iSC$\forall$M Users Guide
Version 1.0

Steven Martell
University of British Columbia
Fisheries Centre
2202 Main Mall
Vancouver, BC
V6T 1Z4
Canada
s.martell@fisheries.ubc.ca

Contents

1 Introduction

The purpose of this users guide is to aid in the development of new assessment models using iSC$\forall$M and to document the code. iSC$\forall$M is written in AD Model Builder and the source code is freely available. This manual was written in LaTeX.

1.1 Overview of iSC$\forall$M

As an AD Model builder program, iSC$\forall$M has several input files and several output files along with the executable program that actually performs the non-linear parameter estimation and all other model calculations. There are three input files required:

1. iscam.dat
2. <data file>
3. <control file>

All three files are required to run iSC$\forall$M, and the files are read in the order presented above. The iscam.dat file contains only the file names of the data file and the control file. The data file contains all of the necessary data for a particular stock including, model dimensions, life-history information, time series data on observed catch, the relative abundance indices and information on age-compositions sampled from each of the fisheries.

The control file contains the necessary information for setting bounds and priors for estimated model parameters, specifying the types of selectivity curves for each of the fisheries, and other miscellaneous controls for producing various outputs and weighing components of the objective function. Note that iSC$\forall$M is intended to have a lot of flexibility, but with this flexibility comes at a cost of being more difficult to rapidly develop models and obtain reasonable parameter estimates.

*i*SC∀M also has a custom command line option for conducting simulation trials based on the observed data set. In a simulation trial, the historical data and known parameter values are used to simulate observed data with known assumptions. Following the simulation, the model then estimates the model parameters. This is an important feature to ensure that your model set up is capable of estimating the true parameter values, or used in simulation-estimation experiments for exploring estimability and parameter bias.

There are a number of standard report files produced by AD Model Builder programs, and in addition to these report files, there are additional custom files for dealing with the MCMC output from *i*SC∀M.

1.2 Obtaining *i*SC∀M

*i*SC∀Mcan be freely obtained from a google code repository `http://code.google.com/p/iscam-project/`. Or by directly emailing Steven Martell.

2 Running *i*SC∀M: input files & command line options

There are three required input files for *i*SC∀M: the `iscam.dat` file, the `datafile`, and the `controlfile`. By default when *i*SC∀Mruns, the first file it looks for is the `iscam.dat` file, unless otherwise specified by using the command line option `-ind`. The following subsections explains the details of each of the data files.

2.1 The `iscam.dat` file

What is required in the `iscam.dat` file is just the name of the data file and the control file, in that order. An example is given below for the `PHake2010.dat` and `Phake2010.ctl` data and control files.

```
PHake2010.dat #Data file name
PHake2010.ctl #Control file name
```

Note that it is necessary to have the `*.dat`, and `*.ctl` extensions, as *i*SC∀M will read in the entire filename including the extension. Also note that the `#` symbol acts as a comment line, and *i*SC∀M will ignore the contents of the remaining line when reading in data.

2.2 The data file

The data file and how it is set up is very important in ensuring that iSC∀M works correctly. In this section I have broken down the description of the data file into several blocks that deal with model dimensions, age-schedule information, removal data from each commercial and sampling gears, relative abundance information that may come from one or all of the gears, a description of the age composition, and finally empirical weight-at-age data. There are some data elements that are *mandatory* (e.g., dimensions, age-schedule information, removals for each gear) and some optional data. For example it is not necessary to have relative abundance information for each of the gear-type each year, or even any information on relative abundance. Nor is it necessary to have age-composition each and every year for every gear type; in fact, it's not necessary to have any age-composition data, but in such cases selectivity parameters will have to be fixed.

The data file is composed of several required sections (required in the sense that they must be defined, but do not necessarily have to have data). The first of these required sections is the model dimensions. Below is an example where the model starts in 1977 and the last year is 2009, the youngest age-group is 1 years old, and the oldest age-group is 15 years old and older (i.e., a plus group). The total number of unique gears (including gear that samples fish in surveys is two, and last line is an integer vector that specifies if the gear is a fishery, or a survey (using 1 or 0, respectively). For each gear you must specify 1 (a commercial fishery) or 0 (a fisheries independent survey). Again the `#` is a comment character and *i*SC∀M will ignore the contents after this character. The following is an example of the model dimensions section:

```
##_________________________
##____Model Dimensions____
1977 #first year of data
2009 #last year of data
1 #age of youngest age class
15 #age of plus group
```

```
2 #number of gears (ngear)
## flags for gears
## fishery (1) or
## survey (0) in ngears
1 0
##_________________________
```

The next required section is the age-schedule information pertaining to natural mortality, growth and maturity-at-age. For now, natural mortality is assumed to be age-independent.

```
## ________________________
## ___Age-schedules info___
#natural mortality rate (m)
0.23
#growth parameters (linf,k,to)
52, 0.32, 0
#length-weight allometry (a,b)
5e-6, 3.0
#maturity at age (am=log(3)/k)
#& gm=std for logistic
3.45, 0.35
## ________________________
```

Next is the time series data for the historical catch by year, fishery(ies) and survey(s). Note that it is assumed that catch exists for each year that is specified in the model dimensions section (e.g., 1977-2009). The first column is the year of the catch, and the subsequent columns are catch (in weight) for each fishery or survey. Years where there are no catches (or no fishery) should be replaced by a 0. In cases where surveys did not exist, or there were no removals (e.g., an acoustic survey), specify a zero catch for each year (row).

```
## ________________________
#Time series data
#Observed catch
#(1977-2009, 1,000,000 metric t)
#yr commercial survey
1977 0.132693 0
1978 0.103639 0
1979 0.137115 0
...  omitted data for space
2008 0.321546 0
2009 0.176671 0
## ________________________
```

The next section pertains to the relative abundance index, where first the number (`nit`) specified the number of independent surveys, and the next row specifies the number of observations (`nit_nobs` ,or rows of data for each survey). The first column is an integer vector that is used to index the survey year, the second column is the actual survey abundance index, and the third column is the gear index associated with this gear. The fourth column is the relative weight that should be used for the index. For example, setting wt=0 for a given year will result in omitting the data, or setting wt=2 would imply that the CV is one half of the other values. The last column specifies the fraction of total mortality that has occurred when the survey was conducted (e.g., if the survey is conducted half way through the year then 0.5 implies that 1/2 of $Z_{t,a}$ has occurred when the survey was conducted).

```
## ________________________
#Relative Abundance index from
#independent survey (it) 1970-2008
#nit
1
#nit_nobs
13
#iyr    it gear wt survey timing
1977 1.915  2  1   0.5
1980 2.115  2  1   0.5
1983 1.647  2  1   0.5
1986 2.857  2  1   0.5
...omitted data for space
2007 0.879  2  2   0.5
2009 1.460  2  0   0.5
## ________________________
```

For age-composition information, a 3 dimensional ragged array is used to store the information by gear-type (matrix), by year (rows of each matrix) and by age (columns of each matrix). An example of the age composition data is shown in Table 1 on page 6.

