

Software Design Document

For a Metapopulation Assessment System

Jon Brodziak¹, Matthew Supernaw², Teresa A'mar³ and Clay Porch²

¹Pacific Islands Fisheries Science Center
NOAA Inouye Regional Center, 1845 Wasp Boulevard, Building 176, Honolulu, HI 96818

²Southeast Fisheries Science Center
75 Virginia Beach Drive, Miami, FL 33149-1003

³Northwest Fisheries Science Center
7600 Sand Point Way NE, Seattle, WA 98115-6349

Jon.Brodziak@NOAA.GOV

5 February 2016

1. INTRODUCTION

Spatial processes are one of the keys to understanding and predicting patterns in resource abundance for population and community ecology. The importance of spatial distribution, scale, and connectivity is a general paradigm for ecology (Levin 1992). This provides a strong rationale to focus not on single homogeneous populations, but instead on metapopulations, or populations composed of subpopulations with source-sink dynamics, as the basic unit of population ecology. Ignoring spatial structure and movement and assessing a metapopulation as several independent populations can lead to approximation errors and increase the risk of overexploiting the resource (e.g., Ying et al. 2011, Spies et al. 2015). To understand and predict human impacts on a metapopulation, a stock assessment model needs to be able to characterize the effects of spatial variation on the fishery system. The development and application of spatially-explicit metapopulation models will require a shift towards hierarchically-structured spatial models with greater complexity than standard single-area assessment models.

From the perspective of software design, an object-oriented paradigm provides a lot of essential features and extensibility options (Jana 2005, Stroustrup 2013) for developing metapopulation

models. These features include precise design specifications, shorter development times for new applications, and easier-to-maintain code. Existing code can be maintained based on the definitions of consistent and extended with reusable classes, associated object interfaces, the capacity to add new features to an existing class without rewriting its interface. An object-oriented design provides a natural way to implement complex hierarchical population models in a structured manner. Some object-oriented programming features are implemented in existing assessment software packages, e.g. Stock Synthesis 3 (Methot and Wetzel 2013) and CASAL (Bull et al. 2005). Over the long term, it is important to maintain an ongoing process of total quality improvement for assessment model development. This includes the capacity for model selection, testing, and uncertainty quantification (Morgan and Henrion 1990, Burnham and Anderson 2002, NRC 2012). In this context, the capacity for rapid model prototyping, development, and testing for fishery systems with diverse spatial properties is an important goal.

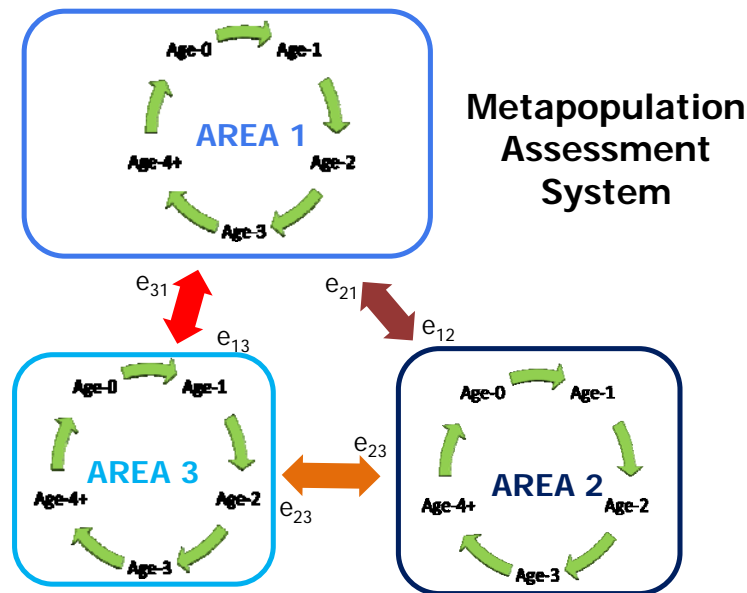
We envision that the functionality for rapid prototyping, development, and testing of models can be successfully achieved through the ongoing development of a metapopulation assessment system consisting of a library of tested modules with extensible templates that implement various types of subpopulation dynamics. The capacity to build new models from existing templates with just-in-time compilation will encourage the ongoing development of structured assessment models. This, in turn, will help to avoid the life cycle issue of maintaining an up-to-date assessment package as science and technology improves through time.

1.1 Purpose

The goal of this software design document is to describe the construction and maintenance of a metapopulation assessment modeling system that incorporates up-to-date modeling capabilities and ongoing technological improvements and is readily extensible to adapt to diverse stock assessment situations (**Appendix, Figure A1**). The software design document presented here provides a preliminary model design which will be iteratively refined to produce a “living” design document, which will include detailed data structures and algorithms. The paradigm for this system is that modeling the spatial dynamics of fishery resource and the fishing fleets that harvest it are essential for improved understanding and predictive capacity. An inherent design to handle the spatial structure and its impacts on population dynamics and the fishery system exploiting the stock is a necessary condition for metapopulation stock assessment. In this context, the preliminary design document focuses on spatial population dynamics as approximated by a box-transfer model, noting that other models of spatial dynamics may be needed in some cases (Goethel et al. 2011). Accounting for spatial structure can have substantial impacts on the pattern of fishery selectivity and on the interpretation of relative abundance indices and structured population dynamics and we need to be able to readily include spatial structure in the population dynamics. As a result, the new modeling system is named a “metapopulation assessment system”, or MAS, to reflect this key design element.

In MAS, each instance of a metapopulation model consists of one or more structured subpopulations and one or more habitat areas (or locations or patches, depending upon the ecological context). Here we note that there is an emphasis on constructing alternative credible models versus the expectation of being able to create a single model that provides a best approximation to the set of all plausible movement dynamics of the metapopulation. For example, one might wish to compare a single subpopulation moving between two areas with an alternative two subpopulation model with feeding migrations and natal homing to a single area. Here we use the term “subpopulation” to represent a group of related animals with linked recruitment patterns and similar demographic parameters, including for example, growth characteristics, stock-recruitment resilience, and survival rates. Each subpopulation will have a set of movement probabilities (e_{ij}) among areas in a given time step (e_{ij} is the probability of movement from area i to area j) where a given subpopulation could be very mobile and potentially move into any of the areas ($e_{ij} > 0$) as well as be very sessile and not move at all ($e_{ij} = 0$ for $i \neq j$) from a single habitat area (Figure 1). Information on the local structure and characteristics of each habitat area and associated environmental forcing is to be included in the environment class.

Figure 1. A schematic of a subpopulation within a metapopulation inhabiting three areas and with differing environmental characteristics and movement rates among areas.



The capacity to accommodate subpopulations that have a clumped distribution in space as well as subpopulations that can move through the entire spatial domain is a key design element of MAS. This design feature allows for the modeling of highly migratory stocks, such as salmon and tunas, as a metapopulation comprised of subpopulations that have the capacity to transit

through and mix in various areas. Similarly, the design also allows for the modeling of sedentary populations, such as scallops and surf clams that are comprised of subpopulations with overlapping larval distributions and sessile adult spatial distributions. Over this design feature allows for a wide variety of movement patterns with interannual and seasonal variability to be modeled. Allowing for diverse movement patterns is one of the basic elements in MAS and is also one of the reasons an object-oriented system design is necessary.

1.2 Scope

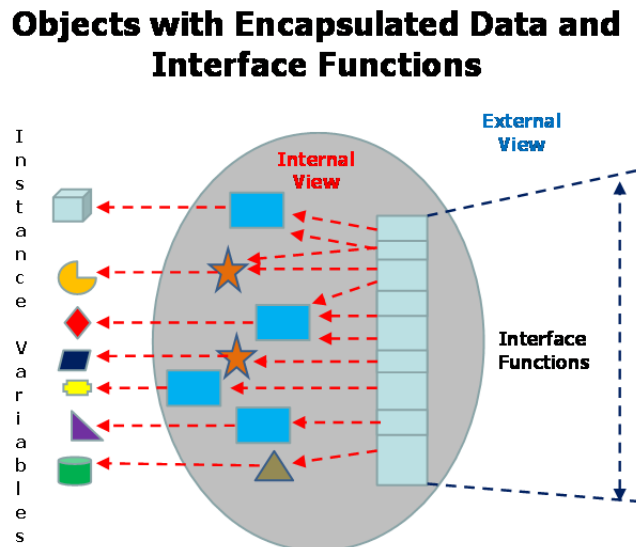
The scope of the MAS design document is to provide modeling and programming specifications for alternative data structures and model parameterizations of metapopulation dynamics. In brief, here are some of the design features that are envisioned, noting that these features will be implemented subject to resource constraints. Movement dynamics among areas can be modeled for highly migratory that move among areas or for relatively sedentary populations with or without exchange of recruitment among areas. The metapopulation size can be a large commercially-important stock with hundreds of millions of individuals or a small bycatch-impacted stock with hundreds of individuals. The spatial resolution of the model areas can be coarse or granular and the size of areas can be uniform or irregular. The parameter estimation process can be based on Bayesian, frequentist, or other approaches, including hierarchical Bayes or random effects for spatial parameters. Metapopulation models can be developed for age- or size-based population structures. Fishery systems can be comprised of multiple fleets operating over multiple areas or consist of a single fleet with changing characteristics. The environmental characteristics of areas can be stationary or time-varying and subpopulations can be subject to local area or global multiple-area environmental forcing. Models can be compared and tested using simulated data sets under alternative operating models and harvest strategies for management strategy evaluation. Best-fitting models or ensembles of the most credible models can be used for making forecasts of annual catch limits and other quantities of interest under alternative future states of nature.

1.3 Overview

This software design document provides specifications for the development of a metapopulation assessment system. This system will employ an object-oriented programming paradigm to ensure software maintenance and flexibility and to improve model extensibility through time. The object-oriented paradigm focuses on the development of class structures to represent the assessment data and function objects needed to construct, optimize, compare, and simulation test a set of metapopulation models for stock assessment applications. The class structures will use data encapsulation and will also include functions with an internal operations view (implementation) for acting on instance variables of the data sets (Figure 2) as well as an external

outward view (specification) with interface functions to share processed data sets with other class objects.

Figure 2. A schematic of data encapsulation for objects with a class structure.



The metapopulation assessment system will consist of a set of libraries of class objects and associated functions to fit fishery system observations. The system will employ just-in-time compilation to produce executable software for specific applications. The system will have the capacity to parallelize some aspects of the model construction, selection and testing, and projection analyses to facilitate large-scale simulations to evaluate model performance under uncertainty. This next section provides an overview of the initial MAS including descriptions of the general system architecture and along with the data and model structure and design of input/output objects.

2. SYSTEM OVERVIEW

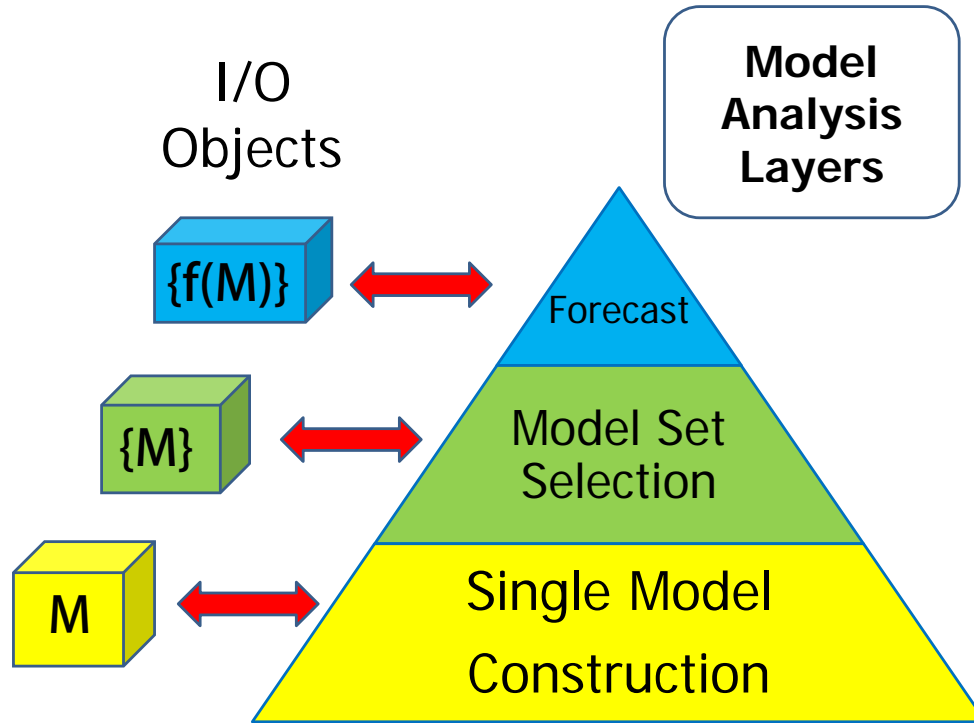
Assessment systems are typically designed to answer management questions using models that are approximations of nature. Assessment analyses typically produce historic and current estimates of stock abundance, mortality, and productivity and may use a variety of model structures fitted to alternative data scenarios. Assessments also need to produce forecasts of stock conditions and fishery yields and other quantities of interest under alternative harvest scenarios. As a result, there are multiple types of modeling analyses that comprise a stock assessment and these can be categorized into analysis layers.

The MAS consists of three analysis layers which correspond to the primary applications and analyses needed to construct and complete a full stock assessment for a metapopulation. The first analysis layer of MAS produces a single metapopulation assessment model which is created and fitted in the single model construction layer (Figure 3). There may be several alternative models, perhaps on the order of dozens, developed in a given stock assessment effort. The process of comparing, contrasting, and selecting a best fitting model, or best-fitting set of models comprises the second model set selection layer (Figure 3). This layer also includes analysis functions to perform sensitivity analyses and quantify uncertainty about quantities of interest.

