Generic size based stock assessment model for Alaskan crab stocks

Prepared by:
Mark Maunder, Ph.D.
Quantitative Resource Assessment, LLC

Prepared for:
Bering Sea Fisheries Research Foundation
23929 22nd Drive SE
Bothell, WA 98021

August, 2012

Abstract
This report documents the equations needed to implement a generic size-based crab stock assessment model that can be used to assess Alaskan crab stocks. The model has flexibility similar to other general stock assessment models in terms of data used and processes modeled. There are several different individual characteristics (size, gender, maturity status, location, and shell condition), called partitions, that are modeled to represent the population dynamics or available data. Not all partitions need to be present in an application. Population and fishing processes can differ or be shared among or within partitions. A seasonal time structure is described to allow flexibility in how many times a process occurs a year and the order and timing of the processes. Emphasis is given to the reuse of size-based functions for different population dynamics and fishery processes. The model should be implemented using AD Model Builder and components should be implemented as general functions that can be reused by different parts of the model and by other ADMB modelers.

Overview
The following are the equations needed to implement a generic size-based crab stock assessment model that can be used to assess Alaskan crab stocks. The model has flexibility similar to other general stock assessment models. Many of the features of this model are taken from the Stock Synthesis general stock assessment program (Methot 2009; Methot and Wetzel in press). The major difference between the generic crab model

and Stock Synthesis is that the crab generic model does not keep track of age and therefore all processes are based on the size of the individuals. A main advantage of the size-based model is that size-based processes such as selectivity can change the (unmodeled) size-at-age composition of the population. In Stock Synthesis, like most other age-structured models, the distribution of size at age is not impacted by these size-based processes (except for specific modeling of growth) unless approximated using the Stock Synthesis "platoon" option. Despite these differences, many components of Stock Synthesis can be used in a size-based model (e.g. length-based selectivity curves, general control setup for parameter estimation options). An effort has been made to follow formats of Stock Synthesis to retain some familiarity for the user.

This description is only a guideline for the development of the general crab model. Development of the code and application of the model will identify modifications and additions that are not described here. It will therefore be necessary to apply the generic model to several case studies to beta test the code. Steve Martell has set up a repository for the crab models that he and Jim Ianelli have been developing and they have kindly offered the use of the code that they have developed.

One of the major tasks in creating a generic model is to allow for the many different modeling options that users need to implement in their stock assessments. There are several different individual characteristics (e.g. gender, maturity status, location) that may need to be modeled to represent the population dynamics or available data (see Figure 1). These characteristics are called partitions (e.g. gender) in the terminology of this description and the states within these partitions are called partition categories (e.g. female or male). Not all assessments will have the same data components or adequate information to model all these characteristics. Therefore, each application will include different characteristics depending on the data available and therefore may not use all the partition types. Stocks with different characteristics might have different population dynamics requiring different parameter values. To implement a flexible general framework, a model that partitions the population into individuals of different characteristics, using size-based functions for the population dynamics and fishery processes, is required. A general implementation should allow the same size based functions to be used for any process. A general approach also allows modeling differences in the processes among partition categories (e.g. between females and males or immature and mature). Some of the main components of this approach are

a) Processes for all categories in a partition are automatically given a value of one for each size bin. This facilitates defining the processes as a product of multiple quantities.
b) The size-based effect is the product of the effect for all appropriate partition categories from the different partitions.
c) Multiple size-based functions can be applied to the same partition (e.g. to model selectivity and availability)
d) Each size-based function can be applied to multiple partition categories
e) Several transformations are applied to the size-based functions depending on the processes being modeled
    a. An additional estimated scaling parameter

b. A logistic transformation to ensure the value is between zero and one
c. Scaling so the maximum is one
d. Scaling so that the sum across all size bins is one

f) Both functions (e.g. logistic) and probability distributions (e.g. gamma) are treated the same and can be used for all processes (i.e. for both functions (e.g. selectivity) and distributions (e.g. proportioning recruitment to size categories)).

One approach to implement the size-based processes is based on the following inputs:

i) Number of size functions for the process
ii) Number of partitions (and/or fisheries) that use each size function
iii) The partitions and partition categories that use each size function (a special value e.g. 0 or -1 can be used to specify that all partitions or categories within a partition use this selectivity curve)
iv) Parameter controls for each of the parameters of each size function

Implementation

The model should be implemented using AD Model Builder (Fournier et al. 2012). As much as possible, components should be implemented as general functions that can be reused by different parts of the model and by other ADMB modelers. The functions and their documentation should follow the format outlined by the ADMB project (http://www.admb-project.org/developers/using-doxygen). The size structured model is accommodated well by matrix algebra and in many cases matrix algebra will be efficient in either the memory used or the number of calculations performed, or both. Therefore, matrix algebra should be used where possible. When implementing the matrix algebra care needs to be taken when setting up the transition matrices, the order that the calculations are conducted in, and the use of multi-dimensional matrices for representing the different population characteristic partitions. A basic population dynamics model can be represented in matrix notation

$$N_t = (\boldsymbol{G\varphi_{t-1}})N_{t-1} + R_t \qquad \text{(Eq. 1)}$$

Where
N_t is a column vector of numbers at size at the start of time *t*
G is the growth transition matrix where the columns represent the proportion of individuals transitioning to the different size classes from a single size class (i.e. they sum to one)
φ is the survival matrix with the diagonals equal to survival for each size class
R_t is a column vector of recruitment for each size class in time *t*

The implementation of the general sized based model will be more complicated due to the various population characteristic partitions that are used. The program may have to loop over partitions implementing the population dynamics model for each partition and model transition among partitions.

Model structure

The individuals represented in the model can be differentiated by several characteristics, which are called partitions (or structures) (Figures 1 and 2; Table 1). The model allows the processes acting on these partitions to differ among categories (e.g. male-female; mature-immature) within partitions (e.g. gender, maturity status) (i.e. the parameters used to define the processes can be shared or be different among categories within partitions). Size is a required partition, while the other partitions are all optional. This partition flexibility avoids representing unused elements in the model. The partitions will need to be treated such that different combinations of partitions may have different characteristics (similar to an interaction term in a GLM). This could be implemented by having different size- specific functions for each partition category and multiplying them together to form the overall effect for the individuals represented by the combination of partition categories. Some combinations of partitions may not be permissible in some applications (e.g. old-shell and immature). A large programming task will be developing indexing to link the parameters and transition matrices with the modeled partitions. The indexing will have to include a time component to allow for time-varying parameters. A terminal molt option needs to be implemented and could be implemented as part of the shell condition or maturity partitions. Maturity can be explicitly modeled as a partition by modeling the proportion maturing by size or as simply a maturity schedule (the proportion mature-at-size). In the notation below $f(L)$ refers to a general function of size.

Table 1. The model partitions and associated partition categories.

Partition	Categories
Size (length)	User defined
Sex	Male, Female
Maturity	Immature, Mature
Shell condition	New Shell, Old Shell I, Old Shell II, …
Area (zone)	User defined

