

# Technical Documentation for the integrated statistical catch age model



Steven J. D. Martell and Catarina Wor Lima  
International Pacific Halibut Commission  
2320 West Commodore Way,  
Seattle, WA  
98199-1287  
USA  
`stevenm@iphc.int`

July 1, 2015



# Contents

Contents . . . . .	2
<b>1 Introduction</b>	<b>2</b>
<b>2 Model Description</b>	<b>3</b>
2.1 General overview . . . . .	3
2.2 List of symbols . . . . .	3
2.3 Input data . . . . .	5
2.3.1 Model dimensions . . . . .	5
2.3.2 Age-schedule information . . . . .	5
2.3.3 Historical removals . . . . .	6
2.3.4 Relative abundance . . . . .	6
2.3.5 Composition information . . . . .	6
2.3.6 Empirical weight-at-age data . . . . .	6
2.4 Analytic methods . . . . .	6
2.4.1 Deterministic age-structured model . . . . .	6
2.4.2 Stochastic age-structured model . . . . .	7
2.4.3 Fisheries selectivity . . . . .	11
2.4.4 Natural mortality . . . . .	11
2.4.5 Options for selectivity . . . . .	11
2.5 Residuals, likelihoods & objective function value components . . . . .	12
2.5.1 Catch data . . . . .	13
2.5.2 Relative abundance data . . . . .	13
2.5.3 Relative abundance data with time-varying catchability . . . . .	14
2.5.4 Age composition data . . . . .	15
2.5.5 Stock-recruitment . . . . .	16
<b>References</b>	<b>18</b>

## 1 Introduction

This document provides the technical details and model description for a generalized Integrated Statistical Catch Age Model (ISCAM). The model is largely based on the original work of (Fournier and Archibald, 1982) with a number of options for alternate model specification. The code is implemented in AD Model Builder and the source code is available at <https://github.com/smartell/iSCAM>.



## 2 Model Description

### 2.1 General overview

As an overview of the model description, there are four major components to this integrated assessment model:

- Input data,
- analytical methods for fishery and stock dynamics,
- statistical criterion for estimating model parameters,
- and model outputs including reference points and decision tables.

First, this section provides a list of symbols use to annotate various model compoents. It then goes onto to describe the anaytical methods, followed by the statistical components (negative loglikelihoods, penalties or prior densities). Lastly, outputs based on maximum likelihood estimates or posterior samples are then presented. These outputs include reference points, and quantities that are necessary in constructing a decision table.

All of the model equations are laid out in tables and are intended to represent the order of operations, or pseudocode, in which to implement the model. The assessment model is implemented in the C++ programming language and uses AD Model Builder version 11.1 (Fournier et al., 2011) as the primary tool for non-linear parameter estimation and numerically integrating the joint posterior distribution.

The general dimensions and structure of the population dynamics and fishery dynamics sub routines are also adopted for the operating model. The operating model is written as a separate library, and has it's own independent set of input files. By design, this organization of the models allows the operating model to have completely different structural assumptions than the assessment model.

Lastly, we note that the acronym for the Integrated Statistical Catch At Age Model (or ISCAM) does not instill a lot of confidence among industry participants. Throughout this document, we occasionally refer to the model as ISCAM and we are still searching for a more appropriate name.

### 2.2 List of symbols

A documented list of symbols for the program is found in Table 1 on page 4.



Table 1: A list of symbols, constants, and description for model variables.

Symbol	Value	Description
<u>Indexes</u>		
$f$		index for area
$g$		index for group
$h$		index for sex
$i$		index for year
$j$		index for age
$k$		index for gear
<u>Model dimensions</u>		
$F$		number of model areas
$G$		number of biological stocks
$H$		number of sexes
$I$		number of years
$J$		number of age-classes ( $J$ is a plus group age)
$K$		number of gears (fisheries dependent & independent)
<u>Observations (data)</u>		
$C_{f,g,h,i,k}$		catch by area, stock, sex, year, and gear.
$I_{k,t}$		relative abundance index for gear $k$ in year $t$ .
$p_{k,t,a}$		observed proportion-at-age $a$ in year $t$ for gear $k$ .
$w_{f,g,h,i}$		empirical weight-at-age data.
<u>Age-schedule variables</u>		
$\iota_j, \hat{\iota}_j$		unfished and fished survivorship to age $j$
$l_j$		mean length at age $j$
$v_j$		mean vulnerability to fishing mortality at age $j$
$f_j$		fraction of age- $j$ biomass that is sexually mature
<u>Estimated parameters</u>		
$R_o$	$> 0$	Age- $j_1$ recruits in unfished conditions
$\ddot{R}$	$> 0$	average age- $j_1$ recruitment in year $i_1 - 1$
$\bar{R}$	$> 0$	average age- $j_1$ recruitment from year $i_1$ to $I$
$\kappa$	$> 1$	recruitment compensation
$M$	$> 0$	instantaneous natural mortality rate
$\tau_k^2$	$> 0$	total precision associated with observation error for gear $k$
$\varsigma^2$	$> 0$	total precision of the total error
$\gamma_k$		vector of selectivity parameters for gear $k$
$\bar{f}_k$		logarithm of the mean fishing mortality for gear $k$
$\ddot{\omega}$		age- $j_1$ deviates from $\ddot{R}$ prior to initial year
$\omega$		age- $j_1$ deviates from $\bar{R}$ for all years
$\varphi_t$		logarithm of annual change in natural mortality rate
Continued on next page		