The third analysis layer is the model forecasting layer in which the best model or set of models are forecast under alternative future states of nature and fishery system harvest control rules, or harvest strategies (Figure 3). This layer includes management strategy evaluation analyses in which alternative assessment models and harvest control approaches are applied to data sets simulated from known operating models for evaluating their efficacy at forecasting quantities of interest. Each of the three primary analysis layers includes a flow control algorithm that specifies the individual processes that can occur in the layer.

Input and output objects need to be constructed to create and distribute information produced in each analysis layer. Ideally, these objects are designed to provide ease of use and interpretability, flexible capacity for simulation testing and model verification, and allow for standard GUI-based model development in MAS. For example, the model selection layer needs to provide input and output processes for model comparison and uncertainty quantification. In this context, one desirable aspect of MAS would be to implement structured model input and output objects that can be efficiently combined and disaggregated to provide compatibility for diverse types of input or output information, e.g., for outputting quantities of interest in specific formats. It is proposed that the JSON (JavaScript Object Notation) syntax be used to construct input and output objects for MAS because of its generality and simplicity. This notation provides a standard data-interchange format based on a subset of the JavaScript Programming Language, namely the Standard ECMA-262 3rd Edition - December 1999. JSON uses a text format that is language independent but also uses two structures, hash tables and arrays, which are basic information structures used in the C-family of languages and structured programming in general. Overall, JSON is a natural data-interchange language for defining input and output objects for the MAS.

Figure 3. Schematic of analysis layers for MAS including model construction, selection, and forecasting.



3. SYSTEM ARCHITECTURE

The model structure is represented in the class `Model` that defines a model as an abstract object. The `Model` class defines the structure of an individual metapopulation model and how it is comprised of data and modeling structures. The `Model` class includes case structures for each of the analysis layers, noting that the single model construction layer is the minimal representation required for a MAS model, i.e., a single assessment model fit using a fixed input data set and fixed model structural assumptions.

The model case structures are important because they specify the logical choice being made for each possible feature in a given model. For the single model construction layer (Figure 3), these choices specify the modeling information needed to represent: (1) the subpopulations within the metapopulation (**Subpopulation**), (2) the set of observational data to be fit by the model (**Observation**), (3) the set of environmental information used to represent the effects of spatial and temporal variation on fishery and population dynamics within areas (**Environment**), and (4) the set of choices made for the analyses of the set of selected input information (**Analysis**). The

model case structures correspond to arrays with a fixed set of possible values for each array element. Here each array element represents a specific model feature, for example, which type of selectivity function is being used to fit length observations for a given fishing fleet. There are model case structures for each of the four primary classes that comprise an instance of the **Model Class** for a metapopulation (Figure 4). These are:

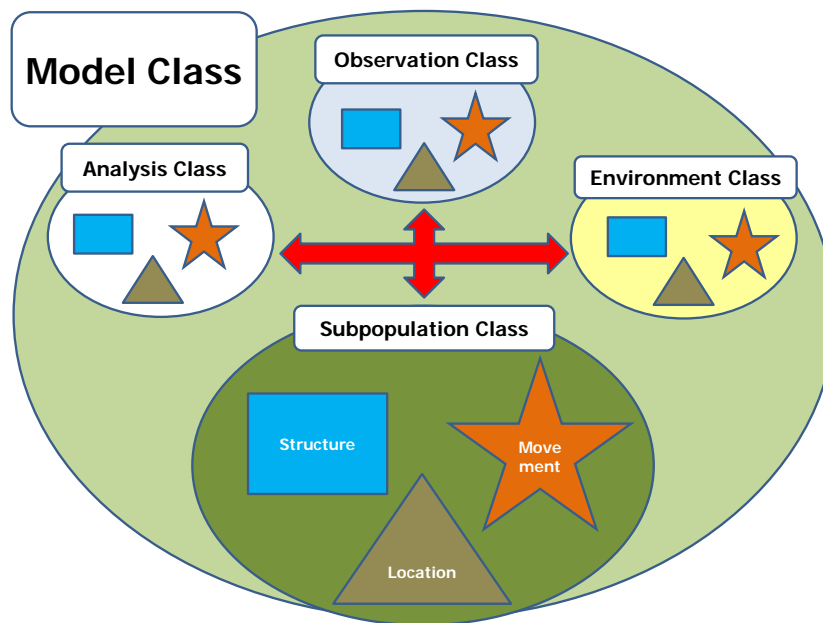
- (1) a **Subpopulation Class** structure that includes subpopulation characteristics, location, and movement attributes (having only one subpopulation, not moving or completely mixing are common assessment assumptions, for example);
- (2) an **Observation Class** with subclass structures that include fisheries data, survey data, results of research studies like estimates of growth, maturity, or fecundity curves or individual observations of tagging or ageing data;
- (3) an **Environment Class** that includes information on habitat in areas and the effects of environmental forcing on subpopulation characteristics and observations with subclasses that include biotic entities or abiotic factors, constant or time-varying predator fields or La Nina/El Nino conditions, for example;
- (4) an **Analysis Class** that has subclasses for parameter input and output, creating simulated input data, likelihood components, model setup and domain parameters, model diagnostics, data weighting and iterative reweighting, model selection and multimodel inference, model averaging, constraints or priors and hyperpriors on model parameters, and uncertainty quantification methods including covariance estimation, bootstrapping, and Markov Chain Monte Carlo simulation and other approaches. Information is transmitted between classes via messages, or interface functions, as indicated by the arrows (Figure 4).

Each of the four primary classes includes data objects and interface functions to transfer information to other classes in order to model the dynamical structure of the subpopulations, including their distribution in space-time and their movement rates between areas. In this context, the spatiotemporal distribution of a subpopulation is the set of areas it inhabits through time. That is, each subpopulation is associated with a particular set of areas through, which could range from one area to all of the areas incorporated in the model. Subpopulations can migrate as groups of individuals, or propagules, through areas. As a result, abundances of subpopulations within specific areas may be reduced through emigration or be increased through immigration.

For the model set selection layer, the model case structures are contingent on the single model construction layer specifications. This hierarchy allows for the rapid development and comparison of alternative models from a single template model that includes the full set of available information for the metapopulation assessment.

Similarly, for the model forecast layer, the model case structures are contingent on the model set selection layer specifications. This feature allows for the rapid development of alternative forecast of models or model ensembles.

Figure 4. Schematic diagram of an assessment model object showing the four classes comprising a metapopulation model.



In summary, the Model Class holds the basic logic needed to put a particular instance of a model object together using model case structures. In addition, the four primary classes needed for constructing an instance of Model Class are: (1) Subpopulation Class, this is where the population dynamics information is contained; (2) Observation Class, this is where the observed data are stored; (3) Environment Class, this is where the biotic and abiotic habitat and environmental information that affect population and fishery dynamics are stored; and (4) Analysis, this is where the likelihood structure, model fitting, and data analysis information is stored.

3.1 Architectural Design

3.1.1 Subpopulation Class

From a top-down perspective, the subpopulation class follows the well-established premise of conservation of biomass (Russell 1931) and numbers within each area of the model. In what follows, we briefly describe the conservation of biomass principle and the conservation of subpopulations numbers principle.

In general, the biomass of the p^{th} subpopulation ($B_{t,s}^{(p)}$) changes through time, indexed by t , in each spatial location, indexed by s , based on biomass increases due to somatic growth of surviving biomass ($GB_{t,s}^{(p)}$), recruitment biomass ($RB_{t,s}^{(p)}$), and immigrating biomass ($IB_{t,s}^{(p)}$) and based on biomass decreases due to natural deaths of biomass ($DB_{t,s}^{(p)}$), fishery catch biomass ($CB_{t,s}^{(p)}$), and emigrating biomass ($EB_{t,s}^{(p)}$) via

$$(1) \quad B_{t+1,s}^{(p)} = B_{t,s}^{(p)} + \left(GB_{t,s}^{(p)} + RB_{t,s}^{(p)} + IB_{t,s}^{(p)} \right) - \left(DB_{t,s}^{(p)} + CB_{t,s}^{(p)} + EB_{t,s}^{(p)} \right)$$

This is the conservation of biomass for a given subpopulation in a given area and time period. Summing the biomass changes in equation (1.1) for an individual subpopulation across areas produces the net change in the total biomass of the subpopulation.

$$(2) \quad \sum_s B_{t+1,s}^{(p)} = \sum_s B_{t,s}^{(p)} + \sum_s \left\{ \left(GB_{t,s}^{(p)} + RB_{t,s}^{(p)} + IB_{t,s}^{(p)} \right) - \left(DB_{t,s}^{(p)} + CB_{t,s}^{(p)} + EB_{t,s}^{(p)} \right) \right\}$$

This is the conservation of biomass for a given subpopulation in a given time period. Summing the subpopulation biomass changes in equation (1.2) across subpopulations produces the net change in metapopulation biomass through time in a given area as

$$(3) \quad \sum_p \sum_s B_{t+1,s}^{(p)} = \sum_p \sum_s B_{p,s,t} + \sum_p \sum_s \left\{ \left(GB_{t,s}^{(p)} + RB_{t,s}^{(p)} + IB_{t,s}^{(p)} \right) - \left(DB_{t,s}^{(p)} + CB_{t,s}^{(p)} + EB_{t,s}^{(p)} \right) \right\}$$

This is the conservation of biomass for the metapopulation in a given time period. Similarly, computing total metapopulation biomass for sex-structured population dynamics requires an additional sum over genders while computing total biomass for a structured population dynamics model requires summing over the structured variable which could be age classes, length classes, or stage classes.

In equations (1.1) to (1.3), the subpopulation biomass information is derived from the subpopulation numbers at age and mean weights at age through time. In this context, the subpopulation numbers at age are initialized for the first period in the time horizon based the model choices from the subpopulation, observation, analysis, and environment classes. The subpopulation class instances are then projected forward with time-varying recruitment though

the assessment time horizon. As the projections sequentially occur, the observation class generates observations through time as fishing fleets and environmental processes remove fish from subpopulations through harvest and expected natural mortality rates or episodic negative environmental conditions.

Similarly, the number of fish in the p^{th} subpopulation ($N_{t,s}^{(p)}$) changes through time in each spatial location based on population size increases due to recruitment ($R_{t,s}^{(p)}$), and immigrating numbers ($I_{t,s}^{(p)}$) and based on population size decreases due to natural deaths in numbers ($D_{t,s}^{(p)}$), fishery catch numbers ($CN_{t,s}^{(p)}$), and emigrating numbers ($E_{t,s}^{(p)}$) via

$$(4) \quad N_{t+1,s}^{(p)} = N_{t,s}^{(p)} + (R_{t,s}^{(p)} + I_{t,s}^{(p)}) - (D_{t,s}^{(p)} + CN_{t,s}^{(p)} + E_{t,s}^{(p)})$$

This is the conservation of population numbers for a given subpopulation in a given area and time period. Summing the population size changes in equation (2.1) for an individual subpopulation across areas produces the net change in the total population size of the subpopulation.

$$(5) \quad \sum_s N_{t+1,s}^{(p)} = \sum_s N_{t,s}^{(p)} + \sum_s \{(R_{t,s}^{(p)} + I_{t,s}^{(p)}) - (D_{t,s}^{(p)} + CN_{t,s}^{(p)} + E_{t,s}^{(p)})\}$$

This is the conservation of population size for a given subpopulation in a given time period. Summing the subpopulation size changes in equation (2.2) across subpopulations produces the net change in metapopulation size.

$$(6) \quad \sum_p \sum_s N_{t+1,s}^{(p)} = \sum_p \sum_s N_{t,s}^{(p)} + \sum_p \sum_s \{(R_{t,s}^{(p)} + I_{t,s}^{(p)}) - (D_{t,s}^{(p)} + CN_{t,s}^{(p)} + E_{t,s}^{(p)})\}$$

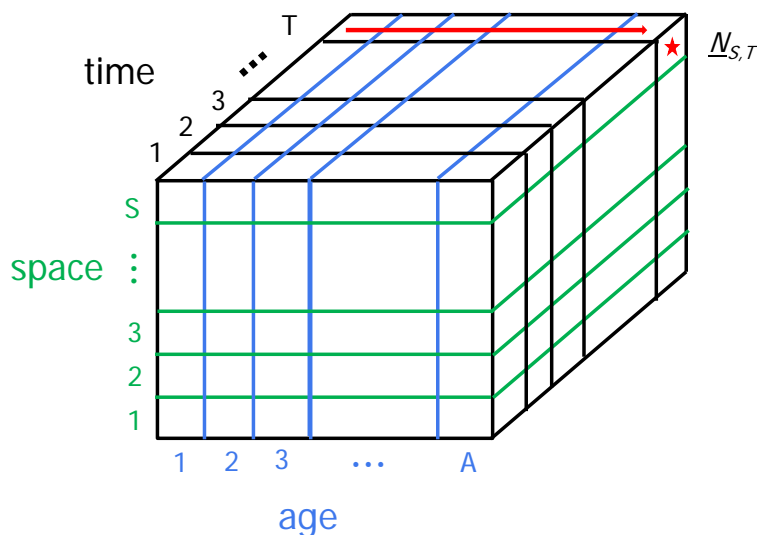
This is the conservation of population size for the metapopulation in a given area. Similarly, computing total metapopulation size for sex-structured population dynamics requires an additional sum over genders while computing total population size for a structured population dynamics model requires summing over the structured variable which could be age classes, length classes, or stage classes.

The structure of the class subpopulation in MAS includes multidimensional arrays for storing the numbers ($N^{(p)}$) of fish in each subpopulation (indexed by p) in particular space, age-size, and gender cells through time. The space dimension consists of S categories for spatial location. The convention of using S areas gives a total of $S - 1$ full degrees of freedom for the empirical multinomial distribution in this dimension. Similarly, there are Y time categories for year and T time categories for season if seasonal time periods are included in the subpopulation dynamics. There are a total of A age- or size-structured categories for subpopulation structure, and a total of

G categories for gender structure with associated degrees of freedom of $Y - 1$, $T - 1$, $A - 1$, and $G - 1$, respectively.

