Stock Assessment: Operational Models in Support of Fisheries Management

Chapter · February 2009

DOI: 10.1007/978-1-4020-9210-7_9

1 author:

Richard D. Methot
National Oceanic and Atmospheric Administration

Chapter 9
Stock Assessment: Operational Models in Support of Fisheries Management

Richard D. Methot Jr.

Abstract Fishery stock assessment models connect ecosystem data to quantitative fishery management. Control rules that calculate annual catch limits and targets from stock assessment results are a common component of US Fishery Management Plans. Ideally, the outcome of such control rules are updated annually on the basis of stock assessment forecasts to track fluctuations in stock abundance. When the stock assessment – fishery management enterprise achieves this level of throughput, they truly are operational models, much as the complex physical models used to routinely update climate forecasts. In reality, many contemporary assessments are closer to an individual scientific investigation than to an operational model. As a result, the review of each stock assessment is extensive and the lag between data acquisition and quota adjustment may extend to several years. If the future stock assessment process is to move towards an operational status, there will need to be changes in three aspects of the process. First, key data streams will themselves need to be made more operational and corporate so that relevant data are immediately available and trusted. Second, stock assessment models need to be made more capable of including diverse relevant data and comprehensively calculating levels of uncertainty, while also being more completely tested, documented, and standardized. The class of models called integrated analysis has these characteristics and is described here, with emphasis on the features of the Stock Synthesis model. Areas of future model development, especially to include more ecosystem and environmental factors, are explored. Third, increased throughput of assessment updates will require streamlining of the extensive review process now routinely required before stock assessment results can serve as the scientific basis for fishery management. Emphasizing review of broadly applicable assessment data and methods, rather than each final result, is a logical step in this streamlining, while maintaining public trust in the final results.

Keywords Fish stock assessment · population dynamics · fishery management

R.D. Methot Jr.
NOAA Fisheries Service, Seattle, WA 98112, USA

R.J. Beamish and B.J. Rothschild (eds.), *The Future of Fisheries Science in North America*, 137

9.1 Introduction

Modern fisheries stock assessment models (Quinn 2003) are the nexus between our growing scientific capability to understand factors affecting the population dynamics of harvested fish stocks, and the expanding demands for quantitative management of their fisheries. These models assimilate a diverse collection of data, produce forecasts linked to historical estimates, and provide a framework for comprehensive evaluation of model uncertainty and risk assessment for proposed management actions (Maunder et al. 2009; Schnute et al. 2007). Assessment models are increasingly able to incorporate spatial structure (Punt et al. 2000) and the influence of environmental and ecosystem factors (Maunder and Watters 2003). Multispecies stock assessment models are beginning to appear. This paper describes the role of fish stock assessment models in providing ongoing quantitative advice for fishery management and provides an overview of the rapidly evolving capabilities of a class of assessment models termed statistical catch-at-age analysis or integrated analysis, with particular emphasis on the Stock Synthesis model (Methot 1989, 2000). Areas of future model development, including increased linkage to environmental and ecosystem data, will be explored.

Quantitative management of US marine fisheries received a large impetus with the 1976 Magnuson Fishery Conservation and Management Act. The 1996 reauthorization of the Act required fishery management plans to prevent overfishing, rebuild previously overfished stocks, and obtain optimum yield from the fishery (NOAA 1996). Quantitative criteria related to the reproductive potential of the stock were required to: gauge the occurrence of overfishing, determine whether a stock was overfished (depleted) and in need of rebuilding, forecast potential rates of rebuilding for previously overfished stocks, and guide management towards a level of catch that will produce optimum yield but is no greater than the level that would produce maximum sustainable yield. NOAA responded with an update to guidelines for implementation of the Act and subsequent technical guidance (Restrepo et al. 1998). The 2006 reauthorization of the MSA upped the ante further by requiring establishment of annual catch limits in each fishery such that overfishing does not occur, and that these annual catch limits be based upon the scientific recommendations of the Fishery Management Council's Scientific and Statistical Committees or an established peer review process. In addition, as more fisheries are managed with individual fishery quotas, the demand for more precision in total quota determinations will only go up.

The need for expansion and improvement of the stock assessment enterprise (NRC 1998) led to development of the Marine Fisheries Stock Assessment Improvement Plan (Mace et al. 2001) by NOAA's National Marine Fisheries Service. The plan identified three tiers of improvement: (1) mining existing data to extend basic assessments to as many stocks as possible; (2) expanding data collection and assessments to provide adequate assessments for major fish stocks and at least baseline monitoring of minor stocks; and (3) reaching to an ecosystem level of assessment for representative stocks in each region. In addition to direct investment in data collection and assessments, NMFS initiated programs such as the Sea Grant Fellowship in Population Dynamics to train new assessment scientists, the Stock Assessment Toolbox to provide a standardized interface for many assessment models, and the Center for Independent Experts (Brown et al. 2006) to increase the rigor of assessment reviews.

The fact that fish are components of marine ecosystems and are influenced by ecosystem and environmental factors is not news (Hjort 1914). Yet, over the mid–late twentieth century, the fisheries assessment community evolved methods that analyzed data often solely collected from the fishery itself and which incorporated simplifying assumptions that left little room for direct incorporation of environmental and ecosystem factors. The focus on empirical description of the state of the stock was a logical outcome of the need to provide quantitative criteria to enable science-based fishery management decisions. Although such models have provided useful short-term guidance regarding adjustments in fishery regulations and the sustainability of a general level of fishery catch, their black-box nature made them poor candidates to serve as direct tools to understand and investigate the non-fishery factors that also influence the abundance of fish stocks. In parallel, fisheries science continued a strong emphasis on studying factors that affect the growth, mortality, and reproduction of fish stocks in an ecosystem context, but opportunities to directly incorporate this growing body of knowledge lagged.

Fish assessment models do not ignore the fact that fish stocks respond to environmental and ecosystem conditions; they just treat it as a reaction to be measured but not predicted. For example, empirical measurements of annual body weight-at-age (i.e., growth) are directly incorporated as detailed data into many age-structured assessment models. Fishery and survey age composition data are used by the models to estimate the annual level of recruitment of young fish into the population. But these methods for dealing with environment-caused variations in growth and recruitment are entirely empirical. The results, particularly for recruitment time series, provide input for subsequent investigation of possible environmental causes of the fluctuations, but it is only recently that stock assessment models have begun to include the environmental information directly as an additional source of information about the fluctuations (Maunder and Watters 2003; Schirripa and Colbert 2006). Some ways in which environmental and ecosystem information can improve fish stock assessments will be explored later in this paper.

9.2 Stock Assessment Overview

A stock assessment is the collection, analysis and reporting of demographic information to determine the effects of fishing on fish stocks (Mace et al. 2001). There are three basic categories of information that must be provided in order to produce accurate assessments: total catch, abundance trend, and life history characteristics. Deficiencies in one cannot be overcome by excessive data in another.

First, there must be accounting of the total catch. Historically, assessments would use the landed commercial catch. For fisheries of interest, this was usually the dominant component of the total catch and was the component that was most completely available for the entire time series. While simple assessment models require only catch biomass as a sufficient description of the fishery impact, more detailed models require that the catch be broken down into catch numbers-at-age to more precisely assign age-specific mortality and to calculate the time series of annual recruitments that must have occurred in order to have supported this catch (Pope 1972). As fisheries have evolved and scrutiny of their total impact has increased, so has the requirement for complete catch accounting. Today's assessments use total catch by fleet, including commercial and recreational landed and discarded catch in target fisheries and bycatch in other fisheries. Further, studies of discard mortality are used to calculate the total mortal catch. Biological samples characterize the age/size/gender of the catch. Fleets are separated to provide the ability to calculate the differential demographic impact of fleets that principally harvest younger/smaller versus older/larger fish. Where this demographic sampling level is high, the resultant time series of fishery catch-at-age is an influential source of information on fishery removal patterns and the time series of recruitment to the fished stock.

Second, there must be some measure of abundance. Assessments that lack a measure of the level, or at least trend, in stock abundance will not be able to achieve confident results (NRC 1998). Ideally, there will be a time series of survey observations, each calibrated to provide an absolute estimate of stock abundance. Absolute calibration is difficult to attain and the typical goal is to have a time series of observations that track the relative trend in stock abundance. From a statistical perspective, this relative abundance trend is best obtained from a fishery-independent survey so that the sampling protocols can be highly standardized and applied over the range of the stock in a statistically-based sampling strategy. Some surveys have been conducted by a single vessel, or its directly calibrated replacement. Others have relied upon use of multiple chartered vessels and have absorbed the added variability of between vessel variability into the total variability of the survey results (Helser et al. 2004). In some cases, basic fishing technology such as longlines or bottom trawls have been adapted and standardized for use as a survey sampling tool. In other cases, specialized technologies have been developed such as hydroacoustics deployed from Fisheries Survey Vessels, egg and larval surveys, or underwater imaging systems deployed from Remotely Operated Vessels or even Autonomous Underwater Vehicles (ref this book). As with the fishery catch, there should be sufficient demographic sampling of the survey catch to describe the life history segment of the total population that is being monitored and to provide more detailed information on the trends in abundance by age and size. Fishery-independent resource surveys conducted from larger, multiple capability vessels are also a valuable platform on which to piggyback ecosystem observations.

Standardized, fishery-independent surveys are not available for many stocks and the fallback is to use a proxy measure of average fish density per unit area calculated by statistical processing of fishery catch rate (CPUE) data obtained from logbooks or observers (Maunder and Punt 2004). While such CPUE data can appear highly precise due to the thousands of fishery logbook observations included in some analyses, the shortcoming is the inability to confidently assert that the units of fishing effort can be standardized over each year of the time series to the degree that fishery-independent survey methods are standardized. Increasingly, assessment methods are able to relax the assumption of constant catchability to make best advantage of fishery CPUE data (Bence and Wilberg 2006). The topic of time-varying catchability will be explored further in the model section.

