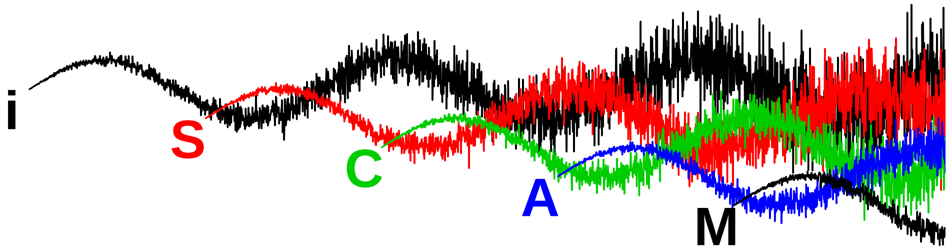

^{i}SCAM Users Guide
Version 1.0

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

Preface

This document is the users guide for the fisheries stock assessment model ^{i}SCA$_{\mathrm{M}}$, or Integrated Statistical Catch Age model. This assessment package was written by Steven Martell and may be freely used by others, but in no way am I responsible for the mess that may or may not happen if you use this software to do your job. Although I try hard, I cannot guarantee that this application is 100% free of bugs/coding errors so double check your own work and see if it makes sense. If you find a bug, fix it, recompile the code and continue on. Or let me know about the bug and I'll happily fix it for you, if I have time.

Contents

1 Introduction

The purpose of this users guide is to aid in the development of new assessment models using $^{i}\mathrm{SCA}_{\mathrm{M}}$ and to document the code. $^{i}\mathrm{SCA}_{\mathrm{M}}$ is written in AD Model Builder and the source code is freely available.

1.1 Overview of $^{i}\mathrm{SCA}_{\mathrm{M}}$

As an AD Model builder program, $^{i}\mathrm{SCA}_{\mathrm{M}}$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 $^{i}\mathrm{SCA}_{\mathrm{M}}$ 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 $^{i}\mathrm{SCA}_{\mathrm{M}}$ 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.

$^{i}\mathrm{SCA}_{\mathrm{M}}$ 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 $^{i}\mathrm{SCA}_{\mathrm{M}}$.

1.2 Obtaining $^{i}\mathrm{SCA}_{\mathrm{M}}$

$^{i}\mathrm{SCA}_{\mathrm{M}}$can be freely obtained from (website). Or by directly emailing Steven Martell.

2 Running $^{i}\mathrm{SCA}_{\mathrm{M}}$; input files & command line options

There are three required input files for $^{i}\mathrm{SCA}_{\mathrm{M}}$: the `iscam.dat` file, the `datafile`, and the `controlfile`. By default when $^{i}\mathrm{SCA}_{\mathrm{M}}$runs, 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.

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

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
```

2.2 The data file

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 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}SCA$_{\mathrm{M}}$ 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 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 on the following page.

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, you must specify the first age-class of the data, and on the next row specify the oldest age-class of the data. On the example in the next page, 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.

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.

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}SCA$_{\mathrm{M}}$ checks to ensure it is 999. If there is an error reading the datafile, ^{i}SCA$_{\mathrm{M}}$ will break and report that there was an error reading the data.

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

```
#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 1. 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 1: 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

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}SCA$_{M}$ and the details of each are explained further in the model documentation section (see page 10). 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 ^{i}SCA$_{\mathrm{M}}$. It's also a place holder to add additional features to ^{i}SCA$_{\mathrm{M}}$ 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 ^{i}SCA$_{\mathrm{M}}$.

2. Switch between Beverton-Holt recruitment (T4.13) and Ricker recruitment (T4.14).

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, ^{i}SCA$_{\mathrm{M}}$ 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 3.3.3 on page 13.

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.

```
## ______________________OTHER MISCELLANEOUS CONTROLS_________________________ ##
0           ## verbose ISCAM output (0=off, 1=on).
1           ## recruitment model (1=beverton-holt, 2=ricker).
0.05        ## std in observed catches in initial phases.
0.01        ## std in observed catches in last phase.
0           ## Assume unfished in first year (0=FALSE, 1=TRUE).
0.01        ## Minimum proportion to consider in age-proportions for dmvlogistic.
0.20        ## Mean fishing mortality for regularizing the estimates of Ft.
0.01        ## std in mean fishing mortality in initial phases.
5.00        ## std in mean fishing mortality in last phase.
-1          ## phase for estimating m_deviations (use -1 to turn off mdevs).
0.1         ## std in deviations for natural mortality.
## ------------------------------------------------------------------------- ##
```

2.4 Command line options

Currently there are two custom command line options available in ^{i}SCA$_{\mathrm{M}}$ 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}SCA$_{\mathrm{M}}$ 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}SCA$_{\mathrm{M}}$ the user simply needs to specify the command line option of `-sim 000`, where the '000' argument specifically instructs ^{i}SCA$_{\mathrm{M}}$ not to add any random variation to the observation or process errors. Invoking this command line option will run ^{i}SCA$_{\mathrm{M}}$ 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}SCA$_{\mathrm{M}}$ 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. ^{i}SCA$_{\mathrm{M}}$ was implemented in AD Model Builder version 9.0.0 (Otter Research, 2008; ADMB Project, 2009).

3.1 Age-structured population model: equilibrium considerations

For the steady-state conditions represented in Table 2, we assume the parameter vector Θ in (T2.1) is unknown and would eventually be estimated by fitting ^{i}SCA$_{\mathrm{M}}$ to time series 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). Note, however, there are alternative selectivity functions implemented in ^{i}SCA$_{\mathrm{M}}$, 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 (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 lifetime 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 (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 a Beverton-Holt model of the form:

$$R_e = \frac{s_o R_e \phi_e}{1 + \beta R_e \phi_e}$$

where

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

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

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}.

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

$$\Theta = (B_o, \kappa, M_a, \hat{a}, \hat{\gamma}) \quad \text{(T2.1)}$$

$$B_o > 0; \kappa > 1; M_a > 0 \quad \text{(T2.2)}$$

$$\Phi = (l_\infty, k, t_o, a, b, \dot{a}, \dot{\gamma}) \quad \text{(T2.3)}$$

Age-schedule information

$$l_a = l_\infty(1 - \exp(-k(a - t_o))) \quad \text{(T2.4)}$$

$$w_a = a(l_a)^b \quad \text{(T2.5)}$$

$$v_a = (1 + \exp(-(\hat{a} - a)/\gamma))^{-1} \quad \text{(T2.6)}$$

$$f_a = w_a(1 + \exp(-(\dot{a} - a)/\dot{\gamma}))^{-1} \quad \text{(T2.7)}$$

Survivorship

$$\iota_a = \begin{cases} 1, & a = 1 \\ \iota_{a-1} e^{-M_{a-1}}, & a > 1 \\ \iota_{a-1}/(1 - e^{-M_a}), & a = A \end{cases} \quad \text{(T2.8)}$$

$$\hat{\iota}_a = \begin{cases} 1, & a = 1 \\ \hat{\iota}_{a-1} e^{-M_{a-1} - F_e v_{a-1}}, & a > 1 \\ \hat{\iota}_{a-1} e^{-M_{a-1} - F_e v_{a-1}}/(1 - e^{-M_a - F_e v_a}), & a = A \end{cases} \quad \text{(T2.9)}$$

Incidence functions

$$\phi_E = \sum_{a=1}^{\infty} \iota_a f_a, \quad \phi_e = \sum_{a=1}^{\infty} \hat{\iota}_a f_a \quad \text{(T2.10)}$$

$$\phi_B = \sum_{a=1}^{\infty} \iota_a w_a v_a, \quad \phi_b = \sum_{a=1}^{\infty} \hat{\iota}_a w_a v_a \quad \text{(T2.11)}$$

$$\phi_q = \sum_{a=1}^{\infty} \frac{\hat{\iota}_a w_a v_a}{M_a + F_e v_a} \left(1 - e^{(-M_a - F_e v_a)}\right) \quad \text{(T2.12)}$$

Steady-state conditions

$$R_o = B_o/\phi_B \quad \text{(T2.13)}$$

$$R_e = R_o \frac{\kappa - \phi_E/\phi_e}{\kappa - 1} \quad \text{(T2.14)}$$

$$C_e = F_e R_e \phi_q \quad \text{(T2.15)}$$

3.1.1 MSY based reference points

^{i}SCA$_M$ calculates F_{MSY} based reference points by taking finding the value of F_e that results in the zero derivative of (T2.15). This is accomplished numerically using a Newton-Raphson method where an initial guess for F_{MSY} is set equal to $1.5M$, then use (1) to iteratively find F_{MSY}. Note that the partial derivatives in (1) can be found in Table 3.

