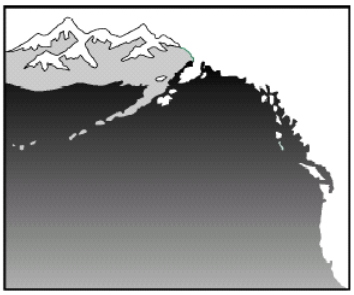
SAFS-UW-0116
Revised May 2003

COLERAINE
A Generalized Age-Structured Stock Assessment Model

USER'S MANUAL VERSION 2.0

R HILBORN, M MAUNDER, A PARMA, B ERNST, J PAYNE, P STARR



University of Washington
SCHOOL OF AQUATIC & FISHERY SCIENCES

CONTENTS

1. Introduction

1.1. General Overview of the Program Coleraine

Coleraine is a user-friendly, general age-structured model for fisheries stock assessment. It combines a familiar Excel environment with a general and powerful AD Model Builder (Otter Research Ltd. 1999) application.

Coleraine has a statistical age- and sex-structured model with a very general structure. It allows for several fisheries to be modeled at once and can be simultaneously fitted to many different sources of information, like catch-at-age and/or -size data from the fishing fleet, and surveys and several indices of abundance (commercial fishery and survey). The estimation is performed using maximum likelihood theory in a first step and a Bayesian approach in a second.

Prior information on model parameters may be incorporated, given the Bayesian framework of this statistical approach. Uncertainty around the estimates of the derived parameters of interest can be assessed directly from the bayesian posteriors.

Once the model is fitted, *Coleraine* allows the user to do policy evaluation by assessing the consequences of different harvest strategies (harvest rates or catch levels) on certain statistics of interest (e.g., predicted vulnerable biomass), which are reported as Bayesian posteriors.

Other salient features of this model are as follows:

- Temporal changes in the selectivity of the fishing fleet.
- Temporal changes in the catchability of the fishing fleet.
- Survey selectivity modeled as age-or size-based.
- The model simultaneously fitted to length and age data.
- Robust normal likelihood function for proportions
- Automated process for saving condensed information on different runs.

1.2. General Overview of the Estimation Model

A general description of the different components of the estimation model (*Colera20.exe*) is presented in the following sections of this manual. The following notation is used throughout Section 1:

Subscripts:	*a*	Age
	l	Length
	t	Time
Superscripts:	*g*	Gear (Fishery or Survey)
	s	Sex

1.2.1. *Abundance dynamics by sex*

Abundance at age and sex is propagated according to the following difference equation

$$N^s_{a+1,t+1} = N^s_{a,t} e^{-M^s}\left(1 - u^s_{a,t}\right) \qquad \text{for } a = 1, \cdots, A$$

where *M* is the instantaneous rate of natural mortality, age *A* is a “plus group", and $u^s_{a,t}$ is an age-specific exploitation rate for all gears combined, which is obtained by summing over all gear types

$$u^s_{a,t} = \sum_g u^{s,g}_{a,t}$$

The exploitation rate for each gear is a product of its age-specific selectivity, $s^{s,g}_{a,t}$, and the exploitation rate of fully selected fish at a specific time

$$u^{s,g}_{a,t} = s^{s,g}_{a,t}\, u^g_t$$

Formulations below are identical whether *g* refers to a fishery component or to a survey, except that the mortality induced by the surveys is negligible and can be ignored. The alternative approaches used for the selectivity function are explained in a later section.

Assuming that total commercial catches in biomass for each gear C^g_t are known without error, and that fishing takes place in a short time interval in the middle of the year, the annual exploitation rate by gear is given by

$$u^g_t = \frac{C^g_t}{\sum_s e^{-0.5M^S} \sum_a s^{s,g}_{a,t}\, N^s_{a,t}\, w^s_{a,t}}$$

which is equal to the ratio of total catch to vulnerable biomass in the middle of the year.

1.2.2. *Initial conditions*

The initial condition assumptions built into the model allow for the estimation of three parameters: $\boldsymbol{R}_0$ (virgin recruitment), $\boldsymbol{\omega}$ (fraction of $\boldsymbol{R}_0$ in the first year), and u_0 (exploitation rate for the first year). The initial vulnerability-at-age pattern by sex has to be incorporated by the user in the "Fixed Parameter Section" (item 13). Also the fraction of $N_{1,1}$ and more generally $N_{1,j}$ (*j* = year) that recruits to each sex is represented by a user-defined constant (λ). Thus the initial population age structure is represented by

$$N^s_{1,1} = \omega\, C^s R_0$$
$$where,\ C^1 = \lambda; C^2 = (1-\lambda)$$
$$N^s_{a,1} = N^s_{1,1} \mathrm{e}^{-M^S(a-1)} \prod_{i=1}^{i=a-1} (1 - {}_I v^s_{i\ I} u^s) \quad \text{for } a = 2, \cdots A-1$$

The plus group for the initial year is given by

$$N^s_{A,1} = N^s_{1,1}\, \mathrm{e}^{-M^S(A-1)} \frac{\prod_{a=1}^{a=A-1} (1 - {}_I v^s_{a\ I} u^s)}{1 - \mathrm{e}^{-M^S} (1 - {}_I v^s_{A\ I} u^s)}$$

Uncertainty in the initial age structure is incorporated by using log-normal error:

$$N^s_{a,1} = N^s_{a,1}\, \mathrm{e}^{-({}_I\varepsilon_a - \frac{{}_I\sigma^2}{2})} \quad where \quad {}_I\varepsilon_a \sim N(0, {}_I\sigma^2); \quad \text{for} \quad a = 2, \cdots A-1$$

The plus group has an independent error component ${}_P\varepsilon_A$ (with its own variance), where *P* stands for plus group and *I* for initial.

1.2.3. *Stock–recruitment*

Recruitment follows a Beverton–Holt stock–recruitment relationship with log-normal error structure

$$N^s_{1,t+1} = C^s \frac{S_t}{\alpha + \beta S_t} e^{\left({}_R\varepsilon_t - {}_R\sigma^2/2\right)}$$

where ${}_R\varepsilon_t$ is the recruitment residual for year *t* (${}_R\varepsilon_t \sim N(0, {}_R\sigma^2)$), and S_t is the spawning biomass in year *t* computed as

$$S_t = \sum_a w^f_{a,t} \Phi_a N^f_{a,t}$$

where Φ_a (maturity ogive) is the fraction of females that have reached maturity at age *a* and $w^f_{a,t}$ is female weight at age and time.

Recruitment at equilibrium in the absence of fishing equals (Mace 1994)

$$R_0 = \frac{SpR - \alpha}{\beta SpR} \text{where} SpR = \lambda \sum_a w^f_{a,t} \Phi_a e^{(-M^f(a-1))}$$

where *SpR* is the spawning biomass per recruit (a function of the surviving proportion, weight at age, and maturity ogive of females). The model was re-parameterized with a

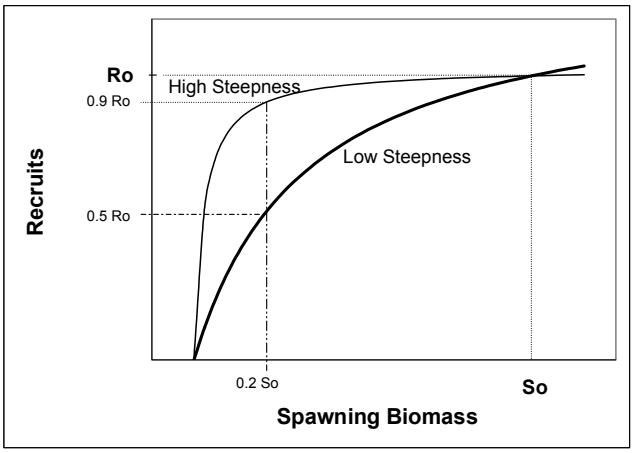
steepness parameter *z*, the proportion of the virgin recruitment that is realized at a spawning biomass level of 20% of the virgin spawning biomass (Francis 1992).

Thus both parameters can be formulated as a function of z, R_0 *and* SpR,

$$\alpha = S_0 \frac{1-z}{4z\ R_0}$$

$$\beta = \frac{5z-1}{4zR_0}$$

$$\boldsymbol{S_0 = R_0\ SpR}$$



This graph shows the Beverton–Holt relationship, formulated as a function of the steepness and virgin recruitment. This parameterization is very convenient because *z* is clearly defined between [0.2, 1].

Note: If the total catch (landings) and weight-at-age data are in the same units, then R_0 *is going to be in numbers. In other situations,* R_0 *should be multiplied by the ratio of the units of the catch to the weight-at-age in order to scale it to the correct units (this is not a necessary task, but is important for understanding the meaning of the units of the virgin recruitment).*

1.2.4. *Growth*

Fish growth is modeled according to a von Bertalanfy model with mean size at age given by

$$L_a^S = L_\infty^s(1-\exp(-k^s(a-t_0^s)))$$

We assume that the distribution of size at age is log-normal with standard deviation sd_a^s, which is a linear function of mean size at age

$$sd_a^S = {}_{L_1}\sigma^S + \left[\frac{{}_{L_n}\sigma^S - {}_{L_1}\sigma^S}{L_n^S - L_1^S}\right](L_a^S - L_1^S)$$

This is basically a linear interpolation between the standard deviation of the mean length at the first (L_1^S) and last (L_n^S) age. The distribution of log(*L)* at age (length–age relationship) by sex is symbolized by $\boldsymbol{\varphi}\left(\mathbf{log}(\boldsymbol{L}) \mid \boldsymbol{\mu}_{\boldsymbol{a}}^{\boldsymbol{s}}\, \boldsymbol{\sigma}^{\boldsymbol{s}}{}_{\boldsymbol{a}}\right)$, and has mean μ_a^s and standard deviation $\boldsymbol{\sigma}_{\boldsymbol{a}}^{\boldsymbol{s}}$, respectively equal to

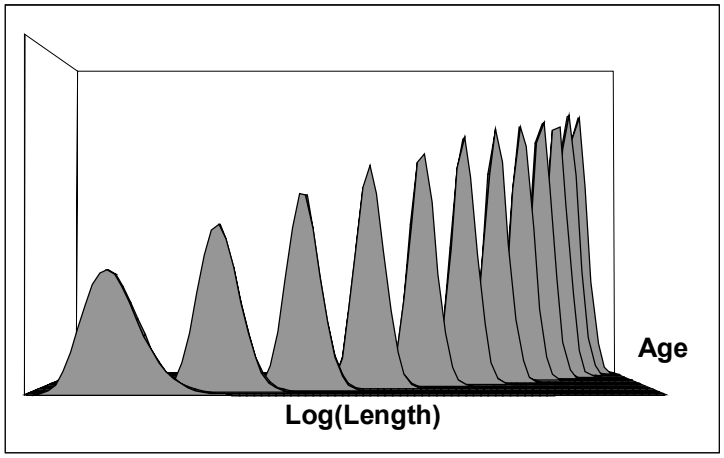
$$\mu_a^s = \log\left(L_a^s\right) - \frac{\sigma_a^{s\,2}}{2}$$

$$\sigma_a^{s^2} = \left(\frac{sd_a^S}{L_a^s}\right)^2$$

The length proportions at age can be approximated as

$$f_{l|a}^s = \frac{\varphi\left(\log L_l \mid \mu_a^s, \sigma_a^{s\,2}\right)\Delta_l}{\sum_{l=1}^{n_L}\varphi\left(\log L_l \mid \mu_a^s, \sigma_a^{s\,2}\right)\Delta_l}$$

where Δ_l is the width of the interval in log scale. This relation can be visualized in the following graph:



The proportions of length at age are used in many sections of the model, depending on the nature of available data. They are used to compute the predicted size compositions, to convert a length-based selectivity into a selectivity-at-age, and to compute the mean weight at age, when the selectivity function of the survey is a function of length.

1.2.5. *Weight at age relationship*

Weight at age is a vital piece of information in the assessment, because it is involved in the vulnerable biomass calculations. It can be directly incorporated into the model as observed data (design-based estimators) or by using a model-based approach (parameters of the von Bertalanffy growth curve and the weight–length power function).

By default the program uses the observed data. The rest of the temporal weight-at-age information arises from the following calculations:

a. If selectivity is a function of age, mean weight at age is predicted from the following equation

$$w_{a,t}^{s} = b_i^S (L_a^S)^{b_{ii}^S} e^{\left(\frac{b_{ii}^S \sigma_a^{s^2} \left(b_{ii}^2 - 1\right)}{2} \right)}$$

where the exponential is a correction for the variance of the log-normal distribution of size at age. If the survey selectivity is based on age, than the weight at age for the commercial fleet is the same as the one for the surveys.

However, selectivity can be modeled as a function of fish size (only for surveys), in which case the mean weight at age for the surveys is affected by selectivity at size and the length–age relationship according to

$$w_{a,t}^{s,g} = \frac{\sum_l b_i^S (L_l)^{b_{ii}^S} s_{l,t}^{s,g} f_{l|a}^s}{\sum_l s_{l,t}^{s,g} f_{l|a}^s}$$

Note: If you do not have model-based weight-at-age estimates, you will need to input observed data for each year. The vulnerable biomass computation for the last year + 1 is based by default on model-based estimates of weight-at-age, so if you do not have those be sure to add one more line of observed weight-at-age data. If your model is in numbers (catch, biomass, etc), you should enter n_{year}+1 vectors of weight-at-age data filled out with values 1.

1.2.6. *Selectivity*

Selectivity is a process that can be modeled based on age or size. This model supports an age-based selectivity for the fishing fleet and a size- or age-based selectivity for the surveys. In this model the only sex-specific variation in the selectivity function arises from the difference between ages of full recruitment.

1.2.6.1 *Selectivity as a function of age*

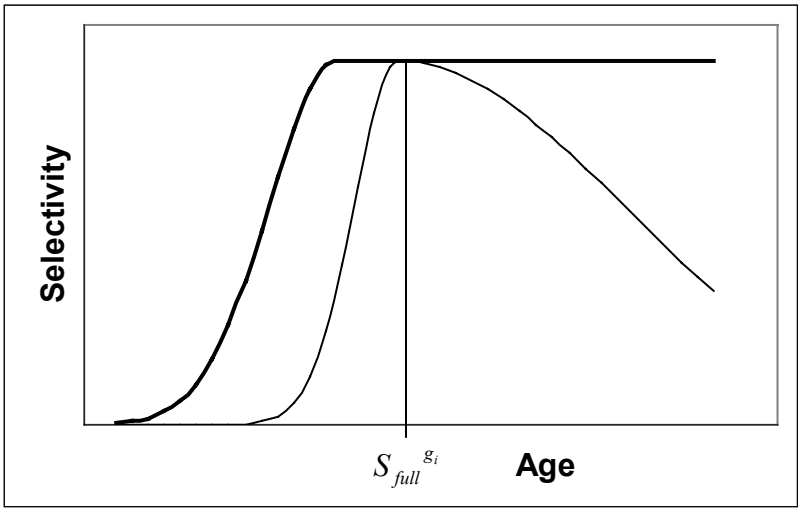
The selectivity function implemented in the model is a double half-Gaussian function of age:

$$s_{a,t}^{s,g} = \begin{cases} \exp\left\{\dfrac{-\left(a - S_{full}^{s,g}\right)^2}{{}_L v^g}\right\} & \text{for} \quad a \le S_{full}^{s,g} \\ \exp\left\{\dfrac{-\left(a - S_{full}^{s,g}\right)^2}{{}_R v^g}\right\} & \text{for} \quad a > S_{full}^{s,g} \end{cases}$$

$$S_{full}^{s,g} = (S_{full}^{g} + (1 - j)\,\Delta_{S_{full}}^{g})$$

where *j* is a dummy variable with value *1* for females and *0* for males, and $\Delta_{S_{full}}^{g}$ is the sex specific difference in age of full recruitment for each gear.

The next graph shows some of the shapes that this three-parameter model can adopt. The thick line represents an asymptotic right-hand side curve (very high right-hand variance), as opposed to the thinner line, which has a declining right-hand limb (smaller right hand variance).



Survey selectivities are assumed to be constant over time while commercial selectivities are allowed to change over time. Residuals are estimated for the periods when we do have catch at age data

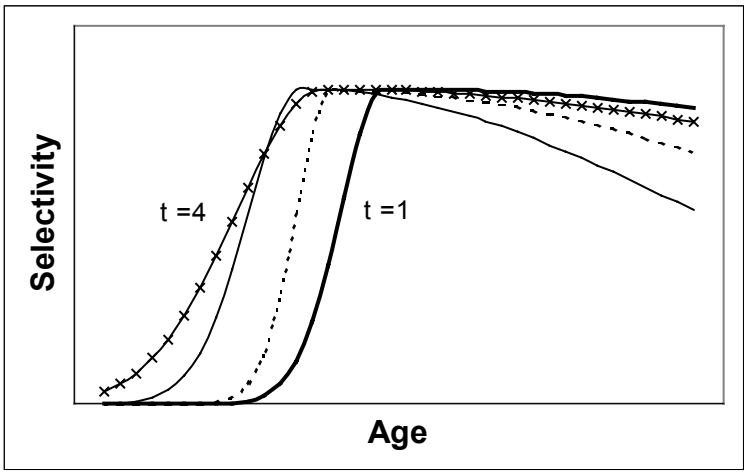
$$S_{full_t}^{S,g} = S_{full_t}^{S,g} e^{{}_{S_{full}}\varepsilon_t^g} \qquad where \;\; {}_{S_{full}}\varepsilon_t^g \sim N(0,\, {}_{S_{full}}\sigma^{g^2})$$

$${}_j v_{t+1}^g = {}_j v_{t+1}^g e^{{}_{j v}\varepsilon_t^{g_i}} \qquad {}_{j v}\varepsilon_t^{g_i} \sim N(0, {}_{j v}\sigma^{g^2})$$

where *j* is the right or left side variance.

Trends in selectivity have been associated with changes in spatial allocation of fishing effort (Jacobson et al. 1997), and the variation considered in this approach is independent of sex.

