LIBRARY Secretariat of the Pacific Community

Preliminary analyses of South Pacific albacore catch, effort and lengthfrequency data, using an age-structured, length-based model with spatial structure

David Fournier

Otter Research Ltd P.O. Box 265, Station A Nanaimo, B.C. V9R 5K9, Canada

and

John Hampton

Oceanic Fisheries Programme South Pacific Commission B.P. D5, Noumea, New Caledonia



ICCAT 25th ANNIVERSARY TUNA SYMPOSIUM 10-18 June 1996, Ponta Delgada, Sao Miguel Island, Azores, Portugal

1. INTRODUCTION

Assessment of fish stocks using commercial fisheries data is usually carried out using one of two basic classes of models – biomass dynamics models and age-structured models (Hilborn and Walters 1992). Of the latter, virtual population (or cohort) analysis (see review in Megrey 1989) has been the technique most frequently applied to fish stocks.

While computationally straight-forward, VPA requires certain assumptions that are difficult to test and/or justify in many situations. Catch-at-age data are required and are usually assumed to be correct (unless aging errors are incorporated using Monte-Carlo simulations). The natural mortality rate and the fishing mortality rate for one age class of each cohort must be assumed. "Tuning" the analysis with catch per unit effort (CPUE) data or other abundance index time series is often employed to avoid the need to specify "terminal" fishing mortality rates for each cohort, but this invariably involves arbitrary assumptions about catchability. Additionally, VPA involves the computation of one fishing mortality rate for each non-terminal catch-at-age observation and the initial abundance for each cohort. The model is thus fully saturated (no degrees of freedom) and, apart from the statistical errors associated with the tuning procedure, there is no notion of statistical uncertainty in the results.

Statistical catch-at-age models (e.g. Doubleday 1976; Paloheimo 1980; Fournier and Archibald 1982; Pope and Shepherd 1982; Dupont 1983; Deriso et al. 1985) can potentially avoid some of these assumptions. While catch-at-age data are still required, age- or time-related constraints on fishing mortality enable a statistical estimation of initial cohort sizes, fishing mortality rates or related parameters and, in theory, natural mortality rate, by minimizing an objective function based on a statistical criterion such as least squares. Variance estimates, and therefore confidence intervals, for the estimated parameters conditional on the catch-at-age data and the model can also be obtained.

Both VPA and statistical catch-at-age models rely on catch-at-age data typically derived from the analysis of annuli on various body parts of individual fish. These methods are often inappropriate or too expensive for routine application, particularly to large-scale tuna fisheries. For many fisheries, catch-at-length data may provide a convenient and less expensive alternative for analysis by age-structured models.

Most catch-at-age and catch-at-length models consider a spatially-aggregated population and fisheries. However, for many fish stocks, population parameters may not be spatially homogeneous. In such cases, assuming that fisheries which operate in different portions of the stock range exploit a common population may lead to biased results. To avoid such problems, spatial structure can be incorporated into the model.

In this paper, we describe a length-based, age-structured, likelihood model that circumvents many of the difficulties associated with sequential analyses such as VPA. The model incorporates the following features:

• Growth and age structure of the catch are estimated simultaneously with population parameters such as recruitment, selectivity, catchability and natural mortality. Approximate confidence intervals are therefore conditional not on catch-at-age, but on catch-at-length data.

- One-dimensional spatial structure is included in the model, with the distribution of fish among regions (latitudinal bands) controlled by hypotheses concerning the location of recruitment and fish movement.
- Missing data and data of different temporal resolutions are allowable and are internally managed by the model.
- Auxiliary data (such as tagging data) can be incorporated into the model, as appropriate.
- Various structural hypotheses, such as density-dependent growth, time-series trends in catchability and seasonal catchability, can be incorporated into the model and tested.

The model is applied to catch, effort and length frequency data for South Pacific albacore, *Thunnus alalunga*. These data were collected from several different sources (fisheries), have inconsistent temporal resolution and include periods of missing effort and/or length frequency data.

2. BACKGROUND INFORMATION ON SOUTH PACIFIC ALBACORE

Albacore are thought to comprise a discrete stock in the Pacific Ocean south of the equator (Lewis 1990). Adults (larger than about 80 cm FL) spawn in tropical and sub-tropical waters between about 10°S and 25°S during the austral summer (Ramon and Bailey 1994), with juveniles recruiting to surface fisheries in New Zealand coastal waters and in the vicinity of the sub-tropical convergence zone (STCZ) in the central Pacific 1–2 years later. The distribution of South Pacific albacore is summarized in Figure 1.

Longline fleets of distant-water fishing nations (DWFN) (mainly Japan, Korea and Taiwan) and several Pacific Island countries catch primarily adult albacore virtually throughout this range. Also, a troll fishery for juvenile albacore has occurred in New Zealand coastal waters since the 1960s. In the mid-1980s, a fleet of troll vessels from the United States began fishing in the central Pacific in the region of the STCZ, and were soon joined by vessels from New Zealand, Fiji and French Polynesia. At about the time the STCZ troll fishery began, driftnet vessels from Japan also entered the fishery, fishing in the central Tasman Sea and in the central Pacific in the STCZ. Taiwanese driftnetters joined the fishery in the late 1980s. These surface fisheries are highly seasonal, occurring mainly during December to April. The distribution of the fisheries is shown in Figure 2.

The sizes of albacore captured by these fisheries increases from south to north, with the longline fishery typically capturing the largest fish and the troll and driftnet fisheries the smallest. (Typical length frequency distributions are shown later in Figure 10.)

Annual catches have varied between 20,000 t and 52,000 t since the 1960s (Figure 3). The longline fishery accounts for the majority of the catch, about 30,000 t per year on average. The troll fisheries are relatively small, generally producing less than 10,000 t per year. The driftnet catch reached 22,000 t in 1989, but has since declined to zero following a United Nations moratorium on industrial-scale driftnetting.