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

where

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

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

The algorithm usually converges in less than 10 iterations depending on how close the initial guess of F_{MSY} is to the true value. A maximum of 20 iterations are allowed in ^{i}SCA$_M$, 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 F_{MSY}, other reference points such as MSY are calculated use the equations in Tabel 2 where each of the expressions is evaluated at F_{MSY}. A graphical representation of MSY based reference points for two alternative values of the recruitment compensation parameter κ is show in Figure 1.

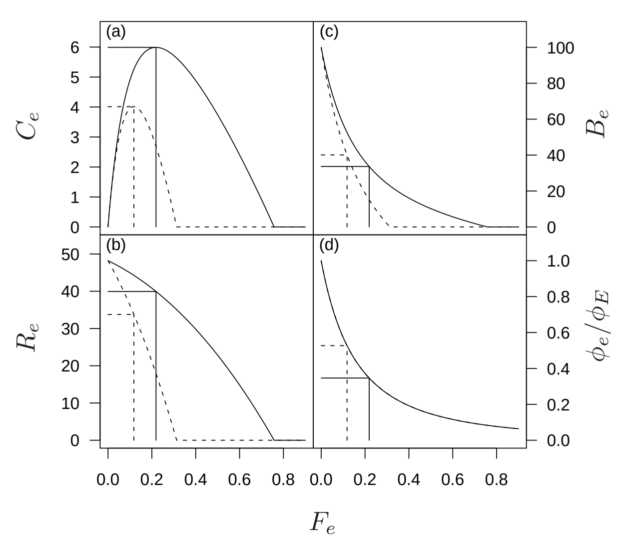

Figure 1: Equilibrium yield (a), recruits (b), biomass (c) and spawner per recruit (ϕ_e/ϕ_E) (d) versus instantaneous fishing mortality F_e for two different values of the recruitment compensation ratio ($\kappa = 12$ solid lines, $\kappa = 4$ dashed lines). Vertical lines in each panel correspond to $\mathrm{F_{MSY}}$ and horizontal lines correspond to various reference points that would achieve MSY.

Table 3: Partial derivatives, based on components in Table 2, required for the numerical calculation of $\mathrm{F_{MSY}}$ using (1).

Mortality & Survival

$$Z_a = M_a + F_e v_a \tag{T3.1}$$

$$S_a = 1 - e^{-Z_a} \tag{T3.2}$$

Partial for survivorship

$$\frac{\partial \hat{\iota}_a}{\partial F_e} = \begin{cases} 0, & a = 1 \\ e^{-Z_{a-1}} \left(\frac{\partial \hat{\iota}_{a-1}}{\partial F_e} - \hat{\iota}_{a-1} v_{a-1} \right), & a > 1 \end{cases} \tag{T3.3}$$

Partials for incidence functions

$$\frac{\partial \phi_e}{\partial F_e} = \sum_{a=1}^{\infty} f_a \frac{\partial \hat{\iota}_a}{\partial F_e} \tag{T3.4}$$

$$\frac{\partial \phi_q}{\partial F_e} = \sum_{a=1}^{\infty} \frac{w_a v_a S_a}{Z_a} \frac{\partial \hat{\iota}_a}{\partial F_e} + \frac{\hat{\iota}_a w_a v_a^2}{Z_a} \left(e^{-Z_a} - \frac{S_a}{Z_a} \right) \tag{T3.5}$$

Partial for recruitment

$$\frac{\partial R_e}{\partial F_e} = \frac{R_o}{\kappa - 1} \frac{\phi_E}{\phi_e^2} \frac{\partial \phi_e}{\partial F_e} \tag{T3.6}$$

3.2 Age-structured population model: state-dynamics

The estimated parameter vector in $^{i}\mathrm{SCA_M}$ is defined in (T4.1), 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 (T4.2).

The unobserved state variables (T4.3) 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 (T4.4) and the annual recruits (T4.5) 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 (T4.6), and the annual fishing mortality for each gear is the product of the average fishing mortality ($\bar{F}_k$) and the annual fishing mortality deviation.($\delta_{k,t}$).

There is an option to treat natural mortality as a random walk process (T4.7), where the natural mortality rate in the first year is the estimated leading parameter (T4.1) 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 T4.9–T4.12, where the spawning biomass is the product of the numbers-at-age and the mature biomass-at-age (T4.9). The total mortality rate is given by (T4.10), and the total catch (in weight) for each gear is given by (T4.11) assuming that both natural and fishing mortality occur simultaneously throughout the year. The numbers-at-age are propagated over time using (T4.12), 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 (T4.13) or a Ricker model (T4.14) where the maximum juvenile survival rate in either case is defined by κ/ϕ_E. For the Beverton-Holt model, β is derived by solving (T4.13) 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 4: Statistical catch-age model using the Baranov catch equation and C^* and F^* as leading parameters.

Estimated parameters

$$\Theta = (R_0, \kappa, M, \bar{R}, \rho, \vartheta^2, \gamma_k, \bar{F}_k, \delta_{k,t}, \{\omega_t\}_{t=1-A}^{t=T}, \{\varphi_t\}_{t=2}^{T}) \tag{T4.1}$$

$$\sigma^2 = \rho/\vartheta^2, \quad \tau^2 = (1-\rho)/\vartheta^2 \tag{T4.2}$$

Unobserved states

$$N_{t,a}, B_t, Z_{t,a} \tag{T4.3}$$

Initial states

$$N_{t,a} = \bar{R}e^{\omega_{t-a}} \exp(-M_t)^{(a-1)}; \quad t = 1; 2 \leq a \leq A \tag{T4.4}$$

$$N_{t,a} = \bar{R}e^{\omega_t}; \quad 1 \leq t \leq T; a = 1 \tag{T4.5}$$

$$v_{k,a} = f(\gamma_k) \tag{T4.6}$$

$$M_t = M_{t-1} \exp(\varphi_t), \quad t > 1 \tag{T4.7}$$

$$F_{k,t} = \bar{F}_k \exp(\delta_{k,t}) \tag{T4.8}$$

State dynamics (t>1)

$$B_t = \sum_a N_{t,a} f_a \tag{T4.9}$$

$$Z_{t,a} = M_t + \sum_k F_{k,t} v_{k,t,a} \tag{T4.10}$$

$$\hat{C}_{k,t} = \sum_a \frac{N_{t,a} w_a F_{k,t} v_{k,t,a} \left(1 - e^{-Z_{t,a}}\right)^{\eta_t}}{Z_{t,a}} \tag{T4.11}$$

$$N_{t,a} = \begin{cases} N_{t-1,a-1} \exp(-Z_{t-1,a-1}) & a > 1 \\ N_{t-1,a} \exp(-Z_{t-1,a}) & a = A \end{cases} \tag{T4.12}$$

Recruitment models

$$R_t = \frac{s_o B_{t-k}}{1 + \beta B_{t-k}} e^{\delta_t - 0.5\tau^2} \quad \text{Beverton-Holt} \tag{T4.13}$$

$$R_t = s_o B_{t-k} e^{-\beta B_{t-k} + \delta_t - 0.5\tau^2} \quad \text{Ricker} \tag{T4.14}$$

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

At present, there are six alternative age-specific selectivity options in ^{i}SCAM. 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·T parameters. For time-varying selectivity ^{i}SCAM implemented the uses of cubic and bicubic splines to reduce the number of estimated parameters. Prior to parameter estimation, ^{i}SCAM 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

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

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 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 2). 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.

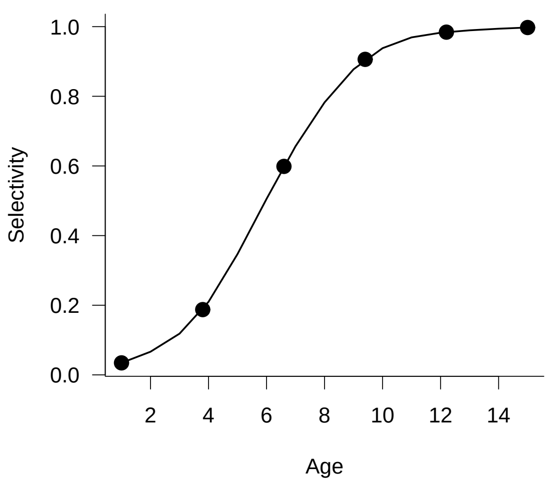

Figure 2: Example of a natural cubic spline interpolation for estimating selectivity coefficients. In ^{i}SCA$_{\text{M}}$ the user specifies the number of nodes (circles) to estimate; then age-specific selectivity coefficients are interpolated using a natural cubic spline.

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 3). 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 January 17, 2011, this has not been implemented).

3.3 Residuals, likelihoods & objective function value components

There are 3 major components to the overall objective function that are minimized while ^{i}SCA$_{\text{M}}$ 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 log-likelihood that is minimized.

