Bayesian surplus production model with the Sampling Importance Resampling algorithm (BSP): a User's Guide.

Murdoch K. McAllister[1]
Elizabeth A. Babcock[2]

Summary

The BSP program fits either a Schaefer model or a Fletcher / Schaefer model to CPUE data using the Sampling /Importance Resampling algorithm. Required inputs are catch for all years (missing catch data in the first years of the fishery are allowed), at least one catch rate (CPUE) index of abundance, with CV's if available (missing data are allowed). The parameters that can be fit are carrying capacity (K), the intrinsic rate of population growth (r), the biomass in the first modeled year defined as a ratio of K (*alpha.b0*), the shape parameter for the surplus production function for the Fletcher/Schaefer fit (n), the average annual catch for years before catch data were recorded (*cat0*), , variance parameters for each CPUE series, depending on the method used to weight the CPUE series, and the catchability (q) for each CPUE series. Priors can be used for all of the fitted parameters for a Bayesian fit. The program can be used to project the biomass trajectory under any constant catch or constant F harvest policy, with confidence bounds. Program outputs include decision tables showing the probability of stock rebuilding and other indicators of policy performance at specified time horizons.

[1] Renewable Resources Assessment Group, Department of Environmental Science and Technology, Imperial College, Royal School of Mines Building, Prince Consort Road, London SW7 2BP UK. +44(0)207 5949330 fax +44(0)207 5895319
[2] Wildlife Conservation Society, Marine Conservation Program, 2300 Southern Blvd., Bronx, NY 10460. (718) 220-2151. fax (718) 364-4275

Table of Contents

1 Introduction

This Bayesian Surplus Production model program (*BSP*) fits either a Schaefer model or a Fletcher / Schaefer model to CPUE data using the Sampling /Importance Resampling algorithm (McAllister and Kirkwood 1998, McAllister and Ianelli 1997, Gelman et al 1995, McAllister et al. 1994). Required inputs are catch for all years (missing catch data in the first years of the fishery are allowed), at least one catch rate (CPUE) index of abundance, with CV's if available (missing data are allowed). The parameters that can be fit are carrying capacity (*K*), the intrinsic rate of population growth (*r*), the biomass in the first modeled year defined as a ratio of *K* (*alpha.b0*), the shape parameter for the surplus production function for the Fletcher/Schaefer fit (*n*), the average annual catch for years before catch data were recorded (*cat0*), variance parameters for each CPUE series, depending on the method used to weight the CPUE series (see below), and catchability (*q*) for each CPUE series. Priors can be used for all of the fitted parameters for a Bayesian fit. The program can be used to project the biomass trajectory under any constant catch or constant harvest rate strategy, to produce credibility intervals and decision tables. The program is a stand-alone Windows executable file.

The example input files provided with the code will recreate previously published model runs for Atlantic large coastal sharks in 1998 (NMFS 1998, McAllister et al. 2001c), North and South Atlantic Swordfish in 1999 (McAllister et al. 2000, ICCAT 2000) and Western Atlantic bluefin tuna in 2000 (McAllister et al. 2001a). For applications to other species, it may be necessary to change the source code, for example, to set up prior distributions that were not used in any of the previous model runs. The source code is available from Beth Babcock (bbabcock@wcs.org), as well as input files for other applications (e.g. Babcock and McAllister 2003).

2 Model formulation

2.1 Population dynamics model

The continuous time formulation of the Schaefer surplus production model (Prager 1994) is:

$$\frac{dB_t}{dt} = rB_t - \frac{r}{K}B_t^2 - F_t B_t \quad (1)$$

where *r* is the intrinsic rate of increase, *K* is the carrying capacity, B_t is the abundance, and F_t is the fishing mortality rate at time *t*. In the Schaefer model, the biomass level that sustains the maximum sustainable yield (*Bmsy*) is at one half of *K*.

A generalized version of this model that allows *Bmsy/K* to vary includes a shape parameter, *n*, as well as the parameters *K* (carrying capacity) and *m* (maximum sustainable yield) (Fletcher 1978):

$$\frac{dB_t}{dt} = gm\frac{B_t}{K} - gm\left(\frac{B_t}{K}\right)^n - F_t B_t \quad (2)$$

where

$$g = \frac{n^{n/n-1}}{n-1} \quad (3)$$ and the inflection point is given by:

$$\phi = \frac{B_{MSY}}{K} = \left(\frac{1}{n}\right)^{1/n-1} \quad (4)$$

At $n = 2$, the inflection point occurs at 0.5 *K* and this model is identical with the Schaefer model (Prager 2002). However, this model has the anomalous property of predicting for values of $n \le 1$ ($B_{MSY}/K \le 1/e$) near infinite rates of surplus production per capita as abundance decreases to low levels (Quinn and Deriso 1999; Prager 2002). To address this anomaly the BSP model fits a synthesis of the Fletcher and Schaefer models that can take on sensible

values for *r* at all possible inflection points (referred to here as the Fletcher-Schaefer model). The Fletcher-Shaefer model is a composite model, in which:

(a) For $n > 2$, the original Fletcher model holds. That is, over all values of B_t/ B_{MSY}, the Fletcher model for dB_t/dt is applied. Note that for $n > 2$, $r = gm/K$ and $r < \infty$

(5) $$\frac{dB_t}{dt} = gm\frac{B_t}{K} - gm\left(\frac{B_t}{K}\right)^n - F_t B_t$$

where $g = \frac{n^{n/n-1}}{n-1}$

(b) For n<2 and for $B_t/B_{MSY} > 1$, the Fletcher model also applies.

(c) For n<2 and for $B_t/B_{MSY} \le 1$, the functional form of the Schaefer model holds:

(6) $$\frac{dB_t}{dt} = rB_t - \frac{r}{h}B_t^2 - F_t B_t$$

where *r* is the intrinsic rate of increase, $h = 2\phi K$, and

(7) $$\phi = \frac{B_{MSY}}{K} = \left(\frac{1}{n}\right)^{1/n-1}$$ (as in the original Fletcher model, Equation 2).

In this model, $r < \infty$ for all values of n and B_t/ B_{MSY} and $r = gm/K$ only for $n > 2$. Furthermore, for $B_t/B_{MSY} \le 1$, dB_t/dt can be no more productive than the least productive of the two functional forms for surplus production over the entire range of *n*. See McAllister et al (2000) for details. Note that ϕ is not defined when *n*=1.

2.2 Fitting the model to the data

The BSP model uses a Bayesian parameter estimation method, in which the joint probability distribution of the parameters given the data (called the posterior distribution) is calculated from the likelihood of the data and the prior probability distribution of the data using Bayes theorem (McAllister et al. 1994, Gelman et al. 1995, Berger 1985, McAllister and Ianelli 1997, McAllister et al 2001c, McAllister et al 2002):

(9) $$P(\theta_i \mid y) = \frac{L(y \mid \theta_i)p(\theta_i)}{\int L(y \mid \theta_i)p(\theta_i)d\theta}$$

where $P(\theta_i \mid y)$ is the posterior distribution of the parameter vector θ_i given the data *y*, $p(\theta_i)$ is the prior distribution of the parameters and $L(y \mid \theta_i)$ is the likelihood of data given the parameters.

The BSP model fits the Schaefer or Fletcher/Schaefer model to one or more time series of CPUE data. The CPUE data are assumed to be lognormally distributed about the model predicted values, so that the log likelihood function of the CPUE data is:

(8) $$\ln(L(y \mid \theta_i) = -\sum_{j=1}^{nind}\sum_{t=iyr}^{icur} J_t \frac{0.5}{\sigma_j^2}\left[\log\left(\frac{I_{j,t}}{q_j B_t}\right)\right]^2 - \log(\sigma_j) - 0.5\log(2\pi) - \log(I_{j,t})$$

where $I_{j,t}$ is the CPUE value for index j in year t, q_j is the constant of proportionality or catchability coefficient for CPUE series j and σ_j^2 is the variance in the lognormal likelihood function of data series j, $J_t = 1$ if $I_{j,t} > 0$, otherwise $J_t = 0$. The variance σ_j^2 can vary depending on the weighting method selected (see inputs section below).

The prior probability distribution of the parameters $p(\theta_i)$ can take a variety of forms (see inputs section below).

The BSP program approximates the posterior distribution using a numerical integration method called the sampling-importance resampling (SIR) algorithm. The SIR algorithm is based on the idea that the posterior distribution of a parameter of interest $g(\theta)$ can by reformulated so that it is defined in terms of its expected value ($E^{h(\theta)}$) with respect to an arbitrary probability density function $h(\theta)$, called the importance function (McAllister et al 1994, McAllister et al 2002). The expected value can then be approximated by taking a large number (m) of draws from the importance function, so that the marginal posterior of $g(\theta)$ becomes:

$$E^{P(\theta|y)}\left(g(\theta)\right) \cong \frac{\sum_{k=1}^{m} g(\theta_k) w(\theta_k)}{\sum_{k=1}^{m} w(\theta_k)} \quad (10)$$

where:

$$w(\theta_k) = \frac{L(y \mid \theta_k) p(\theta_k)}{h(\theta_k)} \quad (11)$$

is called the importance ratio or the weight of draw k.

In other words, the importance ratios form an approximation of the posterior distribution, placing on each θ_k the following mass:

$$F(\theta_k \mid y) = \frac{w(\theta_k)}{\sum_{k=1}^{m} w(\theta_k)} \quad (12)$$

The marginal distributions can be calculated directly from the importance draws, or a subsample can be taken to facilitate computation (the resampling part of the SIR).

For a sufficiently large number of draws equation (3) converges on the posterior distribution of $g(\theta)$ for any importance function $h(\theta)$ which allows a non-zero probability of sampling each point in the posterior distribution. In general, an importance function that closely approximates the posterior distribution will be more efficient, meaning that it will converge more rapidly to the posterior distribution. Also, an importance function that places too little weight in the tails of the posterior distribution will be inefficient in approximating the posterior distribution. See the "running the model" section, and McAllister et al (2002) for more information on diagnostics of convergence.

3 Setting up the program

Unzip the program files into any directory, using the folder information so that all the directories will be set up. The directory containing the program files (BSP1.exe) should also contain a required input file called techinputs.txt, as well as directories called /data for the input files, /output for the output files from the model fit, and /decis for the output files from the projection and decision analysis. The /data directory should contain the example input files described in Appendix 3. The /output directory should include files containing the best fit parameter values for the

examples in appendix 3. First-time users are recommended to run the model with one of the example input files before setting up your own input files (see section 5 for how to run the model).

4 Setting up input files

Five input files are required: (1) a file labeled techinputs.txt, (2) the general inputs file, (3) the catch data, (4) the cpue data, (5) a file containing input values of the parameters.

4.1 Techinputs.txt

A file named techinputs.txt must be in the same directory as BSP1.exe. The following parameters are defined in the file techinputs.txt. These values work for all of the example input files, so they do not need to be changed for the example runs.

```
#Default tech inputs
Fmin            0.01
stepsize        0.000001
eps             0.0000001
maxlikefunc     10000
```

Fmin is the smallest value of the fishing mortality rate *F* allowed when using the continuous *F* formulation of the model (Eq. 1). This can be changed to a smaller number if there are any years in the data with high biomass and very low total catch. *Fmin* must be greater than zero. *Stepsize* controls the *fletch* minimization routine's search algorithm. Smaller numbers are more likely to find a local minimum. *Eps* controls how much the parameter values are varied by the *hess* function for calculating the hessian and variance/covariance matrices. *Maxlikefun*c allows the program to throw out importance draws that do not violate any boundary parameters but are extremely unlikely.

4.2 Inputs.txt file

A file containing the inputs to set up the model must be in the */data* directory. It must follow exactly the format shown in Table 1. For variables that are not used in a particular run (for example the variables associated with *cat0* when *cat0* is not being estimated), the value given in the input file has no effect, but a value must be given as a place holder. The input file has two columns (Table 1), column 1 contains the value of the each input specification, and column 2 contains the name of the variable in the source code. The program does not actually use the text in column 2, but expects the second column to exist. See the example input files for the correct format.

