

Fisheries Ecology and Management

CARL J. WALTERS AND STEVEN J.D. MARTELL

CHAPTER 5

An Overview of Single-Species Assessment Models

WHETHER THE AIM OF FISHERIES MODELING and analysis is to provide information to manage each species better while pretending it lives in isolation, or to provide historical reconstructions to help in assessing the parameters for ecosystem models, there will always be a critical need to analyze single-species historical data. There are now a number of useful texts on the methods and pitfalls of single-species assessment (Ricker 1975; Seber 1982; Hilborn and Walters 1992; Hilborn and Mangel 1997; Quinn and Deriso 1999). These texts describe a bewildering variety of methods, not because of the complexity of population-dynamics processes but because we must generally base assessments on a variety of largely fragmentary measurements that have been gathered over the years by biologists hoping to interpret what is going on in natural populations without having the tools or financial resources to directly measure the components of dynamic change in abundance and composition. Further, there is often much confusion about what the aims of an assessment should be in the first place, and about what criteria should be used in comparing alternative assessment techniques for the same data (e.g., statistical criteria versus criteria that measure performance at providing management advice).

In this chapter, we begin by asking the reader to think carefully about the alternative and sometimes conflicting objectives of single-species data analysis. We then review the state-observation dynamics structure that is rapidly becoming a standard way of thinking about assessment methods and of unifying apparently disparate methods and data. Then we examine two of the most difficult decision problems faced by stock-assessment scientists, namely, whether to use simple or complex models and what statistical criteria to use in the parameter estimation and assessment of uncertainty. We review the two main approaches to the reconstruction of population-dynamics histories from long-term data on population composition and relative abundance, virtual population analysis and stock synthesis, and provide advice on which of these to use under various circumstances. Finally, we discuss approaches to dealing with persistent change or “nonstationarity” in production dynamics.

5.1 OBJECTIVES OF SINGLE-SPECIES ASSESSMENT

Single-species assessment methods can be aimed at answering at least three quite distinct questions. The methods for answering each of these can be quite different, and confusion about this point has led to many published assessments that fail to address the question(s) for which they were funded

PRINCETON UNIVERSITY PRESS

PRINCETON AND OXFORD

by management agencies, to much time wasted by scientists trying to answer questions that no one asked or even needed to ask, and even to assessment results that are actively misleading due to statistical problems related to "overparameterization." The main questions or objectives that we might try to answer with single-species assessment are these:

1. **The Status Question:** What is the current level of harvesting impact on the stock, measured by the fishing-mortality or exploitation rate $F = (\text{Catch})/(\text{Abundance})$, in relation to goals or standards that may come from an analysis of the data from the stock or from other stocks for which there is historical experience about sustainable exploitation rates? Note that catch alone is never a meaningful measure of current fishing impact, since the catch cannot be sensibly interpreted without any information about how large it is relative to the stock size (10,000 tons may sound large, but it would likely be insignificant if taken from a stock of 1,000,000 tons). Often the most efficient way to answer this question would be to abandon the existing machinery for the analysis of historical data and to invest instead in direct stock-size and exploitation-rate estimation methods such as tagging experiments. However, there is typically much resistance to such proposals due to fears about investing in new, unproven field methods.
2. **The Mean Productivity Question:** What is the average long-term relationship between the fishing-mortality rate, stock size, and yield (how does the average net production rate of the stock vary with abundance and the harvest rate?), assuming that there is a "stationary" relationship between the average production rate and stock-size or exploitation rate? This question raises the main concerns about model complexity and overparameterization, since to estimate average response parameters it can often be best to deliberately use an oversimplified, "biased" model that sacrifices realism (in its representation of the sources of variation) for precision in its estimation of the key mean response parameters.
3. **The Stock Reconstruction Question:** How have the components of net production (growth, recruitment, natural mortality rate) varied over time, and what alternative hypotheses about the impact of stock size, fishing, and habitat/environmental changes could explain these variations? The most elaborate, complex stock-assessment models in use today are aimed at this question and use a variety of information sources about changes in relative abundance and size/age composition to say as much as possible about historical total abundances and, particularly, changes in recruitment rates.

The biggest single mistake that is being made routinely by today's inexperienced stock-assessment scientists is to assume that answering the third question is equivalent to, or the best route to, answering the first two questions. Complex stock-reconstruction models generally have 10s to 100s of unknown parameters to be estimated from the data, and many biologists appear to assume that having lots of parameters means that a model can both explain (fit) the data better and also filter out (correct for) more causes of variation so as to make the mean response parameters more visible. This

intuition is often wrong: explaining the noise generally means admitting more, not less, uncertainty about the underlying average responses; i.e., the complex models admit that there are many more ways to explain the observations—meaning that there is more uncertainty about which particular way (if any) is the correct one.

One traditional division of approaches to this analysis involves a hierarchy of increasing detail as more information becomes available about a particular fishery, starting with a simple surplus-production-analysis (the net balance of recruitment + growth – natural mortality) and moving later to an estimation of the age-size and spatial components of these processes. This division has come with the presumption that methods should be used in progression over time, with more detailed methods being obviously "better" as more information becomes available. However, a more important distinction in practice is about whether composition information (size/age structure data) will be used in the analysis. We can always model or explain the overall biomass/numbers dynamics in terms of recruitment, growth, and natural mortality components, whether or not these component effects on past production can be separated by using more detailed data. In such modeling "beyond the data," we should not claim that a "better" model structure is needed to explain what data are available. Rather, we should admit that there are policy concerns and options that cannot be addressed without a more detailed calculation (e.g., the effects of changing size selectivity), and we should provide the capability to at least perform the calculations without pretending that the data justify them. That is, we should not pretend that the limitations of the available data should dictate or limit the range of policy concerns that we attempt to evaluate (policy concerns and options should drive the choice of models, not the other way around!).

5.2 STATE-OBSERVATION COMPONENTS

There has been a major change over the last 20 years in how we analyze historical fisheries data in order to estimate the past impacts of fishing and other factors. If you examine one of the older books on stock assessment and the analysis of fisheries statistics, e.g., Ricker (1975) or even Hilborn and Walters (1992), you will see mainly a collection of disjointed "recipes" for the analysis of bits of data such as catch curves, single tagging experiments, and time series of catch and CPUE data. These recipes were developed mainly by starting out with particular, usually simple population models and assumptions about how observations are related to population characteristics, then trying to interpret the predicted observations as having come from some standard statistical model such as a linear regression. That is, we tried to turn our data-analysis problems into familiar statistical problems. Examples include regression estimates of abundance from depletion experiments and estimates of the total mortality rate from catch curves.

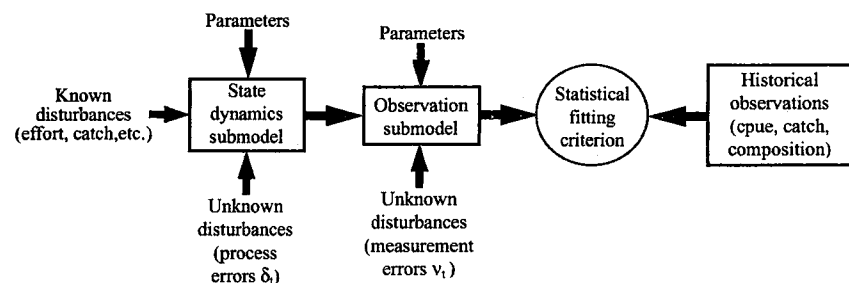


FIGURE 5.1: Most stock-assessment models are built with the general systems notion that to analyze a set of incomplete data on any dynamic system, we must hypothesize (assume) at least two distinct types of “submodels”: one for the (generally unobserved) dynamics of system state in nature, and one for the relationship between system state and the available measurements. Arrows show the flow of calculations or dependencies (e.g., predicted observations depend on predicted states, plus observation submodel parameters). Note that catches are sometimes treated as input disturbances (estimation “conditioned on catch”) or predicted outputs (estimation “conditioned on effort”).

This changed dramatically in the 1980s, with the idea that we could build quite general models incorporating structural knowledge about population processes, use these to predict a rich variety of observations all at once, and use more complex statistical tools and criteria (nonlinear estimation, likelihood functions) to see how much could be said about all the data at once (fig. 5.1). At first, this change in approach was presented in an almost apologetic way, with titles such as Deriso et al. (1985), “catch-age analysis with auxiliary information.”

A key computer innovation that helped promote the change was the spreadsheet. Using a spreadsheet, it is easy to lay out a block of cells representing known historical “drivers” of change (fishing effort, catches, hatchery stocking rates, etc.), a block of cells representing a possibly quite complex population model for response to these drivers, another block of cells representing predictions of a variety of observations, and blocks of cells with actual data. Obvious spreadsheet functions can then be used to calculate the statistical agreement between observations and predictions and to search for parameter values that give better agreement. Ray Hilborn was the first to show how general and powerful this approach could be, in his work on Hudson River fishes. For that work, he not only “drove” a set of spreadsheet population models with fishing disturbance patterns over time, but also estimated the historical mortality rates of larval fish due to entrainment in power plants. He predicted not only patterns of catch and age composition for commercially harvested species, but also relative-abundance trends for larval and juvenile index sampling. His models were set up so as to include a rich variety of parameters for the effects that had been hypothesized, e.g., whether there is compensatory improvement in the survival of young fish after some have been killed by power plant entrainment.

The logical structure of figure 5.1 has also helped us understand that most stock-assessment models are, in fact, just generalizations of a very old and simple model known as a “depletion” or “removal” model. The idea in a depletion model is to remove a known number or biomass of animals from a population, while measuring how this removal affects the measures of relative abundance. As a very simple example of this idea, suppose 1000 tons of fish have been removed from an area, and this removal has reduced some index of abundance (catch per effort by fishers or by survey gear, acoustic target count, etc.) by 20%; we would then say that 1000 tons represents 20% of the initial stock, so the initial stock must have been 5000 tons. More generally, the removal estimator for multiple removals and samples from a closed population (no other gains or losses) is constructed by using the state and observation submodels:

$$N_t = N_0 - K_t \quad (5.1)$$

where N_t = numbers or biomass at time t , K_t = total removals before t ,

$$Y_t = qN_t \quad (5.2)$$

for the index of abundance at the start of the removal period t . Substituting the state equation for N_t into the observation equation for Y_t results in a simple linear regression model, or “Leslie method” (Leslie and Davis 1939) for estimating q and N_0 . Note that this simple model can go wrong for two kinds of reasons: *biological*—there is gain or loss of animals over time not accounted for by the state equation; and/or *observational*—the index Y_t is not proportional to N_t either because it measures the abundance of a subset of relatively more vulnerable animals that are removed at early times (the index shows “hyperdepletion”) or because it stays high as N declines (the index shows “hyperstability”). Much of the machinery of stock assessment has been developed by basically ignoring problems of the second kind and concentrating instead on representing components of biological change. Simple “surplus production” models are constructed, e.g., by replacing the closed-population state equation 5.1 with one that includes a function $f(N_t)$ for net production (recruitment + growth – natural mortality) as a function of population size:

$$N_{t+1} = N_t + f(N_t) - C_t \quad (C_t = \text{catch taken during time step } t) \quad (5.3)$$

which can be solved recursively over time for any starting N_0 and parameters of $f(N)$ to generate a set of predicted states N_t and predicted observations Y_t (using equation 5.2). Note that as we construct more complex models like equation 5.3, we are still using the catches C_t to provide population scale information, but recognizing that we might be able to obtain information about $f(N)$ by observing whether changes in Y_t were less than predicted assuming no production, as Y_t declined following the initial C_t removals. An obvious next step has been to recognize that removals are likely to affect the population composition (relative numbers of animals of different

ages and/or sizes) as well as the relative abundance, through the effects on mortality rates. To model these compositional changes (to predict observed changes in composition data), we replace equation 5.3 with a more complex accounting model for changes in size/age composition and extend the predicted observation submodel accordingly. But no matter how complicated we get, we remain reliant for most analyses (for which the historical abundance change has not been measured directly) on using the models to predict the relative impacts of changing the total removals C_t (or removals by size/age), i.e., we remain reliant on the basic idea of a removal estimator.