Third, there should be sufficient information regarding the stock's life history characteristics. Although biomass-dynamics models operate with just a time series of catch and catch per unit effort, basic age-structured models require the capability to determine fish age from some biological structure such as otoliths, fin rays, or scales; and a measure of body weight-at-age and natural mortality (usually assumed constant across the age range available to the fishery). Because the goal is to analyze the impact of the fishery on the reproductive potential of the fish stock, a normal additional requirement is information on percentage mature at age. Spawning biomass so calculated is still a crude measure of reproductive potential and a more accurate measure will take into account fecundity-at-age and even larval quality if it differs by age (Bobko and Berkeley 2004). Some species such as hermaphrodites and nest-breeders require information on the male's contribution to reproductive potential. A major challenge is maintaining sufficient sampling over time to track changes in growth and maturity. This is especially important if these changes are density-dependent or have long-term trends due to environmental or other factors.

One highly influential factor, natural mortality, is more ecological than biological. Biological measurements of individuals may indicate a fish's relative predisposition to predation, parasitism, disease, and other causes of natural mortality or may provide measurements of growth and reproductive factors that appear correlated with natural mortality. However, natural mortality itself is the average probability of death from non-fishery causes, so is not directly observable from individual fish. Most direct estimates of natural mortality have been obtained by sampling the age composition from pre-fishery periods or lightly fished components of the stock, but such estimates do not directly measure the natural mortality occurring in the current, fished component of the stock.

From the catch, relative abundance, and life history information, assessment models can infer the abundance of the population that must have existed in order to exhibit the observed trend in the abundance indicator while producing the observed level of catch. An adequate assessment should provide an estimate of the time series of stock abundance and fishing mortality and an analysis to determine sustainable levels of fishing mortality and the resultant expected level of catch and stock abundance. The accuracy and precision of the results depends on the quality and quantity of the data, and also on the characteristics of the stock's history. If there is little contrast in the time series of catch and relative abundance, then a wide range of combinations of average fishing mortality and average stock abundance may be consistent with the available data. But if the stock has been monitored through at least one major cycle of lowered abundance and subsequent rebuilding, then more precise estimates of stock abundance and productivity can be obtained from stock trend and absolute catch data. There are two corollaries to this situation: the maximum productivity of a newly fished stock cannot be well determined until it has been fished at a moderate level for a sufficiently long time, nor can the rebuilding target of an overfished stock be well estimated if significant monitoring does not begin until after the stock has already been depleted to a low level of abundance. However, if fishery-independent surveys can employ technologies that are directly calibrated to provide measures of absolute stock abundance, rather than a relative trend, then just a few years of surveys provides immediate stock assessment information regardless of the level of fishery exploitation.

9.3 Scientific Advice for Fishery Management

The results of stock assessments serve as the basis for long-term and short-term fishery management decisions. First, the assessment provides the basis for status determinations. These status determination criteria are specified in regional fishery management plans guided by the National Standard 1 Guidelines of the Magnuson-Stevens Sustainable Fisheries Act and technical guidance. Loosely they entail: (1) determining whether overfishing is occurring by comparing the current level of fishing mortality to a limit level that is based upon the level that would produce maximum sustainable yield; and (2) comparison of current reproductive potential (usually measured just as spawning biomass) to a limit level (usually set to approximately half the level that would produce maximum sustainable yield) as a measure of stock depletion and a trigger for development of a rebuilding plan. Second, assessments provide forecasts of the expected future catch and stock abundance associated with proposed harvest policies. Thus they provide the basis for calculation of the expected time period for rebuilding of previously overfished stocks and for implementation of the harvest policy that will produce optimum yield from the fishery. Finally, the time series of abundance, mortality, and productivity produced by single-species stock assessments provide input to ecosystem food web models. Indeed, the multi-decadal stock assessment results are among the most quantitative and well-documented results available for such ecosystem models.

A single stock assessment can provide sufficient information to serve as the basis for a one-time status determination and for setting fishery management targets. However, stock assessments are also expected to serve as a core component of an ongoing fishery management system. Status determinations need to be updated, rebuilding of overfished stocks needs to be tracked, and catch levels need to be adjusted to maintain fishing mortality targets. Achievement of these additional goals means that assessments must be updated frequently to track changes in stock conditions due to natural and fishery factors. In effect, they become part of the operational model used to provide fishery management advice. Harvest control rules serve to translate stock assessment forecasts into target and limit levels of fishery catch (Fig. 9.1). The term, operational model, distinguishes such assess-

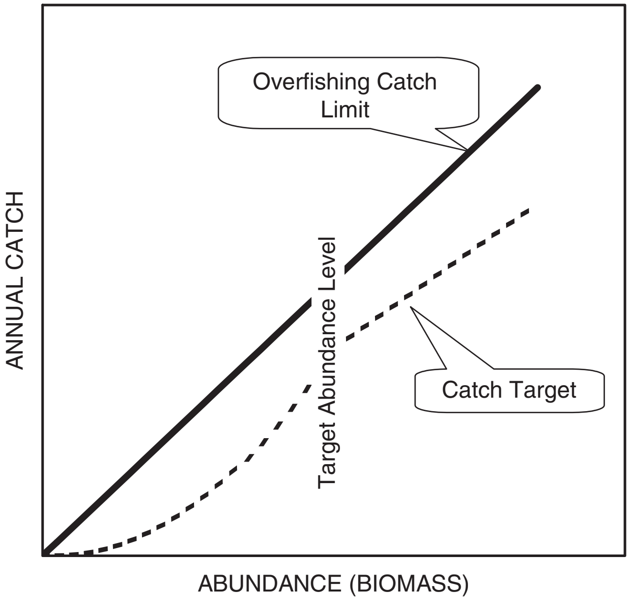

Fig. 9.1 Harvest control rules calculate limit and target levels of catch from short-term forecasts of stock abundance. In this hypothetical example, the target is a smaller fraction of the limit at lower levels of stock abundance in order to guard against further depletion

ments from one-time, stand-alone scientific investigations. An operational model provides timely updates to inform a set of clients of rapidly changing conditions. For example, the frequently updated forecast of the path and intensity of a hurricane is calculated using a complex model of the system calibrated with data collected on the time scale of hours. The results, including probability distributions, are rapidly produced and disseminated to the public in an easily understood graphical format. The fishery assessment operational model is identical in concept. The major differences being that the fishery biological model is less well-understood than the physical model used for hurricane forecasting; the time scale is months–years, not hours–days; and the output of the fishery model feeds directly into a regulatory framework rather than a public information framework.

The requirement for fishery assessments to serve as part of an operational model for management of US marine fisheries has increased with passage of the Magnuson-Stevens Fishery Conservation and Management Reauthorization Act of 2006. By 2010, Fishery Management Plans must specify annual catch limits for each fishery, based on scientific advice and at a level such that overfishing does not occur. Clearly, quantitative stock assessments are key to implementation of these requirements. Stock assessments can provide estimates of the level of catch that would be considered overfishing, and can provide a probability distribution for the chance of overfishing relative to a range of possible annual catch limit levels (Prager et al. 2003).

It is feasible for a simple model that only tracks recent trends in stock abundance to serve as the basis for adjustment of fishery target levels of catch. However, such a simple model is not well suited to assure that the target level is itself correct, nor is it well suited to integrating multiple current data sources into a forecast of future stock conditions. Further, when affected constituents see a simple model's inability to track some observed trends in the data, the importance of environmental and ecosystem factors are usually invoked, and the demand for more complex analysis begins. Although a model should not be more complex than necessary to assimilate the available data, the best solution is not necessarily to start with a simple model and to move to a more complex model as data allow. Instead, a smoothly scalable approach is to build a fully detailed model and to collapse its details down to the level that is estimable with the available data. Further, the complex model provides the framework for more comprehensively calculating the uncertainty in model results due to factors for which there are insufficient data.

9.4 Integrated Analysis Assessment Models

A class of models that has evolved over the past 25 years to meet the growing assessment challenge is termed integrated analysis or statistical catch-at-age analysis. A recent review can be found in Schnute et al. (2007). Such models were first developed in the 1980s (Fournier and Archibald 1982; Methot 1989) and began to see widespread use and rapid evolution by the late 1990s (McAllister et al. 1994; Quinn 2003). Integrated analysis models work as a simulation of the underlying population dynamics calibrated with the available data. They tend to cast the goodness-of-fit to the model in terms of data elements that retain the statistical characteristics of the raw data. This distinguishes integrated analysis from models such as Virtual Population Analysis that work more as a transformation of a particular type of preprocessed data, in this case fishery catch-at-age. Most integrated analysis models have an age-structured population dynamics sub-model, but the class itself is more general and includes strictly length-structured population models (Chen et al., 2005). The current generation of integrated analysis models has broad and flexible capabilities: age and/or length structured; spatial structure; environmental inputs; and other features that have them evolving towards multispecies models (Livingston and Methot 1998; Sitar et al. 1999). The general characteristics of integrated analysis models are described here, with particular emphasis on the features incorporated in the Stock Synthesis model.[1]

Integrated analysis models incorporate a linked set of sub-models (Fig. 9.2). The core sub-model contains the population dynamics. This is where the processes of birth, death, and growth create the time series of estimated population abundance

[1]NOAA Fisheries Toolbox Version 2.10, 2006. Stock Synthesis 2 Program, Version 2.00c. [Internet address: http://nft.nefsc.noaa.gov].