The inputs are grouped into five categories:

(1) Files: names of input and output files that are not automatically declared. File names must include the path from the default directory. Some files take a default value if a –1 is entered instead of a path and filename (Table 1).
(2) General specifications: Basic information such as whether the model is Fletcher/Schaefer or Schaefer and the year the fishery started, and how to set up importance sampling.
(3) Parameter specifications. For all parameters, whether the parameter is to be estimated, whether it should be log transformed for fitting (this is advisable for parameters that have different orders of magnitude), and the minimum and maximum allowable value of the parameter,
(4) Set up projections: Sets up the management policies to be projected for decision analysis.
(5) Set up priors: Sets up the priors for each estimated parameter, if a Bayesian method is used.

An example input file with comments is shown in Table 1. More details about the options for model specification follow the table.

Table 1. Components of input file with example values from a run for a Fletcher/Schaefer model for North Atlantic Swordfish (input10d.txt).

Row	Input file column 1 (usually value of	Input file column 2 (usually name of	Description of row/ variable

	variable)	**variable)**	
1	"## Input file for Bayesian surplus production model"		Comment line, must be in quotes if it contains blanks
2	"Swordfish North Atlantic 1999"		Description of run, must be in quotes if it contains blanks
3	**Files**		Comment line
4	10d	fap$	Name of the run for labeling files
5	1	fap2$	Additional identifier of run (usually a one digit number)
6	data/param10d1.out	filparin$	File to input parameter values, including the path from the directory containing the program. A value of -1 will default to output/param##.out, where ## is the value in fap$. You can also choose the parameter input file at run time from the user interface
7	-1	filparout$	File to contain the best fit parameter estimates, including path from the directory containing the program (-1 to default to output/param## .out)
8	data\north_swo_cpue.dat	cpue1f	Input CPUE data, including path from the directory containing the program
9	data\north_swo_cat.DAT	catf$	Input Catch data file, including path from the directory containing the program
10	**"General specifications"**		Comment line, must be in quotes if it contains blanks
11	1	bayesian	Whether to use priors (1 for yes, 0 for no)
12	1	fletcher	Whether to use a shape parameter for the surplus production function (1=yes,0=no, use Schaefer model) (see below)
13	1	F_iterate	1 uses the continuous time *F* formula, 0 uses discrete time (see below)
14	1	impfunc	0 uses a multivariate *t* importance function, 1 draws from the priors as an importance function.
15	1	isetcov	Set to 1 to set covariances in importance function to 0 and set to 0 to use the estimated covariances from the hessian
16	1	expand_imp	The importance function variance will be multiplied by this number if using a multivariate *t* importance function
17	0.00001	degf	degrees of freedom for importance function, must be>0. Smaller numbers imply a wider importance function
18	3	iwted	Method to weight CPUE indices, values of 1 through 10 correspond to the weighting methods explained below
19	1	nind	number of CPUE or relative abundance indices, must be equal to the number of indices in the CPUE data file, must be ≤30
20	1963	ifyrdata%	first year of CPUE data (must be 1940 or later)
21	1998	iendyrdata%	last year of CPUE data
22	1950	ifyr	initial year of model (must be 1940 or later, can be earlier than the first year of CPUE data, and can be earlier than the first year of catch data if cat0 is estimated)
23	1950	ifyrobscat	initial year of catch observations (must be 1940 or later)
24	1999	icur	current year
25	**"Parameter specifications"**		Comment line, must be in quotes if it contains blanks
26	1	estn	1 to estimate shape parameter n, 0 for fixed n
27	1	ltransn	1 to log transform n, 0 otherwise
28	0.01	aminn	minimum allowable value of n
29	5	amaxn	maximum allowable value of n
30	1	estr	1 to estimate population growth rate r, 0 for fixed r
31	1	ltransr	1 to log transform r, 0 otherwise
32	0.01	aminr	minimum allowable value of r
33	5	amaxr	maximum allowable value of r
34	0	estcat0	1 to estimate constant catch between ifyr and ifyrobscat, 0 for fixed cat0
35	0	ltranscat0	1 to log transform cat0, 0 otherwise

36	0	amincat0	minimum allowable value of cat0
37	0	amaxcat0	maximum allowable value of cat0
38	1	estk	1 to estimate carrying capacity, 0 for fixed K
39	1	ltransk	1 to log transform K, 0 otherwise
40	-1	aminK	minimum allowable value of K
41	800000	amaxk	maximum allowable value of K
42	1	estab0	1 to estimate Biomass in ifyr over K, 0 for fixed alpha.b0
43	1	ltransab0	1 to log transform alpha.b0, 0 otherwise
44	0.05	aminab0	minimum allowable value of alpha.b0
45	148.413	amaxab0	maximum allowable value of alpha.b0
46	1	ltransig	1 log transform variance parameters, 0 otherwise
47	0.1	aminsig	minimum allowable value of variance parameters
48	5	amaxsig	maximum allowable value of variance parameters
49	0	estq	whether to estimate q values (1) or use the numerical shortcut (0)
50	0	ltransq	1 log transform q, 0 otherwise
51	0	aminq	minimum allowable value of q
52	0	amaxq	maximum allowable value of q
53	**"Set up projections"**		Comment line, must be in quotes if it contains blanks
54	11	npol	Number of policies to project (can be up to 20)
55	0	0	For each policy, the first column is 0 for constant quotas, 1 for constant F, and 2 for constant F as a fraction of the median Fmsy, and 3 for a constant F that is calculated as a fraction of Fmsy for each simulation The second column is the quota (for case 0), F (for case 1) or fraction of Fmsy (for case 2 and 3) Number of rows must equal npol (in this case 11)
56	0	4000	
57	0	6000	
58	0	8000	
59	0	9000	
60	0	10000	
61	0	10700	
62	0	11800	
63	0	12000	
64	2	0.9	
65	2	1#	
66	1	iDoCIy	1 to calculate confidence bounds/credibility intervals for the B/K trajectory of the projections, 0 otherwise
67	0.1	lci	Lower C.I. bound
68	0.9	uci	Upper C.I. bound
69	0	binvar	Which variable to bin across in decision analysis, 0 is no bins, 1 is r, 2 is MSY
70	0	lowbin	Lower limit of first bin
71	0	binwidth	Width of bins
72	1	ibins	Number of bins plus one for the total expected value (should be one for no binning)
73	10	tyr1	Time horizons for decision analysis. There must be three, in increasing order. The current year (icur) plus the last time horizon must be ≤ 2100.
74	20	tyr2	
75	30	ihz	
76	1993	refyear	Reference year for decision analysis. P(Bfin>Bref) is a performance indicator.
77	**"Set up priors"**		Comment line, must be in quotes if it contains blanks
78	0.875	alphamean	Mean of alpha.b0 lognormal prior
79	0.25	alphasd	Standard deviation of alpha.b0 lognormal prior
80	0	catmean	Mean of cat0 lognormal prior
81	0	catsd	Standard deviation of cat0 lognormal prior

82	1	sigmaprior	1 to use lognormal priors for variance parameters, zero for uniform prior
83	0.3	sigmamed	Median of lognormal prior for variance parameters
84	0.2	sigmasd	Standard deviation of lognormal prior for variance parameters
85	2	rprior	How to calculate prior for r. 0= no prior distribution, 1 = lognormal distribution, 2= t or multivariate t
86	0.41	armean	Mean of r for either the multivariate t or lognormal prior
87	1.382	anmean	Mean of n multivariate t prior
88	0.141	avarn	Variance of n multivarate t prior
89	0.0287	avarr	Variance of r lognormal or multivariate t prior
90	-0.00029	acovrn	Covariance of r and n in multivariate t prior
91	6.62	tdegf	Degrees of freedom for the multivariate t prior of r and n
92	1	logKprior	1 for the prior of K to be uniform on log K, 0 to be uniform on K
93	0	qprior	1 for the prior of q to be uniform on log K, 0 to be uniform on q (only if q is estimated directly instead of by the numerical shortcut)
94	end		The word "end" must be on the last line

4.2.1 bayesian

Setting *bayesian* false (*bayesian*=0) causes the model to use no priors, so that the posterior distribution is calculated entirely from the likelihood of the parameters given the data. This option allows the program to be used for maximum likelihood estimation. Under this option it is not necessary to include the "Set up priors" section of the input file (i.e. the last line before the word "end" can be *refyear*)

4.2.2 fletcher

If *fletcher* is false (*fletcher* = 0) then the model fit is a Shaefer model. Under this option, it is not necessary to input a starting guess for the shape parameter in the parameter input file (Section 3.4). However, the lines in the inputs file (Table 1, lines 26-29 and 87-90 if using priors) referring to the shape parameter must be present.

4.2.3 F_iterate

If *F_iterate* is true (*F_iterate*=1) then the model uses the continuous time version of the Schaefer or Fletcher/Schaefer model (Prager 1994, McAllister et al. 2000). If *F_iterate* is false (*F_iterate*=0), then the discrete-time approximation is used, for the Schaefer model this is:

$$B_{t+1} = rB_t - \frac{r}{K}B_t^2 - C_t \quad (13)$$

were C_t is the catch in time t. In the discrete time version of the model, the CPUE in year t is fit to the biomass in middle of year t, calculated as the average of the biomass in year t and year t+1. For the continuous time version, the CPUE is fit to the average biomass over the year (B=C/F)

For the Fletcher model, if the continuous time option is chosen, the model approximates continuous time by simulating 50 time steps in each year. The F is adjusted until the total catch from the 50 time steps is within 0.000001 of the inputted total catch for that year (Prager 2002).

4.2.4 impfunc, isetcov, expand_imp, degf

If *impfunc*=1, then the importance function used is simply the priors of the parameters. If any of the priors have been given a prior with a Uniform distribution, it is necessary to use limits that are not too far outside the possible values of the parameter. Otherwise, drawing from the priors is very inefficient, and the importance sampling algorithm must throw out most of the draws for being outside the boundary conditions.

If *impfunc* = 0 then the importance function used is a multivariate t-distribution with degrees of freedom defined by *degf*, means at the mode of the posterior distribution, and variances calculated from the Hessian matrix at the mode,

as modified by *isetcov* and *expand_imp*. If *isetcov* = 1, then the covariances of the t-distribution are all set to 0, which makes the importance function wider. The importance function is also multiplied by the value of *expand_imp*, to make it wider. If a multivariate t-distribution is used as an importance function, it is very important to make sure that the importance function is wide enough that it does not underestimate the variance of the parameters (see McAllister et al 2002). During importance sampling, the CVs of the weights should be less than the CVs of the likelihood*priors. If the CV of the weights is more than twice the CV of the likelihood *priors, then the importance function must be expanded with *degf*, *isetcov, expand_imp* or all three (see diagnostics section below).

4.2.5 iwted

Ten methods to weight the CPUE data points (Table 2) are available by changing the value of *iwted* (*iwted*=1 through 10). These methods differ in how they estimate the variance (σ^2) associated with each CPUE data point, and hence the weight that each data point takes in the total likelihood. McAllister et al (2001a) provides more detail about the equations for these likelihood functions, as well as advice for choosing a weighting method. Most of these weighting methods have only been tested using non-Bayesian model fits, but lognormal priors are available.