**Table 1 – continued from previous page**

Symbol	Constraints	Description
$\Psi_{i,k}$		matrix of fishing mortality rate deviates in year $i$ for gear $k$
<u>Standard deviations</u>		
$\sigma_M$	0.1	standard deviation in random walk for natural mortality
$\sigma$		standard deviation for observation errors in survey index
$\tau$		standard deviation in process errors (recruitment deviations)
$\sigma_C$	0.0707	standard deviation in observed catch by gear
<u>Residuals</u>		
$\delta_t$		annual recruitment residual
$\eta_t$		residual error in predicted catch

## 2.3 Input data

The input data for the assessment consists of 6 major components: 1) model dimensions, 2) age-schedule information, 3) historical removals, bycatch, and discards, 4) relative abundance data, 5) age or size composition information, and 6) empirical observations on mean weight-at-age. The resolution of the data is as fine as the model dimensions. For example, if there are multiple areas defined in the model dimensions (i.e.,  $F > 1$ ), then removal data will indicate what area the catch was removed. It is also possible to have a data set that spans more than one dimension. For example it is possible to create two-sex model without knowing the sex ratio of the catch.

### 2.3.1 Model dimensions

The model has six dimensions that define the number of areas ( $F$ ), the number of biological stocks ( $G$ ), the number of sexes ( $H$ ), the number of years ( $I$ ), the number of age categories ( $J$ ), and the number of fishing and sampling gears ( $K$ ). Setting these dimensions ultimately determines the resolution of the input data, and the minimum number of parameters that are allocated to the problem.

### 2.3.2 Age-schedule information

Absent empirical size-age information, growth parameters for the von Bertalanffy growth model are required to predict the mean length-at-age in the model, as well as, allometric relationships to translate mean length to mean weight-at-age. The other critical piece of information is the maturity-at-age schedules. Maturity-at-age is either a parametric function based on the parameters  $\hat{a}$  and  $\hat{\gamma}$  and a logistic curve (see equation T2.7), or a vector of proportion mature-at-age is supplied. It is assumed that fecundity-at-age is proportional to the mean weight-at-age.



### 2.3.3 Historical removals

The model is fit to the historical removal data (e.g., catch and discards from all fisheries combined). Catch data may be in units of number, or weight, or even refer to the removal of roe in spawn on kelp type fisheries. For each catch observation, the gear, area, and stock it was harvested from must be specified. Also sex information (no sex, male or female) must also be specified. The catch is assumed to be measured with error and the CV in the errors must be specified.

### 2.3.4 Relative abundance

Trend information from commercial CPUE indices and survey information are a key piece of information for fitting the models. For each entry the area, stock, and gear type must be defined. In addition, information on sex must also be specified. The timing of the survey relative to the fraction of total removals removed by the fishery is also required. Lastly the relative weights of each observation must be specified, where 0 would indicate no weight or ignore the observation.

### 2.3.5 Composition information

The model can be fit to both size composition information or age composition information. In the case of age composition data, the proportions-at-age sampled by each gear, in a given year, area, and stock must be specified. The data can be entered in as proportions or numbers, the latter is necessary if fitting the data assuming multinomial sampling, where the sum of the numbers is the assumed effective sample size.

### 2.3.6 Empirical weight-at-age data

It is not uncommon for the mean size-at-age, or weight-at-age, to change over time. The mean weight-at-age data is used to convert numbers-at-age to biomass in the model and also to determine the spawning stock biomass when multiplied by the proportion mature-at-age. Rather than predict the changes in weight-at-age and fit to empirical observations, the empirical data (if available) can be used directly to determine biomass related measures. In cases where there are missing years of data, the data can be backfilled using a -ve year input; otherwise, missing data will be filled in using the mean weight-at-age based on the growth curve from the age-schedule information.