3.3.1 Catch data

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

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

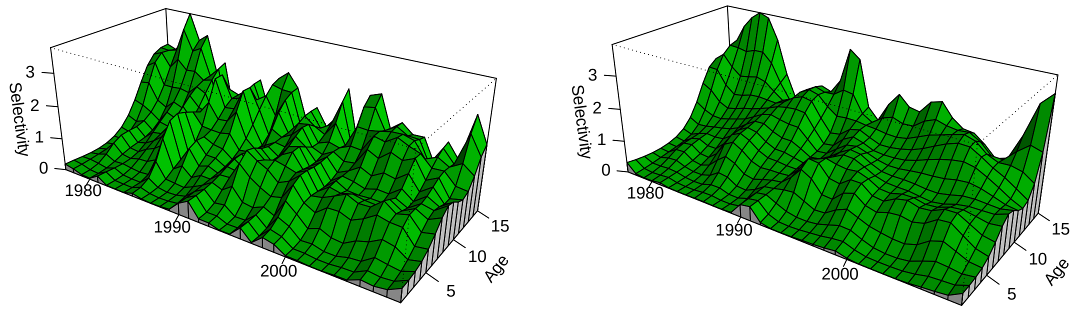

Figure 3: Example of a time-varying cubic spline (left) and bicubic spline (right) interpolation for selectivity as applied to the Pacific hake data. The panel on the left contains 165 estimated selectivity parameters and the bicubic interpolation estimates 85 selectivity parameters, or 5 age nodes and 17 year nodes. There are 495 actual nodes being interpolated.

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

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

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

3.3.2 Relative abundance data

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

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

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. For now, ^{i}SCA$_{\text{M}}$ assumes that the index is measured at the start of each year before any significant mortality takes place. The residuals between the observed and predicted relative abundance index is given by:

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

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 Walters and Ludwig, 1994, for more information). The negative loglikelihood for relative abundance data is given by:

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

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, ^{i}SCA$_{\mathrm{M}}$ 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.

3.3.3 Age composition data

Sampling theory suggest that age composition data are derived from a multinomial distribution (Fournier and Archibald, 1982); however, ^{i}SCA$_{\mathrm{M}}$ assumes that age-proportions are obtained from a multivariate logistic distribution (Schnute and Richards, 1995; Richards et al., 1997). The main reason ^{i}SCA$_{\mathrm{M}}$ 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 (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:

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

The conditional maximum likelihood estimate of the variance is given by

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

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 observed 0 proportions for older, or young, age classes. ^{i}SCA$_{\mathrm{M}}$ 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 ($\widehat{\tau}^2$) by the number of groups created.

3.3.4 Stock-recruitment

There are two alternative stock-recruitment models available in ^{i}SCA$_{\mathrm{M}}$: the Beverton-Holt model and the Ricker model. Annual recruitment and the initial age-composition are treated as latent variables in ^{i}SCA$_{\mathrm{M}}$, 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 (T4.13) or (T4.14), and k is the age at recruitment. Note that a bias correction term for the lognormal process errors is included in (T4.13) and (T4.14).

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.

4 Example: Simulation based on Strait of Georgia Pacific herring

The purpose of this example is to demonstrate how to use ^{i}SCA$_\text{M}$ to simulate data with known parameter values and then demonstrate the ability of the model to estimate the unknown parameters with and without observation and process errors. This example is based on data from the Strait of Georgia Pacific herring fishery.

There are three distinct commercial fishing fleets for Pacific herring in the Strait of Georgia that have operated, and continue to operate, between 1951 and 2010 (Figure 4). The first of these fleets is a purse seine fishery that generally operates in the winter months and historically used to catch herring for a reduction fishery and since the 1970s is now a much smaller bait fishery operation. Since the early 1970s a much more valuable seine fishery for sac roe has been in operation along with a gill net fishery that also targets spawning female herring for its roe.

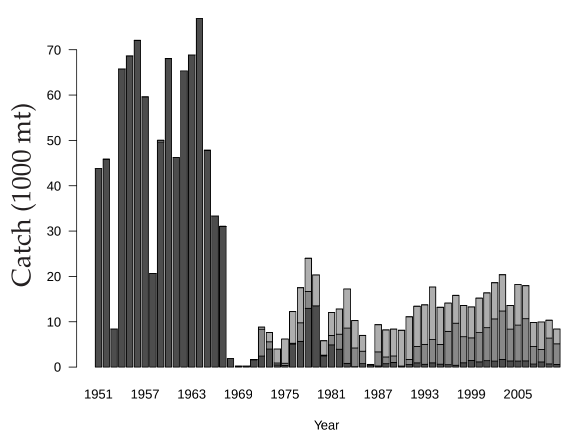

Figure 4: Total landings of Pacific herring in the Strait of Georgia stock assessment region by purse-seine (dark bars), Sn-roe (medium) and gill net (light bars).

The available data for the herring fishery consist of commercial landings for each of the three fisheries, age-composition data from each of these fisheries, average weight-at-age of the catch, and a fisheries independent spawning biomass index (based on spawn deposition). The spawn survey index is split into two separate series that correspond to a change in methods for estimating spawning deposition between 1987 and 1988. The model is fit to all of these data and uses the empirical weight-at-age data to convert numbers-at-age to biomass.

In this example, the command line option `-sim 000` is invoked to simulate fake data based on parameter values specified in the control file. Recall the command line option `-sim` is used to tell ^{i}SCA$_\text{M}$ to overwrite the existing observations in memory with simulated values (in this case with zero error because the random number seed is set to 000), and then attempt to estimate these model parameters. If all is working correctly, the `iscam.par` file should have nearly identical estimates for the model parameters as the initial values that are specified in the control file.

The control file used in this example is shown here, and there are a couple of things that should be highlighted with simulating data with no error. First off, the phase for the variance and variance partitioning parameters (ϑ and ρ, respectively) should be set to a negative value and not be estimated. Second ensure that the upper bound for ϑ (vartheta or the total precision) is set to a very high value (say 5000) and the initial value is set close to the upper bound (4999 in this case). The reason for fixing the parameters should be obvious, there is no error in the data to begin and thus its not necessary to estimate the total variance. The reason to set the initial value of ϑ to a large value is to minimize the a slight bias due to the lognormal bias correction in the stock-recruitment relationship (i.e., the $-0.5\tau^2$ in T4.13 or T4.14) during the parameter estimation phase. If you do not specify a large value of ϑ then it is unlikely that you will obtain nearly exact estimates of the unfished recruitment R_o.

Control file for the SOG herring example

```
## ------------------------------------------------------------------------ ##
##                           SOG HERRING 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
## ------------------------------------------------------------------------ ##
    7.60        -5.0    15      4       0       -5.0    15      #log_ro/msy
    0.70        0.2     1.0     4       3       1.1     1.1     #steepness/fmsy
    -0.7985     -5.0    0.0     2       1       -0.7985 0.2     #log.m
    7.40        -5.0    15      3       0       -5.0    15      #log_avgrec
    0.05        0.001   0.999   -3      3       1.01    1.01    #rho
    4999        0.01    5000    -3      4       1.01    1.01    #vartheta
## ------------------------------------------------------------------------ ##

## ------------------------------------------------------------------------ ##
## ______________________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 1:3 fishery:  Gear 4-5 survey
## isel_type
    1       1           1       1       1
## Age at 50% selectivity (logistic)
    1.5     2.0         2.5     2.05    2.05
## STD at 50% selectivity (logistic)
    0.75      0.5       0.2     0.05    0.05
## No. of age nodes for each gear (0 to ignore).
    5       5           5       0       0
## No. of year nodes for each gear (0 to ignore).
    12      3           10      0       0
## Estimation phase
    2       2            2      -2      -2
## Penalty weight for 2nd differences w=1/(2*sig^2)
    12.5    12.5        12.5    12.5    12.5
## Penalty weight for dome-shaped selectivity 1=1/(2*sig^2)
    3.125   200.0       200.0   200.0   200.0
## ------------------------------------------------------------------------ ##

## ------------------------------------------------------------------------ ##
##                           Priors for Survey q                            ##
## ------------------------------------------------------------------------ ##
## nits  #number of surveys
    2
## priors 0=uniform density     1=normal density
    0       0
## prior log(mean)
    0       0
## prior sd
    1       1
## ------------------------------------------------------------------------ ##

## ______________________OTHER MISCELLANEOUS CONTROLS______________________ ##
0           ## verbose ADMB output (0=off, 1=on)
1           ## recruitment model (1=beverton-holt, 2=ricker)
0.0025      ## std in observed catches in first phase.
0.0001      ## std in observed catches in last phase.
0           ## Assume unfished in first year (0=FALSE, 1=TRUE)
0.00        ## Minimum proportion to consider in age-proportions for dmvlogistic
0.05        ## Mean fishing mortality for regularizing the estimates of Ft
0.01        ## std in mean fishing mortality in first phase
5.00        ## std in mean fishing mortality in last phase
-3          ## phase for estimating m_deviations (use -1 to turn off mdevs)
0.01        ## std in deviations for natural mortality
## ------------------------------------------------------------------------ ##


## eofc
999
```

5 Example Assessment: the Namibian hake case study

As a simple example of fitting ^{i}SCA$_\mathrm{M}$ to CPUE data only, we use the Namibian hake case study from chapter 10 in the Ecological Detective (Hilborn and Mangel, 1997). In this example the available data consist of catch (thousands of tons) and CPUE (tons per standardized trawl hour). Hilborn and Mangel (1997) provide three alternative models to the data that range from simple 4 parameter Schaefer production models (observation & process error only) and a 5 parameter lagged recruitment, growth/survival model. In this example they assume the stock is at an unfished state in 1967.