3. MODEL DESCRIPTION

3.1 Data structures

The fundamental data structure of the model is based on the notion of a fishery, which is thought of as a collection of fishing units which operate in a particular region defined in the model, and which have similar catchability and selectivity characteristics. For the South Pacific albacore analysis, we have defined three latitudinal bands as regions: $0-10^{\circ}$ S (Region A), 10° - 30° S (Region B) and 30° - 50° S (Region C). The DWFN longline fleets fishing in these regions are defined as separate fisheries by region, but are aggregated across nationality. The fleets of small-scale, domestic longliners (using monofilament gear) that have developed in several Pacific Island countries in recent years are also defined as a fishery. This fishery operates in Region B. The troll fishery in New Zealand coastal waters, that operating in the STCZ and the driftnet fishery (all Region C) are classified as different fisheries. The definition of fisheries in the present analysis is therefore:

Fishery 1 (Region A): DWFN longline, 0-10°S Fishery 2 (Region B): DWFN longline, 10-30°S Fishery 3 (Region C): DWFN longline, south of 30°S Fishery 4 (Region B): Domestic longline Fishery 5 (Region C): New Zealand troll Fishery 6 (Region C): STCZ troll Fishery 7 (Region C): Driftnet

The spatial configuration of the fisheries is shown in Figure 4.

Each occurrence of a fishery at a particular time is termed a fishing incident. In reality, fishing is more or less continuous, so the data for each fishery need to be aggregated over appropriate time intervals. For the longline fisheries (1-4), which occur more or less continuously throughout the year, quarterly time periods are sufficient to capture the seasonal variation. The surface fisheries (5-7) tend to operate during the summer months only, therefore monthly time periods are used for these fisheries. The history of effort and CPUE for each of these fisheries is shown in Figure 5.

3.2 The catch equations

It is assumed for simplicity of notation in this description that there is only one fishery operating in each region and that there is only one fishing incident per fishery per year. The model is designed to accommodate a variable number of different fisheries per region and fishing incidents per fishery per year; the equations that follow could easily be generalized in this way.

The catch equations relate the numbers of fish in the population to the numbers of fish in the catch of the fisheries. The form of the catch equations used in the model is described by the following relationships:

$$C_{ijk} = \frac{F_{ijk}}{Z_{ijk}} \left[1 - \exp\left(-Z_{ijk}\right) \right] N_{ijk} \quad \text{for} \quad 1 \le i \le n, \quad 1 \le j \le a, \quad 1 \le k \le r$$
(1)

$$T_{i+1,j+1,k} = \exp(-Z_{ijk}) N_{ijk} \quad \text{for} \quad 1 \le i \le n, \quad 1 \le j < a, \quad 1 \le k \le r$$
(2)

$$T_{i+1,ak} = \exp(-Z_{i,a-1,k}) N_{i,a-1,k} + \exp(-Z_{iak}) N_{iak} \quad \text{for} \quad 1 \le i < n, \quad 1 \le k \le r$$
(3)

$$T_{i1k} = \gamma_k R_i$$
 for $1 \le i < n$, $1 \le k \le r$ where $\sum_k \gamma_k = 1$ and $\gamma_k \ge 0$ (4)

$$N_{ijk} = \sum_{l} \beta_{jkl} T_{ijl} \quad \text{for} \quad 1 \le i \le n, \quad 1 \le j \le a, \quad 1 \le k \le r, \quad 1 \le l \le r$$
(5)

$$Z_{ijk} = F_{ijk} + M_{ijk} \quad \text{for} \quad 1 \le i \le n, \quad 1 \le j \le a, \quad 1 \le k \le r$$
(6)

$$C_{i\cdot k} = \sum_{j} C_{ijk} \quad \text{for} \quad 1 \le i \le n, \quad 1 \le j \le a, \quad 1 \le k \le r$$
(7)

where

- n is the number of years of fishing,
- *a* is the number of age classes in the population,
- r is the number of regions,
- C_{ijk} is the catch (in number of fish) of age class j fish in region k in year i,
- C_{ik} is the total catch observed in region k in year i,
- F_{ijk} is the instantaneous fishing mortality rate for age class j fish in region k in year i,
- M_{ijk} is the instantaneous natural mortality rate for age class j fish in region k in year i,
- Z_{iik} is the instantaneous total mortality rate for age class j fish in region k in year i,
- T_{ijk} is the number of age class j fish in the population in region k at the beginning of year i before movement has taken place,
- N_{ijk} is the number of age class j fish in the population in region k at the beginning of year i after movement has taken place,
 - R_i is the recruitment at the beginning of year *i*,

 γ_k is the proportion of recruitment occurring in region k, and

 β_{jkl} is a k by k diffusion matrix **B**_j for age class j fish.

For each *j*, the elements of \mathbf{B}_j must be specified. In the case of South Pacific albacore, we use a one-dimensional diffusion model operating in three regions (k=3). In this case, the elements of \mathbf{B}_i are given by

$$\begin{bmatrix} 1+\theta_j & -d_2\theta_j & 0\\ -\theta_j & 1+2d_2\theta_j & -d_3\theta_j\\ 0 & -d_2\theta_j & 1+d_3\theta_j \end{bmatrix}^{-1}$$
(8)

where 1, d_2 and d_3 ($d_2 > 0$ and $d_3 > 0$) specify the relative distribution of cohort abundance among regions at equilibrium and θ_j is the age-dependent diffusion rate. We employ a flexible parameterization of θ_j which can result in increasing or decreasing diffusion rate with increasing age:

$$\theta_{j} = \phi_{0} \exp\left\{\phi_{1}\left[-\left(-\kappa_{j}\right)^{\phi_{3}}\right]\right\} \text{ where } \phi_{0} \ge 0, \ \phi_{1} \ge 0 \text{ and } \kappa_{j} < 0$$

$$\theta_{j} = \phi_{0} \exp\left\{\phi_{1}\kappa_{j}^{\phi_{3}}\right\} \text{ where } \phi_{0} \ge 0, \ \phi_{1} \ge 0 \text{ and } \kappa_{j} \ge 0$$
(9)

where $\kappa_{j} = \frac{2(j-1)}{a-1} - 1$.

