
biomass to be arbitrarily smaller or larger than the virgin one. Run 8 was performed to examine the effect of assuming no autocorrelation in recruitment residuals ($\rho = 0$), and Runs 9 and 10 used different choices for the assumed variance in recruitment residuals. In all runs, we set $\sigma^2$ equal to 0.1, and set $\lambda_2$ and $\lambda_3$ equal to 1.

For Runs 1 and 2, we approximated the variance of outputs of interest by bootstrapping (Efron 1982) the indices of abundance from their residual distributions. The algorithm used for estimation was that of Nelder and Mead (1965), with multiple restarts.

Results and Discussion
A summary of outputs for Runs 1-10 is given in Table 3, where we focus on some quantities that are of interest to management. Some general conclusions can be derived from these results:

1. Maximum sustainable yield (MSY) resulting from recent (1990-1995) fishery selectivity patterns is estimated at 6,000-7,000 t, or about 2.5 times recent yield levels. The potential MSY has changed considerably over time as a result of changes in the overall selectivity effected by the various gears (Fig. 2). (Note that long-term changes in MSY for a stock could be due to a number of other factors, including changes in survival or productivity. In this case, due to the model’s assumptions and use of a constant natural mortality rate, the estimated shifts in MSY levels can only be attributed to changes in selectivity). Recent MSY levels are higher than those attainable with the 1960s-1970s selectivities, probably as a result of ICCAT minimum size recommendations and subsequent demise of purse seine harvest of small bluefin tuna (Fig. 1).
Table 1. Biological parameters assumed for bluefin tuna, held constant for all years.

<table>
<thead>
<tr>
<th>Age</th>
<th>Midyear weight (kg)</th>
<th>Spawning biomass (kg)</th>
<th>Natural mortality (yr⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.4</td>
<td>0.0</td>
<td>0.14</td>
</tr>
<tr>
<td>2</td>
<td>10.6</td>
<td>0.0</td>
<td>0.14</td>
</tr>
<tr>
<td>3</td>
<td>20.5</td>
<td>0.0</td>
<td>0.14</td>
</tr>
<tr>
<td>4</td>
<td>37.5</td>
<td>0.0</td>
<td>0.14</td>
</tr>
<tr>
<td>5</td>
<td>56.6</td>
<td>0.0</td>
<td>0.14</td>
</tr>
<tr>
<td>6</td>
<td>80.6</td>
<td>0.0</td>
<td>0.14</td>
</tr>
<tr>
<td>7</td>
<td>111.3</td>
<td>0.0</td>
<td>0.14</td>
</tr>
<tr>
<td>8</td>
<td>143.0</td>
<td>131.3</td>
<td>0.14</td>
</tr>
<tr>
<td>9</td>
<td>176.1</td>
<td>164.8</td>
<td>0.14</td>
</tr>
<tr>
<td>10+</td>
<td>292.8</td>
<td>267.2</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Table 2. Age-structured production model runs performed.

<table>
<thead>
<tr>
<th>Run</th>
<th>Type</th>
<th>( v_t ) estimated</th>
<th>( v_{t=0} )</th>
<th>( \rho )</th>
<th>( \sigma^2 \eta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Deterministic</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>2</td>
<td>Stochastic</td>
<td>1950-1995</td>
<td>0</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>Stochastic</td>
<td>1960-1995</td>
<td>0</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>Stochastic</td>
<td>1970-1995</td>
<td>0</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>5</td>
<td>Stochastic</td>
<td>1980-1995</td>
<td>0</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>6</td>
<td>Stochastic</td>
<td>1950-1995</td>
<td>-0.1</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>7</td>
<td>Stochastic</td>
<td>1950-1995</td>
<td>0.1</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>8</td>
<td>Stochastic</td>
<td>1950-1995</td>
<td>0</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>9</td>
<td>Stochastic</td>
<td>1950-1995</td>
<td>0</td>
<td>0.5</td>
<td>0.05</td>
</tr>
<tr>
<td>10</td>
<td>Stochastic</td>
<td>1950-1995</td>
<td>0</td>
<td>0.5</td>
<td>0.4</td>
</tr>
</tbody>
</table>

See text for description of symbols.
Table 3. Results from the age-structured production model runs (defined in Table 2). $R^2$ is the coefficient of determination for the nine fitted indices of abundance.

<table>
<thead>
<tr>
<th>Run</th>
<th>$R^2$</th>
<th>MSY (t)</th>
<th>$S_{96}/\gamma$</th>
<th>$S_{96}/S_{MSY}$</th>
<th>$R_{92-95}/R_0$</th>
<th>$F_{8+,92-95}$</th>
<th>$SPR_{92-95}$</th>
<th>$SPR_{crash}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.23</td>
<td>6,079 (3.2)</td>
<td>0.028 (14.7)</td>
<td>0.083 (13.9)</td>
<td>0.21 (7.6)</td>
<td>0.35 (12.1)</td>
<td>0.21</td>
<td>0.35</td>
</tr>
<tr>
<td>2</td>
<td>0.28</td>
<td>6,502 (8.2)</td>
<td>0.035 (23.5)</td>
<td>0.107 (21.3)</td>
<td>0.16 (22.7)</td>
<td>0.37 (19.3)</td>
<td>1.30</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.27</td>
<td>7,290</td>
<td>0.029</td>
<td>0.091</td>
<td>0.170</td>
<td>0.46</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.26</td>
<td>6,665</td>
<td>0.033</td>
<td>0.101</td>
<td>0.181</td>
<td>0.44</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.25</td>
<td>6,369</td>
<td>0.036</td>
<td>0.105</td>
<td>0.150</td>
<td>0.35</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.28</td>
<td>6,776</td>
<td>0.033</td>
<td>0.099</td>
<td>0.156</td>
<td>0.38</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.28</td>
<td>6,490</td>
<td>0.036</td>
<td>0.105</td>
<td>0.152</td>
<td>0.35</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.27</td>
<td>7,357</td>
<td>0.028</td>
<td>0.087</td>
<td>0.198</td>
<td>0.46</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.26</td>
<td>7,224</td>
<td>0.030</td>
<td>0.092</td>
<td>0.212</td>
<td>0.41</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.30</td>
<td>5,943</td>
<td>0.040</td>
<td>0.113</td>
<td>0.131</td>
<td>0.31</td>
<td>0.82</td>
<td></td>
</tr>
</tbody>
</table>

The results present estimates of maximum sustainable yield (MSY); 1996 biomass ($S_{96}$) relative to virgin ($\gamma$) and MSY levels; recent recruitment ($R$) relative to virgin levels ($R_0$); recent average fishing mortality rate for ages 8+; and equilibrium spawning potential ratio (SPR) expected under current exploitation relative to the limit at which the stock is expected to collapse ($SPR_{crash}$). Numbers in parentheses are coefficients of variation (percent).

2. The resource is perceived to be severely depleted, as has been suggested by ICCAT assessments for well over one decade. Current spawning biomass is estimated to be at 3-4% of virgin levels, or 8-11% of the level that can sustain MSY. Recent recruitment is also low, estimated to be 13-21% of virgin levels for the most recent years.

3. The estimated annual fishing mortality rate for the spawners (ages 8 and older) in recent years ranges from 0.31 to 0.46 for the various runs. The ratio of current equilibrium spawning potential ratio ($SPR$) to $SPR_{crash}$ (equation 7) suggests that this level of fishing mortality is very high or even unsustainable (ratio values smaller than 1.0 are not sustainable).

Figure 3 contrasts various trajectories between Runs 1 and 2. The recruitment tendencies are generally similar although, as expected, the stochastic fit shows more variability. Tendencies in spawning biomass and fishing mortality relative to MSY levels are very similar, with the stochastic model estimating a larger decline in biomass between the early 1960s and early 1980s. The stock-recruitment trajectories are in general agreement between the two models (Fig. 3).

Figure 4 compares trends in estimated recruitment and spawning biomass since 1970 for the deterministic (Run 1) and stochastic (Run 2) ASPM fits, and for the VPA used in the ICCAT assessment of 1996. The 1970s biomass estimated by the VPA is substantially higher than that from either ASPM fit. However, the VPA estimates were obtained with an algorithm that is sensitive to the choices made to model fishing mortality in the plus group (Hiramatsu 1992). Thus, the estimates of biomass from the VPA...
Figure 2. Estimates of potential maximum sustainable yield (MSY) conditional on year-specific selectivity patterns for the deterministic (Run 1) and stochastic (Run 2) ASPM fits.

Figure 3. Estimated trajectories of recruitment (upper left), stock-recruitment pairs (lower left), the ratio of spawning biomass to that at MSY \((S/S_{\text{MSY}})\), upper right), and the ratio of fishing mortality to that resulting in MSY \((F/F_{\text{MSY}})\), lower right). Results are shown for the deterministic (Run 1) and stochastic (Run 2) ASPM fits.
could change substantially if the plus group were handled differently. Similarly, biomass estimates from the ASPM could be affected by changes in inputs that are assumed to be known exactly (e.g., weights at age, selectivities).

Recruitment levels for the two most recent years estimated with the ASPM are about 3 times larger than those from the VPA (Fig. 4), a difference that is potentially important in making projections into the future. We conducted deterministic projections with the results from Run 1, and stochastic projections with the results from 100 bootstraps of Run 2. In the latter case, for each bootstrap set of starting conditions, the population was projected forward at different catch levels using 100 realizations of future stochastic recruitments (equation 18). The results of these projections are shown in Fig. 5 as the median, 10th and 90th percentiles of spawning biomass and recruitment (also shown are the deterministic

Figure 4. Comparison of recruitment and spawning biomass estimates since 1970 for the VPA assessment, and the deterministic (Run 1) and stochastic (Run 2) ASPM fits.
Figure 5. Projected recruitment (in thousands) and spawning biomass (in thousand t) under four different constant catch levels (TAC). The thick solid and dashed lines give the median and approximate 80% confidence intervals from the stochastic (Run 2) projections. The circles correspond to the deterministic (Run 1) projections. The horizontal solid line is the spawning biomass at MSY level ($S_{MSY}$), used as a recovery target.
projections from Run 1). We note that these projections are not directly comparable to those of ICCAT because some of the choices we made differ from those made by the assessment working group (see ICCAT 1997 for details).

Results of the stochastic ASPM projections suggest that constant catch levels above 3,000 t are not sustainable and that catch levels below 1,000 t are (Fig. 5). Projections made with intermediate levels of catch indicate that yields of 2,500 t are probably not sustainable, but 2,000 t may be (as indicated by the median, Fig. 5). However, there is great uncertainty in these projections as suggested by the approximate 80% confidence intervals. This uncertainty is due to the variance of estimated 1996 stock sizes as well as future recruitment. The deterministic projection is clearly more pessimistic than the stochastic one. The difference can be attributed in part to the deterministic model’s failure to estimate relatively large 1988-1989 year classes (Fig. 4), which is reflected into lower adult stock sizes at the start of the projection period.

The results presented above indicate that the general perception about the depleted status of bluefin tuna obtained from application of the ASPM is quite similar to that reached by ICCAT using other methods. The main methodological focus of our work, to extend the basic ASPM in order to better model interannual changes in recruitment, also makes the stock projections more compatible with those of ICCAT (1997) (although there are differences in the median estimates of projected stock trajectories, some of the choices we made for making the projections differ from ICCAT’s). We do not propose that the stochastic ASPM presented here be used as the primary assessment method for the bluefin stock because the ASPM requires several stringent assumptions, such as known selectivities for the fishing gears. However, we recommend that the ASPM be used as an auxiliary tool to examine the stock’s trajectory in relation to MSY levels.

Acknowledgments

Support for this study was provided by the University of Miami’s (UM) Cooperative Unit for Fisheries Education and Research through National Oceanic and Atmospheric Administration Cooperative Agreement NA67RJ0149, and by the National Marine Fisheries Service (NMFS) Office of Science and Technology. The conclusions presented here are entirely the authors’ and are not necessarily endorsed by UM or NMFS. We are grateful to Clay Porch, Joe Powers, Gerry Scott, Steve Turner, and two anonymous reviewers for helpful comments on an earlier draft.
Symposium on Fishery Stock Assessment Models

References


When Lengths Are Better Than Ages: The Complex Case of Bocaccio

Stephen Ralston
National Marine Fisheries Service, Southwest Fisheries Science Center, Tiburon, California

James N. Ianelli
National Marine Fisheries Service, Alaska Fisheries Science Center, Seattle, Washington

Abstract

Bocaccio (*Sebastes paucispinis*) has historically been the most important rockfish harvested in the California groundfish fishery. A stock assessment of bocaccio in the Eureka-Monterey-Conception area indicated that in 1996 spawning biomass was 5-10% of that present in 1970. This finding was based on the application of the Stock Synthesis model to a split-sex population, assuming length-dependent gear selectivities for four distinct fisheries. A variety of fishery-dependent and fishery-independent data sources were used to model population biomass, including (1) landings from the trawl, setnet, hook-and-line, and recreational fisheries, (2) trawl catch-at-age data for the period 1980-1985 using surface otolith ages, (3) trawl catch-at-age data for 1988, 1991, and 1994 using break-and-burn otolith ages, (4) a probability transition matrix for conversion of age types, (5) length composition data from each fishery over the period 1980-1994, (6) an effort index in the recreational fishery, (7) triennial shelf trawl survey CPUE and length-frequency data, (8) a spawning biomass index derived from larval abundance in CalCOFI surveys, and (9) an index of year-class strength from a midwater trawl survey of young-of-the-year pelagic juvenile abundance.

An evaluation of these diverse sets of information indicated that the age composition data were in fundamental disagreement with all other data sources. This discrepancy was apparently due to bias and imprecision in bocaccio ages, which resulted in uninformative age composition data
that were incapable of resolving a highly variable pattern of recruitment to the fishery. For this purpose, length composition data were much more useful, especially including those from the trawl and recreational fisheries.

**Introduction**

The need to incorporate diverse sets of information into the statistical analysis of fish population dynamics has led to the development and evolution of flexible stock-assessment models (e.g., Fournier and Archibald 1982; Deriso et al. 1985; Methot 1989, 1990). Along the West Coast of the United States, Methot’s Stock Synthesis model has become the standard analytical tool for estimating the population status of groundfish stocks. Within the framework offered by the synthesis model, the inclusion of catch-at-age data has been the cornerstone of most groundfish assessments. Like Fournier and Archibald’s (1982) model, Stock Synthesis plausibly treats age composition data as measured with a multinomial error structure, but it is unique in that errors attributable to reader mis-aging can also be included in the model.