First you must specify the number of gears for which age-composition data exists. If there are no data, then set this to 0. On the next line you must specify the number of years of age-composition data there are for each gear type. Next, for each gear type for which age-composition data is available, you must specify the first age-class of the data, and on the next row specify the oldest age-class of the data. In the example on page 6, there are two gears, the first gear has 33 years of observations, and the second gear has 13 years of observations. Each gear has the youngest age-class at 2 years and the oldest age-class at 15 years. This means there are 14 columns of age-compositions for each gear type.

*i*SC∀M treats the age-composition data as a ragged object to avoid having to read in years of missing age-composition data. The year and gear indexes in the first two columns are used to extract predicted age-proportions to be used in the statistical comparison (negative loglikelihoods).

The first two columns of the age-composition data refer to the year and gear type from which the data were obtained. So in the example on the next page, the first 33 rows of the matrix (some of which is missing so it could fit on the page) corresponds to the years 1977-2009 for gear type 1, and from 1977 to 2009 every 2-3 years for gear type 2. At present *i*SC∀M weights each row for each gear type equally, future versions of *i*SC∀M will probably have a third column here where relative weights based on effective samples sizes can be specified for each observed age-composition data.

Empirical weight-at-age data can be optionally specified immediately following the age-composition data. By default, *i*SC∀M first constructs the observed weight-at-age data based on the age-schedule information specified earlier in the data file. If there is a partial or complete set of empirical weight-at-age data available, then the default weights-at-age are overwritten for the years in which empirical data are available. First you must specify the number of years of observed weight-at-age data as this dimensions the matrix to read in the data. Following is a matrix where the first column specifies the corresponding year the data were collected, and for each age class defined in the model dimensions (youngest age class to plus group age class), the observed mean weights at age must be specified. Note that the units must be in kilograms so that unit consistency in the conversion from numbers-at-age to weight-at-age can be maintained.

```
#n_wt_obs
5
#Empirical mean weight-at-age in kilograms
#A$yr  V1   V2   V3   V4   V5   V6   V7   V8   V9
1951 0.04 0.08 0.11 0.13 0.15 0.17 0.20 0.18 0.18
1952 0.04 0.08 0.11 0.13 0.15 0.17 0.16 0.17 0.18
1953 0.03 0.07 0.09 0.12 0.14 0.15 0.13 0.17 0.18
1954 0.04 0.08 0.10 0.12 0.15 0.17 0.17 0.18 0.18
1955 0.04 0.08 0.10 0.12 0.14 0.17 0.17 0.17 0.18
```

The last component of the data file is an end of file "eof" marker, which is set to 999. This is the last number read in from the datafile and *i*SC∀M checks to ensure it is 999. If there is an error reading the datafile, *i*SC∀M will break and report that there was an error reading the data.

```
## ________________________
#eof
999
## ________________________
```