The information on the subpopulation numbers in space and time will be stored in multidimensional arrays with dimensions for year, time within year, or season, space, age, and gender. The convention for ordering the subscripts for these multidimensional arrays is $(\underline{Year}, \underline{Time}, \underline{Space}, \underline{Age}, \underline{Gender})$. This ordering conforms to the focus on temporal changes in the spatial distribution of the metapopulation as subpopulations within areas. It also provides the most efficient sequence of accessing computer memory within a time loop. In what follows, we will use a simpler subscript set consisting of $(\underline{Time}, \underline{Space}, \underline{Age})$, noting that if subscripts are omitted, then the index ordering convention remains the same and also noting that numbers by gender can be summed over space and age to produce total subpopulation numbers at any time. The proportion of the i^{th} subpopulation in area s during time period t at age a is denoted by $P_{t,s,a}^{(i)}$ and the corresponding subpopulation number of fish in area s at time t is $N_{t,s}^{(i)} = \sum_a N_{t,s,a}^{(i)} = \sum_a P_{t,s,a}^{(i)} N_{t,s}^{(i)}$. Similarly, the total number of fish in the i^{th} subpopulation at time t is $N_t^{(i)} = \sum_a \sum_s N_{t,s,a}^{(i)} = \sum_a \sum_s P_{t,s,a}^{(i)} N_{t,s}^{(i)}$. Thus, the information needed to specify the i^{th} subpopulation numbers at any time includes the 3-dimensional subpopulation proportion array $\underline{\underline{P}}^{(i)} = (P_{t,s,a}^{(i)})$ which contains the subpopulation proportions at age by area and the 2-dimensional subpopulation size array $\underline{\underline{N}}^{(i)} = (N_{t,s}^{(i)})$ which contains the total subpopulation numbers in each area through time.

Figure 5. A three dimensional representation of an age-structured subpopulation in space through time.



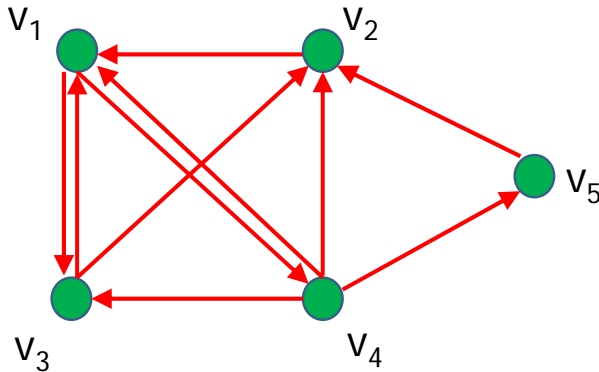
Age composition data can be observed to characterize the age structure of the subpopulation in a given location and time. An age composition sample represents the age structure contained in a 1-dimensional subarray of the subpopulation array, as for example, the array slice $\underline{N}_{T,S}$ shown the red star and arrow (Figure 5) with the subpopulation superscript omitted.

An important element of the subpopulation class is the set of areas the subpopulation can inhabit and have groups within the subpopulation move from area to area. In this case, we are focusing on box-transfer movement (Goethel et al. 2011) in this design document. This set of inhabitable areas is the habitat of the subpopulation. The set of habitable areas for a subpopulation can be represented as a directed graph G in which the connected vertices (v) are the habitable areas and the edges (e) are the directed movement paths between areas. Here is an example of the directed graph for a subpopulation inhabiting five areas and with one-step movement paths (Figure 6). The information on the nearness of the areas within a subpopulation's habitat is represented with an adjacency matrix $\underline{\underline{A}}$. If the directed habitat graph has S vertices or areas, numbered v_1, v_2, \dots, v_S , then the adjacency matrix $\underline{\underline{A}} = (a_{i,j})_{S \times S}$ is $S \times S$ with entries $a_{i,j} = 1$ if there is an edge with initial vertex v_i and terminal vertex v_j in the directed graph and $a_{i,j} = 0$ if there is no edge. The adjacency matrix describes the set of possible one-step movements of the subpopulation within its habitat.

Here is the adjacency matrix $\underline{\underline{A}}$ for the example of a subpopulation with a habitat consisting of five areas (Figure 5).

$$(7) \quad \underline{\underline{A}} = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Figure 6. Directed movement graph of the habitat for a subpopulation inhabiting five areas.



In general, areas that are not part of the subpopulation habitat are represented as a row with zero entries. As a result, the adjacency matrices for each of the subpopulations can be taken without loss of generality to have full dimension $S \times S$. This provides a succinct representation of how subpopulations can move through their habitat in one step or on a path of length one.

The adjacency matrix can also be used to determine the number of paths of any length between two areas. This follows from the fact that the entries of $\underline{\underline{A}}$ raised to the power $m \geq 0$ count the number of distinct paths of length m between two vertices. In particular, if $a_{i,j}^m$ denotes the $(i, j)^{th}$ entry in the matrix $\underline{\underline{A}}^m$, then $a_{i,j}^m$ is the number of paths of length m between the vertices v_i and v_j .

It is useful to note that the number of paths of length zero is given by $\underline{\underline{A}}^0 = \underline{\underline{I}}$, where $\underline{\underline{I}}$ is the identity matrix. This identity matrix corresponds to subpopulation groups that do no move between areas noting that a path of length zero connects an area to itself. Thus, the sum of the adjacency matrix and the identity matrix $\underline{\underline{A}}^0 + \underline{\underline{A}}$ defines the set of all possible connecting paths of edges between areas for the subpopulation movement probabilities.

The movement probabilities define the flow of the individuals in the subpopulation among areas and are associated with the edges in the subpopulation habitat graph. The probability that an individual fish of age a from subpopulation p in area s_i moves to area s_j during time period t is denoted by $e_{t,i,j,a}^{(p)}$, where these probabilities are all nonnegative and satisfy the constraint that they sum to unity for a given subpopulation (p), source area (i), age structure (a), and time period (t). That is,

$$(8) \quad \sum_{j=1}^S e_{t,i,j,a}^{(p)} = 1$$

The movement probabilities can be represented as time-specific and age-specific matrices $\underline{\underline{E}}_{t,a}^{(p)}$

$$\text{where } \underline{\underline{E}}_{t,a}^{(p)} = \begin{pmatrix} e_{t,1,1,a}^{(p)} & \cdots & e_{t,1,S,a}^{(p)} \\ \vdots & \ddots & \vdots \\ e_{t,S,1,a}^{(p)} & \cdots & e_{t,S,S,a}^{(p)} \end{pmatrix}_{S \times S}$$

and $\underline{\underline{E}}_{t,a}^{(p)}$ is the subpopulation movement matrix for structure or age class a and time period t .

Each row of the subpopulation movement matrix $\underline{\underline{E}}_{t,a}^{(p)}$ that is part of the subpopulation habitat has nonnegative entries that sum to unity as in equation (3). Rows of $\underline{\underline{E}}_{t,a}^{(p)}$ that are not part of the

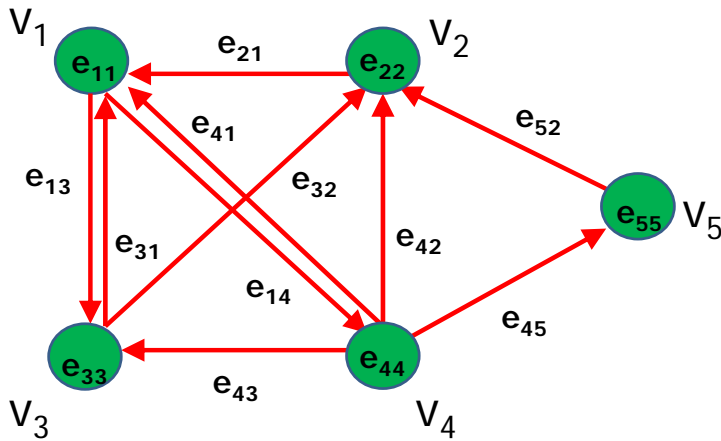
subpopulation habitat have zero entries. The movement probability matrices can also be constant through time or constant across age structure or both, although the most general setting is for the probabilities to vary through time and by age structure.

Here is an example of a movement matrix $\underline{\underline{E}}$ corresponding to the five area subpopulation habitat along with the associated habitat graph (Figure 7) where the movement probabilities are constant in time and by age structure

$$(9) \quad \underline{\underline{E}} = \begin{pmatrix} e_{11} & 0 & e_{13} & e_{14} & 0 \\ e_{21} & e_{22} & 0 & 0 & 0 \\ e_{31} & e_{32} & e_{33} & 0 & 0 \\ e_{41} & e_{42} & e_{43} & e_{44} & e_{45} \\ 0 & e_{52} & 0 & 0 & e_{55} \end{pmatrix}$$

and satisfy the five probability constraints $e_{11} + e_{13} + e_{14} = 1$, $e_{21} + e_{22} = 1$, $e_{31} + e_{32} + e_{33} = 1$, $e_{41} + e_{42} + e_{43} + e_{44} + e_{45} = 1$, and $e_{52} + e_{55} = 1$.

Figure 7. Directed movement graph of the habitat for a subpopulation inhabiting five areas along with the associated movement probability matrix $\underline{\underline{E}} = (e_{ij})$.



This shows the minimal architectural elements for the subpopulation class, a structured subpopulation distributed in space and time along with the associated movement probabilities. Recruitment processes for a subpopulation are modeled in a similar manner noting that only the spatial distribution of new recruits among areas is needed. The subpopulation class interfaces with the information class as well as the observation and environment classes to describe the metapopulation dynamics through time.

3.1.2 Observation Class

The observation class is the set of observed data for the metapopulation and fishery system. This includes fisheries data, survey data, and information from studies like estimates of growth, maturity, or fecundity curves or individual observations of tagging or ageing data. Model parameters are fit to the observed data through interface with the subpopulation, environment, and analysis classes. The observation class is comprised of the basic data and the objects used to store this information and interface with the subpopulation, environment, and information classes. The structure of the observation class is described in the Data Design section.

3.1.3 Environment Class

The environment class characterizes the information on the metapopulation habitat by area. This includes the effects of environmental forcing on subpopulation characteristics and observed data. This class includes biotic entities or abiotic factors, constant or time-varying survival and growth conditions dependent on the existing environmental conditions in each area. The environment class describes the exogenous conditions affecting the metapopulation dynamics by area and interfaces with the subpopulation, observation, and information classes. Details of the observation class are described in the Data Design section.

3.1.4 Analysis Class

The analysis class stores the information on the functions and algorithms needed to fit the parameters of a general discrete-time metapopulation model structure that can include process and observation dynamics (Aoki 1990, Schnute 1994). This class includes information needed for parameter input and output, creating simulated input data, likelihood components, model setup and domain parameters, model diagnostics, data weighting and iterative reweighting, model selection and multimodel inference, model averaging, constraints or priors and hyperpriors on model parameters, and uncertainty quantification methods including covariance estimation, bootstrapping, Markov Chain Monte Carlo simulation as well as other approaches. In general, the model structure and information requirements depend on a state space formulation of the metapopulation dynamics, which is described below.

The general discrete-time state space model depends on the relationships between the system state vector (\underline{S}), the observed data vector (\underline{D}), and the parameter vector to be estimated ($\underline{\Theta}$). We note that the set of vectors can be represented by as a single vector or structured as scalars, vectors, and subarrays in a linked list. The state vector $\underline{S}(t)$ has dimension p and represents the set of unobserved metapopulation and fishery system characteristics at time t . The data subvector $\underline{D}(t) \in \underline{D}$ for time period t has dimension n_t and represents the set of quantities actually observed at time t . If the set of control variables (e.g., catch or fishing effort) for the system are empirical observations, then these variables will be included in the data vector in contrast to

being included in the parameter vector. The parameter vector $\underline{\Theta}(t)$ has dimension z and represents the combined set of all parameters used to describe the process and observation dynamics as well as fitting of the state space model. This includes parameters for population model dynamics ($\underline{\Theta}_\theta$), parameters for prior distributions ($\underline{\Theta}_\pi$), parameters for hyperprior distributions ($\underline{\Theta}_\varphi$), parameters for process error distributions ($\underline{\Theta}_\sigma$), parameters for observation error distributions ($\underline{\Theta}_\tau$), parameters for system control ($\underline{\Theta}_C$) if appropriate, and nuisance parameters ($\underline{\Theta}_\eta$) if parameter constraints or likelihood penalties are used to improve stability of the numerical computations and as a result, the overall parameter vector is

$\underline{\Theta} = (\underline{\Theta}_\theta, \underline{\Theta}_\pi, \underline{\Theta}_\varphi, \underline{\Theta}_\sigma, \underline{\Theta}_\tau, \underline{\Theta}_C, \underline{\Theta}_\eta)$ and the estimated parameter subvector ($\underline{\Theta}_E$) is

$\underline{\Theta}_E = (\underline{\Theta}_\theta, \underline{\Theta}_\sigma, \underline{\Theta}_\tau)$. This gives a full accounting of the parameters to be estimated versus assumed for the state space model, although we would also add that there may be situations where it is reasonable given the information constraints to assume structural dependence of the observation process on the observed data, i.e., an empirical Bayes framework, and that we do not in any way discount the utility of such extensions of the multimodel approach in that manner.

The state space model structure ensures that the time dimension can be represented as a memory-less process in which the next system state only depends on the current system state and where time-lagged variables are included in the state and parameter vectors as needed. It also implies that the model structure is set up for Markov Chain Monte Carlo simulation which can be used to numerically integrate the posterior distribution of a given model (Gilks et al. 1996, Gelman et al. 2005) and also for conducting future forecasts which account for uncertainty in the estimate of system state (Brodziak et al. 1998, Patterson et al. 1999). Last, it is important to note that the state space model structure is completely general and can be applied to single- and multiple-species models with spatial structuring.