While biologists have tended to concentrate their efforts on developing better models for state dynamics, the main thing that appears to be going wrong in assessment with the general state dynamics/observation dynamics structure today is not with the population-dynamics models but, rather, with the observation models. Two main problems have caused dangerously misleading overestimates of abundance, recruitment, and net production during population declines and the onset of overfishing: (1) the use of commercial catch per unit effort (CPUE) or other relative-abundance indices that are not, in fact, proportional to abundance (Harley et al. 2001), and (2) changes over time in size/age selectivities that confuse the interpretation of population composition data. In particular, fishers often have the habit of targeting smaller fish as the abundance of larger ones declines; this can help to prevent the CPUE from declining and can shift the composition data so as to make it appear to a scientific observer that recruitment must still be healthy because plenty of small fish are still being caught. To a naïve reader, the recent stock-assessment literature appears to represent much progress in model development and the statistical representation of uncertainty (modern, equation-filled papers certainly look impressive), and few stock-assessment scientists have been forthright about pointing out that it is not better estimates that we are getting but, rather, better evidence that the data are actually telling us much less than we used to think, i.e., the estimates are often nonsense or even actively misleading. Unfortunately, this “defense of method” often makes it look as if we can get by in the future with the same kinds of data (abundance trend, composition) and, hence, directs attention away from the critical need to start gathering new kinds of data, especially direct estimates of exploitation patterns from methods like tagging experiments.

Still more unfortunately, many fisheries scientists are easy to fool: a lot of biologists and inexperienced assessment people seem to think that assessment methods will work if most of the assumptions about the data are correct, so long as there is plenty of data. That is, they seem to think that assumptions and observation-model components have an additive effect on assessment performance, similar to adding additional explanatory variables in a multiple-regression analysis. Unfortunately, this thinking is very wrong; all it takes to cause a misleading assessment and interpretation of stock changes is for even one key assumption to be wrong, particularly about stock

trends as evidenced by the relative abundance indices. To put it vividly, a few bad assumptions, and sometimes even just a few bad observations, can poison a whole stock assessment. Indeed, it is a bit worrisome that the 2003 Lowell Wakefield symposium (21st Lowell Wakefield Fisheries Symposium Anchorage, Alaska, USA, October 2003; The Wakefield Symposia are a series that has been wonderful about bringing assessment scientists together to share methods and experience) pretends to be specifically about “data limited” assessment situations, as though these were special cases. In fact, when we examine the really critical data and observation-model assumptions that drive assessments, and how frequently the most basic data cannot be trusted, we really should be thinking about all fisheries as severely data limited situations.

Nowadays there is a lot of quibbling about the computer implementation tactics for doing general-state/observation-dynamic analysis. This has been driven largely by numerical problems with nonlinear estimation procedures when we get ambitious about the number of “nuisance parameters” that we try to include in any descriptions of observation processes and size/age-vulnerability patterns of fish-to-fishing and survey sampling. One school of thought is that we should continue to try to use spreadsheet implementations as much as possible, because they help to keep assumptions transparent, to present the results to wider audiences, and to move from historical reconstruction to future prediction (just add spreadsheet rows or columns to represent dynamic changes into the future). The other school of thought is that we should use the best available statistical and nonlinear search technology, best exemplified by David Fournier’s AD Model Builder (AD = automatic differentiation, Otter Research 1994). Thankfully, software is gradually appearing that integrates the best features of both of these approaches.

5.3 ESTIMATION CRITERIA AND MEASURING UNCERTAINTY

Given a set of state-dynamics/observation-dynamics assumptions and their associated unknown parameter values, there are basically two ways to proceed with an estimation of the parameter values. First, we can seek the parameter values that maximize or minimize some intuitive measure of how well the model fits the data, such as weighted sums of squared deviations or absolute deviations between predicted and observed values. Second, we can adopt the maximum-likelihood approach of finding the parameter values for which the observations are most probable (from which the observations are most likely to have come). We now use the second approach most often, even when we cannot be confident about having chosen the right form of probability distributions for variability in the data. The likelihood approach offers several key advantages over ad hoc estimation criteria. These include

- formal methods for measuring possible overparameterization;
- a ready linkage to Bayesian methods for incorporating prior (e.g., other case) information about parameter values into the calculations and for providing outcome-probability statements for management-decision analysis; and
- the “automatic” and objective weighting of heterogeneous data types that may be combined in the likelihood function.

Unfortunately, it is no simple matter to construct a full likelihood function for data gathered from a nonlinear dynamical system that has been subject to nonlinear observation processes and both “process” and “observation” errors. While numerical methods for full, nonlinear “state-space” estimation methods are becoming available, most practical analyses are based on computationally convenient likelihood function approximations. Three main approximation approaches are now commonly used:

1. **Mean trajectory or observation error approach:** solve the state-dynamics equations omitting all process errors, as an approximation of the mean trajectory of the stochastic state dynamics, and treat the combined process + observation deviations from the predicted observations obtained this way as independent “observation errors.”
2. **Nuisance parameter or errors-in-variables approach:** treat the process errors (δ_t in fig. 5.1) as unknown, arbitrary past disturbances, and try to estimate them as “nuisance parameters” along with the parameters representing the mean dynamic response. This generally involves making some a priori assumption about the relative magnitudes of observation and process errors, but at least avoids the often silly pretense that the process errors were, in fact, drawn from some simple statistical distributions (Mother Nature is not that generous).
3. **Full state-space approach:** calculate the total likelihood of each observation integrated over the possible state-variable values that may have resulted from process errors. This generally involves a complex recursive calculation of the likelihood function over time, with either an extended Kalman filter or numerical integration schemes for the state dynamics (Pella 1993; Reed and Simons 1996; Schnute 1994).

Recent simulation studies by Schnute and Kronlund (2002) compare approaches (2) and (3) and show that these approaches give similar results and estimation performance at least for stock-recruitment state dynamics. This hints that the errors-in-variables approach may generally be the better one to take, especially considering the computational difficulties with the full state-space approach and with its requirements to make dubious assumptions about the statistical properties of the process disturbances δ_t . Further, an informative byproduct of the errors-in-variables approach is the time-series estimates of the process errors, which can be critical in detective work about the causes of the “errors,” forecasts of future variation, and the design of harvest strategies for coping with variation.

Whichever of the above approaches to likelihood formulation is taken, we generally have to deal with the likelihood components for two main types of observations: quantitative indices of abundance and/or total outputs (catches) and proportional contributions to observed composition data. Including observation errors, we most often assume observation models for quantitative variables to be of log-normal form:

$$Y_t = qX_t e^{\nu_t} \quad (5.4)$$

where X is any state variable or quantitative output, Y is an observation that we hope is proportional to X , and ν_t is normally distributed with mean zero and some unknown variance σ_ν^2 . We use the log-normal form because most quantitative observations in fish dynamics arise as a product of component and proportional observation/collection processes, and the sum of the logs of such proportions is likely to be normally distributed because of Central Limit Theorem effects. For any collection n of such independent observations included in an analysis, the log-likelihood function component for the collection, evaluated at the conditional (on all other parameters) maximum likelihood estimates of q and σ_ν^2 is just a constant that depends only on the data (and hence can be ignored for all parameter search and comparison calculations), plus the following term (see box 5.1):

$$\log L = -(n/2)\ln(SS) \quad (5.5)$$

Here, SS is just the sum of squared deviations

$$SS = \sum_{i=1}^n (z_i - \bar{z})^2 \quad (5.6)$$

where $z_i = \ln(Y_i/X_i)$ and \bar{z} (the arithmetic mean of the z_i) is the conditional maximum likelihood estimate of $\ln(q)$ (also SS/n is the conditional maximum likelihood estimate of σ_ν^2). For predictions of absolute quantities ($q = 1$), omit \bar{z} from the calculation of SS .

For observations of sample proportions (e.g., the proportions of catch that were of different ages), where n_i creatures of type i were observed and the state/observation model predicts that these should be a proportion p_i of the observations ($\sum p_i = 1$), the log-likelihood component needed for search and comparisons of the parameters that predict p_i from a state change is just

$$\log L = \sum_i n_i \ln(p_i) \quad (5.7)$$

This equation must be used with caution because the effective sample size for composition sampling in fisheries is most often much less than the number of fish ($\sum n$) actually measured (Fournier et al. 1990; Fournier et al. 1991; Zheng et al. 1995; Quinn et al. 1998). Together, the simple equations 5.5 and 5.7 cover most log-likelihood components that are likely to be added up (over data types) to form the circled statistical fitting criterion in figure 5.1.

BOX 5.1

EFFICIENT CALCULATION OF LIKELIHOOD AND BAYES POSTERIOR PROBABILITY FUNCTIONS

The likelihood function for a set of independent observations $Y = \{Y_1 \dots Y_n\}$ given a set of parameters $P = \{P_1 \dots P_m\}$ is just the product of the probability distributions of the individual observations given the parameter values, and the log-likelihood function is just the sum of the logs of these probability distributions. For example, if each observation Y_i is assumed to have arisen from a normal distribution with mean $\mu_i(P)$ (i.e., μ_i is a function of the parameters P), so the probability of Y_i is

$$p(Y_i|P) = \frac{1}{\sigma\sqrt{2\pi}} \exp - \left[\frac{(Y_i - \mu_i)^2}{2\sigma^2} \right]$$

then the log-likelihood for the entire set Y is

$$\ln L = -n \left[\ln(\sigma) + \frac{1}{2} \ln(2\pi) \right] - \sum_{i=1}^n \frac{(Y_i - \mu_i)^2}{2\sigma^2}$$

When we want to compare alternative parameter estimates, either by likelihood ratios (log-likelihood differences) or in search procedures for the maximum $\ln L$, the calculations can be simplified in at least two ways. First, we can drop any multiplicative terms in $p(Y|P)$, or additive terms in $\ln L$, that depend only on the data, since these do not change during parameter searches and cancel in likelihood ratios. Second, we can evaluate “nuisance” parameters like σ^2 at their conditional maximum likelihood estimates given the other P , provided we can obtain analytical formulae for these parameters. For example, if we differentiate $\ln L$ above with respect to σ^2 and set the resulting derivative to zero, the maximum likelihood estimate of σ^2 is easily shown to be $\frac{SS}{n} = \frac{\sum (Y_i - \mu_i)^2}{n}$. Substituting this back into $\ln L$, then dropping any constant, additive terms that will not affect $\ln L$ comparisons across P values, we get the likelihood “kernel” $\ln \hat{L} = -\frac{n}{2} \ln(SS)$ since the second term in $\ln L$ above just becomes the constant $-\frac{n}{2}$, and the first term becomes $-(\frac{n}{2}) \ln(\frac{2\pi SS}{n})$, which is equal to the constant $-(\frac{n}{2}) \ln(\frac{2\pi}{n}) - (\frac{n}{2}) \ln(SS)$.