Table 1: Example of age composition data in the data file.

```
#Age composition data by year, gear (ages 2-15+)
#na_gears
2
#na_nobs
33 13
#a_sage
2 2
#a_page
15 15
#yr  gear       V1       V2       V3       V4       V5       V6       V7       V8       V9      V10      V11      V12      V13      V14
1977    1 0.091087 0.039290 0.208628 0.028500 0.053160 0.211179 0.078270 0.079949 0.063640 0.058483 0.043761 0.029639 0.007592 0.006823
1978    1 0.022968 0.101932 0.068633 0.199094 0.033354 0.071961 0.208406 0.084622 0.072156 0.073040 0.024682 0.021006 0.013116 0.005030
1979    1 0.049457 0.089640 0.100254 0.046571 0.191908 0.071243 0.159754 0.158389 0.056370 0.037676 0.016184 0.010295 0.006469 0.005789
1980    1 0.009331 0.254593 0.042151 0.054263 0.050507 0.143816 0.065236 0.087843 0.169471 0.046122 0.037636 0.023076 0.008874 0.007079
1981    1 0.091224 0.062768 0.280898 0.012851 0.045430 0.047641 0.148751 0.062707 0.066417 0.125977 0.031183 0.012419 0.009671 0.002062
1982    1 0.181412 0.025886 0.016978 0.318964 0.032603 0.045648 0.045099 0.131034 0.027439 0.033879 0.119575 0.010972 0.006862 0.003648
1983    1 0.000322 0.327381 0.030386 0.021774 0.318861 0.034486 0.037515 0.044368 0.095257 0.024331 0.017871 0.037722 0.007340 0.002385
1984    1 0.000000 0.010415 0.546489 0.035445 0.072340 0.185115 0.023775 0.020842 0.014283 0.045333 0.009533 0.007920 0.024390 0.004121
1985    1 0.006798 0.006334 0.065169 0.607023 0.070421 0.058060 0.132423 0.011557 0.006879 0.007111 0.013539 0.002836 0.000000 0.011849
1986    1 0.111570 0.031159 0.007757 0.034088 0.485333 0.058011 0.043959 0.122124 0.022909 0.026576 0.014536 0.026627 0.004392 0.010957
1987    1 0.000000 0.264654 0.016305 0.003861 0.017893 0.540852 0.032262 0.016639 0.080708 0.003902 0.001822 0.005542 0.009811 0.005748
1988    1 0.002907 0.002881 0.325484 0.012085 0.007047 0.010794 0.464716 0.021331 0.009870 0.101698 0.001949 0.004157 0.001274 0.033806
1989    1 0.026833 0.022546 0.009612 0.452262 0.010250 0.004556 0.006132 0.394579 0.015267 0.006758 0.044542 0.000903 0.001179 0.004583
1990    1 0.048604 0.255566 0.024077 0.002273 0.251121 0.006576 0.001663 0.000990 0.323920 0.003924 0.000212 0.072414 0.000146 0.008513
1991    1 0.034754 0.176910 0.169392 0.027073 0.007271 0.316749 0.012094 0.001274 0.001349 0.206127 0.003853 0.000000 0.036791 0.006363
1992    1 0.035191 0.044184 0.126581 0.177710 0.021788 0.007533 0.344623 0.006212 0.001264 0.003920 0.198907 0.004982 0.000449 0.026655
1993    1 0.007327 0.219650 0.032109 0.141618 0.169717 0.014288 0.007544 0.287667 0.008052 0.001062 0.000425 0.104591 0.000492 0.005457
1994    1 0.000419 0.033794 0.194593 0.013819 0.121828 0.200067 0.013059 0.004773 0.307047 0.002355 0.004118 0.000280 0.096116 0.007732
1995    1 0.015172 0.001676 0.067824 0.247580 0.011946 0.076025 0.204514 0.017753 0.003065 0.259156 0.002369 0.003815 0.000000 0.089107
... some missing data removed here to fit on page.
2005    1 0.008720 0.004799 0.070427 0.055023 0.684012 0.084118 0.021823 0.028355 0.019809 0.010432 0.008069 0.002582 0.000360 0.001470
2006    1 0.016047 0.109332 0.016100 0.086023 0.047267 0.606611 0.050565 0.017944 0.019738 0.012433 0.009263 0.004693 0.001532 0.002454
2007    1 0.135250 0.030604 0.145496 0.015585 0.070675 0.041936 0.441809 0.059055 0.018388 0.018549 0.012342 0.004254 0.004551 0.001507
2008    1 0.086419 0.307710 0.023174 0.134343 0.009449 0.035456 0.033322 0.305151 0.032058 0.010867 0.008882 0.005414 0.003330 0.004426
2009    1 0.007237 0.201241 0.298293 0.044466 0.140682 0.014182 0.025967 0.022153 0.193496 0.036166 0.005012 0.004290 0.003855 0.002961
1977    2 0.054308 0.051673 0.322415 0.029524 0.041387 0.358094 0.049372 0.036486 0.020920 0.019594 0.010201 0.003792 0.000997 0.001237
1980    2 0.004557 0.555127 0.053761 0.032569 0.026590 0.117668 0.043603 0.093838 0.037630 0.022180 0.003734 0.006424 0.001338 0.000983
1983    2 0.000265 0.785009 0.026011 0.007869 0.103384 0.016545 0.011402 0.008131 0.022356 0.005273 0.006223 0.006489 0.001042 0.000000
1986    2 0.604601 0.015879 0.002792 0.019748 0.266035 0.028628 0.022778 0.029920 0.003627 0.003812 0.000276 0.001440 0.000463 0.000000
1989    2 0.169990 0.058515 0.012874 0.526835 0.011735 0.004161 0.007554 0.179632 0.009473 0.000722 0.017782 0.000000 0.000000 0.000726
1992    2 0.089253 0.011915 0.069071 0.176823 0.021856 0.008862 0.432238 0.013086 0.007872 0.003964 0.149487 0.007606 0.000000 0.007967
1995    2 0.324964 0.043475 0.012039 0.212541 0.009810 0.032765 0.148871 0.002177 0.000000 0.158452 0.000354 0.006429 0.000000 0.048122
1998    2 0.168351 0.187074 0.157169 0.195749 0.014026 0.055093 0.087607 0.010731 0.015903 0.048868 0.003121 0.001999 0.042448 0.011861
2001    2 0.709921 0.089531 0.052761 0.056572 0.026180 0.026069 0.014190 0.008255 0.005804 0.002446 0.002162 0.004212 0.000400 0.001496
2003    2 0.029781 0.025334 0.640666 0.109500 0.027623 0.060058 0.039723 0.021949 0.022287 0.007181 0.004232 0.004367 0.003083 0.004214
2005    2 0.239916 0.024324 0.072095 0.051813 0.482518 0.052666 0.017966 0.024352 0.013884 0.011229 0.004744 0.002436 0.000323 0.001734
2007    2 0.428146 0.024375 0.101876 0.011527 0.041221 0.026044 0.289941 0.030229 0.013473 0.013191 0.007185 0.006086 0.002778 0.003928
2009    2 0.001881 0.229516 0.423131 0.024861 0.091878 0.007856 0.018074 0.024434 0.128613 0.029027 0.009417 0.005566 0.005402 0.000343
```

2.3 The control file (still under development)

The first section of the control file pertains to the leading parameter vector which is summarized in Table 2. For now, there are 6 leading parameters for which the initial values (ival) lower (lb) and upper bounds (ub) and estimation phase must be specified. Each of these parameters also have parameters for the corresponding prior distributions defined by the prior_type, and parameters p1 and p2.

Table 2: Controls for estimated parameters in the control file.

```
## ___________________________________________________________________________ ##
##                              PACIFIC HAKE CONTROLS
## __________________CONTROLS FOR ESTIMATED PARAMETERS________________________ ##
##  Prior descriptions:
##                      -0 uniform (0,0)
##                      -1 normal (p1=mu,p2=sig)
##                      -2 lognormal (p1=log(mu),p2=sig)
##                      -3 beta (p1=alpha,p2=beta)
##                      -4 gamma(p1=alpha,p2=beta)
## ___________________________________________________________________________ ##
6   ## npar
##  ival        lb      ub      phz     prior   p1      p2      parameter name
## ___________________________________________________________________________ ##
    1.6         -5.0    15      4       1       0.9     0.5     #log_ro/msy
    0.65        0.2     1.0     4       3       3       2       #steepness/fmsy
    -1.469      -5.0    0.0     2       1       -1.469  0.05    #log.m
    1.6         -5.0    15      1       0       -5.0    15      #log_avgrec
    0.2         0.001   0.999   3       3       3.75    12      #rho
    1.25        0.01    500     3       4       1.01    1.01    #kappa (precision)
## ___________________________________________________________________________ ##
```

2.3.1 Prior type distributions

As of now there are 5 different prior types that can be specified and these are given by the integer values 0–4. The following list describes the prior types and the parameter values for the distributions:

0 A uniform prior between lb and up.

1 A normal prior p1 = mean, and p2 = standard deviation

2 A lognormal prior p1 = log(mean), and p2 = log standard deviation

3 A beta prior p1 = alpha, and p2 = beta with lb and ub transformed to a 0-1 scale.

4 A gamma prior with p1=alpha and p2=beta

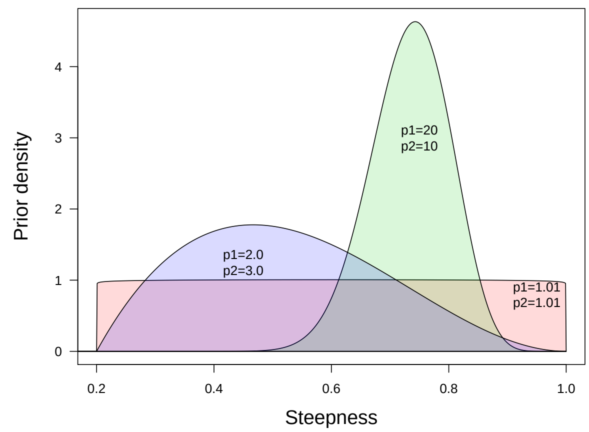

Figure 1: Three alternative beta prior distributions with corresponding values of p1 and p2 specified on the distribution. Note that specifying values of 1.01 results in a more or less uniform prior distribution for steepness in the Beverton-Holt stock recruitment model.