The process dynamics of the state space model depend on the previous system state and parameter vector through the predictive function \underline{F} with its range consisting of a smooth region $R \in \mathbb{R}^p$ where

$$(10) \quad \underline{S}(t+1) = \underline{F}(\underline{S}(t), \underline{\Theta}(t) | \underline{D}(t)) + \underline{\Lambda}_p(t, \underline{\Theta})$$

The likelihood function $\underline{\Lambda}_p(t, \underline{\Theta}) = (\Lambda_{p,1}(t, \underline{\Theta}), \dots, \Lambda_{p,N_p}(t, \underline{\Theta}))$ represents the zero mean process error distributions, whether process errors are modeled as additive normal distributions as in (5), or multiplicative lognormal distributions or other distributional forms. Similarly, the system observation dynamics are governed by the predictive function \underline{G} with its range consisting of a smooth region $R \in \mathbb{R}^{n_t}$ where the predicted information depends on the system state and parameter vector at time t

$$(11) \quad \underline{D}(t) = \underline{G}(\underline{S}(t), \underline{\Theta}(t)) + \underline{\Lambda}_o(t, \underline{\Theta})$$

and the likelihood function $\underline{\Lambda}_o(t, \underline{\Theta}) = (\Lambda_{o,1}(t, \underline{\Theta}), \dots, \Lambda_{o,N_o}(t, \underline{\Theta}))$ represents the zero mean observation error distributions. There are two points about this general representation of the process and observation dynamics to keep in mind. First, the time dynamics are structured to include a discrete Markov time step and given this constraint, if spatial model formulations of the population characteristics variables are included, then these variables need to have informative movement rates among spatial regions within the state space model. Second, while not specifying the form of spatiotemporal constraints for the state space model, we note that this is a ripe and open field for investigation given the general nature of the process and observation dynamical framework.

The joint likelihood $L(\underline{D} | \underline{\Theta})$ of the metapopulation model is constructed from the individual likelihoods that compare model predictions to observed data or process information through time. That is, the joint likelihood is

$$(12) \quad L(\underline{D} | \underline{\Theta}) = \prod_{j=1}^{N_p} \prod_{t=1}^T \Lambda_{p,j}(t, \underline{\Theta}) + \prod_{k=1}^{N_o} \prod_{t=1}^T \Lambda_{o,k}(t, \underline{\Theta})$$

where N_p and N_o are the number of process and observation likelihood terms and noting that terms with no information have the value unity and do not affect the calculation of the joint likelihood value.

The objective function $\Psi_m(\underline{\Theta} | \underline{D})$ for a likelihood-based model (m) is based on the joint likelihood and additional prior distribution information or constraints, if applicable. The objective function is optimized using a numerical optimization package, e.g., ADMB.

Two primary estimation approaches are supported in MAS. Under a frequentist estimation approach, the maximum likelihood estimate of the parameter vector $\underline{\Theta}_{MLE}$ is obtained by minimizing the objective function $\Psi_m(\underline{\Theta} | \underline{D})$. The frequentist objective function is comprised of the negative loglikelihood and the sum of any additional penalty terms for system control $\Upsilon_c(\underline{S} | \underline{\Theta})$ or nuisance parameters $\Upsilon_n(\underline{S} | \underline{\Theta})$. That is, the frequentist objective function is expressed as

$$(13) \quad \Psi_m(\underline{\Theta} | \underline{D}) = -\log(L(\underline{D} | \underline{\Theta})) + \Upsilon_c(\underline{S} | \underline{\Theta}) + \Upsilon_n(\underline{S} | \underline{\Theta})$$

In this context, mixed-effects models can be accommodated through restricted maximum likelihood information with structural nuisance parameters for variances.

Under a Bayesian estimation approach, the posterior distribution of the model parameters is sampled through numerical simulation. For simple Bayes models without hierarchical structure, the posterior distribution $P(\underline{\Theta} | \underline{D})$ is proportional to the product of the prior distributions

$\pi_k(\underline{\Theta}_\pi)$ and the joint likelihood $L(\underline{D} | \underline{\Theta})$ expressed as $P(\underline{\Theta} | \underline{D}) \propto \prod_k \pi_k(\underline{\Theta}_\pi) \cdot L(\underline{D} | \underline{\Theta})$. In simple case, the objective function consists of the negative logarithm of the product of the priors and the joint likelihood along with the sum of penalty terms. That is, the simple Bayes objective function is expressed as

$$(14) \quad \Psi_m(\underline{\Theta} | \underline{D}) = - \sum_k \log(\pi_k(\underline{\Theta}_\pi)) - \log(L(\underline{D} | \underline{\Theta})) + \Upsilon_c(\underline{S} | \underline{\Theta}) + \Upsilon_\eta(\underline{S} | \underline{\Theta})$$

For hierarchical Bayes models, the posterior will also include information from the hyperprior distribution $\varphi(\underline{\Theta}_\varphi)$ and the posterior distribution is proportional to

$$(15) \quad P(\underline{\Theta} | \underline{D}) \propto \prod_k (\varphi_k(\underline{\Theta}_\varphi) \cdot \pi_k(\underline{\Theta}_\pi | \underline{\Theta}_\varphi)) \cdot L(\underline{D} | \underline{\Theta})$$

In the hierarchical case, the objective function consists of the negative logarithm of the product of the hyperpriors, priors and the joint likelihood along with the sum of penalty terms. That is, the hierarchical Bayes objective function is expressed as

$$(16) \quad \begin{aligned} \Psi_m(\underline{\Theta} | \underline{D}) = & - \sum_k \log(\varphi_k(\underline{\Theta}_\varphi)) - \sum_k \log(\pi_k(\underline{\Theta}_\pi | \underline{\Theta}_\varphi)) - \log(L(\underline{D} | \underline{\Theta})) \\ & + \Upsilon_c(\underline{S} | \underline{\Theta}) + \Upsilon_\eta(\underline{S} | \underline{\Theta}) \end{aligned}$$

Under either the frequentist or Bayesian estimation approaches, parameter uncertainty can be characterized using standard techniques, e.g., estimating the asymptotic covariance matrix, bootstrapping, or Markov Chain Monte Carlo simulation, given the parameter estimate $\hat{\underline{\Theta}}$ and estimate of the parameter covariance matrix $\hat{\underline{\Sigma}}$.

The capacity to conduct stock projections depends on the estimates of model parameters and their estimated covariance matrix. In practice, it is also important to have input and output structures that can efficiently support model verification through simulation testing as well as management strategy evaluation analyses. As noted previously, we are using JSON to produce the necessary input and output structures for MAS.

3.2 Subsystem Description

The Model Class requires a model parameter array for organizing the information on the definition of each model parameter. Under a given estimation procedure (i.e., MLE, Partial Bayes, REML, Simple Bayes, Hierarchical Bayes, ...), the information needed to characterize each parameter includes: (1) Is the parameter is to be fixed (constant in space and time) or to be estimated?; (2) If estimated, what is the order of the parameter's estimation in the estimating procedure?; (3) First, second, or third order derivative information, as needed; (3) Prior information, as needed; (4) Space- or time-varying effects, as needed; (5) Place holder for other information, as needed. This setup provides a flexible and extensible model parameter array.

Given the setup of the model parameter array, the set of subsystem objects and associated parameters for the four primary classes are described below.

3.2.1 Subpopulation Class

○ Structure

- Two gender ($g=2$) or pooled-sex ($g=1$) population dynamics model similar to that used in Methot and Wetzel (2013).
- Subpopulation (p) numbers by space (s), age (a) or length, and sex (g) through time (t), $N_{t,s,a,g}^{(p)}$. That is, $N_{t,s,a,g}^{(p)}$ is the number of age- a and gender- g fish in area- s from subpopulation p at the start of time period t .
- Subpopulation biomass by area through time, $B_{t,s}^{(p)} = \sum_a \sum_g B_{t,s,a,g}^{(p)}$.
- Subpopulation growth processes by spatial location, age or length, and sex through time, $G_{t,s,a,g}^{(p)}$.
 - Von Bertalanffy growth model by spatial location (s). Length at age a is denoted by $L_s(a)$ and is given by

$$(17) \quad L_s(a) = L_{\infty,s} \left(1 - \exp(-k_s(a - a_{0,s})) \right)$$

- Schnute (1981) growth models.
Case I. Length at age a is given by

$$(18) \quad L(a) = \left\{ L_{\min}^{\beta} + (L_{\max}^{\beta} - L_{\min}^{\beta}) \cdot \frac{1 - \exp(-\alpha(a - a_{\min}))}{1 - \exp(-\alpha(a_{\max} - a_{\min}))} \right\}^{\beta^{-1}}$$

for parameters $\alpha, \beta, L_{\min}, L_{\max}$ where L_{\min} is estimated size at the minimum age a_{\min} and L_{\max} is estimated size at the maximum age a_{\max}

Case II. Length at age a is given by

$$(19) \quad L(a) = L_{\min} \cdot \exp \left(\log \left(\frac{L_{\max}}{L_{\min}} \right) \frac{1 - \exp(-\alpha(a - a_{\min}))}{1 - \exp(-\alpha(a_{\max} - a_{\min}))} \right)$$

for parameters $\alpha, L_{\min}, L_{\max}$

Case III. Length at age a is given by

$$(20) \quad L(a) = \left\{ L_{\min}^{\beta} + (L_{\max}^{\beta} - L_{\min}^{\beta}) \cdot \frac{a - a_{\min}}{a_{\max} - a_{\min}} \right\}^{\beta^{-1}}$$

for parameters $\beta, L_{\min}, L_{\max}$

Case IV. Length at age a is given by

$$(21) \quad L(a) = L_{\min} \cdot \exp \left(\log \left(\frac{L_{\max}}{L_{\min}} \right) \frac{a - a_{\min}}{a_{\max} - a_{\min}} \right)$$

for parameters L_{\min}, L_{\max}

Growth can also be independent or dependent on environmental processes or population biomass by area or other factors.

- Example of density-dependence of von Bertalanffy asymptotic length ($L_{\infty,s}^{(p)}$) on subpopulation biomass by spatial location (s).

$$(22) \quad L_{\infty,s}^{(p)} = L_{\infty} - \alpha^{(p)} \cdot B_s^{(p)}$$

for a metapopulation asymptotic length L_{∞} and subpopulation density-dependence parameter $\alpha^{(p)}$.

- Subpopulation recruitment processes by area through time, $R_{t,s}^{(p)}$. For example, subpopulation recruitment numbers can be modeled using a Beverton-Holt stock-recruitment relationship with lognormal variation. In this case, total subpopulation recruitment, $R_t^{(p)}$, is a function of total subpopulation spawning biomass, $SB_t^{(p)}$, as

$$(23) \quad R_t^{(p)} = \frac{4h^{(p)}R_0^{(p)}SB_t^{(p)}}{SB_0^{(p)}(1-h^{(p)}) + SB_t^{(p)}(5h^{(p)}-1)} \cdot e^{-0.5\sigma_{R^{(p)}}^2 + \Delta_{R_t^{(p)}}} \quad \text{where } \Delta_{R_t^{(p)}} \sim N(0, \sigma_{R^{(p)}}^2)$$

and where $h^{(p)}$ is stock-recruitment steepness of the p^{th} subpopulation, $SB_0^{(p)}$ is unfished spawning biomass, $R_0^{(p)}$ is unfished recruitment, $\Delta_{R_t^{(p)}}$ is the log-scale recruitment deviation in year t , and $\sigma_{R^{(p)}}^2$ is the log-scale variance of recruitment deviations.

In general, recruitment can be independent or dependent on environmental processes and will depend on maturation and fecundity processes for the calculation of spawning biomass (see, for example, Subbey et al. 2014).

- Subpopulation fishing mortality, natural mortality, and total mortality rates by area, age, and gender through time, $F_{t,s,a,g}^{(p)}$, $M_{t,s,a,g}^{(p)}$, and $Z_{t,s,a,g}^{(p)}$ with

$$(24) \quad Z_{t,s,a,g}^{(p)} = M_{t,s,a,g}^{(p)} + F_{t,s,a,g}^{(p)}$$

- Subpopulation p has movement probabilities from area i to area s for age structure a and gender g in time t and these probabilities are denoted by $e_{t,i,s,a,g}^{(p)}$

- Initial numbers at age and gender by subpopulation and area

- The number of age- a and gender- g fish in area- s from subpopulation p at the start of the assessment time horizon, time $t=0$, is $N_{0,s,a,g}^{(p)}$
- Recruitment (age-0 fish) in subpopulation p in space s of gender g , $N_{0,s,0,g}^{(p)} = f_s^{(p)}(\Theta)$, at time step $t=0$, where f is a general function of model parameters and data. In practice, one can treat the initial recruitment value as a parameter to be estimated given a constant sex ratio r_g , that is,

$$(25) \quad N_{0,s,0,g}^{(p)} = r_g \cdot N_{0,s,0}^{(p)}$$

- True age classes ($a = \text{age-1 to age-(A-1)}$ fish), $N_{0,s,a,g}^{(p)} = f_{s,a,g}^{(p)}(\Theta)$, at time step $t=0$, where f is a general function of model parameters and data.
- Plus-group ($a = \text{age-A and older}$) $N_{0,s,A,g}^{(p)} = f_{s,A,g}^{(p)}(\Theta)$, at time step $t=0$, where f is a general function of model parameters and data.
- Example: Consider the p^{th} subpopulation using a pooled-sex model with constant age- and space-specific movement probabilities from area i to area s ($e_{i,s,a}^{(p)}$) and space-specific natural mortality rates ($M_{s,a}^{(p)}$).

How does one calculate the initial numbers of fish in subpopulation p in area s of age a at time $t=0$, $N_{0,s,a}^{(p)}$? This is an iterative calculation to account for the age-specific movement rates of fish between areas. That is, the numbers at age are computed sequentially for ages $a = 0, 1, \dots, A-1, A$.