The Bayes posterior probability for a particular parameter value P is defined as $p(P|Y) = L(Y|P)p_o(P)/k$, where $p_o(P)$ is a prior probability for P and k is a constant (the total probability of the data Y integrated over all P) that can be ignored in comparing probabilities for alternative values of P since it is the same for every P value. The log of this posterior probability is just $\ln L + \ln(p_o(P)) - \ln(k)$. In comparing posterior probabilities for alternative P estimates, we can again ignore both $\ln(k)$ and any additive terms in $\ln p_o$ that do not depend on P . So, e.g., if the prior probability for P_j is taken

(Continued)

(BOX 5.1 continued)

to be normal with mean $P_j^{(0)}$ and variance σ^2 , the only term of this normal distribution that we need retain in comparisons of log posterior probabilities for alternative values of P is $-(P - P_j^{(0)})^2/2\sigma^2$.

Likelihoods are almost always very tiny values that cannot be meaningfully interpreted by themselves or computed accurately. That is why we always do calculations using logarithms of them, and why we use only ratios of them in comparisons of how well alternative parameter values describe the data. This also means we can subtract any convenient constant, e.g., $\ln L_{max}$, the log-likelihood evaluated at the maximum likelihood estimates of P , from the log-likelihood or log-posterior probability when we want to summarize relative probabilities versus P rather than relative log probabilities, e.g., in plots of posterior probability distributions for key parameter values. ■

With a bit of practice, one can learn very quickly to set up state/observation models in the format of figure 5.1 and to perform the nonlinear statistical searches needed to find the maximum likelihood estimates of those state/observation/disturbance parameters admitted to be uncertain in the problem formulation. But this is where the work really begins, because the next step in the analysis is to provide some honest assessment of how good the parameter estimates, particularly those with policy implications, really are. Two things can go wrong assessing of uncertainty in the parameter estimates:

Numerical uncertainty A very wide (or even infinite) range of parameter combinations may equally well explain the data (the same or nearly the same likelihood for a wide range of combinations); in particular, the data may be equally well explained by assuming they came from a large, unproductive population or from a small, productive one.

Structural uncertainty The presumption of knowledge implied by the choice of model structure (and whatever parameters are treated as fixed knowns) may lead to apparently very precise but actually biased parameter estimates.

To identify and approach the first problem, we somehow need to map how the likelihood function (or likelihood times prior for Bayesian analyses) varies with changes in the parameter estimates. To approach the second problem we need to construct a set of models of increasing complexity and determine whether “freeing up” potentially weak assumptions (allowing more parameters to vary) leads to different results.

There are basically four approaches to studying how the likelihood function varies with parameter values, so as to estimate parameter uncertainty: (1) *Brute force*—map the likelihood function over a grid of parameter combinations. (2) *Information matrix*—use measures of the curvature of the likelihood function near the maximum likelihood estimates to construct confidence regions. (3) *Profile*—vary one important (“leading”) parameter across a grid of values, and plot the likelihood maximized over all other parameters for each value of this leading parameter. (4) *Sampling*—randomly sample parameter combinations using an importance sampling (SIR) or Metropolis-Hastings (MCMC) algorithm (Punt and Hilborn 1997; McAllister and Ianelli 1997), and display the frequency distributions of the important parameter values from this sample. Of these, the brute force approach is best for small problems (few parameters) but is impractical for most of the models used in fisheries assessment. The information matrix approach is the one commonly used to generate approximate confidence limits for parameters, but it can be misleading for fisheries problems because nonlinear model structures do not lead to confidence regions that are even approximately elliptic. The profile and sampling approaches typically give very similar results, provided the likelihood has one unique maximum for each value of the leading parameter. The choice among these approaches is largely a matter of what the analyst finds most convenient, except in cases for which the likelihood maximization procedure is very tedious or likely to fail entirely. Note that when the likelihood does not have a unique maximum (some parameters being completely confounded in their effects), the sampling approach amounts (deliberately or inadvertently) to assigning a Bayes prior probability distribution for the parameters, via at least the range of parameter values that are allowed in the sampling (e.g., a uniform prior distribution).

In our experience, it is very easy to waste time in the analysis of parameter uncertainty just to end up producing a collection of pretty probability distributions or confidence region maps that are hopelessly overoptimistic due to the failure to consider structural uncertainty. There has been much concern in recent years about how to measure structural uncertainty in assessments and about how to use alternative quantitative criteria for comparing alternative models and assumptions (Burnham and Anderson 1998; Helu et al. 2000; Patterson et al. 2001). Suggested approaches have mainly involved using simple statistical measures of prediction-error variance for alternative model structures, the Akaike (1973) criterion (AIC) or a very similar Bayes information criterion (BIC) suggested by Schwarz (1978).

$$AIC = -2\ln L_{max} + 2p \quad (5.8a)$$

$$BIC = -2\ln L_{max} + p\ln(n) \quad (5.8b)$$

where p is the total number of parameters treated as variable in the search for the maximum log likelihood $\ln L_{max}$, and n is the total number of ob-

servations. Unfortunately, as we will see in the next section, there are two things that can go wrong with using such simple and scientifically appealing criteria when the goal of assessment is to find the best model for *providing management advice*. First, they help identify the model structure that will likely result in a minimum short-term prediction variance, which is not the same as obtaining the best estimates of particular parameters that are important for policy. Second, the variance of a parameter estimate can be misleading when there are asymmetric costs of estimation errors (e.g., when it is much more costly to overestimate a key policy variable such as MSY than to underestimate it). Ultimately, the only way to fully evaluate a proposed assessment model compared to other models is to simulate its “closed loop” performance: simulate the full data gathering, estimation, and harvest regulation process. That is a very complex and tedious matter, and it can also give misleading results if the closed-loop simulation procedure inadvertently makes overly optimistic assumptions about the state dynamics and observation processes.

It is important to understand that uncertainty about population-dynamics parameter estimates does not necessarily imply corresponding uncertainty about the best harvest policy. For example, if the goals of management imply that a fixed harvest rate should be maintained (see chapters 3 and 4), the estimates of the best harvest rate depend only on those population-dynamics parameters that affect per-capita productivity (e.g., the population intrinsic rate of increase “ r ” in surplus-production models, or the slope of the mean stock-recruitment relationship in age-structured models), and not at all on the parameters that define absolute population scale (e.g., the carrying capacity K of the logistic production model, or the mean asymptotic recruitment at high stock size). If the goal of assessment is to provide a reference point for overall harvest (e.g., MSY) that is a function of several population parameters, this function may be more or less uncertain than the component parameters. For example, in surplus-production models with rate and scale parameters r and K , the calculated MSY is roughly (depending on the shape of the assumed production function) $rK/4$. Thus, MSY appears to be poorly estimated since it is the product of two uncertain parameter estimates. But in fact, the estimates of r and K are often highly negatively correlated (we can explain the data equally well as having come from a large, unproductive stock or a small, productive one), and in this case the product rK can end up being quite precisely estimated.

5.4 MODELING OPTIONS

How complex should the state-dynamics model in figure 5.1 be, given options ranging from simple surplus-production equations (e.g., 5.1) to detailed age-structure accounting, and given various possible forms for observation

relationships (e.g., proportional, hyperstable, hyperdepleting)? As noted in previous sections, there are two very different ways to answer this question:

1. **Scientific-statistical**—use a model that is optimally complex with respect to the available data (use the most parsimonious model that will do the “job”);
2. **Policy relevance**—use whatever model is most likely to provide the best policy advice and/or to stimulate a search for better policy options.

What question (1) is really about is how to maintain one’s credibility with scientific colleagues by appearing to be a knowledgeable statistical analyst and avoiding such mistakes as overparameterization. What question (2) is about is how to open doors for people who make policy decisions to compare their choices and to identify new choices that may be better than any used in past management, whether or not such choices can be immediately justified through a statistical analysis of the historical data.

Were policy relevance the only criterion for deciding assessment-model complexity, we would immediately recommend that every assessment model contain at least age/size-structured accounting (to allow for the exploration of policies that reduce the risk of growth and recruitment overfishing by avoiding the capture of juvenile fish). Considering that spatial policy options are becoming more popular in general (e.g., marine protected areas), we would even be tempted to recommend spatial models of the types discussed in chapter 8. But to move in these directions without careful thought about the aims of a particular assessment is as silly as to presume that model complexity should be dictated strictly by statistical concerns about overparameterization relative to the available data.

As noted in the previous section, one (difficult) way to compare alternative models for assessment is to simulate their “closed loop” performance when applied by a simulated fishery manager to a reference or “operating” model of realistic biological and observational complexity. Comparisons of this sort were first published by Hilborn (1979). He demonstrated the very surprising result that it can often be best from a management perspective to deliberately base assessments on an oversimplified model that is known to be wrong, if that model gives precise (low variance) estimates of policy parameters with a bias that is in a direction that helps avoid overfishing. Similar results have since been obtained using more elaborate likelihood estimation methods (Ludwig and Walters 1985), demonstrating that Hilborn’s findings were not an artifact of the particular methods he used for simulating statistical analysis of the modeled data. Figure 5.2 shows an intuitive way to understand such results, involving a hypothetical comparison of three estimation methods: one that is biased in a costly direction, one that is less precise but is biased in a safe direction (as was the method studied by Hilborn 1979), and a third that is unbiased but realistic in terms of its representation of the sources of uncertainty (and hence prone to produce even more variable parameter estimates). In this example, the first method is clearly