The following figure shows a declining pattern in the right side of the selectivity curve over time. It also shows a decrease in age of full selectivity between the first and the last time period.



1.2.6.2 *Selectivity as a function of size*

Only the selectivity of the survey is allowed to be size-based. A double-Gaussian function of size, with time invariant parameters, is used. The selectivity at age is computed by integrating the selectivity at size over the size proportions at age. Thus

$$s_{a,t}^{s,g} = \int_{-\infty}^{\infty} s_t^{s,g}(L)\varphi\left(\log L \mid \mu_a^s, \sigma^{2s}_a\right) \mathrm{d}\log L$$

The integral above can be approximated by discretizing the size distribution into n_L size classes, denoted as l, as

$$s_{a,t}^{s,g} = \sum_{l=1}^{n_L} s_{l,t}^{s,g} \; f_{l\|a}^{s}$$

where $s_{l,t}^{s,g}$ is the size-selectivity function evaluated at L_l, the length at the mid-point of interval l . For converting the size-based selectivity into an age-based selectivity, we weight the selectivity at size by the size proportion at the respective age. If we do not rescale the "new" selectivities at age, very likely no age has been fully selected. This would not affect the estimation procedure but would be reflected in the catchability coefficient.

1.3. Data

1.3.1. *Predicted abundance indices*

Commercial CPUE and survey indices, here denoted as I_t^g, are assumed to be directly proportional to the vulnerable biomass in the middle of the year

$$I_t^g = q_t^g \left(\sum_s e^{-0.5M^s} \sum_a s_{a,t}^{s,g} N_{a,t}^s w_{a,t}^g \right)$$

where q_t^g is the gear-specific catchability. The temporal index for the catchability coefficients is incorporated only for the commercial CPUE (catchability coefficients of the surveys are not allowed to have a temporal variation).

A random walk model is used to model the temporal changes, thus

$$\log\left(q_{t+1}^g\right) = \log\left(q_t^g\right) + {}_q\varepsilon_t^{CPUE_i}$$

where ${}_q\varepsilon_t^{CPUE_i} \sim N(0, {}_{q^g}\sigma^2)$. The parameter ${}_{q^g}\sigma^2$ is used to control the amount of year-to-year variation allowed in q_t^g.

1.3.2. *Predicted age and size composition*

The predicted age composition (in proportions) of the catch at time *t* by *sex* and *gear*, is represented by the following equation

$$P_{a,t}^{s,g} = \frac{s_{i,t}^{s,g} N_{a,t}^s}{\sum_s \sum_i s_{i,t}^{s,g} N_{i,t}^s} M_{AxA}^{pool} \Omega^S$$

where Ω^S represents a matrix of age misclassification and $M_{A\,x\,A}^{pool}$ pools the age frequencies for ages $a \geq A_{pool}$ into a plus group.

Real age

Incorrect age

	1	2	3	4	5	6	7	8	9	10
1	**0.8**	0.2	0	0	0	0	0	0	0	0
2	0.1	**0.8**	0.1	0	0	0	0	0	0	0
3	0	0.2	**0.6**	0.1	0.1	0	0	0	0	0
4	0	0	0.2	**0.7**	0.1	0	0	0	0	0
5	0	0	0	0.1	**0.8**	0.1	0	0	0	0
6	0	0	0	0	0.1	**0.8**	0.1	0	0	0
7	0	0	0	0	0	0.1	**0.8**	0.1	0	0
8	0	0	0	0	0	0	0.1	**0.8**	0.1	0
9	0	0	0	0	0	0	0	0.1	**0.8**	0.1
10	0	0	0	0	0	0	0	0	0.1	**0.9**

The above figure shows how to set up the misclassification matrix. If no information on age misclassification is available, an identity matrix (i.e. diagonal of 1) has to be used.

Similarly, size compositions are predicted as

$$P_{l,t}^{s,g} = \frac{s_{l,t}^{s,g} \sum_{a} f_{l|a}^{s} N_{a,t}^{s}}{\sum_{s} \sum_{l} s_{l,t}^{s,g} \sum_{a} f_{l|a}^{s} N_{a,t}^{s}}$$

when selectivity is a function of fish size, or as

$$P_{l,t}^{s,g} = \frac{\sum_{a} s_{a,t}^{s,g} f_{l|a}^{s} N_{a,t}^{s}}{\sum_{s} \sum_{a} s_{a,t}^{s,g} \sum_{l} f_{l|a}^{s} N_{a,t}^{s}}$$

when selectivity is a function of mean length at age.

The range of values of the predicted proportions-at-age, are determined by the dimensioning parameters specified in the input form (same for length) and do not need to match with the ones specified for the dynamics.

1.4. *Objective Function*

Different sources of information contribute to the overall objective function. This can be summarized as follows:

- Survey index: *Relative or absolute index of abundance for each index type.*
- CPUE: Fishery related index of abundance *(by commercial fishing gear index)*.

- Catch-at-length: *Survey* (by gear, length, time, sex, + undetermined sex); *Commercial fleet* (by sex, gear, length, time)

- Catch-at-age: *Survey (by* sex, gear, age, time); *Commercial fleet (by* sex, gear, age, time)

The objective function includes likelihood components for the different data types, and penalties on the variability of the stochastic parameters as specified by their bayesian *prior* distributions.

1.4.1. *Robust normal likelihood for proportions:*

We use the robust likelihood formulation proposed by Fournier et al (1998) for the age-sex and size-sex catch compositions. The observed frequency data is incorporated to the likelihood function as proportions at age and sex, $\widetilde{P}_{a,t}^{s,g}$, or at length, $\widetilde{P}_{l,t}^{s,g}$. The robust-normal model was selected instead of the more traditional multinomial error model because it is more robust to outliers (Fournier et al 1990):

$$\ln L_{\text{age}}^{g} = -0.5 \sum_{t=1}^{Nyears} \sum_{s} \sum_{a=A_{initial}}^{Aplus} \log\left[\left(\tilde{P}_{a,t}^{s,g}(1-\tilde{P}_{a,t}^{s,g}) + .1/A\right)\right]$$
$$+ \sum_{t=1}^{Nyears} \sum_{s} \sum_{a=A_{initial}}^{Aplus} \log\left[\exp\left\{\frac{-(\widetilde{P}_{a,t}^{s,g} - P_{a,t}^{s,g})^2}{2(P_{a,t}^{s,g}(1-P_{a,t}^{s,g}) + .1/A)\tau^{g}}\right\} + 0.01\right]$$

where *A* and τ^{g} are respectively the number of classes and the inverse of the assumed sample sizes. *Aplus*, $A_{initial}$ and N_{years} are the age of the plus group, the initial age observed in the samples and the number of available age-composition samples, respectively. A similar formulation is used for the size-sex compositions and is applicable for survey or commercial data.

1.4.2. *Abundance indices*

Different likelihood functions can be used for the commercial and survey indices. These are normal, log-normal, robust normal and robust log-normal distributions.

The robust log-normal likelihood function has the following representation:

$$\log L_{I}^{g} = \sum_{t} \log\left[\exp\left(-0.5 \frac{{}_{I^g}\varepsilon_t^2}{{}_{I^g}\sigma_t^2}\right) + 0.01\right]$$

In all the cases the variances are entries specified by the user (not estimated within the model) and residuals the difference between the observed and expected values (logarithm of the observed and predicted values, for the previous).

1.4.3. *Total likelihood*

The total log-likelihood is the result of the sum of the individual log-likelihood components

$$\log L = \sum_{n_I} \log L_I^g + \sum_{n_{CPUE}} \log L_{CPUE}^g + \sum_{n_{age}^{Survey}} \log {}^{Survey}L_{\text{age}}^g + \sum_{n_{length}^{Survey}} \log {}^{Survey}L_{\text{length}}^g + \sum_{n_{age}^{Comm}} \log {}^{Comm}L_{\text{age}}^g + \sum_{n_{length}^{Comm}} \log {}^{Comm}L_{\text{length}}^g$$

1.4.4. *Penalties*

Several penalties might be affecting the overall objective function, depending on different model assumptions. In general the penalties correspond with prior assumptions made about some of the stochastic processes involved—specifically, recruitment variability and variability in the initial age structure

$$PSS_R = 0.5 \sum_t \frac{{}_R\varepsilon_t^2}{{}_R\sigma^2}$$

$$PSS_I = 0.5 \sum_t \frac{{}_I\varepsilon_a^2}{{}_I\sigma^2}$$

time-series trends in catchability by gear

$$PSS_q^g = 0.5 \sum_t \frac{{}_q\varepsilon_t^2}{{}_q\sigma^2}$$

and time-series trends in the parameters of the age-selectivity functions for the different commercial fisheries,

$$PSS_{\text{Sfull}}^g = 0.5 \sum_t \frac{{}_{\text{Sfull}}\varepsilon_t^{g\,2}}{{}_{\text{Sfull}}\sigma^2},\ PSS_{{}_L\nu}^g = 0.5 \sum_t \frac{{}_{L\nu}\varepsilon_t^{g\,2}}{{}_{L\nu}\sigma^2} \text{ and } PSS_{{}_R\nu}^g = 0.5 \sum_t \frac{{}_{R\nu}\varepsilon_t^{g\,2}}{{}_{R\nu}\sigma^2}$$

Hence the overall penalty would be the sum of the individual components

$$\text{penalties} = PSS_R + PSS_I + \sum_g PSS_q + \sum_g PSS_{\text{Sfull}}^g + \sum_g PSS_{\nu 1}^g + \sum_g PSS_{\nu 2}^g$$

1.4.5. *Prior*

Prior information on main parameters of the model can be incorporated by using three different density functions (uniform, normal and lognormal). If the parameter is being estimated (active) and has a pre-specified prior, the natural logarithm of this density will be added to the global objective function.

1.4.6. *Global objective function*

Parameter estimates are obtained by minimizing the overall objective function

$$f = -\log L + \text{penalties} + prior$$

1.5. *Policy evaluation*

Coleraine can also be used to evaluate alternative management policies by simulating future stock trajectories *n* years into the future. A Bayesian approach is

implemented to carry out this task. The Monte Carlo Markov Chain (MCMC) technique is used to generate samples from the joint posterior probability distribution, and the marginal distribution of any model and derived parameter can be readily approximated from it. We use the AD Model Builder's (Otter Research Ltd. 1999) implementation of MCMC, which is based on the Hasting-Metropolis algorithm (Gelman et al 1995).

Two types of harvest strategies are implemented in *Coleraine*, namely *constant catch* and *constant harvest rate* strategies. Details on the entries can be found in section 6.1.

The projections are carried out using one selectivity ogive ($s_a^{\mathrm{s,proj}}$**)**. This selectivity pattern is computed as an average selectivity at age for the different fishing fleets, weighted according to the gear-specific exploitation rates in the final year covered by the assessment:

$$s_a^{\mathrm{s,\ proj}} = \frac{\sum_g s_{a,t}^{s,g} u_t^g}{\sum_g u_t^g}$$

The user can also specify other weights to compute the average selectivity (Section 6.1). There is a constraint on the total exploitation rate to be less than 0.99 during the entire projection.

Two sources of uncertainty are incorporated in the projections: (1) uncertainty in current population size and (2) process uncertainty. The uncertainty in the current population size is determined during the MCMC simulation and is a function of all the uncertainty associated to the main parameters of the model. Process uncertainty, on the other hand, is simulated during the execution of *mceval* by allowing for

i. Variability in recruitment, which is modeled as

$$\boldsymbol{N}_{1,t+1}^{s} = \boldsymbol{\lambda} \frac{\boldsymbol{S}_t}{\boldsymbol{\alpha} + \boldsymbol{\beta}\, \boldsymbol{S}_t} \boldsymbol{e}^{\left(_R\varepsilon_t - _R\sigma^2/2\right)}$$

ii. Implementation error due to errors in future assessments (only for harvest rate strategy).

We assume that future estimates of exploitable biomass, $\hat{B}_t$ would be log-normally distributed around the true value, so that

$$\hat{B}_t = \left[\sum_s \mathrm{e}^{-0.5M^s} \sum_a s_a^{\mathrm{s,proj}} w_a^s N_{a,t}^s \right] \mathrm{e}^{_{ass}\varepsilon_t}$$

and $_{\mathrm{ass}}\varepsilon_t$ represents the log of the assessment error in year *t*. The $_{\mathrm{ass}}\varepsilon_t$ are assumed to be normally distributed with variance $_{\mathrm{ass}}\sigma^2$ and serial autocorrelation ρ, such that

$$_{\mathrm{ass}}\varepsilon_{t+1} = \rho\ _{\mathrm{ass}}\varepsilon_t + \xi_{t+1} \qquad \xi_t \sim N(0,\ _{\mathrm{ass}}\sigma^2(1-\rho^2))$$

2. Getting Started

2.1. Hardware Requirements

This program should run on almost any Pentium machine with at least 16 Mb of RAM. More memory will improve the performance of the program. The program requires about 1.6 Mb of hard-disk space. We recommend the use of Excel 2000 because of some well-documented problems with the memory handling that existed with previous versions when a large number of graphs were created.

2.2. Software Requirements

Windows 95/98/2000/NT and Excel 97/2000 are the only software required. Our program's user interface consists of macros written in VBA for Excel 97/2000, a scaled-down version of Visual Basic included with Microsoft Excel 97/2000. In some cases, the user may have to load this part of the program into Excel by using the "Add-Ins" menu.

2.3. Outline of Working Environment

This package consists of two parts: (1) an estimation program (*Colera20.exe*) written with an application called "AD Model Builder" (Otter Research Ltd. 1999) that does all the model parameter estimation and policy evaluation, and (2) a user-friendly Excel 97/2000 interface for entering data and viewing the model outputs. The estimation program was described in Section 1.2. and comes to you as a compiled and unalterable file.

The Excel user interface is controlled by macros written in Visual Basic for Applications (VBA). Our objective was to allow the user to work with Excel's full capabilities, while automating some complicated and time consuming procedures such as setting up a correctly dimensioned input file for the estimation program or graphing output data. Therefore, the user interacts directly only with regular Excel files, and Excel operates normally most of the time. The only exception is that it pauses while macros are running, and a new menu is added to the command bar that allows the user to call macros more easily. This menu is active as long as the Excel file (*Maincode.xls*) containing the macros is open.

2.4. Necessary Files

Five files are critical to running this program:

1. ***Colera20.exe***. A compiled version of the ADMB estimation program. Your "Path" statement in *Autoexec.bat* must include the directory where this program resides. It may be easiest for you just to put it in the working directory.

2. ***Maincode.xls***. Excel file that contains all the macros; it must be open for the macros to function. It can be anywhere Excel can find it and it should not be modified.

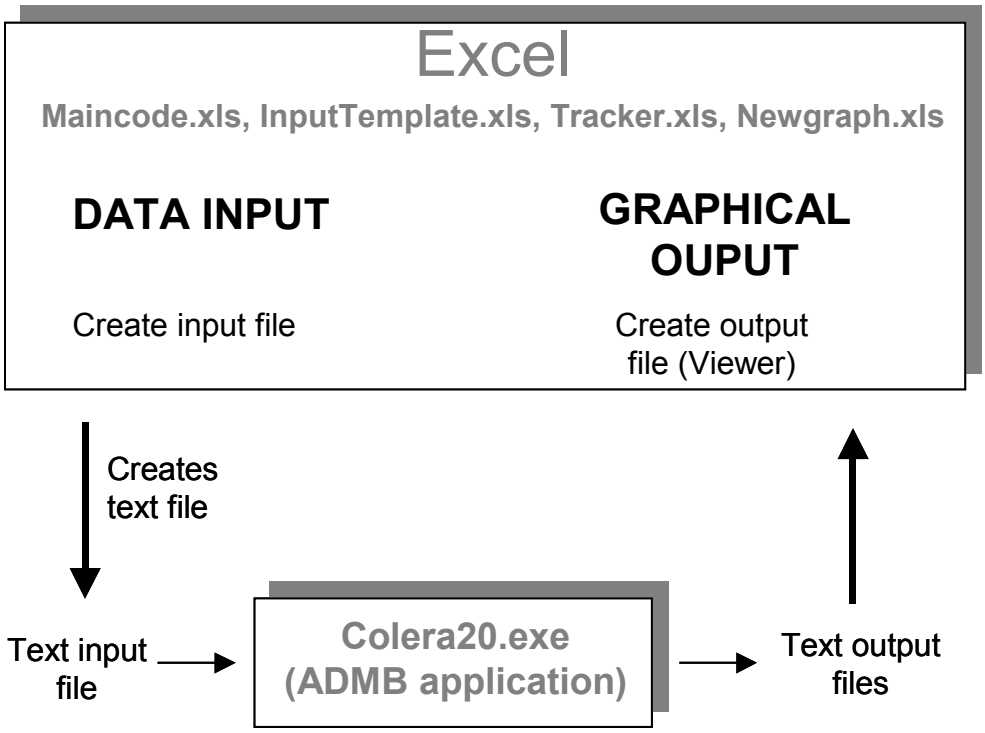
3. ***InputTemplate.xls***. Needed to create new input sheets. It is invoked from *maincode.xls*.

4. ***Tracker.xls***. Excel file that stores vital information (e.g., path information, file names, parameter values, likelihoods) for each run.

5. ***Newgraph.xls***. Used to create new output files. It is automatically called from *maincode.xls*.

2.5. File Management

The use of an Excel interface for data entry and graphing adds an extra step to file management, since *Colera20.exe* requires text-format input, and produces text-format output. In brief, the process of working with a dataset runs as follows:



Therefore, each run of the model has four data files associated with it: two *input* files (one in Excel format, and one in text format), and two *output* files (again: Excel and text formats).

Note: Excel format input and output files may remain open while the model runs, and the Excel file named "Maincode.xls", which contains the macros controlling the user interface, must remain open for the user interface to work.