## 2.4 Analytic methods

### 2.4.1 Deterministic age-structured model

The following description of the steady-state, or equilibrium, age-structured model ignores subscripts associated with sex, area, and group for the purposes of clarity. For the age-structured model represented in Table 2, we assume the parameter vector  $\Theta$  in (T2.1) is



unknown and would eventually be estimated by fitting ISCAM to data. For a given set of growth parameters and maturity-at-age parameters defined by (T2.3), growth is assumed to follow von Bertalanffy (T2.4), mean weight-at-age is given by the allometric relationship in (T2.5), and the age-specific vulnerability is given by a logistic function (T2.6). The growth function can also be substituted with empirical weight-at-age data. Note also that alternative selectivity parametric and non-parametric functions can also be used and the logistic function is one option. Mean fecundity-at-age is assumed to be proportional to the mean weight-at-age of mature fish, where maturity at age is specified by the parameters  $\hat{a}$  and  $\hat{\gamma}$  for the logistic function.

Survivorship for unfished and fished populations is defined by (T2.8) and (T2.9), respectively. It is assumed that all individuals ages  $A$  and older (i.e., the plus group) have the same total mortality rate. The incidence functions refer to the life-time or per-recruit quantities such as spawning biomass per recruit ( $\phi_E$ ) or vulnerable biomass per recruit ( $\phi_b$ ). Note that upper and lower case subscripts denote unfished and fished conditions, respectively. Spawning biomass per recruit is given by (T2.10), the vulnerable biomass per recruit is given by (T2.11) and the per recruit yield to the fishery is given by (T2.12). Unfished recruitment is given by (T2.13) and the steady-state equilibrium recruitment for a given fishing mortality rate  $F_e$  is given by (T2.14). Note that in (T2.14) we assume that recruitment follows either a Beverton-Holt or a Ricker model in the forms:

$$R_e = \begin{cases} \frac{s_o R_e \phi_e}{1 + \beta R_e \phi_e}, & \text{Beverton-Holt} \\ s_o R_e \phi_e \exp(-\beta R_e \phi_e) & \text{Ricker} \end{cases}$$

where the maximum juvenile survival rate is the same for both forms of the recruitment model and is given by:

$$s_o = \kappa / \phi_E,$$

and the density-dependent term is given by:

$$\beta = \begin{cases} \frac{(\kappa - 1)}{R_o \phi_E}, & \text{Beverton-Holt} \\ \frac{\ln(\kappa)}{R_o \phi_E}, & \text{Ricker} \end{cases}$$

which simplifies to (T2.14). The equilibrium yield for a given fishing mortality rate is (T2.15). These steady-state conditions are critical for determining various reference points such as  $F_{MSY}$  and  $B_{MSY}$ , or  $F_{SPR\%}$  and  $B_{SPR\%}$ .

## 2.4.2 Stochastic age-structured model

The system of equations for the stochastic age-structured population model is laid out in Table 3. The estimated model parameters are defined by (T3.1), where the over all population



Table 2: Steady-state age-structured model assuming unequal vulnerability-at-age, age-specific natural mortality, age-specific fecundity and Beverton-Holt type recruitment.

Parameters	
$\Theta = (R_o, \kappa, M_j, \hat{a}, \hat{\gamma})$	(T2.1)
$B_o > 0; \kappa > 1; M_j > 0$	(T2.2)
$\Phi = (l_\infty, k, t_o, a, b, \dot{a}, \dot{\gamma})$	(T2.3)
Age-schedule information	
$l_j = l_\infty(1 - \exp(-k(j - t_o)))$	(T2.4)
$w_j = a(l_j)^b$	(T2.5)
$v_j = (1 + \exp(-(\hat{a} - j)/\gamma))^{-1}$	(T2.6)
$f_j = w_j(1 + \exp(-(\dot{a} - j)/\dot{\gamma}))^{-1}$	(T2.7)
Survivorship	
$\iota_j = \begin{cases} 1, & j = 1 \\ \iota_{j-1}e^{-M_{j-1}}, & j > 1 \\ \iota_{j-1}/(1 - e^{-M_j}), & j = A \end{cases}$	(T2.8)
$\hat{\iota}_j = \begin{cases} 1, & a = 1 \\ \hat{\iota}_{j-1}e^{-M_{j-1}-F_e v_{j-1}}, & j > 1 \\ \hat{\iota}_{j-1}e^{-M_{j-1}-F_e v_{j-1}}/(1 - e^{-M_j-F_e v_j}), & j = A \end{cases}$	(T2.9)
Incidence functions	
$\phi_E = \sum_{j=1}^{\infty} \iota_j f_j, \quad \phi_e = \sum_{j=1}^{\infty} \hat{\iota}_j f_j$	(T2.10)
$\phi_B = \sum_{j=1}^{\infty} \iota_j w_j v_j, \quad \phi_b = \sum_{j=1}^{\infty} \hat{\iota}_j w_j v_j$	(T2.11)
$\phi_q = \sum_{j=1}^{\infty} \frac{\hat{\iota}_j w_j v_j}{M_j + F_e v_j} (1 - e^{(-M_j - F_e v_j)})$	(T2.12)
Steady-state conditions	
$B_o = R_o \phi_B$	(T2.13)
$R_e = R_o \begin{cases} \frac{\kappa - \phi_E/\phi_e}{\kappa - 1} & \text{Beverton-Holt} \\ \frac{\ln(\kappa) - \ln(\phi_E/\phi_e)}{\ln(\kappa)} & \text{Ricker} \end{cases}$	(T2.14)
$C_e = F_e R_e \phi_q$	(T2.15)