Table 2. Methods to weight CPUE data points

Weighting method	Description	Parameters to estimate	Notes
1	No weighting or inputted equal weighting	None	All points have the same σ=1
2	Weighted by the MLE estimate of variance σ^2 for each series	None*	This should be identical to iterative reweighting (method 4)
3	The inverse variance method with annual observations proportional to the inputted annual CV^2 and the average variance for each series equal to the MLE estimate	None*	
4	Iterative re-weighting.	σ for each series	The σ for each series is treated as a free parameter
5	Input variances multiplied by a scale parameter	1 scale parameter	
6	Input variance: simply dividing the likelihood function by the inputted variances or CVs	None	
7	Additional variance method, input variances plus a σ^2 that is an estimable parameter for each series	σ for each series	
8	Inverse variance weighting with a variance input for each year	None	analogous to method 6
9	Input variances for each data point plus an estimated scale parameter	1 scale parameter	
10	Equal weighting, but a single sigma is estimated for all points.	1 global σ	

The ten alternatives are listed below and the resulting $\log_e$ of the likelihood functions (Ln*L*) provided. Unless otherwise specified *q* was estimated as:

$$\hat{q}_j = \exp\left(\frac{1}{n}\sum_y \ln(I_{j,y}) - \ln(\hat{B}_y)\right) \quad (14)$$

* A sigma for each series is estimated internally to the model using the numerical shortcut of Walters and Ludwig (1994)

where j refers to the subscript for the series and y to that for years, $I_{j,y}$ is the observed index of abundance for series j in year y, $\hat{q}_j$ is the model predicted constant of proportionality for time series j, and $\hat{B}_y$ is the model predicted abundance in year y

Method 1. No weighting or inputted equal weighting (σ^2 =1)

$$\ln L = -\sum_j \sum_y \frac{\left[\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right]^2}{2\sigma^2} \quad (15)$$

Method 2. Weighted by the MLE estimate of variance $\hat{\sigma}_j^2$ for each series. This is similar to iterative re-weighting (see below).

$$\mathrm{Ln}\, L = -\sum_j \sum_y \left[\frac{\left(\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right)^2}{2\hat{\sigma}_j^2} + \ln(\hat{\sigma}_j) \right] \quad (16)$$

$$\hat{\sigma}_j^2 = \frac{1}{n} \sum_y \left(\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right)^2 \quad (17)$$

Method 3. The inverse variance method with annual observations proportional to the inputted annual CV^2 and the average variance for each series equal to the MLE estimate (NMFS 1998. McAllister et al 2001a).

$$\ln L = -\sum_j \sum_y \frac{0.5}{c_j CV_{j,y}^2 \hat{\sigma}_j^2} \left[\log\left(\frac{I_{j,y}}{q_j B_y} \right) \right]^2 - 0.5 \log\left(c_j CV_{j,y}^2 \hat{\sigma}_j^2 \right) \quad (18)$$

where

$$\hat{q}_j = \exp\left(\frac{\sum_y \left(\ln(I_{j,y}) - \ln(\hat{B}_y)\right) \Big/ c_j CV_{j,y}^2 \hat{\sigma}_j^2}{\sum_y 1 / \left(c_j CV_{j,y}^2 \hat{\sigma}_j^2 \right)} \right) \quad (19)$$

and c_j is a constant for each series whose value is chosen such that the average variance for each data series equals its estimated average variance, $\hat{\sigma}_j^2$.

Method 4. Iterative re-weighting. The sigma for each series is treated as a free parameter

$$\mathrm{Ln}\, L = -\sum_j \sum_y \left[\frac{\left(\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right)^2}{2\hat{\sigma}_j^2} + \ln(\hat{\sigma}_j) \right] \quad (20)$$

Method 5. Input variances re-adjusted by expert judgment, $\sigma'^2_{j,y}$ (same as that in method 9 below), but multiplied by a scale parameter $\hat{c}$

$$\mathrm{Ln}\, L = -\sum_j \sum_y \left[\frac{\left(\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right)^2}{2\left(\hat{c}\sigma'^2_{j,y}\right)} + \ln\left(\sqrt{\hat{c}\sigma'^2_{j,y}} \right) \right] \quad (21)$$

$$(22) \qquad \hat{q}_j = \exp\left(\frac{\sum_y \left(\ln(I_{j,y}) - \ln(\hat{B}_y)\right) / \hat{c}\sigma'^2_{j,y}}{\sum_y 1/\left(\hat{c}\sigma'^2_{j,y}\right)} \right)$$

Method 6. Input variance: simply dividing by the inputted variances or CVs (often done in the VPA):

$$(23) \qquad \ln L = -\sum_j \sum_y \frac{\left[\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right]^2}{CV^2_{j,y}}$$

Method 7. Additional variance method. $\hat{\sigma}^2_{A,j}$ is an estimable parameter for each series

$$(24) \qquad \mathrm{Ln}\, L = -\sum_j \sum_y \left[\frac{\left(\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right)^2}{2\left(\hat{\sigma}^2_{j,y} + \hat{\sigma}^2_{A,j}\right)} + \ln\left(\sqrt{\hat{\sigma}^2_{j,y} + \hat{\sigma}^2_{A,j}} \right) \right]$$

$$(25) \qquad \hat{q}_j = \exp\left(\frac{\sum_y \ln(I_{j,y}) - \ln(\hat{B}_y) / \left(\hat{\sigma}^2_{j,y} + \hat{\sigma}^2_{A,j}\right)}{\sum_y 1/\left(\hat{\sigma}^2_{j,y} + \hat{\sigma}^2_{A,j}\right)} \right)$$

Method 8. Inverse variance weighting with a variance input for each year, analogous to method 6.

$$(26) \qquad \mathrm{Ln}\, L = -\sum_j \sum_y \left[\frac{\left(\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right)^2}{2\sigma^2_{j,y}} + \ln\left(\sigma_{j,y}\right) \right]$$

$$(27) \qquad \hat{q}_j = \exp\left(\frac{\sum_y \left(\ln(I_{j,y}) - \ln(\hat{B}_y)\right) / \left(\sigma^2_{j,y}\right)}{\sum_y 1/\left(\sigma^2_{j,y}\right)} \right)$$

Method 9. Input variances re-adjusted by expert judgment, $\sigma'^2_{j,y}$, plus an estimated scale parameter $\hat{\sigma}^2_A$

$$(28) \qquad \mathrm{Ln}\, L = -\sum_j \sum_y \left[\frac{\left(\ln(I_{j,y}) - \ln(\hat{q}_j \hat{B}_y)\right)^2}{2\left(\sigma'^2_{j,y} + \hat{\sigma}^2_A\right)} + \ln\left(\sqrt{\sigma'^2_{j,y} + \hat{\sigma}^2_A} \right) \right]$$

$$(29) \qquad \hat{q}_j = \exp\left(\frac{\sum_y \left(\ln(I_{j,y}) - \ln(\hat{B}_y)\right) / \left(\sigma'^2_{j,y} + \hat{\sigma}^2_A\right)}{\sum_y 1/\left(\sigma'^2_{j,y} + \hat{\sigma}^2_A\right)} \right)$$

The average input variance of each time series is set according to the scientists' expert judgment on the relative reliabilities of the different series as indices of trends in abundance. The GLM or other estimates of variances for individual data points in a time series are incorporated by making the assigned variance the sum of some unaccounted for variance, a_j, plus the GLM variance for each data point. Thus, an average variance, $\overline{\sigma}^2_j$, is

assigned by expert judgment to each series, j. The input variances, $\sigma'^2_{j,y}$, for each data point are initially adjusted by the estimated sample variances, $\sigma^2_{j,y}$, e.g., from GLM analysis, assuming that the sample variances and the unaccounted for sources of variance are additive.

(30) $$\sigma'^2_{j,y} = a_j + \sigma^2_{j,y}$$

We solve for the constant a_j for each data series such that the mean assigned variance results:

(31) $$\bar{\sigma}^2_j = \frac{1}{n}\sum_{y=1}^{n_j}\left(a_j + \sigma^2_{j,y}\right)$$

Therefore the constant for each series, a_j is obtained by:

(32) $$a_j = \bar{\sigma}^2_j - \frac{1}{n}\sum_{y=1}^{n_j}\left(\sigma^2_{j,y}\right)$$

Thus each input variance is given by:

(33) $$\sigma'^2_{j,y} = \bar{\sigma}^2_j - \frac{1}{n}\sum_{y=1}^{n_j}\left(\sigma^2_{j,y}\right) + \sigma^2_{j,y}$$

Method 10. Estimated global variance. This method is the same as method 1, except that the σ^2 is estimated instead of being fixed at $\sigma^2 = 1$.

4.2.6 Set up projections

The available harvest policies are TAC (poltype=0), constant *F* (Poltype=1), constant fraction of the median Fmsy (Poltype =2), and constant fraction of the Fmsy calculated for each simulation (Poltype=3). For the poltype 1 and 2, the F for each policy is calculated once at the beginning of the policy simulations. For poltype 3, it is reculculated so that the harvest is the same as the MSY harvest level for each simulation. The trajectories predicted by poltype 2 and poltype 3 are very similar, but the probability of rebuilding to *Bmsy* can be quite different.

4.2.7 Set up priors

Setting *bayesian* =0 in the "general parameters" section means that no priors will be used. Not specifying a prior is equivalent to a specifying a uniform prior on each parameter. If *bayesian*=1, then the following priors are available:

lognormal on alpha.b0
lognormal on cat0
lognormal on r
t-distribution on r
multivariate t distribution on r and n
uniform on log (K)
lognormal on the variance parameters (sigma) for some weighting methods
uniform on log(q) if q is estimated

For the lognormal priors of alpha.b0 and cat0, the necessary inputs are the mean and standard deviation of the lognormal distribution. Thus, if you wish to have a lognormal distribution with a mean of mean(x) and a standard deviation of sd(x), the correct inputs are mean(x) and sd(log(x), given by:

$$(34) \qquad sd(\ln(x)) = \sqrt{\ln\left(1+\left(\frac{sd(x)}{mean(x)}\right)^2\right)}$$

For the lognormal prior on r (rprior = 1), the lognormal variance should be entered $(sd(\ln(x))^2)$, instead of the lognormal standard deviation. The mean is the same as for alpha.b0 and cat0. For a t or multivariate t prior on r (rprior=2), the mean and variance of the t distribution must be provided. For the mulitvariate t prior for r and n, the means and variances of each parameter, as well as the covariance of r and n must be provided.

Techniques to develop priors are beyond the scope of this user's guide. See McAllister et al. (2001c), Babcock and McAllister (2003), Punt and Hilborn (1997), Gelman et al. (1995), Hilborn and Mangel (1997) and Myers et al. (2002).

4.3 Catch data file

The catch data must be in the file named catf$ from inputs.txt. The file should have no header. The first column should be the 4 digit year, the second column should be the associated total annual catches. The catches can be in any units. They are generally given in weight, but can be numbers if the CPUE series are also in numbers. If the catch and CPUE data are given in numbers, then the outputs labeled biomass (B/K, etc.) are in numbers not biomass, and K is in units of numbers.

4.4 CPUE data file

The CPUE data must be in a the file named cpue1f$ from inputs.txt. The file should have no header. The first column is the index number of the CPUE series (which should be 1 for the first series, 2 for the second, etc.), the second is the 4 digit year, the third is the CPUE value, and the fourth is the associated CV. The CVs should be set equal to one if no CVs are available. All the CPUE series must be in the same file.

4.5 Parameter inputs

The file named filparin$ in inputs.txt (usually called param###.out), has two different uses. First, it is used to input starting guess of the parameters to be used in calculating the mode of the posterior distribution. Second, after the mode has been calculated, it is used to input the posterior modal estimates of the parameters, which are the input to other components of the model.

While running the "estimate modes" part of the program, you can try more than one starting guess of the parameter values by changing the name of filparin (and filparout if necessary) in inputs.txt. Alternatively, you can chose a parameter input file at run time from the user interface.

Once you have found the mode of the posterior distribution, you must change filparin to be the same file as filparout, so that the correct parameter estimates are input for the other components of the program. The format of filparin$ must be identical to the format of the first few lines of filparout$, which is as follows:

SS	
1	K
2	alpha.b0
3	r
4	cat0
5	n
6	variance parameters
7	q's

The first line contains a single number (the total log likelihood times priors in filparout, not used in filparin, but must be there as a placeholder). The remaining lines must each contain two numbers, the first is the parameter number in filparout (not used in filparin but must be there as a placeholder), the second is a starting guess for each

parameter. For parameters that are in the model but fixed, the starting value given in filparin$ will be the fixed value. Parameters that can be fixed are K, alpha.b0, r and n (if fletcher = 1). Note that n cannot take the value of 1, so do not use 1 as a starting value for n.

Do not include a starting guess for parameters that are not included in the model. If estcat0=0, then there should be no row for cat0. If fletcher=0 there should be no row for n. If the chosen weighting method does not require estimated parameters, there should be no rows for the variance parameters. The order of the parameters is as given above (K, alphaB0, r, cat0,n,variance parameters, q's). See the following examples for clarification.

Example input files

1. For a model estimating r (3), K (1), alpha.B0 (2) and cat0 (4)

0	
1	100000
2	1.0
3	0.2
4	1000

2. Model estimating K (1) and r (3) with a fixed alpha.B0 (2), and using weighting method 10 (estimating a single overall sigma (4)).

0	
1	100000
2	1.0
3	0.2
4	0.01

3. Fletcher model estimating all q's and a variance for each series, with 3 CPUE series. Model is estimating K (1), alpha.B0 (2), r (3), n (4), 3 sigmas (starting guesses 0.1), and 3 q's (starting guesses 0.0001).