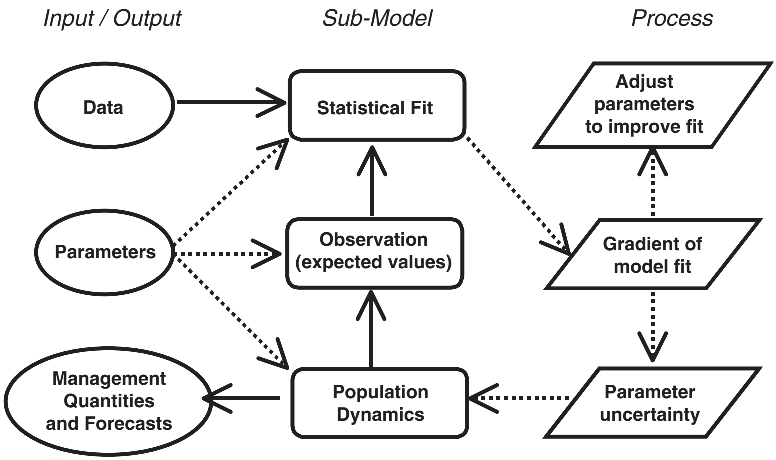

Fig. 9.2 Integrated analysis models consist of a linked set of sub-models: population dynamics, observation, and statistical. The statistical model compares expected values from the observation sub-model to the data and calculates the gradient of the goodness-of-fit with regard to the parameters in order to iteratively adjust the parameter values to achieve the best fit

and mortality. Some of these processes are represented by fixed input quantities and others are calculated from parameters being estimated by the model. For example, annual recruitment of young fish into the modeled population and annual fishing mortality by each fleet are usually calculated from model parameters. On the other hand, natural mortality is usually based on a fixed input value because data available to include in the model are not informative about the exact levels. Growth (e.g., body weight-at-age) is fixed in some models and estimated within others. As integrated analysis models evolve, there has been a move towards estimation of more parameters within the models and to utilize Bayesian methods to provide additional information in the form of a prior probability distribution for the value of the parameters. The estimation of more parameters has also evolved towards development of approaches that allow some parameters to have values that vary over time either as freely fluctuating quantities or as linkages to additional model inputs such as ecosystem or environmental factors.

Next is the observation sub-model where the processes of catchability, selectivity, aging imprecision, and other factors are modeled to create expected values for the types of available data. Like the population dynamics sub-model, the observation sub-model represents some processes with fixed inputs and others as relationships that incorporate estimated model parameters. This observation sub-model is where integrated analysis obtains much of its strength. Rather than preprocess and adjust the data so that it is in terms of the underlying population dynamics, integrated analysis models build knowledge of the observation process into the creation of expected values for the data. For example, the virtual population analysis model assumes that fish ages are determined without any error, so if the otolith reading process is known to have some variability between readers, the inverse of this reader variance should be applied to the age data before feeding the adjusted data into the VPA assessment model. When doing so, the variance associated with reader imprecision is not carried through to final management quantities. Besides, it is difficult to sharpen data that have already been blurred by the observation process. In integrated analysis, the opposite approach is taken. Information about reader imprecision is used to blur the expected value of model estimates of age composition so the model's estimates are blurred to the same degree that it is believed that the data have been blurred. Because the blurring process is built into the model, its effect on variance of model outputs is fully incorporated.

Third is the statistical sub-model where the goodness-of-fit to the data is calculated in terms of the negative of the logarithm of the probability of the observations, termed the negative log-likelihood or NLL for short. The NLL for each diverse type of data is basically calculated by scaling each observation's deviation from the predicted value according to the statistical form and magnitude of the error distribution for that data source.

This NLL basis means that the degree to which the model doesn't exactly fit each type of data is scaled in comparable terms. So, even though the model may contain tens of NLL components (age composition from fishery A, length composition from fishery B, % discard from fishery A, abundance from survey C, catch per unit effort from fishery B, recruitment index from survey D, etc.), the NLL components can be added together into a total NLL that is a meaningful measure of the total goodness-of-fit to all the data. Key to a successful model is inclusion of all relevant processes that have contributed to the observed data so that all the deviations are due to measurement error, and using the correct level of variance for these measurement errors. Of course, determining when the model is at the "sweet-spot" of complexity such that hidden processes are not misinterpreted as high measurement error is part of the art of model building.

Many integrated analysis models are written in the C++ computer language using ADMB, which was developed in the private sector by Otter Research (http://otter-rsch.com/) to facilitate the development of complex models. It employs automatic differentiation so that the gradient of the NLL with respect to each parameter's value can be calculated analytically, thus greatly speeding the iterative search for the set of parameters that maximizes the goodness-of-fit. When the gradient is large, this means that the parameter is influential and is relatively far from the value that maximizes the NLL. As the model searches for the set of parameter values that maximizes the NLL, it is searching for values at which the gradient goes to zero, hence where no further improvements can be made (Fig. 9.2). At that point, the model also calculates the curvature of the NLL surface with respect to the parameters. Where the curvature is strong, this means that small movements of the parameters away from the best fitting point will have large degradation in the NLL, thus meaning that the best value of the parameter is precisely determined. Where the curvature is weak, this means that the parameter's value has little effect on the NLL, which means that the data included in the model do not have much information about the best value of that particular parameter. It is not uncommon for the NLL surface to form a ridge with respect to a pair of parameters.

The strength of this ridge represents the correlation between these two parameters. The model may have tens to hundreds of parameters being estimated, so the multidimensional shape of the NLL surface is complex indeed. Because the curvature of the NLL surface is not necessarily parabolic and symmetric, as assumed by the normal distribution theory for obtaining confidence intervals, integrated analysis methods also use nonparametric approaches to calculate the shape of the NLL surface. In the Monte Carlo Markov Chain approach, after finding the best-fitting set of parameter values, the model then semi-randomly moves around the parameter space, each time calculating the NLL and building up an empirical representation of the multidimensional NLL surface.

The present state-of-the-art for assessment modeling routinely incorporates two forms of uncertainty in forecasts of stock abundance and potential fishery yield: (1) uncertainty in model parameters and current stock conditions based on goodness-of-fit between the model and the available data; and (2) expected future year-to-year fluctuations in productivity (recruitment) (Prager et al. 2003; Brodziak et al. 1998). These two components of variability may capture most of the total uncertainty, but there are other factors to consider: model structure, management implementation, and ecosystem factors.

Every model's structure is an approximation of the myriad of processes actually affecting fish stocks and creating the set of available data. Alternative models will make different assumptions and process the available data in different ways. The use of alternative models is important for understanding the basis for and robustness of any model's results. Considering a range of complex and simple model structures can clarify the additional insight that the more complex model provides as it incorporates a richer set of data. Where data are highly informative about stock conditions, good alternative models should produce similar results. But as the quantity and quality of data weakens, alternative model assumptions will have more influence on the results. Model-averaging and decision tables (Patterson et al. 2001) are two principal approaches to dealing with model structure.

Imperfect implementation of forecast catch levels is an additional factor to consider when conducting medium- or long-term forecasts of the stock's response to fishing. When the fishery is managed principally through input controls such as number of licenses or number of days at sea, the implementation error occurs because such measures are imperfect at holding fishing mortality to exactly the prescribed level. When the fishery is managed principally through the output control of quotas, then there is implementation error in controlling the fishery catch to that level and implementation error in forecasting the correct level of a future quota based on imperfect knowledge of stock abundance. The potential impact of these implementation errors are principally investigated through Management Strategy Evaluations that simulate the entire system of stock dynamics, imperfect assessment, imperfect management implementation, and feedback of actual catch to stock dynamics (Butterworth and Bergh 1993; Smith et al. 1999; Patterson et al. 2001).

An additional aspect of uncertainty is with regard to ecosystem factors. In particular, adult natural mortality and juvenile natural mortality (e.g., the mean spawner–recruitment relationship) are plausibly related to whole ecosystem conditions and it is conceivable that a multispecies ecosystem model will someday be able to estimate how they change and include this information in each stock's assessment. However, today's single-species assessment models have none of these data so these factors are assumed to be constant at some average level. A more complete characterization of the uncertainty in single-species assessments should seek a means to acknowledge the suppressed uncertainty caused by keeping these factors constant.

9.5 Generalized Model: Stock Synthesis

One of the highly generalized integrated analysis models is termed Stock Synthesis (Methot 2000), now implemented as Stock Synthesis 2 in the NOAA Fisheries Assessment Toolbox and popularly known as SS2. SS2 is a third-generation integrated analysis model. The first was developed in the mid-1980s specifically for assessment of anchovy off the coast of California (Methot 1989). The second was a generalized model developed principally for assessment of groundfish off the US west coast and Alaska (Methot 2000). It existed in two versions: one was a length-age-structured model developed for situations with predominately length data, and the other was an age-structured model with capability for movement between geographic regions. This third-generation model merges the length-age and age-area second-generation models, adds additional features, and is coded in ADMB to gain speed and powerful methods for evaluating uncertainty. There are three major aspects of SS2's adaptability that have contributed to is widespread use: (1) it is highly flexible in its ability to have multiple fisheries and surveys with diverse characteristics; (2) its parameters have a rich set of controls to allow prior constraints, time-varying flexibility, and linkages to environmental data; and (3) its structure allows it to be scaled down to simple, data-limited cases using only two estimated parameters, and up to complex data-rich situations requiring hundreds of parameters.

In the population sub-model of SS2, annual total recruitment is calculated as a deviation from an estimated spawner–recruitment curve, which in turn describes the central tendency of the time series of recruitments (Fig. 9.3). The magnitude of each recruitment deviation is informed by the data, including environmental data, in the model yet constrained by an overall distribution function so that estimates of historical, data-limited, and forecast recruitment deviations will have the same distribution properties as the recruitment deviations that are well informed by the data (Fig. 9.4) (Maunder et al. 2006). In this regard, SS2 performs similar to stochastic stock reduction analysis (Walters and Martell 2004), but SS2 also includes a full observation sub-model to make advantage of more complex data where it is available.