In a broader context, the use of catch-at-age data in fish stock assessments has been reviewed by Megrey (1989). It is widely presumed that estimates of the age composition of the catch are the most informative and useful data one can obtain when modeling the effects of fishing on a stock. This is particularly true of species that show a variable pattern of recruitment, as is typical of the rockfishes (*Sebastes*). Even so, a number of studies have highlighted the benefits of including “auxiliary” data in age-structured stock assessments, especially in terms of constraining the fits of population models to catch-at-age data (e.g., Bence et al. 1993; High-tower 1996).

Here we report on certain findings from a recently completed stock assessment of bocaccio (*Sebastes paucispinis*) which employed the Stock Synthesis model (detailed results available in Ralston et al. 1996). The assessment was notable in that a large number of fishery-dependent and fishery-independent data sources were involved in the analysis. Two of the fishery-independent data sources were new and had not been used previously in groundfish stock assessments conducted on the West Coast. We also included new break-and-burn age-frequency distributions in the assessment, as recommended by Bence and Rogers (1992), and evaluated the relationship between those data and the surface age composition information that had been used in the last stock assessment.

Bocaccio is an important species of rockfish that has a long history of exploitation in California (Fig. 1; Lenarz 1987, Ralston et al. 1996). It is most abundant off southern and central California and is uncommon between Cape Mendocino and Cape Blanco. A second population center exists near the Oregon-Washington border, and extends north to Cape Flattery (Gunderson and Sample 1980, Ralston et al. 1996). Bocaccio frequents an
exceptional diversity of habitats, including kelp forests, rocky reefs, midwater, and open, low relief bottoms (Eschmeyer 1983). Even though subadult growth can be very rapid in absolute terms (24 cm at age 1), adults grow slowly \((K = 0.11-0.13 \text{ yr}^{-1})\ [\text{Wilkins 1980}]\). Moreover, growth is sexually dimorphic, with females reaching much larger sizes than males (i.e., 90 versus 70 cm).

**Sources of Data**

**Fishery-Dependent Data**

The assessment was restricted to California because Oregon landings of bocaccio are virtually nil and few biological samples were collected for the small catches taken from the Washington subpopulation. California commercial landings statistics for the period 1980-1995 were summarized using procedures outlined in Erwin et al. (1997) and Pearson and Erwin (1997). Estimates of recreational landings, which have been significant, were extracted from the national Marine Recreational Fishery Survey Statistics (MRFSS) database. The catch time series was lengthened to encompass the 1950-1995 period using information and methods detailed in Ralston et al. (1996). During the last half century, bocaccio have been harvested in four distinct sectors, i.e., the trawl, hook-and-line, setnet, and recreational fisheries (Fig. 1). Each fishery has been characterized by a distinct exploitation pattern, and in the assessment each was modeled independently of the others.

Bocaccio age composition data from the trawl fishery were available for use in the assessment, although all data for the 1980-1985 period were based on surface ages (Table 1), which were thought to be biased low for older fish (Beamish 1979). To estimate the age bias of these fish, 612 bocaccio that had been surface-aged in 1983-1984 were re-aged using the break-and-burn method. That study showed that at a break-and-burn age of 10-yr the mean surface age of bocaccio was 8-yr, while at a break-and-burn age of 20-yr, surface age averaged 14-yr. These results were further analyzed and a probability transition matrix was developed to transform model age composition vectors to predicted surface age composition vectors. The matrix was included in the Stock Synthesis model as a means of generating predicted surface age data from the underlying dynamics of the model. The 1980-1985 surface age data were also supplemented with break-and-burn age data from the trawl fishery for the years 1988, 1991, and 1994 (Table 1). Last, the precision of break-and-burn ages was evaluated by re-examining 25% of all the aged fish. Based on these 275 otoliths, percent agreement between readings declined from ~90% for age-1 fish to ~10% agreement for age-20 fish. The pattern of decline reflected an exponential decay in the precision of age estimates with increasing age (Fig. 2).

Sex-specific length compositions were also available for each year and each of the three commercial fisheries for the period 1980-1994 (Pearson
Figure 1. Estimated landings of bocaccio in the Eureka-Monterey-Conception INPFC areas during the last half century.

Figure 2. Precision of bocaccio break-and-burn age data, as measured by percent agreement to the year among re-examined otoliths. Lines represent linear and exponential fits.
Table 1. Sources of data and information used in the Stock Synthesis model of the Eureka-Monterey-Conception bocaccio stock during the 1969-1996 period.

<table>
<thead>
<tr>
<th>Year</th>
<th>6 7 7 7 7 7 7 8 8 8 8 8 8 8 9 9 9 9 9 9 9 9 9 9 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6</th>
</tr>
</thead>
</table>

**Fishery Dependent**
- Trawl landings<sup>a</sup>  
  - Trawl length composition<sup>a</sup>  
  - Trawl surface age composition<sup>c</sup>  
  - Trawl break-and-burn age composition<sup>c</sup>  
- Hook-and-line landings<sup>a</sup>  
- Hook-and-line length composition<sup>b</sup>  
- Setnet landings<sup>a</sup>  
- Setnet length composition<sup>b</sup>  
- Recreational landings<sup>a</sup>  
- Recreational length composition<sup>a</sup>  
- Recreational fishing effort<sup>a</sup>

**Fishery Independent**
- Triennial trawl survey CPUE<sup>a</sup>  
- Triennial trawl survey length composition<sup>b</sup>  
- Pelagic juvenile rockfish survey<sup>a</sup>  
- CalCOFI larval abundance index<sup>a</sup>

<sup>a</sup> Primary data types (see Model Results section).
<sup>b</sup> Secondary data types.
<sup>c</sup> Tertiary data types (i.e., age composition data).

Note that recreational landings from 1990-92 were interpolated (+).
and Erwin 1997). In contrast, length composition data for the combined-sex recreational fishery were available from the MRFSS database for 1981-1989 and 1993-1995, as was a recreational fishing effort series (Table 1).

**Fishery-Independent Data**

Three sources of auxiliary survey information were used in the bocaccio assessment. These included the Alaska Fisheries Science Center’s triennial shelf survey, the Southwest Fisheries Science Center’s (SWFSC) pelagic juvenile rockfish midwater trawl survey, and the California Cooperative Oceanic Fisheries Investigation’s (CalCOFI) ichthyoplankton surveys. A brief description of each follows.

The triennial bottom trawl survey has been completed once every three years since 1977 (Table 1). The survey, which has found very widespread use in Pacific coast groundfish stock assessments, samples continental shelf habitats in the 55-366 m depth range using a high-opening Nor’easter bottom trawl equipped with bobbin roller gear (Wilkins 1996). In this study, only standard trawls conducted in the southern area (Eureka, Monterey, and Conception International North Pacific Fisheries Commission areas) were used. Although the survey is often used to provide swept-area estimates of absolute biomass, we treated the survey as a relative index of bocaccio abundance (Fig. 3). The triennial survey indicates that a substantial reduction in bocaccio biomass has occurred over the last two decades. Catch-weighted estimates of year-specific and sex-specific length compositions from the survey were also used as input data to the model.

The pelagic juvenile rockfish midwater trawl survey is designed to estimate the relative year-class strength of a group of 10 rockfish species, including bocaccio. The survey has been conducted every year since 1983 and uses a modified Cobb midwater trawl. A series of 36 standard stations are sampled during three repetitive occupations of a 110-mile study area along the central California coast. Stratified means are calculated that represent the average number of 100-day-old fish taken during a standard trawl, with the maximum value among the three occupation means providing an estimate of year-class strength (see Ralston and Howard 1995). In this instance the time series was shifted forward by one year, representing the relative abundance of age-1 bocaccio recruits (Table 1, Fig. 4). Note that the relatively low abundance of age-1 fish in 1984 and 1993 was due to the adverse effects of the 1983 and 1992 El Niños on rockfish reproductive success.

Within the California Current ecosystem, CalCOFI data have been collected over a grid of north-south lines and onshore-offshore stations since 1951. Within that time period cruises are typically identified by the year and the principal month of sampling. At occupied stations, plankton samples have been collected using both bongo and ring nets; samples are later sorted in the laboratory. The ichthyoplankton are identified and
Figure 3. Alaska Fisheries Science Center’s triennial shelf trawl survey catch-per-unit-effort (CPUE) of bocaccio in the Eureka-Monterey-Conception INPFC areas. Error bars represent ± 1.0 standard error.

Figure 4. Relative year-class strength of age-1 bocaccio in year t based on the abundance in year t-1 (data from the pelagic juvenile rockfish midwater trawl survey). Note that low survey catches in 1983 and 1992 were associated with El Niño events.
enumerated when possible and the information entered into the CalCOFI database (Moser et al. 1993). Bocaccio larvae are relatively easy to identify, but have not been sorted from the entire time series of CalCOFI collections (see Table 1 for available years). Jacobson et al. (1996) describe the use of the log-transformed data to index the abundance of bocaccio larvae using a General Linear Model (GLM), which included terms for year, month, line, station, and all non-year interaction terms. Because the survey primarily samples very young larvae, year effects from the GLM can be used to provide an index of spawner abundance (Fig. 5).

**Model Structure**

The Stock Synthesis model is a forward-projecting, separable, age-structured population model. The separability assumption requires that the fishing mortality rate experienced by fish of age \( a \) in year \( t \) (\( F_{a,t} \)) is defined by the product of a year-specific full-selection instantaneous fishing mortality rate (\( F_t \)) and an age-specific value of selectivity (\( s_a \)), i.e., \( F_{a,t} = F_t s_a \). Key features of the model are that it incorporates a multinomial error structure for both age and length composition data, it explicitly models aging errors when constructing predicted age composition data, and it conveniently allows a variety of data elements to be combined and evaluated under one umbrella formulation. In particular, all data types are combined in a total log-likelihood equation of the form:

\[
\ell_{\text{Total}} = \sum_{i=1}^{m} \ell_i \lambda_i
\]

where \( \ell_{\text{Total}} \) is the total log-likelihood of the model and the \( \ell_i \) are the individual log-likelihoods for each of the \( m \) data components used by the model. These are weighted by “emphasis” factors (\( \lambda_i \)), such that in combination the various data sources used by the model can be controlled. To reduce the influence of one data type the particular \( \lambda_i \) can be reduced to a nil emphasis (e.g., 0.001).

The model is typically configured to treat observations of age composition data to be measured with a multinomial sampling error structure. In particular, a log-likelihood component for the \( i \)th type of age data takes the form:

\[
\ell_i(p | \hat{p}) = \sum_t n_{i,t} \sum_a p_{i,a,t} \log_e (\hat{p}_{i,a,t})
\]

where \( p_{i,a,t} \) is the observed proportion of fish that are age \( a \) in samples collected in year \( t \), \( \hat{p}_{i,a,t} \) is the model’s prediction of that proportion, and \( n_{i,t} \) is the year-specific sample size upon which the observed proportions are based. The model then performs an iterative search for values of \( \hat{p}_{i,a,t} \).
that will maximize $\ell_i$. Length composition data are fitted in a similar manner. Survey data, however, are usually modeled with a lognormal error term, i.e.,

$$
\ell_i = -\sum \log_e(\sigma_{i,t}) + \frac{\log_e(I_{i,t} / \hat{I}_{i,t})^2}{2\sigma_{i,t}^2}
$$

where $\ell_i$ is the loge-likelihood component for the $i$th survey, $I_{i,t}$ is the observed value of the survey index in year $t$, $\hat{I}_{i,t}$ is the model's prediction of the index value, and $\sigma_{i,t}$ is the standard error of the statistic on loge-scale.

During the development of a baseline model for bocaccio, we explored a number of different configurations. Since the level of data and model complexity was very high (four fisheries, split sexes, unbiased and biased ages, three surveys, etc.), we decided some simplification was needed. First, because there were only estimates of total catch for the period 1950-1968, we elected to exclude that period and we modeled the population from 1969 to the present. Even so, data from the omitted period were used to establish historic catch levels. Although some CalCOFI data were

Figure 5. Long-term patterns in the abundance of larval bocaccio, as estimated by a General Linear Model (GLM) applied to the CalCOFI data base. Error bars indicate 95% confidence intervals.
available from the mid-1950s (Fig. 5), preliminary analyses indicated they had little or no effect on the model’s final estimate of current stock size.

In the prior assessment, Bence and Rogers (1992) showed that bocaccio selectivity patterns were inadequately described by the age-based Stock Synthesis model (see Methot 1990). Like them, we used the length-based implementation of the model, although the selectivity curve for the pelagic juvenile rockfish survey was modeled as full vulnerability at age-1 and no vulnerability at any other age. Similarly, in the previous assessment, a component of the trawl fishery selectivity function was allowed to vary with time. We initially explored a model with constant selectivity and compared this to models where the ascending inflection point of the selectivity curve was allowed to vary. When interannual variation in that parameter was fully expressed (i.e., a value was estimated for each year), a gain of 49 log-likelihood units was realized at the cost of eleven new parameters, representing a significant improvement in fit. We were able to further simplify the model and reduce the total number of parameters by pooling five of these, without substantially affecting the total log-likelihood of the model. A similar procedure was used for the selectivity curve in the recreational fishery, except that two parameters were allowed to vary with time (i.e., the ascending inflection point and the initial selectivity).

Thus, the final “baseline” version of the bocaccio model included eleven log-likelihood components (i.e., $m = 11$) that together controlled the fit of the model to the data. These were components for: (1) the trawl fishery length composition data, (2) trawl fishery surface age composition

### Table 2. Summary of parameters estimated in the baseline Stock Synthesis model of bocaccio ($M$ was fixed at 0.15 yr$^{-1}$).

<table>
<thead>
<tr>
<th>Model element</th>
<th>Number of parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trawl selectivity</td>
<td>Stationary 8</td>
</tr>
<tr>
<td></td>
<td>Time-varying 8</td>
</tr>
<tr>
<td>Setnet selectivity</td>
<td>9</td>
</tr>
<tr>
<td>Hook-and-line selectivity</td>
<td>9</td>
</tr>
<tr>
<td>Recreational selectivity</td>
<td>Stationary 5</td>
</tr>
<tr>
<td></td>
<td>Time-varying 21</td>
</tr>
<tr>
<td>Triennial survey selectivity</td>
<td>3</td>
</tr>
<tr>
<td>Recruit survey selectivity</td>
<td>1</td>
</tr>
<tr>
<td>CalCOFI survey selectivity</td>
<td>0</td>
</tr>
<tr>
<td>Recreational fishery effort</td>
<td>1</td>
</tr>
<tr>
<td>Growth</td>
<td>5</td>
</tr>
<tr>
<td>Recruitments</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td>98</td>
</tr>
</tbody>
</table>
data, (3) trawl fishery break-and-burn age composition data, (4) hook-and-line fishery length-frequency data, (5) setnet fishery length-frequency data, (6) recreational fishery length-frequency data, (7) a recreational fishing effort series, (8) the triennial trawl survey time series of catch rate, (9) triennial survey length compositions, (10) the pelagic juvenile rockfish survey, and (11) the CalCOFI larval abundance survey (assumed $\propto$ spawning biomass). To fit the baseline bocaccio model a total of 98 parameters were estimated (Table 2). Also note that, once converged, the model solves for the fishing mortality rate that is required to produce an exact match to the observed landings, subject to the specific constraints imposed by the estimated parameter set.