scale is determined by the initial recruitment  $\ddot{R}$  and average recruitment  $\bar{R}$ . The productivity of the stock is defined by the instantaneous natural mortality rate  $M_h$  and model parameters that define growth (T2.3), which may be sex-specific. Survival from egg to recruitment is assumed to be density-dependent, and the density-dependent juvenile survival rate is determined by the steepness of the stock-recruitment relationship, defined here as the relative improvement in juvenile survival rate as the spawning stock tends to 0. This term is often referred to as the recruitment compensation parameter (Goodyear, 1993), defined herein as  $\kappa$ . The variance terms  $\tau_k^2$  represent the observation errors for each of  $k$  relative abundance indices, and  $\sigma_R^2$  is the variance in the annual recruitment variation. A vector of selectivity parameters  $\gamma_k$  are also part of the estimated parameter set and are discussed in more detail in Section 2.4.3 on page 11. Lastly, a vector of scaling parameters for fishing mortality are also estimated ( $\bar{f}_k$ ).

The stochastic elements of the model are defined by the remaining terms in (T3.1), where  $\ddot{\omega}$  and  $\omega$  represent recruitment deviations in the initial state and annual values, respectively. For the initial conditions, deviations for each age class older than the age-at-recruitment are estimated and the vector  $\omega$  is constrained to sum to 0 to ensure that the overall scaling parameter  $\bar{R}$  remains estimable.

The joint estimation of natural mortality rates in statistical catch-at-age models can be problematic due to potential confounding with other structural assumptions (i.e., changes in fisheries selectivity, especially selectivity on older animals). In addition to estimating natural mortality, there is an additional stochastic element that allows for the estimation of changes in natural mortality over time. This is represented in the vector  $\varphi$ , where time-varying changes in natural mortality rates can be modeled as a random walk process (see Section 2.4.4).

Annual instantaneous fishing mortality rates for each gear  $k$  are based on a vector of deviations about the mean fishing mortality rate. A ragged matrix  $\Psi$  with  $K$  rows (one for each gear) and a column for each year where the catch for that gear is greater than 0. To ensure the mean fishing mortality rates are estimable each row in  $\Psi$  is constrained to sum to 0.

Given initial estimates of the parameters in (T3.1), the state variables in (T3.2) are then derived based on equations T3.3-T3.8. Age- and year-specific fishing mortality rates are defined by (T3.3), and the total mortality (T3.4) is the sum of natural mortality and fishing mortality from all sources.

The age-structured population model is initialized based on the initial recruitment  $\ddot{R}$  and age-specific total mortality rates ( $Z_{h,i,j}$ ) and initial recruitment deviates  $\ddot{\omega}_j$ . Note that (T3.5) assumes the value of recruitment deviates are the same for both sexes, and that age  $J$  is a plus group containing individuals age- $J$  and older.

The dynamics of the population numbers-at-age are recursively updated based on (T3.7), where age-1 recruits each year is based on the estimated vector defined by (T3.6). The total catch, in numbers, is based on instantaneous rates and assumes both natural and fishing mortality occur simultaneously (T3.8). Spawning biomass in year  $i$  ( $B_i$ ) is defined by (T3.9).



Table 3: Stochastic age-structured population model assuming instantaneous mortality rates.