Note that in this form of the catch equations, the last (a th) age class consists of all the older fish in the population. This is useful when, as often occurs, the aging estimates are especially inaccurate for the older age classes (Fournier and Archibald 1982). For catch-at-length data it is useful to group the older age classes together after the fish reach an age where they essentially stop growing (Fournier et al. 1991).

3.3 Constraints on natural and fishing mortality rates

Natural mortality

To reduce the number of free parameters in the model, it is often assumed that the instantaneous natural mortality rate is a constant, independent of the year, age and region. We shall make this assumption and denote the instantaneous natural mortality simply by M. However, it should be noted that other (possibly more realistic) models of M variability could be posed and tested within the framework of the model.

and the second second second second

Fishing mortality

To further reduce the freedom of the parameters, we restrict the variation in the instantaneous fishing mortality rates F_{ijk} . Consider for simplicity an individual fishery (i.e. drop the k subscript). We have assumed that

$$\log_{e}(F_{ij}) = \log_{e}(s_{ij}) + \log_{e}(q_{ij}) + \log_{e}(E_{ij}) + \varepsilon_{ij}$$
(10)

and

$$\log_{\bullet}(q_{i+1}) = \log_{\bullet}(q_i) + \eta_i \tag{11}$$

where

- s_j is the selectivity for age class j (assumed constant over time),
- q_i is the catchability in year *i*,
- E_i is the observed fishing effort in year *i*,
- ε_i are normally distributed random variables representing large transient deviations in the effort-fishing mortality relationship, and
- η_i are normally distributed random variables representing small permanent changes in catchability.

The notion, as implied in equation (10), that fishing mortality consists of a "separable" agedependent effect (selectivity) and a time-dependent effect (catchability) was first introduced by Doubleday (1976) and later elaborated upon by Paloheimo (1980) and Fournier and Archibald (1982). Details of the treatment of selectivity and catchability in the model are as follows:

Selectivity

It is sometimes possible to model selectivity as a function of age class, for example using a gamma function (Deriso et al. 1985). We have preferred to allow the s_j to be separate parameters but have applied a transformation that essentially makes selectivity a length-based rather than age-based concept. The transformation is as follows:

$$s_{j} = \sum_{k=-2}^{2} \omega_{k} \left\{ t \Big[\psi_{1} \Big(\mu_{j} + k\sigma_{j} \Big) \Big] + \psi_{2} \Big(\mu_{j} + k\sigma_{j} \Big) \Big(t \Big[\psi_{1} \Big(\mu_{j} + k\sigma_{j} \Big) + 1 \Big] - t \Big[\psi_{1} \Big(\mu_{j} + k\sigma_{j} \Big) \Big] \right\}$$
(12)

where

- ω_k are weights determined from the normal distribution of length at age k standard deviations from the mean,
- ψ_1 is the integer part of the age class number corresponding to length $\mu_j + k\sigma_j$,
- ψ_2 is the fractional part of the age class number corresponding to length $\mu_j + k\sigma_j$,

- μ_j is the mean length of age class j fish,
- σ_i is the standard deviation of length of age class j fish, and
 - t is an estimated parameter.

This transformation effectively ensures relatively small differences in s_j between adjacent age classes having large overlap of their length distributions, as would be expected where selectivity is fundamentally length-based.

Catchability

Catchability is allowed to vary slowly over time. We assume that the q_i have the simple time series structure of a random walk (equation 10), which is the simplest statistical model of a slowly varying random quantity. The assumption that catchability has a time series structure was introduced by Gudmundsson (1994) for the analysis of catch-at-age data. Gudmundsson also included trend components in his time series formulation.

We make the prior assumption that the variance of η_i is small compared to ε_i , i.e. the ε_i represent relatively large transient effects (noise) while the η_i represent relatively small permanent changes in the catchability.

In this simple example of annual fishing incidents, η_i modifies catchability at each successive fishing incident. In general, each step of the random walk can be taken less frequently, as might be appropriate when multiple fishing incidents by one fishery occur within a year. In the albacore analysis (where the frequency of fishing incidents is quarterly for the longline fisheries and monthly for the surface fisheries), random walk steps are taken annually for all fisheries.

Where the frequency of fishing incidents is quarterly or more, we allow catchability within a year to vary with a regular seasonal pattern. Equation (10) then becomes

$$\log_{\bullet}(F_{ij}) = \log_{\bullet}(s_j) + \log_{\bullet}(q_i) + \log_{\bullet}(E_i) + c_1 \sin[24\pi(m - c_2)] + \varepsilon_i$$
(13)

where *m* is the month in which the fishing incident occurred and c_1 and c_2 are the seasonality parameters.

3.4 Assumptions about the length-at-age

The assumptions concerning the length distribution of the fish are:

- 1. The lengths of the fish in each age class are normally distributed around their mean length (see equation 14).
- 2. The mean lengths-at-age lie on a von Bertalanffy growth curve (see equation 16) modified to include, where appropriate:
 - sampling bias for the first age class (see equation 17) (other age classes are randomly sampled without bias); and
 - density-dependent growth (see equation 19).
- 3. The standard deviations of the actual lengths about the mean lengths-at-age are a simple function of the mean length-at-age (see equation 20).

The following symbols are used in the mathematical expression of these assumptions:

- α subscript indexing the length frequency intervals.
- N_1 the number of length intervals in each length frequency data set.
- S_i the number of fish in the *i* th length frequency data set.
- $f_{\alpha i}$ the number of fish whose lengths lie in the α th length interval in the *i* th length frequency data set.
- $p_{ij\alpha}$ the probability that an age class j fish picked at random from the fish which were sampled to get the *i* th length frequency data set has a length lying in length interval α .
- $Q_{\alpha i}$ the probability that an animal picked at random from the fish which composed the *i* th length frequency data set has a length lying in length interval α .
- $\widetilde{Q}_{\alpha i}$ the observed proportion of fish in the *i* th length frequency data set having a length lying in length interval α .
- μ_{ii} the mean length of the age class j fish in the i th length frequency data set.
- σ_{ij} the standard deviation of the length distribution of the age class j fish in the i th length frequency data set.
- x_i the midpoint of the *i* th length frequency interval.
- w the width of the length frequency intervals.
- L_1 the mean length of the first age class on the von Bertalanffy curve in month 1.
- L_r the mean length of the last age class on the von Bertalanffy curve in month 1.
- K the von Bertalanffy K parameter.
- ρ the Brody growth coefficient $(K = -\log_{e}(\rho))$.
- b the coefficient of sampling bias in the first age class.
- λ_1, λ_2 parameters determining the standard deviations $\sigma_{i\alpha}$.
 - $\xi_{i\alpha}$ parameters determining the relative variances of the sampling errors within the *i*th length frequency data set.
 - τ parameter determining the overall variance of the sampling errors in all the length frequency data sets.