**Model Results**

All of the age data used in the previous assessment (Bence and Rogers 1992) were based on surface ages. In our assessment we attempted to incorporate new age data that were derived from the break-and-burn method. This had the unfortunate effect of adding a new level of complexity in interpreting the age data. First, it became evident that, given the spread at older ages in the transition matrix, very little information could be extracted from surface ages (Fig. 6). Second, the percent agreement between reexamined break-and-burn samples was rather poor (Fig. 2). Finally, specific age determination criteria for bocaccio had not been rigorously vali-
dated and the two age readers indicated that, at least relative to other species of rockfish, bocaccio otoliths were difficult to decipher. The lack of validation implied that the break-and-burn ages could be biased. These issues led us to question the fundamental reliability of the different data sources, particularly the age composition data.

Concerns over the age composition data component were verified when the model was fit with (1) all the age data fully emphasized, (2) reliance on break-and-burn ages only, and (3) nil emphasis on any of the age composition data (Fig. 7). In the last case the model indicated a strong recruitment event of age-1 fish occurred in 1978 and that minor but above average events occurred in 1985 and 1989. These patterns were also evident in the trawl length composition data (Fig. 8), the recreational length composition data, and in the recruit survey (Fig. 4). The strong 1977 year class was, moreover, well known to California Department of Fish and Game biologists based on a tremendous influx of small fish in the 1977 and 1978 nearshore recreational fishery. In contrast, in the first two cases, wherein the age data were allowed to influence the fit of the model, the estimated time series of recruitments was blended in the 1980s and no dominant 1977 year class was evident.

Based on these findings, we classified the data into three general groups: primary, secondary, and tertiary (Table 1). This was done to improve the robustness of model outputs and to avoid model sensitivity to data that were noisy or were otherwise questionable. The primary data types were used in the estimation of growth, year-class strength, and population trend. The secondary data were used only to estimate selectivity patterns for the different gear types and the tertiary data were effectively omitted from the analysis except to highlight their deficiencies. To fit the model to the different data classes, we followed a simple, iterative scheme, i.e., (1) fitting the model with only the primary data types emphasized, (2) fixing parameters estimated in the first step and estimating selectivity parameters for secondary data types, and (3) repeating steps one and two until the model showed no further tendency to change. The resulting baseline model was characterized by trajectories of bocaccio summary biomass and spawning output that showed severe declines over the course of the modeled period, with terminal year values in the range of 5-10% of their maxima, which occurred in 1969 (Fig. 9). The obvious “bump” in these downward trends represents the strong 1977 year-class passing through the population.

**Discussion**

The assessment of bocaccio was complicated by a number of factors. In particular, we modeled the effects of four distinct fisheries on the abundance of male and female bocaccio. In addition, the trawl and recreational fisheries were marked by significant time-varying effects on selectivity.
Figure 7. Time series of age-1 recruitments estimated with different emphasis levels on the age composition data.
Figure 8. Relative length composition distributions of bocaccio caught in the trawl fishery (1980-1994).

Figure 9. Estimated trends in summary biomass and spawning potential from the baseline bocaccio Stock Synthesis model.
We also included three separate sources of fishery-independent information in the model. Perhaps most complex, however, was the inclusion of surface and break-and-burn age information in the assessment and unraveling the conflicting signals of those data with other information in the model. In the case of bocaccio, the age data were of questionable utility, apparently due to problems with bias and imprecision (Figs. 2 and 6). In contrast, the length composition data were very informative and carried clear, unambiguous signals of strong year-classes passing through the fisheries (e.g., Fig. 8). Biologically, we believe that these two observations are related. The clear progression of modes in the length data was due to the rapid absolute growth of young bocaccio and the relatively brief seasonal expression of spawning. However, these two features exacerbated the interpretation of bocaccio otoliths. Rapid growth of subadult fish resulted in the proliferation of false annuli and accessory check marks in the otoliths, which were difficult to interpret, resolve, and validate through the application of marginal increment analysis.

We also utilized two new sources of information in a groundfish stock assessment, i.e., the SWFSC midwater trawl survey of pelagic juvenile rockfish abundance and the CalCOFI database of larval bocaccio abundance.

Figure 10. Relationship between the Southwest Fisheries Science Center's pelagic juvenile rockfish midwater trawl survey index of bocaccio abundance and year-class strength estimated from the Stock Synthesis model with no emphasis on those survey data.
These complemented the traditional triennial groundfish trawl survey, which has been used extensively in previous stock assessments. To substantiate and validate the midwater trawl survey index as a relative index of recruitment, we also fitted our final model with zero weight on the recruit survey (Fig. 10). In this independent comparison the survey index was in full agreement with the model’s interpretation of year-class strength, which was largely based on the available length-frequency data, particularly those from the trawl and recreational fisheries. It is our belief that in future applications this index will provide a reliable basis for projecting trends in biomass into the near future. Likewise, the CalCOFI database was used for the first time to index the relative spawning biomass of a groundfish. Except for the obvious outlier in 1970, it tended to closely follow the overall pattern of decline in spawning output observed in the base run model (Figs. 5 and 9). Based on our experience with these information sources, we encourage and advocate the use of both these fishery-independent auxiliary data sets in future groundfish stock assessments. However, one should always closely examine models for assumptions made about data quality.

Acknowledgments

We would like to acknowledge the collaboration and assistance of a number of people who contributed significantly to this work, including Brian Culver, Jerry Gray, Kevin Hill, Larry Jacobson, Jerry Kobylinski, Bill Lenarz, Alec MacCall, Rick Methot, Rachael Miller, Don Pearson, Jean Rogers, Dave Thomas, Wade Van Buskirk, and Mark Wilkins.

References


Integrated Catch-Age Mark-Recapture Model: Application to B.C. Sablefish Stocks

Vivian Haist
Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, B.C., Canada

Abstract
An integrated catch-age mark-recapture model is described along with its application to two British Columbia sablefish stocks. The model can be implemented with explicit migration or not, and the model structure includes age- and sex-specific depth stratification. Model parameters estimated through the analyses include recruitments, fishery parameters, emigration rates, tag loss rates, and tag reporting rates. The parameters of the catch-age and mark-recapture components of the model are estimated simultaneously. The results indicate a high degree of correlation between tag loss and reporting rate but the estimated values for these parameters are precise and robust. The tag attrition rate is greater than would be expected from natural and fishing mortality rates alone. The model accounts for this with high values for the instantaneous tag loss rate (0.26-0.27). These values are significantly higher than would be expected if tag shedding were the major component of tag loss. Implementation of the migration component of the model does not provide a satisfactory fit to the observed tag recovery data, again because the tag attrition rate for migrant fish is greater than model predictions. A multi-stock model that explicitly accounts for immigration as well as emigration may resolve some of the remaining questions.

Introduction
The use of catch-age models in fisheries stock assessments has become common practice for stocks where a reasonable time-series of catch and age composition data is available. However, catch and age composition data alone are generally inadequate to reconstruct the history of a fish population and catch-age analyses require auxiliary information to accu-
rately estimate absolute abundance (Doubleday 1976, Pope and Shepherd 1982, Deriso et al. 1985). Survey-based relative or absolute abundance indices are commonly used to tune catch-age analyses. In the absence of fishery-independent auxiliary data, catch-effort statistics from fisheries are often used. This requires the assumption that the catch per unit effort is proportional to stock abundance, an assumption which is often violated (Hilborn and Walters 1992, Chapter 5).

Alternate methods to catch-age analysis for estimating fishing mortality rates and stock abundance include various forms of mark-recapture analysis (Ricker 1975, Seber 1982, Pollock 1991). Mark-recapture analyses generally require independent estimates of model parameters such as tag loss rates and tag reporting rates because these parameters are confounded with fishing and total mortality estimates (e.g., Kleiber et al. 1983, Myers et al. 1997). Ricker (1975, pp. 113-120) suggests that tag loss can be calculated as the difference between the total mortality rate for the population as estimated from catch curve analysis and the total mortality estimates for the marked fish. However, catch curve analysis requires assumptions of regularity in recruitment and constant fishing mortality rates, assumptions that are rarely met in commercial fishery situations. These assumptions can be eliminated by integrating mark-recapture analysis in a catch-age model.

For British Columbia sablefish stocks, time-series of catch and age composition data are available; however, commonly used auxiliary data are either unavailable or are considered unreliable indices of stock abundance. Tagging studies were conducted on these stocks in the early 1980s and again during the 1990s, so a reasonable time-series of mark and recapture data is available to provide auxiliary information for catch-age analysis. In this paper I describe an integrated catch-age mark-recapture model that was developed to assess the abundance of B.C. sablefish stocks. Parameters of the catch equations and the mark-recapture equations are estimated simultaneously. The model can be implemented as a migration model, with explicit immigration and emigration parameters, or as a non-migration model. Analyses of the B.C. sablefish data are presented for both implementations of the model.

**Model Description**

*Population Dynamics and Catch Equations*

Biological data indicate that sablefish are both age- and sex-stratified. Younger fish are relatively more abundant in shallow waters and the proportion of females in samples collected in shallow waters is higher than in samples from deeper waters (Saunders et al. 1997). The model developed for the sablefish stock reconstructions includes age- and sex-specific segregation of the population by depth strata. The model structure assumes partial recruitment of younger age classes and that all recruited fish are
vulnerable to the fisheries. The catch equations, which relate the numbers of fish in the population to the numbers of fish in the catch, are described by the following relationships:

\[
C_{ijks} = \frac{F_{ik}}{Z_{iks}} \left[ 1 - \exp(-Z_{iks}) \right] p^d_{jks} p^r_{js} N_{ijs} 
\]

\[
N_{i+1,j+s} = \sum_k \exp(-Z_{iks}) p^d_{jks} p^r_{js} N_{ijs} + \exp(-M_s) (1 - p^r_{js}) N_{ijs} \quad \text{for } j < a
\]

\[
N_{i+1,a+s} = \sum_k \exp(-Z_{iks}) p^d_{a-1ks} p^r_{a-1s} N_{i,a-1s} + \exp(-M_s) (1 - p^r_{a-1s}) N_{i,a-1s} + \\
\sum_k \exp(-Z_{iks}) p^d_{a-1ks} p^r_{as} N_{ias} + \exp(-M_s) (1 - p^r_{as}) N_{ias} \quad \text{for } j = a
\]

\[
Z_{iks} = F_{ik} + M_s + G 
\]

\[
C_{ik} = \sum_j \sum_s C_{ijks} w_{js} 
\]

\[
E_{ik} = \sum_j \sum_s p^d_{jks} p^r_{js} N_{ijs} 
\]

where,
- \(i\) indexes year,
- \(j\) indexes age,
- \(a\) is the number of age classes,
- \(k\) indexes depth stratum,
- \(s\) indexes sex,
- \(p^d_{jks}\) is the proportion age class \(j\) and sex class \(s\) fish in stratum \(k\), where \(\sum_k p^d_{jks} = 1\) and \(p^d_{jks} \geq 0\),
- \(p^r_{js}\) is the proportion age class \(j\) and sex class \(s\) fish that are recruited to the exploitable population, where \(0 \leq p^r_{js} \leq 1\)
- \(C_{ijks}\) is the catch (in numbers) of age class \(j\) and sex \(s\) fish in stratum \(k\) in year \(i\)
- \(C_{ik}\) is the total catch (in biomass) in stratum \(k\) in year \(i\)
- \(F_{ik}\) is the instantaneous fishing mortality rate in stratum \(k\) in year \(i\)
- \(M_s\) is the instantaneous natural mortality rate for sex \(s\) fish,
- \(G\) is the instantaneous net migration rate,
- \(Z_{iks}\) is the instantaneous total mortality rate for sex \(s\) fish in stratum \(k\) in year \(i\)
- \(N_{ijs}\) is the number of fish in age class \(j\) and sex class \(s\) in year \(i\)
- \(E_{ik}\) is the number of exploitable fish in stratum \(k\) in year \(i\)
- \(w_{js}\) is the average weight of fish of sex \(s\) and age \(j\).
Parameterizing Partial Recruitment, Depth Stratification, and Mortalities

The partial recruitments are defined in terms of parameters $c_{ms}$ such that they are constrained between 0 and 1, are non-decreasing with age, and are fixed at 1.0 for age classes $a'$ and older. That is,

$$p_{js}^r = \frac{\sum_{m=1}^{a} c_{ms}}{\sum_{m=1}^{a'} c_{ms}} \quad \text{for } j \leq a'$$

$$p_{js}^r = 1.0 \quad \text{for } j > a'$$

The parameter $a'$ is fixed at 9 for all analyses presented here. For each sex, the proportion of an age class occurring in a depth stratum is parameterized as a linear trend from the first age class through the penultimate age class. The trend parameters, $t_{jks}$,

$$t_{jks} = b_{ks}^1 + \frac{j}{a-1} b_{ks}^2$$

are re-scaled so that for each sex the sum of the proportions for each age class are 1.0,

$$p_{jks}^d = t_{jks}/\sum_k t_{jks}$$

The proportions by depth strata for the final age class are independent of those for younger age classes.

$$p_{aks}^d = b_{ks}^3/\sum_k b_{ks}^3$$

The estimated instantaneous natural mortality rate for B.C. sablefish stocks has been estimated as 0.08 (Sunders et al. 1994). For this analysis, the average instantaneous fishing mortality rate ($M$) is fixed at 0.08, but a sex-specific deviation ($m$) from the average rate is estimated. That is,

$$M_1 = M + m \quad \text{and} \quad M_2 = M - m,$$

where $M_1$ and $M_2$ are the instantaneous natural mortality rates for males and females respectively.

The fundamental parameters, i.e., those which are estimated when minimizing the log-likelihood function, are the $\ln(N_{j1})$, $\ln(N_{j1})$, $\ln(F_{ik})$, $c_{ms}$, $b_{ks}^1$, $b_{ks}^2$, $b_{ks}^3$, $G$, and $m$. The number of females (sex 2) are assumed to be equal to the number of males in the population for the first age class.
(N_{i11} = N_{i1,2}). For the first year, the numbers of females at each age are
scaled relative to the number of males based on an estimate of their cu-
mulative total mortality. The cumulative total mortality for an age and sex
class is calculated under the assumption that fishing mortality rates for
years prior to the first year of the analysis are equal to the fishing mortal-
ity estimates for the first year and that the estimated partial recruitment
and age-sex specific depth stratification hold for earlier years.