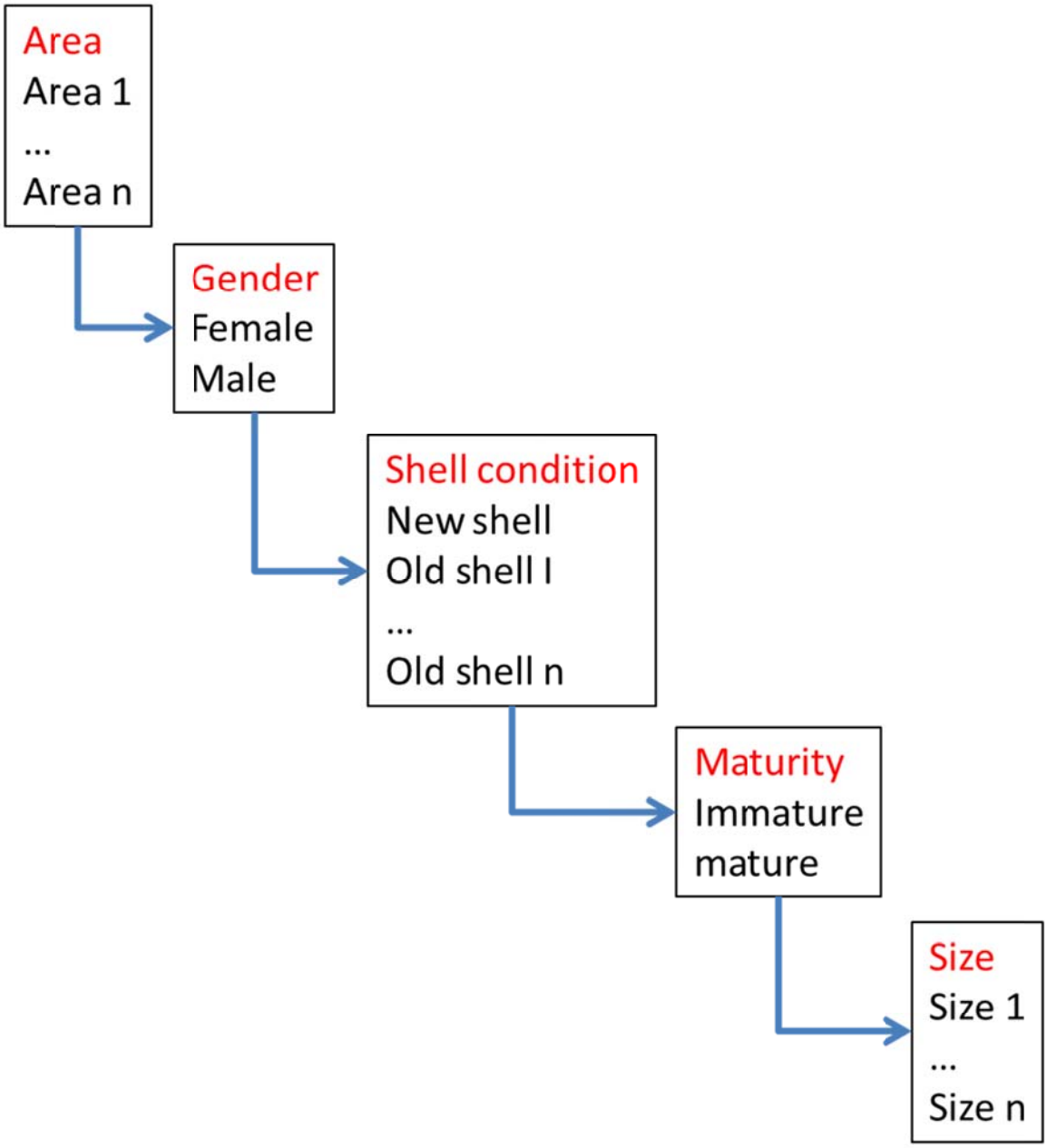

Figure 1. Hierarchical diagram of the different model partitions (structures). In the model implementation, each model partition described in a box will be repeated for each model partition to its immediate left. The model partitions described in a single box should not differ for each category in the partition to it's immediate left (this constraint may need to be relaxed, but will complicate the programming). Not all partitions need to be modeled in a single application.

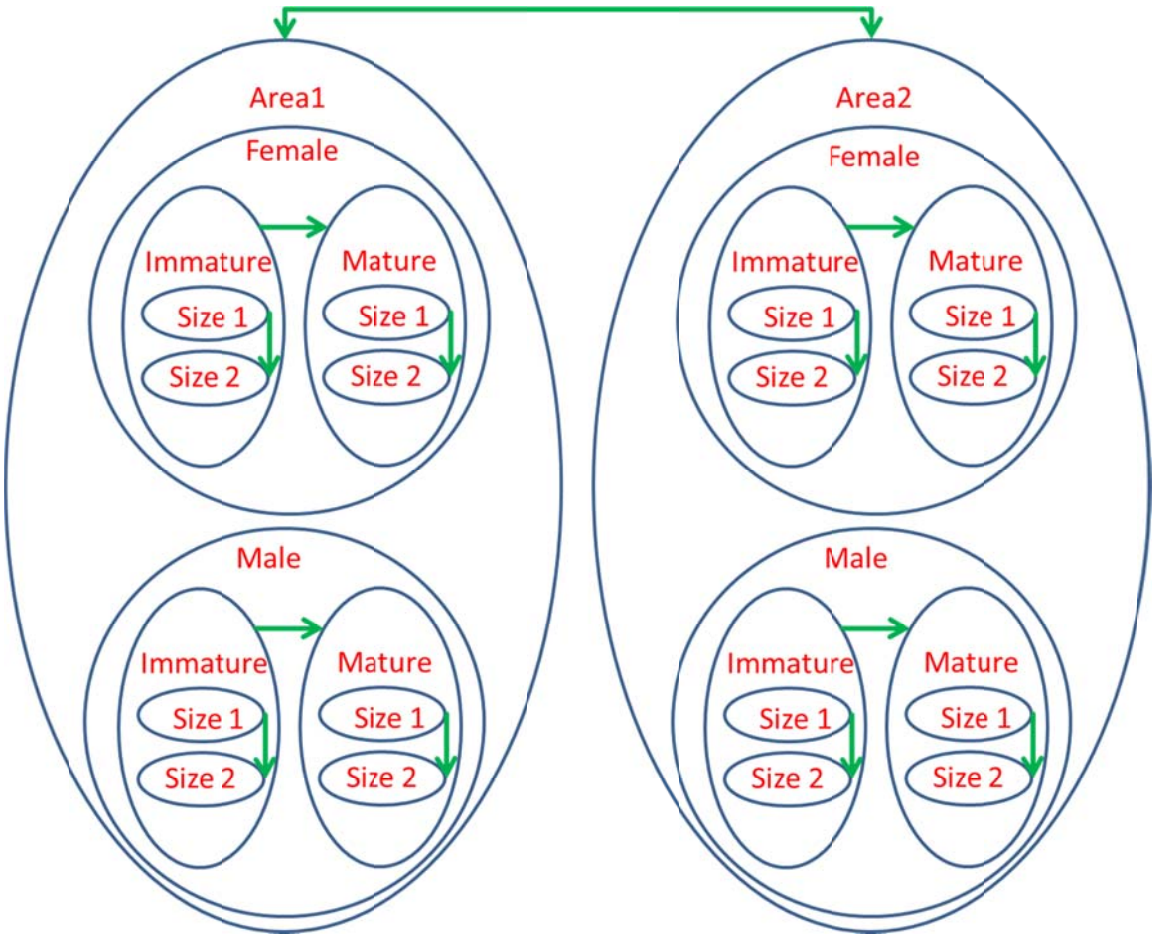

Figure 2. Illustration of the nesting of partitions (structures) for a model with two areas, sex, maturity, and size (two size classes) partions, but no shell condition partition. Possible transitions are indicated by the arrows. The diagram assumes that individuals can not shrink in size and mature individuals cannot become immature. Possible shell condition transitions include transitions from newshell and oldshell I (perhaps to other oldshell conditions as well), transition from one oldshell condition to the next older oldshell condition (perhaps also to older shell conditions), and from any oldshell condition to newshell.

Seasons

The model should have two types of time-steps (see Figure 3). The first time step relates to aging, which will most likely be year (referred to by year and indexed by y in the equations), and the second time step will model intermediate processes between aging events (e.g. seasons or months; referred to by season and indicated by t). The year time step allows intermediate processes to be repeated. Not all processes have to occur in each season time step, so the season time-step has a flexible number and order of processes. The input data used to control the season time step includes (based on the example in Figure 3)

#Number of seasons
3
#Number of processes per season
2 3 2
#Processes per season (input data correspond to events listed below in the order they occur)
5 1
3 4 1
1 2

The program will essentially loop through a vector of index values that identify the processes to be modeled in that season. The duration of each season time-step does not have to be equal (the main point is to get the magnitude of the processes within a time step correct). In general, the proportion of natural mortality assigned to the time-step will determine the length of the time-step.

Possible events within a season

1) Survival (fishing and natural mortality)
2) Movement
3) Growth (including transition among old-shell and new-shell stages as well as size classes)
4) Maturity
5) Recruitment