3. Getting Your Data into *Coleraine*

3.1. General Issues

Colera20.exe reads in all data from a file before running, and is very discriminating about data structure. A single missing value will shift all the following arrays to the left and, if you are lucky, cause the program to crash. Therefore, *most arrays must be filled with dummy data even if they are not used in the calculations*. We have designed the following procedure to minimize problems:

a. Enter into a template the small set of numbers that define the dimensions of the rest of the data arrays (e.g., number of years of data, maximum age and length of fish, number of sexes)
b. Use the filled-in template to create an input form in the proper dimension
c. Enter the rest of the data into the form
d. Check for blanks and other common problems

3.2. Default Values

Entering the required priors is tedious and can be confusing, so our program pastes in default values to the prior section. Our intention is to help you remember what type of values should go in each cell; however*, each value should be carefully checked before running the model*. On the other, the results of the estimation might be sensitive to the initial parameter values; therefore, a menu option allows you to quickly do a deterministic run of the model to ensure that the estimation model starts at reasonable values.

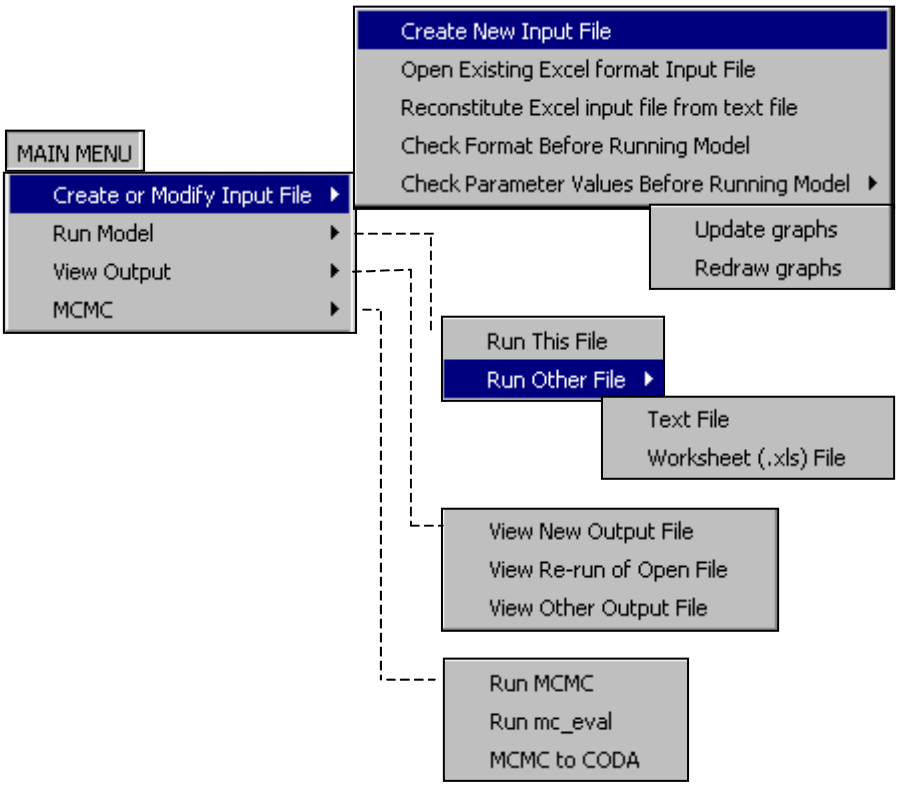
3.3. Two Warnings

Do not attempt to enter dummy data yourself. We have designed the program to do it for you where data are missing.

You will notice that the input form has several sheets. They should not be deleted or renamed. The sheet called “NameList” defines the sizes of each data range in the input sheet, and the sheet called “Defaults” shows the defaults used. “Output” is used in creating a text output file and is cleared each time the main program runs. “Graphs” is used by the parameter checking routine. Note that many cells contain formulas and many are named. Generally speaking, loss or change of names on ranges or sheets will cause the Excel macros to crash. If you want to change default values, try changing the numbers in “Defaults”, but always check the name and formula boxes to makes sure you are not changing names or formulas. Both “Defaults” and “NameList” are protected, but may be unprotected using the Excel menu Tools → Protection → Unprotect sheet.

3.4.The “Main Menu”.

The following figure describes all the different components of the Main Menu that are added to the principal Excel toolbar. Each component will be explained in later sections of this manual.



The three main procedures in the Main Menu are related to the creation or modification of files, running of the estimation program, and viewing the model outputs.

3.5. Step-By-Step Procedure for Entering Data:

Adhering to the following procedure will minimize problems:

a. Open *MainCode.xls*. You will notice a new menu item is added to the toolbar at the top of your file, called “MAIN MENU”. It will disappear when you close *Excel*. (*Note:* the menu becomes inactive when an embedded object such as a chart is activated. If the menu suddenly disappears, simply click on any non-object, such as a cell on the sheet.)

b. To start working with a new dataset, choose *Main Menu → Create* or *Modify Input File → Create New Input File*.

This opens a template.

c. Fill in the appropriate dimensioning data. See Appendix 2 for clarification of the meanings of each value. Every dimension is critical, so make sure you understand the distinctions made.

d. When done, push the button labeled "Create the Rest of the Form".

This will create the rest of the form, using the dimensions you entered. Beware that pushing this button again would erase any information on the page below the button, and redraw the form.

e. Now enter your data. We recommend pasting in data from other spreadsheets, although cells must contain numbers and not formulae (Use "Paste Special"→Values). It is critical to avoid renaming cells by mistake (if you do not use paste special all the time, this might happen). Note before pasting:

   - All outlined areas must be filled with numbers. See Appendix 2 for clarification of the data types and formats required.

   - Some arrays are filled with dummy data. Do not change them.

   - Some arrays may already have zeros in some positions: e.g., if your catch-at-age data begin at age 6, ages 1 through 5 will appear as columns of zeros in the catch-at-age data section. This is the required format.

   - Priors and other default data can and *should be* overwritten.

f. Go to *Main Menu → Create* or *Modify Input File → Check Format Before Running Model*

This calls a macro that will search for blanks and *color* them *yellow*. It will also check the format of years, which must have 4 digits (they are colored red). It can be run repeatedly with no ill effects.

g. Then go to *Main Menu → Create* or *Modify Input File → Check Parameter Values Before Running Model.*

This calls a macro that runs population projections, using the given initial parameter values (i.e., the values in column 7 of the prior section). It graphs the results in a sheet called "graphs." You can go back and change the default values and run this macro again until you have reasonable starting parameter values.

h. At this point you are ready to run the estimation program*. If you are going to run the file immediately, there is no need to save the file or to close it*—just go to *Main Menu → Run Model → Run This File*. The invoked macro does three things: (1) creates a slightly modified text-only version of the file you just made, 2) saves both the Excel and text versions (you will be prompted for names), and (3) calls the kernel AD Model Builder program, passing it the text file. The Excel format input file you created will stay open so that you can go back to it when you have seen the output. *But if you want to save it for later*, use the regular Excel "Save" menu and save it as an Excel workbook (.xls). Later, you can check or modify it using *Main Menu → Create* or *Modify Input File → Open Existing Input File*, *Main Menu → Create* or *Modify Input File → Reconstitute Excel* input file from text file, or run it directly from *Main Menu → Run Model → Run Other File → Workbook* (.xls) file.

4. Running *Coleraine*

4.1. General Issues

Once you have entered all your data, you may want to fit the model to the available information. The instructions for running an open or another file were given in the section 3.5 g-h of this manual. When you choose any of these options, the VB macros internally call the estimation executable program (Colera20.exe) for you. Always remember that the estimation results are a function of the *fixed parameter value*, *input data, likelihood type* and *priors*. For some entries, *Maincode.xls* gives you default values. This was done to help you input the correct type of data, but we strongly advise you to ensure all values are related to your own stock or match your assumptions about the dynamics.

4.2. Parameter Estimation

The statistical model used in the parameter estimation process uses a formal maximum likelihood approach. Given the complexity and non-linear structure of the model, non-linear estimation procedures (using a Quasi-Newton minimization algorithm) are used to fit the model to the available data. It has been recognized (Seber and Wild 1989) that the final parameter estimates might be a function of the initial parameter values in the non-linear search. This is generally true for situations with a very flat, joint likelihood surface and where it is likely that the minimizer gets stacked at different local minima. Therefore, we strongly recommend that you try different sets of starting values.

4.3. Phases in the Estimation

Trying to estimate all the model parameters simultaneously in a non-linear model situation may not be advisable. It is convenient to keep some of the parameters fixed during the initial part of the minimization process and carry out the minimization over a subset of parameters. The other parameters are included in the minimization process in a number of phases until all of the parameters have been included (ADMB–Otter Research Ltd.).

The first entry (first column) of each prior (in the Prior Section) controls the phase in which the parameter will enter the minimization process. A negative number implies that the parameter will not be estimated and the value of a positive number determines the phase in which a particular parameter enters the minimization process.

A suggested phase order would be the following (read Appendix C):

Phase 1: q to fit the index data.
Phase 2: $\boldsymbol{R}_0$ to adjust the trajectory.
Phase 3: Selectivity parameters.
Phase 4: Recruitment residuals.
Phase 5: Changes in selectivity.
Phase 6: Deviation in initial conditions.

4.4. Obtain Bayesian Posteriors

Bayesian posteriors for the derived parameters of interest are obtained by using the Markov Chain Monte Carlo (MCMC) method (Otter Research Ltd. 1999). To run

MCMC go to *MCMC → Run MCMC* (Figure Section 3.4). The program will start by fitting the model to the data using maximum likelihood. To continue with the bayesian part, the variance–covariance matrix must be available. An example on how to use the MCMC and Mceval routines is shown in Appendix 1.

4.5. Policy Evaluation

After the estimation is completed, different policies can be evaluated. The necessary parameters for running the projection under different assumptions are incorporated in the *Projection Parameter* section during the input of the data. Once MCMC is executed, the *colera20.psv* should be in your directory. It contains important information on the uncertainty of the current year biomass, which is used in policy evaluation. One nice feature of *Colera20.exe* is that the user can modify at any time the projection conditions without having to rerun MCMC. To run *Mceval* go to *MCMC → Run mc_eval*. You can do this immediately after running MCMC or at any time in the future.

After running *Mceval*, some new text files will be generated in your working directory:

project.out: Contains the posteriors of the virgin vulnerable biomass, virgin spawning biomass, projected time series of spawning biomass, and vulnerable biomass for each chosen harvest strategy level.

vbio.pst: Posteriors of the time series of estimated vulnerable biomass for the first fishing gear.

mbio.pst: Posteriors of the time series of estimated spawning biomass.

recs.pst: Posteriors of the time series of estimated recruiments.

explrate.pst: Posteriors of the time series of estimated exploitation rates for the first fishing gear.

params.pst: Posterior probability of the virgin spawning biomass, virgin recruitment, recruitment steepness, natural mortality by sex, catchability coefficient per fleet, and catchability coefficient of each survey index.

5. Processing Output

5.1. Overview

The estimation program produces about 20 files each time the program is run. *Only a few of the files generated are relevant to the user, and none need be opened, since our viewer macros copy the main results directly into Excel*. The benefit of using the *Viewer* is that it takes the estimation output file, which is usually large (800 lines or more), sparsely labeled, and dense, and makes it easy to find and display information from it. The process of viewing output is entirely independent, as far as the macros go, of the process of creating input files and running the model. Essentially all the viewer does is to copy the "*Results.dat*" file into an Excel worksheet, and then add 11 more worksheets to the Excel file, including a graph-making template to make it easy to graph the data and manipulate the graphs in blocks.

Memory problems: Graphs take up an enormous amount of memory, and for some undetermined reason, Excel does not reallocate RAM after a graph is deleted. Therefore, after you have created and deleted a number of graphs (e.g., 50 on a laptop), Excel may run out of memory, *even if no graphs are currently visible*. This typically happens when you are trying to make more graphs. The solution is to end the macro (just choose "End"), and close the file. The memory is released and you can then reopen the file (this problem seems to be fixed for Excel 2000). Closing Excel itself seems unnecessary.

5.2. Text Output Files Produced by Colera20.exe

The relevant output files are in text format, and among them only one is used by the viewer: *Results.dat*.

Results.dat contains an exhaustive listing of the maximum likelihood estimates for the model and derived parameters, a re-listing of some of the data and fixed parameters, and a listing of most of the prediction made. This includes numbers at age, fecundity, vulnerable biomass, survey trajectories, and so on. It is always placed in the same directory as the text input file. If there is an existing *Results.dat* file in that directory, it will be overwritten.

Other output files generated during the parameter estimation process include *Colerain.par* (contains the maximum likelihood estimates of the free parameters), *Colerain.cor* (shows standard deviation and correlations between the estimated parameters), and *Colerain.std* (standard deviation of the estimated parameters).

5.3. Using the Viewer, Part 1: Create or Open a Results File

A. Viewing the results of a file you just ran: Choosing Main Menu → View Output → View New Output File will run a macro that opens a Results.dat file and creates an Excel workbook with 11 sheets in it. If the model just ran, the program should still have the name of the input file in its memory. In this case, it will take Results.dat from the same directory, since any older Results.dat would have been overwritten. Otherwise it will ask you to define the Input File used.

B. Viewing a different Excel output file: The option Main Menu → View Output → View Other Output File will open an existing 12-sheet Excel workbook that was previously created using the View New Output File option.

C. Viewing the results of a re-run open file. We wanted to make it possible to quickly modify parameter values, run the model, and look at the output without going through the process of creating a new Excel output workbook, and without losing modifications that might have been made to the graphs. The menu option called Main Menu → View Output → View ReRun of Open File is designed to be used when a model has been re-run, and a 12-sheet Excel viewer file is currently open. It simply pastes a new Results.dat file into the sheet called Master. It is important to realize that any change to a parameter that would change the dimensions of a data array would distort the viewer badly.

5.4. Using the Viewer, Part II: Making Graphs

Once a *Viewer file* is open, you are in a regular Excel file, with the only exception being that the buttons appearing on the sheets are linked to macros in the *MainCode.xls* workbook (which must also be open for them to work). The graphs can be customized individually as you would do with any Excel graph, but the viewer macros also make it easy to manipulate them in blocks.

A. Making changes to one sheet: Anything highlighted in blue can be changed by the user. The changes will go into effect when you press the Make Graphs button again. In rows 1-8 there are a few parameters that control all the graphs on the sheet (height, width, etc.). In the block of data starting in row 12, the options affect each graph individually. Typing "Y" in the blue cells below columns marked Female or Male will activate that graph; anything else will cause the graph to be skipped. It is possible to delete all the graphs on a sheet using the Delete Graphs button, but it is not normally necessary since existing graphs on a sheet are removed each time you press the Make Graphs button.

B. Making changes to all of the sheets at once: The sheet called Graphmaster is essentially a template for all other sheets. Options such as graph height or width, pool length or age, or names of survey indices or commercial fishing methods can be changed for all sheets by changing them on the Graphmaster sheet, and then pressing the yellow button called Remake all Graph Masters. This will cause the other sheets to be re-written and you will lose any changes you have made to individual graphs.

C. The worksheets, one by one: The different worksheets pasted into the viewer are as follows:

> ***General****:* graphs of vulnerable biomass vs. catch, exploitation rate, spawning biomass vs. recruits and spawning biomass vs recruits. Note that the graphs with two series each have two Y axes.

The following sheets all graph predicted against observed data:

> ***SurvNoSexCL***: catch-at-length survey data in which gender is undetermined, by survey index and year.

SurvC@L: survey catch at length data, by survey index, year, and gender.

ComC@L: commercial catch-at-length data by method, year, and gender.

SurvC@A: survey catch-at-age data by survey index, year, and gender.

ComC@A: commercial catch-at-age by method, year, and gender.

Surveys: surveys by survey index and gender. It compares predictions with observed data.

SurvSel: survey selectivity. There is a graph for each gender and method, and a different series for each year. The survey years are shown in row 12, extending to the right of column 9. To make the graph more readable, one can delete series by replacing years in row 12 with any text, which will cause that year to be omitted from the graph.

CommSel: commercial selectivity, by method, year and gender. Uses the same method as SurvSel for deleting series. c

CPUE: catch per unit effort. Graphs the observed CPUE against CPUE index trajectories.

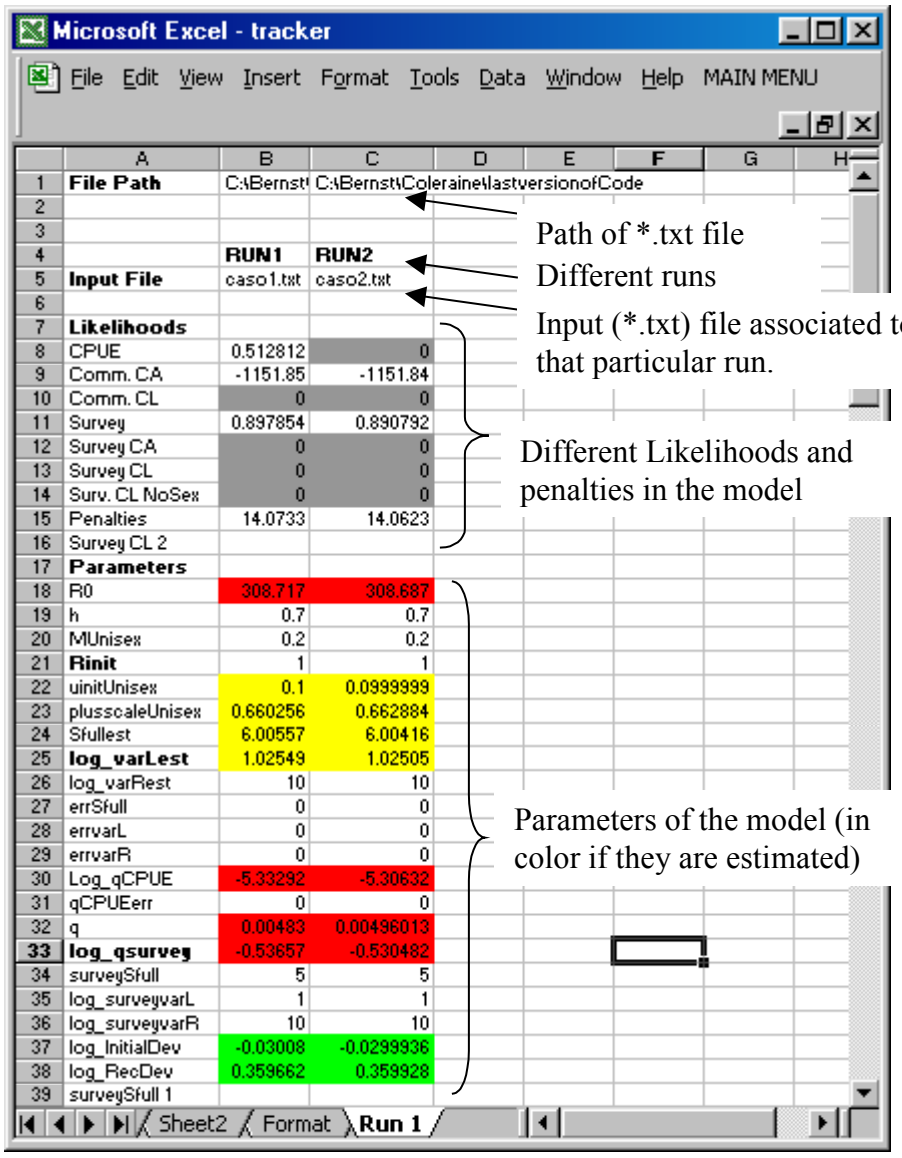
Master: a direct copy of the results file, *Results.dat*, but the viewer program has located and named each data range so the graph sheets can find the data to graph.