2.3.2 Selectivity controls

The next table of numbers in the control file contains the options for selectivities for each of the gear types (both fisheries and surveys). Currently there are 6 options implemented for selectivities in *i*SC∀M and the details of each are explained further in the model documentation section (see page 14). The following is an excerpt from the Pacific hake control file with selectivities defined for two gears:

```
## _________________________SELECTIVITY PARAMETERS______________________________ ##
## OPTIONS FOR SELECTIVITY:
##      1) logistic selectivity parameters
##      2) selectivity coefficients
##      3) a constant cubic spline with age-nodes
##      4) a time varying cubic spline with age-nodes
##      5) a time varying bicubic spline with age & year nodes.
##      6) fixed logistic (set isel_type=1, and estimation phase to -1)
## Gear 1 fishery:  Gear 2 survey
## isel_type
   5        1
## Age at 50% selectivity (logistic)
   3.5      4.0
## STD at 50% selectivity (logistic)
   1.0      0.5
## No. of age nodes for each gear (0 to ignore).
   5        5
## No. of year nodes for each gear (0 to ignore).
   11       3
## Estimation phase
   2        2
## Penalty weight for 2nd differences w=1/(2*sig^2)
   12.5     12.5
## Penalty weight for dome-shaped selectivity 1=1/(2*sig^2)
   3.125    200.0
## ___________________________________________________________________________ ##
```

There are two gears specified in this case, the first gear uses the time varying bicubic spline option with 5 age nodes and 11 year nodes and is estimated in phase 2 of the parameter search routine. The second fishery (second column) is a survey with a logistic selectivity function with initial values of 4.0 and 0.5 as the mean and standard deviation that is assumed in the first phase; in the second phase these values are then treated as estimated parameters. The last two rows of the selectivity controls defines the penalty weights used for the selectivity ogives where the 2nd differences controls the smoothness of the curve and the dome-shaped penalty limits how much the selectivity decline with older ages (dome-shaped). Note that these two penalties are ignored for the logistic (option 1 and option 6) forms of the selectivity curve.

2.3.3 Priors for survey catchability

Although the scaling parameters for surveys or relative abundance indices are not directly estimated, it is possible to specify prior distributions for the conditional maximum likelihood estimates of these parameters. Priors are specified by the following four lines in the control file, there `nits` is simply the number of relative abundance indices. In the following rows, you must specify a '0' or '1' for a uniform prior or an informative prior distribution. Note that if there is more than 1 survey, then you'll have to specify a 0 or 1 for each of the surveys (i.e., columns for each survey). The final two rows specify the log mean of the normal prior and the standard deviation (if they prior type is uniform you must still specify these values, however they are ignored in the objective function calculation; future versions may specify lower and upper bounds of a true uniform density). Again, in the case of multiple surveys, you must have a mean and standard deviation specified (in columns) for each of the surveys.

```
## ------------------------------------------------------------------------------ ##
##                              Priors for Survey q                               ##
## ------------------------------------------------------------------------------ ##
## nits  #number of surveys
   1
## priors 0=uniform density     1=normal density
   0
## prior log(mean);
   0
## prior sd
   1
## ------------------------------------------------------------------------------ ##
```

2.3.4 Other miscellaneous controls

The following is an ordered list of controls that turn various switches on and off or set up alternative structural assumptions such as Ricker recruitment or time varying natural mortality rates in iSC∀M. It's also a place holder to add additional features to iSC∀M as the model continues to evolve over time. The following is an ordered list describing in more detail each of the miscellaneous controls.

1. The first row of the miscellaneous controls is a flag that turns on and off the verbose output of iSC∀M.
2. Switch between Beverton-Holt and Ricker recruitment (T6.14).
3. The assumed standard deviation (in log space) in the observed catch in all phases except the last phase of the parameter estimation scheme. Note that this value must be greater than 0. Slightly larger values (say 0.05) will speed up convergence in earlier phases.
4. The assumed standard deviation (in log space) in the observed catch in the last phase of the parameter estimation scheme. Note that this value must be greater than 0. Slightly smaller values (say 0.01) will increase precision in the estimates of F but generally slow down convergence.
5. The next item is a flag to initialize the model at an unfished state in the initial year, otherwise, iSC∀M estimates the numbers at age in the first year.
6. Age-composition data are pooled into plus groups if the observed proportions-at-age are less than the specified percentage (e.g., <1% in the example control file below). See description of age-cmposition data, specifically the last paragraph in section 3.4.3 on page 17.
7. During the initial phases of the parameter estimation, a large penalty is used to regularize the estimates of the annual fishing mortality rates and then in the last phase this penalty is relaxed. The penalty is on deviations from the average fishing mortality rate (0.20 in the example below) for all fishing fleets.
8. The assumed standard deviation (in log space) in the fishing mortality rate penalty in the initial phases.
9. The assumed standard deviation (in log space) in the fishing mortality rate in the last phase, (should be a large value (e.g., 5 or greater), otherwise the penalty could reduce the true variation in the estimated F_t's.
10. The option to estimate changes in natural mortality rates via a random walk process is implemented by selecting a positive phase (negative

values imply a constant M) and annual deviations in M are not estimated.

11. If annual deviations in natural mortality are estimated, then the standard deviation for the normal prior for deviations in M_t are specified here.

12. Number of nodes to use in the cubic spline interpolation for the random walk in M

13. Fraction of the total mortality rate prior to spawning taking place (this adjust the spawning biomass downwards by ΔZ).