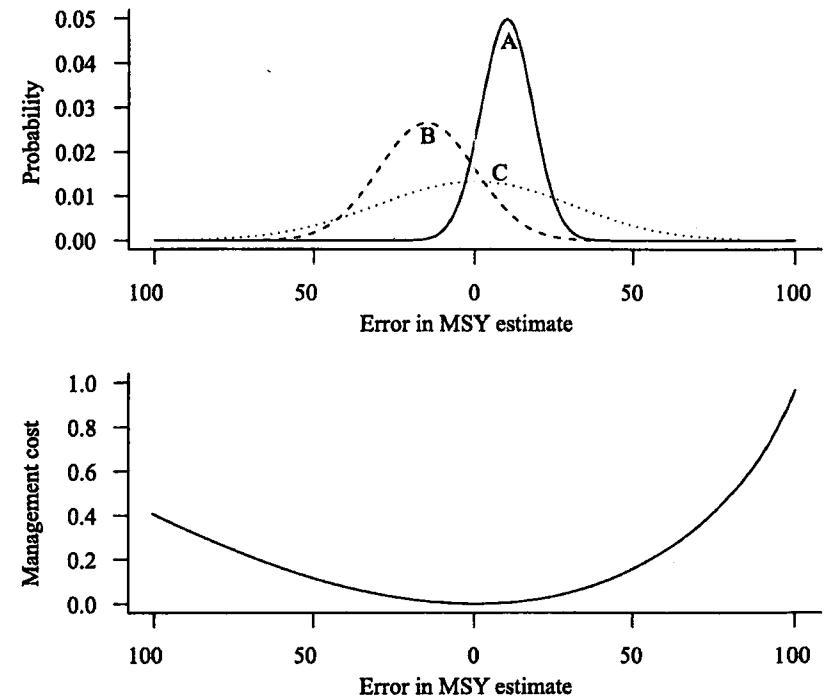


FIGURE 5.2: Variability and bias in the estimation of a key policy parameter (MSY) from three possible assessment models. Model A has very few parameters and so gives precise estimates, but biased upward in a way that would frequently result in a high management cost (due to overfishing) if the method were used in practice. Model B has more parameters and so gives a less precise estimate of MSY (admits more possible values of MSY given the data) that happens to be biased downward so as to give high-cost overestimates only rarely. Model C is a complex, fully “honest” representation of variability in the data; it is unbiased but even more imprecise than model B. Note that like Model A, Model C results in a relatively high probability of overestimating MSY and, hence, leads to a high management cost.

irresponsible: it results in a high probability of overfishing. The distinction between the other two methods is more subtle and becomes apparent only when the methods are applied many times (when MSY is estimated many times, as in multiple simulations to evaluate expected performance). The unbiased model with its realistic admission of uncertainty ends up with the lower average performance (and the higher average cost of applying it) because it more often results in high-cost (overestimation of MSY) outcomes.

An example of the precision-accuracy trade-off shown in figure 5.2 occurs in interpreting catch and effort data from the “one way trip” (Hilborn and Walters 1992) fishery developments shown in the simulated example

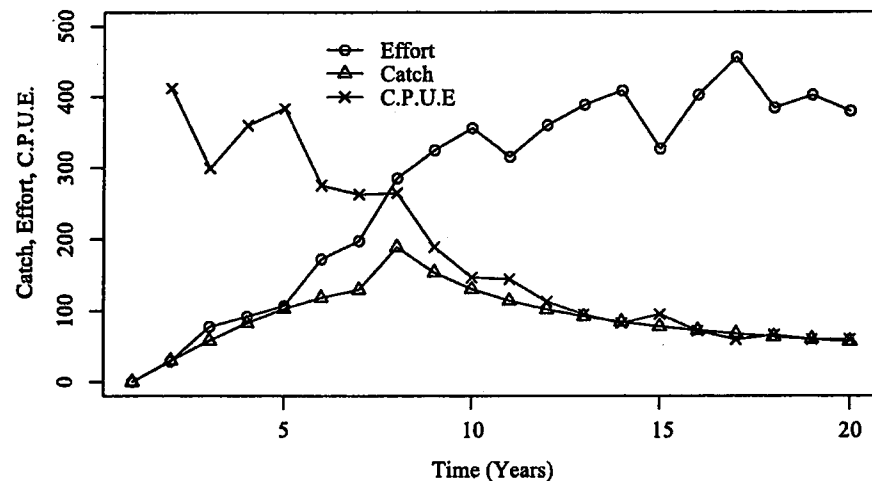


FIGURE 5.3: A simulated one-way-trip fishery development in which effort builds up enough to cause a severe decline in catch per effort, and apparent overfishing, over a 20-year period. Note that units are arbitrary in this example.

(fig. 5.3). Such data sets have typically been analyzed using simple surplus-production models and observation assumptions, e.g., the logistic biomass (B_t) model:

$$B_{t+1} = B_t + rB_t(1 - B_t/K) - qE_tB_t^\beta \quad (5.9)$$

$$Y_t = qE_tB_t^\beta \quad (\text{catch}) \quad Y_t = qB_t^\beta \quad (\text{catch per effort}) \quad (5.10)$$

Here the parameter β represents the possible changes in catchability with stock size ($\beta < 1$ implies hyperstability in the catch per effort, $\beta = 1$ implies proportionality, and $\beta > 1$ implies hyperdepletion). It is easy to demonstrate with simulated data that this model has far too many parameters (B_1, r, K, q, β) to be estimated uniquely from one-way-trip data. To reduce the number of parameters, we have typically assumed $B_1 = K$ (the stock initially near the unfished equilibrium) and $\beta = 1$. The “Gulland method” (Gulland 1961), which has been applied very widely in the developing world, invokes making an additional, even more restrictive assumption that development has been slow enough so that biomass has remained near equilibrium with respect to catch, i.e., $B_{t+1} = B_t$. In that method, the state equation 5.10 is replaced with the assumption $rB_t(1 - B_t/K) = qE_tB_t$, and a simple two-parameter linear regression relationship ($\text{CPUE} = a - bE_t$, $a = K/q$, $b = q^2K/r$) is used to estimate the effort E_{MSY} that would produce the maximum sustained yield on average (as the effort that would result in CPUE equal to $\frac{1}{2}$ the regression intercept; efforts are typically averaged so as to supposedly reduce the effect of the erroneous $B_{t+1} = B_t$ assumption).

Now, suppose we generate 100 20-year data sets as in figure 5.3 that exactly satisfy the state equation 5.10, i.e., no process errors or effects of population structure, and provide three alternative assessment procedures with simulated CPUE data where $Y_t = qB_t^\beta e^{v_t}$, v_t = normally distributed with standard deviation 0.2 (a typical amount of unexplained variation in the catch rate). Let “Procedure A” be the Gulland method, let “Procedure B” be a general procedure based on figure 5.1 with the correct state equation (5.10) but assuming $B_1 = K$ and $\beta = 1$, and let “Procedure C” be the same as Procedure B except being “scientifically honest” by admitting uncertainty about β (in Procedure C, the conditional maximum likelihood estimates corresponding to equation 5.5 of $\ln(q)$ and β are the slope and intercept, respectively, of the linear regression of $\ln(Y_t)$ on $\ln(B_t)$; see Walters and Ludwig 1994). Since β cannot generally be estimated uniquely under Procedure C, suppose we provide the likelihood function with an additional term $-(\beta - 1)^2/0.3$ representing a normal prior-probability distribution for β with a mean of 1.0 and a variance of 0.15 (the metaanalysis results of Harley et al. 2001 suggest we should assume this variance, but a less optimistic mean of 0.8).

Suppose we then judge the performance of the three procedures not by how well they fit the data (an award that most often goes to the Gulland method), nor by AIC or BIC criteria (which most often indicate Procedure B to be best), but by two management criteria: (1) how well they advise management about the best long-term fishing effort (for some objective like MSY) and the effort in year 20 relative to this optimum, and (2) how robust they are at continuing to provide good advice under the dangerous hyperstability case $\beta \ll 1.0$. As shown in figure 5.4, it is obvious that the precise and simple Procedure A throws the baby out with the bath water: it results in a dangerous upward bias in the estimates of optimum and current effort/optimum for all β , which just get worse and worse for more hyperstable CPUE cases. Surprisingly, Procedures B and C do not consistently result in overestimates of the optimum effort for hyperstable CPUE cases as might be expected, given that the simulated CPUE decline in these cases does not reflect the severity of actual stock decline; however, both procedures grossly underestimate the current (year 20) exploitation rate relative to its true value in the hyperstable cases, where catchability increases greatly at low stock sizes. The surprise about Procedures B and C is due in part to the way we generated the simulated data and is a warning about how the details of a simulation protocol for evaluating management procedures can affect the results: we used the same fixed, known exploitation-rate history for each simulation run, then calculated different effort sequences (depending on β) needed to have caused this fixed history; this protocol results in lower mean efforts for simulations with lower β and, hence, lower efforts needed to explain a CPUE decline. Despite a favorable protocol, Procedure C results in a higher variance in the optimum effort estimates among the 100 trials, meaning it has a somewhat

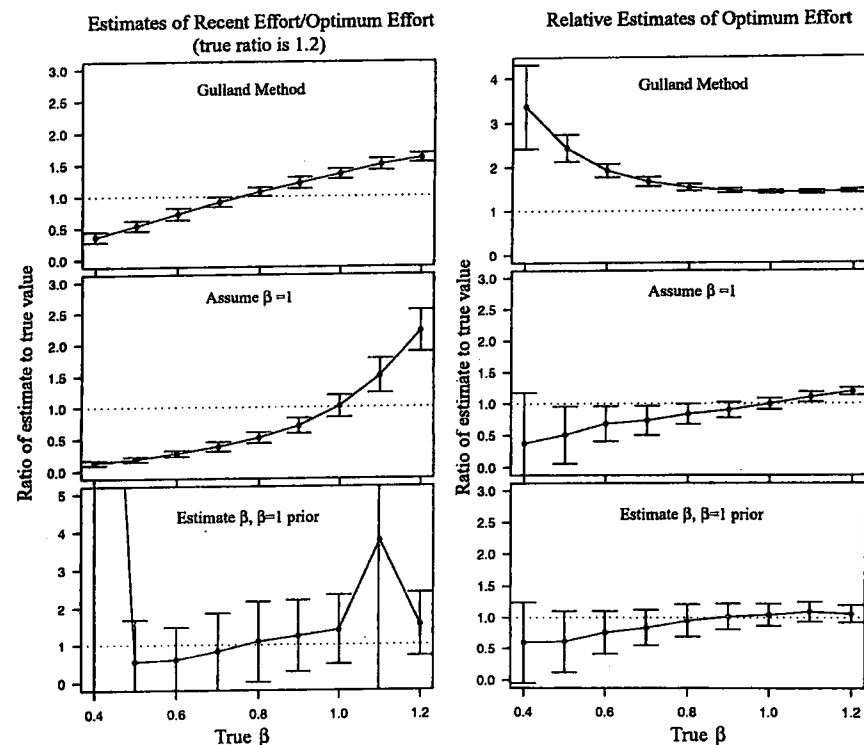


FIGURE 5.4: The simulated performance of three procedures for estimating optimum fishing effort and current exploitation rate for simulated data sets similar to figure 5.3. For each true value of the hyperstability parameter β , the results show the mean and range of estimates for 100 simulated development histories. Note that low values of the ratio of estimated to actual exploitation rates mean that the procedures underestimate the current impact of fishing.

higher probability of resulting in overestimates of the optimum effort (i.e., of leading to a recommendation that will result in overfishing). That is, it is similar to Model C in figure 5.2. If we change the prior distribution for β in Procedure C to have a mean of 0.8, to be more faithful to the empirical results of Harley et al. (2001), the procedure actually ends up doing worse in the sense that its bias is reduced for intermediate $\beta < 1$ cases, but it is much more likely to result in an overestimate of the optimum fishing effort.