Growth of individuals is defined to follow a von Bertalanffy function, the parameters of which are estimable in the model when sufficient size and age data are included. In fact, SS2 could be configured to estimate only the growth parameters

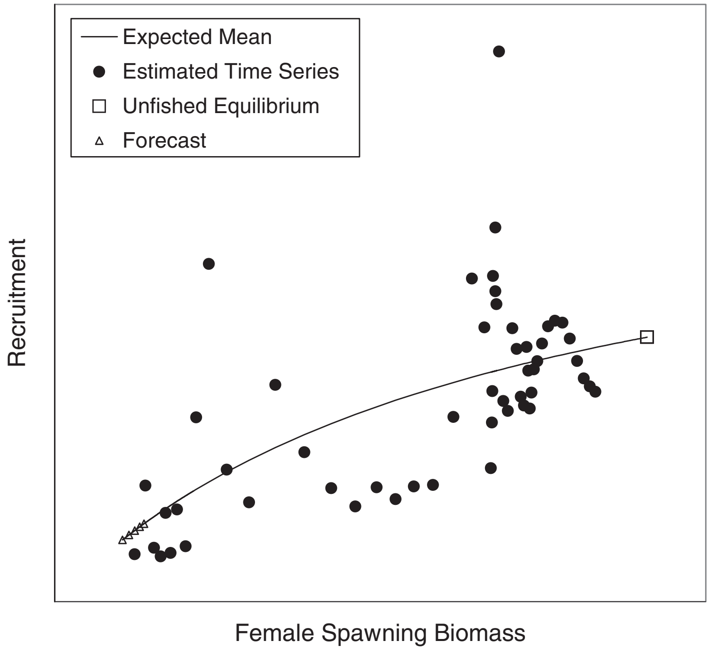

Fig. 9.3 Annual recruitment is defined as a deviation from the estimated spawner–recruitment relationship. The relationship defines the historical unfished equilibrium and provides a basis for calculating maximum sustainable yield and forecasting future levels of recruitment. In this example, during early years (high biomass) there are no data to inform the model about recruitment fluctuations, so estimated recruitments remain close to the mean relationship

and then calculate a yield per recruit analysis from a data set that had no catch and just a single size composition observation that had informative size modes. Of course, the full model capability is realized when it is allowed to estimate growth parameters from a time series of observations while taking into account the influence of size-selectivity sampling, aging error, and other processes that can otherwise bias estimates of growth. As with all SS2 parameters, the growth parameters can vary over time using a variety of methods including random annual deviations, separate parameter values for specified time blocks, and functional linkage to environmental data. For growth, an additional feature is the calculation of a year-class-specific growth deviation for situations such as abundant year classes having density-dependent suppression of growth.

SS2 tracks population numbers-at-age within each of several possible subdivisions, termed platoons (Goodyear 1984). A platoon in SS2 is a collection of individuals that share the same biological characteristics and probability of being captured by a fishery or observed by a survey. Each platoon is defined to have a normal distribution of size-at-age that interacts with the size selectivity of each fishery and survey to create the unique observed size-at-age for that fishery/survey (Fig. 9.5). While a single platoon model is feasible to configure, it is more common

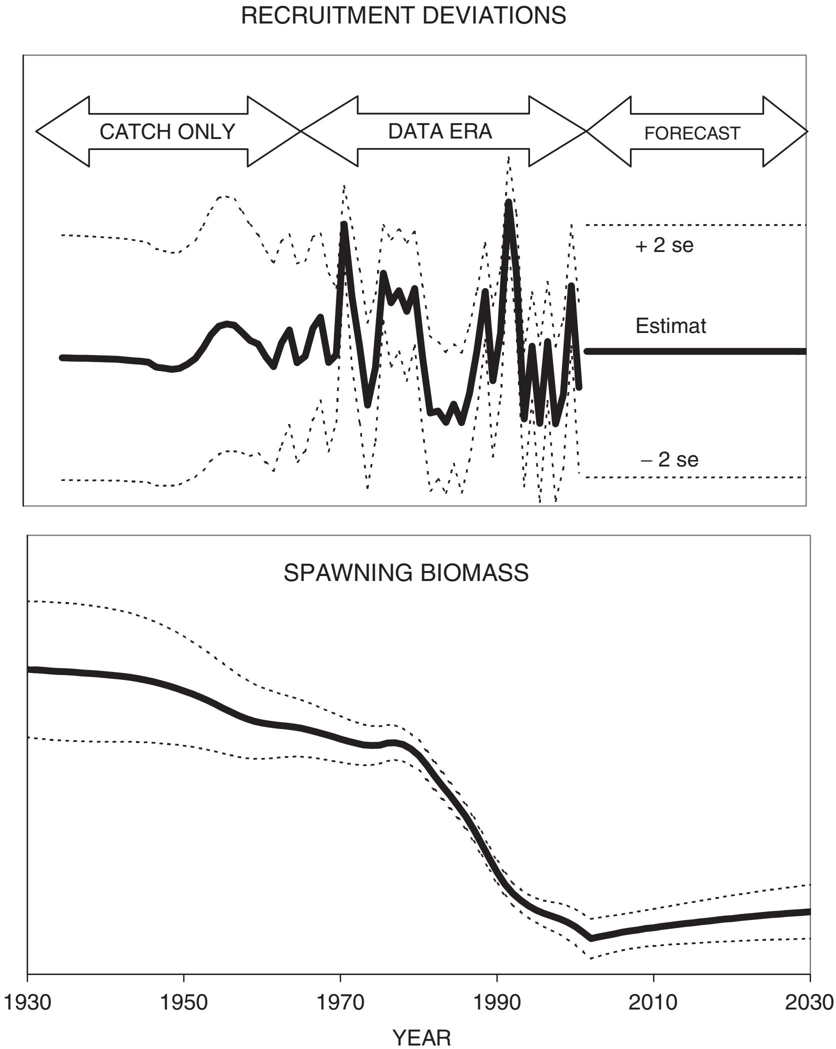

Fig. 9.4 The underlying spawner–recruitment relationship in Fig. 9.3 allows the model to produce population estimates, with variance, during far historical periods with no data other than catch, a data-rich era, and a forecast period. The transition between these periods is mostly transparent with the data phasing in and then back out

to partition the population into male and female platoons so that their dimorphic growth, mortality, and fishery selectivity characteristics can be calculated. When recruitment occurs in multiple seasons of the year, each such birth season adds platoons. In addition, it is possible to define multiple growth patterns, each with unique growth characteristics and receiving a fraction of the total number of recruits. A configuration with multiple areas and multiple growth patterns allows

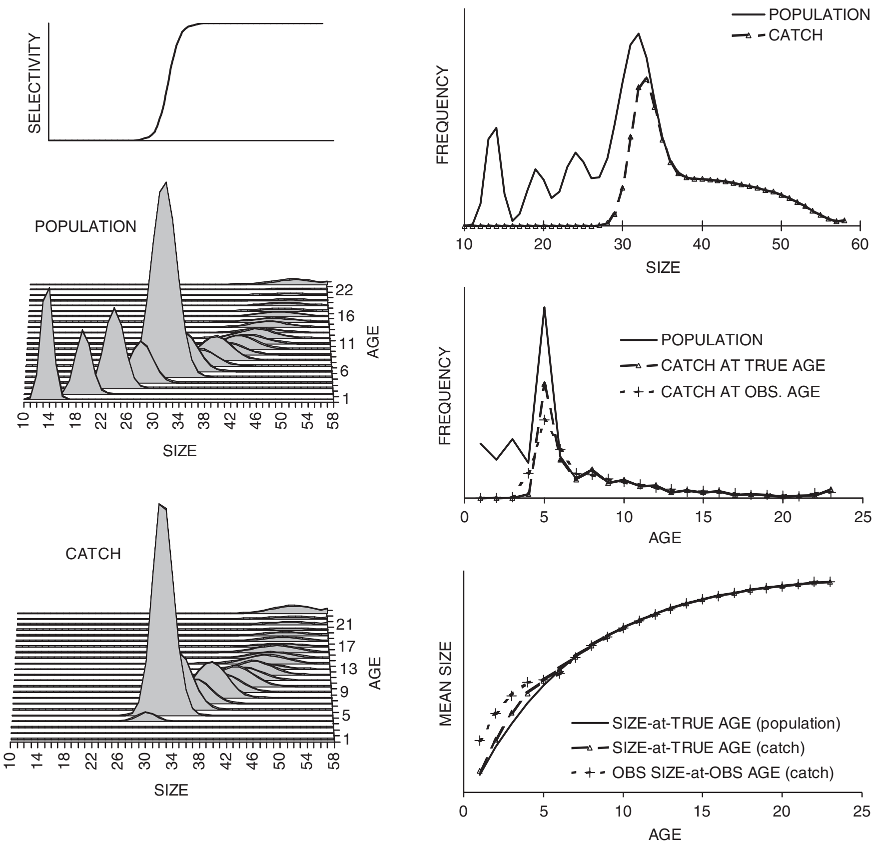

Fig. 9.5 Size-selectivity acts upon the population's total size composition to create the size composition of catch. It also acts upon the normal distribution of size-at-age to create a unique observed size-at-age for that fishery that is used to calculate that fishery's body weight at age. In the observation sub-model, aging error will blur the occurrence of a strong year class into adjacent ages, thus affecting the observed size-at-age for those weaker year classes

investigation of geographic clines in growth. Finally, each gender, birth-season, growth pattern platoon can be further subdivided into up to five sub-platoons to better track the consequences of size-selective mortality. We cannot distinguish such sub-platoons in the data, but we can assert that an underlying growth process that is built up from multiple platoons is a more accurate representation of the natural range of individual growth trajectories (Kristensen et al. 2005) than what can be provided by a single, homogeneous platoon (Fig. 9.6). Fast-growing sub-platoons will enter a size-selective fishery at a younger age, and thus experience greater cumulative fishing mortality and resultant reduced survival to older ages.

