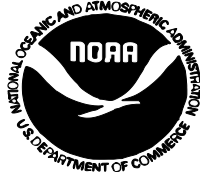


NOAA Technical Memorandum NMFS-NWFSC-43



Technical Description of the Stock Synthesis Assessment Program

August 2000

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service

NOAA Technical Memorandum NMFS Series

The Northwest Fisheries Science Center of the National Marine Fisheries Service, NOAA, uses the NOAA Technical Memorandum NMFS series to issue informal scientific and technical publications when complete formal review and editorial processing are not appropriate or feasible due to time constraints. Documents published in this series may be referenced in the scientific and technical literature.

The NMFS-NWFSC Technical Memorandum series of the Northwest Fisheries Science Center continues the NMFS-F/NWC series established in 1970 by the Northwest & Alaska Fisheries Science Center, which has since been split into the Northwest Fisheries Science Center and the Alaska Fisheries Science Center. The NMFS-AFSC Technical Memorandum series is now being used by the Alaska Fisheries Science Center.

Reference throughout this document to trade names does not imply endorsement by the National Marine Fisheries Service, NOAA.

This document should be cited as follows:

Methot, R.D. 2000. Technical description of the stock synthesis assessment program. U.S. Dept. Commer., NOAA Tech. Memo. NMFS-NWFSC-43, 46 p.

Technical Description of the Stock Synthesis Assessment Program

Richard D. Methot

Northwest Fisheries Science Center
Fishery Resource Analysis and Monitoring Division
2725 Montlake Boulevard East
Seattle, Washington 98112

August 2000

U.S. DEPARTMENT OF COMMERCE

Norman Y. Mineta, Secretary

National Oceanic and Atmospheric Administration

D. James Baker, Administrator

National Marine Fisheries Service

Penelope D. Dalton, Assistant Administrator for Fisheries

Most NOAA Technical Memorandums NMFS-NWFSC are available on-line at the Northwest Fisheries Science Center web site (<http://www.nwfsc.noaa.gov>)

Copies are also available from:

National Technical Information Service

5285 Port Royal Road

Springfield, VA 22161

phone orders (1-800-553-6487)

e-mail orders (orders @ntis.fedworld.gov)

CONTENTS

LIST OF FIGURES	v
LIST OF TABLES	vi
EXECUTIVE SUMMARY	vii
ACKNOWLEDGMENTS	viii
INTRODUCTION	1
BASIC POPULATION DYNAMICS	3
Fishing Mortality	4
Selectivity Functions	4
Recruitment and Initial Age Composition	9
Migration	11
OBSERVATION MODEL	13
Abundance Index	13
Age Composition	14
Split-Sex Features	16
STATISTICAL MODEL	17
Log-Likelihood Components	17
Abundance Index	19
Age or Size Composition	19
Spawner Recruitment	20
Parameter Priors	21
SIZE-AGE MODEL	23
Overview	23
Growth Model	23
Modifications for size-specific survivorship	25
Size-at-age in accumulator age	25
Variability in size-at-age	26
Probability distribution of size-at-age	27
Size-Age Selectivity	28
Body Weight-at-Age	31
Discard and Market Categories	31
Observation Process in Size-Age Model	32
Size and age composition	32
Size-at-age	36

PARAMETER ESTIMATION	38
DISCUSSION	40
CITATIONS	41
APPENDIX A: STOCK SYNTHESIS NOTATION	43

LIST OF FIGURES

Figure 1. Examples of selectivity function shapes derived from the double logistic function.	6
Figure 2. Examples of selectivity function shapes derived from the constrained logistic function. First example is a linear segment with 3 free parameters (selectivity at min and max age, age-at-peak selectivity).	8
Figure 3. Population numbers at size/age. Values were extracted from intermediate calculations within synthesis.	29
Figure 4. Selectivity as a function of both size and age.	30
Figure 5. Population numbers at size, and sampled numbers at size after application of the size and age selectivities in Figure 2 and summing across ages.	34
Figure 6. Population age composition, age composition in sample after application of the size and age selectivities in Figure 2 and summing across sizes, and bin composition after applying the ageing imprecision matrix to the sample age composition.	35
Figure 7. Population mean size-at-age, mean size-at-age in a sample with size-selectivity as in Figure 2, and mean size-at-bin calculated after applying the ageing imprecision matrix to the sample age composition.	37

LIST OF TABLES

Table 1. Parameters used in creating double logistic selectivity functions.	7
Table 2. Log-likelihood components.	18

EXECUTIVE SUMMARY

Synthesis is an age-structured population assessment tool. It includes a population simulation model to calculate the abundance and mortality of a harvested population, an observation model to relate this population model to observable data from the population, and a statistical model to adjust parameters of the population model and observation model to achieve the best fit to all the data. Synthesis has both an age-structured and a size/age-structured version. Both are capable of simultaneously examining data from several fisheries and several surveys, each with its own pattern of selectivity. Synthesis calculates selectivity with modified logistic functions. Parameters of these functions can take on time-specific values, thus allowing flexibility to track changes in fishery selectivity. The goodness of fit is quantified in terms of a log-likelihood function composed of independent terms for each kind of observation from each type of fishery or survey. Synthesis estimates the best parameter values through numerical calculation of parameter derivatives and application of a modified Newton method. This provides estimates of parameter variances and covariances, and provides for calculation of the variance of spawning biomass through application of the delta method. Both versions of synthesis allow inclusion and estimation of spawner-recruitment functions. When detailed age composition data are lacking, the estimated spawner-recruitment curve can be used to generate the entire time series of recruitments, thus turning synthesis into a simple production model. At the other extreme, inclusion of the spawner-recruitment curve in data-rich models allows estimation of this curve while taking into account all available information. Although most population modeling is done within unit stocks, age synthesis provides the capability to model up to three geographic areas and to estimate the degree of migration between areas. Size synthesis adds a body size dimension to the population. This allows fuller use of size-based data, and allows estimation of growth curves while taking into account size-selectivity of the sampler and taking into account the effect of variability in the age determination process.

ACKNOWLEDGMENTS

I wish to thank the many users of synthesis who, over the past 10 years, have helped me to evolve this tool to best meet the needs of the stock assessment community. Special thanks to Jon Brodziak, Jim Ianelli, Chris Legault, Victor Restrepo, and Rick Deriso who provided useful recommendations on this manuscript.

INTRODUCTION

The stock synthesis model¹ (Methot 1989, 1990) was developed to provide a bridge between biomass-based assessment methods and full age-structured methods (Deriso et al. 1985; Fournier and Archibald 1982). Subsequently, synthesis evolved to a flexible tool (Methot 1990) and was used for many west coast and Alaska groundfish stock assessments during 1988-2000. Synthesis consists of a forward population projection model that simulates the dynamics of a stock within a statistical estimation framework. Between the population model and the statistical model is an observation model designed to derive expected values for various fishery and survey data. An underlying philosophy of synthesis is the modeling of potential biases and variability in the observation process, rather than requiring that data be highly preprocessed before analysis by synthesis. This “bring the model to the data” philosophy creates an appearance of high model complexity, but the underlying population model is not complex and this approach allows realistic use of a wide variety of data.

There is an age-only and a size/age version of synthesis (these will be referred to as the age and the size versions for simplicity). Both models have a large number of similar characteristics, but there are also distinct differences.

In the age version, selectivity patterns are defined by:

- functions of age,
- body weight-at-age supplied as input vectors, and
- up to three geographic areas with explicit migration patterns.

In the size version, the population model includes:

- age and size dimensions through inclusion of an explicit growth function,
- selectivity patterns as functions of size and/or age, and
- body weight-at-age calculated from the interaction of size-selectivity and the modeled probability distribution of size-at-age.

Geographic structure is not yet included in the size model. In both configurations, synthesis maintains a full age-structured description of the population and employs conventional equations to describe the population dynamics. A wide hierarchical range of model complexity can be defined for either version (Methot 1998). At one extreme, the model can be used with no age or size data (with external estimation of selectivity and weight-at-age) and the estimation process can be condensed to a few stock-recruitment parameters that mimic stock-reduction analysis (Kimura and Tagart 1982). On the other extreme, inclusion of a vector of parameters to allow time-varying fishery selectivity allows synthesis to closely track fluctuations in a complete fishery catch-at-age database.

¹ The overview of stock synthesis given here complements and updates that in Methot (1990). A more detailed user manual is available from the author.

The performance of synthesis in a variety of situations has been evaluated through its application to simulated data. Bence et al. (1993) investigated the importance of adult surveys versus recruitment surveys in stabilizing the results of model results. Sampson (1993) explored the importance of providing sufficient flexibility in the specification of fishery selectivity. Methot (1994) showed that the size model was capable of extracting information on population characteristics from size composition data. In 1996-1997, the National Research Council's evaluation of stock-assessment methods (NRC 1998) demonstrated that complex models such as synthesis were necessary to account for biases in data. Most recently, Sampson and Yin (1998) investigated the model's performance under a range of data variability and population characteristics.

BASIC POPULATION DYNAMICS

The population simulation, in its simplest form, specifies the numbers-at-age in the beginning year of the simulation, the numbers of recruits in each subsequent year, and the survival rate for each cohort as it moves through the population. These dynamics apply equally to the age and size models. Variables and notations identified here and throughout the document are listed in Appendix A.

a	ages for 1 a A	
y	years for 1 y Y	
j	fisheries for 1 j A_4	
M_a	instantaneous rate of natural mortality	
W_{yaj}	body weight-at-age for fishery or survey j	
s_{aj}	selectivity at age for fishery or survey j	
f_{yj}	annual fishing mortality factor for fishery j	
$F_{yaj} = f_{yj} s_{aj}$	fishing mortality at age for fishery j	(1)
$Z_{ya} = M_a + \sum_j (F_{yaj})$	total mortality rate	(2)
N_{ya}	population numbers at start of year y for age a	
$\bar{N}_{ya} = N_{ya} (1 - e^{-Z_{ya}}) / Z_{ya}$	mean numbers in year y	(3)
$c_{y,a,j} = \bar{N}_{y,a} F_{y,a,j}$	catch numbers for fishery j	(4)
$C_{yj} = \sum_a (c_{y,a,j} W_{yaj})$	catch biomass for fishery or survey j	(5)
$N_{y+1,a+1} = N_{ya} e^{-Z_{ya}}$	survivors, for $a < A$	(6)
$N_{y+1,A} = N_{y,A-1} e^{-Z_{y,A-1}} + N_{yA} e^{-Z_{yA}}$	survivors, for $a = A$	(7)

The above equations define a situation in which both sexes are identical and mortality is continuous throughout the year. Throughout this report, with a few exceptions, the notation and discussion will refer to age and to size, but not to sex. However, synthesis can be used in a combined sex or a separate sex mode, with the latter allowing males to differ from females in mortality, growth, and selectivity.

Also, time is referred to only as year, except for the monthly steps used to calculate growth in the size model. However, both the age and the size model are capable of defining up to four time periods within the year so that seasonal fisheries can be more accurately modeled. The month of spawning is identified so that mortality occurring during the early part of the year is accounted for when calculating spawning biomass. Similarly, each survey is assigned a specific month of occurrence so that pre-survey mortality and growth can be accounted.

Fishing Mortality

The total catch biomass, C_{yj} , is typically known with high precision relative to other types of information. Because there is a one-to-one correspondence between the level of C_{yj} and f_{yj} , the values of the f_{yj} can be continuously adjusted within the model so that the calculated C_{yj} will nearly exactly match the observed C_{yj} . In this typical case there is nil deviation between observed and expected C_{yj} , and the likelihood contribution for the fit to the C_{yj} is nil. However, alternative approaches to specifying and estimating the f_{yj} are available in synthesis. Any number of the f_{yj} can be estimated as free model parameters or assigned a fixed value; in these cases there may be some deviation between the observed and calculated C_{yj} . Finally, the f_{yj} can be made a linear function of input data on fishing effort. This latter option can be elaborated further to include predators and cannibalism in the model (Livingston and Methot 1999).