In the above example, we have fixed the data available for assessment, then applied progressively more realistic models to these data. It is comforting in such examples that statistical measures such as AIC and BIC tend to agree with management-performance simulations in predicting that the optimum model should be one of intermediate complexity. However, this does not by any means complete the discussion about simple versus complex models: what happens to conclusions about optimum model complexity when we can

not only compare simple versus complex models but also include additional “auxiliary” data to help estimate parameters for the more complex model options? In particular, even in poorly monitored fisheries there is often some information on changes in the average size of fish caught, and average size is often a good indicator of changes in mortality and/or recruitment rates. The simplest population-dynamics model that can predict such changes in the average size is the “delay-difference” model of Deriso (1980) and Schnute (1985, 1987), which exactly predicts changes in the vulnerable biomass (B_t) and vulnerable numbers (N_t) for age-structured populations, in which fish have an age-independent total survival rate S_t (the product of natural survival rate times 1-exploitation rate), grow over ages a according to the Ford-Brody body-weight equation $w_{a+1} = \alpha + \rho w_a$, and are equally vulnerable to fishing after some age k (see Hilborn and Walters [1992] for derivations):

$$B_{t+1} = S_t(\alpha N_t + \rho B_t) + w_k R_{t+1} \quad (5.11)$$

$$N_{t+1} = S_t N_t + R_{t+1} \quad (5.12)$$

$$\text{Predicted mean body weight at time } t: \bar{w}_t = B_t/N_t \quad (5.13)$$

To generate surplus production dynamics with a recruitment overfishing risk under this relatively simple age-structured model, we must assume some functional relationship between mean recruitment R_{t+1} and past biomass(es) B_t, B_{t-1} , etc., e.g., the Beverton-Holt form $R_{t+1} = k_1 B_{t-1}/(1 + k_2 B_{t-1})$ if fish recruit at age $k = 2$. Even if we can accurately estimate the natural survival component of S_t and the growth parameters α and ρ from independent analyses of size and composition data, we apparently still need to estimate or assume values for at least the following parameters: B_1, N_1, k_1, k_2 , plus any observation-model parameters (q and β in this case for representing the observation submodel for relative abundance). In this model, the recruitment-function slope parameter k_1 (along with the body-growth parameters) plays the role of “ r ” in the surplus-production models: it determines the intrinsic rate of population increase when abundance is very low, and is a key determinant of the optimum exploitation rate. The recruitment parameter k_2 plays the same scaling role as the carrying-capacity parameter K , determining the absolute maximum recruitment rate and, hence, the overall population size in the absence of fishing. We can use these facts to eliminate B_1 and N_1 from the estimation set if we are willing to assume the age-structured analog of $B_1 = K$ (the stock was initially at unfished equilibrium), i.e., we can calculate B_1 and N_1 from the other parameters for the unfished equilibrium case. But even under all these restrictive assumptions, we are still left with at least 3 or 4 parameters to estimate from time-series data: k_1, k_2, q , and possibly β .

Suppose we repeat the performance simulations described above for the logistic-production model but replace the state-dynamics with the delay-difference system (eq. 5.11 and 5.12) and include noisy annual measurements (c.v. = 0.2) of average body weight \bar{w}_t in the likelihood function. This results in a spectacular improvement in both the precision and the apparent

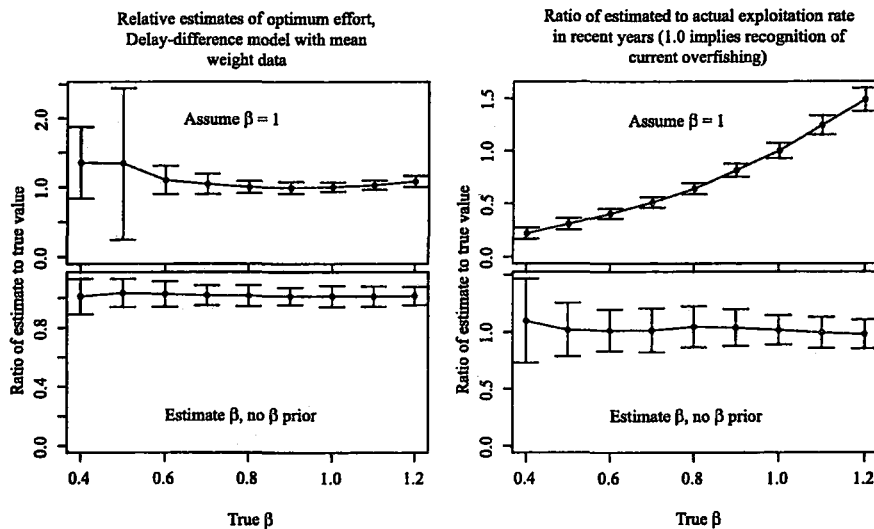


FIGURE 5.5: Performance simulations similar to those shown in figure 5.4 on page 106, but with an assessment procedure based on a delay-difference age-structured model that predicts (and is provided data on) changes in the average size of fish caught over time.

robustness of the assessment to hyperstability in the CPUE data (fig. 5.5). The procedure is able to estimate β quite well, and in fact, the simpler procedure with β assumed equal to 1.0 displays very obvious signals of failure when β is actually small, in the form of “contradictory data” (the model cannot predict both the correct CPUE trend and the mean-body-weight trend at the same time). For readers who worry about process versus observation errors, it is easily demonstrated that the procedure continues to perform quite well even if we include substantial process errors in the form of log-normal variations in the recruitment rates (c.v. 0.5), without even modeling these errors explicitly using one of the more elaborate likelihood formulations.

Simulation studies reported in NRC (1998) involved generating artificial data sets of increasing realism (violating simple model assumptions), then applying various age-structured stock-assessment methods including the delay-difference model (eq. 5.11–5.13) to these artificial data. Surprisingly, but as might be expected from the results in figure 5.5, the delay-difference model performed almost as well as the much more complex assessment methods discussed in the next section (at least at reconstructing stock-biomass histories), despite using only simple data on changes in average body size rather than full catch-at-age data.

Unfortunately, simulation tests as shown in figure 5.5 or reviews in NRC (1998) might well be hopelessly optimistic about an assessment performance with real data. Trends in the average size of fish measured from catches are actually influenced by four main factors: (1) total mortality rates (higher

rates lead to smaller average size), (2) persistent recruitment trends (downward trends lead to larger average size), (3) changes in size selectivity due to fisher targeting practices (can change the average size in unpredictable directions), and (4) sampling procedures (it is difficult to collect a random sample of fish sizes from even one vessel, let alone a whole fleet). Factor (2) can easily mask the effects of fishing that should be evident via the effects of fishing on factor (1), if there have in fact been downward recruitment trends not correctly captured through assumed stock-recruitment relationships, e.g., due to persistent environmental “regime shifts.” Factor (3) can also mask any mortality and recruitment effects, especially if fishers target smaller fish as larger ones become less abundant during stock declines (higher catches of smaller fish can easily be misinterpreted as evidence of recruitment increase). Factor (4) is not just a sample-size issue: when field personnel are allowed to grab fish “at random” for measurement from holds, buckets, or sorting trays, larger fish almost always end up being overrepresented in the sample in ways that can mask the impacts of changing mortality (small fish slither to the bottom of containers, larger fish come to hand first, etc.).

It is very easy to fall into the trap of worrying about simple versus complex models from only an ecological perspective (i.e., what state dynamics submodel to use), and indeed, we have done this in the paragraphs above by emphasizing comparisons of surplus-production versus age-structured assessment models. We have acknowledged problems with observation submodels (assumptions about the data) by examining the effects of hyperstability in the relative-abundance indices and by making a few points about the difficulties of interpreting average-size/size-composition data and about effective sample sizes in multinomial composition sampling. And the examples certainly support the usual assumptions that an “optimum” model complexity depends on what data are available for analysis and that adding additional types of data (and corresponding complication in models to predict these types) can, in principle, help correct for bias problems caused by incorrect assumptions. But it is hard to escape the intuitive feeling that all we are doing in such examples is chasing assumptions with more bad assumptions, grounded largely in wishful thinking about both population dynamics and available fisheries data. This is not just an issue in relation to the analysis of fishery-dependent data; we should probably be equally suspicious about fishery-independent survey data. Such data commonly show interannual variances far higher than would be expected on the basis of variability among observations within surveys and may be prone to hyperstability problems in catch rates due to incorrect definitions of sample survey areas (frames) relative to changing fish distributions.

In a variety of recent assessment exercises, we have found it productive in terms of both data analysis and policy development to abandon the simpler state models (e.g., eq. 5.10–5.12) in favor of general age-structured models initialized using the leading parameter structure discussed in box 3.1 on page 56. In that structure, we must provide a collection of age schedules

(vulnerability, body size, fecundity), an assumed form of mean stock-recruitment relationship, and two key leading parameters representing population scale (unfished biomass B_0) and recruitment performance at low stock size (compensation κ parameter), which play the roles of K and r in simpler models. Assuming that the age schedules can be estimated through an analysis of even fragmentary historical data on the fecundity and size/age compositions of catches (such data are almost always available these days), the age-structure model can still be used to predict relatively simple observations including the catch per effort and the mean size of fish harvested. With fixed vulnerability and growth schedules, such models behave essentially the same (in terms of estimation performance) as delay-difference models. But they have the added advantages of (1) options to explore the impacts of past and future changes in age-selectivity patterns even if such changes were not precisely measured; (2) the representation of smoothing effects on the recruitment of partial vulnerability patterns; and (3) options to examine the impacts of possible past changes in growth patterns, e.g., possible genetic impacts that selective removal of larger fish may have had early in the fishery's development (Conover and Munch 2002).

5.5 USING COMPOSITION INFORMATION

Most major fish stock assessments today are based on the analysis of age- and size-composition data. Such data have been painstakingly collected since at least the late 1970s to early 1980s, when age-structured assessment models first became widely available and showed great promise for improving our ability to reconstruct the historical impacts of fisheries. The basic idea in such models is to combine information on trends in relative abundance from fishery or survey catch rates, with information on changes in age and/or size composition. While some models attempt to compare observed versus predicted size compositions directly, in most cases we use estimated age compositions obtained by combining the sampling of ages with larger size-composition samples (using length-age keys to infer ages from sizes). Table 2 shows a typical catch-at-age data set for the main cod stock off Newfoundland (2J3KL or "Northern cod" stock, Baird et al. 1992).

Such catch-at-age tables generally show four basic kinds of obvious "treatment effects": (1) age selectivity, with young fish usually appearing to be less vulnerable than old ones; (2) yearly exploitation, with higher overall catches in some years than others; (3) cumulative mortality, with older fish ultimately being less numerous than younger ones; and (4) recruitment or cohort, with some year-classes or cohorts (down diagonals of the table) being stronger than others. The published equations used for representing these effects often look very complex, with exponential terms and such for dealing with the overlap in the time of fishing and natural mortality losses over each year. We generally obtain about the same results with a very obvious and simple

TABLE 2: Catch numbers at age (in thousands) used for stock assessments on the Northern cod stock (2J3KL) off Newfoundland (data from Baird et al. 1992).