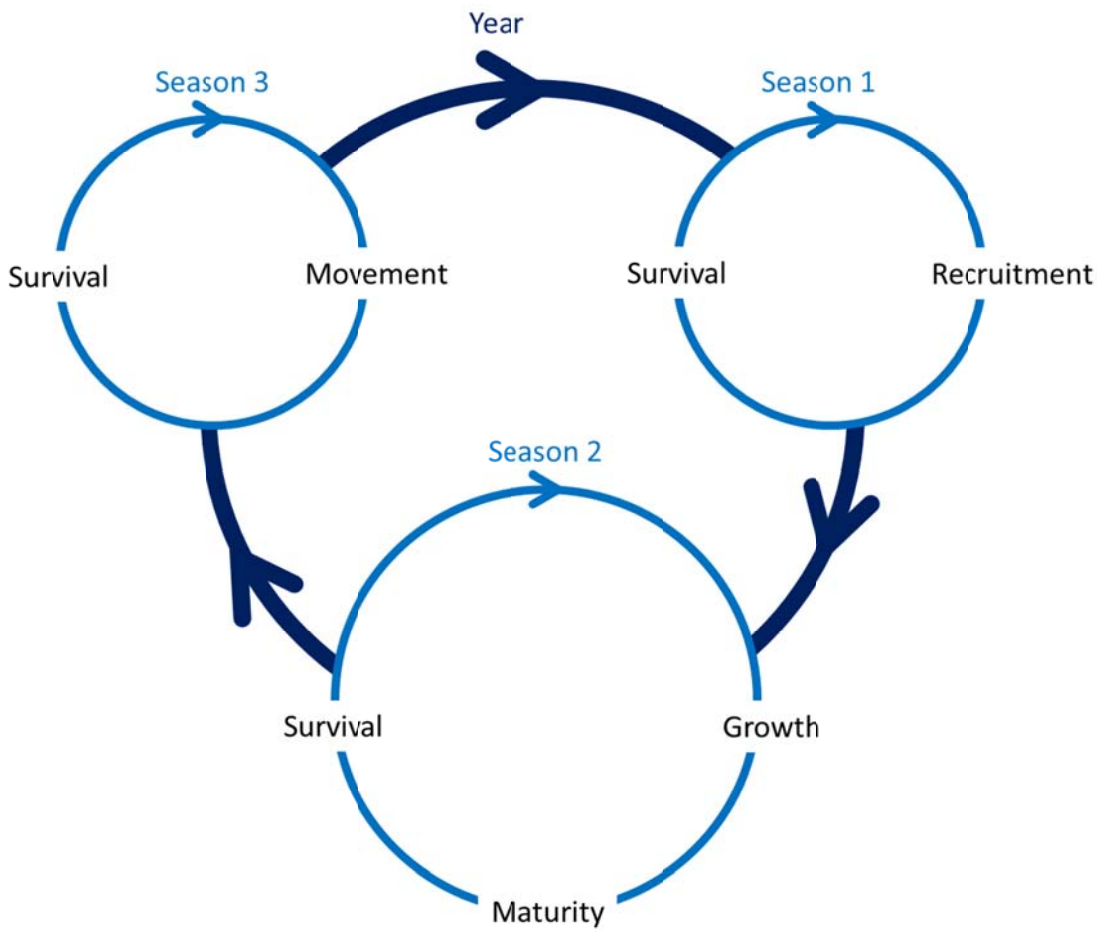

Figure Z. Illustration of the different temporal components in the model. A year can be comprised of one or more seasons. The seasons can be of different lengths. Each season can have one or more processes. Process can be different among seasons and can be repeated in multiple seasons. The order of the processes does not have to be the same in each season. Survival includes both fishing and natural mortality because they can be modeled as occurring simultaneously. The proportion of the annual natural mortality occurring in a season generally represents the length of the season. Natural or fishing mortality can be set to zero for a particular season to eliminate that mortality component even though it is still modeled.

Notation

Population partition (structure) indexing

The notation will include time (y and t) and size (l) indices as subscripts and other indices as superscripts (g = gender, m = maturity, o = shell condition, z = area). The notation for the number of individuals will simply be before and after the processes independent of

the process being modeled (the number of individuals after the process being notated with a "~").

$$N_{y,t,l}^{g,m,o,z}$$

Table 2. Definition of variables used in the model equations

States
N numbers at the start of a process
$\widetilde{N}$ numbers at the end of a process
R recruitment

Transition among partitions (population dynamics processes)
ψ movement
P probability of molting
G growth given molting
Φ survival
Ω maturity

Other quantities
M natural mortality
F fishing mortality
E fishing effort
S selectivity
V retention
T number of time periods
I index of abundance
C catch
p proportion in catch
W average weight by size
L average size (length)
B biomass
q catchability
d is the survival rate of discarded crab
η is the nonlinear relationship parameter for the index of abundance
Δ is the proportion of the lognormal bias correction factor used for the recruitment deviate
τ is the variance weighting factor for the Fournier et al. (1998) composition likelihood function.
δ is the fraction of natural mortality for that time period (season)

Index
l size class
y year
t season (within year time periods)
g gender (male/female)

m maturity (mature/immature)
o shell condition (old shell/new shell or shell conditions I, II, III, …)
z zone (area)
f fishery (or survey where relevant)

General model transition

The following equations describe the main population dynamics processes. To improve readability the equations ignore subscripts and superscripts for the partitions, unless needed for clarification, and therefore require some additional interpretation. A period "." is used to indicate that all categories of that partition are included in the array. In cases where multiple partitions are modeled, the equations would need to be repeated for each partition. In most cases *N* is a vector of numbers-at-size unless noted by an *l* subscript in which it is scalar. There are two types of transition matrices: 1) Partitions that do not allow transitions among categories within a partition for which all non-diagonal terms are set to zero. The diagonal terms are set equal to the appropriate values (e.g. survival). 2) Partitions that allow transition among categories within a partition for which the non-diagonal terms can be non-zero (e.g. growth). Differentiation between these two types of partitions may allow for more efficient programming. The tilde notation indicates the abundance after the process, but does not follow through to the equation for the next process.

Survival

$$\widetilde{N}_{y,t,.}^{g,m,o,z} = \Phi_{y,t,.}^{g,m,o,z} N_{y,t,.}^{g,m,o,z} \qquad \text{(Eq. 2)}$$

Where Φ is a diagonal matrix with the diagonal terms equal to the proportion surviving at size, off diagonal terms equal to zero, and dimensions equal to the number of size bins.

Movement

$$\widetilde{N}_{y,t,l}^{g,m,o,.} = \psi_{y,t,l}^{g,m,o,.} N_{y,t,l}^{g,m,o,.} \qquad \text{(Eq. 3)}$$

Where *N* is a vector of numbers in each zone, and ψ is a square matrix with values equal to the proportion of individuals moving among areas and dimensions equal to the number of zones. ψ may vary by size, sex, year, or other characteristic.

Growth

$$\widetilde{N}_{y,t,.}^{g,m,o,z} = G_{y,t,.}^{g,m,o,z} P_{y,t,.}^{g,m,o,z} N_{y,t,.}^{g,m,o,z} + \left(1 - P_{y,t,.}^{g,m,o,z}\right) N_{y,t,.}^{g,m,o,z} \qquad \text{(Eq. 4)}$$

Where *P* is a square matrix were the diagonal terms are equal to the probability of molting (growing) and the off diagonal terms are equal to zero. *G* is a square matrix where the terms are equal to the growth transition among size classes. P and G may differ by gender, year, or other characteristic. For example, P might be equal to zero for all sizes for mature individuals (i.e. terminal molt to maturity).

Maturity
An individual is assumed to only have the possibility of maturing if it has molted/grown and is therefore modeled at the same time as growth if a maturity partition is included.

$$\begin{aligned}\widetilde{N}_{y,t,.}^{g,mat,o,z} = \Omega_{y,t,.}^{g,o,z} G_{y,t,.}^{g,imm,o,z} P_{y,t,.}^{g,imm,o,z} N_{y,t,.}^{g,imm,o,z} \\ + G_{y,t,.}^{g,mat,o,z} P_{y,t,.}^{g,mat,o,z} N_{y,t,.}^{g,mat,o,z} \\ + \left(1 - P_{y,t,.}^{g,mat,o,z}\right) N_{y,t,.}^{g,mat,o,z}\end{aligned} \qquad \text{(Eq. 5)}$$

$$\begin{aligned}\widetilde{N}^{imm} = \left(1 - \Omega_{y,t,.}^{g,o,z}\right) G_{y,t,.}^{g,imm,o,z} P_{y,t,.}^{g,imm,o,z} N_{y,t,.}^{g,imm,o,z} \\ + (1 - P) N^{imm}\end{aligned} \qquad \text{(Eq. 6)}$$

Where Ω is a square matrix where the diagonal terms are equal to the probability of maturing at size and the off diagonal terms are zero. *imm* = immature and *mat* = mature.

Recruitment

$$\widetilde{N}_{y,t,.}^{g,imm,new,z} = N_{y,t,.}^{g,imm,new,z} + R_{y,t,.}^{g} \qquad \text{(Eq. 7)}$$

Where R is a vector with size equal to the number of size bins and *new* = new shell

Time transition

$$N_{y,t+1,.}^{g,m,o,z} = N_{y,t,.}^{g,m,o,z} \quad if\ t < T \qquad \text{(Eq. 8)}$$

$$N_{y+1,1,.}^{g,m,o,z} = N_{y,T,.}^{g,m,o,z} \qquad \text{(Eq. 9)}$$

Where T is the number of seasons.

F method

The fishing mortality is implemented using two approaches, where the first approach has two options. The choice between the two approaches should be based on the characteristics of the fisheries and the computational demands of the application. Method 1a is appropriate when the fisheries operate over the whole time period, while Method 1b is appropriate if the fishery operates in a short period during the time period. Method 2 is useful when the application is large and the computational demands are outside the computational capacity available or a large number of model runs are required. Note that seasons can be used to further refine the temporal distribution of the fisheries. Also note that selectivity curves are scaled to have a maximum of one, which improves interpretation of parameters and may reduce confounding among some parameters.