14. A switch to choose between the multivariate logistic or multinomial likelihood for age-composition data.

```
## ------------------------------------------------------------------------ ##
## OTHER MISCELANEOUS CONTROLS                                              ##
## ------------------------------------------------------------------------ ##
0           # 1  -verbose ADMB output (0=off, 1=on)
1           # 2  -recruitment model (1=beverton-holt, 2=ricker)
0.100       # 3  -std in observed catches in first phase.
0.0707      # 4  -std in observed catches in last phase.
0           # 5  -Assume unfished in first year (0=FALSE, 1=TRUE)
0.00        # 6  -Minimum proportion to consider in age-proportions for dmvlogistic
0.20        # 7  -Mean fishing mortality for regularizing the estimates of Ft
0.01        # 8  -std in mean fishing mortality in first phase
2.00        # 9  -std in mean fishing mortality in last phase
-3          # 10 -phase for estimating m_deviations (use -1 to turn off mdevs)
0.1         # 11 -std in deviations for natural mortality
12          # 12 -number of estimated nodes for deviations in natural mortality
0.50        # 13 -fraction of total mortality that takes place prior to spawning
1           # 14 -switch for age-composition likelihood (1=dmvlogistic,2=dmultinom)
## ------------------------------------------------------------------------ ##
```

2.4 Command line options

Currently there are two custom command line options available in *i*SC∀M in addition to the standard command line options provided by the AD Model Builder libraries (see help command line options -? for more information on the ADMB command line options).

The custom command line options are:

-sim N use this option turn the model into a simulation model, where N is the random number seed.

-retro N use the option for retrospective analysis where the last N years of data are ignored in the likelihood calculations.

There two random number seeds for the simulation model that the user should be aware of. The first is if the random number seed is set to 000, then *i*SC∀M will actually simulate data with no errors whatsoever. That is, the values of σ and τ (observation error and process errors, respectively) will be set equal to 0 and the simulation model will run as a deterministic model with no observation errors in the relative abundance index or age composition data. This option allows the user to check to ensure that the model parameters are in fact estimable with perfect information.

The second unique random seed number is 99, and this seed number is used for the simulation example in this manuscript. It specifies a unique time-varying selectivity curve for the commercial fishery that goes from dome-shaped to asymptotic.

2.4.1 Running the simulation model

Again, one of the first steps in conducting any assessment should be to first run the model on simulated data with no error to be certain that the model is capable of estimating the specified model parameters. To do so in *i*SC∀M the user simply needs to specify the command line option of `-sim 000`, where the '000' argument specifically instructs *i*SC∀M not to add any random variation to the observation or process errors. Invoking this command line option will run *i*SC∀M as normal, where the data from the data file is first read into memory, then the information from the control file is then read in. However, before proceeding straight into the non-linear parameter estimation procedure, *i*SC∀M first runs a simulation model based on the specified parameters listed in the control file. This simulation model will then replace the existing data in memory with simulated data, then perform the non-linear parameter estimation procedure and attempt to estimate the model parameters. If all is working well the estimated parameters listed in the parameter file should be very close, if not exactly, to the initial values specified in the control file.

3 Model Documentation

The section contains the documentation in mathematical form of the underlying age-structured model, and its steady state version that is used to calculate reference points, the observation models used in predicting observations, and the components of the objective function that formulate the statistical criterion (i.e., the objective function) that is used to estimate model parameters. 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. iSC∀M was implemented in AD Model Builder version 10.0 (**??**).

3.1 List of symbols

A documented list of symbols for the iSC∀Mprogram is found in Table 3 on page 10.

Table 3: A list of symbols, constants and description for variables used in iSC∀M.

Symbol	**Value**	**Description**
Indexes		
a		index for age
t		index for year
k		index for gear
Model dimensions		
$\acute{a}, A$	2, 10	youngest and oldest age class (A is a plus group)
$\acute{t}, T$	1951, 2010	first and last year of catch data
K	5	Number of gears including survey gears
Observations (data)		
$C_{k,t}$		catch in weight by gear k in year t
$I_{k,t}$		relative abundance index for gear k in year t
$p_{k,t,a}$		observed proportion-at-age a in year t for gear k
Estimated parameters		
R_o		Age-$\acute{a}$ recruits in unfished conditions
κ		recruitment compensation
M		instantaneous natural mortality rate
$\bar{R}$		average age-$\acute{a}$ recruitment from year $\acute{t}$ to T
$\ddot{R}$		average age-$\acute{a}$ recruitment in year $\acute{t}-1$
ρ		fraction of the total variance associated with observation error
ϑ		total precision (inverse of variance) of the total error
$\vec{\gamma}_k$		vector of selectivity parameters for gear k
$\digamma_{k,t}$		logarithm of the instantaneous fishing mortality for gear k in year t
$\ddot{\omega}_a$		age-$\acute{a}$ deviates from $\ddot{R}$ for year $\acute{t}$
ω_t		age-$\acute{a}$ deviates from $\bar{R}$ for years $\acute{t}$ to T
φ_t		logarithm of annual change in natural mortality rate
Standard deviations		
σ_M	0.1	standard deviation in random walk for natural mortality
σ		standard deviation for observation errors in survey index
τ		standard deviation in process errors (recruitment deviations)
σ_C	0.0707	standard deviation in observed catch by gear
Residuals		
δ_t		annual recruitment residual
η_t		residual error in predicted catch

3.2 Analytic methods: equilibrium considerations

For the steady-state conditions represented in Table 4, we assume the parameter vector Θ in (T4.1) is unknown and would eventually be estimated by fitting iSC$\forall$M to time series data. For a given set of growth parameters and maturity-at-age parameters defined by (T4.3), growth is assumed to follow von Bertalanffy (T4.4), mean weight-at-age is given by the allometric relationship in (T4.5), and the age-specific vulnerability is given by a logistic function (T4.6). Note, however, there are alternative selectivity functions implemented in iSC$\forall$M, the logistic function used here is simply for demonstration purposes. 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 $\dot{a}$ and $\dot{\gamma}$ for the logistic function.

Survivorship for unfished and fished populations is defined by (T4.8) and (T4.9), 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 (ϕ_E) or vulnerable biomass per recruit (ϕ_b). Note that upper and lower case subscripts denote unfished and fished conditions, respectively. Spawning biomass per recruit is given by (T4.10), the vulnerable biomass per recruit is given by (T4.11) and the per recruit yield to the fishery is given by (T4.12). Unfished recruitment is given by (T4.13) and the steady-state equilibrium recruitment for a given fishing mortality rate F_e is given by (T4.14). Note that in (T4.14) we assume that recruitment follows either a Beverton-Holt or a Ricker model in the forms:

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

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:

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

which simplifies to (T4.14). The equilibrium yield for a given fishing mortality rate is (T4.15). These steady-state conditions are critical for determining various reference points such as $\mathrm{F_{MSY}}$ and $\mathrm{B_{MSY}}$.

Table 4: Steady-state age-structured model assuming unequal vulnerability-at-age, age-specific natural mortality, age-specific fecundity and Beverton-Holt type recruitment.