Each fishery or survey included in a SS2 configuration has a pattern of selectivity that can be in terms of age, size, or both and to include gender differences. Selectivity defines the fraction of a particular size (or age) that is captured relative

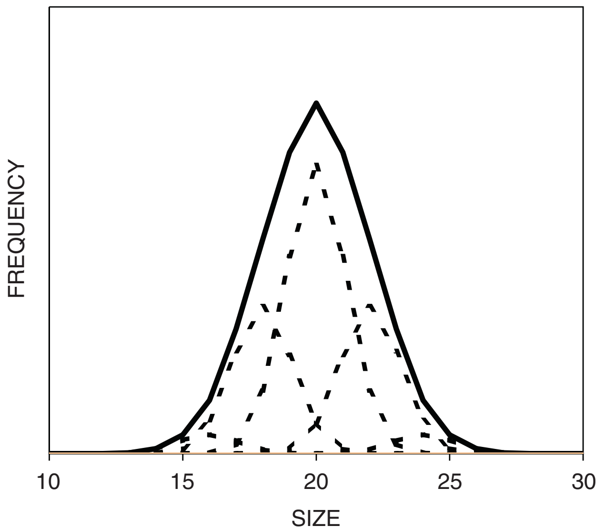

Fig. 9.6 Each platoon of fish in the model can be divided into 1, 3, or 5 independent sub-platoons with a specified fraction of the total variability in size-at-age between sub-platoons versus within sub-platoon. Here five sub-platoons have the amount of variability within sub-platoon equal to 70% of the variability between sub-platoons

to capture rate for the size (or age) that has a selectivity of 1.0. By providing both age and size options, the model can, for example, be configured to estimate small fish selectivity as a function of size if it is believed that it is mostly a function of gear technical characteristics (e.g., mesh size), while also estimating older fish selectivity as a function of age if it is believed that it is mostly due to an age-based diffusion into microhabitats that are relatively inaccessible to the fishery or survey sampling gear. Various parameterizations of selectivity are available. These can be as simple as specification of knife-edge selectivity occurring at a particular age (no estimated parameters). Functional forms include a two-parameter logistic function, a six-parameter double normal (Fig. 9.7), nonparametric forms with a separate parameter for each age, and others.

SS2 incorporates two options for modeling the fishery catch. The first option employs Pope's (1972) approximation to calculation of fishery mortality. Here, the population's numbers-at-age are decayed to the middle of the season according to natural mortality alone. Then the catch-at-age for each fishery is calculated as a harvest rate times the selectivity at age. Each fisheries harvest rate is calculated such that the total catch (either in numbers or biomass) matches the observed catch. The survivors after all fishery removals are then decayed to the end of season. In this first approach, the harvest rates are simply an array of values to match the observed catch and do not enter the model as explicit parameters. The second option treats fishing mortality as a continuous process simultaneous with natural mortality. Here the fishing mortality rates are estimated as model parameters. The first option is faster, especially in models with long time series and large number of fisheries, and the second option performs more robustly when fishing mortalities are high. When fishing mortalities are low or multiple seasons are used to reduce the cumulative mortality within any season, the two approaches produce equivalent results. Fishery catch can be in terms of numbers or biomass and different fisheries can be in different units.

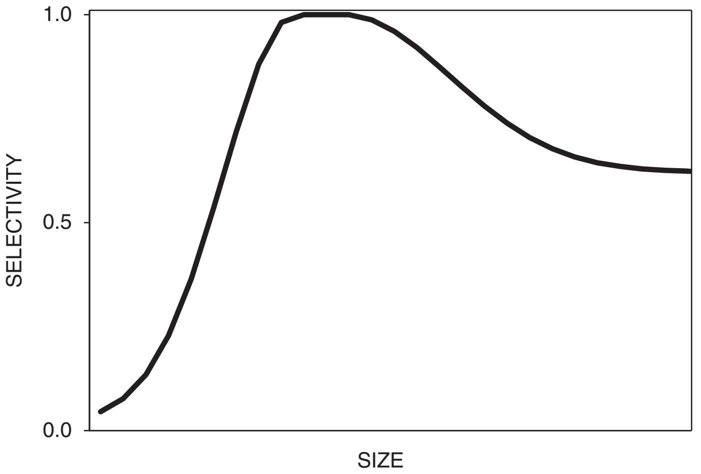

Fig. 9.7 The double normal selectivity function is commonly employed in SS2 when a dome-shaped selectivity pattern is needed. Six parameters control the function's shape

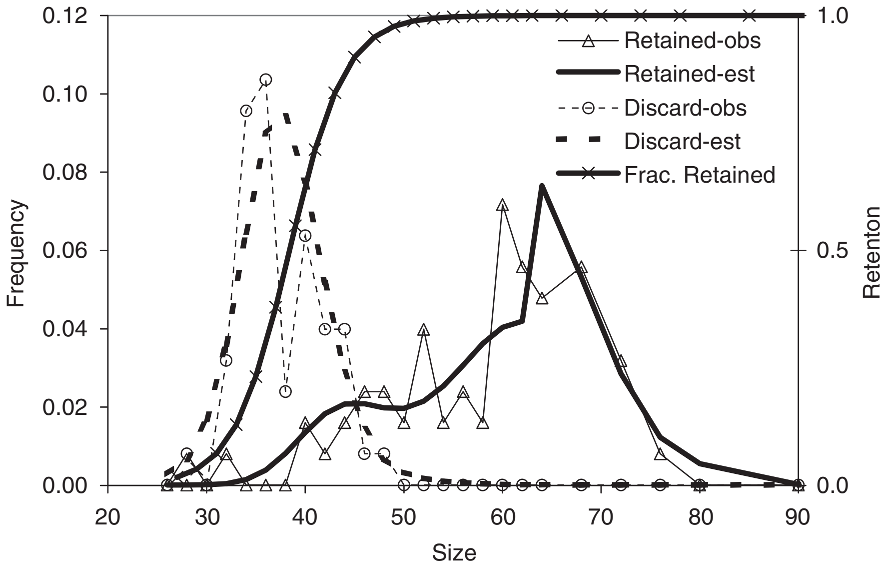

Fig. 9.8 SS2 allows for partitioning of the total catch into discarded and retained components and the calculation of expected values (est) for data from each component. Inclusion of discard and retained size composition data (obs) in this example allowed estimation of the retention function within SS2

Fishery catch can be partitioned into discarded and retained components (Punt et al. 2006) (Fig. 9.8). Further, the discarded partition can have a length-specific survival function defined. Thus the total fishing mortality for a given fleet is the retained catch plus the mortal fraction of the discarded fish. By partitioning the total catch into discard and retained components, SS2 is then able to develop expected values for discard only, retained only, or total catch samples of size and age composition.

With the above description of discarding, we have partly transitioned from the population sub-model to the observation sub-model. SS2 can produce expected values for several kinds of common fishery and survey data. All of these expected values start from a size/age/gender array of selected fish calculated by applying the size/age/gender selectivity for each gear to the size/age/gender population array at that point in time (Fig. 9.5). When summed across ages, the result is the expected size composition, which can then be compared to an observation of the size composition for that fishery or survey. When summed across sizes, the result is the expected value of the sampled age composition. But stopping at this stage would omit the effect of aging imprecision. The process of determining age from otoliths or scales or other structures involves some uncertainty that is expected to blur the observed age composition (Tyler et al. 1989; Kimura and Lyons 1991). Rather than try to remove this blurring effect from the data before providing the data to the model, integrated analysis models like SS2 build the blurring process into the model so that the expected values are blurred to the same degree that the data are blurred. While it is feasible to provide the model with both size and age composition data, this will tend to double weight the information from some fish. An alternative approach available in SS2 is to examine the age composition data conditional on being within a subset of the size/gender range (Methot 2000; Punt et al. 2000). Thus, the model considers the additional information provided by age data over and above the information already provided by the fish size.

The size and age composition expected values can be compared to data for each gender, summed across genders, or treating the size/gender information as a joint distribution that preserves information on sex ratio. When multiple platoons and sub-platoons are used, the selectivity and resultant mortality is applied to each and the results are summed across platoons because the difference between platoons is invisible to the observation process. This is analogous to creating a two-gender model to deal with a known difference in growth between males and females, then summing the expected values across the two genders to provide a combined gender expected value where data have not been partitioned into males in females.

SS2 can also include information from surveys of stock abundance. The sum of the age/size/gender "selected" fish from a particular fishery or survey is the estimated total abundance of this selected component of the population. The sum can be in terms of numbers of fish, or in terms of biomass by incorporating the model's estimate of body weight at age. This sum times a scaling factor is the expected value for the survey observations. This scaling factor is also termed a catchability coefficient, q. Ideally, q would be independently measured and calibrated, as is possible for some acoustic and visual surveys. Some bottom trawl surveys where the area-swept, herding, and escapement has been measured may also be analyzed as a fully calibrated survey (Somerton et al. 2007).

Most surveys, however, have intangible factors in the catchability that have defied direct calibration to date. Where standardization of methods has allowed assertion that q has remained constant at some unknown value, then it is feasible to treat the survey as a measure of the stock's relative trend and to allow the assessment model to estimate q internally as a scaling factor between the units of the survey measurement and the units of estimated, selected population abundance. In this case, it takes accumulation of a time series of observations before any meaningful assessment information can be obtained from the survey. In the worst case, the intangible components of q cannot be comfortably asserted to be constant over time. This is most common when fishery catch per unit of effort is used as an index of population abundance. However, even in this situation it is not uncommon for the assessment configuration to maintain the assumption of constant q or a prespecified drift in q over time.