To conduct the assessment using ^{i}SCA$_\mathrm{M}$ with the same unfished assumption in 1965, the "Assume unfished in first year (0=FALSE, 1=TRUE)" flag must be set to 1 (see the following control file). ^{i}SCA$_\mathrm{M}$ is an age-structured model, and in this example there are no available age-composition data to compare with. Therefore we must also assume a selectivity curve for this fishery. In this example, selectivity was assume to follow a logistic function with the 50% vulnerability-at-age equal to 3.5 years with a standard deviation of 1.0 years. It is also necessary in this case to turn off the estimation of the selectivity parameters by setting the estimation phase to a negative number (e.g., -1).

For the estimated leading parameters, two of the six parameters are not estimated `#log_m` and `#rho`, which is the instantaneous natural mortality rate and the proportion of the total error that is associated with observation errors. A bounded uniform prior is assumed for R_o and a beta prior for steepness h with an expected value of 0.6. The natural mortality rate M is assumed known and fixed at a value of 0.345. A uniform bounded prior is assumed for the log of the average recruitment level, and a non-informative gamma prior is assumed for the total precision κ. In this example we assume that the total error is allocated to observation and process error equally ($\rho = 0.5$).

```
Control file for the Namibian hake data.
## ------------------------------------------------------------------------ ##
##                          NAMIBIAN 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
## ------------------------------------------------------------------------ ##
    7.1         -5.0    15      4     0         -5.0    15      #log_ro/msy
    0.95        0.2     1.0     4     3         1.01    1.01    #steepness/fmsy
```

```
    -1.06421   -5.0    0.0      -2     2         -1.469  0.05   #log.m
    7.1         -5.0    15       1     0         -5.0    15     #log_avgrec
    0.50        0.001   0.999   -3     3         3.75    12     #rho
    5.00        0.01    500      3     4         1.01    1.01   #kappa (precision)
## ___________________________________________________________________________ ##

## ________________________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
## isel_type
    1
## Age at 50% selectivity (logistic)
    3.5
## STD at 50% selectivity (logistic)
    1.0
## No. of age nodes for each gear (0 to ignore).
    5
## No. of year nodes for each gear (0 to ignore).
    17
## Estimation phase
    -1
## Penalty weight for 2nd differences w=1/(2*sig^2)
    12.5
## Penalty weight for dome-shaped selectivity 1=1/(2*sig^2)
    3.125
## ___________________________________________________________________________ ##

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

## ______________________OTHER MISCELLANEOUS CONTROLS_________________________ ##
0           ## verbose ADMB output (0=off, 1=on)
1           ## recruitment model (1=beverton-holt, 2=ricker)
0.05        ## std in observed catches in first phase.
0.025       ## std in observed catches in last phase.
1           ## Assume unfished in first year (0=FALSE, 1=TRUE)
0.01 ## Minimum proportion to consider in age-proportions for dmvlogistic
0.20 ## Mean fishing mortality for regularizing the estimates of Ft
0.01 ## std in mean fishing mortality in first phase
5.00 ## std in mean fishing mortality in last phase
-1 ## phase for estimating m_deviations (use -1 to turn off mdevs)
0.1 ## std in deviations for natural mortality
## ___________________________________________________________________________ ##
## eofc
999
```

To convert numbers-at-age to biomass, growth was based on the von Bertalanffy growth parameters in the `NamibianHake.dat` file and the allometric relationship $w_a = a(l_a)^b$. Maturity-at-age is based on the logistic function with age-4 being the age at 50% maturity and 0.2 is the standard deviation. The plus group age was assumed to be 25 years, and there is only one fishing gear exploiting this stock.

Catch is taken by a single gear each year between 1965 and 1987, and the relative abundance index is based on the catch per standardized hour of trawling for the commercial gear. It is assumed that each CPUE observation is assumed to have the same error distribution, and the relative weights of each observations are all set equal to one.

There is no age-composition data to speak of, but `#na_gears` must have a value of 1 in order to proceed with reading the remaining portion of the data file.

Data file for the Namibian hake data.

```
#NB The data herein were taken from the Ecological detective.
## ________________________
## ____Model Dimensions____
1965 #first year of data
1987 #last year of data
1 #age of youngest age class
25 #age of plus group
1 #number of gears (ngear)
## flags for fishery (1) or survey (0) in ngears
1
## ________________________

## ________________________
#Age-schedule and population parameters
#natural mortality rate (m)
0.345
#growth parameters (linf,k,to)
111, 0.23, 0
#length-weight allometry (a,b)
3.65e-6, 3.0
#maturity at age (am=log(3)/k) & gm=std for logistic
4.0, 0.2
## ________________________

#Time series data
#Observed catch (1965-1987, 1000 metric t)
#yr commercial survey
#Year Catch
 1965    94
 1966   212
 1967   195
 1968   383
 1969   320
 1970   402
 1971   366
 1972   606
 1973   378
 1974   319
 1975   309
 1976   389
 1977   277
 1978   254
 1979   170
 1980    97
 1981    91
 1982   177
 1983   216
 1984   229
 1985   211
 1986   231
 1987   223


#Relative Abundance index from fisheries independent survey (it)
#nit
1
#nit_nobs
23
#iyr    it gear wt
#  Year CPUE gear wt survey timing
1965 1.78    1  1    0.5
1966 1.31    1  1    0.5
1967 0.91    1  1    0.5
1968 0.96    1  1    0.5
1969 0.88    1  1    0.5
1970 0.90    1  1    0.5
1971 0.87    1  1    0.5
1972 0.72    1  1    0.5
1973 0.57    1  1    0.5
1974 0.45    1  1    0.5
1975 0.42    1  1    0.5
1976 0.42    1  1    0.5
1977 0.49    1  1    0.5
1978 0.43    1  1    0.5
1979 0.40    1  1    0.5
1980 0.45    1  1    0.5
1981 0.55    1  1    0.5
1982 0.53    1  1    0.5
1983 0.58    1  1    0.5
1984 0.64    1  1    0.5
1985 0.66    1  1    0.5
1986 0.65    1  1    0.5
1987 0.63    1  1    0.5

#Age composition data by year, gear (ages 2-15+)
#na_gears
1
#na_nobs
0
#a_sage
0
#a_page
0
#yr  gear       Age1 Age2 Age3 ...

#n_wt_obs
0

#eof
999
```

5.1 Maximum likelihood estimates of the model parameters

Estimates of unfished spawning biomass is 2,877, steepness is 0.79, MSY is 266, and $F_{MSY}$is 0.33. These results are very similar to those obtained by Hilborn and Mangel (1997) for the Schaefer model with observation error. Estimates of the total standard deviation amount to 0.16 which equally breaks down to 0.081 for observation and process errors.