Selectivity Functions

Several approaches to specifying selectivity patterns are available in synthesis. These approaches include using one parameter for each age, selection of a single age (such as a recruitment index), and patterns based upon logistic functions. Selectivity is often modeled in synthesis as the product of two logistic functions,

$$\beta_a = \frac{T_1}{(1 + e^{-\beta_1(a-\beta_2)})(1 + e^{\beta_3(a-\beta_4)})}, \quad (8)$$

where β_1 to β_4 are parameters to be estimated by the model, and the temporary quantity T_1 is a calculated scaling factor such that $\max(\beta_a)$ is 1.0. This four-parameter “double logistic” formulation allows the selectivity pattern to be dome-shaped or asymptotic on either the left or right side (Figure 1). The parameters β_2 and β_4 have values of age (or size in the size model) and are interpreted as inflection points. The parameters β_1 and β_3 affect the steepness of these functions. When either has a low value there will be a stronger interaction between the values of the ascending and descending parameters. The synthesis implementation allows the selectivity parameters to be time-invariant, and specifically, time-invariant within a defined range of years. This synthesis implementation allows various ways to change selectivity over time, as defined by: a specific range of years, year-specific, or as a function of an independent variable (such as mean depth of fishing).

A more complex function is necessary to isolate the ascending from the descending portions of the selectivity curve. Before defining this more general selectivity function, it is useful to define a generic building block termed a “constrained logistic” function. This four parameter function is used as a building block for complex selectivity patterns, for age-specific migration functions in the age model, and for size-specific market category partitions in the size model.

The constrained logistic function, g , requires four parameters, $\beta_1, \beta_2, \beta_3, \beta_4$, and a defined age (or size) interval, T_1 to T_2 . β_1 and β_2 are the values of the function at the interval limits, T_1 and T_2 . β_3 and β_4 are parameters of the logistic function. When the slope parameter, β_3 , goes to zero, the inflection parameter, β_4 , can be ignored and the overall function becomes a simple linear interpolation between the levels β_1 and β_2 . The parameters β_1 and β_2 will both be estimable by the model in some situations, and in other situations one or the other will be set at equal to 1.0.

$$g(a) = \beta_1 + (\beta_2 - \beta_1) \frac{(f(a) - f(T_1))}{(f(T_2) - f(T_1))} \quad (9)$$

using the simple logistic function defined in eq. 10:

$$f(a) = \frac{1}{(1 + e^{-\beta_3(a - \beta_4)})} \quad (10)$$

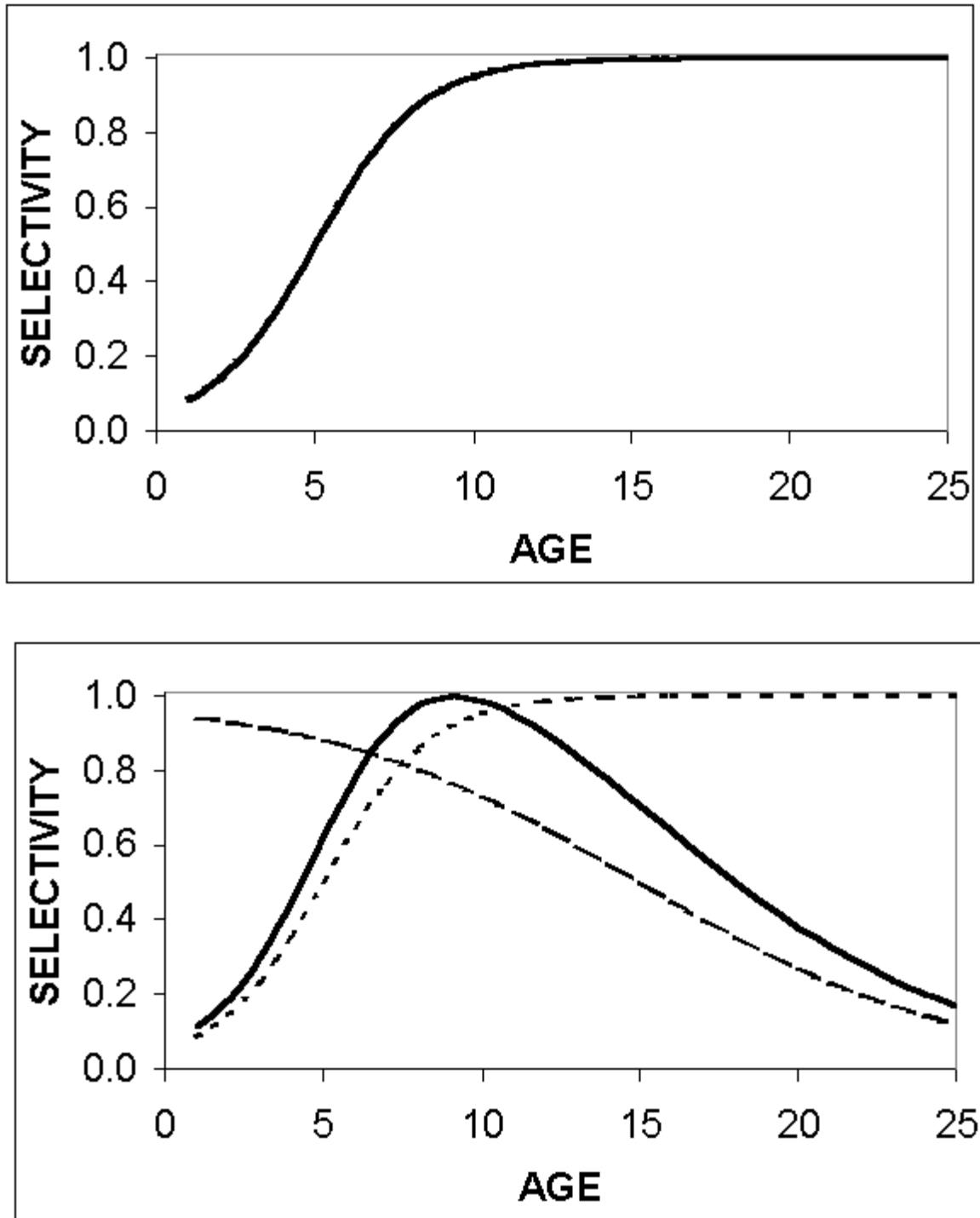


Figure 1. Examples of selectivity function shapes derived from the double logistic function. The dotted line is ascending curve, dashed line is descending curve, and solid line is the product scaled to a maximum of 1.0.

Synthesis is configured to allow the use of the constrained logistic to create ascending only, descending only, or dome-shaped selectivity functions. When it is used in the dome-shaped mode, it is necessary to use an additional parameter (TR) to define the age (size) at which the transition occurs from the ascending to the descending side of the function. In this case, TR becomes the upper limit, T2, for the ascending function; and TR+1 becomes the lower limit, T1, for the descending function. The resulting function can have zero to seven free parameters. The formulation allows use of a nested design to investigate the gain from including more free parameters. Examples of these selectivity functions are in the Table 1 and in Figure 2.

Table 1. Parameters used in creating double logistic selectivity functions.

Side	Parameter Label	Example 1: Complete (7 free parameters)	Example 2: Ascending only (3 free parameters)	Example 3: Line segment (3 free parameters)
Ascending	T1	1	1	1
	T2	TR	A	TR
	Initial selectivity	3a	3a	3a
	Slope	1a	1a	0.0
	Inflection	2a	2a	N/A
	Terminal selectivity	4a 1.0	4a 1.0	4a 1.0
Descending	T1	TR+1	N/A	N/A
	T2	A	N/A	N/A
	Initial selectivity	1d 1.0	N/A	1d 1.0
	Slope	1d	N/A	N/A
	Inflection	2d	N/A	N/A
	Terminal selectivity	4d	N/A	4d

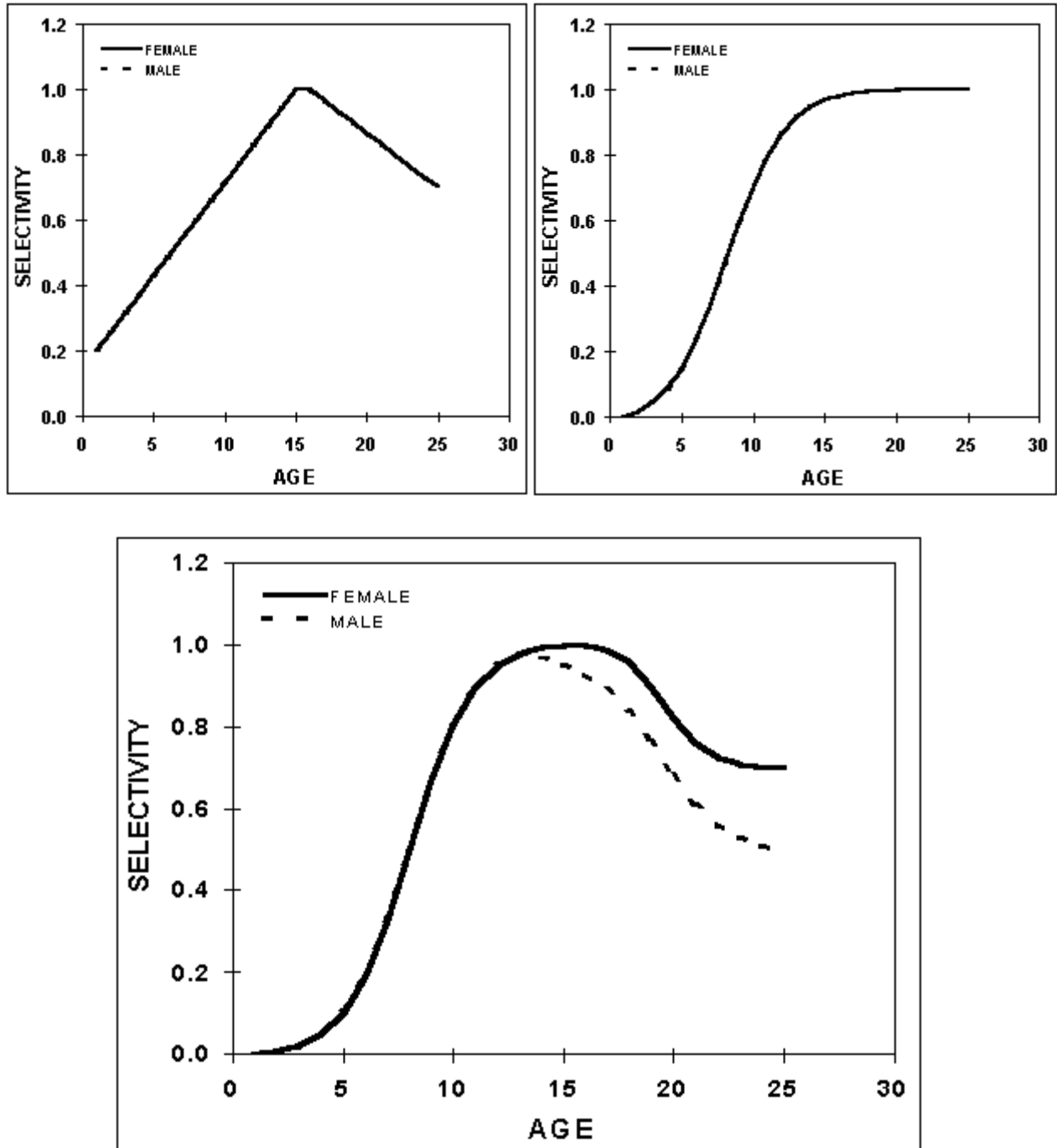


Figure 2. Examples of selectivity function shapes derived from the constrained logistic function. First is a linear segment with 3 free parameters (selectivity at min and max age, age at peak selectivity). Second is ascending selectivity with 3 free parameters (selectivity at min age, age at inflection, relative slope at inflection). Third is the flexible configuration at a maximum, including male selectivity as differing from female selectivity.

When the model is configured to have separate sexes, synthesis uses a three parameter function to define male selectivity relative to female selectivity,

$$\beta_{m,a} = \beta_a \left[1.0 + \left[\frac{(\max(\alpha, \beta_8) - \beta_8)}{(A - \beta_8)} \right]^{\beta_{10}} (\beta_9 - 1.0) \right], \quad (11)$$

where:

- 8 is the age (or size) at which male selectivity begins to deviate from female selectivity (below this age (or size) male selectivity equals that of females),
- 9 is the male selectivity at age A (or maximum size) relative to female selectivity at age A (or maximum size). (Note this can have values greater than or less than 1.0), and
- 10 is an exponent that defines the rate at which male selectivity deviates from female selectivity (a value of 1.0 will cause a linear change in the ratio of male to female selectivity).