Assumption 1: Normal distribution of length at age

If the lengths of the age class j fish in the α th length frequency data set are normally distributed around their mean $\mu_{j\alpha}$ with standard deviations $\sigma_{j\alpha}$, the $p_{ij\alpha}$ can be expressed in terms of $\mu_{j\alpha}$ and $\sigma_{j\alpha}$ by

$$p_{ij\alpha}(\mu_{j\alpha},\sigma_{j\alpha}) = \frac{1}{\sqrt{2\pi\sigma_{j\alpha}}} \int_{x_i-w/2}^{x_i+w/2} \exp\left\{\frac{-(x-\mu_{j\alpha})^2}{2\sigma_{j\alpha}^2}\right\} dx$$
(14)

As long as $\sigma_{j\alpha} > w$, the integral can be approximated sufficiently well by setting

$$p_{ij\alpha}(\mu_{j\alpha},\sigma_{j\alpha}) = \frac{w}{\sqrt{2\pi\sigma_{j\alpha}}} \exp\left\{\frac{-(x-\mu_{j\alpha})^2}{2\sigma_{j\alpha}^2}\right\}$$
(15)

This approximation has been used in the model.

Assumption 2: Relationship of length to age

Parameterization of von Bertalanffy growth

If the mean lengths $\mu_{j\alpha}$ lie on a von Bertalanffy curve, then, using the parameterization given by Schnute and Fournier (1980)

$$\mu_{j\alpha} = L_1 + (L_{N_j} - L_1) \left[\frac{1 - \rho^{j - 1 + (m(\alpha) - 1)/12}}{1 - \rho^{N_{j-1}}} \right]$$
(16)

where L_1 , the mean length of the first age class, L_{N_r} , the mean length of the last age class, and ρ , the Brody growth coefficient, are the three parameters that determine the form of the von Bertalanffy curve, and $m(\alpha) - 1$ is the number of months after the presumed birth month of the fish in the α th length frequency data set.

Sampling bias in the first age class

For some length frequency data sets, the sampling procedure or the fishery does not fully select the smallest fish in the first age class. The effect of this size selectivity is that the mean length of the fish in the first age class in the length frequency data set is larger than the mean length of the fish in the population. If this sampling bias is not accounted for, biased parameter estimates may be produced. We assume that size selective bias only occurs for fish in the first age class and that it decreases linearly with age until the fish reach the second age class, thus

$$\mu_{1\alpha} = L_1 + (L_{N_J} - L_1) \left[\frac{1 - \rho^{(m(\alpha) - 1)/12}}{1 - \rho^{N_{J-1}}} \right] + \frac{b \left[12 - (m(\alpha) - 1) \right]}{12}$$
(17)

and the second second

where b is the sampling bias coefficient.

. . .

Density-dependent growth

For many species it is suspected that individuals of small (in numbers of fish) cohorts may grow more quickly than those of large cohorts (i.e. density-dependent growth). If true, this phenomenon could have a large effect on the conclusions drawn from a length-based stock assessment. To test for evidence of the existence of the dependence of the mean length at age on cohort strength we have incorporated density-dependent growth into the model in the following fashion.

Consider a cohort k at age j in year i. If we denote recruitment as occurring at age 1, the strength of cohort k is N_{k1} , where k = i - j + 1. Let $A = \frac{1}{n} \sum_{k} N_{k1}$ be the average recruitment. The normalized relative cohort strength is given by

$$R_{k} = \frac{(N_{k1} - A)}{\sqrt{\sum_{k} N_{k1}^{2}}}.$$
(18)

The changes in mean length are effected by changing the apparent age of the fish before the length at age is calculated.

If the age class is j the apparent age a is

$$a = j + 1.9 \left[\frac{1}{1 + \exp(-dR_k)} - 0.5 \right]$$
(19)

where d determines the amount of density-dependent growth; if d=0, a=j. Since the standard deviation of the R_k has been normalized to 1, the "generic" variation in the R_k will be about -2 to 2. Thus the difference in a between the largest and smallest cohorts of any given age class will be approximately $1.9 \left[\frac{1}{1 + \exp(-2d)} - \frac{1}{1 + \exp(2d)} \right]$. For d=0.4 (which is close to the estimate for the albacore data) this yields a generic variation of about 0.73 years.

Assumption 3: Relationship of standard deviations in length at age to mean length at age

The standard deviations $\sigma_{j\alpha}$ are parameterized as a simple function of length involving two parameters λ_1 and λ_2 :

$$\sigma_{j\alpha} = \lambda_{1} \exp\left\{\lambda_{2} \left[-1 + 2\left(\frac{1 - \rho^{j-1 + (m(\alpha) - 1)/12}}{1 - \rho^{N_{j-1}}}\right)\right]\right\}$$
(20)

where the term enclosed in square brackets expresses the length dependency of the standard deviations independently of the numerical values of the parameters L_1 and L_{N_1} (cf. equation 16). The two coefficients, λ_1 and λ_2 , transform the rescaled length to the standard deviations. λ_1 determines the magnitude of the standard deviations, and λ_2 determines the length-

dependent trend in the standard deviations. If $\lambda_2 = 0$, the standard deviations are length-independent.

3.6 Maximum likelihood estimation

The parameters of the model are estimated by maximizing the log-likelihood function (or more generally by maximizing the sum of the log-likelihood function and the log of the density of the Bayesian prior distribution). The log-likelihood function consists of the sum of several components, the most important of which correspond to the length frequency data and the total catch estimates.