- Recruitment (age $a=0$): Recruitment is the sum over spatial locations (i) of the number of initial recruits produced in that space ($R_{0,i}^{(p)}$) times the probability that a recruit is transported to area s ($q_{i,s}^{(p)}$). That is

$$(26) \quad N_{0,s,0}^{(p)} = \sum_{i=1}^S R_{0,i}^{(p)} \cdot q_{i,s}^{(p)}$$

In practice, one can estimate a total initial recruitment parameter for the subpopulation ($R_0^{(p)}$) and apportion this recruitment to spatial locations based on a long-term average of early life history stage survey information, e.g. CalCOFI.

- True age classes (ages $a < A$): Subpopulation numbers at age is the sum over spatial locations (i) of the number of survivors, $N_{0,i,a-1}^{(p)} \cdot \exp(-Z_{i,a-1}^{(p)})$, times the probability that a survivor moves to area s , denoted by $e_{i,s,a-1}^{(p)}$. That is

$$(27) \quad N_{0,s,a}^{(p)} = \sum_{i=1}^S N_{0,i,a-1}^{(p)} \cdot \exp(-Z_{i,a-1}^{(p)}) \cdot e_{i,s,a-1}^{(p)}$$

Here the total instantaneous mortality for area i at age $a-1$ includes the natural and fishing mortality rates during the initial time period.

- Plus group (ages $a=A$ and older): Subpopulation numbers at age in the plus group is the sum over spatial locations (i) of the number of age- $(A-1)$ survivors ($N_{0,i,A-1}^{(p)} \cdot \exp(-Z_{i,A-1}^{(p)})$) times the probability that a survivor moves to area s ($e_{i,s,A-1}^{(p)}$) times the plus group survival factor for ages A and older $\left(\frac{1}{1 - \exp(-Z_{s,A}^{(p)})} \right)$. That is,

$$(28) \quad N_{0,s,A}^{(p)} = \sum_{i=1}^S N_{0,i,A-1}^{(p)} \cdot \exp(-Z_{i,A-1}^{(p)}) \cdot e_{i,s,A-1}^{(p)} \cdot (1 - \exp(-Z_{s,A}^{(p)}))^{-1}$$

- Changes in numbers at age and gender through time (**Appendix, Figure A2**)
 - Single Age Class (a). Using a discrete time approximation with catch occurring at the midpoint of each time interval and a box transfer model for fish movement probabilities, the number of age- a and gender- g fish in area- s from subpopulation p at the start of time period $t+1$, $N_{t+1,s,a,g}^{(p)}$, depends on two components:
 - (i) the previous number of age- $(a-1)$ and gender- g fish in area s from subpopulation p at the start of time period t that stayed in area s , survived, and did not emigrate to another area not equal to s , noting that this movement probability is $e_{t,s,s,a-1,g}^{(p)}$
 - (ii) the number of age- $(a-1)$ and gender- g fish from subpopulation p that survive and immigrate into area s in time period t from another area i not equal to area s . For one area $i \neq s$, this movement probability is $e_{t,i,s,a-1,g}^{(p)}$, noting that this formulation of movement assumes that emigration and immigration occur at the end of time period t . For single age classes a , the change in numbers at age can be modeled as

$$\begin{aligned}
 (29) \quad N_{t+1,s,a,g}^{(p)} &= \left[N_{t,s,a-1,g}^{(p)} \cdot \exp\left(\frac{-M_{t,s,a-1,g}^{(p)}}{2}\right) - CN_{t,s,a-1,g}^{(p)} \right] \cdot \exp\left(\frac{-M_{t,s,a-1,g}^{(p)}}{2}\right) \cdot e_{t,s,s,a-1,g}^{(p)} \\
 &+ \sum_{i \neq s} \left[N_{t,i,a-1,g}^{(p)} \cdot \exp\left(\frac{-M_{t,i,a-1,g}^{(p)}}{2}\right) - CN_{t,i,a-1,g}^{(p)} \right] \cdot \exp\left(\frac{-M_{t,i,a-1,g}^{(p)}}{2}\right) \cdot e_{t,i,s,a-1,g}^{(p)}
 \end{aligned}$$

In the expression for changes in subpopulation numbers at age above, the observed catch numbers by subpopulation, area, age, gender, and time $CN_{t,s,a,g}^{(p)}$ is calculated as the product of the catch by area, age, gender, and time $CN_{t,s,a,g}$ and the proportion of subpopulation numbers by area, age, gender, and time $P_{t,s,a,g}^{(p)}$. That is, $CN_{t,s,a,g}^{(p)} = CN_{t,s,a,g} \cdot P_{t,s,a,g}^{(p)}$. Typically, there may be little direct information on the proportion of subpopulation numbers that are caught by area, age, gender, and time. In the absence of observations on the catch composition, each subpopulation proportion of the catch will be calculated based on the assumption that the subpopulation proportion is equal to the subpopulation total numbers by area, age, gender, and time divided by the metapopulation total numbers by area, age, gender, and time. These subpopulation proportions are

tracked in the assessment model and this leads to a default assumption that the subpopulation proportion is

$$(30) \quad P_{t,s,a,g}^{(p)} = \frac{N_{t,s,a,g}^{(p)}}{\sum_p N_{t,s,a,g}^{(p)}}$$

This expression is an approximation because in general, the movement of individuals from subpopulations across area boundaries will operate continuously throughout a unit time step t . In contrast, if there is direct information on the subpopulation proportions then these observations can be formulated as a likelihood component, noting that proportion information for one area affects the proportion information for all other areas.

For the plus-group or accumulator age A , the expression for $N_{t+1,s,A,g}^{(p)}$ includes terms for age- $(A-1)$ and age- A fish. Other than the additional factor of the age- A fish, the derivation of the changes in numbers is the same as for ages $a < A$.

- Plus Group Age Class (A). Using a discrete time approximation with catch occurring at the midpoint of each time interval, the number of age- A and gender- g fish in area- s from subpopulation p at the start of time period $t+1$, $N_{t+1,s,A,g}^{(p)}$, depends on four components:
 - (i) the number of age- $(A-1)$ and gender- g fish in area s from subpopulation p at the start of time period t that stayed in area s and did not emigrate to another area i not equal to s , noting that these fish are the new additions to the plus group and that this movement probability is $e_{t,s,s,A-1,g}^{(p)}$
 - (ii) the number of age- $(A-1)$ and gender- g fish from subpopulation p that immigrate into area s in time period t from another area i not equal to area s . For one area $i \neq s$, this movement probability is $e_{t,i,s,A-1,g}^{(p)}$.
 - (iii) the number of age- A and gender- g fish in area s from subpopulation p at the start of time period t that stayed in area s and did not emigrate to another area not equal to s , noting that these are the existing already in the plus group at time t and that this movement probability is $e_{t,s,s,A,g}^{(p)}$
 - (iv) the number of age- A and gender- g fish from subpopulation p that immigrate into area s in time period t from another area i

not equal to area s . For one area $i \neq s$, this movement probability is $e_{t,i,s,A,g}^{(p)}$. For the plus group age classes A , this can be modeled as

$$\begin{aligned}
 N_{t+1,s,A,g}^{(p)} = & \left[N_{t,s,A-1,g}^{(p)} \cdot \exp\left(\frac{-M_{t,s,A-1,g}^{(p)}}{2}\right) - CN_{t,s,A-1,g}^{(p)} \right] \cdot \exp\left(\frac{-M_{t,s,A-1,g}^{(p)}}{2}\right) e_{t,s,s,A-1,g}^{(p)} \\
 & + \sum_{i \neq s} \left[N_{t,i,A-1,g}^{(p)} \cdot \exp\left(\frac{-M_{t,i,A-1,g}^{(p)}}{2}\right) - CN_{t,i,A-1,g}^{(p)} \right] \cdot \exp\left(\frac{-M_{t,i,A-1,g}^{(p)}}{2}\right) e_{t,i,s,A-1,g}^{(p)} \\
 & + \left[N_{t,s,A,g}^{(p)} \cdot \exp\left(\frac{-M_{t,s,A,g}^{(p)}}{2}\right) - CN_{t,s,A,g}^{(p)} \right] \cdot \exp\left(\frac{-M_{t,s,A,g}^{(p)}}{2}\right) e_{t,s,s,A,g}^{(p)} \\
 & + \sum_{i \neq s} \left[N_{t,i,A,g}^{(p)} \cdot \exp\left(\frac{-M_{t,i,A,g}^{(p)}}{2}\right) - CN_{t,i,A,g}^{(p)} \right] \cdot \exp\left(\frac{-M_{t,i,A,g}^{(p)}}{2}\right) e_{t,i,s,A,g}^{(p)}
 \end{aligned}
 \tag{31}$$

For age-0 fish, the expression for $N_{t+1,s,0,g}^{(p)} = R_{t+1,s,g}^{(p)}$ depends on the recruitment submodel(s) described below.

- Area
 - Spatial connectivity through directed habitat graphs for each subpopulation along with associated adjacency and movement probability matrices.
 - Habitat or environmental effects on subpopulation processes by area.
 - Recruitment distributions of subpopulations among areas.
- Movement
 - Movement probability matrices by subpopulation, age, sex, and time period $\underline{E}_{t,a,g}^{(p)}$.
 - Subpopulation tagging processes by age, sex, and time $T_{t,a,g}^{(p)}$.

3.2.2 Observation Class

- Fisheries Data
 - Catch numbers by area and fleet.
 - Catch biomass by area and fleet.
 - Catch per unit effort by area and fleet.
 - Selectivity by area and fleet.

- Logistic fishery selectivity at age ($S_{x,a}$) by fleet (indexed by x) with age at 50% selection ($a_{x,50}$) and slope (σ_x) parameters.

$$(32) \quad S_{x,a} = \frac{1}{1 + e^{\frac{-(a-a_{x,50})}{\sigma_x}}}$$

- Thompson fishery selectivity at age ($S_{x,a}$) by fleet (indexed by x) with age at 50% selection ($a_{x,50}$), slope (σ_x), and shape (γ_x) parameters.

$$(33) \quad S_{x,a} = \frac{1}{1 - \gamma_x} \cdot \left(\frac{1 - \gamma_x}{\gamma_x} \right)^{\gamma_x} \cdot \frac{e^{\frac{-\gamma_x(a-a_{x,50})}{\sigma_x}}}{1 + e^{\frac{-(a-a_{x,50})}{\sigma_x}}}$$

- Age composition by area and fleet.
- Size composition by area and fleet.
- Sex composition by area and fleet.
- Mean size by area and fleet.
- Tag recovery information by area and fleet.
- Research Survey Data
 - Catch biomass by area and survey.
 - Catch numbers by area and survey.
 - Catch per unit effort by area and survey.
 - Selectivity by area and survey.
 - Age composition by area and survey.
 - Size composition by area and survey.
 - Sex composition by area and survey.
 - Mean size by area and survey.
 - Tag recovery information by area and survey.
- Study Data
 - Size-at-age data by area and study.
 - Maturity-at-age or –size data by area and study.
 - Fecundity-at-age or –size data by area and study.
 - Length-weight data by area and study.

3.2.3 Environment Class

- Subpopulation Effects
 - Recruitment
 - Growth
 - Natural Mortality
 - Structure

- Spatial Distribution and Movement
- Observation Effects
 - Fisheries by area.
 - Surveys by area.
- Information Effects
 - Biotic processes by area.
 - Abiotic processes by area.

3.2.4 Analysis Class

- Input and Output Objects
 - Model object.
 - Subpopulation object.
 - Observation object.
 - Environment object.
 - Model report object.
 - Model projection object.
 - Operational model object.
- Model Predictions
 - Predicted Fisheries Data

- Predicted catch biomass by fleet (x), time (t) and area (s), $\widehat{CB}_{x,t,s}$.

$$(34) \quad \widehat{CB}_{x,t,s} = \sum_p \sum_a \sum_g N_{t,s,a,g}^{(p)} w_{t,s,a,g}^{(p)} \frac{F_{x,t,s}^{(p)} S_{x,t,s,a,g}^{(p)}}{(F_{x,t,s}^{(p)} S_{x,t,s,a,g}^{(p)} + M_{t,s,a,g}^{(p)})} \left\{ 1 - \exp \left[- (F_{x,t,s}^{(p)} S_{x,t,s,a,g}^{(p)} + M_{t,s,a,g}^{(p)}) \right] \right\}$$

for fleet x in area s at time t , where $N_{t,s,a,g}^{(p)}$, $w_{t,s,a,g}^{(p)}$, and $M_{t,s,a,g}^{(p)}$ are numbers-at-age, weight-at-age, and natural mortality-at-age, respectively, for subpopulation p , area s , age a , gender g , at time t ; $F_{x,t,s}^{(p)}$ is the fully-selected fishing mortality for fleet x , subpopulation p , area s , at time t ; and $S_{x,t,s,a,g}^{(p)}$ is selectivity-at-age for fleet x , subpopulation p , area s , age a , gender g , at time t .

- Predicted catch numbers by fleet (x), area (s) and age (a), $\widehat{CN}_{x,t,s,a}$.

$$(35) \quad \widehat{CN}_{x,t,s,a} = \sum_p \sum_g N_{t,s,a,g}^{(p)} \frac{F_{x,t,s}^{(p)} S_{x,t,s,a,g}^{(p)}}{(F_{x,t,s}^{(p)} S_{x,t,s,a,g}^{(p)} + M_{t,s,a,g}^{(p)})} \left\{ 1 - \exp \left[- (F_{x,t,s}^{(p)} S_{x,t,s,a,g}^{(p)} + M_{t,s,a,g}^{(p)}) \right] \right\}$$

for fleet x in area s at time t , where $N_{t,s,a,g}^{(p)}$ and $M_{t,s,a,g}^{(p)}$ are numbers-at-age and natural mortality-at-age, respectively, for subpopulation p , area s , age a , gender g , at time t ; $F_{x,t,s}^{(p)}$ is the

fully-selected fishing mortality for fleet x , subpopulation p , area s , at time t ; and $S_{x,t,s,a,g}^{(p)}$ is selectivity-at-age for fleet x , subpopulation p , area s , age a , gender g , at time t .