Parameters

$$\Theta = (B_o, \kappa, M_a, \hat{a}, \hat{\gamma}) \tag{T4.1}$$

$$B_o > 0; \kappa > 1; M_a > 0 \tag{T4.2}$$

$$\Phi = (l_\infty, k, t_o, a, b, \dot{a}, \dot{\gamma}) \tag{T4.3}$$

Age-schedule information

$$l_a = l_\infty(1-\exp(-k(a-t_o))) \tag{T4.4}$$

$$w_a = a(l_a)^b \tag{T4.5}$$

$$v_a = (1+\exp(-(\hat{a}-a)/\gamma))^{-1} \tag{T4.6}$$

$$f_a = w_a(1+\exp(-(\dot{a}-a)/\dot{\gamma}))^{-1} \tag{T4.7}$$

Survivorship

$$\iota_a = \begin{cases} 1, & a=1 \\ \iota_{a-1}e^{-M_{a-1}}, & a>1 \\ \iota_{a-1}/(1-e^{-M_a}), & a=A \end{cases} \tag{T4.8}$$

$$\hat{\iota}_a = \begin{cases} 1, & a=1 \\ \hat{\iota}_{a-1}e^{-M_{a-1}-F_e v_{a-1}}, & a>1 \\ \hat{\iota}_{a-1}e^{-M_{a-1}-F_e v_{a-1}}/(1-e^{-M_a-F_e v_a}), & a=A \end{cases} \tag{T4.9}$$

Incidence functions

$$\phi_E = \sum_{a=1}^{\infty} \iota_a f_a, \quad \phi_e = \sum_{a=1}^{\infty} \hat{\iota}_a f_a \tag{T4.10}$$

$$\phi_B = \sum_{a=1}^{\infty} \iota_a w_a v_a, \quad \phi_b = \sum_{a=1}^{\infty} \hat{\iota}_a w_a v_a \tag{T4.11}$$

$$\phi_q = \sum_{a=1}^{\infty} \frac{\hat{\iota}_a w_a v_a}{M_a+F_e v_a}\left(1-e^{(-M_a-F_e v_a)}\right) \tag{T4.12}$$

Steady-state conditions

$$R_o = B_o/\phi_B \tag{T4.13}$$

$$R_e = R_o \begin{cases} \dfrac{\kappa-\phi_E/\phi_e}{\kappa-1} & \text{Beverton-Holt} \\ \dfrac{\ln(\kappa)-\ln(\phi_E/\phi_e)}{\ln(\kappa)} & \text{Ricker} \end{cases} \tag{T4.14}$$

$$C_e = F_e R_e \phi_q \tag{T4.15}$$

3.2.1 MSY based reference points

*i*SC∀M calculates $\mathrm{F_{MSY}}$ based reference points by finding the value of F_e that results in the zero derivative of (T4.15). This is accomplished numerically using a Newton-Raphson method where an initial guess for $\mathrm{F_{MSY}}$ is set equal to $1.5M$, then use (1) to iteratively find $\mathrm{F_{MSY}}$. Note that the partial derivatives in (1) can be found in Table 5.

$$F_{e+1} = F_e - \frac{\dfrac{\partial C_e}{\partial F_e}}{\dfrac{\partial^2 C_e}{\partial F_e}} \tag{1}$$

where

$$\frac{\partial C_e}{\partial F_e} = R_e\phi_q + F_e\phi_q\frac{\partial R_e}{\partial F_e} + F_eR_e\frac{\partial \phi_q}{\partial F_e}$$

$$\frac{\partial^2 C_e}{\partial F_e} = \phi_q\frac{\partial R_e}{\partial F_e} + R_e\frac{\partial \phi_q}{\partial F_e}$$

The algorithm usually converges in less than 10 iterations depending on how close the initial guess of $\mathrm{F_{MSY}}$ is to the true value. A maximum of 20 iterations are allowed in *i*SC∀M, however, if $\frac{\partial C_e}{\partial F_e} < 1e-5$ the algorithm stops. Note also, that this is only performed on data type variables and not differentiable variables within AD Model Builder.

Given an estimate of $\mathrm{F_{MSY}}$, other reference points such as MSY are calculated use the equations in Table 4 where each of the expressions is evaluated at $\mathrm{F_{MSY}}$. A graphical representation of MSY based reference points for two alternative values of the recruitment compensation parameter κ is show in Figure 2.

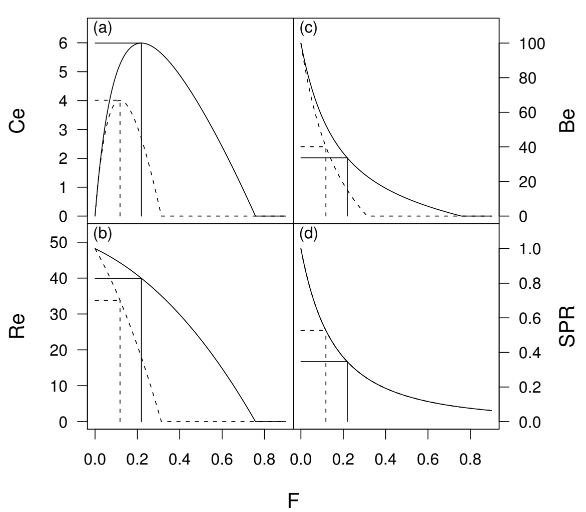

Figure 2: Equilibrium yield (a), recruits (b), biomass (c) and spawner per recruit (ϕ_e/ϕ_E) (d) versus instantaneous fishing mortality F_e for two different values of the recruitment compensation ratio ($\kappa = 12$ solid lines, $\kappa = 4$ dashed lines). Vertical lines in each panel correspond to $\mathrm{F_{MSY}}$ and horizontal lines correspond to various reference points that would achieve MSY.

Table 5: Partial derivatives, based on components in Table 4, required for the numerical calculation of $\mathrm{F_{MSY}}$ using (1).