Recruitment and Initial Age Composition

Synthesis includes the ability to define the initial age composition and all subsequent recruitments with many independent parameters (one for each year and for each age in the initial year), or to define all from a two-parameter spawner-recruitment (S-R) function. The Beverton-Holt spawner-recruitment function is defined according to Kimura (1988) as:

$$R_y = R_2 \frac{\frac{S_{y-h}}{S_0}}{1 - R_1 \left(1 - \frac{S_{y-h}}{S_0} \right)}, \quad (12)$$

where

- h is the number of years between spawning and recruitment,
- R_1 is the parameter defining the degree of density-dependence, and
- R_2 is the parameter defining level of the recruitment curve.

The initial mature female biomass (or egg production) is calculated from R_2 and natural mortality,

$$S_0 = \sum_{a=1}^{A-1} R_2 \left[\prod_{\alpha=1}^{a-1} e^{-M_\alpha} \right] W_{1a0} m_a J_{ya} + R_2 \left[\frac{\prod_{\alpha=1}^{A-1} e^{-M_\alpha}}{1 - e^{-M_A}} \right] W_{1A0} m_A J_{yA}. \quad (13)$$

In subsequent years, mature female biomass (or egg production) that spawned the fish recruiting in year y is

$$S_{y-h} = \sum_{a=1}^A N_{y-h,a} W_{y-h,a0} m_a J_{ya} , \quad (14)$$

where egg production by mature spawners is:

$$J_{ya} = J_1 + J_2 W_{ya0} . \quad (15)$$

With $J_1 = 0$ and $J_2 = 1$, egg production will be equivalent to female spawning biomass.

Intermediate options are available so that some poorly estimable recruitments can be taken from the S-R function while others are estimated as individual parameters. When all recruitments are taken from the S-R function, synthesis essentially becomes an explicitly age-structured stock-reduction analysis (Kimura and Tagart 1982). The manner in which the individually estimated recruitments are compared to the S-R function is described later.

The initial population age composition is first defined as in equilibrium with the initial recruitment level, R_{00} , and with a level of fishing mortality that is either (1) calculated as the level that would produce the initial equilibrium catch, (2) estimated as a free parameter, or (3) set equal to a user-supplied level. The initial recruitment level, R_{00} , may be set as an independent parameter, or set equal to the unfished recruitment level, R_2 , from the spawner-recruitment curve.

$$N_{1,1} = R_{00} \quad a = 1, \quad (16)$$

$$N_{1,a} = N_{1,a-1} e^{-Z_{a-1}} \quad 2 < a < A, \quad (17)$$

$$N_{1,a} = N_{1,A-1} \frac{e^{-Z_{A-1}}}{1 - e^{-Z_A}} \quad a = A, \quad (18)$$

where:

$$\begin{aligned} R_{00} & \text{ is the initial recruitment parameter,} \\ Z_a = M_a + f_{01} & \text{ is the total instantaneous mortality, and} \\ f_{01} & \text{ is the initial fishing mortality factor, calculated to generate a catch} \\ & \text{equal to the historical equilibrium catch level for fishery type 1} \\ & \text{(alternatively, this can be set as a model parameter).} \end{aligned} \quad (19)$$

After calculating the initial equilibrium age composition, zero to all of these numbers-at-age in the initial year can be replaced by a level determined by a specific parameter:

$$N_{1a} = R_{0,a} \quad \text{for } a = A_r, \text{ if } A_r > 0, \quad (20)$$

A_r last age for which individual parameter is used, $0 < A_r < A$, and
 $R_{0,a}$ parameters defining initial numbers-at-age.

Migration

In the age model there can be up to three geographic strata (areas). There are two options for distributing fish between the areas: an annual apportionment (vulnerability) approach and a true migration approach.

In the vulnerability option, the population is apportioned, on an age-specific basis, annually between the areas. Each fishery and survey operates in just one of the areas, and the population is completely mixed and re-apportioned at the beginning of the next year. The age-specific population in each area is proportional to the total population because of the complete re-mixing each year, even though all of the fishing mortality may occur in one area. This vulnerability approach has been used in the Pacific whiting stock assessment to model the annual distribution of the stock across the US-Canada border (Methot and Dorn 1995).

In the migration approach, an age-specific fraction of the fish in each area migrates annually into each of the other areas. In this approach, effects of an intense fishery in one area initially will primarily affect that area and then slowly affect other areas according to the rate of migration.

In both approaches, the constrained logistic function (Equation 9) is used to define the age-specific vector. In the vulnerability configuration, four parameters (for each sex) define the age-specific fraction found in area 1:

- $Y_{1,1}$ fraction in area 1 at $a = \text{minage}$,
- $Y_{2,1}$ fraction in area 1 at $a = A$.
- $Y_{3,1}$ inflection age of logistic function, and
- $Y_{4,1}$ slope of logistic function.

If there are three areas, then a second set of four parameters define the fractions in Area 2, with the remainder being in Area 3.

In the true migration approach, a more complex set of parameters is needed. With two areas and one sex there will be nine parameters:

- 1 parameter - fraction of recruits found in area 1 (assumed same for both sexes);
- 4 parameters - migration from Area 1 to Area 2:

- $Y_{1,1}$ fraction at $a = \text{minage}$ that move from area 1 to area 2,
- $Y_{2,1}$ fraction at $a = A$ that move from area 1 to area 2,

$Y_{3,1}$ inflection age of constrained logistic function, and

$Y_{4,1}$ slope of constrained logistic function;

4 parameters - migration from Area 2 to Area 1 (as described above).

With three areas and one sex there would have to be 26 parameters defined (although many will have values fixed by the user and not be freely estimated by the model):

1 parameter - fraction of recruits found in Area 1;

4 parameters - migration from Area 1 to 2 (as described above);

4 parameters - migration from Area 1 to 3;

1 parameter - fraction of recruits not in Area 1 that are in Area 2 (remainder are in Area 3);

4 parameters - migration from Area 2 to 1;

4 parameters - migration from Area 2 to 3;

4 parameters - migration from Area 3 to 1;

4 parameters - migration from Area 3 to 2.

OBSERVATION MODEL

Abundance Index

The abundance index (G) for a survey is expected to be proportional to the model's estimate of available biomass, B , at the time of year, y , of the survey. Thus, it is algebraically equivalent to fishery catch per unit effort:

$$B_{yj} = \sum_a N_{y+\tau,a} \beta_{aj} W_{yaj} , \quad (21)$$

$$\hat{G}_{yj} = Q_{1j} B_{yj} . \quad (22)$$

If the survey is expanded to a measurement of absolute biomass, then the constant of proportionality (catchability coefficient), Q_{1j} , is 1.0 or some other externally derived value. Even in this case, the selectivities, β_{aj} , still allow some ages to contribute less than fully to the survey. In most cases the survey is interpreted simply as a relative index of population biomass. In this case, the scaling factor is calculated so that the mean log deviation is zero:

$$Q_j = e^{\left[\frac{\sum_y \ln(G_{yj}/\hat{G}_{yj})}{\sum_y 1} \right]} , \quad (23)$$

where the summation is over the years, y , for which observations are available.

For fishery CPUE, the treatment is similar to that for a survey, however it is important to be able to consider a nonlinear relationship between the population abundance and the observed index. This is done by making the catchability coefficient, Q , a power function of population abundance.

$$Q_{jy} = Q_{1j} B_{yj}^{Q_{2j}} \quad \text{is the catchability coefficient for fishery in year } y , \quad (24)$$

$$\hat{G}_{yj} = Q_{jy} B_{yj} \quad \text{is the predicted value for fishery CPUE ,} \quad (25)$$

$$\hat{E}_{yj} = \frac{C_{yj}}{\hat{G}_{yj}} \quad \text{is the predicted value for fishery effort, } E, \quad (26)$$

$$B_{yj} = \sum_a \bar{N}_{ya} \beta_{yaj} W_{yaj} \quad \text{is the estimated biomass available to the fishery, and} \quad (27)$$

$Q_{2,j}$ is the parameter that is commonly set to 0.0, but when estimated provides for a non-linear relationship between fishery CPUE and the biomass available to the fishery.

Equation 26 shows that fishery data can be input as fishery effort rather than fishery CPUE, without any change in the population or observation model equations.

Age Composition

Age composition samples are considered to be random samples of the exploitable (i.e. selected) stock. However, observed fishery or survey age compositions are not compared directly to the model's estimate of sampled numbers-at-age, c_{yaj} . Instead, it is recognized that the observations have some level of ageing imprecision and that the observed "ages" are just bins, i , associated with some age determination method. Ageing imprecision causes strong year classes to smear into adjacent weaker year classes and flattens the overall vector of age composition (Tyler et al. 1989). Consequently, the level of ageing imprecision interacts with the variance of recruitment and the slope of selectivity functions. With high levels of ageing imprecision, the model can estimate high year-to-year fluctuations in recruitment, then blur the resulting population age composition according to the level of ageing imprecision to produce expected sample age compositions that are as blurred as the observed age composition data. Later in the size model, ageing imprecision will be shown to flatten the vector of observed mean size-at-age data also.

Bins of "observed age" can be defined in terms of nearly any transformation of true age, including transforming age to size². These bins have an arbitrary definition and can be of variable width. The transformation of true age into bins is according to an imprecision matrix. Four kinds, k , of imprecision matrices (Methot 1990) can be defined:

1. A one-to-one correspondence between a true age and an bin of measured age.
2. Normal distribution of measured age for each true age.
3. User defined matrix which can incorporate bias and imprecision in conversion of true age to measured age.
4. Transformation of true age into size bins using the growth curve and variability of size-at-age.

² In the age model, all sampling is done in terms of age. The sampled age composition can then be transformed into an estimated size composition as shown here. In the size model, the population has explicit size/age structure and the sampling in terms of age and size selectivity produces a size/age structured sample that will be collapsed across ages to get the sample size composition.

From a particular sample, $c_{y,a,j}$, the model can generate an expected value for each kind of age observation. When there are age data from multiple methods (e.g., otoliths, scales, fish size), the simultaneous inclusion of each method's observation into synthesis assists in cross-calibration of the methods. Application of an ageing imprecision matrix, Ω_{iak} , to the sampled numbers-at-age, c_{yaj} , produces an estimated bin distribution that can be compared to the observed age (or size) bin distribution, p_{yjik} .

$$\hat{p}_{yjik} = \sum_{a=1}^A \Omega_{iak} c_{yaj}, \quad (28)$$

where:

Ω_{iak} is the proportion of age a fish assigned to bin i according to the ageing method k ,

$$\sum_{i=1}^{A_{1,k}} \Omega_{iak} = 1.0, \text{ and} \quad (29)$$

$A_{1,k}$ is the number of bins defined for the k th ageing method.

The level of ageing imprecision for the second kind of age observation is commonly determined by the observed level of agreement between age readers. Synthesis allows the user to input a vector of percent agreement for each true age, or to input two parameters that define a trend in ageing precision. These two parameters can be in terms of standard deviation of observed age, percent correct, or percent agreement. The following relationships allow inter-conversion of ageing standard deviation, percent correct, and percent agreement under the assumption of a normal distribution of observed "age" at each true age:

$$\%_c = \int_{a-0.5}^{a+0.5} \mathcal{N}(a, \sigma_{4,a}^2), \quad (30)$$

$$\%_a = \%_c^2 + 2(\%_{+1})^2 + 2(\%_{+2})^2, \quad (31)$$

where:

$\%_c$ is the probability that a given assigned age will be correct,

$\%_{+1}$ is the probability that two independent ages will be incorrect by 1 year and still agree,

$\%_{+2}$ is the probability that two independent ages will be incorrect by 2 years and still agree,

- $\%_a$ is the probability that two independent ages will agree (ignoring probability of being off by more than 2 years and still agreeing), and
- $4.a$ is the standard deviation of a normal distribution of ageing imprecision.

The expected probability distribution across bins for fish that are in the accumulator age requires special consideration. The accumulator age will contain some fish that have a true age much greater than the exact accumulator age. Presumably these older fish have a lower probability of being mis-aged as younger fish. The expected bin probability distribution for fish in the accumulator age is a discounted contribution from the expected bin distributions of fish that are exactly at the accumulator age plus that of fish at older ages. The level of this discount should take into account natural mortality, fishing mortality, and declining selectivity for older fish. These potential contributions to the discount are simplified by setting the discount factor equal to natural mortality plus a value defined by a parameter specific to this purpose. Typically this discount parameter is set equal to natural mortality so that the overall discount factor is twice the rate of natural mortality.