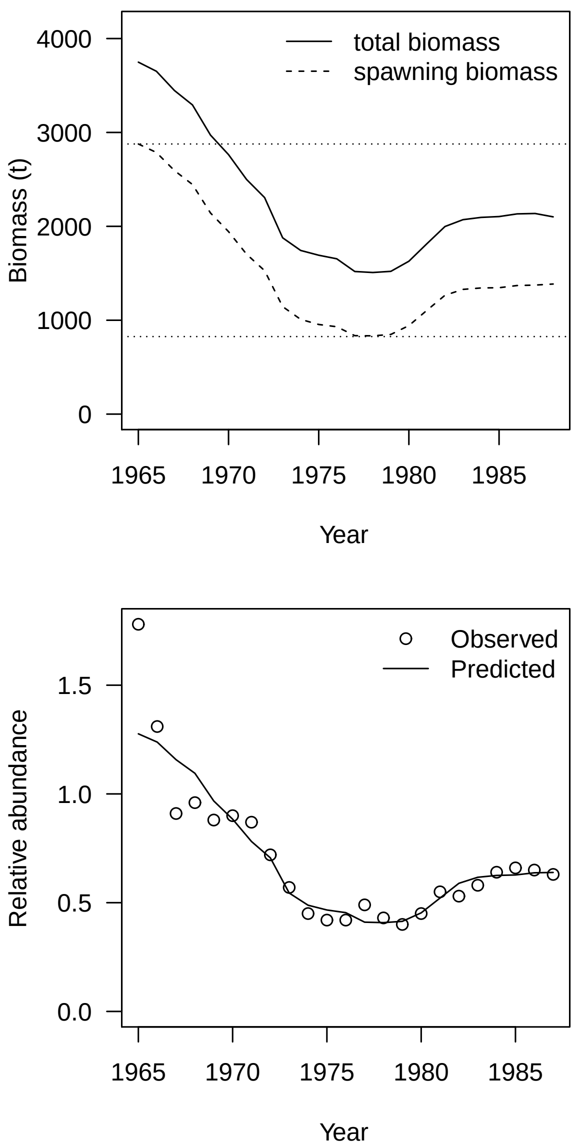

Figure 5: Estimates of total biomass and spawning biomass, observed and predicted CPUE, for the Namibian hake data from ^{i}SCA$_{M}$. Unfished biomass, B_{MSY}, and MSY based depletion levels are shown as horizontal dotted lines.

5.2 Bayesian analysis of model parameters & policy parameters

Marginal posterior distributions of model parameters were constructed by using the metropolis algorithm built into ADMB to sample from the joint posterior distribution. This is accomplished by running ^{i}SCA$_{M}$ in -mcmc mode followed by the -mceval option to produce the `iscam.mcmc` output file. In this example an MCMC chain of length 1,000,000 was run and samples were taken systematically every 500 iterations (`-mcsave 500`), which results in a posterior sample size of 2000.

Uniform prior distributions for the unfished recruitment and average recruitment (R_0 and $\bar{R}$), and non-informative gamma prior for the precision parameter ($1/\vartheta$). In the case of the steepness parameter, a non-informative beta prior was used ($p(h) \sim beta[1.01, 1.01]$), where steepness is re-scaled to the interval 0.2-1.0 (i.e, $(h-0.2)/0.8$) such that a 0 probability was assigned for h values less than 0.2. In comparison to the results obtained by Hilborn and Mangel (1997) using a biomass production model with lagged recruitment and a Beverton-Holt recruitment function, the data here appear to have some information about the steepness parameter (Fig. 6). This owes in part to differences in assumptions about growth, maturity and selectivity between the LRGS model used by Hilborn and Mangel (1997) and this ^{i}SCA$_{M}$ example.

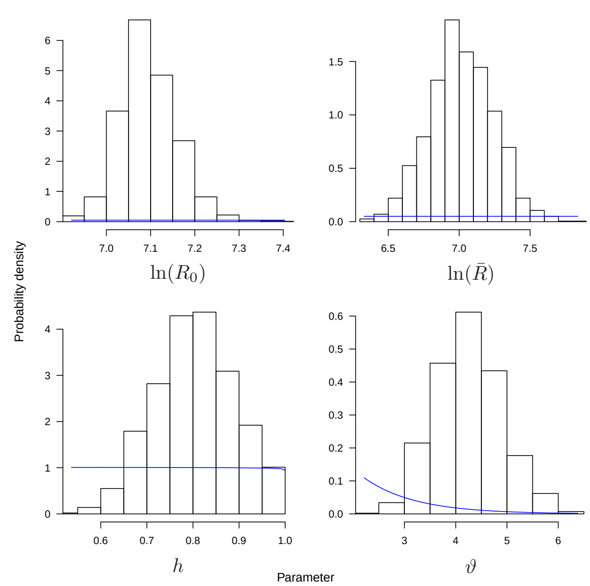

Figure 6: Marginal posterior probability densities (histograms) and prior densities (lines) for unfished recruitment R_0, steepness h, mean recruitment $\bar{R}$

and recruitment compensation κ for the Namibian hake case study.

Marginal posterior densities can also be produced for derived quantities such as MSY based reference points (Fig 7). Again, although not directly comparable due to structural differences between ^{i}SCA$_{\text{M}}$ and the LRGS model, the marginal posterior distributions for MSY and B_0 are very similar. More importantly however is that these marginal distributions can also be used to calculate the probability that the stock is currently overfished and if overfishing is occurring. This is normally represented from a maximum likelihood perspective where the trends in biomass relative to $\text{B}_{\text{MSY}}$and fishing mortality rates relative to $\text{F}_{\text{MSY}}$are plotted against each other (these are known as KOBE plots, Fig 8).

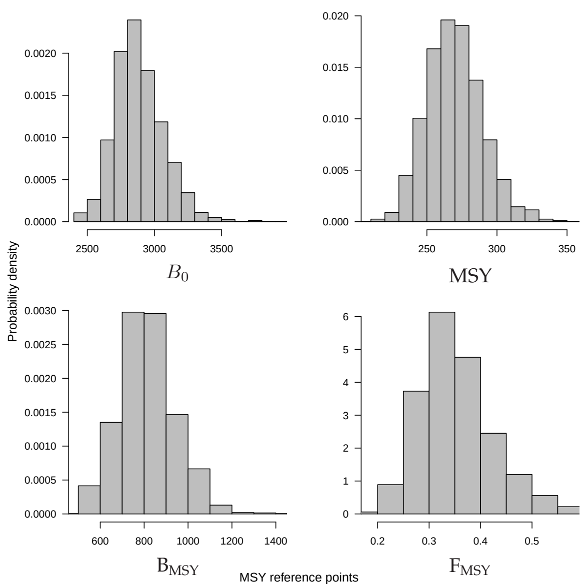

Figure 7: Marginal posterior probability densities for unfished spawning biomass B_0, optimal spawning biomass B_{MSY}, MSY and F_{MSY} for the Namibian hake case study.

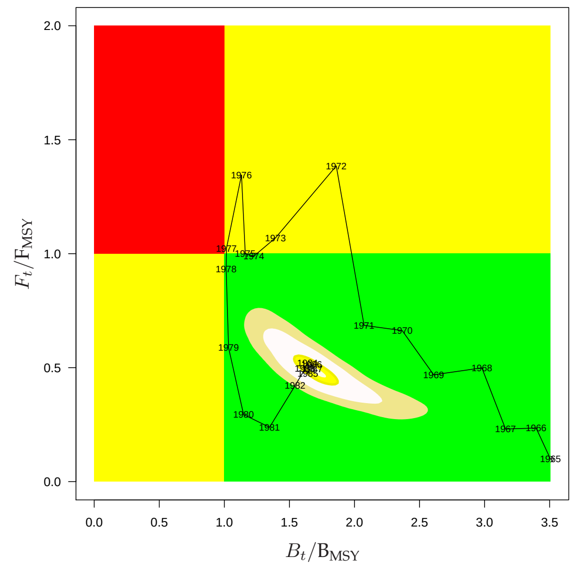

Figure 8: Stock status plot (or Kobe plot) where the "fried egg" represents uncertainty.

6 Example Assessment: the Pacific hake fishery

6.1 Data & assumptions

As more complex example assessment, the data from the Pacific hake fishery is used. Pacific hake (*Merluccius productus*) in the Northeast Pacific has a migratory coastal stock that is harvested by US and Canadian fishing fleets during the summer and late fall. This data is an extension to the previous work in Martell et al. (2008). In this example the data has been restricted to the years 1977-2009, as this was a period when catch-age data from both the Canadian and US fisheries was available and could be aggregated using a weighted average based on the catch proportion from each nation.

The data from this fishery consists of a combined total catch, a relative abundance index from an acoustic survey conducted on a triannual and biannual basis, age-composition data from the commercial fishery, and finally age-composition data from the acoustic trawl survey. The coastal Pacific hake stock undergo an annual migration from spawning grounds in the south near Baja California Sur in the winter to summer feeding grounds to the north; the extent of the northward migration is highly variable and ranges from Oregon–Washington to Southeast Alaska in some years. Larger/older fish tend to migrate further north. Inter-annual variation in the extent of the migration leads to variation in selectivity to the fishery. To accommodate the time-varying selectivity, a total of 85 nodes for a bicubic spline are estimated (17 nodes for the year effect, and 5 nodes for the age effect, see Fig. 3).

In this example it was assumed that recruitment follows a Beverton-Holt relationship, the stock is not at its unfished state in 1977, natural mortality is independent of age and constant over time, and survey selectivity is asymptotic and time-invariant.