Year	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Sum
1962	301	8666	26194	64337	58163	58163	27321	20142	18036	10444	9468	7778	5785	4669	3888	3055	2161	233	403	319457
1963	1446	5746	27577	60234	118112	48996	29349	15520	11612	8248	4204	3942	2913	2928	1737	1263	1352	328	182	355709
1964	2872	19338	27603	57557	60681	100147	50865	20892	12264	8698	6352	4989	4036	2703	1456	1918	1154	501	312	384538
1965	85	5177	28709	46800	66946	64360	68176	33819	14913	6945	3759	3948	3730	2722	1859	575	971	183	226	351873
1966	819	14057	65992	93687	62812	59312	30233	23844	8762	4538	3280	1825	1186	967	806	416	279	486	178	376559
1967	790	15262	77873	100339	96759	54996	38491	17146	16084	3367	2108	1529	1334	685	424	193	107	72	211	432983
1968	288	6142	94291	205805	150541	83808	39443	23171	10984	5591	5249	1939	1334	818	610	127	89	83	26	630339
1969	59	4330	39626	100858	163228	107509	52661	19651	12370	6389	4479	3004	1557	622	567	319	100	46	99	517424
1970	6819	18104	60102	82857	101249	85696	29218	10857	3825	2000	1200	507	224	214	244	124	32	10	34	402816
1971	33	12876	71357	95384	98111	57825	25053	11732	4370	2323	1387	1140	720	355	474	124	128	148	78	382760
1972	236	32877	79809	116552	76106	55944	29553	11750	6393	2987	1660	1388	728	728	606	452	136	195	16	392153
1973	0	3963	40785	94944	59503	35444	27351	14155	7566	3815	2153	1173	450	278	309	82	37	38	8	261963
1974	473	3231	13201	24927	74403	60359	35587	18854	10492	5818	2952	1078	652	249	338	162	113	45	20	263376
1975	420	3968	14101	25570	34436	39105	36985	13921	7314	2215	1179	808	372	165	82	5	8	22	1	179262
1976	15	13767	33727	28049	20858	16811	16022	10951	4637	1462	631	292	251	100	50	40	64	30	20	147797
1977	108	7128	65310	40462	21017	5397	3396	2730	1381	332	296	149	73	42	21	20	14	2	6	139576
1978	0	1323	15336	39206	26319	7711	3078	1530	1083	437	219	105	62	40	21	7	8	2	7	92714
1979	0	2354	12025	28014	20016	10982	3460	1300	757	360	183	116	51	31	38	7	7	4	0	97725
1980	92	2185	7172	13191	24800	22014	11848	3175	779	309	195	125	48	37	13	10	14	4	10	98755
1981	0	2102	31286	19003	14397	25435	16950	11936	1923	338	156	90	153	40	12	20	5	5	5	83918
1982	0	1702	13616	42602	19028	12044	14701	8934	6341	1018	248	90	41	29	11	9	6	2	3	123418
1983	18	2585	14871	31760	38624	12503	7246	8910	4227	2536	451	146	48	41	30	7	4	3	3	122199
1984	3	782	14824	36614	33922	28006	7050	3856	5162	2905	1681	254	107	39	20	17	1	3	5	133096
1985	1	831	13219	44168	43869	26023	14722	3104	2000	1977	1101	574	116	29	18	11	9	2	2	135778
1986	42	2329	9217	32490	49061	28469	19505	5818	1346	676	873	391	200	37	22	3	1	4	0	130334
1987	25	2779	14651	20184	47917	43725	18608	9026	4337	774	422	366	223	100	32	5	10	5	5	165194
1988	8	1096	17659	21150	25212	38708	28499	8696	3640	1695	572	244	180	94	43	4	9	0	1	148090
1989	58	7693	40557	36410	22695	16390	17940	9156	2865	1084	478	103	98	36	25	8	7	1	0	155604
1990	35	3111	31634	53805	29553	9064	6164	4745	1696	641	250	88	39	21	9	3	2	2	0	140882

equation for changes in the number of fish of each age (a) each year (t), based on pretending that the harvest is taken in a short season at the start of the year:

$$N_{a+1,t+1} = s_a(N_{a,t} - C_{a,t}). \quad (5.14)$$

Here, s_a is a (possibly age-dependent) survival rate of the age- a fish that are not caught, and $C_{a,t}$ is the catch. In the model state equations, $C_{a,t}$ can be treated either as a known removal each year (model “conditioned on catch”) or as a predicted output using an equation of the form

$$C_{a,t} = v_{a,t} U_t N_{a,t}, \quad (5.15)$$

where $v_{a,t}$ is a relative vulnerability of age- a fish in year t to the overall exploitation rate U_t for that year. When $C_{a,t}$ is predicted from U_t , we may in turn either try to estimate each U_t as an unknown parameter, or try to predict it from some measure of fishing effort or area swept by gear in year t . The s_a are almost never treated as variable over time unless many tagging data are available and, in most cases, are treated as age independent ($s_a = s = e^{-M}$, where M is an instantaneous natural mortality rate). It is not possible to estimate the s_a or M from catch-at-age models unless the exploitation rates U_t can be well predicted (with strong contrast in the data) from the fishing effort or area swept.

Patterson et al. (2001) provide an excellent review of case studies and alternative approaches that have been taken to reduce the number of parameters for a practical estimation using equations similar to equation 5.14, and for measuring uncertainty in the resulting estimates. From a statistical and logical perspective, these equations admit far more unknowns than measurements. If the catch-at-age table has data for A ages over T years, then it appears that we need to estimate $A \cdot T$ vulnerabilities $v_{a,t}$, T annual exploitation rates U_t , and $A + T - 1$ recruitments and initial abundances $N_{a,1}$ and $N_{1,t}$. Even when we can add considerable additional information in the form of age-specific relative abundances $Y_{a,t} = q_a N_{a,t}$ from surveys that have stable age-selection patterns, the parameter set still needs to be severely restricted to obtain estimates with meaningful (less than infinite) uncertainty. One popular way to simplify the parameter set has been to assume a “separation” of v and U effects, i.e., assuming $v_{a,t} = v_a U_t$; we warn against this assumption, since about the only cases for which it appears to have worked well are a collection of severely overfished stocks in the North Atlantic. For most fisheries, the $v_a U_t$ assumption is likely to result in misleading stock reconstructions due to temporal changes in the way fishers have targeted fish of different sizes and ages. Other options for reducing the number of parameters to be estimated, particularly the $v_{a,t}$, depend on how the state-dynamics equations are solved over age and time.

There are two apparently quite distinctive ways to solve or propagate the state equation 5.14 (fig. 5.6). In “stock synthesis” or “statistical catch-at-age (SCA)” models, we solve equation 5.14 forward over times and ages

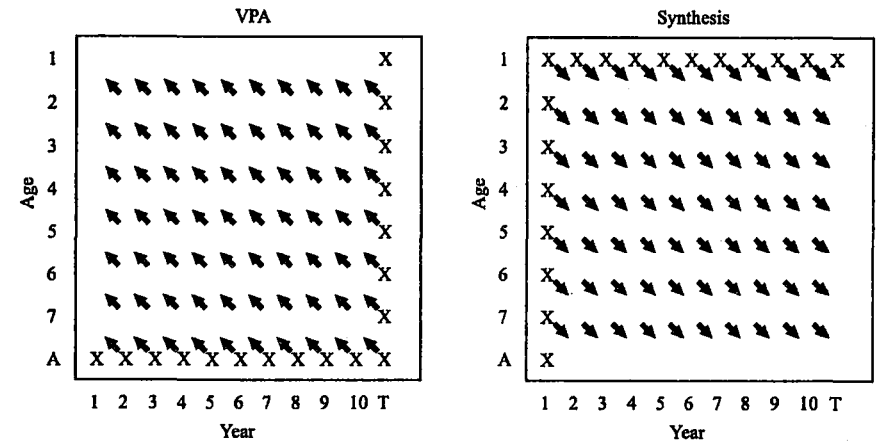


FIGURE 5.6: Assessment models for the reconstruction of historical population changes using catch-at-age data can be solved either backward in time (VPA case) or forward in time (Synthesis case). The flow of the calculations for each of these methods is shown by the arrows, and X's show unknown initial states that must be estimated somehow in the assessment procedure. Note that for the VPA case, we can often assume the X's representing the oldest fish every year to be zero.

just as equation 5.14 shows, either subtracting the observed catches as fixed removals or predicting them from the vulnerability and U_t parameters. In “virtual population analysis (VPA)” models, we propagate the equations backward in time by solving equation 5.14 for $N_{a,t}$, given $N_{a+1,t+1}$, as

$$N_{a,t} = N_{a+1,t+1}/s_a + C_{a,t}. \quad (5.16)$$

There are two main differences between these approaches: (1) error propagation, and (2) which $N_{a,t}$ need to be estimated as boundary-condition parameters (or perhaps predicted from a stock-recruitment relationship in the SCA case), as shown by X's in figure 5.6. Also, VPA cannot be used when there are many missing years in the catch-at-age data table. In the VPA case, errors in the estimation of $C_{a,t}$ are propagated with expanding effect backward in time (small errors at time t are successively divided by numbers less than 1.0, the survival rates, as the calculation proceeds backward over ages), whereas such errors decay over time under forward propagation. But balancing this possible pathology in error propagation, there are usually far fewer boundary $N_{a,t}$ values to be estimated with VPA, since numbers in the oldest age $N_{A,t}$ (where A is the oldest age) can usually be assumed to be zero. Further, except when the exploitation rate is very low (less than 0.1), the numbers in all ages in the last year ($N_{a,T}$) can often be modeled in the VPA setting as

$$N_{a,T} = C_{a,T}/(v_{a,T} U_T). \quad (5.17)$$

Especially if many of the $\nu_{a,T}$ can be assumed to equal 1.0 (older fish being equally vulnerable to the harvest rate U_T), the equation 5.17 “terminal harvest” method for setting $N_{a,T}$ can again substantially reduce the number of parameters to be estimated (to U_T and a subset of $\nu_{a,T}$). Note that when equation 5.17 is used in VPA, only the year- T terminal vulnerabilities $\nu_{a,T}$ need be specified as unknown parameters, since all other $\nu_{a,t}$ ($t < T$) are implied by the backcalculated exploitation rates $C_{a,t}/N_{a,t}$.

Though the origins of SCA methods can be traced to Doubleday (1976), by historical accident much of the recent development of these methods has been by scientists working on the Pacific rim of North America (e.g., Deriso et al. 1985; Fournier and Archibald 1982; Methot 1990; Schnute 1994). In contrast, much of the development of VPA has been by scientists working around the Atlantic, particularly the widely used software ADAPT (Gavaris 1988). It says much about the culture of stock-assessment science, and perhaps about whether scientists really follow the literature, that a recent NRC stock-assessment panel (NRC 1998) found that almost all assessments for Atlantic fisheries were being done with VPA, while almost all Pacific assessments were being done with SCA models. There was much consternation when test-simulated data sets were distributed widely to stock-assessment scientists on both coasts by the NRC panel, and ADAPT was found to sometimes perform poorly compared to SCA methods. This came on the heels of retrospective studies showing that precursors of ADAPT had helped contribute to the Newfoundland cod collapse through gross overestimates of stock size and recruitment (Walters and Maguire 1996). The net result of these findings has been something of an intellectual stampede from VPA- to SCA-based assessments and, in particular, to the use of the AD Model Builder software (Otter Research 1994) that provides extremely efficient and reliable nonlinear search procedures for fitting nonlinear statistical models with large numbers of nuisance parameters (e.g., vulnerabilities $\nu_{a,t}$).