Model Parameters	
$\Theta = (R_o, \ddot{R}, \bar{R}, \kappa, M_h, \tau_k^2, \sigma_R^2, \gamma_k, \bar{f}_k, \ddot{\omega}, \omega, \varphi, \Psi)$	(T3.1)
State variables	
$N_h, \mathbf{Z}_h, \mathbf{F}_h, B_i, R_i, C_i$	(T3.2)
Derived variables	
$F_{h,i,j} = g(\gamma_k, \bar{f}_k, \Psi)$	(T3.3)
$Z_{h,i,j} = f(M_h, \varphi) + \sum_k F_{h,i,j}$	(T3.4)
Initial states ( $i = 1$ )	
$N_{h,i,j} = \frac{\ddot{R}}{H} \begin{cases} \exp(-Z_{h,i,j}(j-1) + \ddot{\omega}_j), & j_1 < j < J \\ \frac{-Z_{h,i,j}(j-1) + \ddot{\omega}_j}{1 - \exp(-Z_{h,i,j} + \ddot{\omega}_j)}, & j = J \end{cases}$	(T3.5)
State dynamics	
$R_i = \bar{R} \exp(\omega_i)$	(T3.6)
$N_{h,i,j} = \begin{cases} R_i/H & j == 1 \\ N_{h,i-1,j-1} \exp(-Z_{h,i-1,j-1}) & i > 1, 1 < j < J \\ N_{h,i-1,j} \exp(-Z_{h,i-1,j}) & i > 1, j = J \end{cases}$	(T3.7)
$C_i = \sum_h \sum_j \frac{N_{h,i,j} F_{h,i,j} (1 - \exp(-Z_{h,i,j}))}{Z_{h,i,j}}$	(T3.8)
$B_i = \sum_h \sum_j N_{h,i,j} f_{h,j}$	(T3.9)



### 2.4.3 Fisheries selectivity

### 2.4.4 Natural mortality

### 2.4.5 Options for selectivity

At present, there are six alternative age-specific selectivity options in ISCAM. The simplest of the selectivity options is a simple logistic function with two parameters where it is assumed that selectivity is time-invariant. The more complex selectivity options assume that selectivity may vary over time and may have as many as  $(A-1) \cdot T$  parameters. For time-varying selectivity, cubic and bicubic splines are used to reduce the number of estimated parameters. Prior to parameter estimation, ISCAM will determine the exact number of selectivity parameters that need to be estimated based on which selectivity option was chosen for each gear type. It is not necessary for all gear types to have the same selectivity option. For example it is possible to have a simple two parameter selectivity curve for say a survey gear, and a much more complicated selectivity option for a commercial fishery.

**Logistic selectivity** The logistic selectivity option is a two parameter model of the form

$$v_a = \frac{1}{1 + \exp(-(a - \mu_a)/\sigma_a)}$$

where  $\mu_a$  and  $\sigma_a$  are the two estimated parameters representing the age-at-50% vulnerability and the standard deviation, respectively.

**Age-specific selectivity coefficients** The second option also assumes that selectivity is time-invariant and estimates at total of  $A-1$  selectivity coefficients, where the plus group age-class is assumed to have the same selectivity as the previous age-class. For example, if the ages in the model range from 1 to 15 years, then a total of 14 selectivity parameters are estimated, and age-15+ animals will have the same selectivity as age-14 animals.

When estimating age-specific selectivity coefficients, there are two additional penalties that are added to the objective function that control how much curvature there is and limit how much dome-shaped can occur. To penalize the curvature, the square of the second differences of the vulnerabilities-at-age are added to the objective function:

$$\lambda_k^{(1)} \sum_{a=2}^{A-1} (v_{k,a} - 2v_{k,a-1} + v_{k,a-2})^2$$

The dome-shaped term penalty as:

$$\begin{cases} \lambda_k^{(2)} \sum_{a=1}^{A-1} (v_{k,a} - v_{k,a+1})^2 & (if) v_{k,a+1} < v_{k,a} \\ 0 & (if) v_{k,a+1} \geq v_{k,a} \end{cases}$$

For this selectivity option the user must specify the relative weights  $(\lambda_k^{(1)}, \lambda_k^{(2)})$  to add to these two penalties.



**Cubic spline interpolation** The third option also assumes time-invariant selectivity and estimates a selectivity coefficients for a series age-nodes (or spline points) and uses a natural cubic spline to interpolate between these nodes (Figure ??). Given  $n + 1$  distinct knots  $x_i$ , selectivity can be interpolated in the intervals defined by

$$S(x) = \begin{cases} S_0(x) & x \in [x_0, x_1] \\ S_1(x) & x \in [x_1, x_2] \\ \dots & \\ S_{n-1}(x) & x \in [x_{n-1}, x_n] \end{cases}$$

where  $S''(x_0) = S''(x_n) = 0$  is the condition that defines a natural cubic spline.

The same penalty functions for curvature and dome-shaped selectivity are also invoked for the cubic spline interpolation of selectivity.