GraphMaster: a template, as explained in the 5.4B.

5.5. Storing important information of different runs in one file (Tracker.xls)

A useful feature of *Coleraine* is the way vital information of different runs gets summarized. Whenever the estimation model has been run and the information has been transferred to the viewer, then the file *Tracker.xls* gets updated with relevant information such as likelihoods, parameter values, phases, file path information. This file is automatically open after you open *maincode.xls*.

The next figure shows the different parts of *Tracker* (using the Northern Cod example). If a likelihood component is not used during the estimation it will be colored gray. Parameters that are estimated will be in color (red = phase 1; yellow = phase 2; green = phase 3, etc).



	A	B	C	D	E	F	G	H
1	**File Path**	C:\Bernst\	C:\Bernst\Coleraine\lastversionofCode					
2								
3								
4		**RUN1**	**RUN2**					
5	**Input File**	caso1.txt	caso2.txt					
6								
7	**Likelihoods**							
8	CPUE	0.512812	0					
9	Comm. CA	-1151.85	-1151.84					
10	Comm. CL	0	0					
11	Survey	0.897854	0.890792					
12	Survey CA	0	0					
13	Survey CL	0	0					
14	Surv. CL NoSex	0	0					
15	Penalties	14.0733	14.0623					
16	Survey CL 2							
17	**Parameters**							
18	R0	308.717	308.687					
19	h	0.7	0.7					
20	MUnisex	0.2	0.2					
21	**Rinit**	1	1					
22	uinitUnisex	0.1	0.0999999					
23	plusscaleUnisex	0.660256	0.662884					
24	Sfullest	6.00557	6.00416					
25	**log_varLest**	1.02549	1.02505					
26	log_varRest	10	10					
27	errSfull	0	0					
28	errvarL	0	0					
29	errvarR	0	0					
30	Log_qCPUE	-5.33292	-5.30632					
31	qCPUEerr	0	0					
32	q	0.00483	0.00496013					
33	**log_qsurveg**	-0.53657	-0.530482					
34	surveySfull	5	5					
35	log_surveyvarL	1	1					
36	log_surveyvarR	10	10					
37	log_InitialDev	-0.03008	-0.0299936					
38	log_RecDev	0.359662	0.359928					
39	surveySfull 1							

Bibliography

Francis, R.I.C.C. 1992. Use of fish analysis to assess fishery management strategies: a case study using orange roughy (*Hoplostethus atlanticus*) on the Chatham Rise, New Zealand. Can. J. Fish. Aquat. Sci. 49: 922-930.

Fournier, D.A, J. Hampton, and J.R. Sibert. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with applications to South Pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci. 55:1-12.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 1995. Bayesian data analysis. Chapman and Hall, New York.

Jacobson, L.D. and five co-authors. 1997. Empirical fishery selectivity estimates for Dover sole, sablefish, and thornyheads in the deep water Dover fishery. U.S. Department of Commerce, National Marine Fishery Service, Southwest Fishery Center, La Jolla, California. 32 p.

Mace, P. 1994. Relationships between common biological reference points used as threshold and targets of fisheries management strategies. Can. J. Fish. Aquat. Sci. 51:110-122.

Otter Research Ltd. 1999. AD Model Builder documentation on line. http://otter-rsch.com/admodel.htm

Seber, G.A.F. and C.J. Wild.1989. Nonlinear Regression. John Wiley & Sons.

Acknowledgements

We would like to thank many people that have contributed to improve the model and/or this manual. Among others we have Juan Valero, Carolina Minte-Vera, Ivonne Ortiz, Vera Agostini, John Field, Andre Punt, Arni Magnuson, Trevor Branch, Viviane Haist, Ian Stewart, Shelton Harley, and Marcus Duke.

The development of this software and manual was supported by the NZ Fishing Industry Board (later NZ Seafood Industry Council) and the NZ Foundation for Research, Science and technology.

APPENDIX A

The Northern Cod case study

Introduction

The motivation for doing this exercise is not to present the best available explanation of the stock dynamics for this particular population, but to familiarize the reader with the use of this software. This section comprises seven parts:

1. Creating the Input Sheet
2. Inputting the Data
3. Running the Estimation Program
4. Creating an Output Viewer
5. Updating Re-run Information on Outputs
6. Forward Projections under Different Harvest Policies
7. Organizing the Projection Outputs

About the data

The data used in this exercise corresponds with a cod population (*Gadus morhua*) of the North Atlantic. The exercise consists of a time series of landings (in hundreds of metric tons) starting in 1962 and ending in 1991. The catch-at-age data cover this entire time range and are represented by individuals from age 2 to 20. Other sources of available information, to which the model will be fit to, are a CPUE index (1978 to 1988) coming from the fishing fleet and relative index of abundance (1981 to 1991) originating from a series of Research Trawl Surveys. Some biological information such as weight-at-age and natural mortality is available or assumed known for this stock.

1. Creating the Input Sheet

The first step is to unzip the file *Coleraine.zip* in a directory on your hard drive. We recommend using a folder with a short path sequence (e.g. C:\Coleraine\Cod\), which will prevent some potential problems with the interface. You should have, by now, four Excel files in your directory (*maincode*, *InputTemplate*, *newgraph* and *tracker*) and one executable (*colera20.exe*).

Open *maincode.xls* and enable the macro option. Go to the new created menu button (MAIN MENU) and select the *Create New Input File* option (Figure A.1.1).

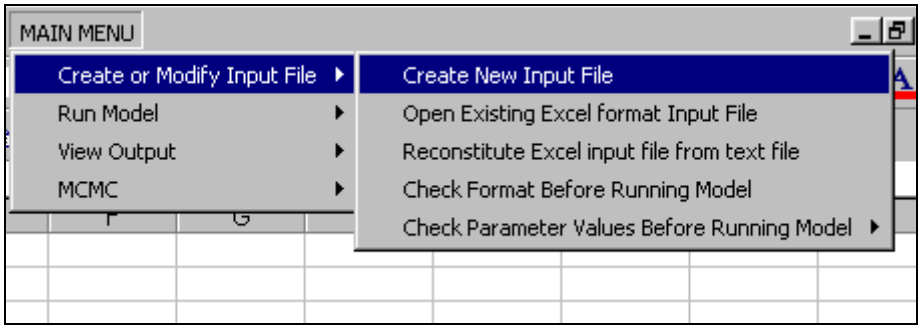


Figure A.1.1. Menu option for creating an Input sheet.

This command creates a new Excel sheet that contains spaces for dimensioning parameters. These entries will structure the rest of the Input sheet in a future step. Start typing in the entries as specified by Figure A.1.2.

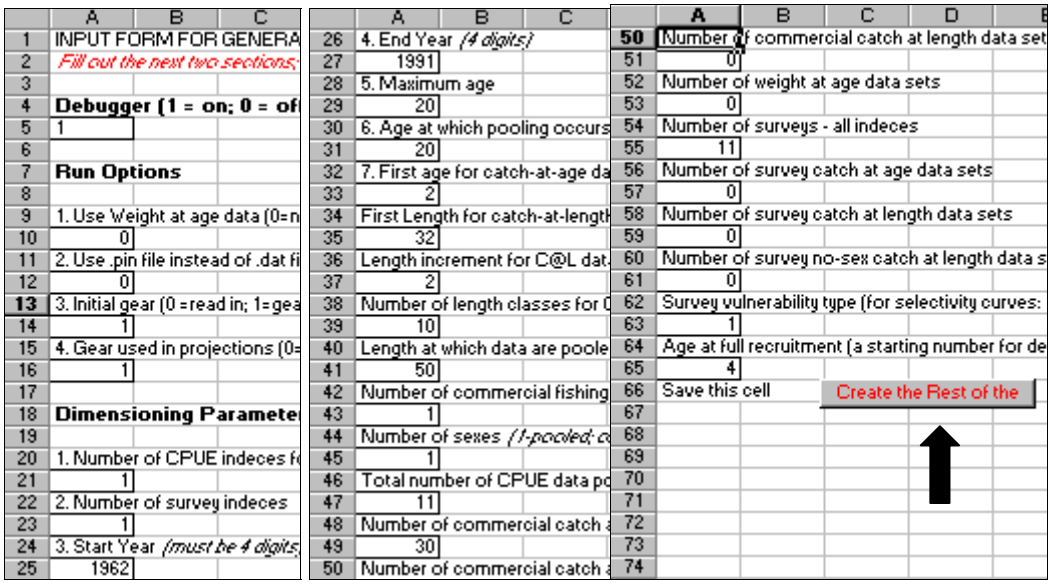


Figure A.1.2. Filling in the dimensioning entries of the *Input Form* for the Northern Cod dataset.

Given that no length-related data are available at this time, entries in rows 35, 39, and 41 represent only dummy data. In order to avoid run-time errors, you should enter a larger number in row 41 than in row 35. Once you are done, click the "*Create the Rest of the Form*" button. The program will create all the necessary spaces for data entry.

2. Inputting the Data

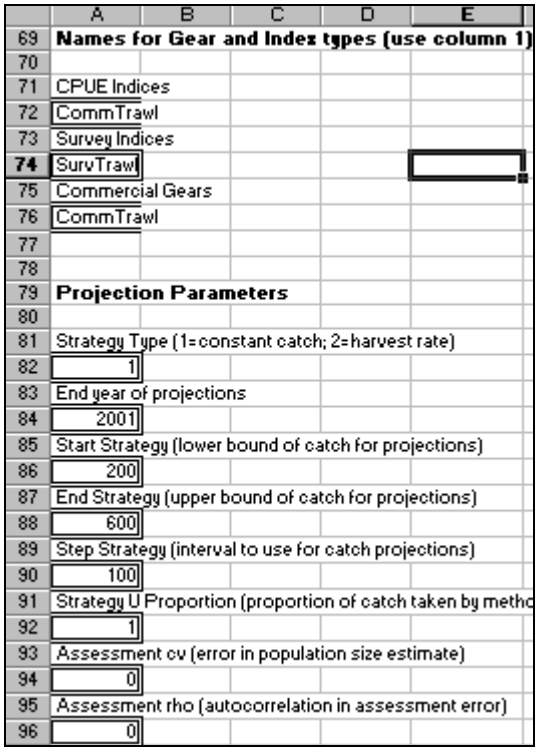
This section comprises the following subsections: (2.1) Gear Type, (2.2) Projection Parameters, (2.3) Priors, (2.4) Likelihoods (2.5), Fixed Parameters, and (2.6) Data. These will be reviewed in sequential order.

Coleraine allows us to specify the *names* of the different gear types and indices (Figure A.1.3). Spaces for these entries are located in rows 72–76. These names will label several graphs in the output viewer and are very useful to keep track of gear-type names in multiple fleet situations.

Figure A.1.3 also shows the *Projection Parameters*. These entries do not play a role in the estimation phase and will be needed in the policy evaluation section. The values must be entered at this stage, though, to get a correct sequence of entries in the input file (more information on the specified values will be given later).

The Priors section is by far the most interactive area of the entire input form. It contains entries for all the relevant information of potentially estimable parameters. The decision to estimate some parameters or not will depend on the available information and our hypothesis about the model structure. In this case, we are estimating the virgin

recruitment (row 102), recruitment deviations from the deterministic Beverton-Holt



	A	B	C	D	E
69	**Names for Gear and Index types (use column 1)**				
70					
71	CPUE Indices				
72	CommTrawl				
73	Survey Indices				
74	SurvTrawl				
75	Commercial Gears				
76	CommTrawl				
77					
78					
79	**Projection Parameters**				
80					
81	Strategy Type (1=constant catch; 2=harvest rate)				
82	1				
83	End year of projections				
84	2001				
85	Start Strategy (lower bound of catch for projections)				
86	200				
87	End Strategy (upper bound of catch for projections)				
88	600				
89	Step Strategy (interval to use for catch projections)				
90	100				
91	Strategy U Proportion (proportion of catch taken by metho				
92	1				
93	Assessment cv (error in population size estimate)				
94	0				
95	Assessment rho (autocorrelation in assessment error)				
96	0				

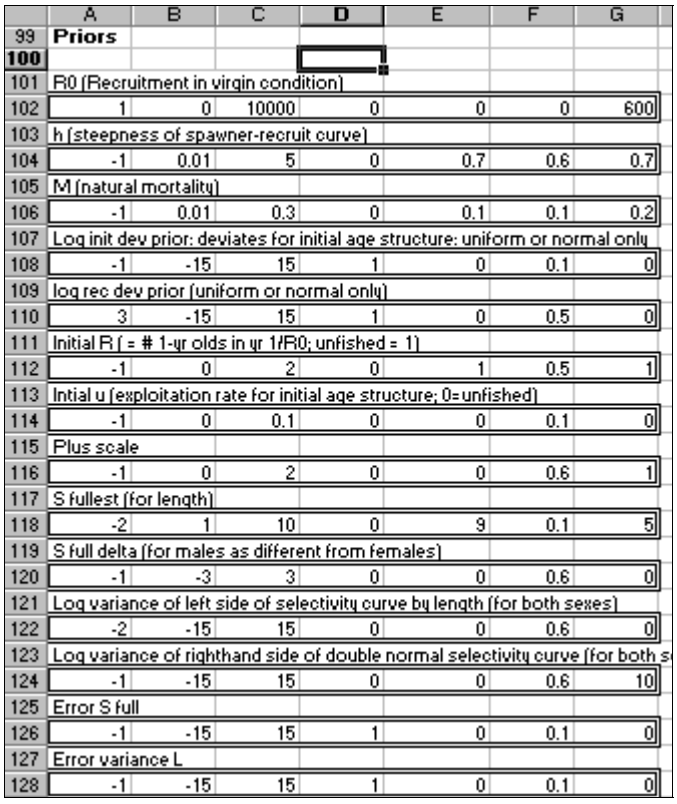
Figure A.1.3: Names for gear types and projection parameters

recruitment model (row 110), and catchability coefficients for the CPUE and relative index of abundance of the Survey (rows 132 and 136). The remaining parameters are fixed and therefore assumed known with values dependent on their initial guesses. Given there was no structured information from the survey data, the survey selectivity parameters were not estimated. For this run the selectivity parameters of the fishing fleet were not estimated. However, this can be done with the long time series of catch-at-age data.

The model currently assumes that the population started at an unfished equilibrium state. This assumption can be readily relaxed by freeing up the initial exploitation rate and the fraction of the virgin recruitment in 1962 (rows 112 and 114). Departures from the deterministic exponential decay in the initial age structure are also allowed by incorporating additional residuals in the estimation (row 108).

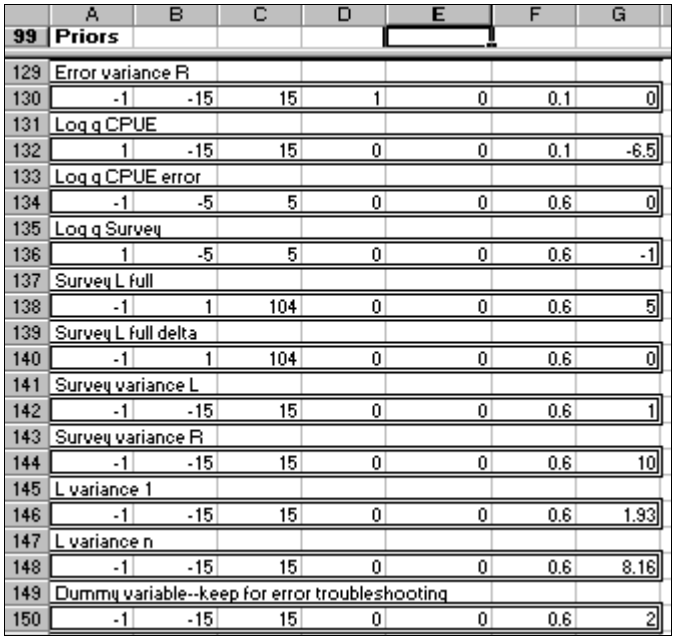
Given that this is a one-sex model (no available sex-structured data), selectivity parameters for rows 120 and 140 are dummy data. It is important to remember that all the residuals, catchability coefficient, and variance parameters for the selectivity ogives are entered in log-scale.

The *Prior* and the *Likelihood* sections can be modified between different runs to explore the consequences of different assumptions. Figure A.1.5 shows the cells for the likelihood entries. This allows us to turn on and off the different likelihood components (in this case CPUE, Survey Index, and Commercial Catch-at-Age) by using the specified values (0: likelihood component is ignored; -1: likelihood component is ignored but the predictions and implied likelihood value is computed; 2: lognormal error structure; 12 robust multinormal for proportions). Cells 164 and 166 determine the nature of the Survey Index and selectivity ogive, respectively.