As an example of the kind of analysis that can be conducted with SS2, consider three survey scenarios. In the first scenario, a single research vessel conducts an annual relative index survey for 20 years and is then replaced by a similar vessel that is calibrated to the first vessel so that a continuous time series of survey observations can be analyzed with the assessment model under a constant q assumption. In the second scenario, several (2–4) chartered vessels conduct replicate surveys each year and the combined results of these surveys are analyzed as a relative index under the constant q assumption. Differences in q between vessels is not directly calibrated and becomes part of the system noise. In the third scenario, thousands of logbook observations from hundreds of vessels are processed using a statistical model to develop an annual CPUE index that is analyzed by the assessment model under the constant q assumption. All three scenarios have assumed a constant q in the assessment analysis and the fishery CPUE scenario may produce a more precise model result because of the large number of observations. What's wrong with this picture and why is there value in single-vessel standardized surveys over statistically standardized fishery data?

A more holistic approach would acknowledge that there is always some fluctuation in q (Millar and Methot 2002; Francis et al. 2003; Bence and Wilberg 2006) and some survey methods are better than others at keeping these fluctuations small. We expect a more constant q from scenario 1 in the preceding paragraph than from scenario 3, so the model should be configured to use this knowledge. In SS2, the q parameter can be specified as having an annual random deviation or an annual random walk, each with prior value with variance. Thus it is feasible to directly incorporate information on the degree of confidence in the constancy of q. In the first scenario above, the annual q parameter could be specified to have only a small random walk from the previous year's q value except in the year of vessel transition in which case the value of the vessel inter-calibration and its variance would provide information on the size of a larger change in q that year, and this degree of change would be updated each subsequent year as more data are collected using the new vessel. In the second scenario, the degree of variability in estimated vessel effect among the chartered survey vessels could be used to constrain the degree of random drift in q for their combined survey result. In the third scenario, the lack of direct standardization of the fishery CPUE data could be used to assign a larger variability to the possible random walk in catchability (Bence and Wilberg 2006).

Of course, allowing too much variability in q means that the population signal possibly found in the survey trend will be lost to the estimation of time-varying q. The result of this more holistic treatment of variability in q would be demonstration of the improved overall precision in assessment model results that can be attained when good standardization and vessel inter-calibration has occurred.

The statistical sub-model of SS2 is generally as described above for integrated analysis models. All data going into the model have an associated level of variance that determines the scaling of the deviations between the data and the model's estimates of expected values. Because these estimates of data variance may themselves be inaccurate and because the structure of the model may not be flexible enough to adequately represent all processes that created the actual data, the SS2 approach provides opportunity for adjustment of the data variance. Model outputs include statistics that compare the average goodness of fit to the level of input data variance. This can then guide changes in the level of model flexibility (more or less parameters) and adjustment of the input variance levels to better represent the subsequent model capability to match these data. With the input and output variances so tuned, the final estimates of variance in model outputs better represent the relative contribution of all sources of data. As the model evolves towards more use of random effects for factors such as annual survey catchability, it will be necessary to develop better protocols for balancing the magnitude of these random effects versus adjustment of the variance terms.

Although model fitting is in terms of the total NLL, it is prudent and necessary to examine the model's fit to each data component individually and graphically (Richards et al. 1997). The best-fitting set of parameters will be a compromise. They will not provide the best possible fit to any one component, nor should they produce an unreasonably poor fit to any influential data. Visualizing and quantifying the residuals in the fit to each type of included data helps the modeler identify places where adjustments need to be made in the model structure. Too tight a fit means that some aspect of the model is too flexible and has too many free parameters. Unreasonable patterns in residuals usually mean that some process affecting the real data has not yet been included in the model. The art of model building is largely about the selection of the best degree of model complexity and best balance of fit among the various data components.

Management quantities and forecasts are an important feature of SS2. Following estimation of the model's parameters, the values of various management quantities are calculated and a forecast is conducted using a selected management policy or specified catch level. By integrating this management layer into the overall model, the variance of the estimated parameters can be propagated to the management quantities and forecast, thus facilitating a description of the risk of various possible management scenarios. Because the entire model works as a simulation, it is feasible for a single model configuration to start in a pre-data, lightly fished era; extend through a data-rich era to the present day; and continue into a forecast era (Fig. 9.4). The transition from the estimation era to the forecast era is transparent and denoted only by the phasing out of the data availability, but even that transition is blurred as very recent environmental data are included in today's models.

The management quantities include the fishing intensity level that would produce a specified level of spawning biomass per recruit, a specified spawning biomass level as a fraction of unfished biomass (taking into account the spawner–recruitment relationship), and the fishing intensity level that would produce maximum sustainable yield. The search for these quantities is across a range of fishing intensity levels conditioned upon a particular allocation of the intensity among fishing fleets and the size/age pattern of selectivity for each fleet. This means that the level of fishing mortality varies between ages in a complex way and a single value representation of fishing mortality is ambiguous. Because of this ambiguity, SS2 reports the level of fishing intensity in terms of spawning potential ratio (Prager et al. 1987). The forecast can use the current fishing intensity or any one of the calculated management quantities. The forecast incorporates biomass-based adjustments to future fishing intensity levels according to the harvest policies in the west coast groundfish Fishery Management Plan (Ralston 2002).

Environmental data can be incorporated into SS2 is two ways. First, any SS2 parameter can be defined to be a function of an input environmental data time series. For example, this could be used to set the annual expected recruitment deviation to be a function of sea surface temperature (Schirripa and Colbert 2006) according to the method described in Maunders and Watters (2003). It could also be used to set a fishery selectivity parameter to be a function of an environmental variable such as wind speed or any other predictor variable such as mean depth of fishing. Survey catchability could be similarly linked to an input variable. Growth could be linked to environmental variables such as temperature or ecosystem productivity. Natural mortality could be linked to predator abundance (Methot 1989; Livingston and Methot 1998).

The first approach to environmental linkage described above is based upon the intuitive concept that the environmental factor has caused changes in the population process being modeled. But this intuition has a degree of naivety. The environmental variables we measure are, hopefully, a good indicator of the myriad and complex factors that actually cause the changes in population processes. But they are only indicators; for example, they are data that may be informative about the process. The second method of including environmental data in SS2 exploits this alternative logic. Currently, SS2 only implements this alternative for recruitment by having the environmental data enter the model as if they are a survey of the age zero recruitment deviations. These data then compete and/or reinforce other data in the model to produce the best-fitting estimates of recruitment. From a statistical perspective, the model does not care if the direct recruitment data come from a fishery-independent survey of 5-month-old juveniles that indexes the numbers of recruits, or from an environmental measurement that indexes the deviation of recruitment relative to the level predicted from the spawner–recruitment relationship. Thus, method two uses environmental data as a correlate to help explain recruitment, whereas method one uses environmental data to cause the recruitment deviations. In the future, SS2 and other integrated analysis models are expected to evolve to more use of such random effects procedures. In this approach, any parameter could be defined as having an annual change, the random effect, and the magnitude of these changes can be estimated through inclusion of conventional as well as environmental data.

9.6 Getting to Ecosystem

The NMFS Stock Assessment Improvement Plan described Tier II as elevating stock assessments to new national standards of excellence and Tier III assessments as reaching to an ecosystem level. What does it mean to achieve an ecosystem level for fishery stock assessments? Characterizing Tier II assessments helps provide a foundation for this discussion. Tier II assessments are empirical descriptions of a stock's status. They measure stock abundance and fishing mortality, compare these to reference levels, calculate fishing mortality levels that are sustainable given current conditions, and translate abundance and fishing mortality targets into a forecast of short-term catch levels. In doing so, the assessment model defines the system as containing the stock of fish and its fishery. Outside influences are recognized to cause random perturbations to the system, but these perturbations are considered measurable within the system and not to require understanding about how the outside influences cause the perturbations. Tier II assessments treat the fishery reference levels as entirely derivable from information inside the "system" and treat the outside influences as providing only random noise without directional trends. Such Tier II stock assessments do have a one-directional link to ecosystem analyses because the time series output of Tier II assessments is valuable input and validation to holistic ecosystem food web models.

Getting to Tier III means expanding the scope of the assessment system so that more of the outside influences become part of the analyzed system. As a first step, fishery assessment models can include more environmental and ecosystem inputs so that factors in the assessment system are explicitly linked to these inputs. Integrated analysis models have already begun evolving in this direction as described in this paper, and some examples of linked multispecies models have appeared. The next step will essentially be a merger of the expanding scope of these integrated analysis models and the increasing detail and data assimilation capabilities of holistic ecosystem food web models. Before reaching that stage, and perhaps even at today's stage of model evolution, it seems relevant to ask whether it would be beneficial to explicitly develop two linked scales of model complexity (Hilborn 2003). The more complex, ecosystem-linked model would be the strategic model used to determine target harvest rates that achieve optimum yield from fisheries while explicitly accounting for ecosystem effects. The less complex model would be the tactical model that uses simple data inputs to guide tweaking the fishery up and down to implement the policies determined from the complex model.

The following section identifies some ways in which Tier II assessments can evolve towards Tier III. In general these fall into two categories: (1) allowing change in factors that now must be assumed to be constant, and (2) improving predictions for factors that are currently allowed to fluctuate in a random manner.

The first category includes natural mortality and the shape and scale of the spawner–recruitment relationship. The second category principally includes annual recruitment deviations and body growth. Environmentally caused fluctuations in survey catchability could be included here also.

Natural mortality (M) is the 900 lb gorilla of stock assessment parameters. It is arguably the parameter that is least estimable from conventional assessment data and the parameter that is most dependent on shifts in the ecosystem predator–prey relationships. Where fishing mortality (F) is much greater than natural mortality, then the exact value of M has little effect on estimates of the trend in abundance of the stock, but well-controlled fisheries are not likely to have F much, if at all, greater than M. While small error in M is unlikely to cause an assessment to misestimate the trend in stock abundance, the absolute level of stock abundance is directly related to the level of M. Natural mortality can be an estimable parameter in an assessment model, but robust performance of such a model generally requires informative and precise survey and age composition data with verifiable selectivity and catchability characteristics. Without data that is truly and unambiguously informative about M, the model will adjust M to attempt to explain other patterns in the data. Consequently, M is usually held fixed in assessment models at a level estimated from the age composition from pre-fishery periods, or lightly fished components of the stock, or from empirical relationships between the few direct estimates of M and more easily obtained life history parameters. Validating the accuracy of these M estimates for today's fully exploited ecosystem and obtaining time and age-varying values is an extreme challenge. Getting contemporary information on M is one of the greatest contributions that ecosystem studies could give to stock assessment.