0	
1	100000
2	1.0
3	0.2
4	0.01
1	0.1
2	0.1
3	0.1
1	0.0001
2	0.0001
3	0.0001

When using an output file from the model (filparout) as an input file (filparin) the format is correct and does not need to be changed.

5 Running the model

5.1 The User Interface

Run the executable file BSP1.exe. You will be prompted for in the input file described in section 4.2. If the file is found, you will arrive at the main user interface, called "Main fletch". The user interface has a pull down menu labeled "Options". The options are:

Number	Option	Description
0	Generate Random numbers	Generates random number seeds for importance sampling
1	Estimate mode	Produces the best point estimate of the parameters
2	Do Hessian	Calculates the variance/covariance matrix of the parameters
3	Importance Sample	Samples from the importance function to get a numerical approximation of the joint posterior distribution
4	Histograms	Produces marginal posterior distributions of model parameters from the importance samples
5	Subsample	Resamples 5000 of the importance draws to be used in the projections
6	Projections	Projects the population dynamics under specified management policies.
	Run	
	Exit	
A1	Grid histograms	These are extra functions that are not yet available
A2	MCMC	
A3	Find initial values	

Option zero generates a random number file (see below).

Options one through six are the components of the population dynamics program. The components must be run in the order shown, as each uses the output of the previous component. If you select more than one of them, they will run in the correct order. However, component 1 (estimate modes) generally has to be run more than once, and the input parameter file has to be changed to the file containing the output from component 1 before the other components can be run. Each component can also be run individually, as long as the previous components have already been run (i.e. the output from 1. is needed for 2. etc.) The input file is read again every time the program is run. Thus, you do not have to re-run the "estimate modes" component after changing the specifications for importance sampling, and you don't have to re-run the importance sampling if you have changed the specifications for the projections. When in doubt, re-run the earlier components. The output files from each program component are described in more detail in Appendix 2.

Once you have selected the components you would like to run, choose Run (or Control-R) to run the program. Choose exit (or Control-E) to stop the run and exit.

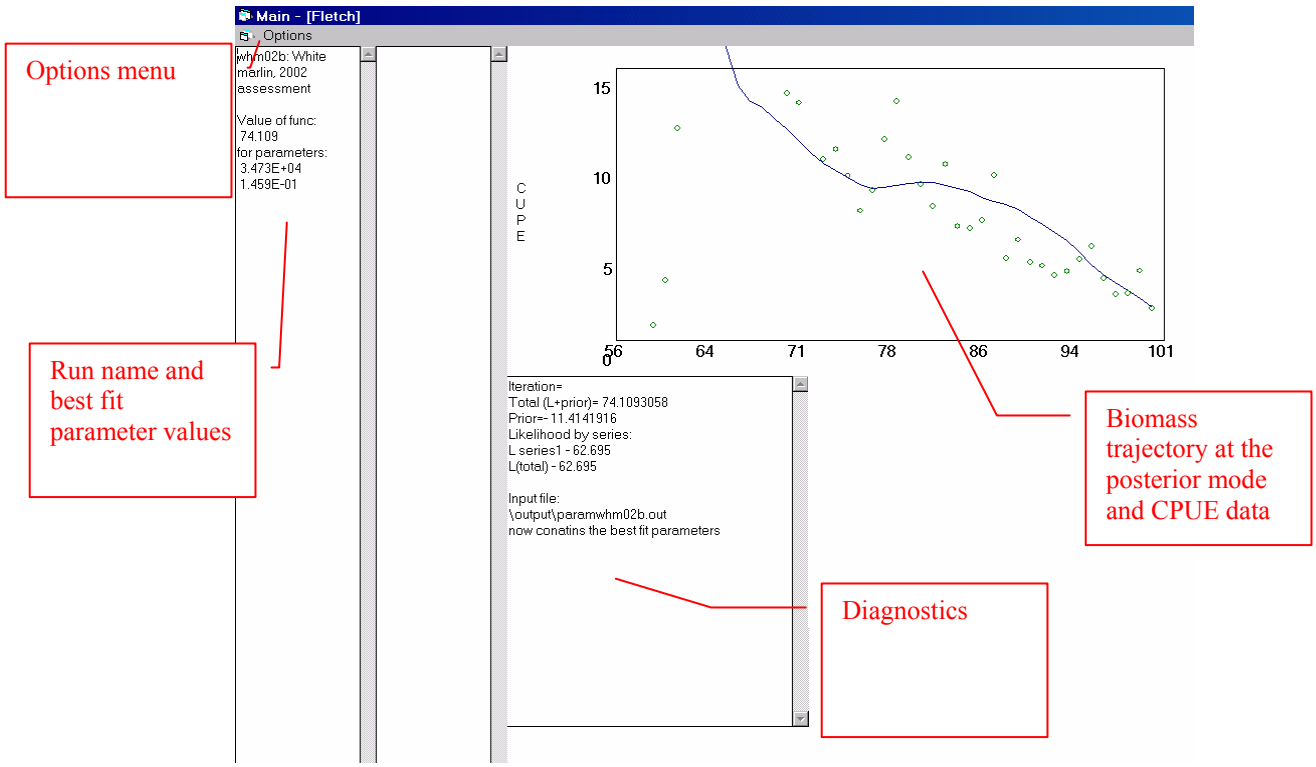
5.2 0. Generate random numbers

This option produces one million double precision random numbers to be used as random seeds in the importance sampling, printed to a file in the /data directory called sam210.dat. This file should be generated once before running the importance sampling subroutine. If the random number seed file is not found the first time the importance sampling subroutine is invoked, the program will automatically generate the random number seed file. The random number seed file does not need to be generated again before each model run, unless you want a different string of random number seeds. Each unique random number determines a unique combination of values for the estimated population dynamics model parameters. When the random number generation is complete, a message appears in the upper left corner of the screen, saying "Random number generation complete".

5.3 1. Estimate modes

When the estimate modes component is called, the program checks for the existence of all the input files and output directories. If a file is missing, you will be prompted for it.

User interface during "estimate mode" operation:



The estimate mode component produces the best point estimate of the parameters, which is the mode of the posterior distribution for Bayesian runs, and the maximum likelihood estimate for non-Bayesian runs. These values are calculated using a gradient search algorithm called *fletch* (Press et al. 1989). The function to be minimized is called *func* (in the main3 module). *Func* returns the value of -ln(likelihood * prior) for a given set of parameter values.

The function minimization algorithm begins with the starting guesses of the parameters from filparin$. The starting guesses of the parameter values must produce a valid biomass trajectory; if the do not, you will receive an error message (see section 6), and be asked which file to use. The easiest way to try other values is to open the starting parameter value file in a text editor, changing the values and re-saving the file. When the BSP program prompts for the file, select "yes". If you select "no" you can open a different starting parameter guess file.

While the fletch algorithm is iterating, the value of *func*, and the associated parameter values are shown in the upper left corner of the screen. The parameters are in the same order as in the parameter input and output files. The text in the middle of the screen shows the total of the natural log of the priors, and the log likelihood for each CPUE series. The graph is the mode of biomass trajectory compared to the original CPUE series.

The best fit estimates are printed to file filparout$. For the rest of the components of the program, filparin$ must contain the parameter values that were output into filparout$ by the 'estimate modes' function. Either use the same name for both filparin$ and filparout$, or run the estimate mode component again selecting filparout$ as the input file before running the rest of the program components. The diagnostics box in the middle of the screen will specify whether the best fit parameters have been printed to the input file, or to a different file. If a different file, the other program components will confirm the parameter input file.

It is worth running the "estimate modes" component several times from several different starting values of the parameters, to make sure that the algorithm has converged on the global mode of the posterior distribution, particularly if the data are uninformative. Once you have a found the mode, it is worth running estimate modes several more times, to increase the precision of the estimate.

The estimate modes component also outputs files with diagnostics of the fit of each CPUE data point. For a run called "10a", the files would be called dev10a.out for the model deviance at each data point and wt10a.out for the estimated variance of each data point, at the mode values of the parameters. These diagnostics are useful for determining the relative weights assigned to various data series, and for comparing the methods of weighting CPUE series (McAllister et al 2001a).

The "estimate modes" component also prints two files that can be used to check the inputs. The file called data\inputcheck##.csv (## is the run label) recreates the input file. If the input file was formatted correctly, then the inputcheck file will be identical to the input file (except for the comment lines). The file called output\priors##.csv produces histograms of all the prior distributions. It is worth plotting the histograms to be sure that priors look as expected.

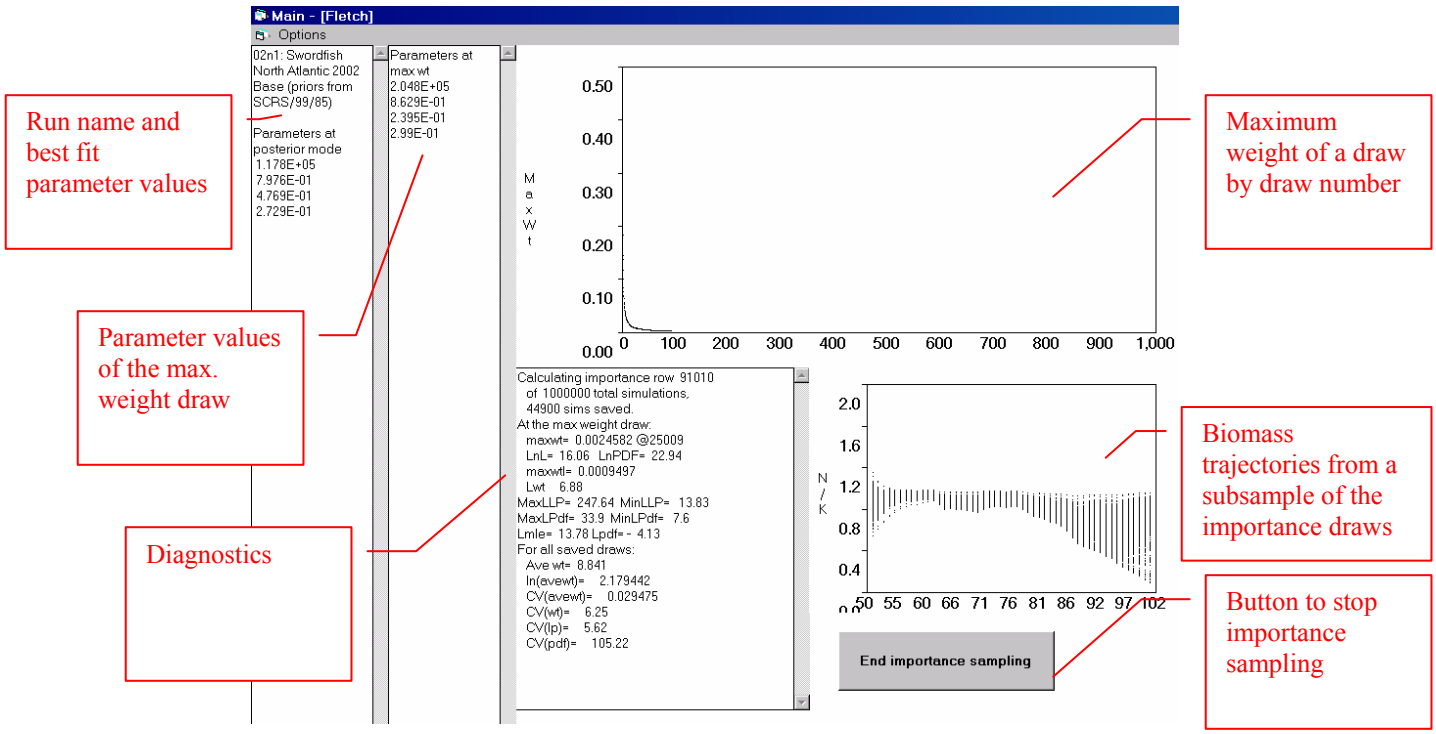
5.4 2. Do Hessian

This function requires that filparin$ contain the best fit parameter estimates that were printed to filparout$ by the "Estimate modes" component. If you select both "estimate modes" and "do hessian", and if the filparin an filparout are not the same file, you will get an error message saying, "Caution: Input parameter values will be used in the Hessian and SIR", and be prompted for the correct input file.