Mortality & Survival

$$Z_a = M_a + F_e v_a \tag{T5.1}$$

$$S_a = 1 - e^{-Z_a} \tag{T5.2}$$

Partial for survivorship

$$\frac{\partial \hat{\iota}_a}{\partial F_e} = \begin{cases} 0, & a = 1 \\ e^{-Z_{a-1}}\left(\dfrac{\partial \hat{\iota}_{a-1}}{\partial F_e} - \hat{\iota}_{a-1}v_{a-1}\right), & 1 < a < A \\ \dfrac{\dfrac{\partial \hat{\iota}_{a-1}}{\partial F_e}}{1-e^{-Z_a}} - \dfrac{\hat{\iota}_{a-1}e^{-Z_{a-1}}v_a e^{-Z_a}}{(1-e^{-Z_a})^2}, & a = A \end{cases} \tag{T5.3}$$

Partials for incidence functions

$$\frac{\partial \phi_e}{\partial F_e} = \sum_{a=1}^{\infty} f_a \frac{\partial \hat{\iota}_a}{\partial F_e} \tag{T5.4}$$

$$\frac{\partial \phi_q}{\partial F_e} = \sum_{a=1}^{\infty} \frac{w_a v_a S_a}{Z_a}\frac{\partial \hat{\iota}_a}{\partial F_e} + \frac{\hat{\iota}_a w_a v_a^2}{Z_a}\left(e^{-Z_a} - \frac{S_a}{Z_a}\right) \tag{T5.5}$$

Partial for recruitment

$$\frac{\partial R_e}{\partial F_e} = \frac{R_o}{\kappa - 1}\frac{\phi_E}{\phi_e^2}\frac{\partial \phi_e}{\partial F_e} \tag{T5.6}$$

3.3 Analytic methods: state-dynamics

The estimated parameter vector in iSC∀M is defined in (T6.1), where R_0, κ and M are the leading unknown population parameters that define the overall population scale in the form of unfished recruitment and productivity in the form of recruitment compensation and natural mortality. The total variance ϑ^2 and the proportion of the total variance that is associated with observation errors ρ are also estimated, then the variance is partitioned into observation errors (σ^2) and process errors (τ^2) using (T6.2).

The unobserved state variables (T6.3) include the numbers-at-age year year t ($N_{t,a}$), the spawning stock biomass (B_t) and the total age-specific total mortality rate ($Z_{t,a}$).

The initial numbers-at-age in the first year (T6.4) and the annual recruits (T6.5) are treated as estimated parameters and used to initialize the numbers-at-age matrix. Age-specific selectivity for gear type k is a function of the selectivity parameters γ_k (T6.6), and the annual fishing mortality for each gear k in year t ($F_{k,t}$). The vector of log fishing mortality rate parameters $F_{k,t}$ is a bounded vector with a minimum value of -30 and an upper bound of 3.0. In arithmetic space this corresponds to a minimum value of 9.36e-14 and a maximum value of 20.01 for annual fishing mortality rates. In years where there are 0 reported catches for a given fleet, no corresponding fishing mortality rate parameter is estimated and the implicit assumption is there was no fishery in that year.

There is an option to treat natural mortality as a random walk process (T6.7), where the natural mortality rate in the first year is the estimated leading parameter (T6.1) and in subsequent years the mortality rate deviates from the previous year based on the estimated deviation parameter φ_t. If the mortality deviation parameters are not estimated, then M is assumed to be time invariant.

State variables in each year are updated using equations T6.9–T6.12, where the spawning biomass is the product of the numbers-at-age and the mature biomass-at-age (T6.9). The total mortality rate is given by (T6.10), and the total catch (in weight) for each gear is given by (T6.11) assuming that both natural and fishing mortality occur simultaneously throughout the year. The numbers-at-age are propagated over time using (T6.12), where members of the plus group (age A) are all assumed to have the same total mortality rate.

Recruitment to age k can follow either a Beverton-Holt model (T6.13) or a Ricker model (T6.14) where the maximum juvenile survival rate (s_o) in either case is defined by $s_o = \kappa/\phi_E$. For the Beverton-Holt model, β is derived by solving (T6.13) for β conditional on estimates of κ and R_o:

$$\beta = \frac{\kappa - 1}{R_o \phi_E},$$

and for the Ricker model this is given by:

$$\beta = \frac{\ln(\kappa)}{R_o \phi_E}$$

Table 6: Statistical catch-age model using the Baranov catch.

Estimated parameters

$$\Theta = (R_0, \kappa, M, \bar{R}, \rho, \vartheta^2, \gamma_k, \digamma_{k,t}, \{\omega_t\}_{t=1-A}^{t=T}, \{\varphi_t\}_{t=2}^{T}) \tag{T6.1}$$

$$\sigma^2 = \rho/\vartheta^2, \quad \tau^2 = (1-\rho)/\vartheta^2 \tag{T6.2}$$

Unobserved states

$$N_{t,a}, B_t, Z_{t,a} \tag{T6.3}$$

Initial states

$$N_{t,a} = \bar{R}e^{\omega_{t-a}} \exp(-M_t)^{(a-1)}; \quad t = 1; 2 \leq a \leq A \tag{T6.4}$$

$$N_{t,a} = \bar{R}e^{\omega_t}; \quad 1 \leq t \leq T; a = 1 \tag{T6.5}$$

$$v_{k,a} = f(\gamma_k) \tag{T6.6}$$

$$M_t = M_{t-1}\exp(\varphi_t), \quad t > 1 \tag{T6.7}$$

$$F_{k,t} = \exp(\digamma_{k,t}) \tag{T6.8}$$

State dynamics (t>1)

$$B_t = \sum_a N_{t,a} f_a \tag{T6.9}$$

$$Z_{t,a} = M_t + \sum_k F_{k,t} v_{k,t,a} \tag{T6.10}$$

$$\hat{C}_{k,t} = \sum_a \frac{N_{t,a} w_a F_{k,t} v_{k,t,a} \left(1 - e^{-Z_{t,a}}\right)^{\eta_t}}{Z_{t,a}} \tag{T6.11}$$

$$N_{t,a} = \begin{cases} N_{t-1,a-1}\exp(-Z_{t-1,a-1}) & a > 1 \\ N_{t-1,a}\exp(-Z_{t-1,a}) & a = A \end{cases} \tag{T6.12}$$

Recruitment models

$$R_t = \frac{s_o B_{t-k}}{1 + \beta B_{t-k}} e^{\delta_t - 0.5\tau^2} \quad \text{Beverton-Holt} \tag{T6.13}$$

$$R_t = s_o B_{t-k} e^{-\beta B_{t-k} + \delta_t - 0.5\tau^2} \quad \text{Ricker} \tag{T6.14}$$

3.3.1 Options for selectivity ($v_{k,t,a}$)

At present, there are six alternative age-specific selectivity options in *i*SC∀M. 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 a may have as many as (A-1)·T parameters. For time-varying selectivity, cubic and bicubic splines are used to reduce the number of estimated parameters. Prior to parameter estimation, *i*SC∀M 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 μ_a and σ_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 3). Given $n + 1$

distinct knots x_i, selectivity can be interpolated in the intervals defined by

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

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

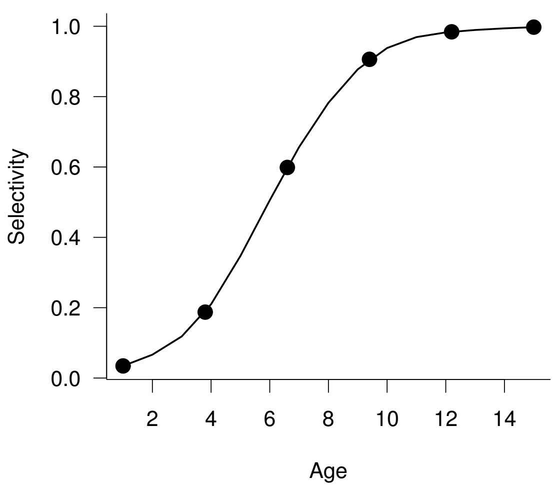

Figure 3: Example of a natural cubic spline interpolation for estimating selectivity coefficients. In *i*SC∀M the user specifies the number of nodes (circles) to estimate; then age-specific selectivity coefficients are interpolated using a natural cubic spline.