**Likelihood of Observed Catch-at-Age**

For fitting the model to the observed catch and age composition data I
assume the error structure assumptions proposed by Fournier and Archi-
bald (1982). That is, the proportion-at-age result from taking a sample of
size \( S_{ik} \) from a population that has a multinomial distribution and the
estimates of total catch result from sampling a lognormal distribution.
The negative log-likelihood function for the data observations is then

\[
- \sum \sum \sum \sum C_{ijk} \ln(p_{ijks}) + \sum \sum \left[ \ln(C_{ik}) - \ln(\hat{C}_{ik}) \right]^2
\]

where \( S_{ijks} \) is the number of fish taken from samples of the catch in depth
stratum \( k \) in year \( i \) that were sex \( s \) and age \( j \), \( \hat{C}_{ik} \) is the observed catch in
depth stratum \( k \) in year \( i \), and the \( p_{ijks} \) are the predicted proportions at age
\( p_{ijks} = C_{ijks} / \sum j C_{ijks} \). The term \( w_C \) is determined by the prior assumption
made about the accuracy of the observed catch data. For the sablefish
analysis, I assumed \( w_C = 5000 \), which is consistent with a coefficient of
variation of about 0.01. The sample sizes, \( S_{ik} \), were scaled relative to the
actual number of samples collected with a maximum value of 200.

**Relative Abundance Data**

For the model as described there is little information in the catch and age
composition data to determine the proportions of the population in each
depth stratum. To provide some stability to these parameters, the model
has a component for the fit to survey estimates of the relative abundance
of fish in each depth stratum (\( \hat{E}_{ik} \)). Because there appear to be environ-
mentally induced anomalies in the survey data for some years, the model
is fit to the average abundance in each depth stratum rather than the time-
series of observations. The average observed relative abundance \( \hat{x}_k \) and
predicted absolute abundance \( x_k \) are

\[
\hat{x}_k = \sum \hat{E}_{ik} / n_k
\]

\[
x_k = \sum E_{ik} / n_k
\]
where \( n_k \) is the number of years with survey observations for depth stratum \( k \). Assuming a lognormal error distribution for the survey abundance data, the contribution to the negative log-likelihood equation is

\[
\ln \left( \frac{\ln(q_{x_k}) - \ln(\hat{x}_k)}{n_k} \right)^2
\]

where \( q \) is an abundance scalar and the constant \( w_p \) reflects the prior assumptions about the accuracy of the survey relative abundance data. For the sablefish analysis, values of \( w_p \) ranging from 5 to 500 were investigated. This range is consistent with coefficients of variation from 0.03 to 0.32.

**Mark-Recapture Equations**

The mark-recapture component of the integrated model tracks tag cohorts, i.e., all the fish tagged in a year. The age and sex of tagged fish is estimated based on the number of tags applied in each depth stratum and the age and sex composition estimated for the stratum. The model explicitly accounts for emigration from the tagged population but not for movement between depth strata. Movement between depth strata is implicit in the model in that the probability of a fish being in a depth stratum is related to its age. Note that the population migration rate is different from the tag cohort migration rate because the population parameter accounts for net movement (i.e., immigration plus emigration) whereas for the tag cohorts only emigration can be estimated. The following equations describe the relationships for the non-migrant component of the tag cohorts, where tag cohorts are referenced by the year in which they were tagged.

\[
T_{lijs} = \sum_k p^d_{jks} p^r_{js} \frac{N_{ik}}{E_{ik}} (1 - s) \hat{U}_{ik} \quad \text{for } i = l
\]

\[
C^T_{lik} = \sum_j \sum_s F^T_{ik} \left[ 1 - \exp(-Z^T_{iks}) \right] p^d_{jks} T_{lijs} \quad \text{for } i > l
\]

\[
Z^T_{iks} = F^T_{ik} + M^T + G^T + L \quad \text{for } i > l
\]

\[
T_{li+1,l+1s} = \sum_k \exp(-Z^T_{iks}) p^d_{jks} T_{lijs} \quad \text{for } i > l, j < a
\]

\[
T_{li+1,a,s} = \sum_k \exp(-Z^T_{iks}) p^d_{a-1ks} T_{lik,a-1s} + \sum_k \exp(-Z^T_{iks}) p^d_{aks} T_{las} \quad \text{for } i > l
\]

\[
I^T_{li} = \sum_j \sum_k G^T_{ik} \left[ 1 - \exp(-Z^T_{iks}) \right] p^d_{jks} T_{lijs} \quad \text{for } i > l
\]
where,

\( T_{ljs} \) is the number of tag cohort \( l \) fish of age \( j \) and sex \( s \) remaining in year \( i \)

\( s \) is the rate of tag loss that occurs immediately after tagging

\( \hat{U}_{lk} \) is the number of fish of tag cohort \( l \) (i.e., tagged in year \( l \)) tagged in depth stratum \( k \)

\( G^T \) is the instantaneous rate of emigration for the tagged fish

\( L \) is the instantaneous rate of tag loss

\( Z_{iks}^T \) is the total instantaneous mortality for tagged fish of sex \( s \) in year \( i \)

\( C_{lik}^T \) is the predicted catch of tagged fish from tag cohort \( l \) in year \( i \) and depth stratum \( k \)

\( I_{li} \) is the number of number of fish from tag cohort \( l \) migrating out of the population in year \( i \)

A number of the relationships described above are defined for a tag cohort only for the years following the tagging year (\( i > l \)). During the same year that tags are applied, the tagged fish are susceptible to only a fraction of the annual mortalities, dependent on the time of year tagging takes place. For the situation, \( i = l \), the quantities, \( F_{ik} \), \( Z_{iks}^T \), and \( G^T \) in the preceding mark-recapture equations are replaced with the following quantities: \( F'_{ik} = t_{ik} F_{ik} \), \( Z_{iks}^T = h_i Z_{iks} \), and \( G^T = h_i G^T \). The \( t_{ik} \) are the proportions of the annual catch in year \( i \) and depth stratum \( k \) which are taken after tagging occurs and \( h_i \) is the proportion of year \( i \) remaining after tagging occurs. The values of the \( t_{ik} \) and \( h_i \) are calculated from the data and are fixed model inputs.

Because migration, as defined, is continuous throughout the year, not all fish that migrate during a year will be vulnerable to the fisheries in the area they migrate to. The model structure assumes that the fish that migrate during a year will incur only half of the fishing mortality for that year. The relationships that describe the dynamics for the migrant fish are

\[
V_{li+1} = \exp(-0.5 h_i F_{i}^{TM}) I_{li} \quad \text{for } i = l
\]

\[
C_{li}^{TM} = \left[1 - \exp(-0.5 h_i F_{i}^{TM})\right] I_{li} \quad \text{for } i = l
\]

\[
V_{li} = \exp(-Z_{i}^{TM}) V_{li} + \exp(-0.5F_{i}^{TM}) I_{li} \quad \text{for } i > l
\]

\[
Z_{i}^{TM} = F_{i}^{TM} + M + L \quad \text{for } i > l
\]

\[
C_{li}^{TM} = \frac{F_{i}^{TM}}{Z_{i}^{TM}} \left[1 - \exp(-Z_{i}^{TM})\right] V_{li} + \left[1 - \exp(-0.5F_{i}^{TM})\right] I_{li} \quad \text{for } i > l
\]
where,

\[ V_{li} \] is the number of tagged fish from tag cohort \( l \) in the migrant pool in year \( i \)

\[ F_{i}^{TM} \] is the instantaneous fishing mortality rate for tagged migrant fish in year \( i \)

\[ Z_{i}^{TM} \] is the instantaneous total mortality rate for tagged migrant fish in year \( i \)

\[ C_{li}^{TM} \] is the predicted catch of migrant fish from tag cohort \( l \) fish in year \( i \)

I do not attempt to estimate all the annual fishing mortality rate parameters for the migrant tagged fish, because the information relative to these parameters is limited (i.e., only tag recovery data) and the parameter values will be highly confounded with other model parameters. Rather, two parameters are estimated, one for the first year and one for the last year (\( F_{80}^{TM} \) and \( F_{96}^{TM} \)), and a linear trend between these two is assumed. More complex formulations than the linear time trend could be developed; however, preliminary analyses suggested that there is not enough information in the data to estimate even an average level of fishing mortality for the migrant tagged fish.

Likelihood of Tag Recoveries

There are two observation error structure models commonly used when fitting tag recovery observations in mark-recapture analysis (Seber 1982, Hilborn 1990). These are the Poisson and the multinomial distribution. I have implemented both in the integrated sablefish model and as reported by Hilborn (1990), found that results are virtually identical for the two. Results presented in this paper are based on analyses conducted with the multinomial implementation. The quantities required to fit the multinomial model are the total tagged for a tag cohort \( (\hat{U}_l) \),

\[ \hat{U}_l = \sum_k \hat{U}_{lk} , \]

the total number of tags in a tag cohort which are not recovered \( (\hat{W}_l) \),

\[ \hat{W}_l = \hat{U}_l - \left( \sum_l \sum_k \hat{C}_{lk} + \sum_l \hat{C}_{li}^{TM} \right) , \]

\( \hat{C}_{lk} \) and \( \hat{C}_{li}^{TM} \) are the number of tags from tag cohort \( l \) returned in year \( i \) from depth stratum \( k \) and from outside the assessment region, respectively, and the predicted proportion of tag releases recovered in each recovery stratum \( (g_{lik} \text{ and } g_{li}^{TM}) \) and the predicted proportion not recovered \( (\hat{g}_l^{w}) \),

\[ g_{lik} = r_{i} C_{lik}^{T}/\hat{U}_l , \]

\[ \hat{g}_l^{w} = 1 - \sum_l g_{lik} . \]
Note that additional mark-recapture parameters are introduced in these equations; $r_i$ is the proportion of tagged fish that are reported in year $i$. Two alternate hypotheses regarding the reporting rates are explored. The first is that reporting rates have been constant over the time-series. In this case the parameters, $r_i$, are replaced with a single parameter (i.e., $r_i = r^*$). Alternately, the $r_i$ have an autoregressive form, that is

$$r_i = r^*$$

$$r_i = \exp[\ln(r_{i-1}) + \delta_i] \quad \text{for} \quad i > 1$$

where the $\delta_i$ are independent normally distributed random variables.

The negative of the log-likelihood for the mark-recapture component of the model, assuming a multinomial error structure for tag recovery data is then,

$$- w_T \sum_i \left\{ \sum_l \left[ \tilde{C}^T_{il} \ln(g_{il}) \right] + \sum_l \left[ \tilde{C}_M^T \ln(g_{il}^M) \right] + \tilde{W}_i \ln(g_{il}^M) \right\} + w_R \sum_{i=2}^n \delta_i^2$$

where the constants $w_T$ and $w_R$ are adjusted to reflect prior assumptions about the variance of the tag recovery data and the trends in reporting rates, respectively. The additional fundamental parameters for the mark-recapture component of the model, i.e., ones that are estimated through the minimization, are $G, G^T, F_{80}^M, F_{96}^M, L, r^*$, and $\delta_r$. The full objective function is the sum of equations 1, 2, and 3. Note that the model as described can also be implemented as a model without migration if the value of the parameters; $G, G^T, F_{80}^M$, and $F_{90}^M$ are fixed at 0. This implementation of the model will be termed the "non-migration" model, as opposed to the "migration" model where all the mark-recapture related parameters are estimated.

**Model Implementation**

The model, as described, is implemented using AD Model Builder software (Otter Research Ltd. 1994). This software uses automatic differentiation for calculating the derivatives needed for finding the posterior mode via a quasi-Newton function minimization routine. The software provides
estimates of the variance-covariance matrix for all dependent and specified independent variables. Likelihood profiles of key quantities of interest are also produced. The following list shows model parameters whose values are either fixed or constrained, except where explicitly stated otherwise.

<table>
<thead>
<tr>
<th>Fixed parameters</th>
<th>Constrained parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w_c = 5,000 )</td>
<td>( G \geq 0 )</td>
</tr>
<tr>
<td>( w_e = 50 )</td>
<td>( G^T \geq 0 )</td>
</tr>
<tr>
<td>( w_i = 0.5 )</td>
<td>( s = 0.15 )</td>
</tr>
<tr>
<td>( w_e = 20 )</td>
<td>( 0 \leq L \leq 1.0 )</td>
</tr>
<tr>
<td>( M = 0.08 )</td>
<td>( 0 \leq r \leq 1.0 )</td>
</tr>
<tr>
<td>( \delta_i = 0.0 )</td>
<td></td>
</tr>
</tbody>
</table>

**Data Sources**

Analyses using the integrated catch-age mark-recapture model were conducted for the northern B.C. and the southern B.C. sablefish stocks with the data segregated into three depth strata (<500 m, 500-800 m, >800 m). The commercial sablefish fishery is composed of three gear types (trap, longline, and trawl), with the trap fishery accounting for 80 to 85% of the total annual landings. There is no program to obtain biological samples from the commercial catch; however, since 1980 biological data have been collected during research trips onboard commercial trap fishing vessels using commercial gear. In some cases the commercial vessels were chartered specifically for research purposes (e.g., tagging and abundance surveys) and in other cases research activities, primarily tagging, occurred during regular commercial fishing operations. Because these research activities used commercial trap gear and fished in commercial fishery locations, age composition samples collected during the research operations should reflect fishing selectivities similar to the commercial trap fishery. However, age and sex-specific fishing selectivity for the longline and trawl fisheries probably differs from the trap fishery. The estimates of the age and sex composition obtained during the research operations are used to approximate the age and sex composition of the entire commercial catch. It is possible that these are biased estimates of the removals from the stock, particularly for the longline and trap components. A potential bias arising from the assumption of common fishery selectivities for the different gear types would be worth exploring, but is beyond the scope of the current analysis. Annual catch data, by stock and depth strata, are shown in Table 1 for the 1980 to 1996 period. Annual age and sex composition data is plotted in Fig. 1 and summarized in Table 2.

Annual relative abundance indices were calculated from survey CPUE (catch in numbers per trap) data for each depth strata. The mean annual CPUEs are assumed to be indices of fish density and these estimates were weighted by the relative size of the depth strata (Saunders and McFarlane
Table 1. Catch in metric tons by depth strata for two B.C. stocks.