Split-Sex Features

When synthesis is configured with two sexes, the sample matrix, c , has an additional dimension of sex. The subsequent generation of an expected value for the size composition data (or age composition data below) can occur for: combined sexes, female only, male only, or split sexes. Of these four options, only the last preserves the sex ratio information in the sample. In addition, when the sexes differ greatly in body size, mis-sexing can affect the distributions of size for each sex. The size version of synthesis includes a mis-sexing parameter to account for this phenomenon. In this case, the expected size composition for one sex is equal to the sum of the size composition for correctly sexed individuals of that sex plus the expected size composition for incorrectly sexed individuals of the other sex.

STATISTICAL MODEL

Log-Likelihood Components

The comparison between observed and expected values is quantified in terms of log-likelihood, l which is a weighted sum of separate log-likelihood components for each data source and kind of observation. The total log-likelihood is the weighted sum of the individual components (indexed by source of data, j , and kind of observation, k), plus special components for spawner-recruitment and parameter priors (Table 2):

$$\lambda = \left[\sum_j \sum_k \omega_{jk} \mathcal{L}_{jk} \right] + \omega_{R1} \mathcal{L}_{R1} + \omega_{R2} \mathcal{L}_{R2} + \omega_{\theta} \mathcal{L}_{\theta}, \quad (32)$$

where:

λ is the total log-likelihood that will be maximized,

ω_{jk} is a weighting factor for each likelihood component, and

\mathcal{L}_{jk} are the individual log-likelihood components for each fishery and survey, j , and each kind of observation, k .

Table 2. Components of log-likelihood.

index	Source	Kind	Error Structure
j,9	fishery j	catch	lognormal
j,0	fishery j	catch per unit effort (CPUE) or effort	lognormal
j,1	fishery j	perfect ages	multinomial
j,2	fishery j	imprecise ages	multinomial
j,3	fishery j	biased ages	multinomial
j,4	fishery j	size	multinomial
j,5	fishery j	mean size-at-age	normal
j,6	fishery j	age distribution in size range	multinomial
j,7	fishery j	market category distribution	multinomial
j,0	survey j	abundance index	lognormal
j,1	survey j	perfect ages	multinomial
j,2	survey j	imprecise ages	multinomial
j,3	survey j	biased ages	multinomial
j,4	survey j	size	multinomial
j,5	survey j	mean size-at-age	normal
j,6	survey j	age distribution in size range	multinomial
R1	spawner-recruit	recruitment deviations	lognormal
R2	spawner-recruit	mean and variance	normal
	priors	parameter deviations	lognormal

Abundance Index

The log-likelihood for fishery catch, fishery CPUE or fishery effort observation, or for a survey abundance observation is defined as:

$$\mathcal{L}_{j0} = -0.5 \sum_y \left[\left(\frac{\ln(G_{yj}/\hat{G}_{yj})}{\sigma_{2,yj}} \right)^2 - \ln(\sigma_{2,yj}) \right], \quad (33)$$

where:

G_{yj} is the observation, and

$s_{2,yj}$ is the standard error of $\ln(G_{yj})$. It is preferable to use standard errors estimated from sampling statistics for each G_{yj} . Alternatively, a single fixed value of $s_{2,yj}$ can be used for the entire time series. Finally, synthesis can do iterative re-weighting by using the root mean squared error (RMSE in Equation 34) of the current fit as the estimate of $s_{2,yj}$.

$$\sqrt{\frac{\sum_y [\ln(G_{yj}/\hat{G}_{yj})]^2}{\sum_y 1}} \text{ is the RMSE.} \quad (34)$$

Age or Size Composition

The log-likelihood for the fit to age composition observations from fishery or survey source j , and ageing method k , is defined according to a multinomial error structure. The multinomial error structure is appropriate because it implicitly assumes greater precision (smaller coefficient of variation) for age (or size) bins with larger proportions than for bins with smaller proportions.

$$\mathcal{L}_{jk} = \sum_y \left[v_{yjk} \sum_i \left[p_{yjik} \ln(\hat{p}_{yjik}) - p_{yjik} \ln(p_{yjik}) \right] \right], \quad (35)$$

where:

u_{yjk} is the assigned sample size for the observation,

p_{yjik} is the observed proportion in each age bin i , and

\hat{p}_{yjik} is the estimated proportion in each age bin i .

The second term in brackets depends only on the observed proportions, so it is a constant that causes the log-likelihood for the observation to approach zero from below as the model fit improves (i.e. as the \hat{p}_{yjik} approach the p_{yjik}).

Although sample size, v , enters this calculation, the absolute value of the sample size (which may be many thousands of fish measured) should not be interpreted literally because of the several additional contributions to variability in the observed age composition (Crone 1995). The recommended maximum level for the sample size, v , was 400 in Fournier and Archibald (1982). In many recent synthesis applications, a value of 200 has been used (which produces an expected coefficient of variation (CV) of approximately 20% for a bin with 10% of the distribution's mass). Where sampling statistics are able to produce an estimate of the precision of the resulting size and age composition data, then the sample sizes assigned in synthesis should be selected to mimic this level of precision (Crone 1995).

The goodness of fit between the observed and expected proportions can be used to calculate an effective sample size for the sample. The expected variance of a proportion is:

$$\text{var}(p) = \frac{p(1-p)}{v} . \quad (36)$$

$\text{Var}(p)$ is also the expected value for $(p-\hat{p})^2$, so the effective sample size for a multinomial observation can be approximated by the ratio estimator:

$$\text{effective } v = \frac{\sum_i p_i(1-p_i)}{\sum_i (p_i-\hat{p}_i)^2} . \quad (37)$$

When this approach was used for analysis of the simulated data sets provided by the National Research Council (NRC 1998), the effective sample sizes calculated by synthesis from the residuals of the age composition fits (Methot 1998) were similar to the actual sample sizes used to generate the data.

Spawner Recruitment

The likelihood components for the recruitment information is composed of two parts. The first of these components is for deviations between recruitment estimates for individual years and predicted values from the estimated spawner-recruitment (S-R) curve. Including this component in the estimation procedure tends to draw individual recruitment estimates towards the S-R curve. Algebraically, this looks no different than to posit that the S-R curve is a recruitment survey, so care must be taken in setting the level of variance, σ_R , for deviations from the S-R curve. The second component is related to the degree to which the estimated S-R curve and recruitment variability parameter are good fits for the observed mean and variability of the individual year recruitment parameters respectively. Either component may be excluded from consideration by setting a nil emphasis ($w \sim 0$).

The first component is defined as:

$$\mathcal{L}_{R1} = -0.5 \sum_y \left[\left(\frac{\ln(R_y/\hat{R}_y)}{\sigma_R} \right)^2 - \ln(\sigma_R) \right], \quad (38)$$

where:

R_y is the estimated recruitment in year y ,

\hat{R}_y is the predicted recruitment in year y from the S-R relationship, and

σ_R is the recruitment standard deviation (a model parameter).

The second component is calculated as:

$$\mathcal{L}_{R2} = -0.5 \left[\frac{(\sigma'_R - \sigma_R)^2}{\sigma_R^2/A_7} \right] - \ln(\sigma_R^2/A_7) - 0.5 \left[\frac{TI}{\sigma_R/A_7} \right]^2 - \ln(\sigma_R/A_7), \quad (39)$$

where:

$$TI \quad \text{is the sum of lognormal deviations} = \sum_y^{A_7} \ln \left(\frac{R_y}{\hat{R}_y} \right), \quad (40)$$

$$A_7 \quad \text{is the number of estimated recruitments, and} \quad (41)$$

$$\sigma'_R = \sqrt{\frac{1}{A_7} \sum_y \left[\ln \left(\frac{R_y}{\hat{R}_y} \right) \right]^2} \quad \text{is the root mean squared error of recruitments.} \quad (42)$$

Parameter Priors

A prior can be established for any parameter and entered into a log-likelihood component by:

$$\mathcal{L}_\theta = -0.5 \sum_n \left(\frac{\ln(\theta_n/\hat{\theta}_n)}{\sigma_{\theta,n}} \right)^2, \quad (43)$$

where:

n is the parameter index for which a prior is defined,

$\hat{\theta}_n$ is the estimated parameter value,

θ_n is the prior value for the parameter, and

$\sigma_{\theta,n}$ is the standard deviation for the parameter's prior.

SIZE-AGE MODEL

Overview

The previous sections describe the basic age-structured population dynamics, the observation model in age synthesis, and the statistical model in synthesis. The observation model in age synthesis includes a method to examine size composition data. In that method, a matrix containing the probability distribution of size-at-age transforms the model's estimate of the sampled true age composition into an expected value for the size composition of that sample (observation kind = 4). The size-age model differs from the age model by doing the size-at-age calculations in the population model rather than the observation model. These calculations are done by using the mean and variance of size-at-age to expand the population numbers-at-age into a matrix of population numbers-at-age and size. Then age- and size-based selectivities are applied to generate a sampled numbers matrix with age and size dimensions. The distribution of sampled numbers at size within each age is used with a weight-at-length function to calculate mean body weight-at-age for each type of sample. This differs from the age model, which treats body weight-at-age as a fixed input vector. The ageing imprecision matrix transforms the sampled numbers-at-age and size matrix into a matrix of observed numbers-at-bin and size. Marginal totals from this last matrix produce expected values for size distribution and for bin distribution. Further, the size-age model is configured to calculate the mean size-at-bin, so mean size-at-age data normally fit outside of assessment models can now be fully integrated into the assessment model while taking into account effects of size selectivity and ageing imprecision. Finally, size synthesis can calculate the bin distribution within a specified size range, which allows synthesis to compare its output to a row of an age-length key.

Growth Model

The growth model follows the von Bertalanffy growth equation as parameterized by Schnute (1981). Synthesis uses the growth curve and variability about this growth curve to calculate the probability distribution of size-at-age in each month of the year:

$$L \approx \mathbb{N}(L_{ya}, \sigma_{3,ya}^2) \quad (44)$$

where:

L_{ya} is the mean size-at-age (Equation 45), and

$\sigma_{3,ya}^2$ is the variability of size-at-age.

Growth is assumed to be continuous throughout the year and is evaluated using fractional ages in monthly steps³. For all ages in the first month of the first year (or for the recruitment age in the first month of each subsequent year) the mean size-at-age is calculated from:

$$L_{ya} = L_{\infty} + (L_1 - L_{\infty})e^{-V*(a-t_1)}, \quad (45)$$

where:

L_{ya} is the mean size at age a , at midpoint of first month (with age taking the non-integer value equal to $a + 1/24$),

t_1 is the reference age near the youngest age well represented in the data,

L_1 is the size at age t_1 (this parameter, like most parameters, can be year-specific so that different cohorts recruit with a different body size),

V is the growth coefficient, and

L is the mean asymptotic size at infinite age (calculated from Equation 46),.

$$L_{\infty} = L_1 + \frac{(L_2 - L_1)}{1 - e^{-V*(t_2 - t_1)}}, \quad (46)$$

t_2 is a reference age which should be near the oldest age well represented in the data, or set at a very large age (like 999) to let the following parameter essentially be L , and

L_2 is the size-at-age, t_2 (this model parameter can be year-specific so that the asymptotic size towards which fish grow can change over time. Note that if L_2 decreases so that L becomes less than the current mean body size of fish at a given age, this cohort retains its current size and does not shrink).

For months 2 through 12 within the year, the mean size-at-age is calculated from:

$$L_{y,a+\tau} = L_{ya} + (L_{ya} - L_{\infty})(e^{-V*\tau} - 1.0), \quad (47)$$

³ Because synthesis uses fractional ages throughout the year, data used for an externally estimated growth function must be indexed appropriately in order to produce a function appropriate for use in synthesis. Typically fish are labeled as being a particular age throughout the year no matter when during the year they are collected. For synthesis, fish need to be identified as their real age relative to a January 1 birth date. For example, a growth curve fit to samples collected in mid-summer would need to have the 4-year-old fish specified as age 4.5.

where:

is elapsed time (in decimal years) since mid-point of the first month of the year.

For the following year, mean size-at-age is calculated as:

$$L_{y+1,a+1} = L_{ya} + (L_{ya} - L_{\infty})(e^{-V} - 1.0). \quad (48)$$

Modifications for size-specific survivorship

When there is a change in fishery size-selectivity within the size range of a particular age, the size composition of the survivors will differ from the unfished size composition. The magnitude of this effect depends upon the steepness of the change in size-selectivity and the magnitude of the fishing mortality. Synthesis provides the option to ignore this effect, or to incorporate it through a change in the mean size-at-age of the survivors. The shape of the distribution of size-at-age is unchanged (no skewness or kurtosis is introduced).