The log-likelihood contribution for the length frequency data

Due to the large variability in the length samples which occurs for real fishery length frequency data, the model employs a robust maximum likelihood estimation procedure. The motivation for using this procedure and the technicalities behind the procedure are described in Fournier et al. (1990) We shall not repeat this discussion here, but for convenient reference we briefly describe the form of the log-likelihood function employed.

If the $\tilde{Q}_{\alpha i}$ are derived from a random sample of size S_i , they would be random variables with means $Q_{\alpha i}$ and variances $(1-Q_{\alpha i})Q_{\alpha i}/S_i$. Two modifications have been made to this formula. If $Q_{\alpha i} = 0$ the formula implies that the variance of $\tilde{Q}_{\alpha i} = 0$. To decrease the influence of areas where no observations are expected we add a small number to the variance formula in such cases. To reduce the influence of very large sample sizes we have assumed that sample sizes >1000 are no more accurate than sample sizes of 1000. Set $\xi_{i\alpha} = (1-Q_{\alpha i})Q_{\alpha i}$ and set $\tau_i^2 = 1/\min(S_i, 1000)$. Assume the variance of \tilde{Q}_{α} is given by $(\xi_{i\alpha} + 1/N_i)\tau_i^2$.

The likelihood function contribution for the length frequency data employed in the model is

$$\prod_{\alpha=1}^{N_{A}} \prod_{i=1}^{N_{I}} \left[\frac{1}{\sqrt{2\pi (\xi_{i\alpha} + 1/N_{I})\tau}} \left(\exp\left\{ -\frac{(\tilde{Q}_{i\alpha} - Q_{i\alpha})^{2}}{2(\xi_{i\alpha} + 1/N_{I})\tau^{2}} \right\} + .01 \right) \right]$$
(21)

Taking the logarithm of equation (21) we obtain the log-likelihood function for the length frequency data:

$$-\frac{1}{2}\sum_{\alpha=1}^{N_{A}}\sum_{i=1}^{N_{I}}\log_{e}\left(2\pi(\xi_{i\alpha}+.1/N_{I})\right) -\sum_{\alpha=1}^{N_{A}}N_{I}\log_{e}(\tau)$$

$$+\sum_{\alpha=1}^{N_{A}}\sum_{i=1}^{N_{I}}\log_{e}\left[\exp\left\{\frac{-(\tilde{Q}_{i\alpha}-Q_{i\alpha})^{2}}{2(\xi_{i\alpha}+.1/N_{I})\tau^{2}}\right\}+.01\right]$$
(22)

Expression (22) is the contribution to the log-likelihood function for the observed length frequency data.

The log-likelihood contribution for the observed total catches

Assuming for simplicity that there is only one fishery per year, the log-likelihood contribution for the observed total catches is given by

$$p_c \sum_{i} \left(\log(C_{i.}^{obs}) - \log(C_{i.}) \right)^2 \tag{23}$$

where p_c is determined by the prior assumption made about the accuracy of the observed catch data.

The log-likelihood contribution for the Bayesian priors on the effort-fishing mortality relationship

The log-likelihood contribution for the Bayesian priors on the η_i and ε_i , (see equations 10 and 11) is given by

$$p_{\eta} \sum_{i} \eta_{i}^{2} + p_{s} \sum_{i} \varepsilon_{i}^{2}$$
 (24)

The size of the constants ρ_{η} and ρ_{ε} are adjusted to reflect prior assumptions about the variances of these random variables.

Fitting the model

The parameters of the model are estimated by maximizing the log-likelihood function (or posterior density in the Bayesian framework) as described above. The maximization was performed with a quasi-Newton function minimizing routine employing exact derivatives with respect to the model parameters. The derivatives were calculated using the C++ class library, AUTODIF, using an extension of the technique known as automatic differentiation (Griewank and Corliss 1991). This approach is especially useful for models with large numbers of parameters. It also provides quick and accurate estimates of the Hessian matrix at the maximum, which can be used to obtain estimates of the covariance matrix and confidence limits for the parameters of interest (see section 3.7).

3.6 Hypothesis testing

It is frequently of interest in statistical modeling to add model structure in the form of one or more hypotheses concerning some process(es) of interest, and to observe the resulting change in model performance. This process can be thought of as attempting to define the "correct" model for a given set of data. Two approaches are taken to the addition and testing of hypotheses – a Frequentist approach and a Bayesian approach.

Frequentist approach

With the Frequentist approach to hypothesis testing, parameters representing a more complex model are added to the simpler model and the resulting improvement in fit is calculated. If this improvement in fit is large enough (as indicated by a likelihood ratio test, for example) the more complicated model is accepted. Otherwise the more complicated model is rejected and the simpler model is accepted as providing an adequate description of the data. Various more complicated models may be investigated in this fashion.

Some of the hypotheses that can be tested in this way are:

- the number of (significant) age classes represented in the data;
- the existence of a length-dependent trend in the standard deviation of the distribution of the length at age around the mean length; and
- the existence of a relative cohort strength component in the mean length at age of a cohort (density-dependent growth).

and the star

Bayesian approach

Some hypotheses that are useful in length-based stock assessment cannot be well represented in a Frequentist context. An example is the existence of a time-series (random walk) trend in catchability for a fishery. For such hypotheses, the results of the analysis are not as clear cut as they are for the Frequentist approach. We do not either accept or reject the existence of a trend in catchability. Instead, the analysis will produce a probability distribution for quantities of interest. For example we can get an (approximate) probability distribution for the ratio of the catchability for the first year of a fishery to the catchability for the last year of the fishery. This can be used to produce, for example, an estimate of the probability that the catchability has increased by 30% or more.

3.7 Estimation of confidence intervals

A great advantage of an integrated model such as this is that the estimates of the uncertainty in the parameter estimates automatically take into account the effect of all of the model's assumptions, such as the uncertainty in the age at length, the possibility of trends in catchability, effects caused by variability in the length frequency data and errors in the estimates of fishing effort.