<table>
<thead>
<tr>
<th>Year</th>
<th>Southern B.C.</th>
<th></th>
<th></th>
<th></th>
<th>Northern B.C.</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;500 m</td>
<td>500-800 m</td>
<td>&gt;800 m</td>
<td>Total</td>
<td>&lt;500 m</td>
<td>500-800 m</td>
<td>&gt;800 m</td>
<td>Total</td>
</tr>
<tr>
<td>1980</td>
<td>422</td>
<td>1,466</td>
<td>302</td>
<td>2,192</td>
<td>384</td>
<td>896</td>
<td>321</td>
<td>1,602</td>
</tr>
<tr>
<td>1981</td>
<td>303</td>
<td>924</td>
<td>106</td>
<td>1,334</td>
<td>660</td>
<td>971</td>
<td>775</td>
<td>2,407</td>
</tr>
<tr>
<td>1982</td>
<td>260</td>
<td>1,116</td>
<td>149</td>
<td>1,526</td>
<td>541</td>
<td>932</td>
<td>864</td>
<td>2,338</td>
</tr>
<tr>
<td>1983</td>
<td>234</td>
<td>1,073</td>
<td>163</td>
<td>1,471</td>
<td>217</td>
<td>1,062</td>
<td>1,308</td>
<td>2,588</td>
</tr>
<tr>
<td>1984</td>
<td>141</td>
<td>1,487</td>
<td>10</td>
<td>1,639</td>
<td>250</td>
<td>1,399</td>
<td>327</td>
<td>1,977</td>
</tr>
<tr>
<td>1985</td>
<td>172</td>
<td>1,813</td>
<td>0</td>
<td>1,986</td>
<td>439</td>
<td>1,496</td>
<td>111</td>
<td>2,047</td>
</tr>
<tr>
<td>1986</td>
<td>930</td>
<td>1,181</td>
<td>70</td>
<td>2,182</td>
<td>742</td>
<td>1,380</td>
<td>143</td>
<td>2,265</td>
</tr>
<tr>
<td>1987</td>
<td>913</td>
<td>1,508</td>
<td>27</td>
<td>2,449</td>
<td>440</td>
<td>1,212</td>
<td>481</td>
<td>2,133</td>
</tr>
<tr>
<td>1988</td>
<td>1,609</td>
<td>1,101</td>
<td>117</td>
<td>2,828</td>
<td>291</td>
<td>1,797</td>
<td>565</td>
<td>2,654</td>
</tr>
<tr>
<td>1989</td>
<td>1,135</td>
<td>1,532</td>
<td>193</td>
<td>2,861</td>
<td>478</td>
<td>2,001</td>
<td>150</td>
<td>2,631</td>
</tr>
<tr>
<td>1990</td>
<td>1,188</td>
<td>949</td>
<td>182</td>
<td>2,319</td>
<td>1,252</td>
<td>1,382</td>
<td>157</td>
<td>2,793</td>
</tr>
<tr>
<td>1991</td>
<td>689</td>
<td>594</td>
<td>84</td>
<td>1,368</td>
<td>754</td>
<td>3,181</td>
<td>164</td>
<td>4,100</td>
</tr>
<tr>
<td>1992</td>
<td>628</td>
<td>245</td>
<td>142</td>
<td>1,016</td>
<td>592</td>
<td>3,462</td>
<td>300</td>
<td>4,355</td>
</tr>
<tr>
<td>1993</td>
<td>457</td>
<td>1,254</td>
<td>54</td>
<td>1,765</td>
<td>297</td>
<td>2,526</td>
<td>519</td>
<td>3,343</td>
</tr>
<tr>
<td>1994</td>
<td>571</td>
<td>1,360</td>
<td>48</td>
<td>1,980</td>
<td>515</td>
<td>2,372</td>
<td>274</td>
<td>3,162</td>
</tr>
<tr>
<td>1995</td>
<td>401</td>
<td>1,236</td>
<td>232</td>
<td>1,870</td>
<td>283</td>
<td>1,416</td>
<td>605</td>
<td>2,305</td>
</tr>
<tr>
<td>1996</td>
<td>564</td>
<td>736</td>
<td>33</td>
<td>1,334</td>
<td>213</td>
<td>1,615</td>
<td>260</td>
<td>2,089</td>
</tr>
</tbody>
</table>

1993, Table 5.6) to generate relative abundance estimates for each stratum. The depth strata area measurement were based on slightly different geographical units from those used for the current analyses, so the abundance indices may not provide accurate information on the relative abundance by depth.

Sablefish were tagged in British Columbia during the late 1970s through early 1980s and from 1991 to the present. Only tag release data since 1980 and recovery data from these taggings are used. Summaries of tag and recapture data are presented in Table 3. Of the tags applied to fish in the southern B.C. stock assessment region, approximately 84% were recovered in the southern region, with the remainder of the recoveries primarily in the northern B.C. assessment region and in Alaska. For the northern B.C. assessment region, 79% of recoveries were in northern B.C. and 12% were in Alaska.

Stock Reconstructions

A series of stock reconstructions were conducted using both the migration and the non-migration implementations of the sablefish model. For the migration version of the model all recoveries of tags which occurred
Figure 1. Observed (open circles) and predicted (filled circles) proportion-at-age and sex for northern and southern B.C. depth zones, 1980-1996. Circle areas are proportional to the proportion-at-age and sum to one across age and sex.
Figure 1. (Continued.)
Table 2. Average (1980-1996) observed percent-at-age and sex, by stock and depth strata.

<table>
<thead>
<tr>
<th>Strata</th>
<th>Sex</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15+</th>
<th>Sex ratio</th>
<th>Sex ratio</th>
<th>15+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern B.C.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shallow</td>
<td>M</td>
<td>0.8</td>
<td>3.0</td>
<td>4.6</td>
<td>4.3</td>
<td>3.9</td>
<td>3.2</td>
<td>3.0</td>
<td>2.4</td>
<td>1.7</td>
<td>1.1</td>
<td>0.8</td>
<td>0.8</td>
<td>0.4</td>
<td>3.7</td>
<td>34</td>
<td>40</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.1</td>
<td>7.1</td>
<td>8.7</td>
<td>10.1</td>
<td>7.7</td>
<td>7.8</td>
<td>6.9</td>
<td>4.5</td>
<td>2.8</td>
<td>1.5</td>
<td>1.7</td>
<td>1.2</td>
<td>0.8</td>
<td>4.4</td>
<td></td>
<td>4.7</td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>M</td>
<td>1.3</td>
<td>3.2</td>
<td>3.9</td>
<td>4.3</td>
<td>4.5</td>
<td>4.1</td>
<td>3.0</td>
<td>2.3</td>
<td>2.1</td>
<td>1.8</td>
<td>1.2</td>
<td>1.5</td>
<td>1.1</td>
<td>11.4</td>
<td>46</td>
<td>46</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.3</td>
<td>4.2</td>
<td>6.0</td>
<td>8.2</td>
<td>6.0</td>
<td>5.0</td>
<td>4.2</td>
<td>3.3</td>
<td>2.4</td>
<td>2.3</td>
<td>1.3</td>
<td>1.1</td>
<td>1.0</td>
<td>8.0</td>
<td></td>
<td>7.6</td>
<td></td>
</tr>
<tr>
<td>Deep</td>
<td>M</td>
<td>0.1</td>
<td>0.6</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>0.9</td>
<td>1.2</td>
<td>0.9</td>
<td>0.9</td>
<td>1.0</td>
<td>1.4</td>
<td>1.0</td>
<td>43.6</td>
<td>53</td>
<td>54</td>
<td>42.1</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0.1</td>
<td>0.7</td>
<td>0.4</td>
<td>0.8</td>
<td>0.9</td>
<td>1.3</td>
<td>1.8</td>
<td>2.9</td>
<td>2.0</td>
<td>2.0</td>
<td>1.7</td>
<td>1.9</td>
<td>28.3</td>
<td></td>
<td></td>
<td>29.5</td>
<td></td>
</tr>
<tr>
<td>Northern B.C.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shallow</td>
<td>M</td>
<td>0.7</td>
<td>0.7</td>
<td>2.1</td>
<td>3.1</td>
<td>2.8</td>
<td>2.7</td>
<td>1.8</td>
<td>1.3</td>
<td>1.0</td>
<td>1.5</td>
<td>0.7</td>
<td>0.7</td>
<td>0.5</td>
<td>4.4</td>
<td>24</td>
<td>26</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0.7</td>
<td>2.1</td>
<td>5.7</td>
<td>8.1</td>
<td>10.7</td>
<td>10.0</td>
<td>8.6</td>
<td>5.6</td>
<td>4.3</td>
<td>4.2</td>
<td>2.4</td>
<td>2.0</td>
<td>1.6</td>
<td>10.1</td>
<td></td>
<td>11.9</td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>M</td>
<td>0.8</td>
<td>1.9</td>
<td>2.2</td>
<td>3.4</td>
<td>3.9</td>
<td>3.5</td>
<td>3.4</td>
<td>2.2</td>
<td>1.7</td>
<td>1.5</td>
<td>1.2</td>
<td>1.1</td>
<td>0.8</td>
<td>8.7</td>
<td>36</td>
<td>41</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.1</td>
<td>2.7</td>
<td>3.9</td>
<td>6.0</td>
<td>7.7</td>
<td>7.8</td>
<td>6.3</td>
<td>5.2</td>
<td>3.8</td>
<td>3.0</td>
<td>2.6</td>
<td>1.9</td>
<td>1.7</td>
<td>10.0</td>
<td></td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td>Deep</td>
<td>M</td>
<td>0.1</td>
<td>1.2</td>
<td>1.3</td>
<td>0.6</td>
<td>1.4</td>
<td>2.2</td>
<td>1.8</td>
<td>1.5</td>
<td>1.7</td>
<td>1.3</td>
<td>1.5</td>
<td>0.8</td>
<td>1.2</td>
<td>18.2</td>
<td>35</td>
<td>34</td>
<td>14.6</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0.2</td>
<td>1.3</td>
<td>1.3</td>
<td>2.3</td>
<td>3.7</td>
<td>5.4</td>
<td>4.7</td>
<td>4.7</td>
<td>4.4</td>
<td>3.5</td>
<td>3.5</td>
<td>3.0</td>
<td>3.3</td>
<td>24.3</td>
<td></td>
<td>22.8</td>
<td></td>
</tr>
</tbody>
</table>

Percentages sum to 100 across age and sex for each depth stratum.
### Table 3. Number of fish tagged by stock, year, and depth strata and number of tag recoveries by tag year and area.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number tagged by depth</th>
<th>Number of recoveries by area</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500 m</td>
<td>800 m</td>
</tr>
<tr>
<td>Southern B.C.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>0</td>
<td>4,410</td>
</tr>
<tr>
<td>1982</td>
<td>1,632</td>
<td>1,091</td>
</tr>
<tr>
<td>1991</td>
<td>0</td>
<td>525</td>
</tr>
<tr>
<td>1992</td>
<td>326</td>
<td>1,030</td>
</tr>
<tr>
<td>1993</td>
<td>530</td>
<td>2,045</td>
</tr>
<tr>
<td>1994</td>
<td>605</td>
<td>618</td>
</tr>
<tr>
<td>1995</td>
<td>1,925</td>
<td>2,258</td>
</tr>
<tr>
<td>Total</td>
<td>5,018</td>
<td>11,977</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Number tagged by depth</th>
<th>Number of recoveries by area</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500 m</td>
<td>800 m</td>
</tr>
<tr>
<td>Northern B.C.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>641</td>
<td>3,555</td>
</tr>
<tr>
<td>1981</td>
<td>123</td>
<td>1,901</td>
</tr>
<tr>
<td>1982</td>
<td>472</td>
<td>2,242</td>
</tr>
<tr>
<td>1991</td>
<td>0</td>
<td>555</td>
</tr>
<tr>
<td>1992</td>
<td>45</td>
<td>755</td>
</tr>
<tr>
<td>1993</td>
<td>170</td>
<td>1,552</td>
</tr>
<tr>
<td>1994</td>
<td>874</td>
<td>3,282</td>
</tr>
<tr>
<td>1995</td>
<td>577</td>
<td>9,959</td>
</tr>
<tr>
<td>Total</td>
<td>2,902</td>
<td>23,801</td>
</tr>
</tbody>
</table>

Proportion of total recoveries by area shown in brackets.

outside of the assessment (and tagging) region were treated as a single migrant pool. For the non-migration version of the model, tags that were recovered in areas outside the assessment area were treated as if they had been recovered within the assessment area. This treatment of the data would be consistent with the stock dynamics if there were no net stock migration (i.e., immigration is equal to emigration) and if the fishing mortality rates on the tagged migrant population were the same as the rates on the tagged non-migrant population. The first series of results presented are from analyses using the migration model.

**Migration Model**

Initial runs using the migration implementation of the model indicated that values for the migrant fishing mortality parameters ($F_{80}^{TM}, F_{96}^{TM}$) were unrealistically high so a series of runs were conducted where the value of these parameters were constrained to maximum levels ranging from 0.08 to 0.45. Estimates of the mark-recapture model parameters for this series of analyses are shown in Table 4. For the southern stock the 1980 fishing
Table 4. Parameter estimates from the migration model for the southern and northern B.C. sablefish stocks.

<table>
<thead>
<tr>
<th>Constraint on $F^{TM}$</th>
<th>$G$</th>
<th>$G^T$</th>
<th>$F^{TM}_{80}$</th>
<th>$F^{TM}_{96}$</th>
<th>$L$</th>
<th>$r$</th>
<th>Function value</th>
<th>1997 biomass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern B.C.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>0.00</td>
<td>0.10</td>
<td>0.08</td>
<td>0.07</td>
<td>0.29</td>
<td>1.00</td>
<td>1173.8</td>
<td>27.6</td>
</tr>
<tr>
<td>0.15</td>
<td>0.00</td>
<td>0.06</td>
<td>0.15</td>
<td>0.13</td>
<td>0.31</td>
<td>1.00</td>
<td>1164.4</td>
<td>28.0</td>
</tr>
<tr>
<td>0.25</td>
<td>0.00</td>
<td>0.04</td>
<td>0.25</td>
<td>0.21</td>
<td>0.32</td>
<td>1.00</td>
<td>1157.2</td>
<td>28.5</td>
</tr>
<tr>
<td>0.35</td>
<td>0.00</td>
<td>0.03</td>
<td>0.35</td>
<td>0.30</td>
<td>0.32</td>
<td>1.00</td>
<td>1152.4</td>
<td>28.9</td>
</tr>
<tr>
<td>0.45</td>
<td>0.00</td>
<td>0.03</td>
<td>0.45</td>
<td>0.39</td>
<td>0.32</td>
<td>1.00</td>
<td>1148.9</td>
<td>29.2</td>
</tr>
<tr>
<td>Northern B.C.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>0.03</td>
<td>0.22</td>
<td>0.04</td>
<td>0.08</td>
<td>0.29</td>
<td>1.00</td>
<td>1595.2</td>
<td>25.6</td>
</tr>
<tr>
<td>0.15</td>
<td>0.01</td>
<td>0.19</td>
<td>0.06</td>
<td>0.15</td>
<td>0.29</td>
<td>0.76</td>
<td>1583.0</td>
<td>19.5</td>
</tr>
<tr>
<td>0.25</td>
<td>0.00</td>
<td>0.15</td>
<td>0.10</td>
<td>0.25</td>
<td>0.29</td>
<td>0.66</td>
<td>1575.8</td>
<td>17.4</td>
</tr>
<tr>
<td>0.35</td>
<td>0.00</td>
<td>0.12</td>
<td>0.13</td>
<td>0.35</td>
<td>0.29</td>
<td>0.63</td>
<td>1573.7</td>
<td>16.7</td>
</tr>
<tr>
<td>0.45</td>
<td>0.00</td>
<td>0.11</td>
<td>0.16</td>
<td>0.45</td>
<td>0.29</td>
<td>0.61</td>
<td>1573.2</td>
<td>16.2</td>
</tr>
</tbody>
</table>

The mortality parameter is at the upper limit for all runs, while for the northern stock the 1996 estimates are consistently at the upper limit. Results from these runs show a high correlation between the fishing mortality parameters and the tag migration rate parameter, although estimates of 1997 exploitable biomass are relatively insensitive to these parameters. The value of the population migration parameter was 0 for most runs, indicating that the best fit to the observed catch and age composition data is obtained with no immigration to the population.