The first step is to calculate the ratio, π , of mean size-at-age of survivors to the initial mean size-at-age. This calculation uses the size-at-age distribution in the month that is the middle of the time period because this is the size-at-age distribution that the fishery (with continuous mortality throughout the time period) acts upon. The time series of π for a cohort is used to modify Equation 48 which calculates the mean size-at-age of a cohort at the beginning of the next year. This modification includes an adjustment to the asymptotic size towards which the cohort is growing under the assumption that the distribution of size-at-age represents a distribution of growth potential. This adjustment to L is the product of all the previous ratios experienced by a cohort during its lifetime.

$$L_{y+1,a+1} = \pi_{ya} \left[L_{ya} + \left[L_{ya} - L_{\infty} \left(\prod_{\alpha=1}^{y-1} \pi_{y-a+\alpha,\alpha} \right) \right] (e^{-V} - 1.0) \right] \quad (49)$$

Size-at-age in accumulator age

For fish in the accumulator age, the calculation of mean size-at-age must take into account the fish that are exactly at the accumulator age, as well as older fish that are larger. The relative contribution of these older fish must be discounted by their cumulative mortality. In the initial year, this discounting uses the natural mortality rate so that the calculated mean size within the accumulator age, A , is:

$$L_{1,A} = \frac{K1}{K0}, \quad (50)$$

where the temporary variables $K0$, $K1$, $K2$, and $K3$ are calculated as:

$$K0 = \sum_{a=A}^{K3} e^{-M(a-A)}, \quad (51)$$

$K3$ is the age at which $e^{-M(K3-A)} = 0.01$, and

$$K1 = \sum_{a=A}^{K3} e^{-M(a-A)} (L_{\infty} + (L_1 - L_{\infty}) e^{-V(a-t_1)}). \quad (52)$$

Also, for later use in calculation of variability in size-at-age:

$$K2 = \sum_{a=A}^{K3} e^{-M(a-A)} [L_{\infty} + (L_1 - L_{\infty}) e^{-V(a-t_1)}]^2. \quad (53)$$

In subsequent years, the fish at age A and age $A-1$ in the previous year are each grown towards the current level of L (creating the temporary size values L' in eq. 55 and 56). The weighted sum of these two sizes is then calculated and used as the mean size-at-age A in the current year:

$$L_{yA} = \frac{N_{y-1,A-1} L'_{y-1,A-1} + N_{y-1,A} L'_{y-1,A}}{(N_{y-1,A-1} + N_{y-1,A})}, \quad (54)$$

where:

$$L'_{y,A-1} = L_{y-1,A-1} + (L_{y-1,A-1} - L_{y-1,\infty})(e^{-V} - 1.0), \quad (55)$$

$$L'_{y,A} = L_{y-1,A} + (L_{y-1,A} - L_{y-1,\infty})(e^{-V} - 1.0). \quad (56)$$

Variability in size-at-age

Variability of size-at-age typically is defined as a normal distribution with variance increasing as a linear function of mean size-at-age. Two parameters are used to indicate the level of variability at sizes L_1 and at L_{∞} , respectively. The level of variability for each age is linearly interpolated from these two parameter values and from the L_{ya} :

$$\sigma_{L,ya} = \sigma_{L,1} + (\sigma_{L,2} - \sigma_{L,1}) \frac{(L_a - L_1)}{(L_\infty - L_1)}. \quad (57)$$

In one alternative option, the calculated $\sigma_{L,ya}$ is interpreted as a CV and multiplied by the L_a to get the estimated standard deviation of size-at-age. This option allows a different rate of change in $\sigma_{L,ya}$ with respect to age, and may more closely match externally derived relationships.

In another option the variance is defined as a lognormal, rather than normal, distribution of size-at-age. This option creates a longer upper tail to the size-at-age distribution to more closely match some observed patterns. Synthesis mimics the standard deviation of the lognormal distribution with the CV of the normal distribution.

An approximation is necessary to propagate the variability in size-at-age into the accumulator age. This variability has one component from the variability of size at each age, and a second component from the variability between the mean sizes of the infinite tail of ages within the accumulator age. The following process is used to calculate the level of variability in the first year:

$$\sigma_{L,1A} = \sqrt{\sigma_{L,2}^2 + \frac{K2 - \frac{(K1 * K1)}{K0}}{K0}}. \quad (58)$$

($K0$, $K1$, $K2$ were calculated in Equations 51, 52 and 53.)

In subsequent years, variability in size-at-age in the accumulator age is calculated from:

$$\sigma_{L,yA} = \frac{N_{y-1,A-1} \sigma_{L,y-1,A-1} + N_{y-1,A} \sigma_{L,y-1,A}}{(N_{y-1,A-1} + N_{y-1,A})}. \quad (59)$$

Probability distribution of size-at-age

Subsequent to calculating the mean and variability of size-at-age, synthesis calculates the probability distribution of size-at-age and parses this distribution into user-defined size bins. These bins may be of variable width, and typically 25 bins are used. The matrix, ϕ , contains the probability distribution of size, s , for each age, a , at a point in time, y .

$$\phi_{yas} = \int_{I_{2,s}}^{I_{2,s+1}} \mathbb{N}(L_{ya}, \sigma_{L,ya}^2) \quad (60)$$

where:

I_{2s} bin boundaries for size s , and

$$\sum_{s=1}^{A_s} \phi_{yas} = 1.0. \quad (61)$$

_____The vector, N_{ya} , contains the population numbers-at-age. The product of this vector and _____ is the population numbers-at-age and size. Figure 3 illustrates the population numbers-at-age/size in a typical model configuration.

Size-Age Selectivity

Size- and age-selectivity will act on the estimated population numbers-at-age/size to produce a matrix of sampled numbers-at-age/size. Simultaneous use of both size- and age-selectivity may be necessary to model relevant phenomena. For example, small fish may have a lower probability of being retained by the fishing gear, and older fish may migrate out of the sampling area. The selectivity for a particular size/age is the product of the size- and age-selectivity, so no age/size bin may have a selectivity equal to 1.0. These selectivity values pose no algebraic problem, but may complicate interpretation of the absolute value of the catchability coefficient to which these selectivities apply. Figure 4 illustrates a size- and age-selectivity pattern similar to that in the assessment for sablefish on the U.S. West Coast trawl fishery⁴. The elements of the sampled size/age matrix depend upon the population numbers-at-age/size, size-selectivity, age-selectivity, and catchability:

$$c_{yajs} = Q_j \beta_{yaj} \beta_{ylj} \phi_{yas} N_{ya}. \quad (62)$$

⁴ Methot, R. D. 1994. Assessment of the west coast sablefish stock in 1994. *In* Status of the Pacific coast groundfish fishery through 1994 and recommended acceptable biological catches for 1995. Pacific Fishery Management Council, Portland Oregon 97201.

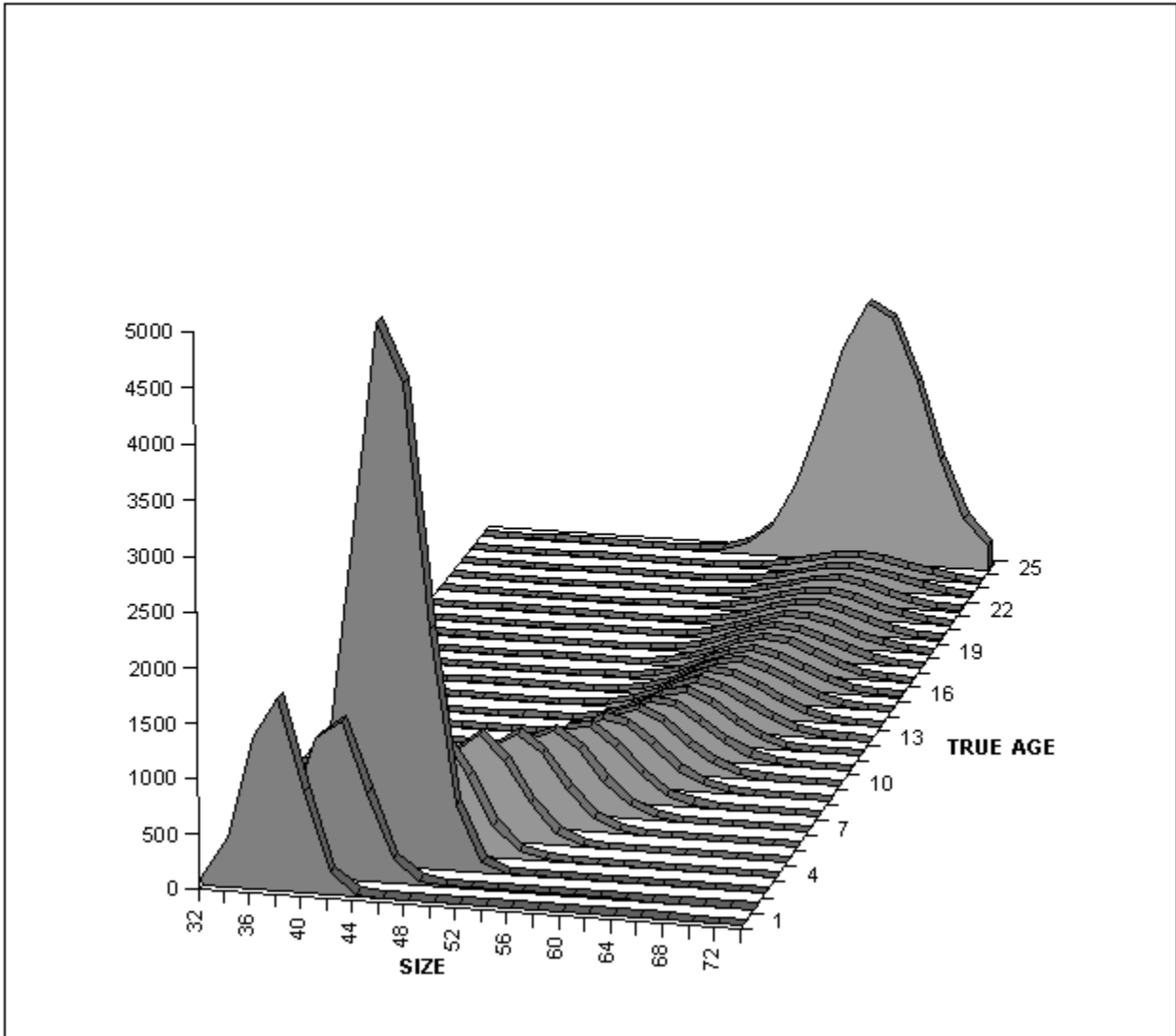


Figure 3. Population numbers at size/age. Values were extracted from intermediate calculations within synthesis. Parameters included natural mortality at 0.07, constant recruitment for all years except 4x recruitment 3 years prior to the sample, negligible fishing mortality.