**Time-varying selectivity with cubic spline interpolation** A fourth option allows for cubic spline interpolation for age-specific selectivity in each year. This option adds a considerable number of estimated parameters but the most extreme flexibility. For example, given 40 years of data and estimated 5 age nodes, this amounts 200 (40 years times 5 ages) estimated selectivity parameters. Note that the only constraints at this time are the dome-shaped penalty and the curvature penalty; there is no constraint implemented for say a random walk (first difference) in age-specific selectivity). As such this option should only be used in cases where age-composition data is available for every year of the assessment.

**Bicubic spline to interpolate over time and ages** The fifth option allows for a two-dimensional interpolation using a bicubic spline (Figure ??). In this case the user must specify the number of age and year nodes. Again the same curvature and dome shaped constraints are implemented. It is not necessary to have age-composition data each and every year as in the previous case, as the bicubic spline will interpolate between years. However, it is not advisable to extrapolate selectivity back in time or forward in time where there are no age-composition data unless some additional constraint, such as a random-walk in age-specific selectivity coefficients is implemented (as of July 1, 2015, this has not been implemented).

## 2.5 Residuals, likelihoods & objective function value components

There are 3 major components to the overall objective function that are minimized while ISCAM is performing maximum likelihood estimation. These components consist of the likelihood of the data, prior distributions and penalty functions that are invoked to regularize the solution during intermediate phases of the non-linear parameter estimation. This section discusses each of these in turn, starting first with the residuals between observed and predicted states followed by the negative loglikelihood that is minimized.



### 2.5.1 Catch data

It is assumed that the measurement errors in the catch observations are log-normally distributed, and the residuals is given by:

$$\eta_{k,t} = \ln(C_{k,t}) - \ln(\hat{C}_{k,t}), \quad (1)$$

The residuals are assumed to be normally distributed with a user specified standard deviation  $\sigma_C$ . At present, it is assumed that observed catches for each gear  $k$  is assumed to have the same standard deviation. To aid in parameter estimation, two separate standard deviations are specified in the control file: the first is the assumed standard deviation used in the first, second, to N-1 phases, and the second is the assumed standard deviation in the last phase. The negative loglikelihood (ignoring the scaling constant) for the catch data is given by:

$$\ell_C = \sum_k \left[ T_k \ln(\sigma_C) + \frac{\sum_t (\eta_{k,t})^2}{2\sigma_C^2} \right], \quad (2)$$

where  $T_k$  is the total number of catch observations for gear type  $k$ .

### 2.5.2 Relative abundance data

The relative abundance data are assumed to be proportional to biomass, or numbers, or spawning biomass (in the case of spawn surveys) that is vulnerable to the sampling gear:

$$V_{k,t} = \sum_a N_{t,a} e^{-\lambda_{k,t} Z_{t,a}} v_{k,a} w_a, \quad (3)$$

where  $v_{k,a}$  is the age-specific selectivity of gear  $k$ , and  $w_a$  is the mean-weight-at-age, or the mean fecundity-at-age, or simply a vector of 1s in cases where the index is proportional to population numbers. A user specified fraction of the total mortality  $\lambda_{k,t}$  adjusts the numbers-at-age to correct for survey timing relative to the fraction of total mortality that has occurred prior to the start of the survey. The residuals between the observed and predicted relative abundance index is given by:

$$\epsilon_{k,t} = \ln(I_{k,t}) - \ln(q_k) + \ln(V_{k,t}), \quad (4)$$

where  $I_{k,t}$  is the observed relative abundance index,  $q_k$  is the catchability coefficient for index  $k$ , and  $V_{k,t}$  is the predicted vulnerable biomass/numbers at the time of sampling. The catchability coefficient  $q_k$  is evaluated at its conditional maximum likelihood estimate:

$$\ln(q_k) = \frac{1}{N_k} \sum_{t \in I_{k,t}} \ln(I_{k,t}) - \ln(V_{k,t}),$$

where  $N_k$  is the number of relative abundance observations in index  $k$  (see [Walters and Ludwig, 1994](#), for more information). The negative loglikelihood for relative abundance data



is given by:

$$\ell_I = \sum_k \sum_{t \in I_{k,t}} \ln(\sigma_{k,t}) + \frac{\epsilon_{k,t}^2}{2\sigma_{k,t}^2} \quad (5)$$

where

$$\sigma_{k,t} = \frac{\rho\varphi^2}{\omega_{k,t}},$$

where  $\rho\varphi^2$  is the proportion of the total error that is associated with observation errors, and  $\omega_{k,t}$  is a user specified relative weight for observation  $t$  from gear  $k$ . The  $\omega_{k,t}$  terms allow each observation to be weighted relative to the total error  $\rho\varphi^2$ ; for example, to omit a particular observation, set  $\omega_{k,t} = 0$ , or to give 2 times the weight, then set  $\omega_{k,t} = 2.0$ . To assume all observations have the same variance then simply set  $\omega_{k,t} = 1$ . Note that if  $\omega_{k,t} = 0$  then equation (5) is undefined; therefore, a small constant to  $\omega_{k,t}$  (1.e-10, which is equivalent to assuming an extremely large variance) to ensure the likelihood can be evaluated.