The variance/covariance matrix is output to the file called output\cov10a.out. The Hessian matrix, inverse Hessian, variance/covariance matrix and correlation matrix are all printed to a file called "output\hess10a.out", for a run called 10a. Check that all of the variances are greater than zero and that none of the parameters have more than 99% correlation. If two parameters have very high correlation (more than 99%), importance sampling may be inefficient. Negative variances, or errors in calculation of the hessian (see section 6) generally mean that the estimate modes component did not converge on the true mode of the posterior distribution. Try re-running "estimate modes" from different starting values.

5.5 3. Importance sampling

In importance sampling mode, the user interface looks like this:



The importance sampling function requires the random number file, best estimates of the parameters and the variance/covariance matrix from the hessian as inputs. The available importance functions are a multivariate t distribution as defined in section 4.1.5 above, and the priors.

The information showed on the screen during importance sampling includes the posterior mode values of the parameters (text box on the left), the values at the draw that has the highest weight (second text box). The top graph shows the maximum weight in any individual draw. The bottom graph shows the biomass trajectories. The text box near the graphs shows various diagnostics of convergence.

Example diagnostics screen	**Description**
Calculating importance row 7610	Number of current importance draw
of 1000000 total simulations,	Maximum number of importance draws
3997 sims saved.	How many draws have been saved (draws outside the range of the parameters are not saved. Should be greater than 20,000
At the max weight draw:	
maxwt= 0.0097046 @6008	Number of draw that has the highest weight, and its weight as a fraction of the total importance weight. Should be less than 0.005
LnL= 15.49 LnPDF= 21.65	LnL =ln(likelihood * priors) at the max weight draw LnPDF = ln(importance function) at the max weight draw
maxwtl= 0.0043351	Ratio of likelihood*priors at the max weight draw to the likelihood*priors at the posterior mode draw as a percent of the total. A large number means that the max weight draw is in the tails of the distribution
Lwt= 6.16	Ln(weight) at max weight
MaxLLP= 172.42 MinLLP= 14.11	Maximum and minimum values ln(likelihood*priors)
MaxLPdf= 35.7 MinLPdf= 10.32	Maximum and minimum values of ln(importance function)
Lmle= 14.01 Lpdf= - 11.73	Lmle = ln(likelihood plus priors) at the posterior mode Lpdf = ln (importance function) at the posterior mode
For all saved draws:	
Ave wt= 12.235	Average weight over all saved draws
ln(avewt)= 1.860351	Ln of average weight
CV(avewt)= 0.04655	CV of average weight
CV(wt)= 4.06	CV(wt)= CV of weights
CV(lp)= 3.5	CV of ratio of likelihood*priors at the posterior mode to the likelihood*priors in each draw
CV(pdf)= 23.16	CV of importance function values

The program is set up to make one million importance draws. However, it is usually not necessary to make that many draws to achieve convergence. In general, the joint posterior distribution is sufficiently well estimated when the maximum weight of any draw (shown on the graph and in the diagnostics) is less than about 0.5%. At least 20,000 samples should be saved. Usually, more than 20,000 simulations must be done because draws outside of the range of the priors are not counted. The CV of the weights CV(wt) should be low (preferably less than 5). It is best if the parameter values with the maximum weight (2nd column) are not in the tails of the posterior distribution. Most importantly, the CV of the weights should not be much larger than the CV of the likelihood*priors. See McAllister et al. (2002) for more details on diagnostics of convergence.

When these conditions are met, hit the button labeled "End importance sampling".

For a run called "10a", the outputs are in the subdirectory output in files sim10a.rn (for resampling for decision analysis) and sim10a.bio (to produce histograms of posterior distributions, means, and CVs for fishery quantities of interest from the Bayesian parameter estimation).

The diagnostics from importance sampling can also be used to calculate the relative weights of two different functional forms of model, given the same input data (McAllister et al. 2001b).

5.6 4. Histograms

This function uses the outputs of the importance sampling to produce marginal posterior histograms of various model parameters. A message prints on the screen when the run is finished. The histograms print to a file called hist10a.csv, for a run called 10a.

5.7 5. Subsample

This function subsamples 5000 draws from the importance draws to be used in the projections.

5.8 6. Projections

This function uses the subsampled draws to project the population dynamics into the future. The projections are done by using the parameter values drawn from each of the sub-sampled importance draws, along with the catch data (up to the present year) and the management policy (in the future) to predict the biomass in each year. There is no process error in the projections.

Output is in the directory /decis. The file labeled dec1 contains management parameters of interest at up to three time horizons (often 5, 10 and 15 years from the current year). Dec2 is various parameters binned across r or MSY, and proj contains all of the projection output for the various management variables including ones projected from the beginning of the time series to the 15 years (or whatever the final time horizon is) in the future. If confidence bounds were requested, they are printed in files with names beginning with "ci". The projection subroutine runs must more rapidly if confidence bounds are not requested.

6 Error messages

The following error messages are built into the code.

Warning: total - log likelihood x priors > maxlikefun, discarding draw. This error occurs when a draw is highly unlikely without being outside the boundary conditions (minimum and maximum parameter values, etc.). This occurs most often when estimating q or the variance parameters, and can be avoided by constraining the estimates of q and sigma within more reasonable bounds. This warning does not end execution.

Invalid random number seed/bad random number seed. If this error occurs, run the random number generator, and make sure the file sam210.dat is present, in the /data directory, and is about 15 MByte in size. This error ends execution.

Parameter # less than minimum value / Parameter # greater than maximum value. This error implies that one of the starting guesses for the parameter values is outside the parameter's range. You will be prompted to open the parameter input file again. Check the correct parameters are present, and in the correct order, and within the range specified in the inputs file. Also, check for errors in the input file.

Input parameters too pessimistic. This error means that either the starting parameter guesses imply a biomass trajectory that reaches zero before the end of the time series, or another boundary condition has been met. You will be prompted to open the parameter input file again. Try varying the values.

Warning: Do you want to use parameter values from (filename). This warning confirms the file to be used for the parameter inputs, if it is different from the file used to output the best fit parameters. If you select no, you can choose a different parameter file.

CPUE input file has wrong number of indices. This error ends execution. Check that the number nind in the inputs file is correct, and that the indices in the CPUE input file are correctly numbered.

Negative determinant for r and n. This error stops execution. It implies that the inputted variance/covariance matrix for *r* and *n* is invalid, in Fletcher/Schaefer model runs.

That prior for r is not available. This error stops execution. Only the multivariate t distribution works for r and n in the Fletcher/Schaefer implementation.

Error on matrix inversion. This error from the Hessian subroutine does not stop execution. It implies that the variance/covariance matrix cannot be calculated, generally because the estimates of the parameters input to the hessian subroutine are not the global mode of the posterior distribution. Try re-running "estimate mode" from different starting values.

Input file has too few lines/Input file has two many lines. The program does not reality check the values in the input file, but it produces this error and stops execution if the input file does not have the correct number of inputs. Check that the input file has the lines specified in table 1, and that the last line contains the word "end". Remember to remove the prior section if bayesian = 0.

In addition, the program checks for the existence of the necessary files and paths, and may prompt for files that are missing. Any error not listed above will stop execution of the program, including sharing violations (if a file needed by BSP is already open), division by zero, attempt to take the log of zero, etc.

7 Acknowledgements

This code has evolved from earlier versions written by Jim Ianelli, Ray Hilborn, John Sibert, and others. Many of the subroutines were originally taken from Numerical Recipes. The present applications were developed with the help of many people, including, Michael Prager, Ellen Pikitch, Enric Cortes, Yiota Apostolaki, Ramon Bonfil, Gerry Scott, the members of the 1998 and 2002 Shark Evaluation Workshops and the ICCAT working groups on swordfish, bluefin tuna, billfish, bycatch and methods.

8 References

Babcock, E.A. and M.K. McAllister. 2003. Bayesian methods for accounting for data contradictions in stock assessment of Atlantic white marlin (*Tetrapturus albidus*). ICCAT Coll. Vol. Sci. Pap. 55:541-558.

Fletcher, R. I. 1978. Time-dependent solutions and efficient parameters for stock-production models. *Fishery Bulletin* 76:377-388.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. 1995. Bayesian data analysis. Chapman and Hall. 552 pp.

Hilborn, R. and M. Mangel. 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press. Monographs in Population Biology 28.

ICCAT. 2000. Detailed Report on Swordfish: Report of the ICCAT Swordfish Stock Assessment Session (Madrid, Spain, September 27-October 4, 1999). International Commission for the Conservation of Atlantic Tunas, *ICCAT Collective Volume of Scientific Papers* 51: 1001 - 1209

McAllister, M,K., E.A. Babcock, E.K. Pikitch and R. Bonfil 2002. Importance sampling issues with the 1998 large coastal shark assessement. Document presented at the 2002 Shark Evaluation Workshop.

McAllister, M. K., E. A. Babcock, E. K. Pikitch. 2001a. Evaluating the relative merits of alternative methods to weight different time series of abundance indices in stock assessment. International Commission for the Conservation of Atlantic Tunas, *ICCAT Collective Volume of Scientific Papers* 52: 1094-1115 (SCRS/00/104).

McAllister, M. K., E. A. Babcock, E. K. Pikitch. 2001b. Using Bayesian methods and decision analysis as a rational basis for dealing with conflicting stock assessment results while providing management advice on stock rebuilding. International Commission for the Conservation of Atlantic Tunas, *ICCAT Collective Volume of Scientific Papers* Vol. 52: 1706-1728 (SCRS/00/34).

McAllister, M. K., E. K. Pikitch and E. A. Babcock. 2001c Using demographic methods to construct Bayesian priors for the intrinsic rate of increase in the Schaefer model and implications for stock rebuilding *Canadian Journal of Fisheries and Aquatic Sciences* 58: 1871-1890.

McAllister, M. K., E. A. Babcock, E. K. Pikitch and M. H. Prager. 2000. Application of a non-equilibrium generalized production model to South and North Atlantic Swordfish: combining Bayesian and demographic methods for parameter estimation. International Commission for the Conservation of Atlantic Tuna, *ICCAT Collective Volume of Scientific Papers* Vol. LI: 1523 - 1551 (SCRS/99/85).

McAllister, M.K., and Ianelli, J.N. 1997. Bayesian stock assessment using catch-age data and the Sampling/ Importance Resampling Algorithm. Can. J. of Fish. and Aquat. Sci. 54: 284-300.

McAllister, M.K. and Kirkwood, G.P. 1998. Bayesian stock assessment: a review and example application using the logistic model. ICES J. Mar. Sci. 55: 1031-1060.

McAllister, M.K., Pikitch, E.K., Punt, A.E., and Hilborn, R. 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. Can. J. Fish. Aquat. Sci. 51: 2673-2687.

Myers, R., N.J. Barrowman, R. Hilborn and D. G. Kehler. 2002. Inferring Bayesian priors with limited direct data: applications to risk analysis

National Marine Fisheries Service. 1998. 1988 Report of the Shark Evaluation Workshop. NOAA, NMFS SEFSC. Panama City, FL.

Pella, J.J. and Tomlinson, P.K. 1969. A generalized stock production model. Bull. Inter-Am. Trop. Tuna Comm. 13: 419-496.

Prager, M.H. 1994. A suite of extensions to a nonequilibrium surplus production model. Fishery Bulletin 92: 374-389.

Prager, M.H. 2002. Comparison of the logistic and generalized surplus-production models applied to swordfish, *Xiphias gladius*, in the north Atlantic Ocean. *Fisheries Research*. 58:41-57.

Press, W. H. et al. 1989. Numerical Recipes: The Art of Scientific Computing. Cambridge University Press. New York.

Punt, A.E., and Hilborn, R. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Rev. in Fish Biol. and Fisheries. 7: 35-63.

Walters, C.J. and Ludwig, D. 1994. Calculation of Bayes posterior probability distributions for key population parameters: a simplified approach. Can. J. Fish. Aquat. Sci. 51: 713-722.

9 Appendix 1. Running the program from the source code.

The BSP program source code requires Microsoft Visual BASIC version 6.0 to run. Unzip the program files into any directory, using the folder information so that all the directories will be set up. The directory containing the program files (BayesianSP1.mak, Main.bas, etc.) must also contain directories called /data for the input files, /output for the output files from the model fit and /decis for the output files from the projection and decision analysis. For the program to run correctly, the working directory must be set to the directory containing the program files (for example, C:/BSP/Program/). To set the working directory, set up a new shortcut to Visual BASIC, right click on the Visual BASIC shortcut to open its properties, and type the working directory path in the "start in" box. To open the program, click on the Visual BASIC shortcut, choose the "existing" tab and select "BayesianSP1.mak". To run the program, hit the F5 key.