1) Treat annual fishing mortality (F) as parameters with an optional lognormal penalty and fitting the model to catch data. This method can be used to allow for uncertainty in

the catch data. The timing of fishing within the season is implemented when calculating the predicted catch.

a) Continuous with *M*:

$$\Phi_{y,t,l} = exp\left(-M_{y,t,l} - F_{y,t,l}\right) \quad \text{(Eq. 10)}$$

$$F_{y,t,l} = \sum_{f} F_{y,t}^{f}\left(S_{y,t,l}^{f} V_{y,t,l}^{f} + S_{y,t,l}^{f}\left(1 - V_{y,t,l}^{f}\right) d_{y,t,l}^{f}\right) \quad \text{(Eq. 11)}$$

b) Occurring during a short period. The first equation treats *F* as an exploitation rate. The two equations differ in how selectivity interacts with natural mortality. The two equations also differ in how catch is calculated.

i)

$$\Phi_{y,t,l} = exp\left(-M_{y,t,l}\right)\left(1 - F_{y,t,l}\right) \quad \text{(Eq. 12)}$$

Or

ii)

$$\Phi_{y,t,l} = exp\left(-M_{y,t,l} - F_{y,t,l}\right) \quad \text{(Eq. 13)}$$

$$F_{y,t,l} = \sum_{f} F_{y,t}^{f}\left(S_{y,t,l}^{f} V_{y,t,l}^{f} + S_{y,t,l}^{f}\left(1 - V_{y,t,l}^{f}\right) d_{y,t,l}^{f}\right) \quad \text{(Eq. 14)}$$

Where the constraint $F_{y,t,l} \leq 1$ might be needed for the first equation

The fishing mortality could be modeled as a function of covariates (e.g. effort), which would be beneficial in the case where catch data is not available for all time priors (e.g. for bycatch fisheries). In this case

$$F_{y,t}^{f} = q^{f}\, E_{y,t}^{f} exp\left(\varepsilon_{y,t}^{f}\right) \quad \text{(Eq. 15)}$$

Where the temporal deviates are optional and the lognormal distributional penalty is applied.

2) Assume that catch is known and occurs during a short period (F is treated as an exploitation rate). The timing of the fishery is implemented through the fraction of natural mortality occurring before the fishery (δ) when calculating the exploitation rate from the catch and biomass, which assumes all fisheries occur at the same time.

$$\Phi_{y,t,l} = exp\left(-M_{y,t,l}\right)\left(1 - F_{y,t,l}\right) \quad \text{(Eq. 16)}$$

$$F_{y,t,l} = \sum_{f} F_{y,t}^{f}\left(S_{y,t,l}^{f} V_{y,t,l}^{f} + S_{y,t,l}^{f}\left(1 - V_{y,t,l}^{f}\right) d_{y,t,l}^{f}\right) \quad \text{(Eq. 17)}$$

Where the constraint $F_{y,t,l} \leq 1$ might be needed

retained catch in weight

$$F_{y,t}^{f} = \frac{C_{y,t}^{f}}{\sum_{l} N_{y,t,l} S_{y,t,l}^{f} V_{y,t,l}^{f} W_{y,t,l} exp\left(-\delta M_{y,t,l}\right)} \qquad \text{(Eq. 18)}$$

retained catch in numbers

$$F_{y,t}^{f} = \frac{C_{y,t}^{f}}{\sum_{l} N_{y,t,l} S_{y,t,l}^{f} V_{y,t,l}^{f} \, exp\left(-\delta M_{y,t,l}\right)} \qquad \text{(Eq. 19)}$$

Catch may not be known for some fleets (e.g. bycatch fisheries) and a different option will be needed or catch assumed.

The hybrid method (uses Pope's approximation to find initials values for *F* and then solves the catch equation iteratively) used in Stock Synthesis (Methot and Wetzel in press.) could be used as an alternative if estimating the fishing mortalities as parameters becomes unstable or computationally inefficient and Pope's approximation is not a sufficient approximation. A user supplied fixed number of iterations should be applied to ensure the derivatives remain valid.

Parameter setup

How parameter estimation is set up is based on the approach used in Stock Synthesis to retain familiarity with users of Stock Synthesis (Table 3). The main difference is that multiple covariates can be used for each parameter rather than single covariates, as implemented in Stock Synthesis, because population processes are typically controlled by multiple factors. There are several components that allow the parameters to change over time (deviates, time blocks, and covariates), requiring that internal data structures for parameter values and their derived quantities will need to have a time component (both year and season). Parameter values can also be shared or different among categories with a partition or between partitions. A system similar to Stock Synthesis should be developed to allow the sharing of parameter values.

If the "covariate" entry is greater than one, then after all the standard parameter control vectors are entered, the corresponding number of covariate control vectors are entered (Table 4).

Covariates are implemented as multiplicative exponents $\theta = \theta_{base} exp(\sum_{i} \beta_i I_i)$. Time blocks are implemented as additive or multiplicative and on the natural scale or exponentiated depending on the parameter.

Table 3. Control setup for parameter estimation. A vector of numbers corresponding to the entries described below will be defined for each parameter and control aspects of the parameter estimation. The setup is very similar to that used for Stock Synthesis.

Control	Description
Lower bound	Lower bound on the parameter value
Upper bound	Upper bound on the parameter value
Initial value	Initial value
Phase	The phase in which the parameter is estimated
Prior Type	Normal, lognormal, zero means no prior
Prior par 1	Parameter of the prior distribution e.g. mean
Prior par 2	Parameter of the prior distribution e.g. sd
Covariate	The number of covariates used for this parameter. The pointers to covariates are provided in another control variable
Deviate type	Random, random walk, zero is none
Deviate sd	sd deviate distribution
Deviate minimum year	Year to start deviate
Deviate maximum year	Year to stop deviate
Time Block	The block definition

Table 4. Control setup for covariates. A vector of numbers corresponding to the entries described below will be defined for each covariate identified in the standard "parameter control vector" and control aspects of the parameter estimation.

Control	Description
Lower bound	Lower bound on the parameter value
Upper bound	Upper bound on the parameter value
Initial value	Initial value
Phase	The phase in which the parameter is estimated
Prior Type	Normal, lognormal, zero means no prior
Prior par 1	Parameter of the prior distribution e.g. mean
Prior par 2	Parameter of the prior distribution e.g. sd
Covariate time series	The number corresponding to the covariate time series entered in the data file

Parameterization of size specific processes (e.g. selectivity, retention, natural mortality, molting probability, movement, initial conditions, maturity, recruitment distribution)

These are a set of functions that are useful for representing processes that change with size. The same function might be useful for modeling several different processes. Both

functions and probability distributions are used interchangeably with appropriate scaling. For selectivity, the functions are rescaled so the maximum is one. These functions can be used to model the mean growth and standard deviation of growth, but modeling the growth distribution will require integration over the appropriate size range.

Logistic

$$\theta_L = \frac{1}{1 + exp\left(-ln(19)\frac{L - \theta_{50}}{\theta_{95} - \theta_{50}}\right)} \qquad \text{(Eq. 20)}$$

θ_{50} is the size at 50%
θ_{95} is the size at 95%

It might be useful to have θ_{95} to be an offset of θ_{50} so that the logistic curve can be constrained to be increasing or decreasing.

Inverse logistic
The same as the logistic with –ln(19) replaced with ln(19)

Broken stick/piecewise linear

This function will need input on the number of breaks and may need to be implemented with smooth joiners to ensure differentiability.

Size-specific parameters with smoothness penalties
A parameter is estimated for each size bin and penalties are applied to control how much the curve changes from one size bin to the next (Maunder and Harley 2011; see Table 5). The method may only be practical for applications with a small number of size bins. This method scales the selectivity to average one, it might be better to also make the maximum equal one.

Table 5. Equations defining the penalties applied to the non-parametric selectivity curves.

Description	Equation	
First difference	$\lambda^1 \sum_{l=1}^{l=N-1} \left[-\ln(s_l) + \ln(s_{l+1}) \right]^2$	(Eq. 21)
Second difference	$\lambda^2 \sum_{l=1}^{l=N-2} \left[\ln(s_l) - 2\ln(s_{l+1}) + \ln(s_{l+2}) \right]^2$	(Eq. 22)
Third difference	$\lambda^3 \sum_{l=1}^{l=N-3} \left[-\ln(s_l) + 3\ln(s_{l+1}) - 3\ln(s_{l+2}) + \ln(s_{l+3}) \right]^2$	(Eq. 23)
Scaling	$\lambda^{s=1} \left[ln\left(\frac{\sum_l s_l}{N} \right) \right]^2$	(Eq. 24)
Monotonic increasing	if $s_l > s_{l+1}$ $\lambda^{\text{mon}} \left[\ln(s_l) - \ln(s_{l+1}) \right]^2$	(Eq. 25)

where λ^1, λ^2, and λ^3 are the weighting factors for the first, second, and third difference, respectively, s_l is the selectivity for an individual of size class l, and N is the number of size classes in the model.