- Catch per unit effort by area and fleet.
- Selectivity by area and fleet.
- Age composition by area and fleet.
- Size composition by area and fleet.
- Sex composition by area and fleet.
- Mean size by area and fleet.
- Tag recovery information by area and fleet.
- Predicted Research Survey Data
 - Catch biomass by area and survey.
 - Catch numbers by area and survey.
 - Catch per unit effort by area and survey.
 - Selectivity by area and survey.
 - Age composition by area and survey.
 - Size composition by area and survey.
 - Sex composition by area and survey.
 - Mean size by area and survey.
 - Tag recovery information by area and survey.
- Predicted Study Data
 - Size-at-age data by area and study.
 - Maturity-at-age or –size data by area and study.
 - Fecundity-at-age or –size data by area and study.
 - Length-weight data by area and study.

- Model Objective Function,

$$(36) \quad \Psi_m(\Theta | \underline{D}) = -\sum_i \log L_i(\Theta | \underline{O}_i) - \sum_j \log L_j(\Theta | \underline{O}_j) + \text{other terms}$$

- Likelihood components for relative abundance indices, $L_i(\Theta | \underline{O}_i)$.
 - Survey relative abundance index with normal observation error and annual coefficients of variation. In this case, the negative loglikelihood $(-\log L_i(\Theta | \underline{O}_i))$ for the observed survey abundance index values from the i^{th} survey by year $(O_{i,t})$, indexed by t , relative to predicted index values for the i^{th} survey by year $(P_{i,t})$ based on an additive normal error distribution with annual coefficients of variation $c_{i,t}$ is

$$(37) \quad -\log L_i(\underline{\Theta} | \underline{O}_i) = T \cdot \log(c_{i,t} P_{i,t}) + 0.5 \sum_{t=1}^T \left(\frac{O_{i,t} - P_{i,t}}{c_{i,t} P_{i,t}} \right)^2$$

- Survey relative abundance index with normal observation error and annual standard deviations. In this case, the negative loglikelihood for the observed survey abundance index values from the i^{th} survey by year $(O_{i,t})$, indexed by t , relative to predicted index values $(P_{i,t})$ based on an additive normal error distribution with annual standard deviations $\sigma_{i,t}$ is

$$(38) \quad -\log L_i(\underline{\Theta} | \underline{O}_i) = T \cdot \log(\sigma_{i,t}) + 0.5 \sum_{t=1}^T \left(\frac{O_{i,t} - P_{i,t}}{\sigma_{i,t}} \right)^2$$

- Survey relative abundance index with lognormal observation error and annual coefficients of variation. In this case, the negative loglikelihood for the observed survey abundance index values from the i^{th} survey by year $(O_{i,t})$, indexed by t , relative to predicted index values $(P_{i,t})$ based on a multiplicative lognormal error distribution with annual coefficients of variation

$$c_{i,t} = \sqrt{\exp(\sigma_{i,t}^2) - 1} \text{ and a bias correction for } E[O_{i,t}] = P_{i,t} \text{ is}$$

$$(39) \quad -\log L_i(\underline{\Theta} | \underline{O}_i) = T \cdot \log(\sigma_{i,t}) + 0.5 \sum_{t=1}^T \left(\frac{\log\left(\frac{O_{i,t}}{P_{i,t}}\right)}{\sigma_{i,t}} + 0.5 \sigma_{i,t} \right)^2$$

- Survey relative abundance index with lognormal observation error and annual coefficients of variation. In this case, the negative loglikelihood for the observed survey abundance index values from the i^{th} survey by year $(O_{i,t})$, indexed by t , relative to predicted index values $(P_{i,t})$ based on a multiplicative lognormal

error distribution with annual coefficients of variation

$$c_{i,t} = \sqrt{\exp(\sigma_{i,t}^2) - 1} \text{ and } E[\log O_{i,t}] = \log P_{i,t} \text{ is}$$

$$(40) \quad -\log L_i(\underline{\Theta} | \underline{O}_i) = T \cdot \log(\sigma_{i,t}) + 0.5 \sum_{t=1}^T \left(\frac{\log\left(\frac{O_{i,t}}{P_{i,t}}\right)}{\sigma_{i,t}} \right)^2$$

▪ Likelihood components for age or size composition data, $L_j(\underline{\Theta} | \underline{O}_j)$.

- Age composition data with multinomial error distribution and annual effective sample sizes. In this case, the negative loglikelihood $\left(-\log L_j(\underline{\Theta} | \underline{O}_j)\right)$ for the observed bin values, indexed by b from the j^{th} composition data set with $n_{bin,j}$ bins by year $(O_{j,b,t})$ relative to predicted bin values $(P_{j,b,t})$ based on a multinomial error distribution with annual effective sample sizes $N_{j,t}$ is

$$(41) \quad -\log L_j(\underline{\Theta} | \underline{O}_j) = -\sum_{t=1}^T N_{j,t} \sum_{b=1}^{n_{bin,j}} O_{j,b,t} \log E_{j,b,t}$$

- Age composition data with a robust multivariate normal error distribution and annual effective sample sizes. In this case, the negative loglikelihood for the observed bin values, indexed by b from the j^{th} composition data set with $n_{bin,j}$ bins by year $(O_{j,b,t})$ relative to predicted bin values $(P_{j,b,t})$ based on a robust multivariate normal error distribution (Fournier et al. 1990, Starr et al. 1999) with annual effective sample sizes $N_{j,t}$ is

$$(42) \quad -\log L_j(\underline{\Theta} | \underline{O}_j) = 0.5 \sum_{t=1}^T \sum_{b=1}^{n_{bin,j}} \log\left((1 - O_{j,b,t}) O_{j,b,t} + 0.1 / n_{bin,j}\right) \\ - \sum_{t=1}^T \sum_{b=1}^{n_{bin,j}} \log\left(\exp\left[\frac{-(O_{j,b,t} - P_{j,b,t})^2}{2((1 - O_{j,b,t}) O_{j,b,t} + 0.1 / n_{bin,j}) / N_{j,t}}\right] + 0.01\right)$$

- Priors, $\pi_k(\underline{\Theta}_\pi)$ or $\pi_k(\underline{\Theta}_\pi | \underline{\Theta}_\varphi)$.
- Hyperpriors, $\varphi_k(\underline{\Theta}_\varphi)$.
- Numerical constraints, $\Upsilon_c(\underline{\mathcal{S}} | \underline{\Theta})$.
- Nuisance parameters, $\Upsilon_\eta(\underline{\mathcal{S}} | \underline{\Theta})$.
- Model Diagnostics and Selection
 - Convergence criteria.
 - Iterative reweighting.
 - Residual analyses.
 - Likelihood profiles.
 - Residual analyses.
 - Parameter correlations.
 - Goodness-of-fit measures.
 - Leverages.
 - Cross validation.
 - Information theoretic measures.
 - Bayesian P-values.
 - Other model diagnostics.
- Model Results
 - Biomass time series.
 - Biomass reference points.
 - Harvest rate time series.
 - Harvest rate reference points.
 - Other quantities of interest.
- Uncertainty Characterization
 - Bootstrapping.
 - Markov Chain Monte Carlo simulation.
 - Covariance matrix.
 - Jackknifing.
 - Probability of having exceeded the OFL as a function of ACL.
 - Model averaging.
 - Other measures of uncertainty characterization.
- Model Projection Types
 - Stochastic or deterministic.
 - Single model or ensemble of models.
 - Catch-based or effort-based harvest.
 - With or without control rule.
 - With or without implementation error.
 - Probability of exceeding OFL under alternative harvests.

- Other projected quantities of interest.
- Management Strategy Evaluation
 - Operating model.
 - True parameter values.
 - Simulated subsets of parameters.
 - Covariance.
 - Bias.
 - Performance measures.
 - Harvest control rules.
 - Management procedures.

3.3 Design Rationale

The MAS class structures were conceived to be flexible enough to produce what we now want or imagine we may want in the future from a spatially-explicit assessment system. The design choice of beginning with the end in mind is an important part of this approach, noting that we have grouped potential features into related categories to facilitate structured system development. Overall, we note that the suggested categorizations of features were intended to facilitate, but not rigidly define, possible model structures and capacities for understanding and prediction.

One goal that we want to achieve with the MAS is the implementation of a well-documented, high quality, reliable, and easily extensible programming system. We chose to implement MAS in C++ to emphasize an object-oriented programming style. This choice will facilitate the maintenance, extensibility, and reusability of code and model structures, and most importantly will facilitate rapid model prototyping and testing. Another key structural component for efficient and reliable coding is the use of input and output structures, where these I/O objects can be tailored to specific needs of specific stock assessment applications using JSON. Overall, the programming design of MAS is integrated with a living design document that includes necessary and sufficient information to define and reconstruct MAS and also to extend it with new modules through stepwise quality assurance and quality control procedures.

The MAS model structure is designed to reflect metapopulation dynamics. Multiple population habitats and areas are part of the system and can be used to represent spatial dynamics for ocean zoning and fisheries management. MAS also has the capacity for frequentist and Bayesian population dynamics processes for model fitting to provide adequate generality and flexibility. In general, it is preferable to employ Bayesian approaches to address parameter uncertainty in a direct manner but it is very important to have the option to avoid subjective prior assumptions in a frequentist context. The structural issue of the mode of inference is also linked to the choice of error structure. Error terms can be broadly dichotomized as being additive and normally

distributed or as being multiplicative and lognormally distributed. Alternative distributional forms for additive errors are the T-distribution and Cauchy distribution and alternative distributional forms for the multiplicative are the gamma and exponential distributions. It is also expected that the MAS needs to provide the capacity to fit both near-symmetric and skewed error structures to be a flexible platform for assessment modeling. Following on this point, there needs to be clear output information for model diagnostics and selection among competing alternative models. MAS also includes options for data weighting, to emphasize or deemphasize information, as well easy to implement data inclusion or exclusion in alternative models. Last, the MAS also needs to provide standard assessment quantities of interest for fishery management, including biological reference points, status determination information, and forecasts.

4. DATA DESIGN

4.1 Data Description

The input data and model structure and analysis information needed to define a particular model are stored in a model object file, which records the analysis, observation, environment, and subpopulation information specific to that model instance. The input data for the model comprise part of the model object file. Input data for MAS consists of the data types described in the observation class (section 3.2.2) and in the environment class (section 3.2.3). It is recommended that the information needed to access the entire set of available input data be included in the model object file to facilitate efficient analyses using subsets of the entire data in alternative model configurations. Each of the input data types needs to be assigned a unique acronym. The unique acronym can also serve as a keyword within the input data file that designates exactly where the data type starts and ends and also where individual instances of the data type start and end. Input information for the analysis and subpopulation classes is also stored in the model object file in a similar manner. By convention, instances of an input data type are contiguous and are listed sequentially through the assessment time horizon.

4.1.1 Subpopulation Class Data

This section describes data structures and variable naming conventions for the subpopulation class data. Under development.

4.1.2 Observation Class Data

This section describes data structures and variable naming conventions for the observation class data. Under development.

4.1.3 Environment Class Data

This section describes data structures and variable naming conventions for the environment class data. Under development.

4.1.4 Analysis Class Data

This section describes data structures and variable naming conventions for the analysis class data. Under development.

4.2 Model Metadata

Each model object file contains sufficient metadata about the structure of itself to reconstruct itself. This metadata includes the unique acronym and location of each data type in the model object file. Most importantly, the metadata describes the entire set of available input data that can be used in the given model. Under development.

4.2.1 Subpopulation Class Metadata

This section describes the metadata needed to specify the subpopulation class for each of the three model analysis layers: (i) model construction; (ii) model set selection; (iii) model projections. Under development.

4.2.2 Observation Class Metadata

This section describes the metadata needed to specify the observation class for each of the three model analysis layers: (i) model construction; (ii) model set selection; (iii) model projections. Under development.

4.2.3 Environment Class Metadata

This section describes the metadata needed to specify the environment class for each of the three model analysis layers: (i) model construction; (ii) model set selection; (iii) model projections. Under development.

4.2.4 Analysis Class Metadata

This section describes the metadata needed to specify the analysis class for each of the three model analysis layers: (i) model construction; (ii) model set selection; (iii) model projections. Under development.

5. MODEL ANALYSES

5.1 Model Construction

The model construction specification consists of a unique sequence of lists, or case structures, that specify the components and relationships of subpopulation, observation, environment, and analysis classes. These model case structures are defined for the Model Class, as well as the Subpopulation, Observation, Environment, and Analysis Classes used in a given model

5.1.1 Input Data

Input data file constructed with a graphical user interface or user-constructed file comprised of a model construction header and input data. The model construction header specifies where data to fill out each model construct are found in the subsequent input data section. The input data section reflects the structure specified in the data design.

5.1.2 Model Structure

The model structure specifies the components and timing of the processes that change or sample biomass in an area during a time step. The time step could be a year or it could be a season within a year. In this context, processes that increase biomass are somatic growth, recruitment, and immigration and processes that decrease biomass are natural deaths, fishery yields, and emigration. Spawning is another process needed to calculate the reproductive output of the subpopulation while tagging is a process that may provide information on movement and harvest intensity.

5.1.3 Model Parameters

This is the unique list that maps the set of parameters to be estimated to the likelihood components in the specified in the information class and encapsulated in the model construct header.

5.1.4 Parameter Estimation

Parameter estimation is based on maximum likelihood-based, simple Bayes, hierarchical Bayes, random effects or other estimation approaches. The maximum likelihood-based estimation approach requires specification of the likelihood components of the model and maximization of the joint likelihood. The simple Bayes approach requires specification of the likelihood components and parameter priors of the model and numerical sampling of the posterior distribution. The hierarchical Bayes approach requires specification of the likelihood components, the parameter priors, and any hyperpriors for the model and numerical sampling of the posterior distribution. The random effects approach requires specification of the marginal likelihood and its maximization.

Parameter optimization is achieved through application of an objective function minimizer or through numerical integration.