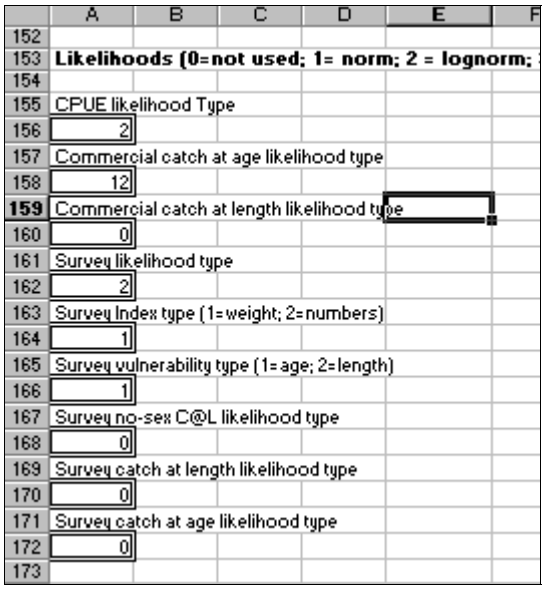
	A	B	C	D	E	F	G
99	**Priors**						
100							
101	R0 (Recruitment in virgin condition)						
102	1	0	10000	0	0	0	600
103	h (steepness of spawner-recruit curve)						
104	-1	0.01	5	0	0.7	0.6	0.7
105	M (natural mortality)						
106	-1	0.01	0.3	0	0.1	0.1	0.2
107	Log init dev prior: deviates for initial age structure: uniform or normal only						
108	-1	-15	15	1	0	0.1	0
109	log rec dev prior (uniform or normal only)						
110	3	-15	15	1	0	0.5	0
111	Initial R (= # 1-yr olds in yr 1/R0; unfished = 1)						
112	-1	0	2	0	1	0.5	1
113	Intial u (exploitation rate for initial age structure; 0=unfished)						
114	-1	0	0.1	0	0	0.1	0
115	Plus scale						
116	-1	0	2	0	0	0.6	1
117	S fullest (for length)						
118	-2	1	10	0	9	0.1	5
119	S full delta (for males as different from females)						
120	-1	-3	3	0	0	0.6	0
121	Log variance of left side of selectivity curve by length (for both sexes)						
122	-2	-15	15	0	0	0.6	0
123	Log variance of righthand side of double normal selectivity curve (for both s						
124	-1	-15	15	0	0	0.6	10
125	Error S full						
126	-1	-15	15	1	0	0.1	0
127	Error variance L						
128	-1	-15	15	1	0	0.1	0

Figure A.1.4. Parameter information required by the estimation model. The different columns represent: (1) phase number, (2) lower bound of the parameter, (3) upper bound of the parameter, (4) type of prior distribution, (5) mean, (6) standard deviation, (7) initial guess.



	A	B	C	D	E	F	G
99	**Priors**						
129	Error variance R						
130	-1	-15	15	1	0	0.1	0
131	Log q CPUE						
132	1	-15	15	0	0	0.1	-6.5
133	Log q CPUE error						
134	-1	-5	5	0	0	0.6	0
135	Log q Survey						
136	1	-5	5	0	0	0.6	-1
137	Survey L full						
138	-1	1	104	0	0	0.6	5
139	Survey L full delta						
140	-1	1	104	0	0	0.6	0
141	Survey variance L						
142	-1	-15	15	0	0	0.6	1
143	Survey variance R						
144	-1	-15	15	0	0	0.6	10
145	L variance 1						
146	-1	-15	15	0	0	0.6	1.93
147	L variance n						
148	-1	-15	15	0	0	0.6	8.16
149	Dummy variable--keep for error troubleshooting						
150	-1	-15	15	0	0	0.6	2

Figure A.1.4 (continued). Parameter information required by the estimation model.



	A	B	C	D	E	F
152						
153	**Likelihoods (0=not used; 1= norm; 2 = lognorm;**					
154						
155	CPUE likelihood Type					
156	2					
157	Commercial catch at age likelihood type					
158	12					
159	Commercial catch at length likelihood type					
160	0					
161	Survey likelihood type					
162	2					
163	Survey Index type (1=weight; 2=numbers)					
164	1					
165	Survey vulnerability type (1=age; 2=length)					
166	1					
167	Survey no-sex C@L likelihood type					
168	0					
169	Survey catch at length likelihood type					
170	0					
171	Survey catch at age likelihood type					
172	0					
173						

Figure A.1.5. Switches controlling the use of different likelihood components.

The *Fixed Parameter s*ection allows the user to input auxiliary information on biological parameters such as weight-at-age parameters, von Bertalanfy growth parameters, sd 1 age, last age, maturity ogive, sex ratio, pre-1962 selectivity ogive, ageing errors, and observed weight-at-age over time. Given the lack of length data for this dataset, many of these parameters are dummy variables (Figure A.1.6).

The last section of the input sheet corresponds with the *data entries*. In row 233 we incorporate the time series of landings. We have only one row of entries because this is a uni-fleet fishery. The number of landing observations should match the number of specified years for this analysis.

Lines 235 and 249 contain the number of CPUE and relative index observations. For the CPUE, column *A* specifies the index type and column *B* the fishing method. The fishing method is related to the selectivity used to compute the vulnerable biomass. The index type, on the other hand, is related to the catchability coefficient. This means that one fishing method can have more than one index type (e.g., temporal variation). Column *E* specifies the coefficient of variation around the yearly point estimates of the CPUE. The Survey entries follow the same pattern, starting with index number in column *A*, year in column *B*, point estimate in *C*, and coefficient of variation in *D*.

Catch-at-age data (in proportions) are entered in lines 265–294 and have the following structure:

Column A: Fishing method
Column B: Sampling year
Column C: Sample size (number of observations in a specific year)
Column D: Proportions-at-age in the catch-at-age matrix. For a two-sex model, the female proportions-at-age entries are entered first and the male proportions immediately after them. The entire row of proportion has to add up to 1 (one).

For this dataset, the rest of the entries are dummy data, but they need to be entered to keep the program from crashing.

	A	B	C	D	E	F	G
174							
175	**Fixed Parameters**						
176							
177	Bi-scalar of length-weight relationship						
178	1E-05						
179	bii exponent of length-weight relationship						
180	3.2						
181	L-infinity of the vonBertanlanffy growth equation						
182	103						
183	k of the vonBertanlanffy growth equation						
184	0.097						
185	t0 of the vonBertanlanffy growth equation						
186	-1.97						
187	Brody parameter						
188	0.2						
189	Mean length of age 1 fish						
190	32						
191	Length at oldest age						
192	32						
193	S.d. of length at age of 1-year old fish						
194	6.4						
195	S.d. of length at age of oldest fish						
196	16						
197	Maturity ogive						
198	0	0	0	1	1	1	1
199	Fraction recruiting of each sex (1 in a 1-sex model; 0.5 0.5 in a two-sex model)						
200	1						

Figure A.1.6. Biological parameters that are entered as auxiliary information to the model.

	A	B	C	D	E	F	G
201	vinit						
202	0	0	0	1	1	1	1
203	Age error (which way does it go???)						
204	1	0	0	0	0	0	0
205	0	1	0	0	0	0	0
206	0	0	1	0	0	0	0
207	0	0	0	1	0	0	0
208	0	0	0	0	1	0	0
209	0	0	0	0	0	1	0
210	0	0	0	0	0	0	1
211	0	0	0	0	0	0	0
212	0	0	0	0	0	0	0
213	0	0	0	0	0	0	0
214	0	0	0	0	0	0	0
215	0	0	0	0	0	0	0
216	0	0	0	0	0	0	0
217	0	0	0	0	0	0	0
218	0	0	0	0	0	0	0
219	0	0	0	0	0	0	0
220	0	0	0	0	0	0	0
221	0	0	0	0	0	0	0
222	0	0	0	0	0	0	0
223	0	0	0	0	0	0	0
224	No. of weight-at-age data sets						
225	1						
226	Weight at age on annual basis (year; sex; a1;a2;a3...)						
227	1963	1	1	1	1	1	1

Figure A.1.6 (continued). Biological parameters that are entered as auxiliary information to the model.

	A	B	C	D	E	F	G
230	**Data**						
231							
232	Catch by method and year						
233	502	509	602	545	524	611	810
234	Total NCPUE data points						
235	11						
236	CPUE (Index; Method; Year; Value; CV)						
237	1	1	1978	1.595	0.7		
238	1	1	1979	2.222	0.7		
239	1	1	1980	2.691	0.7		
240	1	1	1981	3.27	0.7		
241	1	1	1982	3.1	0.7		
242	1	1	1983	3.7	0.7		
243	1	1	1984	4.28	0.7		
244	1	1	1985	5.007	0.7		
245	1	1	1986	4.6	0.7		
246	1	1	1987	3.8	0.7		
247	1	1	1988	4.4	0.7		
248	Number of survey data points						
249	11						
250	Survey summary (Index; Year; Value; CV)						
251	1	1981	519	0.7			
252	1	1982	442	0.7			
253	1	1983	596	0.7			
254	1	1984	553	0.7			
255	1	1985	388	0.7			
256	1	1986	952	0.7			
257	1	1987	451	0.7			
258	1	1988	464	0.7			
259	1	1989	506	0.7			
260	1	1990	436	0.7			
261	1	1991	207	0.7			

Figure A.1.7. Data section for the Northern Cod case study.

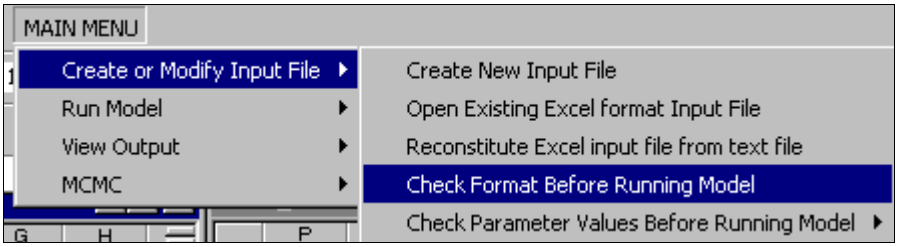
	A	B	C	D	E	F	G
262	Number of commercial catch at age data sets						
263	30						
264	Catch at age data (method; year; sample size; 1; 2 ; 3 ...)						
265	1	1962	50	0	0.0003263	0.0072	0.0427
266	1	1963	50	0	0.0040651	0.01615	0.07753
267	1	1964	50	0	0.0074687	0.05029	0.07178
268	1	1965	50	0	0.0002402	0.01463	0.08113
269	1	1966	50	0	0.0021977	0.03772	0.17708
270	1	1967	50	0	0.0018262	0.03528	0.18002
271	1	1968	50	0	0.0004569	0.00974	0.14959
272	1	1969	50	0	0.000114	0.00837	0.07658
273	1	1970	50	0	0.0169283	0.04494	0.1492
274	1	1971	50	0	0.000086	0.03355	0.18646
275	1	1972	50	0	0.0006018	0.01718	0.20351
276	1	1973	50	0	0	0.01357	0.13969
277	1	1974	50	0	0.001797	0.01228	0.05015
278	1	1975	50	0	0.0023364	0.02207	0.07844
279	1	1976	50	0	0.0001015	0.09315	0.2282
280	1	1977	50	0	0.0007749	0.05114	0.47002
281	1	1978	50	0	0	0.01427	0.18936
282	1	1979	50	0	0	0.01179	0.12649
283	1	1980	50	0	0.0009316	0.02586	0.12177
284	1	1981	50	0	0	0.02543	0.08347
285	1	1982	50	0	0	0.01379	0.2535
286	1	1983	50	0	0.0001484	0.02131	0.11223
287	1	1984	50	0	0.0000246	0.0064	0.12169
288	1	1985	50	0	0	0.00481	0.10973
289	1	1986	50	0	6.42E-06	0.00533	0.0977
290	1	1987	50	0	0.0002794	0.01549	0.06131
291	1	1988	50	0	0.0001513	0.01682	0.08869
292	1	1989	50	0	0.000054	0.01145	0.11911
293	1	1990	50	0	0.0003727	0.04944	0.26064
294	1	1991	50	0	0.0002484	0.02208	0.22468

Figure A.1.7 (continued): Data section for the Northern cod case study.

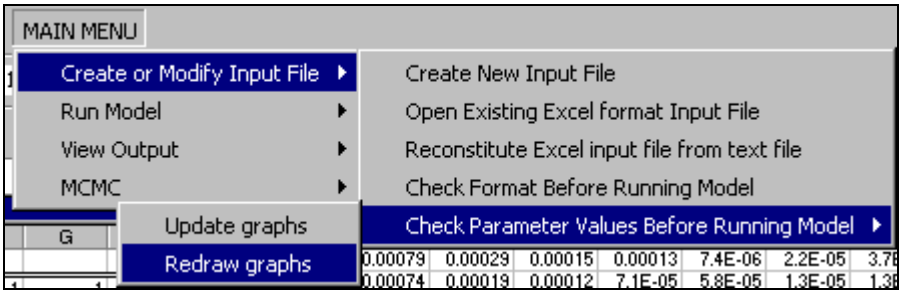
	A	B	C	D	E	F	G	H
295	Nsurvey CA							
296	2							
297	DUMMY DATA: Survey catch at age data (survey; year; sample size; a1; a2; a3...)							
298	1	1964	1	1	1	1	1	1
299	1	1965	1	1	1	1	1	1
300	Number of commercial catch at length data sets							
301	2							
302	DUMMY DATA: Commercial catch at length data (method; year; sample size; 11; 12; 1							
303	1	1964	1					
304	1	1965	1					
305	Number of survey C@L							
306	2							
307	DUMMY DATA: Survey catch at length data (method; year; sample size; 11; 12; 13...)							
308	1	1964	1					
309	1	1965	1					
310	Number of survey no-sex C@L data sets							
311	2							
312	DUMMY DATA: Survey no-sex C@L data (method; year; sample size; 11; 12; 13...)							
313	1	1964	1					
314	1	1965	1					
315	EndOfForm							
316								

Figure A.1.7 (continued). Data section for the Northern cod case study.

At this point, all the data should be in the Input sheet, but before running the estimation model we recommend checking for missing data and inconsistencies in the spreadsheet. *Coleraine* has a built-in command that allows the user to do so: *Check Format Before Running Model*. Potential problems are highlighted in red or yellow. After running this command, check for colored cells in the entire spreadsheet.



The next step is to do some deterministic projections before invoking the estimation program. The motivation behind this is to get reasonable starting values for all the model parameters, conditioned on the available data. This exercise can be easily done by using the following command:



This procedure generates a group of graphs containing model predictions of vulnerable biomass, spawners and recruits, selectivity pattern, CPUE, and Survey Index (Figure A.1.8).

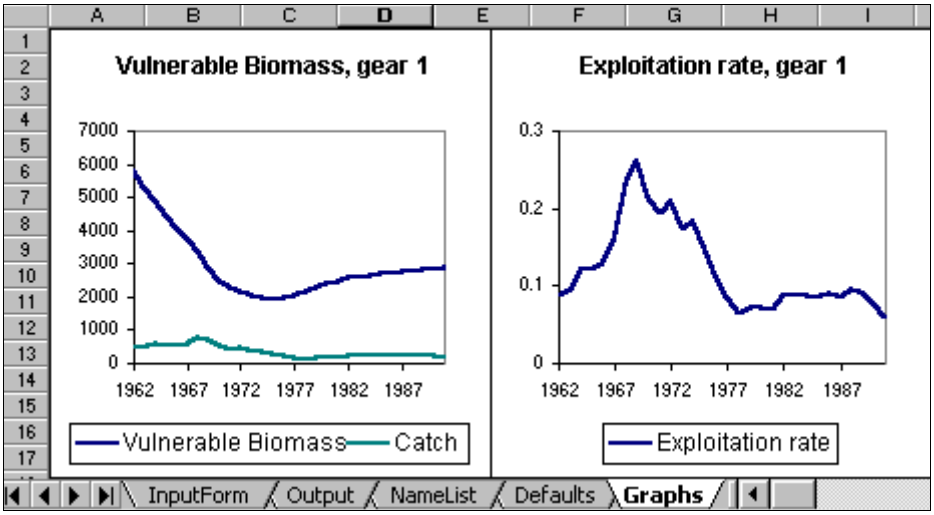


Figure A.1.8. Deterministic model outputs before running the estimation model.

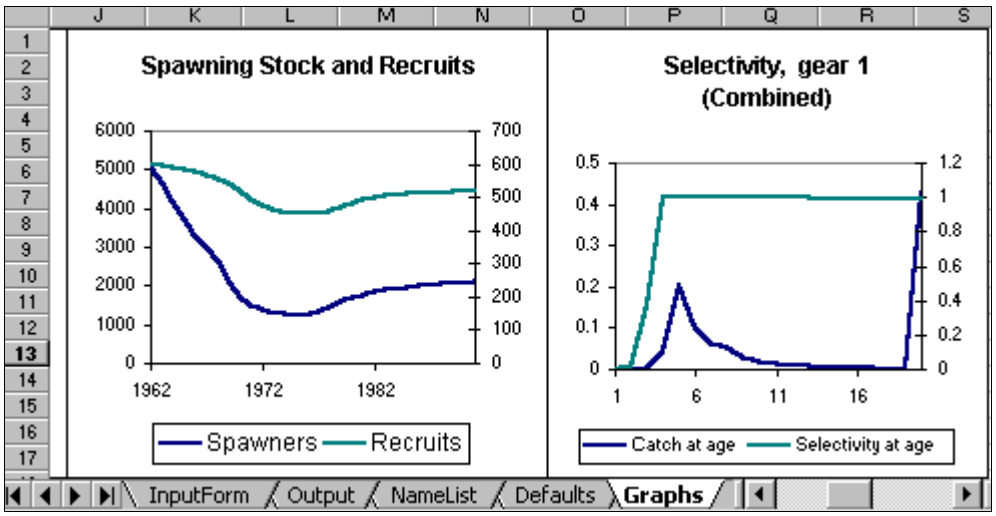


Figure A.1.8 (continued). Deterministic model outputs before running the estimation model.