The Bayesian Surplus Production Visual BASIC project contains six modules, as follows:

1. *Graph* contains subroutines that make the graphs on the user interface. If you are doing a run that generates errors in this module, try commenting out the line 'Call plotfit' where it occurs in the main module. This has no effect on the model outputs, but prevents the graphs from plotting on the user interface.
2. *Histogram* is called when you select "histogram" from the pulldown menu.
The main subroutine in histogram is called *hist*. There is a section of *hist* labeled ' Maximum values in the printed histograms' which it is sometimes useful to change.
3. *Main3* contains most of the subroutines you might need to modify. It includes the following functions:
    genran generates the random number seeds
    func returns the negative log likelihood *priors. The likelihood calculations and the priors section are labeled.
    aproj does the projections.
    *importan*t does the importance sampling.
    ainitmodel, aopenfiles, areaddata and *asettupparams* initialize the model
4. *Optimize* contains various function minimization routines, including the fletch algorithm that is called by the *main* module.
5 and 6. *Stat2* and *Subs* contain various statistical functions, such as functions to generate random numbers from various probability distributions
7. *Subsample* is called when you select "subsample" from the pulldown menu.

When running the program from the source code, if you get the "input parameters too pessimistic" error, try stepping through the function *func* in the *main* module, to see which parameter value is causing the problem (i.e. Put a break point in the first line of *func* (F9), and use F8 to execute *func* line by line). The variable *idontp* is set to 1 when the parameters are too pessimistic.

10 Appendix 2. Output files from the BSP program

This appendix contains example output files for a run called 20a, applied to swordfish in the south Atlantic in 1999.

Input checking files printed by the estimate modes component:

Inputcheck20a.csv. This file recreates the input file that was used for the run.

Comment line	
Swordfish South Atlantic Schaefer 1999	
Files	
20a	fap$
	fap2$
output\param20a.out	filparin$

output\param20a.out	filparout$
data\south_swo_cpue.dat	cpue1f
data\south_swo_cat.DAT	catf$
General specifications	
1	bayesian
0	fletcher
0	F_iterate
0	impfunc
1	isetcov
2	expand_imp
1E-07	degf
4	iwted
2	nind
1970	ifyrdata%
1998	ifyrdata%
1957	ifyr
1957	ifyrobscat
1999	icur
Parameter specifications	
0	estn
1	ltransn
0.01	aminn
5	amaxn
1	estr
0	ltransr
0.01	aminr
5	amaxr
0	estcat0
0	ltranscat0
0	amincat0
0	amaxcat0
1	estk
1	ltransk
3746	amink
800000	amaxk
1	estab0
1	ltransab0
0.05	aminab0
148.413	amaxab0
1	ltransig
0.1	aminsig
5	amaxsig
0	estq
0	ltransq
0	aminq

0	amaxq
Set up projections	
4	npol
0	0
0	13000
0	14600
0	16000
1	iDoCIy
0.1	lci
0.9	uci
2	binvar
10000	lowbin
10000	binwidth
5	ibins
5	tyr(1)
10	tyr(2)
15	tyr(3)
1993	refyear
Priors	
1	alphamean
0.2	alphasd
0	catmean
0	catsd
1	sigmaprior
0.3	sigmamed
0.2	sigmasd
2	rprior
0.405	armean
1.39	anmean
0.1111	avarn
0.0159	avarr
-0.0366	acovrn
3.95	tdegf
1	logKprior
0	qprior
end	

priors20a.csv. This file shows the values of all of informative prior density functions, evaluated at 70 points between the minimum and maximum values of each parameter.

Priors for run20a Swordfish South Atlantic Schaefer 1999	
r prior	
0.01	0.135291
8.13E-02	0.261808

0.152571	0.524985
0.223857	1.05003
0.295143	1.923407
0.366429	2.804359
0.437714	2.849947
0.509	2.00542
0.580286	1.108996
0.651571	0.556247
0.722857	0.276922
0.794143	0.142584
0.865429	7.70E-02
0.936714	4.37E-02
1.008	2.60E-02
1.079286	1.61E-02
1.150571	1.03E-02
1.221857	6.86E-03
1.293143	4.68E-03
1.364429	3.28E-03
1.435714	2.35E-03
1.507	1.72E-03
1.578286	1.28E-03
1.649571	9.66E-04
1.720857	7.41E-04
1.792143	5.75E-04
1.863429	4.52E-04
1.934714	3.60E-04
2.006	2.89E-04
2.077286	2.34E-04
2.148571	1.91E-04
2.219857	1.57E-04
2.291143	1.30E-04
2.362429	1.09E-04
2.433714	9.14E-05
2.505	7.72E-05
2.576286	6.56E-05
2.647571	5.60E-05
2.718857	4.81E-05
2.790143	4.15E-05
2.861429	3.59E-05
2.932714	3.12E-05
3.004	2.72E-05
3.075286	2.38E-05
3.146571	2.09E-05
3.217857	1.85E-05
3.289143	1.63E-05

Output files printed from the "estimate modes' component. All of the information in these files applies to the mode of the posterior distribution.

param20a.out			
5.555664			Sum of –ln(likelihood*priors)
1	150797.7		K
2	0.941685		alpha.b0
3	0.422968		r
1	0.257791		sigma 1
2	0.217082		sigma 2
57	145282.5		Biomass in 1957 (iyr) through 1999 (icur), in the biomass trajectory specified by the parameter values at the mode of the posterior distribution
58	147437.9		
59	148656.4		
60	149090		
61	148788.2		
62	148857.7		
63	148250.2		
64	147279.6		
65	146154.9		
66	146106.3		
67	146451.5		
68	145789.1		
69	143356.3		
70	140922.5		
71	142659.9		
72	143336.1		
73	143257.8		
74	143534.5		
75	143396.5		
76	143561.2		
77	143622.9		
78	143738.9		
79	143291.2		
80	140986.2		
81	140888.1		
82	138357.7		
83	137783.4		
84	133674.3		
85	130509.3		
86	132035.5		
87	132954.6		
88	126659.3		
89	118304		
90	111871.4		
91	110208.9		

92	108954.3		
93	105925.7		
94	99697.99		
95	92333.01		
96	89377.41		
97	86298.97		
98	88435.35		
99	90454.3		
1	7.87E-06	0.2391	For each of 2 CPUE series, the series number, the MLE estimate of catchability (q) and the unweighted MLE estimate of sigma (the standard deviation of the residuals).
2	2.65E-03	0.1001	

sigq20a.out			
1	0.2391	7.87E-06	For each of 2 CPUE series, the series number, the unweighted MLE estimate of sigma and the MLE estimate of catchability (q)
2	0.1001	2.65E-03	

wta20.out			
70	6.65E-02	0	The first column is the 2 digit year. The next column is the weight applied to each data point in CPUE series 1, the next is the weight applied to each data point in CPUE series 2, etc. The weight applied to each point depends on the weighting method used, but it is generally the inverse of the variance estimated for the data point.
71	6.65E-02	0	
72	6.65E-02	0	
73	6.65E-02	0	
74	6.65E-02	0	
75	6.65E-02	0	
76	0	0	
77	0	0	
78	0	0	
79	0	0	
80	0	0	
81	0	0	
82	0	0	
83	0	0	
84	6.65E-02	0	
85	6.65E-02	0	
86	6.65E-02	0	
87	6.65E-02	0	
88	6.65E-02	0	
89	6.65E-02	0	
90	6.65E-02	4.71E-02	
91	6.65E-02	4.71E-02	
92	6.65E-02	4.71E-02	
93	6.65E-02	4.71E-02	
94	6.65E-02	4.71E-02	
95	6.65E-02	4.71E-02	
96	6.65E-02	4.71E-02	
97	6.65E-02	4.71E-02	

98	0	4.71E-02	

dev20a.out

70	0.452381	0	The first column is the two digit year. The next column is the estimated deviance for each residual of CPUE series 1, the next column is the estimated deviance for each residual of CPUE series 2, etc.
71	0.153711	0	
72	0.455691	0	
73	1.374206	0	The sum of the squared deviance residuals can be evaluated using chi-square tests. Values greater than 0.99 or less than 0.01 indicate that the model or the model's likelihood function is mis-specified. Deviance is given by (Gelman et al. 1995):
74	0.779134	0	
75	1.448057	0	
76	0	0	
77	0	0	
78	0	0	$X^2_{n-p} = \sum_i \sum_y \frac{\left(x_{i,y} - E(x_{i,y} \mid \theta)\right)^2}{\mathrm{var}(x_{i,y} \mid \theta)}$
79	0	0	
80	0	0	
81	0	0	where n is the number of data points, p is the number of parameters, $x_{i,y}$ is the observed value of index i in year x, $E(x_{i,y}\mid\theta)$ is the expected value of the index i in year y, given the estimated parameters θ, (i.e. $E(x_{i,y}\mid\theta)$ is the predicted value of index i in year y), and $\mathrm{var}(x_{i,y}\mid\theta)$ is the variance of $x_{,y}$. For lognormally distributed CPUE data, the deviance is calculated as:
82	0	0	
83	0	0	
84	5.244615	0	
85	4.58142	0	
86	0.957582	0	
87	1.399386	0	
88	0.587117	0	
89	0.176528	0	$\chi^2_{n_T-p} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \frac{\left(x_{i,j} - \left(\hat{q}_i \hat{B}_{i,j} \mid \hat{\Theta}\right)\exp\left(\frac{\hat{\sigma}^2_{i,j}}{2}\right)\right)^2}{\left(\exp\left(\hat{\sigma}^2_{i,j}\right)-1\right)\left(\left(\hat{q}_i \hat{B}_{i,j} \mid \hat{\Theta}\right)\exp\left(\frac{\hat{\sigma}^2_{i,j}}{2}\right)\right)^2}$
90	1.52E-02	7.11E-03	
91	0.32933	4.10E-03	
92	0.246773	0.149764	
93	0.080532	0.664038	
94	7.23E-02	2.68E-02	
95	0.115192	0.5646	
96	1.76E-03	0.351118	
97	0.727834	6.20E-02	
98	0	6.22E-03	

cpue20a.out

57	0	0	0	0	0	0	This file prints the input CPUE values, predicted CPUE values and residuals. In the first part of the file, the columns are: 1. 2 digit year 2. input CPUE for index 1 3. predicted CPUE for index 1 4. input CPUE-predicted CPUE for index 1 5. input CPUE for index 2 6. predicted CPUE for index 2 7. input CPUE-predicted CPUE for index 2 The remainder of the file contains the estimated CPUE values for each series in the same format as the input CPUE file. The first column is the index number, the second is the 2-digit
58	0	0	0	0	0	0	
59	0	0	0	0	0	0	
60	0	0	0	0	0	0	
61	0	0	0	0	0	0	
62	0	0	0	0	0	0	
63	0	0	0	0	0	0	
64	0	0	0	0	0	0	
65	0	0	0	0	0	0	
66	0	0	0	0	0	0	
67	0	0	0	0	0	0	

68	0	0	0	0	0	0	year, the third is the predicted CPUE value and the fourth is the input CV value.
69	0	0	0	0	0	0	
70	1.318748	1.118616	0.200132	0	0	0	
71	1.003408	1.115876	-0.11247	0	0	0	
72	0.928282	1.125373	-0.19709	0	0	0	
73	0.782913	1.127726	-0.34481	0	0	0	
74	0.869308	1.128507	-0.2592	0	0	0	
75	0.774613	1.129053	-0.35444	0	0	0	
76	0	0	0	0	0	0	
77	0	0	0	0	0	0	
78	0	0	0	0	0	0	
79	0	0	0	0	0	0	
80	0	0	0	0	0	0	
81	0	0	0	0	0	0	
82	0	0	0	0	0	0	
83	0	0	0	0	0	0	
84	1.713183	1.068166	0.645017	0	0	0	
85	1.626396	1.039543	0.586853	0	0	0	

bio20a.csv									
akv	bval	alphaval	cat0val	rval	ss				akv = K bval = starting biomass (in 1957) alphaval=alpha.b0 cat0val= cat0 rval=r SS=-ln(likelihood * priors) The first 4 columns represent the CPUE series: index=series number esig2(ii) =MLE sigma amleq(ii) =MLE q
150798	142004	0.9417		0.423	5.5557				
index	esig2(ii)	icv(ii)	amleq(ii)	iyr	biom(iyr)	biom(iyr)/k	c(iyr)	fyr	
1	0.239	0	0.000008	57	142004	0.9417	224	0.002	
2	0.1	0	0.002647	58	145282	0.9634	92	0.001	
3	0	0	0	59	147438	0.9777	171	0.001	
4	0	0	0	60	148656	0.9858	459	0.003	
5	0	0	0	61	149090	0.9887	1016	0.007	
6	0	0	0	62	148788	0.9867	769	0.005	
7	0	0	0	63	148858	0.9871	1418	0.01	
8	0	0	0	64	148250	0.9831	2030	0.014	
9	0	0	0	65	147280	0.9767	2578	0.018	
10	0	0	0	66	146155	0.9692	1952	0.013	
11	0	0	0	67	146106	0.9689	1577	0.011	