Here is the ^{i}SCA$_{\mathrm{M}}$ control file for the Pacific hake data, and the data file is provided at the end of this section on page 24. The observed combined landings from both the US and Canadian zones have averaged about 233,000 metric tons between 1977 and 2009, and in the last 10 years has averaged 270,000 mt with a peak in 1994 of 361,000 mt (Fig. 9.)

```
## ______________________________________________________________________________ ##
##                            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)
## ______________________________________________________________________________ ##


## ________________________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
## ______________________________________________________________________________ ##

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

## _______________________OTHER MISCELLANEOUS CONTROLS___________________________ ##
0           ## verbose ADMB output (0=off, 1=on)
1           ## recruitment model (1=beverton-holt, 2=ricker)
0.05        ## std in observed catches in first phase.
0.01        ## std in observed catches in last phase.
0           ## Assume unfished in first year (0=FALSE, 1=TRUE)
0.01        ## Minimum proportion to consider in age-proportions for dmvlogistic
0.20        ## Mean fishing mortality for regularizing the estimates of Ft
0.01        ## std in mean fishing mortality in first phase
5.00        ## std in mean fishing mortality in last phase
-1          ## phase for estimating m_deviations (use -1 to turn off mdevs)
0.1         ## std in deviations for natural mortality
## ______________________________________________________________________________ ##


## eofc
999
```

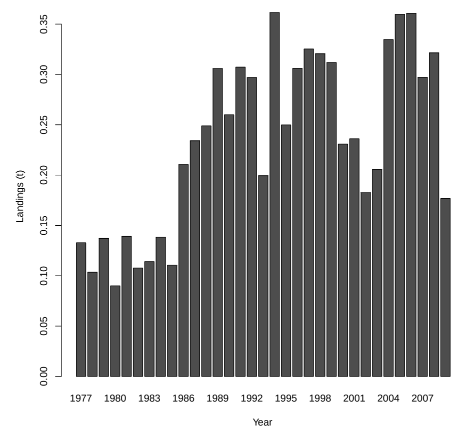

Figure 9: Combined observed landing from the US and CAN fisheries for Pacific hake between 1977 and 2009.

6.2 Maximum likelihood estimates

A total of 176 model parameters were estimated and it took roughly 10 seconds to obtain maximum likelihood estimates, including the calculations for the Hessian matrix on a MacBook Pro, with a 2.66 GHz Intel Core i7 processor.

Maximum likelihood estimates of total biomass and spawning biomass along with estimates of B_0 and $\mathrm{B}_{\mathrm{MSY}}$are shown in Fig. 10. Starting in 1977, estimates of spawning biomass was just slightly less than the estimate of $\mathrm{B}_{\mathrm{MSY}}$. Starting in the 1980's, spawning biomass increased to a maximum in 1990 owing to two very large year classes (1980 and 1984, Fig. 11). Between 1985 and 1999, recruitment ranged between average and median values and the spawning stock biomass declined to less than $\mathrm{B}_{\mathrm{MSY}}$values in 2001 while fisheries removals exceeded 200,000 mt per year. Another significant year class (1999) was responsible for rebuilding the spawning stock biomass up to 2004, and since 2005, the spawning stock biomass has continued to decline.

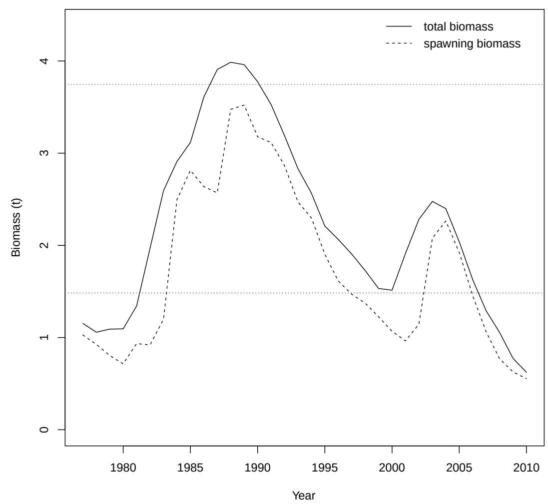

Figure 10: Maximum likelihood estimates of total biomass and spawning stock biomass for Pacific hake along with reference points (dotted lines) for unfished spawning biomass B_0 and $\mathrm{B_{MSY}}$.

Information to estimate age-1 recruitment for Pacific hake comes from the catch-age composition data. Between 1978 and 2009 the average age-1 recruitment is estimated to be 2.72 billion individuals and the median value is 1.16 billion individuals (Fig. 12). The maximum likelihood estimate of the standard deviation in recruitment (τ, see eq. T4.2 on page 10) was 1.29 given the prior information specified in the control file for this assessment.

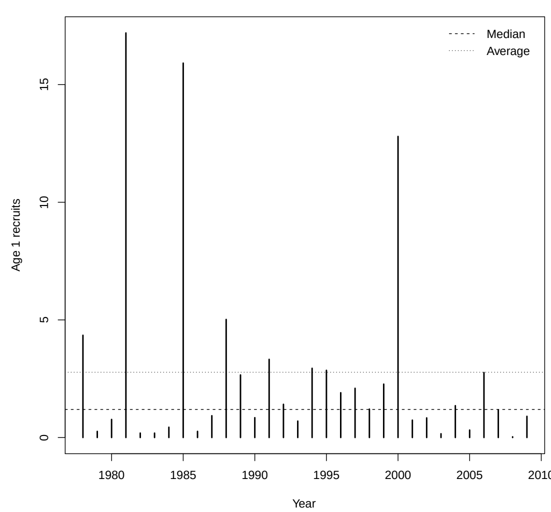

Figure 11: Maximum likelihood estimates of age-1 recruits from 1978 to 2009, with median and average values shown as the horizontal dashed and dotted lines.

Current estimates of stock status relative to $\mathrm{B_{MSY}}$ and the removal rate relative to $\mathrm{F_{MSY}}$is estimated to be in the critical zone in term of the Department of Fisheries and Oceans Canada, Fisheries Management Framework (Fig. 12). Estimates of the spawning stock biomass are less than 80% of $\mathrm{B_{MSY}}$and are currently in the cautious zone. Estimates of fishing mortality rate are roughly 1.5 times the estimate of $\mathrm{F_{MSY}}$. Maximum likelihood estimates of $\mathrm{B_{MSY}}$and $\mathrm{F_{MSY}}$are 1.13 million mt 0.336, respectively.

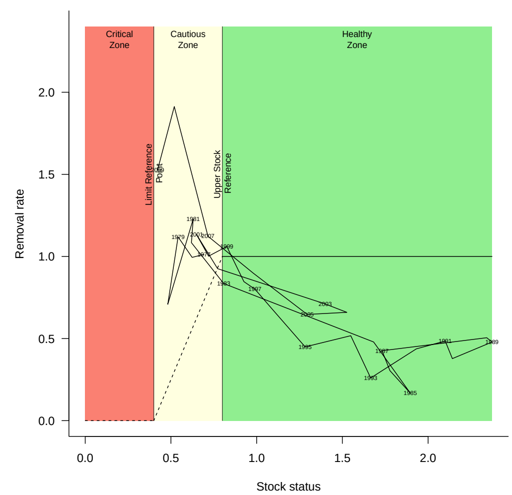

Figure 12: Maximum likelihood estimates of stock status ($B_t/\mathrm{B_{MSY}}$) and removal rate ($F_t/\mathrm{F_{MSY}}$) for Pacific hake relative to the Department of Fisheries and Oceans Canada's Fisheries Management Framework.

Model fit can be partially judged by the residual patterns between the observed and predicted data (Fig. 13). The catch data are assumed to be measured fairly accurately with a small standard deviation ($\sigma_C = 0.025$) in measurement errors; the largest residual in the catch is just less than 100 mt in 1981.

Recall that ^{i}SCA$_\mathrm{M}$ directly estimates annual recruitment values, and the reported residuals in Fig. 13 correspond to the log differences between the estimated recruitment and a Beverton-Holt model prediction where R_0 and steepness h are the estimated parameters for the stock recruitment model. The strong 1980, 1984 and 1999 cohorts, show up as strong positive residuals in 1981, 1985 and 2000 in the residual plot (note that the age-at-recruitment is 1 year). The 2002 and 2004 cohorts appear to be well below the median values in recent years, and the 2005 cohort is currently estimated to be the next largest cohort since 1999.

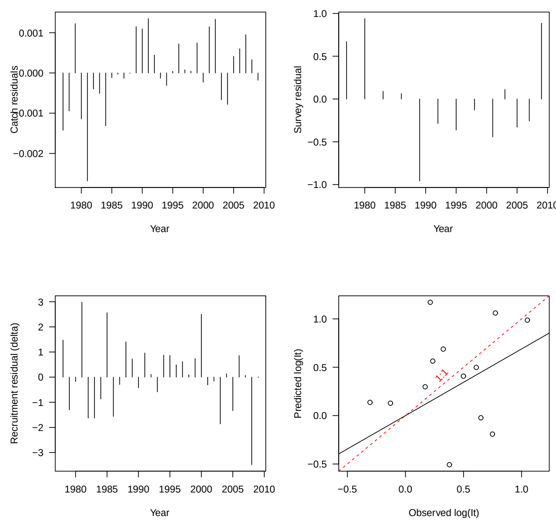

Figure 13: Residuals between the observed and predicted catch, deviations between estimated recruitment and a deterministic Beverton-Holt model, and the observed and predicted relative abundance data from the acoustic survey.