The residuals (standard normal deviates) of the predicted versus observed proportion of tag recoveries by recovery stratum are shown in Fig. 2, for the runs with the maximum fishing mortality parameters fixed at 0.15. For the migrant tag recoveries, the patterns of residuals follow a strong nonrandom pattern with primarily negative residuals for 0 and 1 year-at-large followed by positive residuals for 2 and greater years-at-large. That is, the model is unable to fit the observed attrition of tag recoveries for migrant fish that occurs over time. The analyses with high $F$ values on migrant fish and low migration rates provide a somewhat better fit to the data observations because under this scenario the migrant pool of tagged fish decreases at a faster rate. With higher migration rates and lower $F$ values the ongoing emigration of tagged fish maintains the numbers in the migrant pool. However, the nonrandom pattern of residuals holds even for the runs where the maximum value for the migrant fishing mortality parameter was fixed at 0.45. Clearly, the pattern in the residuals suggests that the model formulation for the migration implementation is inconsistent with the data observations.
Figure 2. Tag recovery residuals by tag cohort and years-at-large for the migration model. Positive residuals are represented by open circles and negative residuals by closed circles. The circle area is proportional to the absolute value of the residual.
**Non-Migration Model**

A series of runs were conducted with the non-migration implementation of the integrated model to evaluate the sensitivity of parameter estimates to the penalty weights for the tag data and the survey abundance data. The weightings for the tag data are 1.0, 0.5, and 0.1, and for the survey data they are 500, 50, and 5. Results from this series of runs are shown in Table 5. For the runs with a high weight on the survey data, the model estimates of the proportions of the population in each depth stratum fit the observations almost exactly. With lower weighting on the fit to the survey data the model estimates a higher proportion of the population in the middle depth stratum for the northern population and a higher proportion in the shallow stratum for the southern population. The tag loss, reporting rate, and 1997 exploitable biomass estimates are relatively insensitive to the weightings, with the exception of analyses for the northern stock with low weight on the fit to the tagging data. The likelihood profiles indicate relatively precise estimates for these model parameters, given the model structure and data observations (Fig. 3). The fit to the age composition data is similar for all weighting combinations, and there is reasonable agreement between predicted and observed proportions-at-age (Fig. 1). For further analyses the penalty weights used are 0.5 for the fit to the tagging data and 50 for the fit to the survey relative abundance data.

Although the estimated values for the tag loss parameter appear to be relatively precise, they are high relative to expected values if this parameter accounts primarily for tag shedding. A series of runs were conducted with the tag loss parameter fixed at values ranging from 0.05 to 0.35 (Table 6). The model fits to the data observations, as measured by the objective function value, deteriorates significantly when the tag loss parameter value is less than 0.20. Also, the reporting rate parameter is highly correlated with the tag loss parameter.

For the analyses described so far, the value of the initial tag loss parameter was fixed at a value of 0.15, as reported by Beamish and McFarlane (1988). This parameter is intended to account for the immediate losses of tags from the tag cohort resulting from tag shedding and tag-induced mortality. Table 7 shows the results of model runs where the value of this parameter was fixed at levels ranging from 0.05 to 0.35. As the results for the northern stock show, this parameter is completely confounded with the reporting rate parameter. That is, as the initial tag loss value is changed, the values for the tag loss parameter, the 1997 exploitable biomass, and the function value remain the same and only the value of the reporting rate parameter changes. For the southern stock, the estimate for the reporting rate parameter is at its upper bound (1.0) so the correlation with the initial tag loss parameter is not apparent.

Because of concerns that the reporting rate for recaptured tags has changed over time, a final set of analyses was conducted with year-dependent
Table 5. Parameter estimates from the non-migration model resulting from alternative weightings for the tag data ($W_T$) and relative survey abundance data ($w_E$).

<table>
<thead>
<tr>
<th>$w_T$</th>
<th>$w_E$</th>
<th>&lt;500</th>
<th>500-800</th>
<th>&gt;800</th>
<th>L</th>
<th>r</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>5</td>
<td>0.09</td>
<td>0.79</td>
<td>0.12</td>
<td>0.26</td>
<td>0.59</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.16</td>
<td>0.72</td>
<td>0.12</td>
<td>0.26</td>
<td>0.59</td>
<td>12.9</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.22</td>
<td>0.65</td>
<td>0.13</td>
<td>0.26</td>
<td>0.59</td>
<td>12.9</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>0.08</td>
<td>0.81</td>
<td>0.11</td>
<td>0.26</td>
<td>0.59</td>
<td>13.6</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.15</td>
<td>0.73</td>
<td>0.11</td>
<td>0.26</td>
<td>0.59</td>
<td>13.6</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.21</td>
<td>0.65</td>
<td>0.13</td>
<td>0.26</td>
<td>0.60</td>
<td>13.8</td>
</tr>
<tr>
<td>0.1</td>
<td>5</td>
<td>0.06</td>
<td>0.83</td>
<td>0.11</td>
<td>0.27</td>
<td>0.63</td>
<td>16.3</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.07</td>
<td>0.82</td>
<td>0.10</td>
<td>0.23</td>
<td>0.51</td>
<td>11.9</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.11</td>
<td>0.77</td>
<td>0.11</td>
<td>0.16</td>
<td>0.35</td>
<td>7.9</td>
</tr>
<tr>
<td>Observed:</td>
<td></td>
<td>0.22</td>
<td>0.64</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$w_T$</th>
<th>$w_E$</th>
<th>&lt;500</th>
<th>500-800</th>
<th>&gt;800</th>
<th>L</th>
<th>r</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>5</td>
<td>0.72</td>
<td>0.11</td>
<td>0.17</td>
<td>0.30</td>
<td>1.0</td>
<td>29.7</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.54</td>
<td>0.20</td>
<td>0.26</td>
<td>0.28</td>
<td>1.0</td>
<td>28.0</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.45</td>
<td>0.22</td>
<td>0.33</td>
<td>0.27</td>
<td>1.0</td>
<td>28.4</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>0.74</td>
<td>0.09</td>
<td>0.17</td>
<td>0.29</td>
<td>1.0</td>
<td>29.4</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.52</td>
<td>0.21</td>
<td>0.27</td>
<td>0.27</td>
<td>1.0</td>
<td>26.6</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.45</td>
<td>0.22</td>
<td>0.33</td>
<td>0.26</td>
<td>1.0</td>
<td>26.8</td>
</tr>
<tr>
<td>0.1</td>
<td>5</td>
<td>0.67</td>
<td>0.14</td>
<td>0.19</td>
<td>0.28</td>
<td>1.0</td>
<td>27.8</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.48</td>
<td>0.21</td>
<td>0.31</td>
<td>0.26</td>
<td>1.0</td>
<td>25.5</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.45</td>
<td>0.22</td>
<td>0.34</td>
<td>0.25</td>
<td>1.0</td>
<td>25.6</td>
</tr>
<tr>
<td>Observed:</td>
<td></td>
<td>0.44</td>
<td>0.22</td>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model parameters are: tag loss rate ($L$), tag reporting rate ($r$), and 1997 exploitable biomass ($B$).
Figure 3. Likelihood profiles for tag loss rates, tag reporting rates, and 1997 biomass estimates. The solid lines and dashed lines are from analyses with penalty weights of 0.5 and 0.1 on the tag data, respectively. The weighting for the survey data is 50 in both cases.
Table 6. Parameter estimates for alternative runs of the non-migration model with the tag loss parameter fixed at various levels.

<table>
<thead>
<tr>
<th>Tag loss</th>
<th>Reporting rate</th>
<th>Function value</th>
<th>1997 biomass</th>
<th>Reporting rate</th>
<th>Function value</th>
<th>1997 biomass</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.59</td>
<td>1,272.8</td>
<td>22.5</td>
<td>0.30</td>
<td>1,718.9</td>
<td>6.9</td>
</tr>
<tr>
<td>0.10</td>
<td>0.82</td>
<td>1,227.0</td>
<td>29.5</td>
<td>0.36</td>
<td>1,673.5</td>
<td>8.4</td>
</tr>
<tr>
<td>0.15</td>
<td>1.00</td>
<td>1,188.1</td>
<td>33.5</td>
<td>0.43</td>
<td>1,643.3</td>
<td>10.0</td>
</tr>
<tr>
<td>0.20</td>
<td>1.00</td>
<td>1,162.8</td>
<td>30.1</td>
<td>0.50</td>
<td>1,625.9</td>
<td>11.6</td>
</tr>
<tr>
<td>0.25</td>
<td>1.00</td>
<td>1,151.2</td>
<td>27.6</td>
<td>0.58</td>
<td>1,619.4</td>
<td>13.3</td>
</tr>
<tr>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td>0.59</td>
<td>1,619.2</td>
<td>13.6</td>
</tr>
<tr>
<td>0.27</td>
<td>1.00</td>
<td>1,149.9</td>
<td>26.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>1.00</td>
<td>1,151.2</td>
<td>25.6</td>
<td>0.66</td>
<td>1,622.1</td>
<td>14.9</td>
</tr>
<tr>
<td>0.35</td>
<td>1.00</td>
<td>1,160.8</td>
<td>23.9</td>
<td>0.74</td>
<td>1,632.8</td>
<td>16.5</td>
</tr>
</tbody>
</table>

The best model fits are underlined.

reporting rate parameters. The time series of estimated reporting rates is shown in Fig. 4. The estimated reporting rates are similar for the northern and southern B.C. stocks for the 1986 to 1996 period, but are higher for the southern stock in the earlier years. Under the variable reporting rate scenario, the estimated stock trajectories change somewhat with a lower 1997 exploitable biomass estimate for the southern stock and a higher terminal biomass estimate for the northern stock.

Discussion

The inability of the migration version of the integrated model to fit observed recoveries of migrant tags is of concern because significant numbers of tags are recovered outside the tagging area, suggesting migration is an important aspect of stock dynamics. The lack of model fit may be caused by a number of factors. The structure of the integrated model assumes that tag migration is a permanent event. That is, after tagged fish leave the population they do not return. There is some evidence suggesting that this assumption is not appropriate for sablefish in the eastern Pacific. That is, analysis of tag return data shows a tendency for smaller fish to move in a northwesterly direction and larger fish to move in a south and easterly direction (Heifetz and Fujioka 1991, Maloney and Heifetz 1997). This pattern suggests that there may be different phases to the ontogenetic dispersion of sablefish where the direction of migration is age-or size-dependent. If the sablefish migration pattern is more complex
Table 7. Parameter estimates for alternative runs of the non-migration model with the initial loss parameter fixed at various levels.

<table>
<thead>
<tr>
<th>Initial loss</th>
<th>Southern stock</th>
<th></th>
<th>Northern stock</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tag loss</td>
<td>Reporting rate</td>
<td>Function value</td>
<td>1997 biomass</td>
</tr>
<tr>
<td>0.05</td>
<td>0.29</td>
<td>1.00</td>
<td>1,140.6</td>
<td>29.6</td>
</tr>
<tr>
<td>0.15</td>
<td>0.27</td>
<td>1.00</td>
<td>1,149.9</td>
<td>26.6</td>
</tr>
<tr>
<td>0.25</td>
<td>0.26</td>
<td>1.00</td>
<td>1,161.7</td>
<td>23.5</td>
</tr>
<tr>
<td>0.35</td>
<td>0.23</td>
<td>1.00</td>
<td>1,177.1</td>
<td>20.3</td>
</tr>
</tbody>
</table>

than what has been modeled, for example, age- or sex-specific migration or seasonal migration, model parameter estimates will likely be biased.

Another model assumption that may account for the lack of fit of the migration implementation is that tag reporting rates are the same for the migrant and non-migrant fish. Beamish and McFarlane (1988) suggest that most tags recaptured in British Columbia are reported, and based on this Heifetz and Fujioka assume a tag reporting rate of 0.90 for British Columbia. For Alaska fisheries, reporting rate estimates range from 0.25 (Heifetz and Fujioka 1991) to 0.50 (Bracken 1983). Results from the analyses presented here suggest higher tag return rates for British Columbia than those estimated for Alaska. If the U.S. reporting rate is lower than the B.C. rate, the integrated model will underestimate migration rates and thus overestimate the numbers of tagged fish at large in both the migrant and non-migrant components of the population.

An extension to the single-stock analytical approach presented here, to a multi-stock model that explicitly allows movement between all stock components may provide more consistent fits to the tagging data. For the single stock model, the migration rate parameters are highly correlated with the fishing mortality rates for migrant fish. In a multi-stock analysis, the additional information from catch-age data for each stock may allow relatively independent estimates for these parameters. Alternately, treatment of all sablefish in the eastern Pacific as a single population, as suggested by Beamish and McFarlane (1988), may be the appropriate geographic scale for stock assessment, given the apparent high migration rates between stock units.

For the non-migration implementation of the integrated model, the parameters representing ongoing tag loss and reporting rates appear to be well determined, given the model structure and data observations. That is, although the parameters are correlated, their estimates are precise and robust to alternative model assumptions. The parameter representing the immediate loss of tags from the population is completely confounded
Figure 4. Estimates of time-dependent reporting rates and resulting exploitable biomass estimates for northern and southern B.C. stocks.
with reporting rates, so either one parameter must be fixed at an assumed value or a single parameter accounting for the joint effect can be estimated. The similarity between the estimates of the annual reporting rates for the northern and southern B.C. stocks suggests that there may be adequate information in the tag recovery data to obtain reasonable estimates for these parameters. The B.C. sablefish fleet fishes both of these assessment regions, so there is no reason to believe that reporting rates should be different in the two areas. If the model estimates of the reporting rates were driven by random noise rather than an underlying signal in the tag return data the high level of agreement in the estimates for the two stocks would not be expected.