12	0	0	0	68	146452	0.9712	2448	0.017	The last five columns represent annual values: iyr= 2-digit year biom(iyr)= biomass biom(iyr)/K = biomass as a ratio of K c(iyr) = catch in each year fyr = fishing mortality rate
13	0	0	0	69	145789	0.9668	4481	0.031	
14	0	0	0	70	143356	0.9507	5426	0.038	
15	0	0	0	71	140922	0.9345	2166	0.015	
16	0	0	0	72	142660	0.946	2580	0.018	
17	0	0	0	73	143336	0.9505	3078	0.021	
18	0	0	0	74	143258	0.95	2753	0.019	
19	0	0	0	75	143535	0.9518	3062	0.021	
20	0	0	0	76	143397	0.9509	2812	0.02	
21	0	0	0	77	143561	0.952	2852	0.02	
22	0	0	0	78	143623	0.9524	2774	0.019	
23	0	0	0	79	143739	0.9532	3294	0.023	
24	0	0	0	80	143291	0.9502	5322	0.037	
25	0	0	0	81	140986	0.9349	3978	0.028	
26	0	0	0	82	140888	0.9343	6446	0.046	
27	0	0	0	83	138358	0.9175	5402	0.039	
28	0	0	0	84	137783	0.9137	9139	0.066	
29	0	0	0	85	133674	0.8864	9585	0.072	
30	0	0	0	86	130509	0.8655	5901	0.045	
31	0	0	0	87	132036	0.8756	6029	0.046	
32	0	0	0	88	132955	0.8817	12949	0.097	
33	0	0	0	89	126659	0.8399	16931	0.134	
34	0	0	0	90	118304	0.7845	17215	0.146	
35	0	0	0	91	111871	0.7419	13877	0.124	
36	0	0	0	92	110209	0.7308	13801	0.125	
37	0	0	0	93	108954	0.7225	15816	0.145	

Output files printed from the "Do hessian" component.

cov20a.out					
5.80E-02					Variance/covariance matrix at the mode of the posterior distribution. This file is used as an input to the importance sampling component.
-1.67E-04	2.90E-03				
-2.03E-02	9.17E-05	1.13E-02			
1.27E-03	5.59E-07	9.22E-05	1.68E-02		
-4.70E-03	-2.28E-05	1.25E-04	-2.86E-04	3.57E-02	
2.41E-01	5.39E-02	1.07E-01	1.30E-01	1.89E-01	Sigmas = square roots of the variances

hess20a.out					
Hessian					Hessian matrix at the posterior mode
-4.81E+01					
-9.30E-02	-3.44E+02				
-8.62E+01	2.62E+00	-2.43E+02			
4.00E+00	0.00E+00	7.67E+00	-5.97E+01		
-6.00E+00	-2.42E-01	-1.05E+01	2.15E-02	-2.88E+01	

Inverse	Hessian,	Determinant			Inverse of the Hessian matrix at the posterior mode
-5.80E-02					
1.67E-04	-2.90E-03				
2.03E-02	-9.17E-05	-0.0113402			
-1.27E-03	-5.59E-07	-9.22E-05	-1.68E-02		
4.70E-03	2.28E-05	-1.25E-04	2.86E-04	-3.57E-02	
Covariance	matrix,	sigma			Variance/covariance matrix
5.80E-02					
-1.67E-04	2.90E-03				
-2.03E-02	9.17E-05	1.13E-02			
1.27E-03	5.59E-07	9.22E-05	1.68E-02		
-4.70E-03	-2.28E-05	1.25E-04	-2.86E-04	3.57E-02	
2.41E-01	5.39E-02	1.07E-01	1.30E-01	1.89E-01	Sigmas (square roots of the variances)
correlation	matrix				Correlations of the parameters.
1					
-0.013	1				
-0.793	0.016	1			
0.041	0	0.007	1		
-0.103	-0.002	0.006	-0.012	1	

Output files from the importance sampling. Many of these files have at least one row for each importance draw, so they can be large. Only the first page of each file is shown.

sim20a.rn		
1068374	-14.8816	Each row is one importance draw. The first column is the random number seed. The second is the ln(likelihood*priors). This file is the input for the subsampling algorithm.
964682.9	-13.6236	
1451983	-14.4575	
1865938	-14.1595	
1871509	-17.1154	
1487486	-14.5931	
1616484	-12.1292	
673702.4	-13.9309	
44406.89	-14.1242	
1404010	-16.5127	
713957.4	-15.2732	
215145.6	-14.1654	
793490.3	-13.0672	
1687729	-14.2559	
1492196	-13.6858	
1492246	-14.58	
626740.8	-14.7224	

sim20a.bio											
198230	175676	141692	0.389	19279	0	99115.05	0	16159.43	-14.882	8.473	The sim###.bio file is the input to the histograms
146143	128974	76919	0.3961	14470	0	73071.41	0	14447.97	-13.624	6.688	
154815	146250	65473	0.3216	12445	0	77407.52	0	12208	-14.458	7.663	

197287	164485	152335	0.4573	22554	0	98643.67	0	16643.06	-14.159	9.055	subroutine. Each row is an importance draw, the columns are the parameter values K, Binit(starting year biomass), Bcur (current biomass), r, MSY, n (the shape parameter), Bmsy, cat0, current replacement yield, ln(likelihood), ln(likelood*priors)
220101	261749	118205	0.2295	12630	0	110050.7	0	12544.72	-17.115	8.506	
210410	196105	175173	0.5308	27920	0	105205.2	0	16686.49	-14.593	7.846	
161891	147066	114242	0.4731	19149	0	80945.38	0	16470.3	-12.129	6.227	
124856	107880	59593	0.4573	14274	0	62427.84	0	14226.67	-13.931	7.265	
182983	169942	138395	0.4713	21561	0	91491.47	0	16638.46	-14.124	7.467	
226892	244992	179418	0.4221	23944	0	113446.2	0	16601.97	-16.513	7.949	
185774	161844	145085	0.4983	23141	0	92887.19	0	16720.68	-15.273	7.992	
209321	193299	171478	0.506	26480	0	104660.7	0	16717.28	-14.165	9.944	
184077	180873	132420	0.4262	19612	0	92038.63	0	16361.3	-13.067	6.126	
176908	186754	139763	0.5368	23743	0	88454.1	0	16748	-14.256	7.906	
149754	153748	103109	0.496	18571	0	74876.78	0	16464.61	-13.686	6.484	
121133	128661	60401	0.4853	14695	0	60566.57	0	14687.37	-14.58	8.268	
136239	116460	82796	0.4813	16392	0	68119.69	0	15860.42	-14.722	7.041	

sim20a.dat						
-1000	1	1068374				sim###.dat prints summary information for each draw that has maximum weight when it is drawn. The numbers in the first row are the weight of the draw, the number of the draw, and the random number seed of the draw. The second row is the ln(likelihood), the ln(likelihood*priors) and the ln(importance function). The third row is the initial biomass and the current biomass. The fourth row is K, followed by the parameter vector.
14.88157	8.473	23.355				
175676	141692					
198230.1	198230.1	0.8333	0.389	0.2049	0.2251	
2903915	6	1871509				
17.11535	8.506	25.621				
261749	118205					
220101.3	220101.3	1.2684	0.2295	0.2546	0.2574	
27108155	404	595981.8				
17.38119	6.819	24.2				
238119	149856					
232177	232177	1.037	0.2718	0.2719	0.2074	
35363337	3611	1754629				
17.41524	7.657	25.072				

sim20a.mwt			
rinc=0	df=0.0000001		rinc is a constant that can be added to the importance function, df is the degrees of freedom of the MVT importance function
200	0.0524	1.54	The rest of sim###.mwt prints diagnostics of convergence for every 200 draws. They are the maximum weight of a draw as a percentage of total weight, and the standard deviation of the weights.
400	0.0275	1.49	
600	0.0246	1.51	
800	0.0184	1.49	
1000	0.0144	1.46	
1200	0.0121	1.43	
1400	0.0104	1.41	
1600	0.0092	1.42	

1800	0.0082	1.4	
2000	0.0072	1.41	
2200	0.0065	1.4	
2400	0.0059	1.38	
2600	0.0055	1.37	
2800	0.0051	1.36	
3000	0.0047	1.36	

The histogram subroutine prints only one file, which contains summary information for the run, as well as all the histograms of various parameters.

hist20a.csv						
Posterior histograms for run20a Swordfish South Atlantic Schaefer 1999						
draws	60201	cv(wt)	2.363144	cv(like*prior)	1.308438	Importance sampling diagnostics.
Lwt@maxWt	20.293	LL@maxWt	10.143	Lpdf@maxWt	30.436	
minLwt	-119.155	maxLwt	20.293			
minLL	5.572	maxLL	155.219			
minLpdf	9.099	maxLpdf	36.195			
%maxpWt	0.453147					
%maxLp	0.012362					
log(average_wt)	-55.3157					
E(K)	190655.3	SD(K)	52142.17	CV(K)	0.273489	Summaries of marginal posterior distributions of various parameters
E(r)	0.415785	SD(r)	0.117916	CV(r)	0.283599	
E(MSY)	18985.44	SD(MSY)	5514.221	CV(MSY)	0.290445	
E(Bcur)	130402.1	SD(Bcur)	51189.91	CV(Bcur)	0.392554	
E(Bcur/K)	0.669716	SD(Bcur/K)	0.112839	CV(Bcur/K)	0.168488	
E(binit1)	184639	SD(binit1)	51665.08	CV(binit1)	0.279817	
E(Bcur/binit1)	0.693483	SD(Bcur/binit1)	0.12285	CV(Bcur/binit1)	0.177149	
E(Ccur/MSY)	0.756914	SD(Ccur/MSY)	0.183541	CV(Ccur/MSY)	0.242486	
E(fcur/fmsy)	0.608403	SD	0.267857	CV	0.440262	
E(bcur/bmsy)	1.339431	SD	0.225678	CV	0.168488	
E(ccur/repy)	0.879147	SD	9.83E-02	CV	0.111864	
E(bmsy)	95327.67	SD	26071.09	CV	0.273489	
E(repy)	15454.2	SD	1418.324	CV	9.18E-02	
posterior (K	r)					
	0	0.04	0.08	0.12	0.16	The rest of this file contains histograms of the marginal posterior distribution of the following: r and K K Bmsy r n (the shape parameter) Bcur MSY
0	0	0	0	0	0	
4900	0	0	0	0	0	
9800	0	0	0	0	0	
14700	0	0	0	0	0	
19600	0	0	0	0	0	
24500	0	0	0	0	0	
29400	0	0	0	0	0	
34300	0	0	0	0	0	
39200	0	0	0	0	0	
44100	0	0	0	0	0	

49000	0	0	0	0	0	Replacement yield
53900	0	0	0	0	0	Bcur/K
58800	0	0	0	0	0	Bmsy/K
63700	0	0	0	0	0	Bcur/Binit
68600	0	0	0	0	0	Binit
73500	0	0	0	0	0	Fcur/Fmsy
78400	0	0	0	0	0	Bcur/Bmsy
83300	0	0	0	0	0	cat0
88200	0	0	0	0	0	
93100	0	0	0	0	0	
98000	0	0	0	0	0	
102900	0	0	0	0	0	
107800	0	0	0	0	0	

The subsampling component prints one output file, which is input to the projections component.