5.1.5 Model Outputs

Model output consists of information on model convergence and fits to the data, model diagnostics, and quantities of interest. The quantities of interest will typically include estimates of metapopulation size and stock status and conservation information.

5.1.6 Additional Model Construction Components

- Data Weighting –observe that Francis (2011) provides an initial set of alternative approaches to data weighting
- Time-Varying Parameters – use Methot and Wetzel (2013) for an initial set of approaches to time-varying parameters
- Multivariate Priors
- Randomization of Initial Parameter Values
- Randomization of Model Input Data
- Randomization of Model Structure

5.2 Model Quality

Evaluating the robustness and reliability of the models used to understand and predict the dynamics of a metapopulation is important, noting that such evaluations will be ongoing process. Therefore, we want to build models that can readily be tested in a simulation framework. In the ideal situation, quality assurance and quality control for metapopulation modeling would be achieved through an iterative process of model verification, validation, and uncertainty quantification (Figure 8). This quality assurance is important for evaluating the reliability of complex models (NRC 2012).

5.2.1 Model Verification

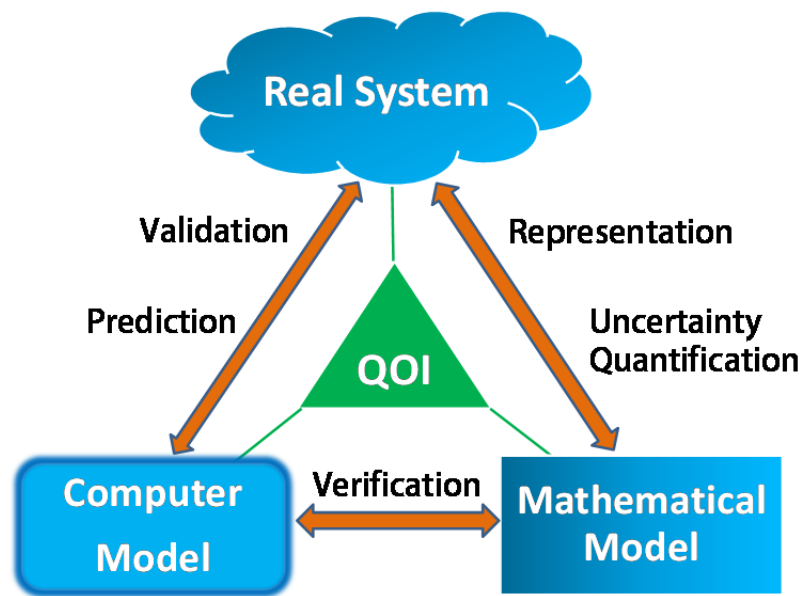
Clearly define the quantities of interest for model verification. Identify the algorithmic structures to be tested. Code to alternative code testing is useful but does not constitute solution verification. This is best done using an analytic solution for a well-posed problem.

5.2.2 Validation, Prediction, and Uncertainty Quantification

Specify the quantities of interest before starting the validation process. What are the important quantities to be well-estimated? Use a broad range of possible inputs to quantify and bound the error from a model application. Use cross validation and other predictive accuracy checks for characterizing model validity and predictive performance. Use physical observations to constrain uncertainties in model inputs. Document assumptions about the sources of uncertainty used to characterize uncertainty in the quantities of interest. Use replication to estimate variability and measurement uncertainty.

Figure 8. Validation, verification, and uncertainty quantification processes for modeling ecosystem impacts on quantities of interest.

Validation, Verification, and Uncertainty Quantification for Quantities of Interest



5.3 Model Set Selection

In this section, a general multimodel inference approach (e.g., Burnham and Anderson 2002) for implementation in MAS is described. Structural uncertainty in approximating models used for inference about metapopulations is an important feature of the incomplete knowledge of complex marine system dynamics. Structural uncertainty, when combined with natural variability and observational error, makes it challenging to provide the understanding and predictive information needed to manage human impacts (Peterman 2004, Link et al. 2012). The multimodel inference approach in MAS uses a state space modeling formulation (e.g., Aoki 1990, Schnute 1994) to represent alternative hypotheses with Markov process dynamics. A frequentist (Burnham and Anderson 2002) or Bayesian estimation framework (Punt and Hilborn 1997, Ellison 2004) can be used to assess the evidence for alternative models, which are formulated as either simple or hierarchically-structured models (Clark 2005) with time as a hierarchical dimension. Information theoretic model selection criteria are used to weight the probability that each candidate model provides the best approximation and predictive information about the system state, given the observed fixed data. Model averaging can be used when the weight of evidence to support a single operating model is ambiguous (Hoeting et al. 1999, Burnham and Anderson 2002), else

the results of a single best operating model are identified and directly applied. This approach can account for the probable range of uncertainty in model structure while both incorporating relevant prior information and providing probabilistic interpretations of results for risk assessment.

First, a general description of the state space model formulation is provided. Next, the multimodel inference approach and the development of a set of candidate models is discussed, including the calculation of model likelihoods and weights, the multimodel inference process, and model averaging when appropriate. Alternative hypotheses can be formulated as either simple or hierarchical Bayes models and fit the alternative models to the data. Model selection and multimodel inference is used to judge the adequacy of the models, including formulating model-averaged results.

The approach we use for model selection and multimodel inference is conceptually based on ideas expressed in Burnham and Anderson (2002). The multimodel inference approach starts with a set of candidate models M_j , denoted by $\underline{M} = \{M_j\}$, that represent the set of plausible hypotheses about metapopulation dynamics. Here the emphasis is on applying “the method of multiple working hypotheses” and putting forward candidate models that have an empirical basis and are grounded in scientific plausibility (e.g., Hilborn and Mangel 1997). In this context, there should be an emphasis on both keeping an open mind about the set of plausible hypotheses and not seeking to derive a dichotomous null versus alternative hypothesis in order to simplify the model selection process. Burnham and Anderson (2002) advocate limiting the set of candidate models for parsimony and this is an important point to reduce the possibility of chance results due to sampling variability. That is, when more models are included in the set of candidate models with a fixed sample size, then the power to detect important differences between models decreases as the relative sample size per model decreases. Further, it is also important to avoid including models that are functionally redundant.

The multimodel inference approach is designed to address the important issue of model selection uncertainty. A key feature of the multimodel inference approach is to avoid the risk of making decisions based on overconfidence and overinterpretation of modeling results due to the underestimation of uncertainty and the a priori acceptance of a single best hypothesis without critical examination. Because our understanding of complex fishery systems is and will always be incomplete, there needs to be an ongoing emphasis on confronting our alternative hypotheses about ecosystem dynamics with observed data (Hilborn and Mangel 1997). In this context, we emphasize a tiered approach to the evaluation of each candidate model as being judged to be a credible model that is appropriate for inference in fishery management application. First and foremost, candidate models have to satisfy convergence requirements and statistical optimization criteria, this is a minimal requirement. Second, standard model goodness-of-fit diagnostics need to provide evidence that each candidate model conforms to assumptions. Last, the selected credible models among the set of candidate models should have quantitative support based on

information theoretic criteria that provide an estimate of model distance from the unknowable true state of nature (Burnham and Anderson 2002). In the case of one selected, best-fitting credible metapopulation model, inference and scientific information for management will be based on that single model.

Alternatively, when more than one credible metapopulation model exists, we apply model averaging to account for model selection uncertainty. In this application, we used the Deviance Information Criterion (DIC , Spiegelhalter et al. 2002) as a plug-in measure of the information-theoretic goodness of fit of the alternative candidate models. We also note that alternative information theoretic criteria are available and that some of these criteria may provide better approximations of the relative goodness of fit among alternative models than DIC in some cases. Regardless, we note that the DIC values used in this study were empirically calculated based on the information generated by the convergent Markov Chain Monte Carlo (MCMC) simulations conducted for each model.

To provide some background on the logical basis for DIC as a measure of goodness of fit or conversely the amount of discrepancy between the candidate model and the data (Burnham and Anderson 2002, Gelman et al. 2005), observe that the model deviance $Dev(\underline{D}, \hat{\underline{\Theta}})$ is typically defined as -2 times the log-likelihood for the data \underline{D} conditioned on the estimated parameters $\hat{\underline{\Theta}}$

$$(7) \quad Dev(\underline{D}, \hat{\underline{\Theta}}) = -2 \cdot \log(L(\underline{D} | \hat{\underline{\Theta}}))$$

The expected deviance, computed by averaging $Dev(\underline{D}, \hat{\underline{\Theta}})$ over the true but unknown sampling distribution is equal to 2 times the Kullback-Leibler information up to a constant that does not depend on the estimated parameters $\hat{\underline{\Theta}}$. This relation to the Kullback-Leibler information implies that the parameters that produce the lowest expected deviance will produce the maximum information and have the highest posterior probability. However, the true sampling distribution is not known and therefore one needs an accurate way to estimate the expected deviance. To this end, note that the discrepancy between the model and the data depends on both data \underline{D} and the set of parameters being estimated in the bootstrapping or MCMC simulation process $\hat{\underline{\Theta}}$. To get a deviance measure that depends only on the data, one can approximate the model deviance value $\overline{Dev}(\underline{D})$ conditioned on a mean point estimate of $\underline{\Theta}$

$$(43) \quad \overline{Dev}(\underline{D}) = \overline{Dev}(\underline{D}, \hat{\underline{\Theta}}(\underline{D}))$$

where the point estimate $\hat{\underline{\Theta}}$ was the mean of the posterior MCMC simulations for $\underline{\Theta}$. That is,

$$(44) \quad \hat{\underline{\Theta}}(\underline{D}) = E[\underline{\Theta}|\underline{D}] = \frac{1}{J} \sum_{j=1}^J \underline{\Theta}^{(j)}$$

where $\underline{\Theta}^{(j)}$ is the jth iterate of $\underline{\Theta}$ in a total of J posterior simulations. This is one estimate of the expected deviance for a fixed point estimate of $\underline{\Theta}$ calculated from the posterior simulations.

Another natural approach to estimating the expected deviance would be to use the expected value of \widehat{Dev} , the estimate of model discrepancy, calculated over the posterior distribution as our estimate of the model deviance

$$(45) \quad \widehat{Dev}(\underline{D}) = E[Dev(\underline{D}, \underline{\Theta})|\underline{D}]$$

Of course, one does not have complete knowledge or information about the true posterior distribution and so an analytic calculation of the integral for the expected value is not generally possible except for simple problems. However one can use the natural plug-in estimate of the expected deviance by taking the average of $Dev(\underline{D}, \underline{\Theta})$ over the posterior simulations to be \widehat{Dev} where

$$(46) \quad \widehat{Dev}(\underline{D}) = \frac{1}{J} \sum_{j=1}^J Dev(\underline{D}, \underline{\Theta}^{(j)})$$

Given this background, we observe that the difference between the posterior mean of the deviance (\widehat{Dev}) minus the deviance evaluated at the posterior mean of the stochastic parameters being estimated (\overline{Dev}) provides a measure of the effect of model fitting and can be used as a measure of the effective degrees of freedom in the model (ρ_D) as argued by Spiegelhalter et al. (2002) under the assumption that the posterior distribution is asymptotically a multivariate normal distribution as

$$(47) \quad \rho_D = \widehat{Dev} - \overline{Dev}$$

The value of DIC was then calculated as twice the posterior mean of the deviance minus the deviance evaluated at the posterior mean of the stochastic nodes

$$(48) \quad DIC = 2\widehat{Dev} - \overline{Dev} = \widehat{Dev} + \rho_D = \overline{Dev} + 2\rho_D$$

The alternative candidate models, indexed by j (M_j), can then be ranked by their DIC values (DIC_j). The best fitting model (M^*) with the minimum value of DIC (DIC_{Min}) produced the

best fit to the observed data. For each model, the difference in the model's value of DIC from the minimum was calculated (Δ_j) as

$$(49) \quad \Delta_j = DIC_j - DIC_{Min}$$

Typically models with values of Δ_j less than 2 have similar support and can be included in the set of credible candidate models that have similar goodness of fit to the observed data. In contrast, models with Δ_k greater than 3 differ from the best-fitting model and have limited support (Spiegelhalter et al. 2002). In this context, evidence indicated that the credible models also provided adequate fits to the observed data and should be considered as viable alternative states of nature in comparison to the best fitting model. The results from the set of credible models, denoted by \underline{M}^* , were model-averaged based on the likelihood $L_j(\underline{D}|\underline{\Theta})$ of each model $M_j \in \underline{M}^*$. In this context, the model likelihood was proportional to the exponential of the Δ_j value through the expression

$$(50) \quad L_j(\underline{D}|\underline{\Theta}) \propto e^{-0.5\Delta_j}$$

One can apply model averaging (MA) to the set of credible models using the DIC -based approximation of the candidate model likelihood. This produced a model-averaged set of results which were used for probabilistic inference about stock status determination through time and were also available for as a distribution for future population projections under alternative harvest patterns and states of nature.

To apply MA, prior model probabilities (π_k) were developed to express the relative belief that each of the credible models represented the true state of nature. One can adopt an objective approach for setting the prior model probabilities in the absence of any information for preferring one credible model over another (Berger 2006). In this case, the prior model probabilities are equal based on the principle of indifference, where for a total of m credible models

$$(51) \quad \pi_j = \pi_k = \frac{1}{m} \text{ and } \sum_j \pi_j = 1$$

Given the integrated likelihoods of the credible models and the prior model probabilities, we calculated the posterior model probabilities (W_j) over the set of credible models as

$$(52) \quad W_j = \frac{\pi_j \cdot e^{-0.5\Delta_j}}{\sum_i \pi_i \cdot e^{-0.5\Delta_i}}$$

The posterior model probability quantified the relative support contained in the sample data and the assumed prior model probabilities. Because we did not have any evidence to assume one model was a priori more likely than another, the posterior model probabilities were in effect based on the observed data. In this context, this approach to MA analyses can be viewed as an objective Bayesian approach (Berger 2006).