Confidence limits for the parameter estimates are calculated by employing the usual second order approximation to the posterior distribution at its mode. Let $\theta_1, \ldots, \theta_n$ denote a minimal set of *n* model parameters from which all model parameters can be calculated, and let $p(\theta_1, \ldots, \theta_n)$ be some parameter of interest, while $L(\theta_1, \ldots, \theta_n)$ is the logarithm of the posterior distribution. Then the estimated standard deviation p_σ for *p* is given by the square root of $\sum_{ij} \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} \Lambda_{ij}$ where $\Lambda = \left(\frac{\partial^2 L}{\partial \theta_i} \frac{\partial \theta_j}{\partial \theta_j}\right)^{-1}$ and the calculations are carried out at the mode of the posterior distribution. Then, 0.95 confidence limits for the *p* are given by $[p-1.96p_\sigma, p+1.96p_\sigma]$. These confidence limits are not invariant under reparameterization. To compensate somewhat for this the confidence limits for parameters which must be positive, such as estimates of biomass, are calculated by computing the confidence limits. This yields the logarithms of these parameters and then transforming the confidence limits. This yields the confidence limits $[p \exp(-1.96 p_\sigma/p), p \exp(1.96 p_\sigma/p)]$.

13

and general and a state of the

4. SOUTH PACIFIC ALBACORE ASSESSMENT

The preliminary results presented below are a limited selection of the fits to the albacore data carried out to date. As yet, not all of the structural hypotheses described in the previous section have been tested. The results presented assume the following model structure:

- nine significant age classes (probably corresponding to ages 2 to 10+);
- three regions, as defined earlier;
- age-independent diffusion of fish among regions;
- recruitment assumed to occur only in region C (the southern-most region);
- time-series trends in catchability for all fisheries; and
- a trend in standard deviation of length at age with increasing mean length at age.

Additional model hypotheses, such as seasonal variation in catchability, density-dependent growth, age-dependent diffusion and recruitment by region, are yet to be formally incorporated into the full analysis. However, some results from preliminary fits incorporating seasonal catchability and density-dependent growth are presented.

We used two different starting values for M (0.2 and 0.4 yr⁻¹). When M was eventually relaxed, both of the fits converged to the same solution. This encourages us to believe that the solution obtained is global, rather than local, at least with respect to M, which is a notoriously difficult parameter to estimate. Ultimately, we will need to use different starting values for other parameters, such as K and the movement parameters, and use different assumed numbers of significant age classes, to increase our confidence that an overall global solution has been found.

4.1 Results to date

Selectivity coefficients

Selectivity coefficients (Figure 6) for the longline fisheries show an increasing trend with age. The differences among fisheries 1, 2 and 3 are in line with observations of increasing average size of longline-caught albacore from south to north. We originally tried fitting the model with the selectivities for fisheries 1, 2 and 3 constrained to be equal. However, a significantly improved fit was obtained when this assumption was relaxed. This suggests that differences in the size composition of longline catches in the three regions are not entirely due to differences in the size structure of the albacore populations in those regions - other factors such as targeting may also have an effect.

The New Zealand troll fishery selectivities also increase with age, in order to accommodate occasional length frequency samples containing larger fish from that fishery. Maximum selectivity for the STCZ troll and driftnet fisheries occurs for age classes 3 and 4 (likely ages 4 and 5).

Catchability coefficients

Estimated catchability time series are plotted with catchability deviations (ε_i), by fishery, in Figure 7. Strong trends in catchability are evident in the longline fisheries. For the DWFN

fisheries (fisheries 1-3), catchability is estimated to have declined substantially during the late 1960s and 1970s. In this analysis, these longline fisheries are aggregated across the Japanese, Taiwanese and Korean fleets, which are known to have targeted albacore to different extents over time. It is possible, therefore, that the estimated trends in catchability reflect both the changing fleet composition in the three regions and changing targeting practices by those fleets over time. In recent years, catchability in fisheries 2 and 3 has increased somewhat. For fishery 4 (Region B), catchability has shown a consistent increasing trend over time. For the troll and driftnet fisheries, no clear trends in catchability are evident.

Catchability deviations show a wide scatter about the estimated catchability trends. Much of this scatter is due to seasonal variation in catchability, as suggested by the seasonally variable CPUE patterns shown in Figure 5. Some preliminary fits incorporating seasonal variation in catchability have been carried out, and strong seasonal signals for all fisheries estimated (Figure 8).

Exploitation rates

Exploitation rates (the proportion of the population harvested per year) for two age groups, corresponding approximately to 3-6 years and 7-10+ years (assuming age class 1 fish are approximately two years of age), have been estimated, along with their 95% confidence intervals (Figure 9). Overall, the exploitation rates are low, and are lower for the younger age group (primarily exploited by the surface fisheries) than for the older age group (primarily exploited by the surface fisheries) than for the older age group (primarily exploited by the driftnet fishery. Prior to the mid-1980s, exploitation of this group was inconsequential, as it was restricted to minor catches of these age classes by longliners. For the older age group, exploitation rate declined through the 1960s and 1970s from a high of around 0.15. Since the mid-1980s, exploitation rate has increased slightly, with the most recent estimate in the vicinity of 0.05. These recent exploitation rates are consistent with the range of estimates obtained from an analysis of albacore tagging data (Bertignac et al. 1996).

Natural mortality rate

The estimate of M is reasonably consistent with expectation, 0.53 yr⁻¹, with 95% confidence intervals of 0.49-0.58 yr⁻¹. As indicated above, this estimate was arrived at from two different starting values, 0.2 and 0.4 yr⁻¹. The confidence interval is surprisingly small, indicating that Mis reasonably well determined in these data. Previous analyses of the same data without spatial structure were not successful in estimating M (the estimates tended to converge to zero). It is therefore likely that the addition of spatial structure is important in resolving this parameter.

In common with many other tuna populations, the sex ratio of adult South Pacific albacore changes rapidly with increasing size to favor males. This raises the possibility that M may be greater for older fish (because of high female mortality). Such possibilities can be investigated within the framework of this model.