This arrangement allows the user to change the initial guesses of the input parameters and redraw the graphs in an iterative process.

3. Running the Estimation Program

At this point, the estimation phase should be ready to start. Line 5 contains a switch for including a debugger into the estimation program during the input process of the data. If it is set to 1, than the user will need to type "*n*" and press "*return*" whenever

the cursor prompts for some entry. This basic mechanism allows us, to some extent, to trace the problem at runtime.

To run the program use the following command:

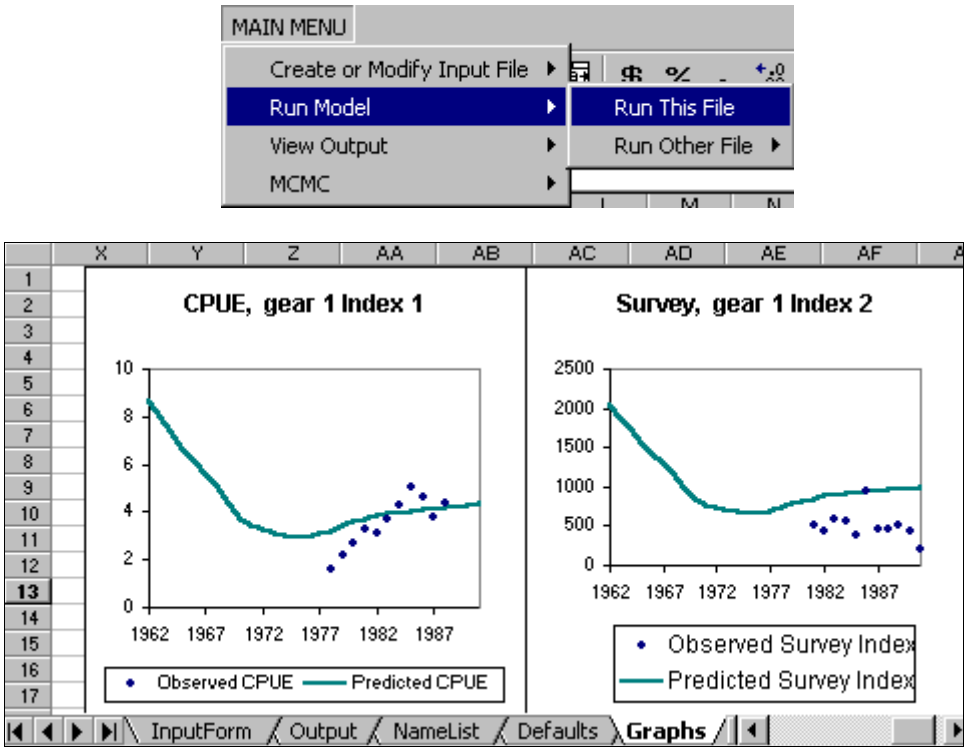
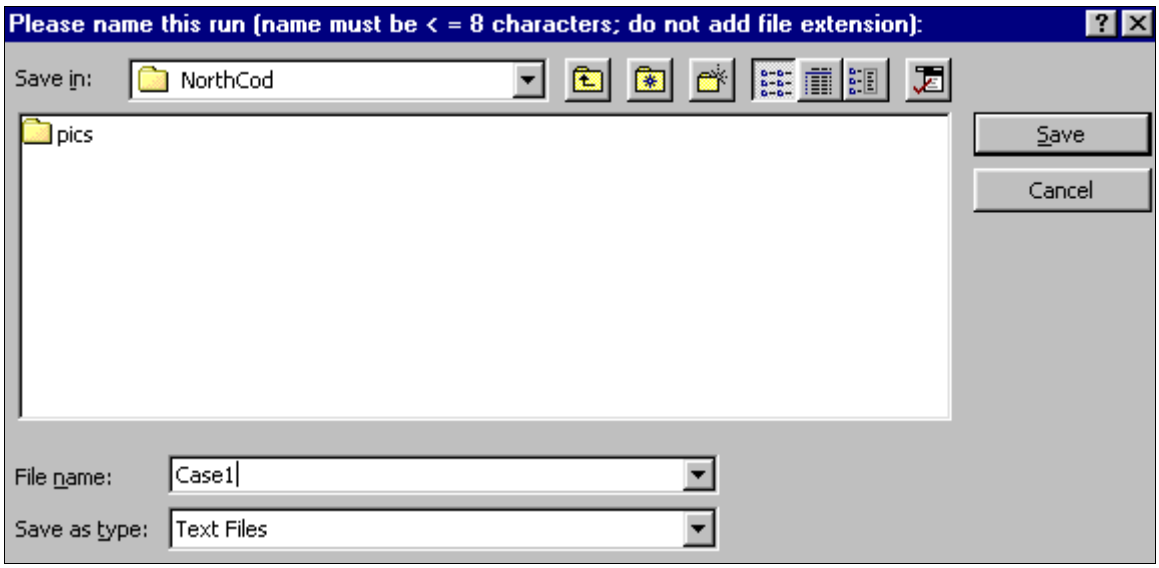
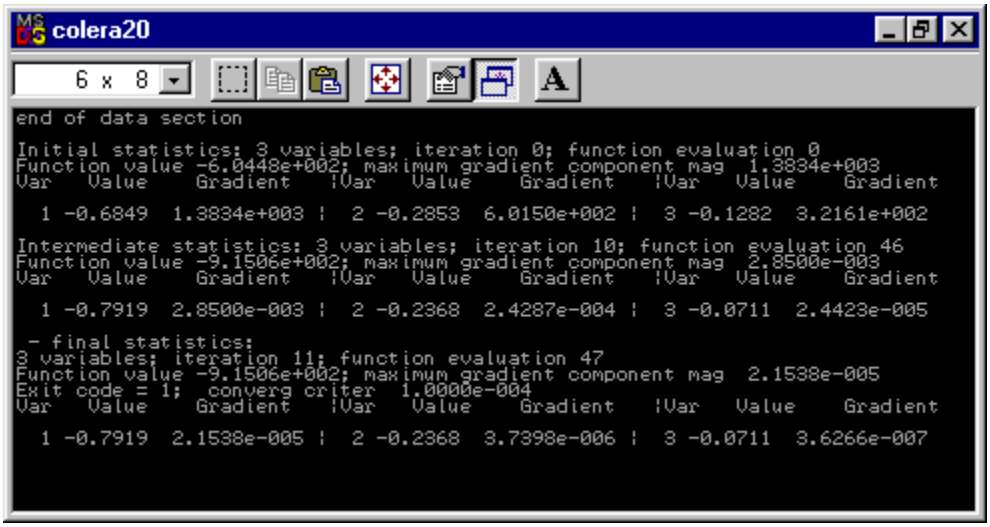


Figure A.1.8 (continued). Deterministic model outputs before running the estimation model.

This process will invoke a dialog box, where you can specify the name of the text file that contains all the information of this run (e.g. Case1).

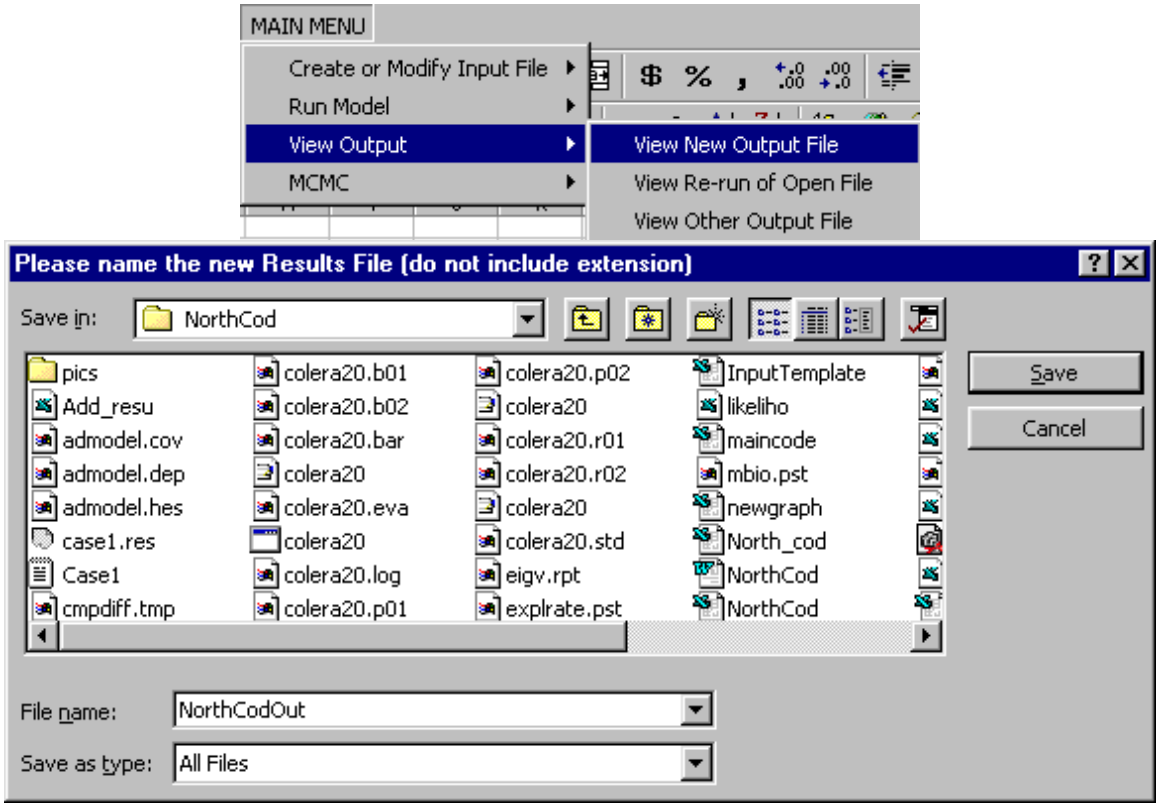


After this is completed, a DOS window will appear, showing valuable information of phases, number of estimated parameters, total value of the objective function and gradients associated with each parameter.



4. Creating an Output Viewer

The output of the estimation model is a text file, and as such it does not have a very user-friendly graphic environment. Coleraine was design to automate the creation of graphic outputs. The creation of this viewer is done using the following commands:



This procedure creates an Excel file that contains several sheets. Some of them have controls to specify the setup of the graphs. The user can also specify which graphs should be created.

The worksheet *General* has several graphs, which are shown in Figure A.1.9. It includes vulnerable biomass, recruitment, spawners, and harvest rate trends. The outputs for the age-structured data are reported in the sheet ComC@A (Figure A.1.10). This sheet has many entries to control the graph setup.

Other outputs of interest for this dataset are shown in Figure A.1.11. Fits of predicted CPUE and relative index of abundance of the surveys to the observed values and selectivity ogives are displayed in different sheets.

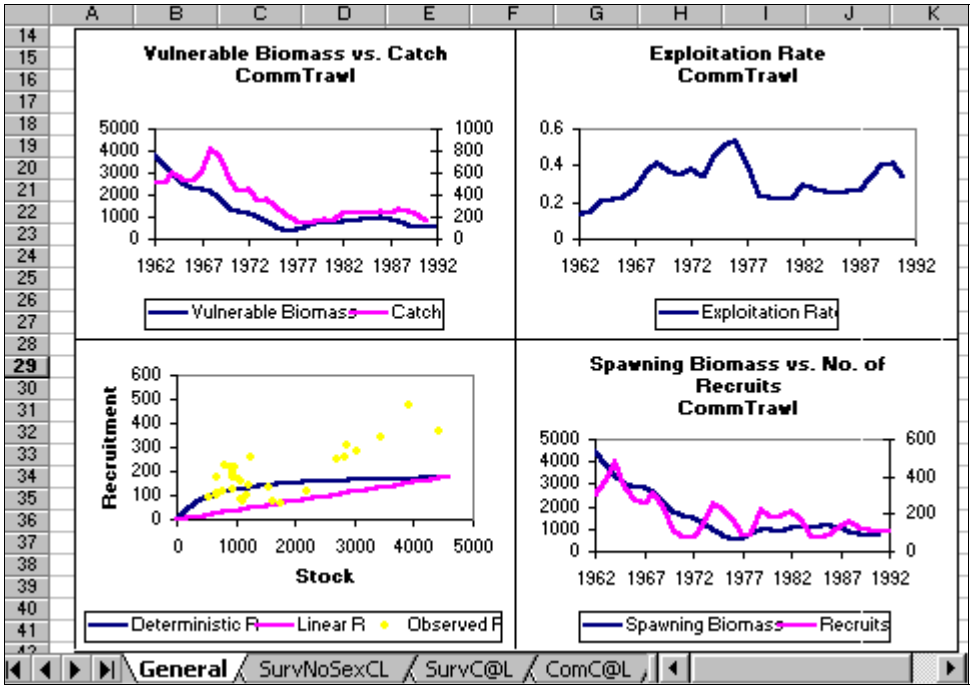


Figure A.1.9. Graphs of some major model outputs.

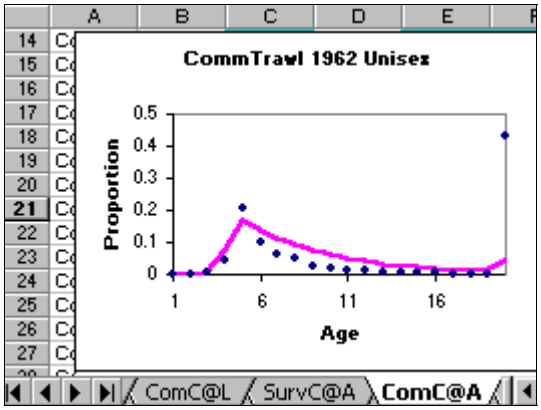


Figure A.1.10. Graphs of commercial observed and predicted catch-at-age data.

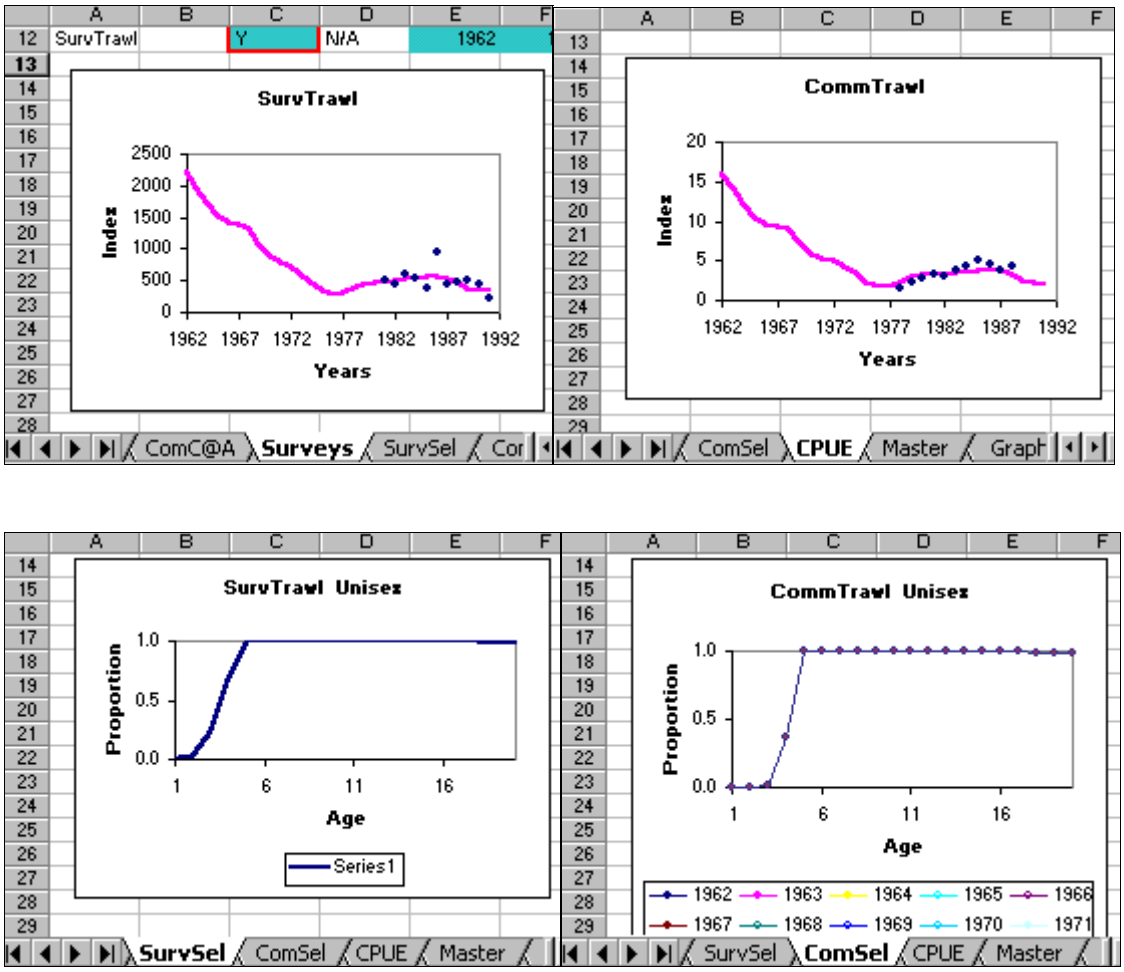
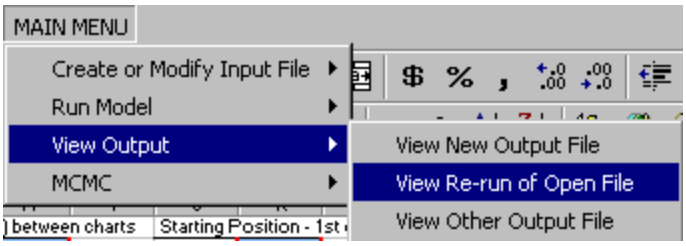


Figure A.1.11. Graphs of predicted and observed survey index and CPUE. Selectivity of the survey and commercial gears are presented in the Sheets *SurvSel* and *ComSel* respectively.