The posterior model probabilities provide the essential model weighting information for MA of the conditional model distributions for any resulting individual state space model parameter or for any derived model output result ϕ . The model-averaged estimate of a model result ϕ is denoted as $\bar{\phi}$ and its distribution depends on the expected estimates $\bar{\phi}_j$ from each candidate model indexed by j . The expected value of the model-averaged estimate of ϕ averaged over the set of credible models is

$$(53) \quad \bar{\phi} = E[\bar{\phi}] = \sum_j W_j \bar{\phi}_j$$

(e.g., Buckland et al. 1997, Burnham and Anderson 2002) and the variance of the model-averaged estimate is

$$(54) \quad \text{VAR}[\bar{\phi}] = \left[\sum_j W_j \sqrt{\text{VAR}[\phi_j] + (\bar{\phi}_j - E[\bar{\phi}])^2} \right]^2$$

The variance of the model-averaged estimation result includes two components, the first is the variance of the individual model estimates and the second is an expression for the variance contribution of model uncertainty in the point estimate of the result ϕ . Given these two fundamental quantities for multimodel inference, one can use the mean and the unconditional variance estimate of the model-averaged result for risk analyses. Alternatively, other measures of model goodness of fit may be applied to subsets of the assessment data when there are differences among the data sets used in alternative models.

5.4 Model Projections

The key features of model projections are listed below.

- Projection Domain Parameters
 - Time Frame
 - Set of Projection Models
 - Population Dynamics
 - Initial Conditions
 - Deterministic Components

- Stochastic Components
- Fishery Dynamics
 - Initial Conditions
 - Deterministic Components
 - Stochastic Components
- Harvest Control Rule or Management Procedure
 - Initial Conditions
 - Deterministic Components
 - Stochastic Components
- Projection Outputs
 - Distributions of Quantities of Interest
 - Individual Forecast Models
 - Combining Forecast Models
 - Measures of Central Tendency of QOI
 - Mean
 - Arithmetic
 - Trimmed
 - Harmonic
 - Geometric
 - Median
 - Measures of Dispersion of QOI
 - Variance, Standard Deviation, CV
 - Interquartile or Other Percentile Range
 - Confidence or Credible Intervals
 - Measures of Bias of QOI
 - Retrospective Patterns
 - Mohn Rho (Mohn 1999)
- *Additional Projection Components*
 - *Find FMSY*
 - *Calculate FREBUILD*
 - *Calculate Probability of Overfishing as a Function of ACL*
 - *Calculate Probability of Depletion as a Function of ACL*
 - *Calculate Probability of Achieving Target F as a Function of ACL*
 - *Calculate Probability of Achieving Target Biomass as a Function of ACL*
 - *Calculations for Multiyear ACLs*
 - *See AGEPRO USERS GUIDE for further information*

6. HUMAN INTERFACE DESIGN

6.1 Overview of User Interface

The key features of the user interface are listed below.

- Lists of Objects
 - Individual Models
 - Set of Models
 - Templates for Model Component Configurations
 - Subpopulation Class
 - e.g., Movement Matrices, Life History Parameters
 - Observation Class
 - Environment Class
 - Analysis Class
 - Quantities of Interest
 - Central Tendency
 - Dispersion
 - Time Series of Distributions
- Lists of Actions
 - Create Model
 - Get Model
 - Copy Model
 - Modify Model
 - Save Model
 - Delete Model
 - Evaluate Model(s) Fit to Data
 - Change Data for Model(s)
 - Compare Models
- Lists of Tools
 - Read Existing Model Tool
 - Construct New Model Tool
 - Model Set Construction Tool, Variations on a Theme
 - Using Different Data
 - Using Different Structure
 - Using Different Data and Structure
 - Random Model Set Construction Tool
 - Model Document Generation Tool
 - Text
 - Tables
 - Figures
 - Appendices

6.2 Screen Objects and Actions

- Single or Multiple Windows
- Model Property Sheet
 - Nested Pull Down Menus
- Action Panel
 - Model Definition GUI Interface
 - Nested Pull Down Lists
- Visualization and Graphics
 - Model Inputs
 - Model Outputs
 - Model Diagnostics
 - Model Projections

7. REQUIREMENTS AND SYSTEM MAINTENANCE

7.1 Requirements

The key system requirements are listed below.

- Optimization Routine Library
- MCMC Simulation Routine Library
- Bootstrap Routine Library
- Probability Distribution Library
- Negative Loglikelihood Function Library
- Growth Function Library
- Recruitment Function Library
- Maturity Function Library
- Fecundity Function Library
- Movement Function Library
- Input Data Function Library
- Output Function Library
- Prior Distribution Library
- Hyperprior Distribution Library
- Operating Model Template Library

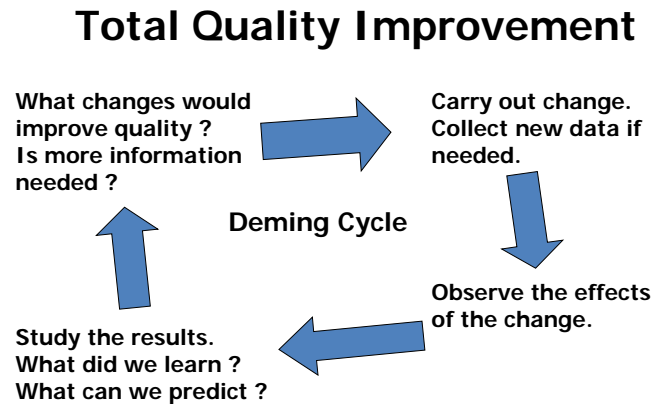
7.2 System Maintenance

The key system maintenance requirements are listed below.

- Standard Operating Protocol for Testing Models.

- Standard Operating Protocol for Total Quality Improvement (Figure 9).
General principles of system maintenance through a continuous evaluation cycle from Deming include:
 - (i) How can the system be improved? What change would improve quality ?
 - (ii) Make the change and observe the results through simulation or sampling.
 - (iii) Are the results better? If yes, implement the change. If no, then identify why the results were not better.
 - (iv) Repeat.

Figure 9. Schematic of the Deming cycle for total quality improvement.



8. REFERENCES

- Ando, T. 2007. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika* 94:443-458.
- Berger, J. 2006. The case for objective Bayesian analysis. *Bayesian analysis*, 3:385-402.
- Brooks, S. P., and Gelman, A. 1998. Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434-455.
- Bull, B., Francis R. I. C. C., Dunn, A., McKenzie, A., Gilbert, D. J., and Smith, M. H. 2005. CASAL (C++ algorithmic stock assessment laboratory): CASAL user manual v2.07-2005/08/21. 2005. NIWA Technical Report 127.
- Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference: A practical information theoretic approach, 2nd ed. Springer, New York, 488 p.
- Clark, J.S. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* 8:2-14.
- Ellis, A. 2004. Bayesian inference in ecology. *Ecology Letters* 7:509-520.
- Fournier, D., Sibert, J., Majkowski, J., and Hampton, J. 1990. MULTIFAN, a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data for southern bluefin tuna (*Thunnus maccoyii*). *Can J. Fish. Aquat. Sci.* 47(2): 301-317.
- Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci.* 68:1124-1138.
- Gelman, A., and Rubin, D. 1992. Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7: 457-511.
- Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics*. Edited by J. Bernardo, J. Berger, A. Dawid, and A. Smith. Vol. 4, Clarendon Press, Oxford, U.K. pp. 169-194.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D.J. [Eds.] 1996. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London. 486 p.
- Goethel, D. Quinn II., T, Cadrin, S. 2011. Incorporating spatial structure in stock assessment: Movement modeling in marine fish population dynamics. *Reviews in Fisheries Science*, 19:119-136.
- Heidelberger, P. and Welch, P. 1992. Simulation run length control in the presence of an initial transient. *Op. Res.* 31: 1109-1144.

- Hilborn, R. and Mangel, M. 1997. The ecological detective. Princeton University Press, Princeton, New Jersey, 315 p.
- Hill, S., Watters, G., Punt, A., McAllister, M., Le Quere, C., and Turner, J. 2007. Model uncertainty in the ecosystem approach to fisheries. *Fish and Fisheries* 8: 315-336.
- Hoeting, J.A., Madigan, D., Raftery, A., and Volinsky, C.T. 1999. Bayesian model averaging: A tutorial. *Statistical Science* 14:382-417.
- Hollowed, A., Barange, M., Beamish, R., Brander, K., Cochrane, K., Drinkwater, K., Foreman, M., Hare, J., Holt, J., Ito, S., Kim, S., King, J., Loeng, H., MacKenzie, R., Mueter, F., Okey, T., Peck, M., Radchenko, V., Rice, R., Schrippa, M., Yatsu, A., and Yamanaka, Y. 2013. Projected impacts of climate change on marine fish and fisheries. *ICES J. Mar. Sci.* 70:1023-1037.
- Jana, D. 2005. C++ and the object-oriented programming paradigm, 2nd Edition. PHI Learning, 548 p.
- Jiao, Y., Cortes, E., Andrews, K., and Guo, F. 2011. Poor-data and data-poor species stock assessment using a Bayesian hierarchical approach. *Ecol. Appl.* 21:2691-2708.
- Kadande, J., and Lazar, N. 2004. Methods and criteria for model selection. *Journal of the American Statistical Association* 99:279-290.
- Levin, S. 1992. The problem of pattern and scale in ecology. *Ecology*, 73:1943-1967.
- Link, J.S., Ihde, T.F., Harvey, C.J., Gaichas, S.K., Field, J.C., Brodziak, J.K.T., Townsend, H.M., and Peterman, R.M. 2012. Dealing with uncertainty in ecosystem models: The paradox of use for living marine resource management. *Progress in Oceanography*.
- McAllister, M. Starr, P., Restrepo, V., and Kirkwood, G. 1999. Formulating quantitative methods to evaluate fishery-management systems: What fishery processes should be modeled and what trade-offs should be made? *ICES J. Mar. Sci.* 56:900-916.
- Methot, R.D., and Wetzel, C.R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fish. Res.* 142:86-99.
- Morgan, M.G., and Henrion, M. 1990. Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, 332 p.
- National Research Council [NRC]. 2012. Assessing the reliability of complex models: Mathematical and statistical foundations of verification, validation, and uncertainty quantification. National Academies Press, 131 p.
- R Development Core Team. 2008. R: a language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Raftery, A., and Zheng, Y. 2003. Discussion: Performance of Bayesian model averaging. *Journal of the American Statistical Association* 98:931-938.

- Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., Punt, A., Restrepo, V., Skagen, D.W., and Stefansson, G. 2001. Estimating uncertainty in fish stock assessment and forecasting. *Fish and Fisheries* 2:125-157.
- Peterman, R.M. 2004. Possible solutions to some challenges facing fisheries scientists and managers. *ICES J. Mar. Sci.* 61:1331-1343.
- Plummer, M., Best, N., Cowles, K., and Vines, K. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*. 6: 7-11.
- Punt, A.E., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. *Reviews in Fish Biology and Fisheries* 7:35-63.
- Schnute, J. (1981). A versatile growth model with statistically stable parameters. *Can. J. Fish. Aquat. Sci.* 38:1128-1140.
- Schnute, J.T. 1994. A general framework for developing sequential fisheries models. *Can. J. Fish. Aquat. Sci.* 51:1676-1688.
- Shepherd, J.G. 1999. Extended survivors analysis: An improved method for the analysis of catch-at-age data and abundance indices. *ICES J. Mar. Sci.* 56:584-591.
- Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. 2002. Bayesian measures of model complexity and fit. *J. R. Statist. Soc. B*, 64:583-639.
- Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. 2003. *WinBUGS User Manual*. Available at: <http://www.mrc.bsu.carn.ac.uk/bugs/winbugs/manual14.pdf>
- Spies, I., Spencer, P., and Punt, A. 2015. Where do we draw the line? A simulation approach for evaluating management of marine fish stocks with isolation-by-distance stock structure. *Can. J. Fish. Aquat. Sci.* 72:968-982.
- Starr, P., Bentley, N., and Maunder, M. 1999. Assessment of the NSN and NSS stocks of red rock lobster (*Jasus edwardsii*) for 1998. New Zealand Fisheries Assessment Document 99/34. Ministry of Fisheries, Wellington, New Zealand.
- Stroustrup, B. 2013. The C++ programming language, 4th Edition. Pearson Education, 1347 p.
- Subbey, S., Devine, J., Schaarschmidt, U., and Nash, R. 2014. Modelling and forecasting stock-recruitment: current and future perspectives. *ICES J. Mar. Sci.* doi:10.1093/icesjms/fsu148.
- Vert-pre, K.A., Amoroso, R.O., Jensen, O.P., and Hilborn, R. 2013. Frequency and intensity of productivity regime shifts in marine fish stocks. *Proc. Natl. Acad. Sci. U.S.A.* 110(5): 1779-1784.

The Western and Central Pacific Fisheries Commission [WCPFC]. 2011. Scientific Committee. Regular session (7th : 2011 : Pohnpei, Federated States of Micronesia) Seventh regular session, Pohnpei, Federated States of Micronesia, 9-17 August 2011 : summary report, 165 p.

Ying, Y., Chen, Y., Lin, L., and Gao, T. 2011. Risks of ignoring fish population spatial structure in fisheries management. Can. J. Fish. Aquat. Sci. 68:2101-2120.

WCPFC. 2012. Scientific Committee. Regular session (8th : 2012 : Busan, Korea) Eighth regular session, Busan, Korea, 7-15 August 2012 : summary report, 163 p.

Zhang, Z., and Perry, R.I. 2005. Use of state-space modeling with a Bayesian approach to estimate target reference points for green sea urchin (*Strongylocentrotus droebachiensis*) stocks in the Queen Charlotte Strait region, British Columbia, Canada. Fish. Res. 74:253-264.

APPENDIX

Figure A1. White board listing some MAS features on 14-March-2014.



Figure A2. White board showing some MAS notation and equations on 25-September-2015.