Unfortunately, there has been a recent tendency to condemn VPA methods in general, on grounds ranging from error propagation to the particular bias problems of ADAPT, without regard to the obvious advantages of using a method for which relatively few parameters need be estimated. Indeed, some workers even appear to take pride in announcing that they are fitting an SCA with more than 100 parameters. They seem to believe either that admitting more unknowns and uncertainties is a virtue in itself (honesty) or that having more parameters is sure to result in “better” estimates of important leading parameters such as the mean recruitment rate and recruitment trend (by accounting for more sources of variation in the data). There has been much retrospective examination of cases in which the much simpler VPA approach has been found to give very misleading stock reconstructions and estimates of recent abundance, i.e., the Newfoundland cod stock. In these cases, the main problem has not been error propagation or some other intrinsic bias

at all but, rather, the two specific difficulties mentioned at the start of this chapter about observation-submodel assumptions:

1. changes over time in the vulnerabilities at age, $\nu_{a,t}$, leading to poor estimates of the terminal vulnerabilities $\nu_{a,T}$ used in the terminal numbers calculation equation 5.16; and
2. fitting (or “tuning”) the models using relative-abundance data that is not proportional to stock size, in particular the use of hyperstable fisheries CPUE data.

(Even VPA does not escape the need to use relative-abundance data in estimations; without information on recent abundance trends, there is no way to decide a best estimate for the terminal U_T in eq. 5.16). Exactly the same problems have plagued SCA in some cases, e.g., the Pacific halibut stock (Parma and Deriso 1990a) in which there has been a large change in the $\nu_{a,t}$ associated with changes in body growth patterns. Both SCA and VPA are prone to give misleading results in any case in which the $\nu_{a,t}$ are changing rapidly in the most recent few years of data, e.g., during stock declines when fishers often start to target smaller fish; in these cases, there is no way in principle of using only the catch-at-age data to decide whether catches of younger fish have been sustained because of relatively high recruitments or because of relatively high $\nu_{a,t}$.

Neither SCA nor VPA methods have lived up to our early expectations about how valuable catch-at-age data would be in providing useful information for management. Consider for a moment the three most important “products” that we hoped they would provide: (1) estimates of the historical exploitation rates $U_{a,t} = C_{a,t}/N_{a,t}$, to evaluate trends in fishery impact and overfishing risk; (2) estimates of trends in recruitment, both to help detect recruitment overfishing and to provide information for scientific studies of the causes of recruitment variation; and (3) estimates of current stock size and age composition, to use in implementing harvest strategy rules. Both approaches are fine for meeting objectives (1) and (2), provided we look only at the results for $t \ll T$ (and in SCA, do not assume away much of the $U_{a,t}$ variation by assuming a constant $\nu_{a,t}$ over time). Both approaches are dangerously misleading for objective (3) when the $\nu_{a,t}$ may be changing rapidly and when we cannot be certain that available age-specific relative-abundance indices $Y_{at} = q_{a,t}N_{a,t}^\beta$ are not in fact hyperstable (i.e., when we cannot be sure that $\beta = 1$ and $q_{a,t} = q_a$ independent of t). But these are just the cases in which good management advice is most important!

Even for the analysis of long-term trend patterns, SCA methods have not turned out to be such a wonderful generalization of early catch-curve techniques (plots of catch over age to obtain total mortality estimates) as we had hoped. They do not really lead to a clear separation of the four types of “treatment effects” mentioned at the start of this section. Changes in

catches over age at any time still exhibit the confounding effects of changes in selectivity, past recruitment trend, and past effects of fishing on mortality rates. Changes in catches over time and age (along the dynamic catch curve for each cohort) still have the confounding of effects of changing selectivity (with both age and time) and of changes in the exploitation rate, despite the subjecting of these observations to the same recruitment "treatment." In some cases, declines in vulnerability with age of older fish (which we usually assume away by treating all older fish as equally vulnerable) have likely masked even the crude trend signals of overall changes in the mortality rate due to fishing. About the only thing that SCA methods sometimes do well is to recover information on strong versus weak recruitments in successive years, something that VPA sometimes misses due to the "smearing" of fish across cohorts caused by aging errors in the $C_{a,t}$ assessment.

So here is our general advice to assessment workers who are worried about whether to use VPA or SCA methods. Use VPA whenever there are reasonably large catch-at-age samples for every year and in places where historical exploitation rates have been relatively high (at least 0.1), so as to take advantage of its compact parameterization, but pay very careful attention to the estimation of the terminal vulnerabilities $v_{a,T}$ and the terminal exploitation rate U_T . Use SCA methods only when there are large gaps in the data, when the $C_{a,t}$ data are highly suspect due to inadequate sample sizes, and/or when recent exploitation rates are suspected to have been very low. When in doubt about whether to trust the $C_{a,t}$ data or the U_t , try both approaches, with particular attention to whether the VPA results suggest changes in vulnerabilities $\bar{v}_{a,t} = (C_{a,t}/N_{a,t})/\bar{U}_t$ (where \bar{U}_t is the mean of $C_{a,t}/N_{a,t}$ over "fully vulnerable" ages a) that it has been necessary to assume away in order to obtain unique maximum likelihood estimates with SCA. For both methods, never pretend that the estimates for recent years ($t > T - A/2$) are accurate or reliable, despite what any statistical measures of goodness of fit or prediction-error variance (AIC, BIC) may suggest.

Some assessment scientists appear to believe that all we need do to make the SCA/VPA methods work well is to invest in collecting good survey data on age-specific relative abundances. This is a very naïve belief that has been used to justify some very expensive but still suspect survey programs (variances are higher than they should be, with possible hyperstability in the survey catch rates). Such advice needs very careful rethinking, to consider whether we should be going after the $v_{a,T}$, U_T problem with other more direct monitoring methods such as tagging programs.

A variety of models now attempt to deal directly with (to predict observations on) length-composition data, hence avoiding possible problems with the estimation of catch-at-age composition from the length data (Fournier et al. 1998; Fu and Quinn II 2000; Sullivan et al. 1990). New methods for analyzing growth from tagging experiments promise better estimates of growth curves (box 5.2). Unfortunately, there is a catastrophic flaw in these models. None of them explicitly represents the cumulative effect of fishing

BOX 5.2

ESTIMATION OF BODY GROWTH PATTERNS FROM AGE-SIZE AND TAGGING DATA

Assessments of body growth are critical in many aspects of fisheries assessment, from the interpretation of length distribution data to evaluation of effects of size limits/size selection on yield (yield per recruit analysis) to the prediction of natural mortality rates when such rates cannot be measured directly. Typically, growth in body length $l(a)$ with age a is described by models of the form $l(a) = l_{\infty}f(a)$, where l_{∞} is the asymptotic (maximum) body length and $f(a)$ is a monotonic increasing function of a , approaching 1.0 as a increases. The most common model is the Von Bertalanffy, for which $f(a) = 1 - e^{-K(a-a_0)}$, where the metabolic parameter K is a useful predictor of the natural mortality rate and a_0 represents an apparent age at the time of hatching. Other suggested forms for $f(a)$ mainly involve changes in K with age for young fish (e.g., Porch et al. 2002). Typically, we expect to see considerable variation among individual fish in the asymptotic length parameter l_{∞} but relatively little variation in the metabolic parameter K .

The average growth parameters l_{∞} and K for fish populations are typically estimated by assuming that representative samples of fish of each of a range of ages has been collected, then either fitting the growth function to these samples directly or to data on changes in the length of fish between the times of tagging and later recapture. Unfortunately, there are two potentially serious problems with the basic assumption of representative sampling: (1) in fact, fish are first collected, then are aged or tagged for later recapture, so that the proportions of fish at age a are not set beforehand "by design"; (2) most fish collecting methods are highly size selective, so that most often the fish of younger ages that do appear in samples are the fastest-growing (highest individual l_{∞}) individuals. Problem (2) is serious: it causes an upward bias in estimates of K and a downward bias in estimates of the mean l_{∞} . Recent work, particularly by Laslett et al. (2002), has built on earlier developments by Wang et al. (1995), Wang (1998), and others to show how to deal with the age-composition-sampling effects along with the individual variation in l_{∞} in maximum likelihood estimations. However, there has as yet been no general recommendation about how to cope with biases caused by size-selective sampling. Simulation tests with the likelihood function suggested by Laslett et al. show that it can provide nearly unbiased estimates of K from tag-recapture data even when the size at tagging for younger fish has been quite size-selective, provided K is not treated as variable with age.

One solution to the size-selectivity problem when only size-at-age data are available is to estimate a size-selection function, along with relative abundances at age, while estimating the growth parameters. Suppose the growth

(Continued)

(BOX 5.2 continued)

data consist of a matrix of observed numbers of fish by discrete length l and age $n_{l,a}$, and that these data represent a random sample from the vulnerable fish $V_{l,a}$ present in the population. Suppose further that $V_{l,a}$ is predicted to be $V_{l,a} = v(l)N_a W(l|a)$, where $v(l)$ is a simple (few parameters) length-vulnerability function, N_a is the relative total number of age- a animals in the population, and $W(l|a)$ is a normal probability density function (or integral over the width of the discrete length measurement interval) with mean $l(a)$ given by the Von Bertalanffy or other growth function (and typically standard deviation proportional to the mean length, implying only one parameter for describing the individual variation in growth). Under these assumptions, the likelihood for the multinomial sample of observed lengths can be written as

$$\log L = \sum_l \sum_a n_{l,a} \ln[V_{l,a}/V_T]$$

where V_T is the total vulnerable population ($V_T = \sum_l \sum_a V_{l,a}$). Parameters are then estimated by maximizing $\log L$ over the parameters of $v(l)$ (vulnerability), $l(a)$ and $W(l|a)$ (growth), and relative numbers at age N_a . We have done extensive simulation tests of this approach and find it surprisingly good at finding the underlying growth parameters even when $v(l)$ implies the sampling of only larger fish. However, here is an important caveat to these findings: the estimation can be fooled if an incorrect shape is assumed for the $v(l)$ function, e.g., it is assumed asymptotic when it is, in fact, dome-shaped.