One of the most useful features of *Coleraine* is the possibility of doing multiple successive runs and organizing and summarizing the output information. The Excel sheet called *Tracker* stores vital information for the different runs (Figure A.1.12), such as file path, name of input files, likelihoods, parameter values, and phases. *Tracker* gets updated during the redraw process of the output viewer.

5. Updating the Information of Re-runs

Once the new entries of the parameters and likelihoods have been specified for another run, then steps 3 and 4 can be repeated. Each new run can have a different name, so that you can return to that file at any time. The only variation in this repeated sequence of events is the use of the *View Re-run of Open file* command. This will just update the graphs and not create new output files.



After updating the output file 3 or 4 times, the program might crash because of memory problems in your system. If that happens close the output file and re-open it, following the updating sequence again.

tracker

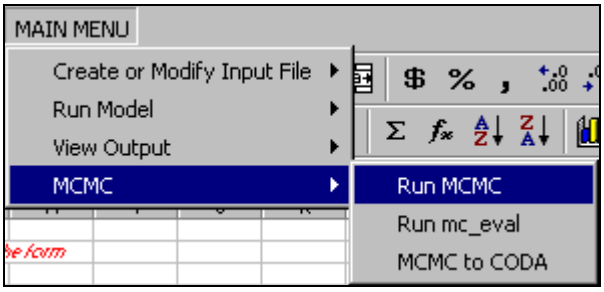
	A	B	C	D	E
1	**File Path**	C:\ernst\RA_ray\M_A_N_U_A_L\NorthCod			
2					
3					
4		**RUN15**			
5	**Input File**	case1.txt			
6					
7	**Likelihoods**				
8	CPUE	0.53603			
9	Comm. CA	-1053.61			
10	Comm. CL	0			
11	Survey	0.83056			
12	Survey CA	0			
13	Survey CL	0			
14	Surv. CL NoSex	0			
15	Penalties	15.0387			
16	Survey CL 2				
17	**Parameters**				
18	R0	195.312			
19	h	0.7			
20	MUnisex	0.2			
21	**Rinit**	1			
22	uinitUnisex	0			
23	plusscaleUnisex	1			
24	Sfullest	5			
25	log_varLest	0			

Figure A.1.12. Summarization of one run in the Excel-sheet *Tracker*.

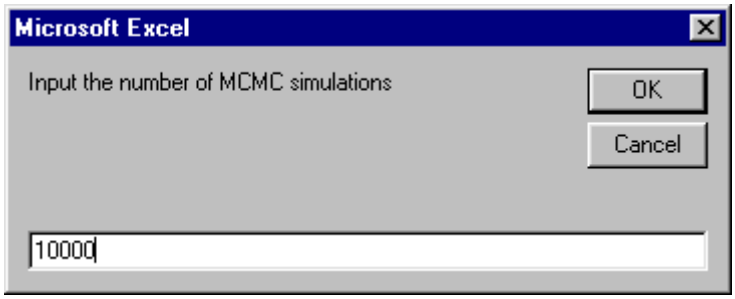
6. Forward Projections under Different Harvest Policies

Coleraine allows the user to explore and evaluate the consequences of future management actions. This is done in a Bayesian framework by using a Markov Chain Monte Carlo (MCMC) method. Having specified priors and likelihoods in the model, the estimation model uses numerical techniques for obtaining the posterior probability distribution for model and output parameters. To specify the simulation conditions, we need to go back to lines 82–96 (Figure A.1.3). These entries allow the user to set the extension of the forward projection and the type of harvest strategy involved in the analysis. In any case, we need to set the starting, ending, and step values of the chosen harvest strategy. This will allow us to evaluate the consequences associated with each strategy value. *Coleraine* will estimate posterior probabilities for the time series of

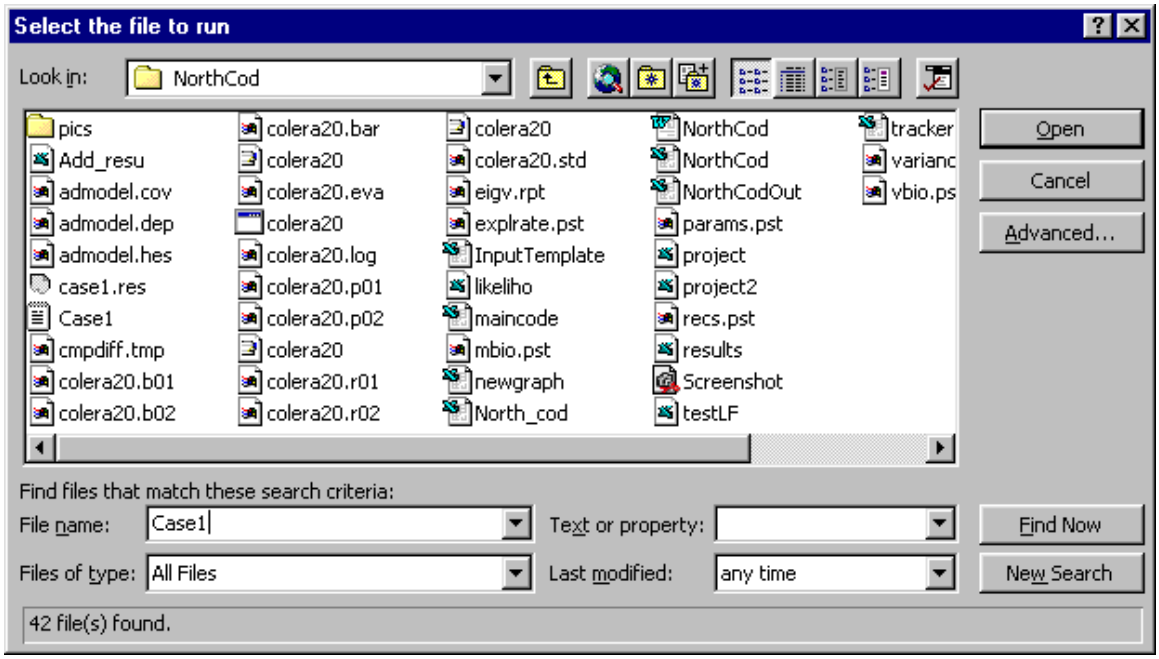
projected spawning stock size and vulnerable biomass. These results are output to a text file called *project.out*. To run the projections, we use the *Run MCMC* command:



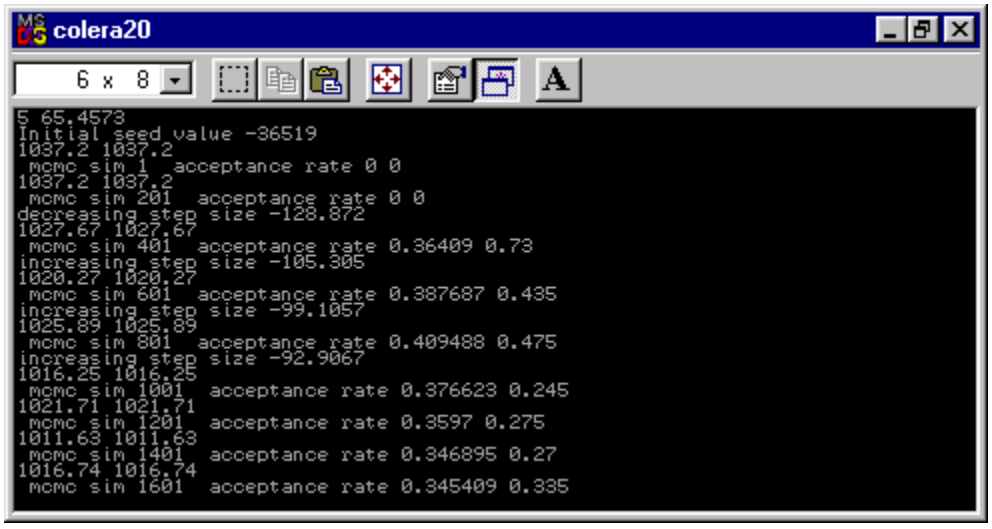
This action will prompt you to enter the number of MCMC simulations. This number will depend on the number of iterations needed to satisfactorily describe the posterior probability of the output parameters. In some cases this could be a long process, taking millions of iterations.



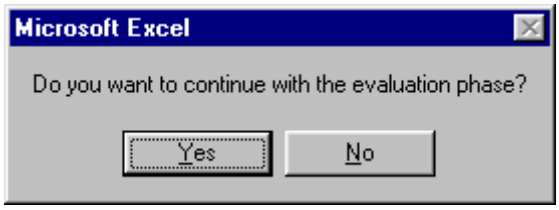
The program will prompt you to enter the text file name of the specific run you are interested in. If you introduced changes to the Excel input sheet, you will need to run the estimation program before running MCMC, in order to pass these modifications to the text file.



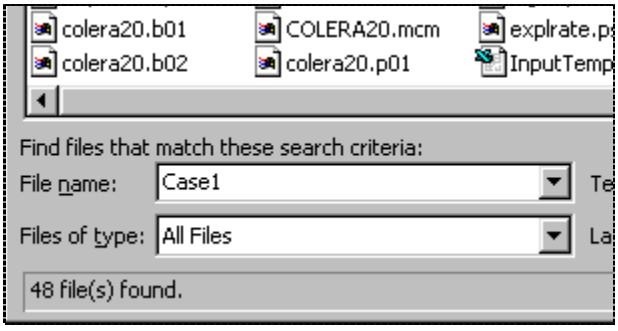
During the execution of this program you will see the following DOS window:



This window displays the simulation number and the acceptance rate. This statistic is a diagnostic element for convergence and should tend towards values between 0.15 and 0.4. When the program is done with the simulations, it will generate a *colera20.psv* file. This file contains values of all the model parameters collected from the Markovian chain. The natural step after the simulation is to generate the prediction given the desire harvest policy. The program will ask the user to continue with the projection phase through the following dialog box:



It asks the user again to input the text file:



7. Organizing the Projection Outputs

The *projection.out* file contains the values for the virgin vulnerable biomass, virgin spawning biomass, projected biomass, and spawning biomass in each year, for each of the chosen harvest levels (100–600).

	A	B	C	D	X	Y
1	Virgin_Vulnerable_Biomass	Virgin_Spawning_Biomass	Proj_Biom_1992	Proj_Biom_1993	Spawners_	Catch_199
2	4269.8	4440.28	544.159	620.441	2024.97	100
3	4269.8	4440.28	544.159	516.653	557.915	200
4	4269.8	4440.28	544.159	412.865	202.179	300
5	4269.8	4440.28	544.159	309.077	185.806	400
6	4269.8	4440.28	544.159	205.289	169.871	500
7	4269.8	4440.28	544.159	159.477	160.888	544.139
8	4269.8	4440.28	544.159	620.437	1076.23	100
9	4269.8	4440.28	544.159	516.649	98.7147	200
10	4269.8	4440.28	544.159	412.861	73.2243	300
11	4269.8	4440.28	544.159	309.073	63.5082	400
12	4269.8	4440.28	544.159	205.285	59.7297	500
13	4269.8	4440.28	544.159	159.473	57.6823	544.139
14	4269.8	4440.28	544.159	620.436	1292.35	100
15	4269.8	4440.28	544.159	516.647	181.563	200

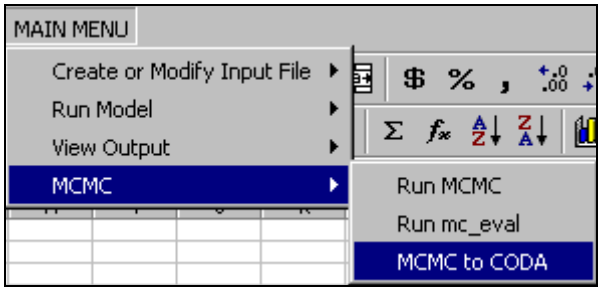
Figure A.1.13. Output file from the policy evaluation (*mceval* phase).

Figure A.1.13 show the set of evaluation policies (rows 2–7, 8–13, etc) for each MCMC simulation that was stored. *Coleraine* automatically specifies the spaces between saved values, in order to get output files with 1000 simulations. In other words, if we have 6 catch levels, we should have 6000 rows of data in our *projection.out* text file.

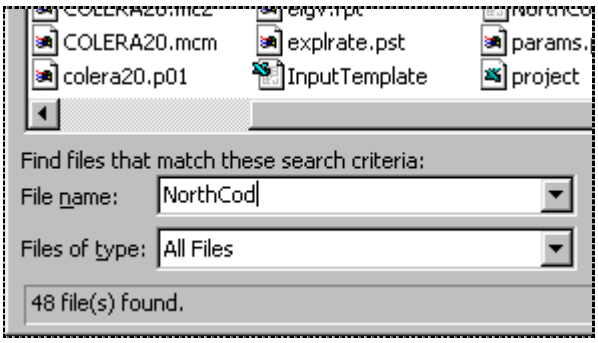
This data can be sorted and analyzed in many different environments. One of these is CODA (Convergence Diagnosis and Output Analysis Software), an S-Plus function written to analyze Gibbs sampling outputs. It has many built-in statistical methods for monitoring convergence of Markov chains. For downloading the software and documentation, the user can visit the following website:

http://www.mrc-bsu.cam.ac.uk/bugs/classic/coda04/readme.shtml.

CODA has a very specific input data structure, which is not compatible with *projection.out* file. *Coleraine* enables us to create two necessary input files (*label.ind* and *data.out*) by using the following command:



You need to specify the Excel input file that you are using in this run.



The next step is to chose the predictions in which you are interested (check off the boxes in the dialog box). The program will create two Excel files and you need to save them as text files with the names *label.ind* and *data.out*.

APPENDIX B

General description of all the necessary input information for running the program.

DEBUGGER

0 = On **1** = Off	Turn the debugger on/off. It is part of the executable code and helps the user to find errors at runtime.

RUN OPTIONS

1	**0** = no / **1** =yes	Use observed weight at age data?
2	**1** = default **user defined value [0, 1]**	Upper bound for the exploitation rate.
3	**0** = Read in **1** = Gear 1 **2** = Gear 2	Gear that was operating at the beginning of the fishery. If value = 0, than it uses ${}_I v_a^S$, otherwise it is one of the specified selectivities.
4	**0** or **1**	In the projections only one gear is used. If multiple gears were used in the estimation, then a weighted average of the current year selectivity ogives is used. (0 = the user specifies the weights assigned to each gear; 1 = use same gear proportions as last year of data (weights are the exploitation rates)

DIMENSIONING PARAMETERS

1		Number of available CPUE indices for the commercial fleet. This value can be greater than the number of fishing gears.
2		Number of Survey indices (number of different types of indices, not the actual number of data points).
3		Start year of the estimation program.
4		End year of the estimation program.
5		Number of ages used in the numbers-at-age matrix.
6		Age at which the catch-at-age data are pooled.
7		First age with effective catch-at-age values in the datasets.
8		First length with effective catch-at-length values in the datasets.
9		Length increment in the catch-at-length dataset (same for Commercial and Survey data).
10		Number of length intervals in the catch-at-length dataset.
11		Length interval at which catch-at-length data are pooled.
12		Number of different commercial fishing methods.
13		Number of sexes considered in this analysis (1 = pooled).
14		Total number of available CPUE data points for all commercial indices.
15		Total number of observed commercial catch-at-age datasets (= <years*gears).
16		Total number of observed commercial catch-at-length datasets (= <years*gears).
17		Total number of weight-at-age datasets. If you are not including the

		growth and allometric parameters of the length–weight relationship, you should enter (Start year—End year) rows of data.
18		Total number of survey index data points (all types combined).
19		Total number of observed survey catch-at-age datasets (= <years * survey gears).
20		Total number of observed survey catch-at-length datasets (= <years* survey gears).
21		Number of observed survey catch-at-length datasets where sex was not determined.
		Survey vulnerability type (1 = age-based; 2 = length-based).
22		Age of full recruitment to the fishery. This information is used for building a default maturity and initial selectivity ogive.

PROJECTION PARAMETERS

After fitting the model to the data, these parameters are used for evaluating policy options.

1		Strategy type used in the forward projections; Constant catch (**1**) and Harvest rate (**2**).
2		Year to which the projections should be carried out.
3		Lower limit for constant catch or harvest rate strategy.
4		Upper limit for constant catch or harvest rate strategy.
5		Increment in the strategy.
6		Proportions used to weight each selectivity curve in the current year to average them out (needs to add up to 1 [one]).
7		Coefficient of variation in the population size estimates of the future assessments (only for constant harvest rate strategy).
8		Degree of autocorrelation in the assessment errors (only for constant harvest rate strategy).