Restricted size range
This curve is equal to one for all sizes within the given size range. This parameterization has no estimated parameters.

Stock Synthesis double normal
This is the recommended Stock Synthesis selectivity curve (Methot and Wetzel in press). It is constructed by combining the peaks of two half normal curves with a plateau. It also allows control of the selectivity at the minimum and maximum size bins. By fixing some of the parameters the double normal can be used to implement an asymptotic selectivity curve similar to a logistic.

$$\theta_l = asc_l\left(1 - j_{1,l}\right) + j_{1,l}\left(l\left(1 - j_{2,l}\right) + j_{2,l}des_l\right) \qquad \text{(Eq. 26)}$$

Where the smooth joiner functions are defined as

$$j_{1,l} = \left(1 + \exp\left(-20\frac{L_l - \theta_1}{1 + \left|L_l - \theta_1\right|}\right)\right)^{-1} \qquad \text{(Eq. 27)}$$

$$j_{2,l} = \left(1 + \exp\left(-20\frac{L_l - \theta_{peak2}}{1 + \left|L_l - \theta_{peak2}\right|}\right)\right)^{-1} \qquad \text{(Eq. 28)}$$

$$asc_l = \left(1 + \exp\left(-\theta_5\right)\right)^{-1} + \left(1 - \left(1 + \exp\left(-\theta_5\right)\right)^{-1}\right)\frac{\exp\left(\frac{-\left(L_l - \theta_1\right)^2}{\exp\left(\theta_3\right)}\right) - \mathrm{t}_{\min 1}}{1 - \mathrm{t}_{\min 1}} \qquad \text{(Eq. 29)}$$

$$dsc_l = 1 + \left(\left(1 + \exp\left(-\theta_6\right)\right)^{-1} - 1\right)\frac{\exp\left(\frac{-\left(L_l - \theta_{peak2}\right)^2}{\exp\left(\theta_2\right)}\right) - 1}{\mathrm{t}_{\min 2} - 1} \qquad \text{(Eq. 30)}$$

$$t_{\min 1} = \exp\left(-\frac{\left(L_{\min} - \theta_1\right)^2}{\exp\left(\theta_3\right)}\right) \qquad \text{(Eq. 31)}$$

$$t_{\min 2} = \exp\left(-\frac{\left(L_{\max} - \theta_{peak2}\right)^2}{\exp\left(\theta_4\right)}\right) \qquad \text{(Eq. 32)}$$

$$\theta_{peak2} = \theta_1 + L_{width} + \frac{\left(0.99L_{\max} - \theta_1 - L_{width}\right)^2}{1 + \exp\left(-\theta_2\right)} \qquad \text{(Eq. 33)}$$

Where

L_l is the midpoint of size bin l

$L_{\min}$ is the midpoint of smallest size bin l

$L_{\max}$ is the midpoint of largest size bin l

θ_1 is the size at which $\theta = 1$ begins

θ_2 is the width for which $\theta = 1$ (θ_{peak2} is the size at which $\theta = 1$ ends)

θ_3 determines the slope of the ascending limb

θ_4 determines the slope of the descending limb
θ_5 is value of θ at L_{min}
θ_6 is value of θ at L_{max}
L_{width} is the width of the population size bins

Cubic splines
See Stock Synthesis implementation (Methot and Wetzel in press).

Gamma

$$f(L) = \frac{L^{\alpha-1} exp(-L/\beta)}{\beta^{\alpha}\Gamma(\alpha)} \qquad \text{(Eq. 34)}$$

Where

$$\mu = \alpha\beta \qquad \text{(Eq. 35)}$$

$$\sigma^2 = \alpha\beta^2 \qquad \text{(Eq. 36)}$$

To formulate the gamma in terms of the mean and variance use the following equations

$$\beta = \frac{\sigma^2}{\mu} \qquad \text{(Eq. 37)}$$

$$\alpha = \frac{\mu^2}{\sigma^2} \qquad \text{(Eq. 38)}$$

Lognormal

$$f(L) = \frac{1}{L\sigma\sqrt{2\pi}} exp\left(-\frac{\left(ln(L) - (ln(\mu) + 0.5\sigma^2)\right)^2}{2\sigma^2}\right) \qquad \text{(Eq. 39)}$$

Normal

$$f(L) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(L-\mu)^2}{2\sigma^2}\right) \qquad \text{(Eq. 40)}$$

Selectivity/catchability

Selectivity is modeled using one of the size-based functions described above (e.g. the Stock Synthesis double normal). An additional "availability" parameter is used so that the

maximum selectivity can differ among categories within partitions. (the selectivities should be scaled, to the extent possible, so the selectivity of the category with the highest selectivity is equal to one). In addition to sharing selectivity among categories within partitions, selectivities can be shared among fisheries (i.e. mirroring). The possibility of multiple selectivities for each fishery or survey should be modeled to allow for the availability to be shared among fisheries or surveys, or to allow the modeling of escape gaps and minimum legal size (MLS) that changes over time.

The use of multiple and shared selectivity curves can be used to implement the relationship between the BSFRF survey and the NMFS survey. The NMFS covers the whole stock, but the BSFRF survey only covers a component of the stock. The size of the individuals in the area covered by the BSFRF survey may differ from that of the whole stock; think of this as size specific availability. The selectivity of the BSFRF survey gear may differ from the selectivity of the NMFS survey gear. This can be modeled using two selectivity curves and one size based availability curve. The whole area NMFS survey can be modeled with the NMFS survey gear selectivity curve, the NMFS survey in the BSFRF survey area can be modeled as a combination of the NMFS survey gear selectivity curve and the availability curve. The BSFRF survey can be modeled using the BSFRF survey gear selectivity curve and the availability curve. This method uses some of the NMFS survey data twice and if this is considered a problem, then the NMFS survey data outside the BSFRF survey area could be modeled as the NMFS survey gear selectivity curve adjusted by some function of the availability curve.

One approach is to input the following control variables:

Number of selectivity curves
Number of fisheries/partition assignments
The fisheries, partitions, and partition categories that use each selectivity curve (a special value e.g. 0 can be used to specify that all fisheries, partitions, or categories within a partition use this selectivity curve) a negative number means all except that fishery/partition/category has that selectivity

```
#Nselect
4
#Nselect_assignments
3
#Select_assigments
#Selectivity Fishery Partition category
1 1 1 1          #categroy 1 of partition 1 has selectivity 1 for fishery 1
2 1 1 -1         #all categories except category 1 of partition 1 has selectivity 1 for fishery
                 1
3 2 0 0          #Fishery 2 has selectivity 2 for all partitions
4 0 0 0          #All fisheries and partitions have this curve, e.g.  restricted size range
```

The selectivity curves and other population processes can be time varying.

Discards

The discards are modeled using a retention curve that is applied to the selectivity curve, where the discard proportion is the difference between the selectivity and retained curves. The retention curve is modeled using one of the size-based functions described above. As with the selectivity curves, the retention curves can be the same or differ among categories within partitions. An additional parameter is used to model the maximum retention. The discard mortality rate is modeled using one of the size-based functions described above with an additional parameter for the maximum discard mortality rate. The selectivity curves and the maximum discard rate can be time varying.

Growth

Growth can be modeled either using growth increments (the increment is added to the original size to define the growth transition matrix) or growth transition, both of which are modeled in terms of probability distributions or equivalently relative frequency distributions. Either way the model converts these into a growth transition matrix. Several options are available for the mean (μ), standard deviation (σ), and probability distribution ($f(L)$). A separate option to pre-specify (input) the complete growth transition matrix should be implemented. The growth parameters can be shared or differ among partition categories. Reliable methods for banding transition matrices (to avoid issues related to computer underflow) should be considered (Jim Ianelli pers. com.).

for growth increment

$$G_{L\to\acute{L}} = \int_{\acute{L}-L-0.5\Delta L}^{L\acute{-}L+0.5\Delta L} f^{inc}\left(\mu^{inc}, \sigma^{inc}, L, \acute{L}-L\right) d\left(\acute{L}-L\right) \qquad \text{(Eq. 41)}$$

for growth transition

$$G_{L\to\acute{L}} = \int_{\acute{L}-0.5\Delta L}^{\acute{L}+0.5\Delta L} f^{trans}\left(\mu^{trans}, \sigma^{trans}, L, \acute{L}\right) d\acute{L} \qquad \text{(Eq. 42)}$$

Note that in general $\mu^{inc} = \mu^{trans} - L$ and $\sigma^{inc} \neq \sigma^{trans}$

The same functions for the mean, standard deviation, and distribution can generally be used for both the growth increment and the transition formulations.