### 2.5.3 Relative abundance data with time-varying catchability

The scaler  $q_k$  is also subject to change over time with gear innovations, changes in the distribution of the stock relative to the geographic extent of the survey, and a number of other reasons (Wilberg et al., 2009). To model changes in  $q_k$  as a random walk process

$$q_{k,t} = q_{k,t-1}e^{\delta_{t-1}},$$

we first compute the conditional vector of coefficients as (omitting the gear index  $k$  for clarity):

$$\ln(q_t) = \ln(I_t) - \ln(B_t). \quad (6)$$

Equation (6) assumes that the deviatoin between the observed index and predicted index is entirely due to variation in  $q$ , and measurement error is negligible. Under the constant  $q$  assumption the variance of the catchability vector  $\sigma_q^2 \rightarrow 0$ . If  $\sigma_q^2 > 0$ , then in a random walk model the conditional maximum likelihood estimates of  $\delta_t$  is the first difference

$$\delta_t = \ln(q_t) - \ln(q_{t+1}). \quad (7)$$

In the case of a random walk in  $q$ , the objective function minimizes the variance in  $\delta_t$ . The negative loglikelihood of the relative abundance data is given by

$$\ell_I = \sum_{t \in k} 0.5 \ln(2\pi) + \ln(\sigma_{q_t}) + \frac{(\delta_t - \bar{\delta})^2}{2\sigma_{q_t}^2}, \quad (8)$$

where

$$\sigma_{q_t}^2 = \frac{\rho\varphi^2}{\omega_t},$$



Note that in (8) the residual terms reflect the changes in catchability from year to year. Actual values of  $q_t$  for the predicted relative abundance index are given by:

$$q_t = \begin{cases} \exp(\ln(I_t) - \ln(B_t)), & t = 1 \\ q_{t-1}e^{\delta_{t-1}}, & t > 1 \end{cases} \quad (9)$$

There may be special cases where there has been a dramatic shift in the catchability coefficient due to sudden changes in gear (e.g., a shift from J-hooks to circle hooks), and the variance in  $q_t$  that year is much greater than the variance in other years. To model this effect, the variance weighting term  $\omega_t$  would be set to smaller value in the year when the dramatic change occurred.

#### 2.5.4 Age composition data

Sampling theory suggest that age composition data are derived from a multinomial distribution (Fournier and Archibald, 1982); however, ISCAM assumes that age-proportions are obtained from a multivariate logistic distribution (Schnute and Richards, 1995; Richards et al., 1997). The main reason ISCAM departs from the traditional multinomial model has to do with how the age-composition data are weighted in the objective function. First, the multinomial distribution requires the specification of an effective sample size; this may be done arbitrarily or through iterative re-weighting (McAllister and Ianelli, 1997; Gavaris and Ianelli, 2002), and in the case of multiple and potentially conflicting age-proportions this procedure may fail to converge properly. The assumed effective sample size can have a large impact on the overall model results.

A nice feature of the multivariate logistic distribution is that the age-proportion data can be weighted based on the conditional maximum likelihood estimate of the variance in the age-proportions. Therefore, the contribution of the age-composition data to the overall objective function is “self-weighting” and is conditional on other components in the model.

Ignoring the subscript for gear type for clarity, the observed and predicted proportions-at-age must satisfy the constraint

$$\sum_{a=1}^A p_{t,a} = 1$$

for each year. The residuals between the observed ( $p_{t,a}$ ) and predicted proportions ( $\widehat{p}_{t,a}$ ) is given by:

$$\eta_{t,a} = \ln(p_{t,a}) - \ln(\widehat{p}_{t,a}) - \frac{1}{A} \sum_{a=1}^A [\ln(p_{t,a}) - \ln(\widehat{p}_{t,a})]. \quad (10)$$

The conditional maximum likelihood estimate of the variance is given by

$$\widehat{\tau}^2 = \frac{1}{(A-1)T} \sum_{t=1}^T \sum_{a=1}^A \eta_{t,a}^2,$$



and the negative loglikelihood evaluated at the conditional maximum likelihood estimate of the variance is given by:

$$\ell_A = (A - 1)T \ln(\hat{\tau}^2). \quad (11)$$

In short, the multivariate logistic likelihood for age-composition data is just the log of the residual variance weighted by the number observations over years and ages.