6.3 Time-varying selectivity

Estimates of time-varying selectivity for the commercial fishery were based on estimating 85 nodes (17 years and 5 ages) and interpolating between these nodes using a bicubic spline. The estimated nodes in ^{i}SCA$_\mathrm{M}$ are equidistant, and the total number of estimated nodes is specified in the control file. Increasing the number of estimated nodes should improve the overall fit to the age-composition data; however, this comes at the expense of increasing the associated uncertainty in overall model parameter estimates. To ensure that the model is not over-fitting the data, there are two additional penalties that are added to the objective function that limit the rate of change in age-effects (penalty weight for second differences), and how much dome-shaped is allowed in the age-effects. Increasing the penalty weight on second differences insures a smoothed increase or decrease in the selectivity-at-age, and increasing the weight on the dome-shaped penalty reduces the amount of dome-shaped selectivity that can occur. Again, these penalty weights are specified in the control file in the selectivity parameters section.

In the Pacific hake example, estimates of selectivity increase with age during the late 1970s and early 1980s (Fig. 14). As the 1980 and 1984 cohorts recruit to the fishery, the selectivity shifts to younger ages, and becomes more dome-shaped. At the peak of the spawning stock biomass in 1990, selectivity increases continuously with age, and is more or less asymptotic until the 1999 cohort enters the fishery. Recent estimates of selectivity indicate that the 1999 cohort (age-10 in the year 2009) is still strongly selected for, but as the biomass of the 1999 cohort erodes there is an apparent increase in selectivity for older ages (Fig. 14).

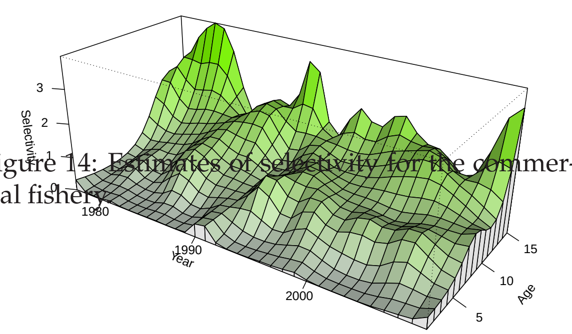

Figure 14: Estimates of selectivity for the commercial fishery.

The residual patterns in the age composition data from the commercial fishery don't appear to have any significant pattern that would indicate a major model mis-specification (Fig. 15). There is a tendency for age-2 proportions to have more negative residuals and age-3 positive residuals, but over all these residuals are fairly small. This is not much of a surprise given the flexibility of the time-varying selectivity that was assumed in the commercial fishery.

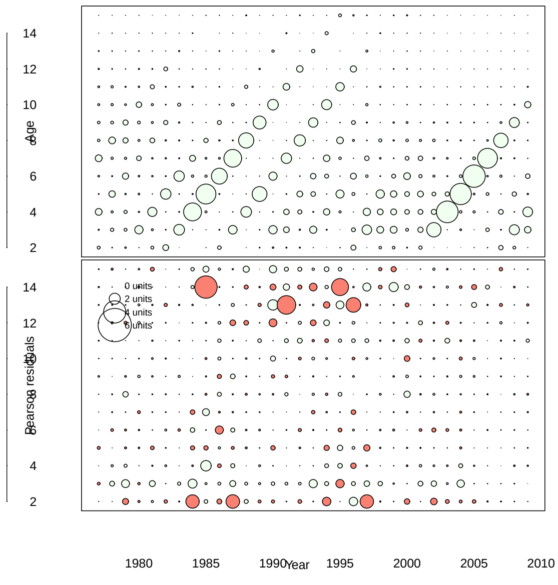

Figure 15: Observed age-composition (top panel) and Pearson residuals between observed and predicted proportions-at-age in the commercial fishery (bottom panel, with negative residuals given by dark circles).

Although not shown here, the residual pattern for the survey age composition also appears to be random, and in this case a time-invariant asymptotic selectivity curve was used for the acoustic survey. Survey data from 1995 to 2007 were assumed to be twice as accurate in comparison to the data collected prior to 1995 when spatial coverage of the survey was incomplete. Also, the 2009 survey carries no weight as this survey was contaminated by the presence of Humboldt squid (*Dosidicus gigas*) during the 2009 survey. Additional details about the data for the Pacific hake assessment and the methods used to aggregate the age-composition for the US and CAN fisheries can be found in Martell (2009).

6.4 Bayesian implementation

To obtain samples from the joint posterior distribution and obtain median values and credible intervals, ^{i}SCA$_{\mathrm{M}}$ was run using the Metropolis-Hastings routine that is built into ADMB. In this example, 2000 systematic samples from a chain of length 1,000,000 was used. The total runtime for conducting a MCMC sample of length 1,000,000 was 39 minutes and 56 seconds with 176 estimated parameters.

The marginal posterior distributions and the corresponding prior distributions are shown in Fig. 16. There is no information in the data about the underlying steepness of the stock recruitment relationship; this is clearly shown by the marginal posterior distribution for h reflects the assumed (*ad hoc*) prior distribution.

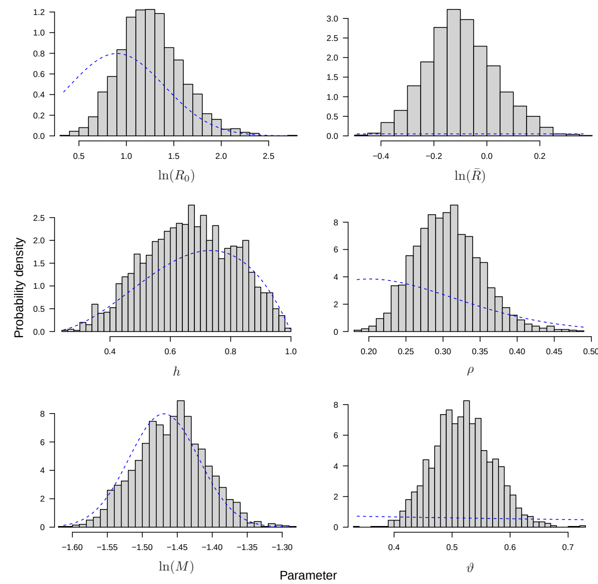

Figure 16: Marginal posterior densities and prior densities for the leading parameters in ^{i}SCA$_{\mathrm{M}}$.

The prior distributions for each of the estimated leading parameters are specified in the control file. In this example, a normal prior was assumed for the unfished recruitment ($\ln(R_0)$) and the log of the natural mortality rate ($\ln(M)$), a beta prior for the steepness (h) and the fraction of the total error that is associated with observation error (ρ), and a non-informative gamma prior for the total precision (ϑ). A uniform prior was specified for the average recruitment ($\ln(\bar{R})$).

Recent trends in the spawning stock biomass, and depletion, along with the associated uncertainty in the form of a credible interval are given in Table 5. Projected estimates of spawning stock depletion at the start of 2010 is 22%, with a lower bound of 7.5% and an upper bound of 53.2%. This translates into a projected spawning stock biomass of 670,000 mt with a 95% credible interval of 255,000 mt to 1,506,000 mt.

Table 5: Median estimate and 5% and 95% credible intervals for spawning stock biomass, and spawning stock depletion. These estimates are based on sampling the joint posterior distribution using MCMC.

	Spawning stock biomass			Depletion		
Year	5%	median	95%	5%	median	95%
2001	0.884	0.997	1.130	0.202	0.336	0.524
2002	1.047	1.194	1.384	0.240	0.403	0.635
2003	1.913	2.150	2.549	0.434	0.727	1.142
2004	2.075	2.340	2.815	0.474	0.795	1.244
2005	1.742	1.994	2.452	0.404	0.679	1.063
2006	1.293	1.524	1.961	0.308	0.521	0.820
2007	0.904	1.139	1.577	0.224	0.389	0.632
2008	0.602	0.848	1.315	0.158	0.288	0.503
2009	0.378	0.729	1.406	0.108	0.242	0.497
2010	0.255	0.670	1.506	0.075	0.222	0.532

Relative to the spawning stock depletion reference points, the median estimate of spawning stock biomass falls in the critical zone (Fig. 17). Estimates of spawn stock depletion is very uncertain; there is a fairly high probability that the stock is also in the critical zone, or less than 40% of $\mathrm{B_{MSY}}$.