Mean size

linear function of size

$$\mu = a + bL \qquad \text{(Eq. 43)}$$

Von Bertalanffy
This can be modeled using the Francis model with b=1 (see below).

Richards/Schnute/Francis

We use Francis (1995) general model, which is based on Baker et al.'s (1991) size-based analog of Schnute's (1981) size-at-age model. The mean size after growth is defined by

$$\mu^{trans} = \left[L^b e^{-a\Delta t} + c\left(1 - e^{-a\Delta t}\right)\right]^{\frac{1}{b}} \qquad a \neq 0, b \neq 0 \qquad \text{(Eq. 44)}$$

Where L is the initial size, $\Delta t = 1$ (for the population dynamics model, but might differ from 1 for interpreting growth increment data from tags if this data is used in the likelihood function), and a, c, λ_1 and λ_2 are variables to make the calculations more convenient.

$$a = \ln\left[\frac{y_2^b - y_1^b}{\lambda_2^b - \lambda_1^b}\right] \qquad \text{(Eq. 45)}$$

$$c = \frac{y_2^b \lambda_1^b - y_1^b \lambda_2^b}{\lambda_1^b - y_1^b + y_2^b - \lambda_2^b} \qquad \text{(Eq. 46)}$$

$$\lambda_1 = y_1 + g_1 \qquad \text{(Eq. 47)}$$

$$\lambda_2 = y_2 + g_2 \qquad \text{(Eq. 48)}$$

The model includes five parameters; y_1 and y_2 are arbitrary fish sizes, small and large respectively, g_1 and g_2 are the mean growth increments for fish of sizes y_1 and y_2, respectively, over a given time period (usually year); and b, which has no biological meaning but describes the curvature in the model. The general growth model reduces to the von Bertalanffy growth model when b is set equal to one.

Polynomial

$$\mu = a + \sum_i b_i L^i \qquad \text{(Eq. 49)}$$

This formulation might need to be modified or constrained to ensure that $\mu \geq 0$ or $\mu_{l+1} \geq u_l$.

Variance of growth distribution

Constant

Can be modeled using the linear function with $b = 0$

Linear function of size

$$\sigma^2 = a + bL \qquad \text{(Eq. 50)}$$

Can also be modeled using the polynomial with $i = 1$.

Polynomial

$$\sigma^2 = a + \sum_i b_i L^i \qquad \text{(Eq. 51)}$$

The polynomial might need a constraint to keep it positive or alternatively the exponentiated polynomial used.

$$\sigma^2 = exp\left(a + \sum_i b_i L^i\right) \qquad \text{(Eq. 52)}$$

Distribution

Gamma

$$f(L) = \frac{L^{\alpha-1} exp(-L/\beta)}{\beta^\alpha \Gamma(\alpha)} \qquad \text{(Eq. 53)}$$

Where

$$\mu = \alpha\beta$$

$$\sigma^2 = \alpha\beta^2$$

To formulate the gamma in terms of the mean and variance use the following equations

$$\beta = \frac{\sigma^2}{\mu}$$

$$\alpha = \frac{\mu^2}{\sigma^2}$$

Lognormal

$$f(L) = \frac{1}{L\sigma\sqrt{2\pi}} exp\left(-\frac{\left(ln(L) - (ln(\mu) + 0.5\sigma^2)\right)^2}{2\sigma^2}\right) \qquad \text{(Eq. 54)}$$

Normal

$$f(L) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(L-\mu)^2}{2\sigma^2}\right) \qquad \text{(Eq. 55)}$$

Recruitment

Recruitment is modeled as mean recruitment for each area with annual (or time period if more than one recruitment event occurs during a year) penalized deviates. When there is spatial partition, an overall annual deviate is computed that is shared by all areas and then area specific annual deviates. This allows some correlation of recruitment among areas. Multiple covariates (I) are also available for either the variation common to all areas or the area specific variation. Inclusion of a stock-recruitment curve requires a definition of age, specifically the age-at-recruitment so that the recruitment can be linked to the spawning biomass. The stock-recruitment curve might be kept separate from the modeling of recruitment so that it can be applied to a flexible range of years. In area-specific models the spawning biomass can be pooled across areas. Therefore, for a specific area, the proportion of the spawning biomass from each area used to create the spawning biomass of the area of interest is input (this can be equal to zero for all areas other than the area of interest to implement area-specific stock-recruitment relationships). In the initial implementation the stock-recruitment relationship is ignored. Recruitment is only to immature new-shell partitions and either the sex ratio fixed or estimated, or the recruitment strength and size composition estimated separately for males and females, with an optional penalty on the sex ratio.

$$R^z_{y,t} = R^z_\mu exp\left(\varepsilon_{y,t}\sigma - \Delta_{y,t}0.5\sigma^2 + \varepsilon^z_{y,t}\sigma^z - \Delta^z_{y,t}0.5 + \sum_i b_i I_{i,y,t} + \sum_j b_j I^z_{j,y,t}\right) \qquad \text{(Eq. 56)}$$

Bias correction of the recruitment deviates (Δ) is implemented with tapering following Stock Synthesis (Methot and Taylor 2011). The equation below is defined for y, but could also differ by t and z.

$$\Delta_y = \begin{cases} 0 & y \le \Delta_1 \\ \Delta_{max}\left(1 - \frac{y - \Delta_1}{\Delta_2 - \Delta_1}\right) & \Delta_1 < y < \Delta_2 \\ \Delta_{max} & \Delta_2 \le y \le \Delta_3 \\ \Delta_{max}\left(1 - \frac{\Delta_3 - y}{\Delta_4 - \Delta_3}\right) & \Delta_3 < y < \Delta_4 \\ 0 & \Delta_4 \le y \end{cases} \qquad \text{(Eq. 57)}$$

$\square_1$ the first year of the bias ramp up adjustment period
$\square_2$ the last year of the bias ramp up adjustment period
$\square_3$ the first year of the bias ramp down adjustment period
$\square_4$ the last year of the bias ramp down adjustment period
$\square_{max}$ the maximum bias adjustment

A choice of distributions is available to distribute recruitment into the size classes. An option for estimating size specific parameters within a given size range should also be implemented.

$$R^z_{y,t,l} = R^z_{y,t} f(L) \qquad \text{(Eq. 58)}$$

Gamma

$$f(L) = \frac{L^{\alpha-1} exp(-L/\beta)}{\beta^\alpha \Gamma(\alpha)} \qquad \text{(Eq. 59)}$$

Where

$$\mu = \alpha\beta$$

$$\sigma^2 = \alpha\beta^2$$

To formulate the gamma in terms of the mean and variance use the following equations

$$\beta = \frac{\sigma^2}{\mu}$$

$$\alpha = \frac{\mu^2}{\sigma^2}$$

Lognormal

$$f(L) = \frac{1}{L\sigma\sqrt{2\pi}} exp\left(-\frac{\left(ln(L) - (ln(\mu) + 0.5\sigma^2)\right)^2}{2\sigma^2}\right) \qquad \text{(Eq. 60)}$$

Truncated or accumulated at the minimum size bin and at the upper size bin.

Normal

$$f(L) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(L-\mu)^2}{2\sigma^2}\right) \qquad \text{(Eq. 61)}$$

Truncated or accumulated at the minimum size bin and at the upper size bin.

Natural mortality

Several options can be used for natural mortality. Any of the size-based functions listed above can be used with the addition of a parameter to scale the absolute level of natural mortality. In addition, a special natural mortality function that is a size-based version of the Maunder–Lorenzen model (see Brodziak et al. 2011) that assumes M is inversely proportional to size up until the size at maturity and follows the logistic curve as individuals mature. The rate of natural mortality may be shared among partitions for juveniles. It also might be desirable to estimate the size at maturity (or first maturity) independent of the real size at maturity (in this case the size at maturity is simply where the natural mortality changes and not related to maturity). Maturity can be sex specific. Two versions of the model are implemented:

$$M_L = M_{juv}\left(\frac{L_{mat}}{L}\right)^{\lambda} + \frac{M_{mat} - M_{juv}}{1+\exp\left[\beta\left(L - L_{50}\right)\right]} \qquad \text{(Eq. 62)}$$

$$M_L = \begin{cases} M_{juv}\left(\frac{L_{mat}}{L}\right)^{\lambda} & L < L_{mat} \\ M_{juv} + \frac{M_{mat} - M_{juv}}{1+\exp\left[\beta\left(L - L_{50}\right)\right]} & L \geq L_{mat} \end{cases} \qquad \text{(Eq. 63)}$$

Where M_{juv}, and M_{mat} are the main parameters to estimate. λ controls the shape of the relationship between size and natural mortality and can be fixed at a standard value from Lorenzen (1996) or estimated. L_{50} is the size at 50% maturity (or alternatively a parameter defining the change in M relationship) and can be fixed or estimated. L_{mat} is the size at first maturity (or alternatively when the relationship deviates from M_{juv}).