The estimated spawner–recruitment relationship (S–R) represents the average level of recruitment expected from a specified level of reproductive output. Walters and Martell (2004) contains a broad examination of the various factors that go into estimation and interpretation of this relationship. They note that R is not produced by a single S–R relationship; rather it is the result of a myriad of sequential life history stages, each with various potentials for density-dependent factors. So, under what conditions can a simple two-parameter S–R relationship adequately describe the historical pattern in the data, serve as the basis for estimation of the long-term productive capacity of the stock, and forecast the expected level of future recruitment?

First is the measurement of spawning biomass. It is not uncommon for fish stocks to exhibit changes in age-specific maturity over time and it is possible that the shift to earlier maturation is part of the stock's compensatory response to the additional adult mortality imposed by the fishery. Yet such changes in maturity are usually not measured as a time series and it is more common for contemporary, "better" estimates of maturity and fecundity to be used to calculate the spawning output throughout the time series. Thus, the estimated curvature in the S–R can be confounded with the degree to which maturation has shifted and whether these temporal shifts have been taken into account in the calculation of spawning biomass. Investigation of the degree to which density-dependent shifts in maturation occur in harvested fish stocks could lead to improved standard practices for dealing with such shifts in the face of inadequate information.

Second is the assertion that the S-R is constant in the face of ecosystem shifts. The S–R basically represents the cumulative mortality through the egg–larval–juvenile stages and how this mortality changes with stock abundance. However, juvenile fish can be prey to many species of fish and many of these species may have exhibited changes in abundance over the same decades that are being analyzed for the S–R of the subject species. In many systems, the S–R data have been collected over a period of time in which the abundance of many species has been reduced due to fishing. Frankly, the coastal ecosystems are coming into a new state with human fishers as an introduced predator. Fishing mortality's primary effect is in reducing the abundance of older fish, which have the greatest tendency to be piscivorous. So it is possible that while fishing has reduced the abundance of spawners that produce juveniles of species A, fishing has also changed the abundance of species B and C that are predators on the juveniles of species A. A major contribution of ecosystem food web studies could be identification of the stocks that are most in need of including ecosystem interactions in their S–R relationships.

A third issue is the sequencing of environmentally caused and density-dependent mortality. The S–R relationship is routinely written such that environmentally caused variability occurs after the action of density-dependent survival. The ecological arguments that would support such a relationship make sense for a species like salmon. For example, when salmon spawners are super-abundant they may spawn in marginally suitable reaches of streams that do not support high egg survival. Subsequent to this density-dependent stage is the estuarine and early ocean period in which it is recognized that highly variable environmental conditions will cause variation in survival of juveniles. However, what makes sense for the predominant life history of marine fish? Isn't it generally accepted that high variability in survival occurs in the early larval stage and that density-dependence is most likely during the subsequent juvenile stage as they settle into limited habitats or otherwise behaviorally interact? If so, shouldn't the S–R relationship be formulated such that most variability occurs before density-dependence acts, in which case the density-dependence should dampen the environmentally caused variability in larval survival? Further consideration of this alternative S–R formulation could perhaps make more sense of the uninformative scatter found in many sets of spawner–recruitment data.

Fourth is the effect of environmental factors, which has both long-term and short-term consequences. Long-term environmental patterns can bias estimates of S–R parameters. As fishing has moved average spawning biomass from a moderately high level to a moderately low level over a period of decades, the observed change in average recruitment is the basis for the estimated S–R relationship. However, if the decadal time scale of some environmental shifts also affects recruitment, then the S–R estimate is confounded with the environmental changes. Unambiguously disentangling the spawner effect from the long-term environmental effect seems nearly impossible until many decades of monitoring are available or until process studies are able to estimate either the spawner effect or the environmental effect without resorting to simple correlation studies.

The short-term effect of the environment on annual recruitment is more a matter of improved forecasting rather than disentangling historical relationships. Conventionally available fishery and survey data are sufficient to estimate annual fluctuations in recruitment. However, these data cannot provide direct information on recruitment until the fish are old enough to appear in these sources of data. They provide precise, accurate estimates but they are fundamentally retrospective estimates. This may be sufficient for long-lived species in which the young recruiting year classes are only a small fraction of the population and fishery, but more timely recruitment estimates are needed for short lived species, species with extremely high recruitment fluctuations, or species with management plans that seek to closely track maximum potential yield. More timely estimates of recent recruitment fluctuations can be made by conducting a survey that measures abundance of pre-recruits at a young age, and/or by determining and measuring environmental covariates that provide good predictions of recruitment. Survey cost, timeliness of estimates, and precision are factors that influence the relative merits of initiating a pre-recruit survey versus initiating a research program to determine a relevant environmental predictor. In practice, a program desiring a better recruitment predictor will probably need to do both in order to provide necessary cross-validation.

Growth and reproduction, like recruitment, are empirically measurable from available data, thus can be allowed to change over time in assessment models. Tier III models will include mechanisms that link growth and reproduction to population abundance and ecosystem/environmental factors.

A final aspect of more realistic Tier III assessment models is spatial structure. The need for spatially explicit models is growing as we consider the dynamics of populations that have a significant fraction of their abundance within marine reserves (Holland 2002; Punt and Methot 2004; Field et al. 2006). Conventional models that treat the stock as if diffusion was infinitely high can produce biased results when applied to populations that have low rates of mixing between areas. Spatial population models may be needed to combine information from multimodal survey programs in which direct observation methods measure the absolute fish abundance on rocky habitats while conventional trawl surveys measure relative trends on adjacent smoother habitats. In principle, it is straightforward to code the model to include multiple geographic zones and to allow fish movement between zones. The data requirements for such a model are feasible in some of our highly monitored fisheries today. However, spatially disaggregating historical data and obtaining information on rates of fish movement are daunting steps.

Some of the above suggestions have substantial new data requirements. In particular, tagging studies to obtain movement rates and predation monitoring programs to measure natural mortality rates are expensive field programs, the value of which should be evaluated against the potential gains in assessment accuracy and precision. The suggestions regarding temporal shifts in maturation, the form of the spawner–recruitment relationship, and long-term climate effects on the spawner–recruitment relationship could, in some instances, be addressed through additional investigation using currently available data. Finally, suggestions for more flexible model capabilities can be implemented as next generation assessment models are developed. In the short term, expanded model capabilities to admit time-varying factors may demonstrate decreased precision in model results, but this will establish a framework for better identification of the data needed to truly improve the precision of model results.

9.7 Operational Model

Stock assessments provide operational support for quantitative management of fisheries. The integrated analysis models described in this paper have the capability to link the estimation of the population's historical abundance, to the inference of biological reference points and forecasts of future population trends and potential catch. In order to provide more timely updates for more stocks, efficiencies are needed at each step in the sequence from collection of data through delivery of results.

The first obvious step is the need for timely access to accurate, precise, and comprehensive fishery and survey data. A great fraction of the total assessment time and energy goes into discovering, quality-checking, and calibrating historical data that don't quite meet the standards of today's data collection systems. Greater efficiency in this process can be obtained by taking a horizontally integrated approach rather than a vertically integrated approach. Most data collection systems collect data on multiple species, so analysis and review of these data is best accomplished at the same time across all relevant species, rather than species by species as they are assessed. Likewise, life history analysis methods are likely to be relevant for many related species, so also can be clustered into a methods-oriented review rather than fully opening the topic for each species as its assessment is reviewed. Timely availability of contemporary data can be improved through better data systems: more electronic recording of data in the field and more sophistication in the databases to quickly accomplish quality checks and delivery of data to end-users.

The second step is the set of models. These must be at the right level of complexity: simple enough to be rapidly updated without extensive diagnostics, and complex enough to adjust for confounding effects of non-fishery factors. Because of the great diversity of data situations and fish life history patterns, standardized models must have a flexible structure that is scalable in complexity to the particular situation being analyzed. Quick tactical models may need to be paired with more complex strategic models to achieve the right mix of overall capabilities. Once developed, such standardized models allow less experienced users to fully participate in assessment modeling, they facilitate improved communication as results are being reviewed, and they enable development of a more comprehensive suite of tools to disseminate model results to a wide range of clients and the interested public. A downside of increased standardization among the models used for management purposes is a stifling of research and creativity. Recognizing this as a possibility can perhaps be turned into greater emphasis on explicit research on model development. We need to decide what technical and ecosystem processes can be included within the modeled system of the operational models, and which must be fire-walled off into exploratory analyses designed to improve the next generation of operational models.

Third is the process in which the model is used to develop technical advice for the regulation of fisheries. This regulatory link creates a high level of controversy for all aspects of the stock assessment enterprise. Fishery data and assessment models receive an extraordinary degree of public scrutiny and formal review. Increased throughput of assessment updates will require streamlining of the extensive review process now routinely required before stock assessment results can serve as the scientific basis for fishery management. Emphasizing review of broadly applicable assessment data and methods, rather than each final result, is a logical step in this streamlining, while maintaining public trust in the final results.

References

Bence, James R. and M. J. Wilberg. 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. Canadian Journal of Fisheries and Aquatic Sciences 63: 2275–2285.

Bobko, S. J. and S. A. Berkeley. 2004. Maturity, ovarian cycle, fecundity, and age-specific parturition of black rockfish (*Sebastes melanops*). Fishery Bulletin 102: 418–429.