spec20a.out		
1404010	1411828	spec###.out contains 5000 lines for the 5000 subsampled draws. The first number on each line is the random number seed for the run.
1404010	523537	
1138113	865561	
688021.4	487561	
688021.4	905773	
289802.6	1270272	
1725437	26838	
3517.866	166342	
209793.1	1824273	
1970803	745255	
312108.2	1943483	
312108.2	1709095	
309895.2	297700	
1206343	1780042	
1312835	125463	

The projections component prints output files to the directory called /decis.

dec120ap.csv									
Horizon	Policy	E(Bfin/ K)	E(Bfin/ Bmsy)	P(Bfin< 0.2K)	P(Bfin> Bmsy)	P(Bfin> Bcur)	P(Ffin< Fcur)	P(Bcur >Bref)	dec1###.csv is a decision table. The columns are the time horizons (in years from icur), the management policies (total catch: TAC=##, constant fishing mortality rate: F=## or constant proportion of Fmsy : F=Fmsy*##), and the measures of policy performance. They are: The expected value of the final
5 -year	TAC= 0	0.94	1.87	0	1	1	1	1	
	TAC= 13000	0.71	1.42	0	0.93	0.93	0.95	0.37	
	TAC= 14600	0.68	1.35	0	0.9	0.76	0.07	0.07	
	TAC= 16000	0.64	1.29	0.01	0.85	0.03	0	0	

10 - year	TAC= 0	0.99	1.97	0	1	1	1	1	biomass over K, expected value of the final biomass over Bmsy, the probability that the final biomass is below 20% of K, the probability of recovery to the Bmsy level, probability of population increase, probability that the final fishing mortality rate is below the current F, and the probability that the biomass is above the level in a reference year defined in the inputs.
	TAC= 13000	0.72	1.44	0.01	0.94	0.93	0.94	0.54	
	TAC= 14600	0.67	1.34	0.02	0.87	0.76	0.15	0.09	
	TAC= 16000	0.62	1.24	0.04	0.8	0.03	0	0	
15 - year	TAC= 0	1	1.99	0	1	1	1	1	
	TAC= 13000	0.73	1.45	0.01	0.94	0.93	0.94	0.62	
	TAC= 14600	0.66	1.32	0.04	0.86	0.76	0.18	0.1	
	TAC= 16000	0.6	1.19	0.09	0.77	0.03	0	0	

dec220ap.csv							
E[cat/yr] 5 -year							Dec2###.csv presents decision tables of various performance indicators (in this case, average annual catch) at the time horizons(in this case 5 years) binned across values of either r or (as in this case) MSY. The first two rows are the values of MSY and the probability that MSY is in that bin. The rest of the rows are the harvest policies. The columns are the values of the parameter in each bin, and the last column is the expected value across bins. The performance indicators presented are the same as the in dec1##.csv, with the addition of the expected average annual catch
	value for MSY						
	<10000.0000	10000	20000	30000	40000.0000+	E(X)	
prob. per bin	0.002	0.6746	0.2796	0.0382	0.0056		
0	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	
13000	1.30E+04	1.30E+04	1.30E+04	1.30E+04	1.30E+04	1.30E+04	
14600	1.46E+04	1.46E+04	1.46E+04	1.46E+04	1.46E+04	1.46E+04	
16000	1.57E+04	1.60E+04	1.60E+04	1.60E+04	1.60E+04	1.60E+04	
E[cat/yr] 10 -year							
	value for MSY						
	<10000.0000	10000	20000	30000	40000.0000+	E(X)	
prob. per bin	0.002	0.6746	0.2796	0.0382	0.0056		
0	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	
13000	1.24E+04	1.30E+04	1.30E+04	1.30E+04	1.30E+04	1.30E+04	
14600	1.29E+04	1.46E+04	1.46E+04	1.46E+04	1.46E+04	1.46E+04	
16000	1.34E+04	1.59E+04	1.60E+04	1.60E+04	1.60E+04	1.59E+04	
E[cat/yr] 15 -year							
	value for MSY						
	<10000.0000	10000	20000	30000	40000.0000+	E(X)	
prob. per bin	0.002	0.6746	0.2796	0.0382	0.0056		
0	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	
13000	1.09E+04	1.30E+04	1.30E+04	1.30E+04	1.30E+04	1.30E+04	

14600	1.12E+04	1.45E+04	1.46E+04	1.46E+04	1.46E+04	1.45E+04	
16000	1.05E+04	1.56E+04	1.60E+04	1.60E+04	1.60E+04	1.57E+04	

ci20a1.csv						
North Stock: Fletcher-Schaefer						Confidence bounds for biomass and other performance indications. Each management policy is in a separate file, labeled ci20a1.csv, ci20a2.csv, etc.
Prior CV on B0/K = 0.2						
Prior CV on r = 0.0159						
Prior CV on n = 0.1111						
Number of simulations = 5000						
policy 1: tac = 0						
Stock biomass: Medians and 80 %CIs and Rescaled CPUEs Using Median q's						Trajectories from the beginning of the fishery to the time horizon of projections for each policy.
57	117787.2	168466.2	226844.2	0	0	
58	122467.2	169088.1	221653.7	0	0	
59	124230.9	170617.1	220506.3	0	0	
60	124936.7	171445.4	219868.5	0	0	The first table under each policy contains the 2-digit year, lower limit of the credibility interval, median and upper limit of the credibility interval for biomass, followed by the CPUEs of each series divided by the median value of q so that they are rescaled to biomass units.
61	124922.2	171634.4	219272.9	0	0	
62	124736.3	171717.3	219072.8	0	0	
63	124519.7	171441.3	218629.6	0	0	
64	123823.5	170799.6	217908.5	0	0	
65	122861.9	169744.4	216828.3	0	0	
66	122377.2	169059.2	216156.4	0	0	
67	122646.4	169149.4	216220.7	0	0	
68	122496.5	169031.9	216006.2	0	0	
69	121000.9	167431.8	214383	0	0	
70	118765.7	164979.9	211844.7	195140.7	0	The rest of the tables for each policy are medians and credibility intervals for: Biomass/Bmsy Fishing mortality rate(F) F/Fmsy
71	118688.1	164522.1	211387.2	148478.5	0	
72	119985.6	165612.2	212429.8	137361.8	0	
73	120280.7	165904	212647.7	115850.9	0	
74	120369.9	166002.1	212721.3	128635.2	0	
75	120428.5	166071.1	212750.9	114622.8	0	
76	120432.1	166052	212767.3	0	0	
77	120539.1	166147.4	212883.3	0	0	
78	120618.7	166216.7	212974.5	0	0	
79	120443	166061.1	212813.4	0	0	

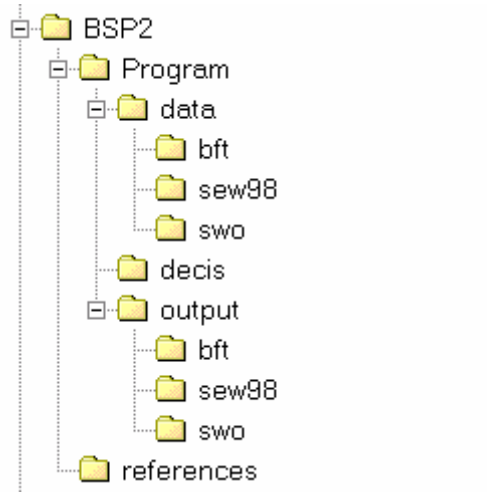
proj20ap.csv							
Swordfish South Atlantic Schaefer 1999							Run summary, proj##.csv contains statistics for each time horizon binned across MSY (or r)
Prior CV on B0/K = 0.2							
Prior mean r = 0.405 Prior var on r = 0.0159							
Number of simulations = 5000							
policy 1: tac = 0							Summary statistics for

							the first policy. All policies are in the same file.
cvK = 0.259							
median K = 174041							
cvBf = 0.259							
median Bfin = 173755							
cv(Bfin/K) = 0.004							
cv(Bfin/Bmsy) = 0.004							
median Bfin/Bmsy = 1.9996							
80 % lower Bfin/Bmsy = 1.9903							
80 % upper Bfin/Bmsy = 2.0000							
median Bfin/K = 0.9998							
80 % lower Bfin/K = 0.9951							
80 % upper Bfin/K = 1.0000							
	10000	10000	20000	30000	40000	E(X)	Summary of various indices of policy performance binned across MSY, for each time horizon.
prob. per bin	0	0.7442	0.2426	0.0106	0.0026		
5-year time horizon							
E[cat/yr]	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	
E[B 2004 in kt]	0.00E+00	1.53E+05	1.92E+05	2.59E+05	6.33E+05	1.65E+05	
E[B 2004 / K]	0.00E+00	9.24E-01	9.84E-01	9.92E-01	9.88E-01	9.40E-01	
E[B 2004 / Bmsy]	0.00E+00	1.85E+00	1.97E+00	1.98E+00	1.98E+00	1.88E+00	
P[Bfin<0.2K]	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	
P[Bfin>MSYL]	0.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	
P[Bfin>Bcur]	0.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	
P[Ffin<Fcur]	0.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	
P(Bfin>B1993)=	0.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	
10-year time horizon							
E[cat/yr]	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	
E[B 2009 in kt]	0.00E+00	1.64E+05	1.95E+05	2.61E+05	6.40E+05	1.74E+05	
E[B 2009 / K]	0.00E+00	9.88E-01	9.99E-01	9.98E-01	9.98E-01	9.90E-01	
E[B 2009 / Bmsy]	0.00E+00	1.98E+00	2.00E+00	2.00E+00	2.00E+00	1.98E+00	
P[Bfin<0.2K]	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	
P[Bfin>MSYL]	0.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	
P[Bfin>Bcur]	0.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	
P[Ffin<Fcur]	0.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	

11 Appendix 3. Example input files.

The input files provided with the program recreate the model runs described in the 1999 ICCAT Swordfish Assessment (ICCAT 2000), the 1998 U.S. Atlantic large coastal shark assessment (NMFS 1998), a paper presented at the 2000 ICCAT bluefin tuna assessment (McAllister et al. 2001a), some of the shark sensitivity analyses in McAllister et al. (2002). Input files for other assessments that have been done with this program are available from Beth Babcock (bbabcock@wcs.org).

The directory structure looks like this:



The example files are in the directories labeled BFT (bluefin tuna), sew98 (sharks) and swo (swordfish). To run the example applications, move the data files for the species you want to the \data directory, and the parameter files to the \output directory.

Large coastal shark example

cat98base.csv	Large coastal shark baseline catch
cpue98base.csv	Large coastal CPUE
cpue98SB.csv	Sandbar CPUE
cat98sens.csv	Large coastal shark alternative catch
cat98BT.csv	Blacktip baseline catch
cat98BT2.csv	Blacktip alternative catch
cat98SB.csv	Sandbar baseline catch
cat98SB2.csv	Sandbar alternative catch
cpue98BT2.csv	Blacktip CPUE
inputsSBbasepriors.txt	Inputs for the sandbar baseline drawing from the priors
inputsLCSbasepriors.txt	Inputs for the large coastal baseline drawing from the priors
inputsBTbasepriors.txt	Inputs for the blacktip baseline drawing from the priors

These input files reproduce the runs with the prior used as the importance function in McAllister et al 2002. To reproduce the SEW 1998 runs change the importance function as described in McAllister et al 2002.

Swordfish example

The input files provided in the /data/swo directory are as follows:

north_swo_cat.DAT	Swordfish catch in the north Atlantic
north_swo_cpue.dat	Swordfish CPUE in the north Atlantic
south_swo_cat.DAT	Swordfish catch in the south Atlantic
south_swo_cpue.dat	Swordfish CPUE in the south Atlantic
inputs10d.txt	North Atlantic "base case" Fletcher/Schaefer model inputs, the prior median of Bmsy/K is 0.43
inputs20a.txt	South Atlantic "base case" Schaefer model inputs

Bluefin tuna example

McAllister et al. (2001a) presented a variety of runs with different weighting methods for western Atlantic bluefin tuna. The model was not Bayesian, and only the MLE estimates were presented. The files provided under /data/bft are:

cat1.csv	Catches of western Atlantic bluefin tuna
cpue1.csv	CPUE indices (7 series) for western Atalntic bluefin tun
inputsb.txt	Input file

To reproduce the runs in McAllister et al. (2001a), change the value of iwted. Note that the numbering of the weighting options is different between the model and the original paper. The user's guide is correct.