PRIORS FOR THE MODEL

Seven entries are required for each row in the *Priors* section:

a. ***Phase*** *(column A): The phase determines the order in which the specified parameter will be estimated during the nonlinear search. (A value of 1 means that this parameter will be estimated in the first phase, 2 in the second and so on. Any negative number implies that the parameter is fixed at the value specified in the starting conditions and will not be estimated by the model.)*

b. ***Bounds****: These correspond with the lower (Column B) and upper (Column C) bounds imposed by the user on the parameter values in the constrained optimization.*

c. ***Prior type*** *(column D): Three types of priors are available in this model: uniform = 0, normal = 1 and lognormal = 2.*

d. ***Mean*** *(column E): Mean of the normal or lognormal distributions. If the prior is uniform, a dummy value must be entered.*

e. ***CV****(column F): Coefficient of variation of the probability distribution. If the prior is uniform, a dummy value must be entered.*

f. ***Default value*** *(column G): Initial guess for the parameter. This data will be used in the maximum likelihood estimation of the parameters. In many cases the final estimates depend on the initial guesses, so several combinations of initial values should be explored.*

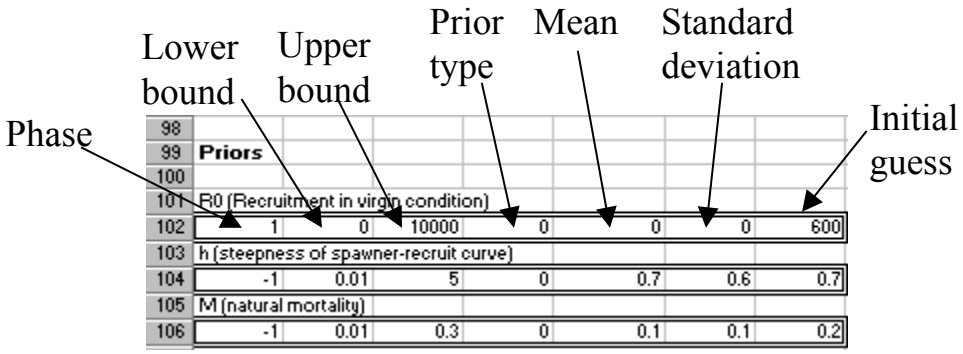


Figure B.2.1. Location of each of the 7 entries for each parameter

1	R_0	Virgin recruitment from the Beverton-Holt recruitment model.
2	h	Steepness parameter from the Beverton-Holt recruitment model. This parameter is defined in the range [0.2, 1.0].
3	M_s	Natural mortality by sex.
4	${}_I\varepsilon_a$	Initial age structure residuals. They go from age 2 to nages -1. They are normally distributed with mean 0 and variance ${}_I\sigma^2$. *The bounds and initial guess are in natural logarithmic scale*.
5	${}_R\varepsilon_t$	Initial age structure residuals. They go from age 2 to nages -1. They are normally distributed with mean 0 and variance ${}_R\sigma^2$. *The bounds and initial guess are in natural logarithmic scale*.
6	ω	Fraction of R_0 in the initial study year. This parameter is defined in the range [0,1].
7	${}_I u$	Exploitation rate in the initial state. This parameter is defined in the range [0 , <1.0].
8	${}_P\varepsilon_A$	Residual term in the plus group at the initial state. They are normally distributed with mean 0 and variance ${}_A\sigma^2$. *The bounds and initial guess are in natural logarithmic scale*.
9	$S_{full}{}^{g_i}$	Age of full selectivity for females of gear type (*I*) of the fishing fleet.
10	$\Delta_{S_{full}}{}^{g_i}$	Age difference between sexes in the age of full selectivity of fishing fleet gear type (*i*).
11	${}_L v^{g_i}$	Variance of the left half Gaussian distribution of the selectivity of gear (*i*). *The bounds and initial guess have to be in natural logarithmic scale.*
12	${}_R v^{g_i}$	Variance of the right half Gaussian distribution of the selectivity of gear (*i*). *The bounds and initial guess have to be in natural logarithmic scale.*
13	${}_{S_{full}}\varepsilon_t^{g_i}$	Residuals of age of full selectivity by gear (*i*) at time (*t*). The number of parameters per gear to be estimated will depend upon the number of

		years catch-at-age data are available. They are normally distributed with mean 0 and variance $_{Sfull}\sigma^2$.
14	$_{L^v}\varepsilon_t^{g_i}$	Residuals of variance of the left half Gaussian distribution of the selectivity of gear (*i*) at time (*t*). The number of parameters per gear to be estimated will depend upon the number of years catch-at-age data are available. They are normally distributed with mean 0 and variance $_{L^v}\sigma^2$.
15	$_{R^v}\varepsilon_t^{g_i}$	Residuals of the variance of the right half Gaussian distribution of the selectivity of gear (*i*) at time (*t*). The number of parameters per gear to be estimated will depend upon the number of years where we do have catch at age data. They are normally distributed with mean 0 and variance $_{R^v}\sigma^2$.
16	q^{CPUE_i}	Catchability coefficient of CPUE series (*i*). *The bounds and initial guesses have to be in natural logarithmic scale.*
17	$_q\varepsilon_t^{CPUE_i}$	Temporal residuals of the catchability coefficient of the CPUE series (*i*). They are normally distributed with mean 0 and variance $_q\sigma^2$
18	q^{S_j}	Catchability coefficient of survey index (*j*). *The bounds and initial guesses have to be in natural logarithmic scale.*
19	$S_{full}^{S_j}$	Age of full selectivity for females of survey gear type (*j*).
20	$\Delta_{S_{full}}^{S_j}$	Age difference between sexes of the age of full selectivity of the survey gear type (j).
21	$_Lv^{S_j}$	Variance of the left half Gaussian distribution of the selectivity of survey gear type (*j*) at time (*t*). *The bounds and initial guesses have to be in natural logarithmic scale.*
22	$_Rv^{S_j}$	Variance of the right half Gaussian distribution of the selectivity of survey gear type (*j*) at time (*t*). *The bounds and initial guess have to be in natural logarithmic scale.*
23		Not used in the model.
24		Not used in the model.
25		This dummy parameter is not involved in the calculations and must be excluded from any estimation phase. It is useful for doing deterministic projections (no estimation involved) and to be able to start in any iteration *mceval* from the maximum likelihood estimates.

LIKELIHOODS

The different entries for the likelihoods are as follows:

-1	*Not used, but likelihoods and predictions are evaluated in last run*
0	*Not used*
1	*Normal*
2	*Lognormal*
3	*Robust normal*
4	*Robust lognormal*
12	*Robust lognormal for proportions*

1		Likelihood type for the CPUE data (from 1 to number of CPUE

		indices).
2		Likelihood type for the commercial catch-at-age datasets (from 1 to n methods).
3		Likelihood type for the commercial catch-at-length datasets (from 1 to n methods).
4		Likelihood type for the Survey indices (from 1 to n Survey indices).
5		The survey indices are in weight (**1**) or numbers (**2**).
6		The survey vulnerability is based on age (**1**) or length (**2**).
7		Likelihood type for the survey catch-at-length not separated by sexes.
8		Likelihood type for the survey catch-at-length separated sexes (1 to n Survey indices)
9		Likelihood type for the survey catch-at-age separated by sexes (1 to n Survey indices).

FIXED PARAMETERS

Rows 1–10 are sex-specific parameters (if using a model separated by sexes). First column = female; second = males.

1	b_i^S	Proportionality parameter of the allometric length–weight relationship.
2	b_{ii}^S	Exponent parameter of the allometric length–weight relationship.
3	L_∞^S	Asymptotic length of the von Bertalanffy growth model.
4	k^S	Von Bertalanffy *k* parameter.
5	t_0^S	Von Bertalanffy t_0 parameter.
6	ρ^S	Brody parameter.
7	L_1^S	Mean length of individuals of the first age.
8	L_n^S	Mean length of individuals of the last age.
9	${}_{L_1}\sigma^S$	Standard deviation of the Gaussian distribution that describes the variability around the mean length of individuals of the first age.
10	${}_{L_n}\sigma^S$	Standard deviation of the Gaussian distribution that describes the variability around the mean length of individuals of the last age.
11	Φ_a	Maturity at age of females in the population (ogive).
12	λ	Female fraction of the total that recruits every year.
13	${}_I v_a^S$	Vulnerability at age (by sex) at the beginning of the fishery.
14	Ω^S	N-ages x N-ages upper triangular aging error matrix.
15		Number of weight-at-age dataset.
16		Annual weight-at-age. First row = **year**; second row = **sex**; others = **weight-at-age** starting from age 1 to age n. If you do not have estimates of parameters 1 - 5, you will need to enter observed weight-at-age data from the first year to the last year + 1. The same applies if you are using a model in numbers (not in biomass units), but you will have to set all weight-at-age values to 1.

DATA

1		Total catch in weight by year. One row for each commercial fishing method. It needs to match with the total number of year of the estimation model.
2		Total number of observed CPUE data points.
3		CPUE matrix. First column = **Index** number; second column = **method** number; third column = **year** (4 digits); fourth column = **point estimate**; fifth column = **coefficient of variation**.
4		Total number of observed Survey index data points.
5		Survey index matrix. First column = **Index** number; second column = **year** (4 digits); third column = **point estimate**; fourth column = **coefficient of variation**.
6		Number of observed commercial catch-at-age datasets.
7		*Commercial catch-at-age data*: The sample size will weight the likelihood. First column = **method** (number); second column = **year** (4 digits); third column = **sample size;** others: Starts at age 1 of females and goes to pool age. In the next entry to the right, it starts with the 1-year-old males and goes to pooled age. These values are proportions that, for both sexes combined, add up to 1 (one).
8		Total number of observed survey catch-at-age datasets.
9		*Survey catch-at-age data*. First column = **index**; second column = **year**; third column = **sample size**; others: Starts at age 1 of females and goes to pooled age. In the next entry to the right it starts with the 1-year-old males and goes to pool age. These values are proportions that for both sexes combined add up to one.
10		Total number of observed commercial catch-at-length data (separated by sex).
11		*Commercial catch-at-length data (separated by sex).* First column = **method**; second column = **year**; third column = **sample size**; others: Starts at the first defined length interval of females and goes to pooled length. In the next entry to the right it starts with the first defined length interval for males and goes to pool age. These values are proportions that for both sexes combined add up to 1 (one).
12		Total number of observed survey catch-at-length data (separated by sex).
13		*Survey catch-at-length data (separated by sex).* First column = **index**; second column = **year**; third column = **sample size**; others: Starts at the first defined length interval of females and goes to pooled length. In the next entry to the right, it starts with the first defined length interval for males and goes to pool age. These values are proportions that for both sexes combined add up to 1 (one).
14		Total number of observed survey catch-at-length data (not separated by sex).
15		*Survey catch-at-length data (not separated by sex).* First column = **index**; second column = **year**; third column = **sample size**; others: catch-at-length data.

APPENDIX C

Steps in model fitting for age-structured models using *Coleraine*

Obtaining MLE estimates

The first step in model fitting is to get the best possible fit to the data. Since *Coleraine* uses both the likelihood of the fit to the data and any penalties based on priors, many people argue that this should not be called Maximum Likelihood but, instead, call it "mode of the posterior estimation" or MPE. However, if you think of priors as additional forms of likelihood, then this is still an MLE fitting—and it is really all semantics anyway unless you have some very strong and informative priors. Thus, I will call it MLE, even if it is really MPE.

Rule 1: Graph the data and the fit

The first task in fitting an age-structured model is to get the model to fit the data. You will have two major ways of judging the goodness of fit: the numerical values of the negative log-likelihoods as found in the master worksheet of the viewer (and *Tracker*), and in the graphs. It is difficult to overemphasize the importance of actually looking at the data to make sure it is fitting. You can't look at a negative log likelihood and say whether the fit is good—they can only be used relatively—you must plot the data and the fits and see if you are coming close to the points. Once you get close, then you can start looking at the negative log likelihoods for fine-tuning. In general it is often difficult to see any differences in the graphs when the likelihoods change by only a few units.

Rule 2: Start with good beginning parameter values

The key issue here is that you usually need to start with very good parameter starting values to ensure that the non-linear search algorithms work well. Generally the important parameters are the average level of recruitment (R_o being the key parameter), the q's for the Surveys and CPUE, and the selectivities of survey and commercial gear. You must get these within 50% or so of the actual MLE best fit to get good performance. The VisualBasic software in *Coleraine* lets you view deterministic trajectories and their fits. Use this feature before starting non-linear fitting.

The key to getting parameters "about right" is to have an average exploitation rate in mind. We recommend 20%. As a starting estimate, assume that the fishery has taken about 20% of the population.

An estimate of exploitation rate lets us start with the parameter R_o. Begin by understanding the relationship between average recruitment (R_o usually) and vulnerable population size and spawning stock size. This is the so-called spawning-biomass-per-recruit (SBPR), or for vulnerable stock it is vulnerable-biomass-per-recruit. If the weight of individual fish is in kilograms, then the usual SBPR will be numbers between 1 and 10. It naturally depends upon natural mortality, growth and age at maturity, but 1 to 10 encompasses most of the SBPR's we have observed. Thus, if SBPR is 5, and R_o is 1,000 then we would have 5,000 units of SBPR in the unfished state. Usually catches are measured in tons, so this mean that 1 recruit (with weights in kilograms) is really a thousand recruits when compared to catches in tons. Thus, if catches are on the order

of 1,000 tons, then a R_o of 1,000 would imply an exploitation rate (on an unfished stock) or 20%. As a starting estimate, choose an R_o so that the exploitation rate would average about 20%.

You should always have these numbers in mind—SBPR (or vulnerable-biomass-per-recruit), average levels of catch—and an estimate of the likely exploitation rate. If you have measured weight of individuals in grams, then each unit of recruitment will be 1,000,000 fish (assuming catches are still in tons).

We can use our target exploitation rate of 20% to help us define the starting estimates of q's for CPUE and surveys as well. Assume that the average CPUE is .5 and the average catch is 1,000 (tons). If exploitation rate is 20%, then the average biomass is 5,000, and thus q is .5/5000 or 1/10000.

Getting starting estimates for the selectivity values is a little tougher. Initially, it is usually simplest to assume that males and females have the same selectivity, and that there is not a descending right-hand limb. The age of full selectivity can be set either to the age at which catch-at-age is maximum or the age that corresponds to the mode of the length frequency data. We have found that setting the left-side standard deviation of the selectivity curve to 2 (log value .69) works well as a starting condition (i.e., it is not knife-edged but it isn't too broad).

Once you have starting values that give a trajectory that does not go extinct from the known removals, and also shows an exploitation rate is the 20% range, then it is time to run the non-linear convergence using the "run this model" option.

Start with the simplest assumptions

It is important to do your first fitting with the simplest possible model. Begin by assuming unfished equilibrium. Estimate R_o, selectivities, fixing right-hand side.

Do not estimate recruitment residuals

Do not estimate initial conditions

Do not estimate M

Do not estimate steepness

The next step is to free up recruitment residuals

Then free up initial conditions

Then explore M and steepness

Phases

First, however, you have to specify the phases for each parameter. This sequence of phases may be very important, or it may not matter to much. In general put the most important parameters in the early phases, and the less important ones in the last phases. For data series that show a decline in abundance as fishing increases (the so-

called one-way-trip), the key parameters will usually be R_o (and thus initial population size), and the q's. As a general rule. we suggest

Phase 1: q to fit the index data.
Phase 2: R_o to adjust the trajectory.
Phase 3: Selectivity parameters.
Phase 4: Recruitment residuals.
Phase 5: Changes in selectivity.
Phase 6: Deviation in initial conditions.

However, in some analyses that have considerable catch-at-age data with clear strong cohorts, you cannot fit the catch-at-age data without getting the recruitment residuals right, and you need to move the recruitment residuals up to the first or second phase. One of the limitations in the current (November 2001) implementation of Coleraine is you cannot specify starting estimates for year class strengths, so if you have dominating catch-at-age data you need to get the year class strengths more or less right (by putting them in phase 1 or phase 2) before trying to estimate the selectivities and possibly even the q's.

Assuring convergence

How do you know if the model is actually finding the best possible fit? Step 1 is first to look at the fit plotted against the data. In general if you are fitting the data well, the model has almost certainly converged. Step 2 is to restart the estimation from different initial points (usually different R_o, q and selectivity values). If the fits converge on the same point, it is a pretty good sign that the model has converged. If you both get a good fit to the data and multiple starting points converge to the same place, you can be fairly sure you have reached a true minimum of negative log likelihood. A third way is to use the AD Model Builder feature of estimating the Hessian matrix (derivative of the negative log likelihood with respect to each parameter). If this matrix is not positive definite, then the model has not converged.

Note checks during fitting

When doing sensitivity runs you always need to be aware that you might not get convergence when you free up additional parameters or change the value of a fixed parameter. Here you can often detect problems by looking at the tracker output. If you free up an additional parameter (for instance freeing up natural mortality or the right-hand side selectivity), and the likelihood does not get better (smaller), then something is wrong. Anytime you make more parameters free, the model must fit as well or better. When looking at *Tracker* you have to look at the total likelihood; some data sources may fit worse, having been sacrificed to fit other sources better. Note that if you change the sample sizes in length or age-frequency data, or take elements out of the likelihoods used, then the total likelihood will be changed and you can no longer directly compare likelihoods.

Another way to check on convergence and model behavior is to do a likelihood profile by manually running the model with different values of a key parameter, such as R_o or M. For instance you might run the model with M fixed at .1, .15, 2, .25, .3 and .35. Then also run the model with M estimated. The likelihood with M estimated should be the lowest

one, and a plot of likelihood vs M should be smooth with a clear minimum at the value estimated when the parameter is free.

Bayesian integration

Finding the MLE is just the beginning—it is one fit to the data that happens to be "best" by the definition of maximum likelihood, but the real parameter values may be something else and we want to find the probabilities associated with other possible parameter values (i.e., explore the uncertainty in the real parameters). Thus, we use the MLE as a jumping off point for MCMC, which does a random walk over the posterior space. Typically we let the MCMC do about 1,000,000 samples from the posterior space, and keep every 1000'th of these samples.

Assuring MCMC convergence

We recommend that you start by doing 1,000,000 MCMC draws and saving 1,000 of these. We know we will have started in the middle of the posterior density because Ad Model Builder begins MCMC at the MLE parameter values.

Your first step should be to plot the total negative log likelihood as it evolves in the MCMC. There are two things you are seeking: First is the absolute value of the value. It should wonder around the MLE value (the lowest you will see) and about 2 times the number of parameters above the MLE. Thus if the MLE is at –30, and you have 10 parameters, the value of the likelihood (really posterior density) should range between –20 and –10 (2*10 units higher). Second, you should look at the evolution of the likelihood. It should look like a random walk, not like a directed walk. The average value in the final 500 saved points should be about the same as the first 500 saved points.

You should also check the evolution of other parameters. Again they should look like random walks, not one-way walks. The parameters shouldn't stay in one range for a long time and then walk back. The excursions from the average values should be brief. You can plot pairs of parameters against each other; this two-dimensional graph should look like a shotgun blast and not have clumps of points in the space.

There are many formal statistical tests for convergence (Gelman et al. 1995). There is also a software package, CODA, that performs many of these statistical tests. Check mean and variance of blocks of data. Check for autocorrelation in samples, ideally not autocorrelated.

Sensitivities to priors

- Average recruitment
- *q*'s
- Sample sizes for multinomial likelihoods.
- Profile current stock size by fixing q and look at likelihood components.
- Look at different sources of data one at a time.