Growth parameters and catch age composition

The estimation of catch age composition from length composition assumes, amongst other things, that albacore grow according to a von Bertalanffy growth curve. We use a parameterization such that growth is specified by three parameters - the mean length of the first age class (47.7 cm), the mean length of the last age class (98.1 cm), and growth coefficient K (0.17 yr⁻¹). These parameters can be transformed to provide the usual von Bertalanffy growth parameter L_{∞} (115.6 cm). These parameters imply a growth increment of about 10 cm after one year for a 50 cm albacore. This growth increment is somewhat less than that derived on the basis of tagging data (13.85 cm - Bertignac et al. 1996), but is probably within the bounds of uncertainty of that analysis.

In determining age composition from length composition, we also assume that the standard deviation (SD) of length at age is a linear function of mean length at age. The SD for the first age class is estimated to be 3.09 cm, while the ratio of the SDs of the first to the last age classes is 0.29. The fit of the model to the length data in terms of the correspondence of estimated mean lengths at age to obvious modes in the length frequency samples appears to be good for most of the fisheries. Examples of the fits to the length-frequency data are shown in Figure 10.

The addition of density-dependent growth to the model makes the growth of individual cohorts dependent on their abundance. Preliminary fits suggest that this effect is quite strong in South Pacific albacore, with more abundant cohorts growing slower than less abundant cohorts. The effect that this has on the estimated mean lengths at age is shown in Figure 11 by the differences between the solid and dashed vertical lines for each age class.

Movement parameters

The estimated movement parameters can be converted to annual exchange rates between regions for ease of interpretation. The annual rate of movement from region C (the recruitment region) to region B is 14.5%, from region B to region A and region C is 2.3%; and from region A to region B is 2.2%. These exchange rates result in a net "flow" of albacore from south to north.

Population biomass and recruitment

The time series of population biomass and recruitment are key outputs of the model from a stock assessment viewpoint. The time series can be looked at in terms of absolute or relative quantities. For the biomass (Figures 12 and 13) and recruitment (Figure 14) time series, we have plotted both of these measures, with the relative measures normalized to the average values over the whole time series. As would be expected, the 95% confidence intervals for the relative measures are much tighter than for the absolute estimates.

The biomass estimates (Figure 12) show a strongly increasing trend up to the late 1970s and a decreasing trend thereafter until about 1990. Biomass in the last two years increases, although the confidence limits about these estimates are relatively wide. Relative biomass shows similar trends in the three regions (Figure 13). Biomass is smallest in region A and largest in region C (the recruitment region). Note that the confidence intervals on the region-specific estimates are considerably wider than for the spatially aggregated time series, reflecting uncertainty in the estimated movement parameters.

The recruitment estimates (Figure 14) show an increasing trend through to the mid-1970s and a decreasing trend from the mid-1970s to about 1990, which appear to drive the biomass trends observed in Figures 12 and 13. The 1992 recruitment is estimated to be close to the

highest of the entire time series, whereas the 1993 level is the lowest. Note, however, that these later estimates of recruitment are associated with the highest uncertainty.

Relatively low estimates of recruitment are obtained for 1980, 1985 and 1990. Assuming that the age of recruitment is approximately two years, the spawning seasons corresponding to the low recruitments match well with the occurrence of *El Nino* episodes (negative values of the Southern Oscillation Index) in the Pacific Ocean (Figure 14). The high recruitments resulting from spawning in the mid-1970s and 1989-90 also seem to correspond to *La Nina* events (positive values of the Southern Oscillation Index). The relationship is not as good over the first half of the time series, but recruitment variability during this period may not be well estimated because of the absence of fisheries directed at small albacore.

The possibility of large-scale environmental effects on recruitment is interesting in several respects. First, the nature of the relationship, if it exists, is interesting from a biological point of view in that it may shed light on the population dynamics of pre-recruit albacore. Also, the existence of such a relationship raises the possibility of using large-scale environmental data to develop a statistical model of recruitment which could be used in population projections and fishery forecasting.

4.2 Conclusions

We have developed an integrated, likelihood-based age-structured model using length data that may, for many fisheries, be a viable alternative to more traditional approaches such as VPA. The statistical approach employed offers the advantage of being able to objectively assess the information content the data. One outcome of this is the ability to construct approximate confidence intervals on the parameters of interest. Another outcome is the ability to test hypotheses regarding alternative model formulations. While the "correct" model may never be found, our approach enables sensible decisions to be made in choosing the *best* model from a range of alternatives.

The results presented here for South Pacific albacore should be considered preliminary, as further questions concerning the analysis need to be explored. These include:

- testing different numbers of assumed age classes
- using different starting values of various model parameters
- incorporating additional model structure, as appropriate, e.g. density-dependent growth, age-dependent M.

In the longer term, we might also investigate improvements that might result from further stratification of the DWFN longline fishery in each region by fleet (i.e. separating Japan, Korea and Taiwan), using logbook coverage for the various fisheries to index the credibility of the effort data and using tagging data to refine model parameters, particularly the movement parameters. Pending the completion of this work, we can offer some observations on the basis of the results obtained to date:

• The estimated exploitation rates on both young and older age groups are relatively low, generally less than 10% per year.

- The trends in recruitment and population biomass are unlikely to have been influenced significantly by the fisheries.
- The low exploitation rate of the younger age classes, even at the height of the driftnet fishery, implies that the surface fisheries are unlikely to have had an observable impact on the longline fishery.
- South Pacific albacore recruitment may be linked to large-scale climatic variation, with lower than average spawning success or larval survival during *El Nino* episodes.

The model should be a useful tool for future management of the South Pacific albacore fishery. Two key uses come readily to mind, and would require only minimal adaptation of the existing computer software. First, it would be relatively straight forward to cast the results of the model in a form suitable for comparison with limit or target reference points, as envisaged by the recent United Nations agreement on straddling and highly migratory fish stocks. This could be done by calculating the probability that a particular reference point would be exceeded under a particular fishing regime. Second, it is possible that the model could be a useful forecasting tool for both the surface and longline fisheries. Given some reasonable model for future recruitment (perhaps linked to large-scale environmental conditions such as *El Nino*), it would be possible to project the stock forward in time. Confidence intervals could also be determined for the projections to capture the uncertainty in future recruitment and the current population state. Such forecasting, if successful, would presumably assist both industry and management decision-making.