Molting

The probability of molting (growing) can be modeled using any of the size-based functions described above and can be shared or differ among partition categories. Shell condition and maturity state are likely to influence the molting probability. A terminal molt will need to be an option such that any individual that reaches this stage does not molt. This may be linked to the shell condition or maturity state.

Movement

Movement parameters can be modeled using any of the size-based functions. In some cases an additional parameter will be used to model the magnitude of the movement.

Movement could be modeled in logit space to ensure that it is between zero and one. Movement in one of the directions (e.g. remaining in the current area) will be calculated analytically rather than as a parameter so that movement sums to one. A penalty may need to be applied to ensure that the movements sum to one. Movement may need to be implemented so it changes over time and among partition categories.

$$\psi_{z \to \acute{z}} = f(L) \qquad \text{(Eq. 64)}$$

$$\sum_{\acute{z}} \psi_{z \to \acute{z}} = 1 \qquad \text{(Eq. 65)}$$

$$\psi_{z \to \acute{z}_N} = 1 - \sum_{z=1 \; to \; N-1} \psi_{z \to \acute{z}} \qquad \text{(Eq. 66)}$$

Where N is number of areas.

Initial conditions (Initial numbers by partition categories)

Several methods are implemented to estimate the initial conditions.

1) Any of the size-based functions can be used to model the numbers-at-size with the addition of a scaling parameter. The initial conditions can be shared (equal size distribution and/or numbers) or different among partition categories. Alternatively, the size component can be the same and the scaling component different.
2) Equilibrium assumptions with parameters such as fishing mortality, average recruitment, and penalized lognormal recruitment deviates (see recruitment section for similar details) following the Stock Synthesis approach. i.e. the average recruitment is modeled over age using growth and survival. A recruitment deviate is added to each age, but not propagated to older ages. The average recruitment used could be the average used in the dynamics or a separate (or offset) parameter. Fishing mortality could be estimated for certain fisheries, or depending on the F method, equal to F in the first few time periods (re: MULTIFAN-CL, Fournier et al. 1998).
3) Similar to (2) except the average F is the same as over the first few years of modeling time frame. This is similar to the MULTIFAN-CL approach (Fournier et al. 1998) and does not work for Pope's approximation method.
4) Recruitment for years [Startyear – x] to [startYear – 1] estimated. Fishing mortalities for each of the fisheries can also be estimated. Recruitment before [Startyear – x] is set to zero.

Maturity

Maturity can either be modeled as the probability of maturing if a maturity partition is specifically modeled (where a transition matrix is needed to model transition from immature to mature) or as the proportion mature-at-size (where the maturity at size is applied to the number at size to determine the number mature at size). Both of these can

be modeled using the size-specific functions presented above. A parameter restricted between zero and one can be used to scale the maximum probability of maturity.

Data

The setup for fitting to data needs to be flexible so that data can either be separate for each partition category or grouped across partition categories.

Catch

The model needs to fit to catch data when using the Fishing mortality approach that treats each fishing mortality as a parameter. The catch for each fishery is fit with either a normal or lognormal distribution and either in biomass or numbers. The catch can be fit as total, retained, and/or discarded. The catch estimates described below are for retained catch in weight, appropriate modifications can be made for catch-in-numbers or total or discarded catch.

a) Continuous with M

$$C_{y,t}^{f} = \sum_{l} \frac{\acute{F}_{y,t,l}^{f}}{Z_{y,t,l}} N \left(1 - exp\left(-Z_{y,t,l}\right)\right) w_l \qquad \text{(Eq. 67)}$$

$$Z_{y,t,l} = M_{y,t,l} + \sum_{f} F_{y,t,l}^{f} \qquad \text{(Eq. 68)}$$

$$\acute{F}_{y,t,l}^{f} = F_{y,t}^{f} S_{y,t,l}^{f} V_{y,t,l}^{f} \qquad \text{(Eq. 69)}$$

b) Occurring during a short period (F is exploitation rate in this formulation)

$$C_{y,t}^{f} = \sum_{l} N_l exp(-\Delta M) \left(1 - exp\left(-\acute{F}_{y,t,l}\right)\right) w_l \qquad \text{(Eq. 70)}$$

$$\acute{F}_{y,t}^{f} = F_{y,t}^{f} S_{y,t,l}^{f} V_{y,t,l}^{f} \qquad \text{(Eq. 71)}$$

Negative log-likelihoods

$$-\ln L\left(\boldsymbol{\theta} \mid C\right) = \sum_{y,t} \left\{ 0.5 \ln\left[\sigma\right] + \frac{\left(C_{y,t} - \hat{C}_{y,t}\right)^2}{2\sigma^2} \right\} \qquad \text{(Eq. 72)}$$

$$-\ln L(\boldsymbol{\theta} \mid C) = \sum_{y,t} \left\{ 0.5 \ln[\sigma] + \frac{\left(\ln[C_t] - \ln\left[\hat{C}_t + 0.5\sigma^2\right]\right)^2}{2\sigma^2} \right\} \qquad \text{(Eq. 73)}$$

Composition

Several options are available for fitting to composition data. There should be flexibility in the definition of the size bins for the observations and the possibility of multiple definitions within an application. This allows for the inclusion of standard size composition data from scientific sampling (e.g. 1mm) as well as commercial size category data (e.g. small, medium, and large), which may be irregular bin sizes. There may be a need to model alternative size transition functions to allow for different measurements (e.g. a transition matrix would be needed to covert from the population size categories to weight categories to fit to weight composition data). The composition data can be treated independently for each partition category or sum to one over the partition categories and/or over partitions (e.g. to provide sex ratio information the size composition data proportions should sum to one over both males and females).

Normal with binomial variance (based on Fournier et al. 1998)

$$-\ln L(\boldsymbol{\theta} \mid data) = \sum_{t,l} \left\{ 0.5 \ln\left[p_{t,l}\left(1 - p_{t,l}\right)\tau^2 \right] + \frac{\left(p_{t,l} - \hat{p}_{t,l}\right)^2}{2 p_{t,l}\left(1 - p_{t,l}\right)\tau^2} \right\} \qquad \text{(Eq. 74)}$$

Where τ is a parameter that scales the variance.

Where p and $\hat{p}$ are the observed and predicted proportions at size and will sum to one across some specified set of partitions and partition categories. τ scales the variance and is related to the effective sample size. The option to estimate τ should be available.

The variance could be calculated using the observed or predicted proportions. A constant can be added to the variance to insure it does not get too small or alternatively some sizes could be pooled.

Multinomial

$$-\ln L(\boldsymbol{\theta} \mid data) = \sum_{t,l} SS_t p_{t,l} \ln\left[\hat{p}_{t,l}\right] \qquad \text{(Eq. 75)}$$

Where SS is the sample size.

This method is problematic if the predicted proportion is zero, a small additive constant may be added if appropriate or alternatively some sizes could be pooled. The multinomial offset might be needed to avoid over/underflow errors.

Lognormal with variance inversely proportional to the proportion
The size frequency data is fit using the method of Punt and Kennedy (1997) that uses a lognormal likelihood function for the proportions at size and assumes that the variance is proportional to the inverse of the predicted proportion.

$$-\ln L\left(\boldsymbol{\theta} \mid data\right) = \sum_{t,l}\left\{0.5\ln\left[\frac{\tau^2}{\hat{p}_{t,l}}\right] + \frac{\hat{p}_{t,l}\left(\ln\left[p_{t,l}\right] - \ln\left[\hat{p}_{t,l}\right]\right)^2}{2\tau^2}\right\} \quad \text{(Eq. 76)}$$

The variance parameter can be estimated analytically:

$$\hat{\tau} = \sqrt{\frac{1}{T*L}\sum_{t,l}\hat{p}_{t,l}\left(\ln\left[p_{t,l}\right] - \ln\left[\hat{p}_{t,l}\right]\right)^2} \quad \text{(Eq. 77)}$$

This method is problematic if the observed data is zero, a small additive constant may be added if appropriate or alternatively some sizess could be pooled. The log normal bias correction factor should also be considered.

Multivariate logistic
Other composition likelihoods should also be considered. Schnute and Haigh (2007) use the logistic-normal (or multivariate logistic; Schnute and Richards, 1995) distribution, which they show can be converted into a multivariate normal with appropriate transformations.