The estimates for the ongoing tag loss parameter are unexpectedly high and suggest that there may be aspects of sablefish population dynamics that are not accounted for in the model. In mark-recapture analysis the parameter that I have termed “tag-loss” is generally attributed to tag shedding. However, the values of the tag loss parameter obtained in these analyses (0.26 – 0.27) are substantially higher than those estimated from double tagging experiments. Beamish and McFarlane (1988) estimated immediate tag shedding at 10% and ongoing tag shedding at 2% per year for sablefish tagged in British Columbia. Similar, low tag shedding rates were estimated by Lenarz and Shaw (1997) for sablefish tagged in the southern U.S. zone. Their estimates were 5% for immediate tag shedding and instantaneous ongoing shedding rates ranging from 0.03 to 0.07. Thus, it is likely that the tag loss parameter in the integrated model is accounting for the disappearance of tagged fish resulting from mechanisms in addition to tag shedding. Other potential sources of tag loss include ongoing tagging-induced mortality and migration to areas where there are no fisheries. Alternately, the model estimates of fishing mortality rates may be biased (i.e., underestimated).

References


Integrating Ecosystem Studies: A Bayesian Comparison of Hypotheses

Milo D. Adkison, Brenda Ballachey, James Bodkin, and Leslie Holland-Bartels
U.S. Geological Survey, Anchorage, Alaska

Abstract

Ecosystem studies are difficult to interpret because of the complexity and number of pathways that may affect a phenomenon of interest. It is not possible to study all aspects of a problem; thus subjective judgment is required to weigh what has been observed in the context of components that were not studied but may have been important. This subjective judgment is usually a poorly documented and ad hoc addendum to a statistical analysis of the data. We present a Bayesian methodology for documenting, quantifying, and incorporating these necessary subjective elements into an ecosystem study. The end product of this methodology is the probability of each of the competing hypotheses. As an example, this method is applied to an ecosystem study designed to discriminate among competing hypotheses for a low abundance of sea otters at a previously oiled site in Prince William Sound, Alaska.

Introduction

Ecosystem approaches are increasingly advocated as a way of improving the science and management of natural systems (Lackey 1998). For instance, studies of the effects of anthropogenic stressors on a species can be misleading if they ignore possible indirect effects acting through predator or prey populations (Higashi and Patten 1989). Further, natural changes in these other components of the ecosystem may cause changes in the focal population, masking or exaggerating the effects of the stressor (Piatt and Anderson 1996). Many studies of the impacts of human actions on a particular species now include research on other components of the ecosystem thought to be important to the focal species.
Nonetheless, there are practical limitations to an ecosystem approach. Because of cost and logistical constraints, not all ecosystem components can be studied and therefore some indirect impacts may be missed. Experimentation or replication may not be possible, and it may thus be difficult to unambiguously assign causes to any observed differences in populations between impacted and non-impacted sites, or before versus after an impact at a single site. It is also highly likely that among the suite of studies, some will give results that are to some degree contradictory. For these reasons, interpreting the results of an ecosystem study requires some degree of expert judgment. Synthesizing the results of numerous studies of parts of a complex problem is difficult, and it may thus be difficult for investigators to reach conclusions in a rational fashion. Further, different scientists faced with the same evidence may arrive at different conclusions. As the subjective interpretation of results tends to be an ad hoc and poorly documented process, the sources of disagreement may be difficult to uncover and resolve. This paper presents a structured method for documenting and quantifying the expert interpretation of the results of an ecosystem study.

Proposed Methodology
The methodology presented here is designed for testing ecosystem-level hypotheses. It integrates studies of diverse components of the ecosystem, summarizing the results as the relative evidence for each hypothesis from each study and the overall evidence for each hypothesis from the ensemble of studies. Its Bayesian features consist of incorporating and quantifying the subjective step of interpreting results, and calculating a probability that each hypothesis is true.

The method consists of the following steps:

1. Generate hypotheses
2. Summarize the experiments and their results
3. Create a table of the expected results under each hypothesis if each experiment were ideal
4. Calculate the probability of the observed result under each hypothesis using statistical considerations
5. Adjust probabilities by considering potential violations of statistical assumptions
6. Adjust probabilities to account for differences between the hypotheses tested and the hypotheses of interest
7. Summarize the evidence for each hypothesis, accounting for dependencies among experiments
Steps 3-6 deal with eliciting statements of probability from experts. Such elicitations can be problematic if experts are unfamiliar with translating their experiments into numerical probabilities (Morgan and Henrion 1990, Ch. 7). Our sequence of steps is designed to overcome such problems by sequentially considering several sources of uncertainty, progressing from the most to least familiar. At each of the seven steps, in particular those where subjective judgment is required, the rationale leading to the decision should be thoroughly documented.

**Step 1. Generate Hypotheses.** The first step is to have the experts identify the hypotheses that are the competing explanations for the phenomenon under investigation. It is important that the hypotheses be both exhaustive and mutually exclusive. If not, the confidence assigned to some hypotheses will be overstated, as the evidence for them will in some respects be counted twice.

Often, there will be reason to believe that several of the hypothesized phenomena might act simultaneously. There are two principal ways of constructing mutually exclusive hypotheses if this is a possibility. The first is to consider a “multiple causes” hypothesis. The second is to redefine the hypotheses to allow minor effects of other factors. For instance, the two hypotheses “effect is produced by factor A” and “effect is produced by factor B” can be made mutually exclusive by redefinition as “effect is principally produced by factor A” and “effect is principally produced by factor B.”

**Step 2. Summarize the Available Data.** In this step, the studies and their results are summarized. For clarity, it is often more useful to use a short verbal description of the results. For instance, a study of differences in prey abundance between control and treatment might be summarized as “much greater abundance found at the control site.”

**Step 3. Consider Ideal Studies.** The third step in this process is to lay out a table with the different hypotheses as the top row and the different experiments as the left-most column (Table 1). Then, have the experts fill out this table as if each study were an ideal experiment; i.e., there was no possibility of either false positive or false negative results.

<table>
<thead>
<tr>
<th>Hyp. 1</th>
<th>Hyp. 2</th>
<th>Hyp. 3</th>
<th>Hyp. 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study A</td>
<td>Positive</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>Study B</td>
<td>Negative</td>
<td>Negative</td>
<td>Positive</td>
</tr>
<tr>
<td>Study C</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
</tr>
</tbody>
</table>

*Table 1. Hypothetical results of a set of ideal experiments.*
Step 4. Statistical Considerations. While ideally the three studies would determine which hypothesis was true with 100% accuracy, in the real world misleading results may be obtained. One of the ways this may happen is through random sampling error. Often, almost any result is possible under any of the hypotheses. Nonetheless, the observed result will be more probable under some hypotheses than others.

The objective of this step is to calculate these relative probabilities, otherwise known as the likelihoods of each of the hypotheses (Gelman et al. 1995, Ch. 1). Often, with continuously distributed variables, the likelihood is a probability density rather than a probability per se. Likelihoods (Table 2) are usually obtained from standard statistical distributions such as the normal or binomial. The exact distribution used depends upon the assumptions made about the experimental data, such as whether each point is independent and identically distributed, whether the sampling variance is constant, etc.

Table 2 shows the first of a series of steps in which experts are asked to assign probabilities to the competing hypotheses. Some experts are unfamiliar with quantitative probability statements and scientists in particular are often uncomfortable making assertions about the relative merits of competing hypotheses without conclusive evidence. This step is important in that it introduces experts to assigning probabilities to the hypotheses, yet does so in a rigorous way using familiar statistical calculations.

Step 5. Account for Possible Biases in the Test or Experimental Results. The assumptions of statistical tests are rarely exactly met. Samples may not be completely independent, important sources of error may not be included in the statistical model (e.g., ignoring error in the measurement of the independent variable), and measurements may have some unknown biases. Historically, statistical confidence tends to overstate the certainty of scientific results (Henrion and Fischoff 1986).
In constructing the table of likelihoods of results, this overconfidence needs to be accounted for. Generally, the effect of such errors is to make the probabilities of the result under each hypothesis more similar. Based on their knowledge of the experiment, experts should determine which assumptions of the test are likely to be violated, and to what degree. These judgments are to some extent subjective, but once made the statistical literature or computer simulations can provide guidance on their likely effects. In consultation with a statistician, the experts should adjust the table of probabilities to account for such violations.

**Step 6. Account for Differences Between the Statistical Hypothesis Being Tested and the Biological Hypothesis That Is Actually of Interest.** Often, an experiment to test a hypothesis tests it only indirectly. The results may thus be ambiguous if the indirect indicator could occur in several ways, some of which are not related to the hypothesis.

For example, if the hypothesis were that some population was affected by an environmental contaminant, an investigator might test the environment for the presence of the contaminant and test individuals for signs of poor health. A positive result in either case would not necessarily implicate the contaminant; the contaminant might be present yet not be causing health effects, or poor health might be due to causes other than the contaminant.

As in step 5, the effect of a difference between the hypothesis tested and the hypothesis of interest is to even further equalize the probabilities of the observed results under each hypothesis. The appropriate amount of adjustment of the table entries depends on the probability of other (possibly unknown) alternative explanations for the test results. Such assessments are unavoidably subjective and require the judgment of experts. Hopefully, by this point in the process the experts are comfortable with assessing the relative probability of the data under each hypothesis and how violations of assumptions may result in misleading experimental results. It is crucial that they consider alternative explanations for their data yet not be paralyzed by such possibilities. They should be willing to examine data that seems to strongly favor one hypothesis and consider whether there are other, possibly unstudied ecosystem pathways that could produce similar results and state how probable they feel such pathways are.

**Step 7. Summarize the Evidence.** In this step, the table of probabilities is summarized to derive the overall weight of evidence for each hypothesis provided by the ensemble of studies. If the studies are independent, then elementary statistical theory says the joint likelihood of each hypothesis is simply the multiplication of its probability under each study (equation 1). The overall likelihood of each hypothesis is then simply the product of its column of probabilities (here \(R_1\), \(R_2\), and \(R_3\) signify the results of experiments 1, 2, and 3, respectively).
Likelihood of hypothesis = \( P(R_1|\text{hyp.}) \times P(R_2|\text{hyp.}) \times P(R_3|\text{hyp.}) \)  

(1)

The different hypotheses can then be compared in terms of their relative likelihoods. This comparison is easier if the likelihoods are re-scaled so that the sum of all of the likelihoods is 1. From a Bayesian perspective, each re-scaled likelihood could then be interpreted as the probability that a hypothesis was true.

**Complication A. Dependencies among Results.** There are two ways that experimental results might not be independent. First, the data from two experiments may have been taken from the same random sample. Second, two experiments may measure the same ecological phenomenon two different ways. In either case, it is not appropriate to treat the results as providing independent evidence bearing on the alternative hypotheses; i.e., simply multiplying the probabilities of the two experiments together will overweight the evidence.

There are several possible methods to account for dependencies among experimental results. If experiments are highly interdependent, they should be lumped and a single probability of each hypothesis calculated for the ensemble results. If experiments are only partially dependent, the correlation of results must be accounted for. If the correlation can be calculated, probability theory provides methods for calculating a joint probability. If not, a value must be obtained from experts, although experts have been found to perform poorly at providing a numerical value for correlation coefficients (Morgan and Henrion 1990, Ch. 7).

A more intuitive method for dealing with partially correlated results is to ask investigators to provide an estimate of the “effective” number of experiments. For instance, investigators may feel that dependence between two experiments is such that they jointly provide only as much evidence as 1.5 independent experiments. Then, the appropriate adjustment would be to raise each of the probabilities to the 0.75 power (e.g., equation 2). In general, if \( N \) experiments are correlated so that the effective number is \( E \), probabilities for hypotheses for each experiment should be adjusted by raising them to the \( E/N \) power.

\[
\text{Likelihood of hypothesis} = P(R_1|\text{hyp.})^{0.75} \times P(R_2|\text{hyp.})^{0.75}
\]

(2)

**Complication B. Prior Probabilities.** Bayesian statistics involves multiplying the likelihoods by a set of prior weights (the prior probabilities) for the hypotheses before re-scaling to calculate the posterior probabilities. In the Bayesian approach, these prior probabilities reflect the weight accorded each hypothesis before the experiments were conducted. Assuming the probability of each hypothesis to be proportional to the joint likelihoods treats each hypothesis as being equally likely a priori, thus letting the data determine the relative probability of each hypothesis. While this is intuitively appealing, it may not be appropriate.
For instance, if the analysis were being used in a legal proceeding, it might be appropriate to give the benefit of the doubt to the defendant by assigning small prior weights to hypotheses implicating the defendant. Similarly, in investigating current scientific theory a high prior weight might be assigned to the currently accepted paradigm, so that a novel competing theory would not get much credence unless the evidence for it was overwhelming. An alternative to using prior weights is to calculate probabilities only from likelihoods, but require a very high probability that a hypothesis is true before acting on it. Whatever the prior weights, if data strongly support one hypothesis over the others the final probabilities will reflect this.

Standard Bayesian practice is to compare the evidence for competing hypotheses using Bayes factors (Kass and Raftery 1995). The Bayes factor is simply the ratio of the posterior probabilities of two competing hypotheses divided by the ratio of the prior probabilities assigned before the experiments were conducted. When the prior probabilities of the hypotheses are equal, this is simply the ratio of the posterior probabilities.

**An Example: Sea Otters after the Exxon Valdez Oil Spill**

On March 4, 1989, the supertanker Exxon Valdez spilled nearly 42 million liters of crude oil in Prince William Sound, Alaska (Spies et al. 1996). This spill is hereafter referred to with the acronym EVOS. Sea otter populations in oiled areas suffered high mortality (Loughlin et al. 1996). Other components of the ecosystem were likewise severely affected. Five years after the spill, residual oil was present in sediments and mussel beds in some areas of the spill (Spies et al. 1996). Even today, residual oil is found in some areas.

The Nearshore Vertebrate Predator (NVP) project (Holland-Bartels et al. 1996), a multi-university and agency investigation funded by the EVOS Trustee Council, is aimed at determining whether top predators in Prince William Sound are still suffering the effects of the oil spill. The question is difficult to answer unambiguously because of the complicated nature of the ecosystem and the lack of data from the period before EVOS. The NVP project studies predator populations from several points of view, and also looks at other components of the ecosystem on which these predators depend. If a population is still being affected by EVOS, the study is designed to ascertain whether the effects are due to the continuing toxic effects of oil, a slow rate of recovery from past mortality, or an indirect effect on some critical ecosystem component.

With limited resources and such an intensive approach, few populations can be studied. Sea otter abundance at Knight Island, which was oiled in 1989, is lower than at Montague Island, which was not. The NVP
sea otter study has focused on these two populations, trying to find the reason for these differences in abundance. The principal hypotheses are:

1. **Direct toxicity of residual oil.** Residual oil is present and reducing the fecundity and/or survival of otters at the oiled site.