There are at least three ways to deal with the abundance-at-age “nuisance” parameters N_a . The first is to assume a simple functional form for these, e.g., $N_a = e^{-Za}$ and include parameters of this form (e.g., Z) in the estimation. The second is to include the $\log L$ likelihood terms for each $n_{l,a}$ sample in the overall likelihood function for some stock-assessment model that includes the estimation of numbers at age (a stock synthesis or VPA model), so that the growth and the size-vulnerability parameters are estimated along with recruitments, etc. The third is to treat each N_a as an arbitrary nuisance parameter and eliminate these nuisance parameters by evaluating $\log L$ at their conditional maximum likelihood estimates whenever estimating the vulnerability and growth parameters. It is easily seen that by differentiating $\log L$ with respect to N_a and setting the derivative to zero, the maximum likelihood estimates of N_a must satisfy $N_a = n_a V_T / (v_a n)$, where n_a is the total number of age- a fish in the sample, n is the total number of fish aged, and $v_a = \sum_l v(l) W(l|a)$ is the mean vulnerability of age- a fish. Substituting this condition into $\log L$ (and dropping the terms involving sample sizes that do not affect the likelihood comparisons) results in the “reduced” likelihood function

(Continued)

(BOX 5.2 continued)

$$\log L' = \sum_l \sum_a n_{l,a} \ln[v(l) W(l|a)/v_a]$$

which depends only on the vulnerability and growth parameters. Simulation tests show that this reduced likelihood gives considerably more variable parameter estimates than can be obtained when it can safely be assumed that the N_a vary in some structured way with a . ■

on the size distributions of animals of different ages. The sampled length frequency at age is assumed to be distributed around some mean (possibly density-dependent) growth curve, but the growth curve is not represented as dependent on the fishing mortality rate. In reality, individual fish tend to be persistently either fast or slow growing, and the faster-growing individuals usually become vulnerable to fishing at earlier ages (i.e., each age cohort consists of growth type “subcohorts” with different growth curves and age-vulnerability patterns). The cumulative removal of these individuals results in a decrease in the mean size and growth rate of the remaining individuals. Thus the apparent growth curve, and length-frequency distributions at age, are likely to be strongly dependent on the fishing mortality rate and size-selection pattern (Sinclair et al. 2002a, 2002b), whether or not there is genetic selection over long periods for slow-growth types as reported by Conover and Munch (2002). Fishery-induced changes in the growth curves can, of course, be monitored on a year-to-year basis to avoid a bias in predicting length-frequency patterns (or an alternative approach can be used to predict length-composition data, see box 5.3 for details), but with the sampling effort required to do this one might as well use age-structured methods in the first place.

BOX 5.3

MODELING CUMULATIVE EFFECTS OF INDIVIDUAL VARIATION IN GROWTH

When we are trying to interpret historical size-distribution changes or predict impacts of size-selective harvesting, it can be misleading to assume that the mean and variance of size at age are independent of mortality rate. Individual differences in growth rates tend to be persistent, so, e.g., if growth is represented by the Von Bertalanffy equation $l_a = l_\infty(1 - e^{-Ka})$, we generally expect individuals to exhibit different maximum body lengths l_∞ but a similar metabolic parameter K (Wang et al. 1995; Laslett et al. 2002). This implies that if larger individuals are subject to higher mortality rates, the mean length at age of surviving fish will shift downward as the fishing mortality rate increases, and the distribution of lengths at age may not remain normal.

To simulate such cumulative effects, we somehow have to keep track of changes in both the size and growth patterns of surviving animals, over age and time. Moving to a pure length-transition probability model does not solve the problem, and moving to a full individual-based model (IBM) is computationally very costly. A simple accounting tactic that avoids most of the cost of an IBM approach is to divide each cohort or year class into a set of growth-type groups or subcohorts g , where the subcohort index g varies from, say, $g = -10, -9, \dots, 0, \dots, 9, 10$. Then for each of these subcohorts, construct a cohort-specific length-at-age table $l_{g,a}$, from a model like $l_{g,a} = (l_\infty + \Delta g)(1 - e^{-Ka})$, where Δ represents a length bin or interval width for the individual variation in l_∞ . Use a length-vulnerability function $v(l, t)$ to calculate a subcohort-and-age table of vulnerabilities to harvest, so that in model year t the fish of subcohort g and age a are assigned the exploitation rate $u_{g,a,t} = v_{g,a,t}u_t$, where u_t is an overall ("fully vulnerable") rate for the year. Also construct a table of weights and a length-weight relationship, and use these to predict a fecundity table $f_{g,a,t}$ so that the total egg production ϵ_t can be predicted each year by summing $f_{g,a,t}$ times the subcohort abundances $N_{g,a,t}$.

This approach simplifies the accounting of an individually based model by representing the stock as a collection of different populations that differ only in their asymptotic length L_∞ . Another way to think of it is that each of the growth-type groups comes from a family, or distribution, of asymptotic sizes, and size-selective fishing mortality is explicitly represented for each growth-type group g . Then the assessment model estimates the distribution of asymptotic lengths in the population (assuming that variation in length increases with age), rather than estimating standard deviations in length for each age class.

(Continued)

(BOX 5.3 continued)

Time simulations then proceed just as in age-structured models, but with each annual recruitment $R(\epsilon_t)$ distributed initially over the subcohorts and all survival calculations carried out over all subcohorts g as well as ages a :

$$N_{g,1,t+1} = R(\epsilon_t)p(g) \quad \text{all } g$$

$$N_{g,a+1,t+1} = s_{g,a,t}(1 - v_{g,a,t}u_t)N_{g,a,t} \quad \text{all } g, a$$

Here, $p(g)$ is a vector of proportions of the total recruitment assigned to subcohorts, with values most likely just proportional to a normal probability density. An important issue is whether to treat the annual natural survival rates $s_{g,a}$ as dependent on subcohort and age and, in particular, as lower for fish with higher growth rates ("Lee's phenomenon"); not including such variation is equivalent to treating growth rate as having zero heritability, since if the $s_{g,a}$ are independent of g , the contributions of fast growing and early maturing fish to ϵ_t will be differentially high.

Note that the useful equilibrium calculations for dynamic-model initialization, yield, etc. involving Botsford's incidence functions over age (box 3.1) apply equally well when there is a subcohort structure in the accounting. Simply include sums over the g index in all incidence function (ϕ) calculations. ■

5.6 DEALING WITH PARAMETERS THAT AREN'T

The things we call "parameters" in population-dynamics models are not, in fact, fixed physiological or physical properties of organisms but, rather, are the complex resultants of myriad fine-scale processes. For example, the slope of a mean stock-recruitment relationship (chapter 7) is a product of the short-time survival rates of juvenile fish through various early life-history stages or stanzas. The interactions that result in each survival rate involve habitat/predation/competition factors that may have a "random" variation but are also likely to exhibit persistent change (nonstationarity, Walters 1987) as well. Likewise, even scientific observation parameters including survey q and v_a are the results of a complex interaction between survey fishing gear and fish, subject to persistent changes (even if the gear is held constant) because of changes in size, behavior, and distribution of fish relative to the survey area.

In stock assessment we have tended to assume away persistent changes in order to get any answer at all (or have failed even to recognize the possibility of them). We have, instead, spent much time agonizing about how to measure and cope with the essentially trivial variation due to the random,

time-independent statistical effects caused by sampling and by singular ecological events such as the occasional production of strong year classes. This is not because it is wise or even necessary to be concerned about simple statistical variation but, rather, because it is what we know how to do. Further, assessment scientists have wasted much time in trying to counter mindless criticisms of the form “things are so complicated and there is so much variation out there that your models cannot possibly work,” coming mainly from people who barely passed introductory biometrics courses and hence have no clue about how to deal with variation of any kind. Of course, pure statistical noise can overwhelm any assessment technique, but that is not generally what happens. If you read over the warnings earlier in this chapter about things that can go wrong with assessment methods, you will see that most of these things involve persistent changes, i.e., parameters that aren’t.

There are at least two assessment tactics for trying to find out whether the available data are consistent with the hypothesis that particular parameters are not (i.e., that these parameters have changed over time in ways that would result in misleading assessments). The simplest tactic is to assume some functional form for the time dependence, then try to estimate the parameters of this form. A more complex approach is to treat the parameter(s) as state variables with incremental, random change-per-time steps in state-space likelihood formulations, then try to track these changes using state-space filtering and smoothing operations (Gelb 1974).

Plausible functional forms to assume in estimating parameter changes over time include (1) piece-wise constant, (2) polynomial (including spline functions), and (3) Fourier (sum of sine curves). The easiest of these to implement in the general estimation framework of figure 5.1 is the piece-wise constant option: divide the historical data period into reasonable blocks, and generate the state/observation predictions with different q ’s, r ’s, or whatever for each block. When change is likely to have been smooth over time, e.g., over a period of gradually changing gear or habitat alteration, an easily interpreted way to represent the main components of change via a polynomial function is to use the form $P_t = P_o + a_1(t - t_o) + a_2(t - t_o)^2 + a_3(t - t_o)^3 + \dots$, where P_o is the parameter value at some base reference time t_o (typically $T/2$), and the parameters a_1 , a_2 , etc. represent linear, quadratic, and higher-order components of temporal variation. Similar patterns interpretable as “low frequency” versus “high frequency” variation can be achieved with Fourier series. Since all of these functions are purely descriptive (with no particular biological or physical meaning or derivation), it is really a matter of convenience for the analyst about which is easiest to interpret and explain. It is not uncommon to see various functional approaches used in statistical catch-at-age (SCA) models, particularly to represent known temporal changes in the vulnerability patterns ($v_{a,t}$) due to gear changes (both survey and fishery) and regulations (e.g., Hampton and Gunn 1998).

Peterman et al. (2000) demonstrate using simulated data that state-space filtering methods may be a very useful way to track (recover information

about past) climate-induced changes in recruitment parameters. Several workers have suggested using state-space methods to track possible parameter changes in statistical catch-at-age models (Sullivan 1992; Schnute 1994; Millar and Meyer 2000).

For both the functional form and state-space approaches, it is generally necessary to make some a priori assumption about relatively how much of the variance in observations is due to possible parameter changes as opposed to observation errors and independent (uncorrelated) process errors. This means setting more or less restrictive prior-probability distributions on parameters of functional forms, and variances for random walk processes in state-space models. We cannot, of course, know ahead of time how restrictive to be. The best practical strategy may be to follow a “relaxation” approach, starting out with very restrictive assumptions and relaxing these to move more and more of the explained variation into an estimated change over time in the parameter values. This strategy will in effect “sketch” out a set of alternative hypotheses about how to interpret the historical data, and these hypotheses are quite likely to have very different policy consequences. In particular, admission of the possibility of nonstationarity in the stock-recruitment parameters means admission that any observed stock declines might have been due to two very different mechanisms: environmental changes/regime shifts or recruitment overfishing. So scientists who attempt to be honest about alternative ways to explain the available data should be aware that their results will contribute to the “Thompson-Burkenroad” (see chapter 7) or “environment vs. fishing” debates that surround almost every major fishery decline. Perhaps some honest modeling will help show that such debates should be fundamentally changed, from arguments about what happened to arguments about how to manage the risks and costs implied by not being able to decide which hypothesis is correct on the basis of the historical data. It is pointless to argue about what the future will bring in such situations, and the wise thing to argue about is what policy to follow given that either outcome is equally likely (equally consistent with available historical data).