Indices of abundance

Indices of abundance (e.g. surveys or CPUE) are fit using a lognormal distribution (other likelihood functions such as the normal and t-distribution should also be considered). Flexibility should be implemented so that the abundance indices can be separated or combined by size groups, partition categories or partitions. There should be an option for multiple catchabilities so that availability or the proportion of the area covered can be modeled and shared among surveys. For example, if the area of a survey changes over time, there could be a catchability for the survey in general, and then a catchability for each time period based on the area covered in that time period. This could be implemented using a method similar to that described above for catchability/selectivity.

$$-\ln L(\boldsymbol{\theta} \mid I) = \sum_{t,l} \left\{ 0.5 \ln[\sigma_t] + \frac{\left(\ln[I_t] - \ln\left[\hat{I}_t + 0.5\sigma_t^2\right]\right)^2}{2\sigma_t^2} \right\} \qquad \text{(Eq. 78)}$$

$$\hat{I}_t = qB_t^{\eta} \qquad \text{(Eq. 79)}$$

Where η is the nonlinear relationship parameter.

The catchability coefficient and standard deviation can be calculated analytically (The following equations assume that the variance is constant over time and need to modify for year specific variance scaling or additive values and the associated lognormal bias correction factor. Equations are also available for Bayesian analysis)

$$q = \exp\left(\frac{\sum_t \ln\left(\frac{I_t}{B_t}\right)}{n} \right) \qquad \text{(Eq. 80)}$$

$$\sigma = \sqrt{\frac{\left(\ln(I_t) - \ln\left(\hat{I}_t\right)\right)^2}{n}} \qquad \text{(Eq. 81)}$$

Discards

The discard data can be modeled as explained in the section on catch data.

Tagging

The following is only an example of how tagging data might be integrated into the model. The tag data could include tag growth increment, tag-recapture, and tag-recapture with growth increment. The catch conditioned movement model of McGarvey and Feenstra (2002) might also be useful.

We use Francis (1995) general method to estimate growth with mark-recapture data, which is based on Baker et al.'s (1991) size-based analog of Schnute's (1981) size-at-age model. Francis' model combines a general size-based growth function that predicts the mean growth increment from the size at release and time at liberty, with an error structure that includes both measurement error and individual variability in growth. This model was used by Maunder (2001) to estimate growth for skipjack tuna in the EPO. The observed growth increment is assumed to be normally distributed with the variance of the growth increment a function of both time at liberty and size at release using Maunder's (2001) generalized growth variation model. To simplify the calculations, measurement error is assumed to either be only in size at recapture or that this assumption is an adequate approximation to measurement error in both the size at release and size at

recapture. A robust version of the normal distribution based likelihood is used to reduce the influence of outliers that may be due to incorrect recording of size or date

$$\ell\left(\theta \mid L, L', \Delta t\right) = \frac{1}{\sqrt{2\pi\left(\sigma_m^2 + \sigma_g^2\right)}} \exp\left(-\frac{\left(G_{obs} - G - \mu_m\right)^2}{2\left(\sigma_m^2 + \sigma_g^2\right)}\right) + 0.01 \quad \text{(Eq. 82)}$$

Where

μ_m and σ_m^2 are the mean and variance of the measurement error, respectively.

Mortality

Total mortality or fishing mortality can be fit using the normal or lognormal likelihood functions. For example, the lognormal likelihood for age and time specific total mortality would be:

$$-\ln L\left(\boldsymbol{\theta} \mid Z^{obs}\right) = \sum_{t,a}\left\{0.5\ln\left[\sigma_{t,a}\right] + \frac{\left(\ln\left[Z_{t,a}^{obs}\right] - \ln\left[Z_{t,a} + 0.5\sigma_{t,a}^2\right]\right)^2}{2\sigma_{t,a}^2}\right\} \quad \text{(Eq. 83)}$$

Weighting and priors

Weighting factors should be included for all likelihood components. These should follow the Stock Synthesis approach where the lambdas are all set to one unless explicitly modified. One addition is to allow a change to influence all fisheries and/or surveys (see the approach for assigning selectivity parameters above).

Priors should be implemented on all model parameters similar to the approach used in Stock Synthesis. The priors should include the Normal, Lognormal, and beta distributions. The beta distribution can be used for parameters that are between zero and one or other bounds such as steepness of the stock-recruitment relationship, which is between 0.2 and one.

Normal

$$-\ln\left[p\left(\theta \mid \mu_\theta, \sigma_\theta\right)\right] = 0.5\left(\frac{\theta - \mu_\theta}{\sigma_\theta}\right)^2 \quad \text{(Eq. 84)}$$

Lognormal

$$-\ln\left[p\left(\theta \mid \mu_\theta, \sigma_\theta\right)\right] = 0.5\left(\frac{\ln[\theta] - \ln[\mu_\theta]}{\sigma_\theta}\right)^2 \qquad \text{(Eq. 85)}$$

Bounded Beta (from Methot and Wetzel in press)

$$-\ln\left[p\left(\theta \mid \mu_\theta, \sigma_\theta\right)\right] = \left(\ln\left(1-\theta'\right) - \ln\left(1-\mu'_\theta\right)\right)\left(\theta_A - 1\right) + \left(\ln\left(\theta'\right) - \ln\left(\mu'_\theta\right)\right)\left(\theta_B - 1\right) \qquad \text{(Eq. 86)}$$

Where

$$\theta' = \frac{\theta - \theta_{\min}}{\theta_{\max} - \theta_{\min}} \qquad \text{(Eq. 87)}$$

$$\mu'_\theta = \frac{\mu_\theta - \theta_{\min}}{\theta_{\max} - \theta_{\min}} \qquad \text{(Eq. 88)}$$

$$\theta_B = \frac{\left(1-\mu_\theta\right)\mu'_\theta}{\sigma_\theta^2} - \mu_\theta \qquad \text{(Eq. 89)}$$

$$\theta_A = \frac{\theta_B}{\mu_\theta} - \theta_B \qquad \text{(Eq. 90)}$$

The standard deviation additive and multiplicative factors should be estimable for all likelihood components. Algebraic (maximum likelihood) solutions should be implemented where possible. Although, there should also be the ability to estimate them as free parameters, particularly if Bayesian MCMC is being applied. The analytic solution for Bayesian methods should also be considered.

Acknowledgements

We are grateful to reviewers from federal and state agencies, industry, and academia for their constructive critique. This report was funded by the Bering Sea Fisheries Research Foundation.

Literature cited

Baker, T.T., Lafferty, R. and Quinn, T.J., II. 1991. A general growth model for mark recapture data. Fisheries Research, 11 (3,4): 257-281.

Brodziak, J., Ianelli, J., Lorenzen, K., Methot Jr., R.D., 2011 Estimating natural mortality in stock assessment applications. NOAA Technical Memorandum NMFS-F/SPO 119 38p.

Fournier, D.A., Hampton, J., Sibert, J.R., 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, Thunnus alalunga. Can. J. Fish. Aquat. Sci. 55, 2105–2116.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J. (2012) AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software, 27: 233-249.

Francis, R.I.C.C. 1995. An alternative mark-recapture analogue of Schnute's growth model. Fisheries Research, 23 (1,2): 95-111.

Lorenzen, K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49:627–647.

Maunder, M.N. (2001) Growth of skipjack tuna (Katsuwonus pelamis) in the eastern Pacific Ocean, as estimated from tagging data. Inter-American Tropical Tuna Commission Bulletin, 22(2): 93-131. (in English and Spanish)

Maunder, M.N. and Harley, S.J. (2011) Using cross validation model selection to determine the shape of nonparametric selectivity curves in fisheries stock assessment models. Fisheries Research 110: 283-288.

McGarvey, R. and Feenstra, J.E. 2002. Estimating rates of fish movement from tag recoveries: conditioning by recapture. Canadian Journal of Fisheries and Aquatic Sciences 59(6): 1054-1064.

Methot Jr., R.D., 2009. Stock assessment: operational models in support of fisheries management. In: Beamish, R.J., Rothschild, B.J. (Eds.), The Future of Fisheries Science in North America, Fish Fish. Ser, pp. 137–165.

Methot, R.D. Jr., I.G. Taylor. 2011. Adjusting for Bias due to Variability of Estimated Recruitments in Fishery Assessment Models. Can. J. Fish. Aquat. Sci. 68: 1744-1760.

Methot, R.D., Wetzell, C.R., (in press). Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 00, 00–00.

Punt, A.E., Kennedy, R.B., 1997. Population modelling of Tasmanian rock lobster, Jasus edwardsii, resources. Mar. Freshw. Res. 48, 967–980.

Schnute, J. 1981. A versatile growth model with statistically stable parameters. Canadian Journal of Fisheries and Aquatic Sciences, 38 (9): 1128-1140.

Schnute, J.T., Haigh, R., 2007. Compositional analysis of catch curve data, with an application to *Sebastes maliger*. ICES J. Mar. Sci. 64, 218–233.

Schnute, J.T., Richards, L.J., 1995. The influence of error on population estimates from catch-age models. Can. J. Fish. Aquat. Sci. 52, 2063–2077.