2. **Reduced forage due to oil effects.** The initial impact of oil or residual oil is reducing prey available to sea otters.

3. **Slow recovery due to demographic limitations.** Aside from the initial otter mortality from EVOS, residual oil is absent or does not affect otters or their food. However, limitations on the maximum growth rate of the population have prevented the population from reaching capacity yet.

4. **Natural differences in capacity.** The oiled site has poorer or less abundant otter habitat.

A variety of studies have been undertaken to determine which hypothesis is the most likely. These include:

1. **Demographic comparisons.** Population abundance, age structure, and reproductive rates were compared between islands.

2. **Individual health.** Otters were captured at both locations. Individuals were weighed and measured, and blood samples taken. In particular, blood cells and serum chemistry were examined for signals of poor health, and a specific signal of exposure to oil (the enzyme P450) was tested for.

3. **Prey abundance and foraging success.** The abundance and size distribution of major prey items of sea otters were compared among islands. In addition, foraging sea otters were observed to determine relative rates of success in obtaining prey items.

Statistical hypothesis tests were performed for many of the studies but are not reported here. We chose not to calculate likelihoods based solely on statistical distributions—step 4 of our methodology—because the limitations imposed by the design of the study tended to emphasize the considerations dealt with in steps 5 and 6. There are multiple predictions from each of the hypotheses, not all of which are distinct. Any particular study result may eliminate some hypotheses but leave several others. More likely, any particular study result would be ambiguous, as there is a small likelihood of almost any result from each hypothesis. In particular, the detection of a phenomenon does not necessarily imply that this was the cause of the difference in abundance between the two islands. For instance, oil could be present but yet not greatly affect survival. Likewise, prey abundance could differ between one site and another but be unrelated to the difference in otter abundance.
Thus, the interpretation of the results of the studies required some 
judgment. Our chief tool was to ask ourselves, “What is the probability 
we would get the result we observed from Study ___ if Hypothesis ___ was 
true?” We attempted to quantify our impression of the strength of each 
piece of evidence by filling out the table of probabilities, sequentially 
considering what the result would mean in an ideal world, what the sta-

tistical tests implied, how the assumptions of the tests might be violated,
and what mechanisms might cause the results to be misleading.

We felt our ability to discriminate among probability levels was fairly 
coarse. Accordingly, we initially filled in the table of probabilities ver-

bally, using the categories “high,” “moderate-high,” “moderate,” “low-

moderate,” and “low,” which we later replaced with 0.9, 0.7, 0.5, 0.3, and 0.1, 
respectively (Table 3).

The result of our first analysis was to assign more than a 98% proba-
bility to the hypothesis that the population differences were due to a de-

mographic limitation in the rate of recovery of the Knight Island popula-
tion from spill mortality. All other hypotheses combined had less than a 1.5%
probability of being true. We were unhappy with this result, as this high degree of confidence did not reflect our personal higher degree of uncertainty. We felt that the evidence for this hypothesis was not that strong.

In examining the reasons for this initial result, we identified three principal sources of error. First, we overstated the power of the studies to discriminate among hypotheses. For instance, we assigned a 0.90 probability of seeing greater prey abundance at the oiled site if demography was limiting recovery, but only a probability of 0.10 under any of the other hypotheses. We did not adequately address step 6 of our methodology; for instance, there would be a fairly good chance of seeing higher prey abundance at the oiled site under several alternative hypotheses.

Second, the range of hypotheses we considered was too narrow. In retrospect, we felt there was a strong possibility that all of the hypotheses might be incorrect, and some other factor might be responsible for differences between areas. This resulted in an unrealistically high probability for the hypothesis most consistent with the data.

Third, we did not adequately account for dependencies among experimental results (step 7, complication A). While we lumped most blood chemistry measures into one result, we kept the assay for the enzyme P450 (a more direct measure of exposure to oil) as a separate experiment. Since this assay could indicate the same phenomenon, and was measured on the same sample of animals, we felt the two results were effectively equivalent to only 1.5 experiments. Similarly, measures of prey size, prey abundance, and foraging success to some extent measured the same phenomenon. In retrospect, we decided to consider them as equivalent to two experiments.

We therefore revised the tabled probabilities, taking what we hoped was a more realistic look at the power of the studies and adding another alternative hypothesis to those we had listed. While we were able to think of several specific alternatives, we felt the true explanation for population differences might be something we hadn’t considered. Therefore, we added only one hypothesis; an “unknown causes” category. Meanwhile, the completion of analyses of blood chemistry and the enzyme P450 suggested that residual oil might be present at the oiled site, and new information became available about the size distribution of prey species (Table 4).

The revised table again supports the hypothesis that the populations differ because the population in the oiled area has not had the time to recover fully from the losses due to the oil spill. However, it shows even greater support for the hypothesis that residual oil is still affecting the population. The hypothesis that some unknown factor accounts for the difference between populations is also quite probable.

Two hypotheses were eliminated from consideration, principally because of the forage abundance studies. Forage was more abundant and foraging success higher at the oiled site. These results were not at all
consistent with the food limitation hypothesis, and were also unlikely if the population at the oiled site had recovered to its carrying capacity. However, it should be noted that the “unknown causes” hypothesis, which has a fairly high probability of being true, is not necessarily related to the spill. Thus it would be inappropriate to say the probability that the population is no longer suffering effects of the spill is only 0.01.

We will refine and expand this analysis as more data become available and more experts are consulted. These results are not our final interpretation, and should be viewed as a preliminary analysis. We provided this example solely to illustrate the use of the methodology.

Table 4. Second attempt at integrating studies.

<table>
<thead>
<tr>
<th>Experiment and (result)</th>
<th>&quot;A&quot; Demogr. limit</th>
<th>&quot;B&quot; Food limit</th>
<th>&quot;C&quot; Oil persist</th>
<th>&quot;D&quot; Recovered</th>
<th>&quot;E&quot; Unknown causes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otter density (K &lt;&lt; M)</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.3</td>
<td>0.9</td>
</tr>
<tr>
<td>Repro rates (equal)</td>
<td>0.9</td>
<td>0.5</td>
<td>0.7</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Blood CBCs &amp; chemistry (weak indication of liver damage at K)</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>P450 (M &lt; K)</td>
<td>0.3</td>
<td>0.3</td>
<td>0.9</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Prey abundance (M &lt; K)</td>
<td>0.9</td>
<td>0.1</td>
<td>0.5</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Prey size (M &lt; K)</td>
<td>0.9</td>
<td>0.1</td>
<td>0.7</td>
<td>0.3</td>
<td>0.7</td>
</tr>
<tr>
<td>Foraging success (M &lt; K)</td>
<td>0.9</td>
<td>0.1</td>
<td>0.7</td>
<td>0.3</td>
<td>0.7</td>
</tr>
<tr>
<td>Joint likelihood</td>
<td>0.1581</td>
<td>0.0011</td>
<td>0.1744</td>
<td>0.0040</td>
<td>0.0764</td>
</tr>
<tr>
<td>Probability of hypotheses</td>
<td>38.2%</td>
<td>0.3%</td>
<td>42.1%</td>
<td>1.0%</td>
<td>18.5%</td>
</tr>
</tbody>
</table>

Top row gives hypotheses, and left column gives experiments with the results in parentheses. “M” refers to Montague Island (control), and “K” to Knight Island (oiled). The main body of the table gives the probability of obtaining each experimental result under each hypothesis. The bottom two rows summarize the result as the product of the probabilities for each hypothesis (i.e. the joint likelihood) and the probability products re-scaled to sum to 100%.
Discussion

The Bayesian aspects of the proposed methodology are (1) use of subjective expert judgment in interpreting indirect tests of hypotheses, and (2) integration of experimental results and expert judgment into an overall probability for each hypothesis using Bayesian probability calculations. A large literature exists on using Bayesian methods to compare hypotheses (Kass and Raftery 1995).

Bayesian methods have been criticized from a variety of standpoints (e.g., Dennis 1996). The principal criticism is that Bayesian methods inject subjectivity into scientific analyses that should be objective. However, in extrapolating from the results of diverse studies on small aspects of a larger question, subjectivity in the form of expert judgment is unavoidable. We propose a methodology that formalizes the intuitive process experts use in interpreting the results of ecosystem studies. This approach clearly distinguishes subjective interpretation from experimental results, and clearly shows the reasoning used.

Our methodology provides a tool for investigators to organize their thinking. The ecosystem and the results of the numerous studies may be too complex to be readily grasped in their entirety. By allowing investigators to approach the synthesis of the studies one element at a time, our method increases the tractability of the process.

The methodology also facilitates openness and discussion, since subjective components of the synthesis of the studies are documented and quantified. It clearly shows why a particular conclusion was reached, and what evidence investigators felt was ambiguous or particularly strong. Areas of disagreement among investigators are also easily identified.

Our methodology is based on principles derived from other methods widely used for eliciting probabilities from experts (summarized in Morgan and Henrion 1990, Ch. 7). Examples of such methods include the Stanford/SRI protocol (Spetlzer and Stael von Holstein 1975, Merkhofer 1987) and the Wallsten/EPA protocol (Wallsten and Whitfield 1986). We've tailored our methodology to the specific goal of summarizing the relative support for alternative hypotheses from an interrelated but necessarily incomplete set of studies.

Most methods for probability elicitation pay great attention to getting experts comfortable with the idea of translating their knowledge and judgment into probability statements, and to overcoming a tendency of experts to give probabilities that overstate the level of certainty (Tversky and Kahneman 1982; Morgan and Henrion 1990, Ch. 7). Our solution to these difficulties is to take experts through a specific sequence of probability elicitation steps. These start with specifying deterministic outcomes, then progress through familiar specifications of probability (likelihood calculations) to less familiar probability specifications (the effects of violation of statistical assumptions and of not directly testing the hypothesis of interest). This sequence gradually introduces the process of making
probability statements. It also sequentially introduces more and more forms of uncertainty, continually forcing the expert to reflect on whether the degree of confidence he's previously expressed is appropriate.

Our example illustrates both the utility and limitations of the methodology. The summary table lists the hypotheses and the experimental results. Probabilities within the table explicitly document the experts' interpretation of the consistency of the results of each experiment with each hypothesis. The summary probabilities excluded two hypotheses but retained three others, one of which appears to be only half as probable as the other two.

However, the 18.5% probability assigned to the “Unknown Causes” hypothesis makes interpretation of the other probabilities somewhat ambiguous. Much of the probability assigned to this hypothesis may indicate that recovery has occurred, and the differences we found are caused by some unknown factor(s) unrelated to the spill. It is also possible that “unknown causes” represents effects related to the spill such as cascading ecological effects. In either case, the results do provide guidance for further research; they suggest that continuing studies should focus on hypotheses “A,” “C,” and “E.”

The necessity for re-evaluating our initial analysis because of unrealistic results is instructive. It reinforces the experience of others who have found that numerical statements of probability given by experts tend to be overly confident (Tversky and Kahneman 1982, Henrion and Fischhoff 1986). Our second try produced a result that we felt better reflected the strength of the evidence provided by the experiments.

There is a danger that allowing such reanalysis could result in investigators juggling numbers to arrive at a result that reflected their preconceptions. However, an honest reappraisal of each element in the table is not inappropriate. Most methods for probability elicitation do recommend that assessors return to an earlier phase in the process whenever questioning reveals that the probabilities elicited clearly don't reflect the expert's judgment (Kadane et al. 1980; Morgan and Henrion 1990, Ch. 7; Laskey 1995). We found the reanalysis of the table caused us to re-examine the basis of our interpretations; rather than reinforcing our preconceptions, it tended to make us change them.

Use of our methodology will make it easier to examine the source of differences in interpretation of a study. For example, a scientist who disagreed with our conclusions might find that the basis of his difference was the weight placed on the blood chemistry results. A sensitivity analysis to alternative interpretations would be easy to perform by replacing the disputed probability with an alternative value to see if this affected the conclusions.

This method is not proposed as a substitute for good experimentation. With scarce, poor quality, and ambiguous data the conclusion reached after applying this method will be that considerable uncertainty remains. However, in such situations this methodology may identify areas of major
uncertainty and suggest fruitful lines of investigation. The major benefit of this approach is the explicit documentation and quantification of the unavoidable subjective interpretation of ambiguous results that arise in many ecosystem investigations. In contrast, when strong experimental designs are available that produce clear evidence, subjective interpretation will be minimized and investigators should reach consensus.

Acknowledgments
The authors wish to thank Tom Dean, Jennifer DeGroot, George Esslinger, Steve Jewett, Dan Monson, Chuck O’Clair, Alan Rebar, Paul Snyder, and Glenn VanBlaricom for their contributions. The EVOS Trustee Council provided financial support for this study.

References


Assessment of Southeast Alaska Pink Salmon Abundance Based on Commercial Catch and Effort and Sex Ratio Data

Jie Zheng
Alaska Department of Fish and Game, Juneau, Alaska

Ole A. Mathisen
University of Alaska Fairbanks, School of Fisheries and Ocean Sciences, Juneau, Alaska

Abstract
Pink salmon (Oncorhynchus gorbuscha) is the most abundant salmon species in Southeast Alaska and supports an important commercial fishery. Like most salmon fisheries in Alaska, pink salmon fisheries in Southeast Alaska are managed by a fixed escapement policy. To achieve a targeted escapement, managers must know the abundance of the incoming spawning run. The accuracy of abundance information acquired inseason substantially affects the manager's ability to achieve management objectives. To improve accuracy of inseason forecasts of southern Southeast Alaska pink salmon runs, we incorporated sex ratio information into inseason forecast models to annually adjust timing and shape of the run timing curves. First, we developed a sex ratio index and subsequently evaluated three inseason forecast models—linear, nonlinear, and combined—using this index and cumulative catch of all gears or cumulative catch per unit effort of the seine fishery from 1983 to 1997. Based on a cross-validation evaluation of forecast accuracy, the nonlinear model outperformed the linear and combined models. Cumulative catch per unit effort was a better predictor than cumulative catch in the first three weeks (weeks 28-30) of a fishing season, and vice versa in the remaining five weeks. Inseason abundance estimations greatly improved the preseason forecasts. Incorporating sex ratios into inseason forecast models correctly adjusted the run timings during a large majority of years and thus improved overall forecasts starting in the second week. In weeks 29-32, the best performing
model using sex ratios improved forecasts more than 30% over the best model without using sex ratios; improvements included averages of relative forecast errors, absolute deviations, or squared residuals. Averages of relative forecast errors in weeks 29-34 were less than 24% for the best performing model using sex ratios and less than 38% for the best model without using sex ratios, compared to 51% for preseason forecasts. Average relative forecast errors from the best model were less than 20% before the run midpoint and less than 14% after the run midpoint.