Brodziak, J., P. Rago, and R. Conser. 1998. A general approach for making short-term stochastic projections from an age-structured fisheries assessment model. In F. Funk, T. Quinn II, J. Heifetz, J. Ianelli, J. Powers, J. Schweigert, P. Sullivan, and C.-I. Zhang (Eds.), Proceedings of the International Symposium on Fishery Stock Assessment Models for the 21st century. Alaska Sea Grant College Program, University of Alaska, Fairbanks, AK.

Brown, S. K., M. Shivlani, D. Die, D. B. Sampson, and T. A. Ting. 2006. The Center for Independent Experts: The National External Peer Review Program of NOAA's National Marine Fisheries Service. Fisheries 31 (12): 590–600.

Butterworth, D. S. and M. O. Bergh. 1993. The development of a management procedure for the South African anchovy resource, pp. 83–99. In S. J. Smith, J. J. Hunt, and D. Rivard (Eds.), Risk Evaluation and Biological Reference Points for Fisheries Management. Canadian Special Publications in Fisheries and Aquatic Science 120.

Chen, Y., M. Kanaiwa, and C. Wilson. 2005. Developing a Bayesian stock assessment framework for the American lobster fishery in the Gulf of Maine. New Zealand Journal of Freshwater and Marine Sciences 39: 645–660.

Field, J. C., A. E. Punt, R. D. Methot, and C. J. Thomson. 2006. Does MPA mean 'Major Problem for Assessments'? Considering the consequences of place-based management systems. Fish and Fisheries 7: 284–302.

Fournier, D. A. and C. P. Archibald. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences 39: 1195–1207.

Francis, R. I. C. C., R. J. Hurst, and J. A. Renwick. 2003. Quantifying annual variation in catchability for commercial and research fishing. Fishery Bulletin 101: 293–304.

Goodyear, C. P. 1984. Analysis of potential yield per recruit for striped bass produced in Chesapeake Bay. North American Journal of Fisheries Management 4: 488–496.

Helser, T., A. E. Punt and R. D. Methot. 2004. A generalized linear mixed model analysis of a multi-vessel fishery resource survey. Fisheries Research 70: 251–264.

Hilborn, R. 2003. The state of the art in stock assessment: where we are and where we are going. Scientia Marina 67 (Suppl 1): 15–20.

Hjort, J. 1914. Fluctuations in the great fisheries of northern Europe viewed in the light of biological research. Rapports et Proces-Verbaux des Réunions du Conseil Permanent International pour l'Exploration de la Mer 20: 1–228.

Holland, D. S. 2002. Integrating marine protected areas into models for fishery assessment and management. Natural Resource Modeling 15: 369–386.

Kimura, D. K. and J. J. Lyons. 1991. Between-reader bias and variability in the age-determination process. Fishery Bulletin 89: 3–60.

Kristensen, K., P. Lewy, and J. E. Beyer. 2005. Maximum likelihood estimation in a size-spectra model. Canadian Journal of Fisheries and Aquatic Sciences 99: 1–10.

Livingston, P. A. and R. D. Methot. 1998. Incorporation of predation into a population assessment model of eastern Bering Sea walleye pollock. In F. Funk, T. J. Quinn II, J. Heifetz, J. N. Ianelli, J. E. Powers, J. F. Schweigert, P. J. Sullivan, and C.-I. Zhang (Eds.), Fishery Stock Assessment Models.. Alaska Sea Grant AK-SG-98-01. University of Alaska Fairbanks 1037 p.

Mace, P. M., N. W. Bartoo, A. B. Hollowed, P. Kleiber, R. D. Methot, S. A. Murawski, J. E. Powers, and G. P. Scott. 2001. National Marine Fisheries Service Stock Assessment Improvement Plan. Report of the NMFS National Task Force for Improving Fish Stock Assessments. NOAA Technical Memorandum NMFS-F/SPO-56. 76 pp.

Maunder, M. N. and A. E. Punt. 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research 70: 141–159.

Maunder, M. N. and G. M. Watters. 2003. A general framework for integrating environmental time series into stock assessment models: model description, simulation testing, and example. Fishery Bulletin 101: 89–99.

Maunder, M. N., S. J. Harley, and J. Hampton. 2006. Including parameter uncertainty in forward projections of computationally intensive statistical population dynamic models. ICES Journal of Marine Science 63: 969–979.

Maunder M. N., J. T. Schnute, and J. Ianelli. 2009. Computers in Fisheries Population Dynamics. In B. Megrey and E. Moksness (eds.), Computers in Fisheries Research. Kluwer Academic Publishers. p 337–372. ISBN: 978-1-4020-8635-9.

McAllister, M. K., E. K. Pikitch, A. E. Punt, and R. Hilborn. 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. Canadian Journal of Fisheries and Aquatic Sciences 51: 2673–2687.

Methot, R. D. 1989. Synthetic estimates of historical abundance and mortality for northern anchovy. American Fisheries Society Symposium 6: 66–82.

Methot, R. D. 2000. Technical Description of the Stock Synthesis Assessment Program. National Marine Fisheries Service, Seattle, WA. NOAA Tech Memo. NMFS-NWFSC-43: 46 p.

Millar, R. B. and R. D. Methot. 2002. Age-structured meta-analysis of the U.S. West Coast rockfish (Scorpaenidae) populations and hierarchical modeling of trawl survey catchabilities. Canadian Journal of Fisheries and Aquatic Sciences 59: 383–392.

National Research Council (NRC). 1998. Improving Fish Stock Assessments. National Academy Press, Washington, DC. 177 pp.

NOAA. 1996. Magnuson-Stevens Fishery Conservation and Management Act. NOAA. Technical Memorandum NMFS-F/SPO-23. Silver Spring, MD.

Otter Research Ltd. 2005. An introduction to AD Model Builder Version 7.1.1 for use in nonlinear modeling and statistics. Sidney, BC, Canada. (http://admb-project.org/)

Patterson, K., R. Cook, C. Darby, S. Gavaris, L. Kell, P. Lewy, B. Mesnil, A. E. Punt, V. Restrepo, D. W. Skagen, and G. Stefansson. 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish and Fisheries 2: 125–157.

Pope, J. 1972. An investigation of the accuracy of Virtual Population Analysis using cohort analysis. ICNAF Research Bulletin 9: 65–74. Also available in D. H. Cushing (Ed.). 1983. Key Papers on Fish Populations. IRL Press, Oxford. 405 p.

Prager, M. H., J. F. O'Brien, and S. B. Saila. 1987. Using lifetime fecundity to compare management strategies: a case history for striped bass. North American Journal of Fisheries Management 7: 403–409.

Prager, M. H., C. E. Porch, K. W. Shertzer, and J. F. Caddy. 2003. Targets and limits for management of fisheries: a simple probability-based approach. North American Journal of Fisheries Management 23: 349–361.

Punt, A. E. and R. D. Methot. 2004. Effects of marine protected areas on the assessment of marine fisheries. American Fisheries Society Symposium 42: 133–154.

Punt, A. E., F. Pribac, T. I. Walker, B. L. Taylor, and J. D. Prince. 2000. Stock assessment of school shark, *Galeorhinus galeus*, based on a spatially explicit population dynamics model. Marine and Freshwater Research 51: 205–220.

Punt, A. E., D. C. Smith, G. N. Tuck, and R. D. Methot. 2006. Including discard data in fisheries stock assessments: Two case studies from south-eastern Australia. Fisheries Research 79: 239–250.

Quinn, T. J. I. 2003. Ruminations on the development and future of population dynamics models in fisheries. Natural Resource Modeling 16: 341–392.

Richards, L. J., J. T. Schnute, and N. Olsen. 1997. Visualizing catch-age analysis: a case study. Canadian Journal of Fisheries and Aquatic Sciences 54: 1646–1658.

Ralston, S. 2002. West coast groundfish harvest policy. North American Journal of Fisheries Management 22: 249–250.

Restrepo, V. R., G. G. Thompson, P. M. Mace, W. L. Gabriel, L. L. Low, A. D. MacCall, R. D. Methot, J. E. Powers, B. L. Taylor, P. R. Wade, and J. F. Witzig. 1998. Technical Guidance on the Use of Precautionary Approaches to Implementing National Standard 1 of the Magnuson-Stevens Fishery Conservation and Management Act. NOAA Technical Memorandum. NMFS–F/SPO–31.

Schirripa, M. J. and J. J. Colbert. 2006. Interannual changes in sablefish (*Anoplopoma fimbria*) recruitment in relation to oceanographic conditions within the California Current System. Fisheries Oceanography 15 (1): 25–36.

Schnute, J. T., M. N. Maunder, and J. N. Ianelli. 2007. Designing tools to evaluate fishery management strategies: can the scientific community deliver? ICES Journal of Marine Science 64: 1077–1084.

Sitar, S. P., J. R. Bence, J. E. Johnson, M. P. Ebener, and W. W. Taylor. 1999. Lake Trout mortality and abundance in Southern Lake Huron. North American Journal of Fisheries Management 19: 881–900.

Smith, A. D. M., K. J. Sainsbury, and R. A. Stevens. 1999. Implementing effective fisheries management systems–management strategy evaluation and the Australian partnership approach. ICES Journal of Marine Science 56: 967–979.

Somerton, D. A., P. T. Munro, and K. L. Weinberg. 2007. Whole-gear efficiency of a benthic survey trawl for flatfish. Fishery Bulletin 105: 278–291.

Tyler, A. V., R. J. Beamish, and G. A. McFarlane. 1989. Implications of age determination errors to yield estimates. 27–35. In R. J. Beamish and G. A. McFarlane (Eds.), Effects of Ocean Variability on Recruitment and an Evaluation of Parameters Used in Stock Assessment Models. Canadian Special Publications in Fisheries and Aquatic Science 108: 27–35.

Walters, C. J. and S. J. D. Martell. 2004. Fisheries Ecology and Management. Princeton University Press, Princeton, NJ. 399 p.