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 4). 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 December 21, 2011, this has not been implemented).

3.4 Residuals, likelihoods & objective function value components

There are 3 major components to the overall objective function that are minimized while *i*SC∀M 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.

3.4.1 Catch data

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

$$\eta_{k,t} = \ln(C_{k,t} + o) - \ln(\hat{C}_{k,t} + o), \qquad (2)$$

where o is a small constant (1.e-10) to ensure the residual is defined in the case of a 0 catch observation. The residuals are assumed to be normally distributed with a user specified standard deviation σ_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.

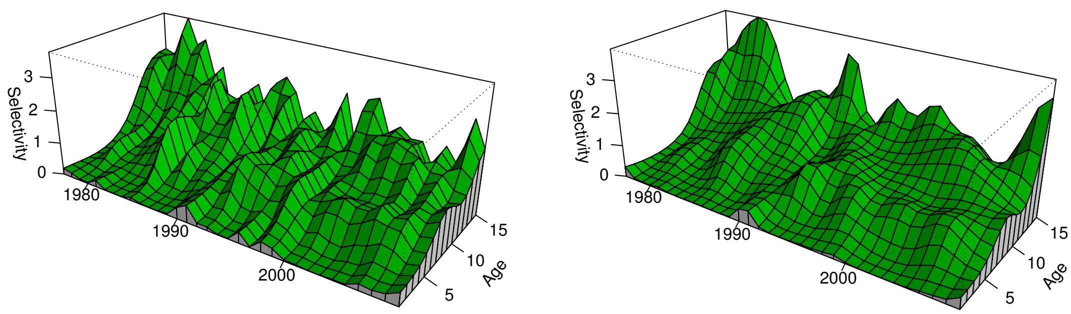

Figure 4: Example of a time-varying cubic spline (left) and bicubic spline (right) interpolation for selectivity as applied to the Pacific hake data. The panel on the left contains 165 estimated selectivity parameters and the bicubic interpolation estimates 85 selectivity parameters, or 5 age nodes and 17 year nodes. There are 495 actual nodes being interpolated.

The negative loglikelihood (ignoring the scaling constant) for the catch data is given by:

$$\ell_C = \sum_k \left[T_k \ln(\sigma_C) + \frac{\sum_t (\eta_{k,t})^2}{2\sigma_C^2} \right], \tag{3}$$

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

3.4.2 Relative abundance data

The relative abundance data are assumed to be proportional to biomass that is vulnerable to the sampling gear:

$$V_{k,t} = \sum_a N_{t,a} e^{-\lambda_{k,t} Z_{t,a}} v_{k,a} w_a, \tag{4}$$

where $v_{k,a}$ is the age-specific selectivity of gear k, and w_a is the mean-weight-at-age. A user specified fraction of the total mortality $\lambda_{k,t}$ adjusts the numbers-at-age to correct for survey timing. The residuals between the observed and predicted relative abundance index is given by:

$$\epsilon_{k,t} = \ln(I_{k,t}) - \ln(q_k) + \ln(V_{k,t}), \tag{5}$$

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 at the time of sampling. The catchability coefficient q_k is evaluated at its conditional maximum likelihood estimate:

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

where N_k is the number of relative abundance observations for index k (see **?**, for more information). The negative loglikelihood for relative abundance data is given by:

$$\ell_I = \sum_k \sum_{t \in I_{k,t}} \ln(\sigma_{k,t}) + \frac{\epsilon_{k,t}^2}{2\sigma_{k,t}^2} \tag{6}$$

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 (6) is undefined; therefore, iSC$\forall$M adds 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.

3.4.3 Age composition data

Sampling theory suggest that age composition data are derived from a multinomial distribution (**?**); however, iSC$\forall$M assumes that age-proportions are obtained from a multivariate logistic distribution (**??**). The main reason iSC$\forall$M 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 (**??**), 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:

$$\eta_{t,a} = \ln(p_{t,a}) - \ln(\widehat{p_{t,a}}) - \frac{1}{A} \sum_{a=1}^{A} \left[\ln(p_{t,a}) - \ln(\widehat{p_{t,a}}) \right]. \tag{7}$$

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:

$$\ell_A = (A-1)T\ln(\widehat{\tau}^2). \tag{8}$$

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 (7), 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. iSC∀M adopts the same approach described by **?** 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 ($\widehat{\tau}^2$) 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 (7) 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. iSC∀M will actually estimate two strong cohorts instead of correctly estimating one strong and one weak cohort in the following year.

3.4.4 Stock-recruitment

There are two alternative stock-recruitment models available in iSC∀M: the Beverton-Holt model and the Ricker model. Annual recruitment and the initial age-composition are treated as latent variables in iSC∀M, 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

$$\delta_t = \ln(\bar{R}e^{w_t}) - f(B_{t-k}) \tag{9}$$

where $f(B_{t-k})$ is given by either (T6.13) or (T6.14), and k is the age at recruitment. Note that a bias correction term for the lognormal process errors is included in (T6.13) and (T6.14).

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

$$\ell_\delta = n\ln(\tau) + \frac{\sum_{t=1+k}^{T}\delta_t^2}{2\tau^2} \tag{10}$$

Equations (9) and (10) are key for estimating unfished spawning stock biomass and recruitment compensation via the recruitment models. The relationship between (s_o, β) and (B_o, κ) is defined as:

$$s_o = \kappa/\phi_E \tag{11}$$

$$\beta = \begin{cases} \frac{\kappa-1}{B_o} & \text{Beverton-Holt} \\ \frac{\ln(\kappa)}{B_o} & \text{Ricker} \end{cases} \tag{12}$$

where s_o is the maximum juvenile survival rate, and β is the density effect on recruitment.