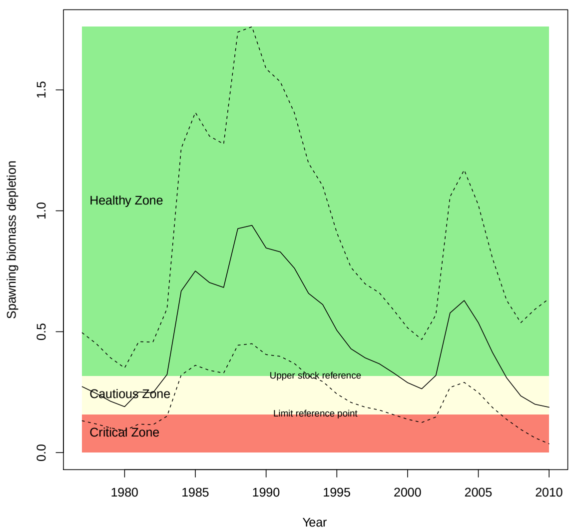

Figure 17: Median estimates of spawning stock depletion and 95% credible interval based 2000 samples from the joint posterior distribution. Transition between the critical, cautious and healthy zones is defined as $0.4\mathrm{B_{MSY}}/B_0$ and $0.8\mathrm{B_{MSY}}/B_0$, respectively

Table 6: Maximum likelihood estimates (MLE) and standard deviations (SD) based on the inverse Hessian for the six leading parameters. Median values and the 95% credible interval based on posterior samples.

	MLE	SD	Median	2.5%	97.5%
$\ln(R_0)$	1.167	0.326	1.238	0.674	1.958
h	0.688	0.214	0.669	0.370	0.932
$\ln(M)$	-1.478	0.049	-1.457	-1.554	-1.363
$\ln(\bar{R})$	-0.168	0.119	-0.103	-0.343	0.177
ρ	0.293	0.043	0.305	0.227	0.405
ϑ	0.525	0.053	0.517	0.422	0.623

Here is the data file for ^{i}SCA$_{\mathrm{M}}$.

```
  #NB The data herein were taken from the 2010 Pacific Hake Assessment using TINSS.
## ________________________
## ____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 fishery (1) or survey (0) in ngears
1 0
## ________________________

## ________________________
#Age-schedule and population parameters
#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
## ________________________

#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
1980 0.089936 0
1981 0.139121 0
1982 0.107734 0
1983 0.113924 0
1984 0.138441 0
1985 0.110401 0
1986 0.210617 0
1987 0.234147 0
1988 0.248804 0
1989 0.305916 0
1990 0.259792 0
```

```
1991 0.307258 0
1992 0.296910 0
1993 0.199435 0
1994 0.361529 0
1995 0.249770 0
1996 0.306075 0
1997 0.325215 0
1998 0.320619 0
1999 0.311855 0
2000 0.230820 0
2001 0.235962 0
2002 0.182911 0
2003 0.205582 0
2004 0.334672 0
2005 0.359661 0
2006 0.360683 0
2007 0.297098 0
2008 0.321546 0
2009 0.176671 0


#Relative Abundance index from fisheries 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
1989 1.238  2  1   0.5
1992 2.169  2  1   0.5
1995 1.385  2  2   0.5
1998 1.185  2  2   0.5
2001 0.737  2  2   0.5
2003 1.840  2  2   0.5
2005 1.265  2  2   0.5
2007 0.879  2  2   0.5
2009 1.460  2  0   0.5

#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
1996    1 0.155201 0.119794 0.007952 0.092629 0.182995 0.011348 0.062955 0.117641 0.007196 0.004787 0.192117 0.000152 0.001182 0.044050
1997    1 0.003320 0.292808 0.225550 0.015138 0.076900 0.137855 0.023476 0.038302 0.073383 0.015622 0.001784 0.063845 0.008784 0.023231
1998    1 0.078999 0.209796 0.176425 0.256621 0.026724 0.051299 0.092465 0.009653 0.017327 0.039036 0.004298 0.001100 0.030568 0.005689
1999    1 0.081647 0.211722 0.181019 0.196830 0.121289 0.024536 0.043645 0.045877 0.009675 0.016122 0.026668 0.006719 0.007091 0.027160
2000    1 0.031229 0.087791 0.141496 0.145797 0.209008 0.117003 0.079190 0.058676 0.020073 0.020671 0.025636 0.014921 0.010884 0.037626
2001    1 0.101860 0.161720 0.147170 0.180056 0.100173 0.138696 0.068275 0.018180 0.019642 0.020480 0.011859 0.011015 0.009238 0.011638
2002    1 0.000361 0.437336 0.158460 0.115792 0.063816 0.051055 0.079963 0.044753 0.010010 0.008598 0.012159 0.001716 0.004788 0.011192
2003    1 0.000558 0.009885 0.662599 0.131923 0.034081 0.055113 0.030260 0.034235 0.019313 0.009771 0.003029 0.005330 0.001132 0.002771
2004    1 0.000371 0.056931 0.078054 0.649859 0.086401 0.023953 0.039404 0.028983 0.013183 0.012676 0.003277 0.002959 0.001744 0.002205
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

#n_wt_obs
0

#eof
999
```

References

ADMB Project (2009). 2009 ad model builder: Automatic differentiation model builder. developed by David Fournier and freely available from admb-project.org.

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.

Gavaris, S. and Ianelli, J. (2002). Statistical Issues in Fisheries' Stock Assessments*. *Scandinavian Journal of Statistics*, 29(2):245–267.

Hilborn, R. and Mangel, M. (1997). *The ecological detective: confronting models with data*. Princeton Univ Pr.

Martell, S. (2009). Assessment and management advice for pacific hake in u.s. and canadian waters in 2009. *DFO Can. Sci. Advis. Sec. Res. Doc.*, 2009/021:iv+54p.

Martell, S. J. D., Pine, W. E., and Walters, C. J. (2008). Parameterizing age-structured models from a fisheries management perspective. *Can. J. Fish. Aquat. Sci.*, 65:1586–1600.

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.

Otter Research (2008). *An introduction to AD Model Builder for use in nonlinear modeling and statistics*. Otter Research Ltd., Nanaimo, B.C.

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.

A Statistical functions & probability distributions

Many of the statistical functions commonly used in R have been written as negative log likelihoods and are in the `stats.cxx` library. In this appendix is the documentation for the available functions in the stats.cxx library. For the most part I have implemented the function based on the description from the R language, so it is possible to use `?function` name in R to learn more about the funciton. Here I provide the formula, the actual code used to implement the function and a description of the variables. Note that some of the functions have been overloaded several times to deal with variables, vectors or a matrix.

dbeta The beta distribution.

$$p(x|a,b) = -\ln(\Gamma(a+b)) + (\ln(\Gamma(a)) + \ln(\Gamma(b))) - (a-1)\ln(x) - (b-1) * \ln(1-x)$$

the mean is given by $a/(a+b)$ and the variance is $\dfrac{ab}{(a+b)^2(a+b+a)}$

```
//beta distribution
dvariable dbeta(const dvariable& x, const double a, const double b)
{
return - gammln(a+b)+(gammln(a)+gammln(b))-(a-1.)*log(x)-(b-1.)*log(1.-x);
}
```

dgamma The gamma distribution.

$$p(x|a,b) = -a\ln(b) + \ln(\Gamma(a)) - (a-1)\ln(x) + bx$$

where the mean and variance are given by $E(x) = a/b$ and $Var(x) = a/b^2$. The following code is implemented in `stats.cxx` library:

```
//gamma
dvariable dgamma(const dvariable &x, const double a, const double b)
{
return -a*log(b)+gammln(a)-(a-1.)*log(x)+b*x;
}
```

dnorm The normal distribution

$$p(x|\mu,\sigma) = 0.5\ln(2\pi) + \ln(\sigma) + 0.5\frac{(x-\mu)^2}{\sigma^2}$$

where the mean is μ and the variance is σ^2.

```
//normal distribution
dvariable dnorm(const dvariable& x, const double& mu, const double& std)
{
double pi=3.141593;
return 0.5*log(2.*pi)+log(std)+0.5*square(x-mu)/(std*std);
}
```

dlnorm The log normal distribution

$$p(x|\mu,\sigma) = 0.5\ln(2\pi) + \ln(\sigma) + \ln(x) + 0.5\frac{(\ln(x)-\mu)^2}{\sigma^2}$$

where the log mean is μ and the log variance is σ^2.

```
//log normal distribution
dvariable dlnorm(const dvariable& x, const double& mu, const double& std)
{
double pi=3.141593;
return 0.5*log(2.*pi)+log(std)+log(x)+square(log(x)-mu)/(2.*std*std);
}
```

B R-code for figures and Tables