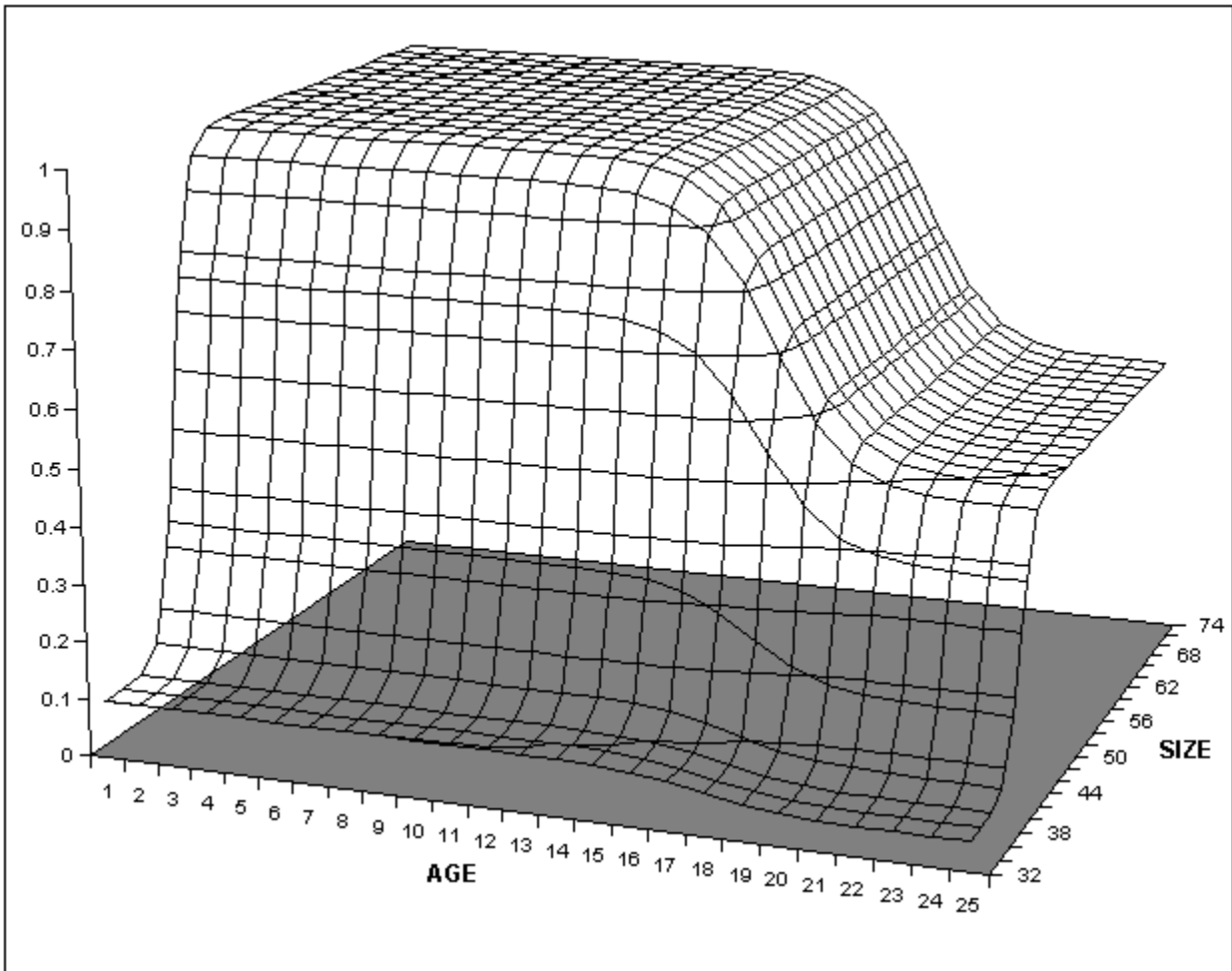


Figure 4. Selectivity as a function of both size and age. Size selectivity is 0.10 for the smallest size category and is 1.0 above 50 cm. Age selectivity is 1.0 for the youngest age and decreases according to a constrained logistic function to 0.5 at the accumulator age category. Overall selectivity is a product of size and age selectivity.

Body Weight-at-Age

The age-version of synthesis uses externally calculated values of body weight-at-age to convert population or catch numbers-at-age into population or catch biomass. Size synthesis calculates these weights-at-age from the probability distribution of size-at-age, size-selectivity, and the weight-length relationship. Because each fishery or survey has its own characteristic size-selectivity, each will have its own characteristic mean body weight-at-age.

$$W_{yaj} = \frac{\sum_{s=1}^{A_2} (W_{3,s} \beta_{ysj} \phi_{yas})}{\sum_{s=1}^{A_2} (\beta_{ysj} \phi_{yas})} \quad (63)$$

where:

$$W_{3,s} = \frac{(W_1 I_{2,s}^{W_2}) + (W_1 I_{2,s+1}^{W_2})}{2} \text{ is the mean body weight-in-size bin } s, \quad (64)$$

W_1 and W_2 are coefficients of the weight-length function, and

$I_{2,s}$ is the body size at the lower edge of the s th size bin.

Discard and Market Categories

In some situations, the catch may be partitioned (sorted) into various categories before samples can be collected. For example, there may be size-specific discard at sea before shoreside samples of retained catch can be collected. Once shoreside, price differentials may cause the catch to be sorted into market categories before random samples can be collected. Normally there is good sampling from all categories and an appropriately weighted sum of these various categories can be calculated and used as data in the assessment model. However, all categories, especially at sea discard, may not be sufficiently well sampled to be included in an expanded total size and age composition. Synthesis provides the capability to model the sorting process so that observations can be compared to expected values specific to each sampled sort category. Use of this capability requires a relatively stable sorting process.

Sort (or market) categories are defined from smallest to largest and are specific to each independent fishery. A four-parameter constrained logistic function (Equation 9) defines the size-specific fraction of a sample assigned to the smallest category; the remainder is assigned to all the larger categories. For each additional category that is defined, an additional constrained logistic function defines the size-specific fraction retained within the category, with the remainder going to the larger categories.

Thus, the catch matrix is subdivided into A_3 market categories:

$$c_{yasjm} = u_{ysjm} c_{yasj}, \quad (65)$$

where:

$$\sum_{m=1}^{A_3} u_{ysjm} = 1.0. \quad (66)$$

Observation Process in Size-Age Model

Several kinds, k , of observations can be made from each sample, just as in age synthesis. Synthesis allows size and age composition observations to be based on a specific sort category, m , if categories are defined, or on the sum across all categories.

First, the total catch biomass by a particular fishery (or abundance index for a survey) is:

$$\hat{C}_{yj} = \sum_{a=1}^A \sum_{s=1}^{A_2} (W_{3,s} c_{yasj}) = \sum_{a=1}^A (W_{yaj} c_{yaj}). \quad (67)$$

The estimated proportion (by weight) of the fishery catch biomass in a particular market category is:

$$\hat{P}_{yjm7} = \frac{\sum_{a=1}^A \sum_{s=1}^{A_2} (W_{3,s} c_{yasjm})}{\hat{C}_{yj}}. \quad (68)$$

These proportions can be used in a likelihood component (kind = 7) based on a multinomial error structure.

Size and age composition

The expected proportions-at-size (Figure 5) are calculated by summing across all ages and dividing by the total catch in the category.

$$\hat{p}_{y_{sjm}4} = \frac{\sum_{a=1}^A (c_{yasjm})}{\sum_{a=1}^A \sum_{s=1}^{A_2} (c_{yasjm})}. \quad (69)$$

The expected proportions-at-size are observation kind = 4, and are compared to the observed sample numbers at size in a log-likelihood formulation (Equation 35).

Observation kinds, $k = 1, 2$ and 3 refer to age composition in the same manner as in the age model. The expected true proportions-at-age are calculated by summing across all sizes and dividing by the total catch in the category:

$$p'_{y_{ajm}} = \frac{\sum_{s=1}^{A_2} (c_{yasjm})}{\sum_{a=1}^A \sum_{s=1}^{A_2} (c_{yasjm})}. \quad (70)$$

Then, just as with the age-based model, these proportions at true age are converted to expected proportions in each age bin, i , for each kind of ageing method, k . The population age composition, sample age composition, and sample bin composition are shown in Figure 6.

One special capability of size synthesis is the ability to examine the fit to the bin composition within a specified range of size bins (observation kind = 6). This could be relevant when age data were not collected from the entire size range, or when there was interest in examining the impact of two-staged sampling for ageing structures. The expected proportions for age composition within a range of size bins (for example size bins 3 through 8 in Equation 71) are:

$$p'_{y_{ajm}} = \frac{\sum_{s=3}^8 (c_{yasjm})}{\sum_{a=1}^A \sum_{s=3}^8 (c_{yasjm})}. \quad (71)$$

These proportions at true age are then processed by the ageing matrix corresponding to observation kind = 2. This process yields expected proportions by bin, $p_{y_{ijm}6}$, for use in a multinomial likelihood calculation. Note that if an observation kind = 6 was defined for each size bin of a sample, synthesis would effectively be fitting each row of the age-at-size matrix. This approach deserves further consideration, especially when the size based sampling is good and the value added from ageing a small number of specimens is uncertain.

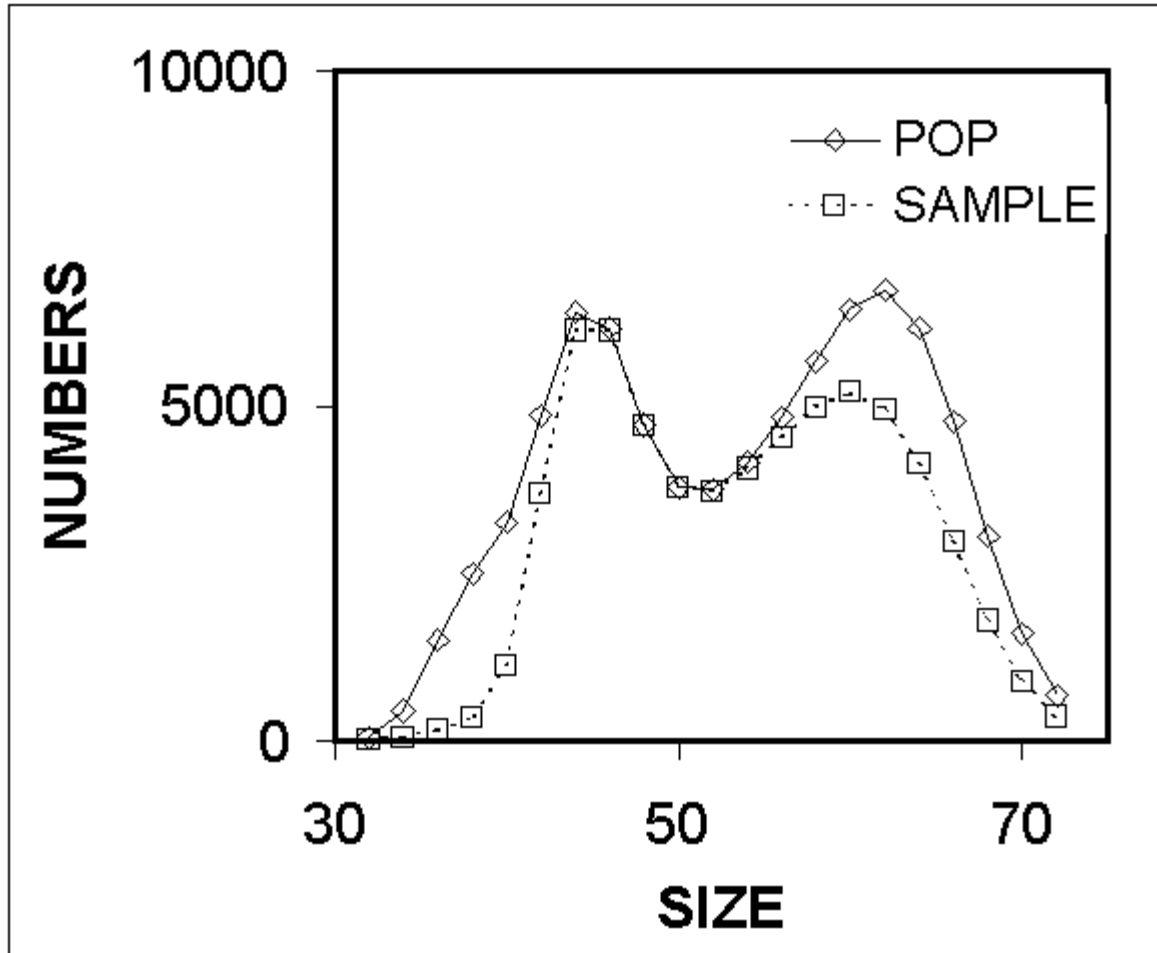


Figure 5. Population numbers at size, and sampled numbers at size after application of the size and age selectivities in Figure 2 and summing across ages.