There is also a technical detail in (10), where observed and predicted proportions-at-age must be greater than 0. It is not uncommon in catch-age data sets to observe 0 proportions for older, or young, age classes. ISCAM adopts the same approach described by [Richards et al. \(1997\)](#) where the definition of age-classes is altered to require that  $p_{t,a} \geq 0.02$  for every age in each year. This is accomplished by grouping consecutive ages, where  $p_{t,a} < 0.02$ , into a single age-class and reducing the effective number of age-classes in the variance calculation ( $\hat{\tau}^2$ ) by the number of groups created. The choice of 2% is arbitrary and the user can specify the minimum proportion (including 0) to consider when pooling age-proportion data. In the case of an exact 0 in the observed age-proportions the pooling of the adjacent age-class still occurs, this ensures that (10) is defined.

A **WARNING** about extremely weak year classes is required here. A potential problem exists if in fact there is a very small cohort relative to the adjacent cohorts such that it never makes up more than say 2% (or whatever minimum is specified) of the age-proportions in any given year. In such cases, the information in the age-composition data about this weak year class relative to of that the adjacent (younger) year class because its always pooled into the younger year class. ISCAM will actually estimate two strong cohorts instead of correctly estimating one strong and one weak cohort in the following year.

### 2.5.5 Stock-recruitment

There are two alternative stock-recruitment models available in ISCAM: the Beverton-Holt model and the Ricker model. Annual recruitment and the initial age-composition are treated as latent variables in ISCAM, and residuals between estimated recruits and the deterministic stock-recruitment models are used to estimate unfished spawning stock biomass and recruitment compensation. The residuals between the estimated and predicted recruits is given by

$$\delta_t = \ln(\bar{R}e^{w_t}) - f(B_{t-k}) \quad (12)$$

where  $f(B_{t-k})$  is given by either (??) or (??), and  $k$  is the age at recruitment. Note that a bias correction term for the lognormal process errors is included in (??) and (??).

The negative log likelihood for the recruitment deviations is given by the normal density (ignoring the scaling constant):

$$\ell_\delta = n \ln(\tau) + \frac{\sum_{t=1+k}^T \delta_t^2}{2\tau^2} \quad (13)$$

Equations (12) and (13) are key for estimating unfished spawning stock biomass and recruitment compensation via the recruitment models. The relationship between  $(s_o, \beta)$  and  $(B_o, \kappa)$





is defined as:

$$s_o = \kappa / \phi_E \quad (14)$$

$$\beta = \begin{cases} \frac{\kappa-1}{B_o} & \text{Beverton-Holt} \\ \frac{\ln(\kappa)}{B_o} & \text{Ricker} \end{cases} \quad (15)$$

where  $s_o$  is the maximum juvenile survival rate, and  $\beta$  is the density effect on recruitment.



## References

- Fournier, D. and Archibald, C. (1982). A general theory for analyzing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences*, 39(8):1195–1207.
- Fournier, D., Skaug, H., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M., Nielsen, A., and Sibert, J. (2011). Ad model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models.
- Gavaris, S. and Ianelli, J. (2002). Statistical Issues in Fisheries’ Stock Assessments\*. *Scandinavian Journal of Statistics*, 29(2):245–267.
- Goodyear, C. (1993). Spawning stock biomass per recruit in fisheries management: foundation and current use. *Canadian Special Publication of Fisheries and Aquatic Sciences*, pages 67–82.
- McAllister, M. K. and Ianelli, J. (1997). Bayesian stock assessment using catch-age data and the sampling: importance resampling algorithm. *Canadian journal of fisheries and aquatic sciences(Print)*, 54(2):284–300.
- Richards, L., Schnute, J., and Olsen, N. (1997). Visualizing catch-age analysis: a case study. *Canadian Journal of Fisheries and Aquatic Sciences*, 54(7):1646–1658.
- Schnute, J. and Richards, L. (1995). The influence of error on population estimates from catch-age models. *Canadian Journal of Fisheries and Aquatic Sciences*, 52(10):2063–2077.
- Walters, C. and Ludwig, D. (1994). Calculation of Bayes posterior probability distributions for key population parameters. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(3):713–722.
- Wilberg, M. J., Thorson, J. T., Linton, B. C., and Berkson, J. (2009). Incorporating time-varying catchability into population dynamic stock assessment models. *Reviews in Fisheries Science*, 18(1):7–24.