REFERENCES

- Bertignac, M., P. Lehodey and J. Hampton. 1996. An analysis of South Pacific albacore tagging data: estimation of movement patterns, growth and mortality rates. Sixth South Pacific Albacore Research Workshop, Working Paper 3. Rarotonga, Cook Islands, 5-7 March 1996.
- Deriso, R.B., T.J. Quinn II, and P.R. Neal. 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42:815-824.
- Doubleday, W.G. 1976. A least squares approach to analyzing catch at age data. Int. Comm. Northw. Atl. Fish. Res. Bull. 12:69-81.
- Dupont, W.D. 1983. A stochastic catch-effort method for estimating animal abundance. Biometrics 39:1021-1033.
- Fournier, D., and C.P. Archibald. 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39:1195-1207.
- Fournier, D.A., J.R. Sibert, J. Majkowski and J. Hampton. 1990. MULTIFAN: a likelihoodbased method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data for southern bluefin tuna (*Thunnus maccoyii*). Can. J. Fish. Aquat. Sci. 47:301-317.

- Fournier, D.A., J.R. Sibert, and M. Terceiro. 1991. Analysis of length frequency samples with relative abundance data for the Gulf of Maine northern shrimp (*Pandalus borealis*) by the MULTIFAN method. Can J. Fish. Aquat. Sci. 48:591-598.
- Griewank, A., and G.F. Corliss (eds). 1991. Automatic differentiation of algorithms: theory, implementation and application. Proceedings of the SIAM Workshop on the Automatic Differentiation of Algorithms, held January 6-8, in Breckenridge, Colorado. Society for Industrial and Applied Mathematics, Philadelphia.
- Gudmundsson, G. 1994. Time-series analysis of catch-at-age observations. Appl. Statist. 43:117-126.
- Hilborn, R. and C.J. Walters. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Chapman and Hall, N.Y. 570 p.
- Lewis, A.D. 1990. South Pacific albacore stock structure: a review of available information. Third South Pacific Albacore Research Workshop, Working Paper 5. Noumea, New Caledonia, 9-12 October 1990.
- Megrey, B.A. 1989. Review and comparison of age-structured stock assessment models from theoretical and applied points of view. American Fisheries Society Symposium 6:8-48.
- Nishikawa, Y., M. Honma, S. Ueyanagi, and S. Kikawa. 1985. Average distribution of larvae of oceanic species of scombroid fishes, 1956–1981. Far Seas Fish. Res. Lab. S Series 12.
- Paloheimo, J.E. 1980. Estimation of mortality rates in fish populations. Trans. Am. Fish. Soc. 109:378-386.
- Pope, J.G., and J.G. Shepherd. 1982. A simple method for the consistent interpretation of catch-at-age data. J.Cons. Int. Exp. Mer 40:176-184.
- Ramon, D., and K. Bailey. 1993. Reproductive patterns of South Pacific Albacore (*Thunnus alalunga*) as indicated by gonosomatic index and meiotic activity. Fifth South Pacific Albacore Research Workshop, Working Paper 13. Papeete, French Polynesia, 29 March-1 April 1993.
- Schnute, J., and D.A. Fournier. 1980. A new approach to length frequency analysis: growth structure. J. Fish. Res. Board Can. 37:1337-1351.

Figure 1. General distribution of the South Pacific albacore stock (light shading) and South Pacific albacore spawning (dark shading). The spawning distribution is interpreted from larval sampling data of Nishikawa et al. (1985). The "?" symbols indicate areas of minimal larval sampling.



Figure 2. Distribution of longline, troll and driftnet fisheries for South Pacific albacore.







Figure 4. Definition of regions and associated fisheries for the analysis of South Pacific albacore data.



Figure 5. Catch and catch per unit effort (CPUE) of South Pacific albacore by fishery and year quarter. Catch is in thousands of fish. CPUE for longline fisheries is in number of fish per 100 hooks; CPUE for troll and driftnet fisheries is in number of fish per operational day.





Figure 5. Continued.

23



1.0 Fishery 5: NZ Troli 0.8 **Region** C 0.6 0.4 0.2 0.0 2 3 5 1 4 6 7 8 Age class 1.0 0.8 Region C 0.6 0.4 0.2



9

2 3 4 5 6 7 8 9 Age class



Figure 6. Estimated selectivity coefficients for each fishery.

Figure 7. Estimated catchability (solid lines), which is assumed to change annually and has no seasonality, and deviations from the effort-fishing mortality relationship (dots), by fishery. The "***" indicate deviations beyond the scale of the figures.



25

ng national search in the search of the search of the search of the search in the search in the search of the The search of the s





Figure 8. Estimated catchability (solid lines), which is assumed to change annually and seasonally, and deviations from the effort-fishing mortality relationship (dots), by fishery. The "***" indicate deviations beyond the scale of the figures.





. . 1



Figure 9. Estimated average annual exploitation rates (heavy lines) and their 95% confidence intervals (thin lines) of presumed ages 3-6 and ages 7-10+.



29

Figure 10. Examples of model fit to the length-frequency data. The vertical lines indicate estimated mean lengths at age. Both the estimated aggregate (upper line) and age-class specific length distributions are shown.



DWFN Longline Region B



DWFN Longline Region C



Domestic Longline Region B

		Δ

Figure 10. Continued.





Figure 11. Example of a fit to the length-frequency data for the STCZ troll fishery for a model incorporating density-dependent growth. The solid vertical lines indicate estimated mean lengths at age in the absence of density-dependent growth. Both the estimated aggregate (upper line) and age-class specific length distributions are shown.



Figure 12. Estimated relative (scaled to the average) and absolute biomass, with 95% confidence intervals.







Figure 14. Estimated relative (scaled to the average) and absolute recruitment, with 95% confidence intervals.



Figure 15. Estimated relative recruitment by presumed spawning season (assuming age 2 at recruitment) and the Southern Oscillation Index (SOI). Negative values of the SOI indicate *El Nino* episodes and positive values *La Nina* episodes.



Assumed spawning season