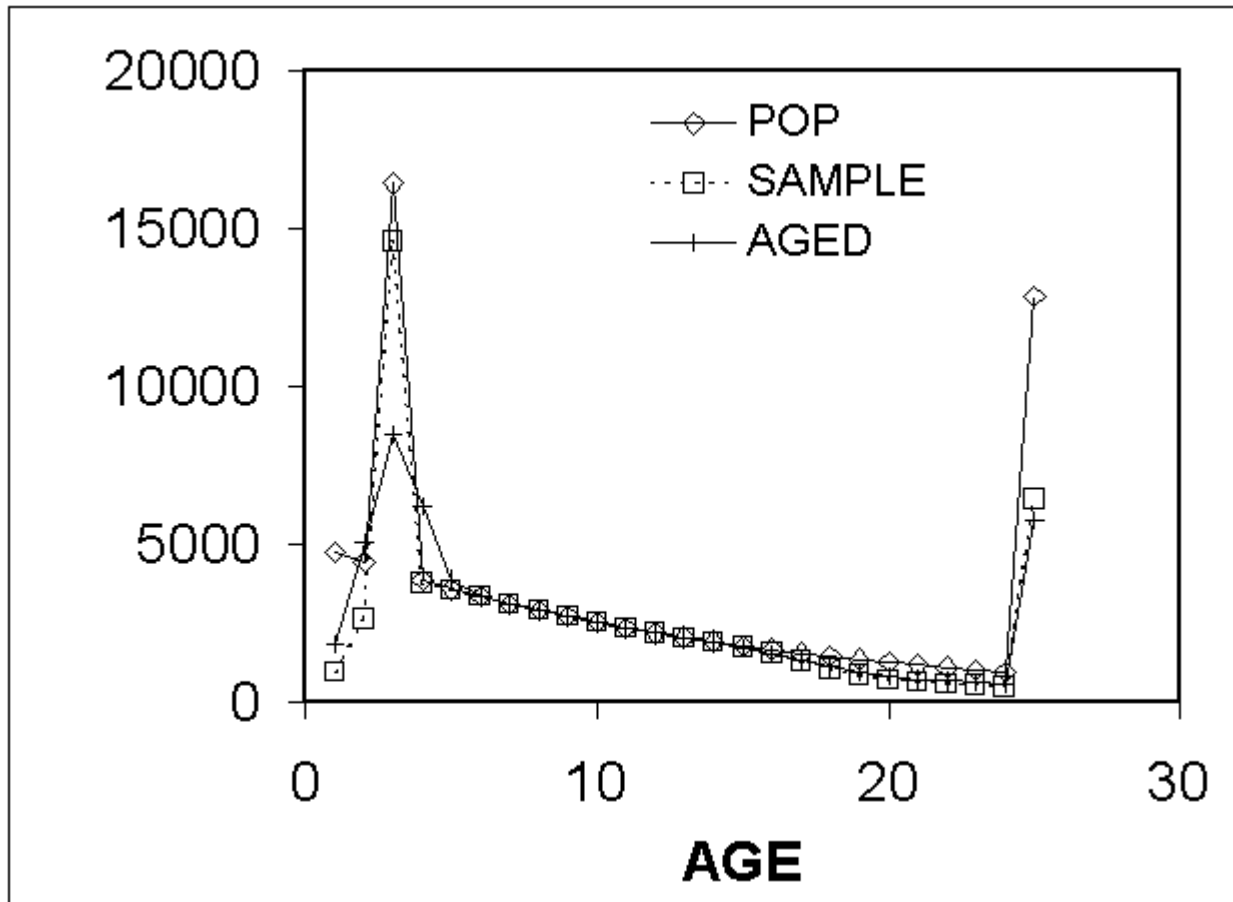


Figure 6. Population age composition, sample age composition after application of the size and age selectivities in Figure 2, and summing across sizes, and sample age-bin composition after applying the ageing imprecision matrix to the sample age composition.

Size-at-age

A final transformation of the catch age/size matrix is calculation of expected values for mean size-at-bin (observation kind = 5). Unlike conventional fitting of a growth curve to mean size-at-age data, fitting the size-at-bin data within synthesis takes size-selectivity and ageing imprecision into account (Figure 7). Expected size-at-bin will differ between each survey and fishery that has different size-selectivity. Also, ageing imprecision will make the expected mean size for a particular bin depend on the mean size and relative abundance of adjacent cohorts (Tyler et al. 1989).

$$\hat{p}_{yijm5} = \frac{X1_i}{X0_i} \text{ is the expected value for mean size in bin } i, \quad (72)$$

where:

$$X0_i = \sum_{a=1}^A \sum_{s=1}^{A_7} c_{yasjm} \Omega_{ia2}, \quad (73)$$

$$X1_i = \sum_{a=1}^A \sum_{s=1}^{A_7} c_{yasjm} \Omega_{ia2} W_{4,s}, \quad (74)$$

$$X2_i = \sum_{a=1}^A \sum_{s=1}^{A_7} c_{yasjm} \Omega_{ia2} W_{4,s}^2, \text{ and} \quad (75)$$

$$W_{4,s} = \frac{(I_{2,s} + I_{2,s+1})}{2} \text{ is mean body size in size bin, } s. \quad (76)$$

The log-likelihood for the mean-size-at-bin i observation is:

$$\mathcal{L}_{yij5} = -0.5 \left(\frac{d}{\sigma_{5,i}} \right)^2 - \text{Ln}(\sigma_{5,i}). \quad (77)$$

Where:

d is the deviation between observed mean size-at-bin and \hat{p}_{yijm5} ,

$$\sigma_{5,i} = \sqrt{\frac{X2_i - \frac{X1_i X1_i}{X0_i}}{p_{N,yijm5} X0_i}} \text{ is the expected standard error of size in bin } i, \text{ and} \quad (78)$$

$p_{N,yijm5}$ is the number of fish in the calculation of observed mean size-at-bin i .

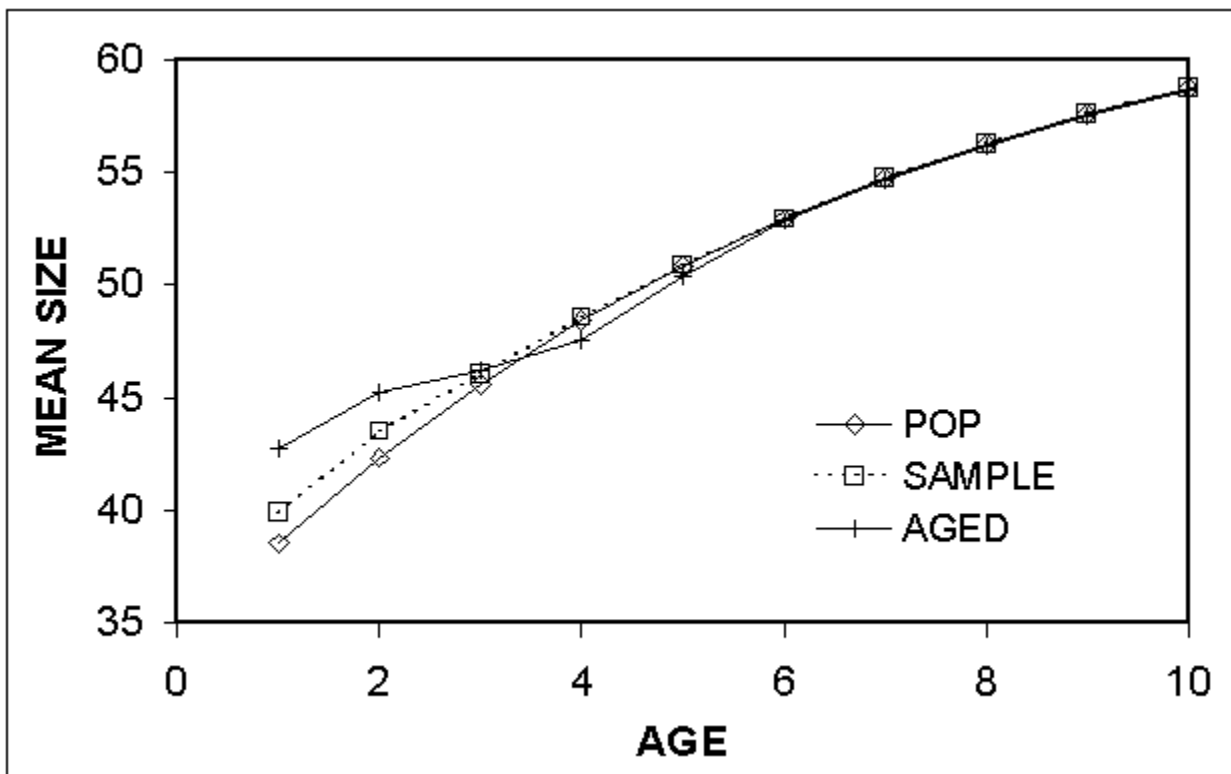


Figure 7. Population mean size-at-age, mean size-at-age in a sample with size selectivity as in Figure 2, and mean size-at-age bin calculated after applying the ageing imprecision matrix to the sample age composition. Note, that the effect of mis-ageing is to flatten the size-at-age vector, especially in the vicinity of the large, 3-year old year-class.

PARAMETER ESTIMATION

The total log-likelihood, λ , is maximized by iterative application of a quasi-Newton method using the inverse Hessian. In order to apply this method, the first and second derivatives of λ with respect to each parameter, θ , and the mixed partial derivatives for each parameter pair are calculated. These derivatives are numerically approximated by recalculating λ after a small delta, ϵ , is added to each parameter sequentially. The Hessian matrix, \mathcal{H} , has the second derivatives on the main diagonal and the mixed partials on the off-diagonal.

The second derivative (curvature) is:

$$\frac{\partial^2 \lambda}{\partial \theta_\eta^2} = \frac{\lambda(\omega, \theta, \epsilon_\eta) + \lambda(\omega, \theta, -\epsilon_\eta) - 2\lambda(\omega, \theta)}{\epsilon_\eta * \epsilon_\eta}. \quad (79)$$

The mixed partial with respect to parameters θ_η and θ_κ is:

$$\frac{\partial^2 \lambda}{\partial \theta_\eta \partial \theta_\kappa} = \frac{\lambda(\omega, \theta, \epsilon_\eta, \epsilon_\kappa) - \lambda(\omega, \theta, \epsilon_\eta) - \lambda(\omega, \theta, \epsilon_\kappa) + \lambda(\omega, \theta)}{\epsilon_\eta \epsilon_\kappa}. \quad (80)$$

Note, that the mixed partial is calculated asymmetrically around the point $\{\theta_\eta, \theta_\kappa\}$. In order to reduce potential for a bias, synthesis moves to alternative quadrants in subsequent calculations (by changing the sign on the tweaks) and it averages newly calculated mixed partials with previous values to achieve quasi-symmetric calculation.

The first derivative with respect to parameter θ_η is:

$$\frac{\partial \lambda}{\partial \theta_\eta} = \frac{\lambda(\omega, \theta, \epsilon_\eta) - \lambda(\omega, \theta, -\epsilon_\eta)}{2\epsilon_\eta}. \quad (81)$$

In a quasi-Newton method, the updated parameter values are calculated from:

$$\theta_{new} = \theta_{cur} + \mathcal{H}^{-1} \frac{\partial \lambda}{\partial \theta_{cur}} \quad (\text{Using matrix notation}). \quad (82)$$

However, \mathcal{H}^{-1} is expensive (in terms of function evaluations) to calculate. Therefore, synthesis employs two modifications to take fullest advantage of the information in \mathcal{H} . First, the off-diagonal elements of \mathcal{H} are multiplied by a scalar, ρ . At each iteration, synthesis scans up to four values of ρ , from 1.0 down to 0.0, to seek the value that maximizes the improvement in λ .

As ϵ approaches 0.0, the method ignores the mixed partials and the maximization method realizes its steepest descent. At each value of ϵ , the program seeks (by golden section search) a second scalar factor, α , which determines the magnitude of parameter changes that maximize the improvement in the likelihood. When θ and gradient are highly informative about the direction and magnitude of maximum improvement, α will have a value near 1.0 and ϵ will have a value of at least 1.0.

The parameter estimation proceeds in stages, and each stage has a progressively smaller convergence criterion defined in terms of the change in l . In a typical application, this convergence criterion has a value of 100 log-likelihood units in the first phase, and declines by a factor of 10 in each succeeding phase until it reaches a final convergence criterion of 0.01 log-likelihood units. Within each phase, the value of the each parameter's delta, δ , is adjusted, within bounds, so that the resulting magnitude of the change in l is similar to the magnitude of the convergence criterion for that phase. This adjustment causes the model to search widely in early phases when it is far from the final solution, and to calculate gradients in a much smaller area during the final stages.

Parameter variances and covariances are obtained directly from the inverse of the information matrix, \mathcal{H}^{-1} , (Richards et al. 1997). In addition, the variance of spawning biomass, S_y , is calculated through the delta method:

$$\text{var}(S_y) = \sum_{\eta=1}^{A_\epsilon} \sum_{\kappa=\eta}^{A_\epsilon} \mathcal{H}_{\eta\kappa}^{-1} \frac{\partial S_y}{\partial \theta_\eta} \frac{\partial S_y}{\partial \theta_\kappa}, \quad (83)$$

where the gradient of spawning biomass on a parameter is calculated from

$$\frac{\partial S_y}{\partial \theta_\eta} = \frac{S_y(\theta, \epsilon_\eta) - S_y(\theta, \epsilon_\eta)}{2\epsilon_\eta}. \quad (84)$$

DISCUSSION

Synthesis was initially designed to include age composition information in a simple biomass-based production model for northern anchovy (Methot 1989). Subsequently, synthesis evolved to a flexible tool to analyze a wide variety of fishery and survey data typically available for groundfish populations off the U.S. West Coast and Alaska (Methot 1990). During the mid-1990s, the age and size versions of synthesis were used for most groundfish assessments in these areas. These assessments were the primary source of information on the status of these species and their potential yield. The ability of synthesis to simultaneously incorporate a wide variety of data sources made best advantage of these data, identified inconsistencies between some data sources, and helped to identify the degree to which these west coast groundfish assessments are handicapped by infrequent and lacking data.

Synthesis assessments have analyzed available fishery and survey data and, in some cases, have included estimation of spawner-recruitment curves to both obtain information about the shape of this curve and to constrain estimates of recruitment in data-poor situations. Assessments which use the size-synthesis model have been able to estimate parameters of body growth as well as estimates of population abundance. These estimates of body growth use information in size composition and size-at-age data, and take into account size-selectivity of gear used to collect samples and imprecision in the process for determining fish age.

Fully testing the performance of synthesis has lagged behind the development of features to address the needs of current stock assessments. Tests of age synthesis include Bence and Hightower (1993), Sampson (1993), Sampson and Yin (1998), and Methot (1998). All show that the model accurately tracks known population trends when good data are available, but imprecise or biased data will lead to high variability in results. The more detailed size synthesis model performed well with high precision simulated data (Methot 1994), but an extensive test of its ability to simultaneously estimate both growth and population parameters from imprecise data is needed.

Synthesis was written in FORTRAN and its coding evolved throughout the 1990s in response to the needs of the West Coast and Alaska stock-assessment community. The coding of the conceptual model described in this document lacks the structure and clarity that can be more easily achieved with today's generation of object-oriented languages. Over the past several years, many concepts of synthesis have been recreated and enhanced in models written in a modern fourth generation language (Ianelli and Fournier 1998). In particular, the newer implementations incorporate much advancement in methods to estimate and portray variance in parameter estimates and model results. Synthesis remains in use today. Over the next few years it is expected to be fully replaced by models written in fourth generation languages which will be more powerful, more fully documented and tested, and still flexible enough to be as adaptable as synthesis has been.

CITATIONS

- Bence, J. R., A. Gordo, and J. E. Hightower. 1993. Influence of age-selective surveys on the reliability of stock synthesis assessments. *Can. J. Fish. Aquat. Sci.* 50: 827-840.
- Crone, P. R. 1995. Sampling design and statistical considerations for the commercial groundfish fishery of Oregon. *Can. J. Fish. Aquat. Sci.* 52: 716-732.
- Deriso, R. B. 1980. Harvesting strategies and parameter estimation for an age-structured model. *Can. J. Fish. Aquat. Sci.* 37: 268-282.
- Deriso, R. B., T. J. Quinn II, and P. R. Neal. 1985. Catch-age analysis with auxiliary information. *Can. J. Fish. Aquat. Sci.* 42: 815-824.
- Fournier, D., and C. P. Archibald. 1982. A general theory for analyzing catch at age data. *Can. J. Fish. Aquat. Sci.* 39: 1195-1207.
- Ianelli, J. N., and D. A. Fournier. 1998. Alternative Age-structured Analyses of the NRC Simulated Stock Assessment Data. *In* V. Restrepo (ed.), *Analysis of simulated data sets in support of the NRC study on stock assessment methods*, p. 81-96. NOAA Tech. Memo. NMFS-F/SPO-30.
- Kimura, D. K. 1988. Stock-recruitment curves as used in the stock-reduction analysis model. *J. Cons. Int. Explor. Mer* 44: 253-258.
- Kimura, D. K., and J. V. Tagart. 1982. Stock reduction analysis, another solution to the catch equations. *Can. J. Fish. Aquat. Sci.* 39: 1467-1472.
- Livingston, P., and R. D. Methot. 1999. Incorporation predation and cannibalism into assessment model. *In* Proc. of the 15th Lowell Wakefield Fish. Symp.: Fishery Stock Assessment Models for the 21st Century. 98:01. Univ. of Alsk. Sea Grant College Program. Anchorage, AL.
- Methot, R. D. 1989. Synthetic estimates of historical abundance and mortality for northern anchovy. *In* Edwards, E. and B. Megrey (eds.), *Mathematical analysis of fish stock dynamics: reviews and current applications*. *Am. Fish. Soc. Symp.* 6: 66-82.
- Methot, R. D. 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. *Int. North Pac. Fish. Comm. Bull.* 50: 259-277.
- Methot, R. D. 1994. Estimates of growth and abundance from simulated size composition data using the stock synthesis model. *Int. Comm. Conserv. Atl. Tuna Working Document*, SCRS/94/66.

- Methot, R. D. 1998. Application of Stock Synthesis to NRC test data sets. *In* V. Restrepo (ed.), Analysis of simulated data sets in support of the NRC study on stock assessment methods, p. 59-80. NOAA Tech. Memo. NMFS-F/SPO-30.
- Methot, R. D., and M. W. Dorn. 1995. Biology and fisheries for North Pacific hake (*M. productus*). *In* J. Alheit and T. J. Pitcher (eds.), Hake: Fisheries, ecology and markets. Chapman & Hall, London, 496 p.
- National Research Council. 1998. Improving Fish Stock Assessments. National Academy Press. Washington, D.C., 177 p.
- Richards, L. J., J. T. Schnute, and Norm Olsen. 1997. Visualizing catch-age analysis: a case study. *Can. J. Fish. Aquat. Sci.* 54: 1646-1658.
- Sampson, D. B. 1993. The assumption of constant selectivity and the stock assessment for widow rockfish. *Fish. Bull.* 91: 676-689.
- Sampson, D. B., and Y. Yin. 1998. A Monte Carlo evaluation of the Stock Synthesis assessment program. *In* Proc. of the 15th Lowell Wakefield Fish. Symp.: Fishery Stock Assessment Models for the 21st Century, p. 315-338. Univ. of Alsk. Sea Grant College Program. Anchorage, AL.
- Schnute, J. 1981. A versatile growth model with statistically stable parameters. *Can. J. Fish. Aquat. Sci.* 38: 1128-1140.
- Tyler, A.V., R. J. Beamish, and G. A. McFarlane. 1989. Implications of age determination errors to yield estimates, p. 27-35. *In* R. J. Beamish and G. A. McFarlane (ed.), Effects of ocean variability on recruitment and an evaluation of parameters used in stock assessment models. *Can. Spec. Publ. Fish. Aquat. Sci.* 108.

APPENDIX A: STOCK SYNTHESIS NOTATION

This section contains a complete list of quantities referred to in the text and equations of this report. The quantities are grouped into six categories. For the last 3 categories, the “type” of quantity is identified as P = parameter that can be kept constant at its input value or included in the list of quantities estimated by the model fitting procedure, I = fixed input coefficient., blank = internally calculated quantity.

Indices and index ranges

a	age
y	year
s	length (size) bin
i	bin for observed age
m	bin for market category
j	index for source of data (fisheries and surveys)
k	index for kind of age/size observation
,	indexes for parameters
r	index for geographic area
h	age at recruitment
A	accumulator age-class for population
$A_{1,k}$	number of bins for observed age by ageing method k
$I_{1,k}$	bin boundaries for ageing method k
A_2	number of bins for size
$I_{2,s}$	bin boundaries for size s
A_3	number of market categories
A_4	number of fisheries
A_5	number of fisheries and surveys
A_6	number of parameters
A_7	number of estimated recruitments
AR	number of ages in initial age composition for which non-equilibrium numbers-at-age are estimated
Y	number of years

Data

C_{yj}	catch biomass for fishery j in year y
G_{yj}	CPUE for fishery or survey j in year y
E_{yj}	fishing effort for fishery j in year y
P_{yijm1}	proportion in bin i from market category m for fishery or survey j in year y , with no ageing error ($k=1$)
P_{yijm2}	proportion in bin i from market category m for fishery or survey j in year y , with normally distributed ageing error ($k=2$)
P_{yijm3}	proportion in bin i from market category m for fishery or survey j in year y , with user-defined ageing error and bias ($k=3$)
P_{ysjm4}	proportion in size bin s from market category m for fishery or survey j in year y ($k=4$)
P_{yijm5}	mean size-at-bin i from market category m for fishery or survey j in year y , with normally distributed ageing error ($k=5$, size model only)
$P_{N.,yijm5}$	sample size for mean size-at-bin i from market category m for fishery or survey j in year y (size model only)
P_{yijm6}	proportion in bin i from market category m for fishery or survey j in year y , with normally distributed ageing error and within a specified range of size bins ($k=6$, size model only)
P_{yjm7}	proportion (by weight) in market category m for fishery j in year y ($k=7$, size model only)
y_{jmk}	sample size used to scale the variability of age and size composition proportions

Parameters and quantities used in estimation

parameter vector, which contains variables identified with a **P** in this table

total log-likelihood

weighting factor for a log-likelihood component

negative Hessian matrix of mixed partial derivatives of with respect to

log-likelihood for a component

Biological Characteristics**TYPE**

M_{ya}	natural mortality	
M_1, M_2, M_3	one to three parameters to define natural mortality as $f(a)$	P
m_a	proportion mature	
m_1, m_2	two coefficients for logistic maturity function	I
W_{yaj}	body weight-at-age for source j in year y	
W_1, W_2	two coefficients for weight-at-length function	I
$W_{3,s}$	mean body weight in size bin s	
$W_{4,s}$	mean body size in bin s	
$J(W)$	eggs per kg body weight for mature spawners (linear function)	
J_1, J_2	two coefficients for egg function	I
L_{ya}	mean body length at age a in year y	
t_1, t_2	two ages at which growth curve parameters are indexed	I
L_1, L_2	two parameters for size at ages t_1 and t_2	P
V	von Bertalanffy growth parameter	P
L	mean asymptotic size at infinite age	
$L_{3,s}$	body size at lower edge of size bin s	I

Population States and Processes**TYPE**

N_{ya}	numbers-at-age at beginning of year y	
\bar{N}_{ya}	mean numbers-at-age in year y	
N'_{ya}	numbers-at-age at the specified time of a particular survey in year y	
N_{1a}	numbers-at-age in first year for ages A_R	P
S_y	female egg production (or spawning biomass) in year y	
R_y	recruitment to first age-class in year y	P
R_{00}	initial equilibrium recruitment level	P
$R_{0,a}$	initial numbers-at-age for $a \in A_r$	P
R_1	density-dependence in spawner-recruitment function	P
R_2	recruitment level in spawner-recruitment function	P
f_{yj}	fishing mortality factor for fishery j in year y	P

y_{aj}	selectivity for age a (or size s) in fishery or survey j in year y	
$_{1,j}$ to $_{10,j}$	selectivity parameters for fishery or survey j	P
T_1 and T_2	age or size interval over which a selectivity function is calculated	I or P
F_{yaj}	fishing mortality rate for age a in fishery j in year y	
Z_{ya}	total mortality rate for age a in year y	
B_{yj}	exploitable (available) biomass for fishery or survey j	
$Y_{rr'a}$	proportion of age a in area r migrating to area r in year y	
$Y_{1,r}$ to $Y_{x,r}$	parameters defining migration vectors	P
$Q_{1,j}$	catchability coefficient for fishery or survey j	P
$Q_{2,j}$	power coefficient for CPUE in fishery j	P
$_{yas}$	proportion of population in length bin s for age a in year y (size model)	
c_{yajs}	catch numbers at age a and length s from fishery or survey j in year y	
p_{yajm}	proportion at true age a from market category m for fishery or survey j in year y	
$_{iak}$	proportion of age a assigned to bin i by ageing method k	I or P
$_{sa4}$	proportion of age a assigned to size bin s by ageing method 4 (age model)	I

Standard Deviation

		TYPE
R	standard deviation of recruitment process error	P
$'_R$	RMSE of estimated recruitments	
$_{2,yj}$	standard deviation of CPUE for fishery or survey j	I
$'_{2,j}$	RMSE of CPUE for fishery or survey j	
$L_{,ya}$	standard deviation of size at age in year y	
$L_{,1}$ $L_{,2}$	two parameters from which $L_{,ya}$ is calculated	P
$_{4,a}$	standard deviation of bin assignment at true age a	
$_{4,1}$ $_{4,2}$	two parameters from which $_{4}$ is calculated	P
$_{5,i}$	standard deviation of size in bin i	
$_{,n}$	standard deviation for prior on parameter n	I